

GENOMIC PROFILING OF ORAL PRECANCEROUS LESIONS AND EARLY STAGES OF ORAL CANCER

By

PRIYANKA GANPAT BHOSALE

[LIFE09200904017]

TATA MEMORIAL CENTRE

MUMBAI

A thesis submitted to

The Board of Studies in Life Sciences

In partial fulfilment of the requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE



February, 2017

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Ms. Priyanka Ganpat Bhosale** entitled "**Genomic Profiling of Oral Precancerous Lesions and Early Stages of Oral Cancer**" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.



13.02.2017

Chairman – **Dr. Pratibha Kadam Amare**

Date:



13/02/2017

Guide/Convener – **Dr. Manoj B. Mahimkar**

Date:



13/02/2017

Member-1 – **Dr. Tanuja Teni**

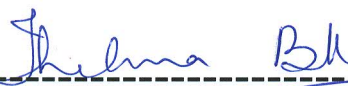
Date:



13/02/17

Member-2 – **Dr. Sanjeev Waghmare**

Date:



Feb. 13, 2017

External Examiner – **Prof. B. K. Thelma**


Date:

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 13 February 2017

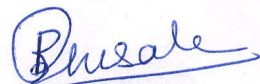
Place: Navi Mumbai


Dr. Manoj B. Mahimkar
(Guide)

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfilment of the requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under the rules of the HBNI.

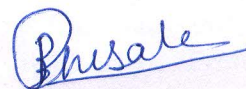
Brief quotations from this dissertation are allowed without special permission, provided that accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



Priyanka Ganpat Bhosale

DECLARATION

I, hereby declare that the investigation presented in the thesis titled "*Genomic Profiling of Oral Precancerous Lesions and Early Stages of Oral Cancer*" has been carried out by me under guidance of Dr. Manoj B. Mahimkar. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.



Priyanka Ganpat Bhosale

List of Publications arising from the thesis

Journal

1. “Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse”, Srikant Ambatipudi*, **Priyanka G. Bhosale***, Emma Heath, Manishkumar Pandey, Gaurav Kumar, Shubhada Kane, Asawari Patil, Girish B. Maru, Rajiv S. Desai, Fiona M. Watt, Manoj B. Mahimkar, PLoS One, **2013**; 8(7):e70688. (PMID: 23936238)
*Equal Contribution
2. “Low Prevalence of Transcriptionally Active HPV in Indian Head & Neck Squamous Cell Carcinoma and Leukoplakia”, **Priyanka G. Bhosale**, Manishkumar Pandey, Rajiv S. Desai, Asawari Patil, Shubhada Kane, Kumar Prabhash, Manoj B. Mahimkar, Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, **2016**; 122(5):609-618 (PMID: 27765330)
3. “Chromosomal Aberrations 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, Manoj B. Mahimkar, (**Manuscript under review**).
4. “Recurring Copy Number Gain at Chromosome 11q22 Plays an Important Role in Lymph Node Metastasis and Radioresistance in OSCC”. **Priyanka G. Bhosale**, Simona Cristea, Manishkumar Pandey, Mickey Shah, Asawari Patil, Alejandro A. Schäffer, Niko Beerenwinkel, Manoj B. Mahimkar (**Manuscript under preparation**)

Chapters in books

1. “Detection of HPV E6/E7 mRNA in clinical samples using RNA *in situ* Hybridization”, Manishkumar Pandey, **Priyanka G. Bhosale**, Manoj B. Mahimkar (Invited book chapter under review for Springer Protocols- Methods in Molecular Biology Laboratory Protocols Series).

Conferences

Conference Proceedings (Published abstracts)

- 1) “Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”, **Bhosale, P. G.**, M. Shah, A. Patil, R. S. Desai, S. Kane and M. B. Mahimkar. Journal of Carcinogenesis (**2015**) 14 (2): 21-38.
- 2) “Challenges of HPV detection in Head & Neck Squamous cell carcinoma: Redefining the algorithm”, Pandey, M., **P. G. Bhosale**, R. S. Desai, A. Patil, S. Kane, K. Prabhash and M. B. Mahimkar. European Journal of Cancer (2016) 54 (1).

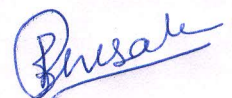
Conferences attended

Oral Presentations:

- Oral Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. At Carcinogenesis 2015, Molecular pathways to therapeutics: Paradigms and challenges in oncology, ACTREC, India.

Poster Presentations:

- Poster Presentation on topic **“Challenges of HPV detection in Head & Neck Squamous cell carcinoma: Redefining the algorithm”** at TMC Platinum Jubilee Celebrations, A Conference of new ideas in Cancer- Challenging dogmas 2016, Mumbai, India.
- Poster Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. EACR conference, Radiation Biology and cancer, From Molecular response to clinics, 2015, Essen, Germany.
- Poster Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. At 5th International Conference on Stem Cells and Cancer (ICSCC-2014): Proliferation, Differentiation, and Apoptosis, New Delhi, India.
- Poster Presentation entitled **“Differential expression of Keratin 76 in Oral Squamous Cell Carcinoma progression”**. 31st Annual convention of Indian Association for Cancer Research & International symposium on Cancer Genomics and its impact in the clinics” 2012, ACTREC, Navi Mumbai, India.



Priyanka Ganpat Bhosale

ACKNOWLEDGEMENTS

Towards the end of my PhD, I am happy to express my sincere gratitude to all those who made this journey possible and making this thesis an unforgettable experience of my life. First of all, I wish to thank my research guide Dr. Manoj B. Mahimkar, for his constant guidance and support. This work would have not been possible without his valuable suggestions and encouragement throughout my PhD tenure.

I would like to thank Prof. Shubhada Chiplunkar (Director, ACTREC), Dr. Rajiv Sarin (ex-Director, ACTREC) and Dr. Surekha Zingde (ex-Dy. Director, ACTREC) for providing an excellent research atmosphere and infrastructure at ACTREC. I thank my doctoral committee members Dr. Pratibha Kadam Amare (my present DC Chairperson), Dr. M Sheshadri (my ex-DC Chairperson), Dr. Tanuja Teni and Dr. Sanjeev Waghmare for giving their valuable suggestions and patient hearing during the meetings.

I thank Dr. Anita Borges (Department of Pathology, Raheja Hospital), Dr. Asawari Patil (Department of Pathology, Tata Memorial Hospital), Dr. Rajiv Kumar (Department of Pathology, Tata Memorial Hospital) for helping us with tumor percentage determination and immunohistochemical scoring. I would especially thank Dr. Borges and Dr. Patil for making me understand the basics of histopathological grading. I acknowledge the help received from our clinical collaborators Dr. Prathamesh Pai (Head and Neck Unit, Tata Memorial Hospital), Dr. Pankaj Chaturvedi (Head and Neck Unit, Tata Memorial Hospital), Dr Rajiv Desai (Department of Oral Pathology & Microbiology, Nair Hospital Dental College, Mumbai) and Dr. S.V. Kane (Department of Pathology, Tata Memorial Hospital). I acknowledge the constant help from the staff of the tumor tissue repository and Pathology Department at Tata Memorial Hospital and

ACTREC for providing leukoplakia and oral cancer specimens and I would like to compliment them for maintaining high standards of tissue collection and storage. I would extend my gratitude to all the members of Nair Dental Hospital, for helping me procure leukoplakia and control tissue samples.

I would like to sincerely acknowledge the Department of Atomic Energy for providing me financial support (PhD fellowships). I am thankful to the Council of Scientific and Industrial Research (CSIR); Department of Biotechnology (DBT); the Department of Science and Technology (DST) and The Terry Fox Foundation for funding my research project. I thank the Director, Tata Memorial Centre and Director, HBNI for providing me The Sam Mistry fund and HBNI fund to attend and present my work at EACR conference, Radiation Biology and cancer, From Molecular response to clinics, Essen, Germany.

Help from Dr. Simona Constantinescu (Computational biology group, ETH, Basel), Prof. Alejandro Schaffer (Computational biology branch, NIH) and Prof. Niko Beerenwinkel (Computational biology group, ETH, Basel) is deeply acknowledged. A special thanks to Simona for extensive analysis of the high throughput assays. Prof. Alejandro and Prof. Niko deserves a special mention for their valuable suggestions during the analysis. I thank Dr. Sussane Gollin (Professor, Univ. of Pittsburgh) for providing us oral cancer cell lines. I would like to thank Dr. Fiona Watt (CRI, Cambridge) and Dr. Girish B. Maru for providing us KRT76-knockout mice and hamster tissues. I thank Dr. Sorab N. Dalal for allowing me to do a few cloning related experiments in his lab.

I would like to thank all the members of the Common Instrumentation facility (Mr. U Dandekar), IT department (Mr. Kanvinde, Mr. Aanand Jadhav), the Microscopy facility (Vaishali, Tanuja

and Jayraj), the Animal house facility (Dr. A Ingle, Dr. Rahul Thorat), the Library facility and the Administrative Department of ACTREC. Mr. P. Chavan and Mr. V. Sakpal from Histology Section deserves a special mention for their constant help during my Ph.D. tenure.

The members of the Mahimkar lab (past and current lab members) have contributed immensely to my personal and professional time at ACTREC. I was fortunate to have a very helpful and supportive senior (Srikant), I can't thank you enough for teaching me lab skills, for all the fun we have had, and for the long discussions (starting from the hard core science, cooking, music, movies, trekking, photography and the list will go on and on). I would like to thank Mrs. Tanuja Samant for her kind help for all these years. I appreciate the help and support by RT (Mr. Ravindra Pawar) throughout my PhD (General lab upkeep, tissue culture room maintenance and sample collection). I very much appreciated his enthusiasm for engaging everyone in puzzles and debates and keeping the lab atmosphere cheerful. I also thank previous lab members Ravindra (Gowda), Tejas, Chintan and Manoj for being so helpful and keeping the lab atmosphere joyful. I am very glad to have amazing juniors at the lab, a special mention to Usha and Mayuri for providing a friendly and joyful atmosphere at work. I am also grateful for their help in numerous ways and accompanying me to have a stimulant (tea) which was a stress buster during the final stages of this Ph.D. I thank the trainees who worked with me- Tabreez, Lalit, Mickey, Shubhra, Avnish, Jayanthan, Akshay, Vishnu, Anjali, Sheel, Mansi and Wajid, I appreciate the help received from them. I have had the pleasure to work with past and present lab members they all have contributed to make Mahimkar lab a very lively place to work and live.

My time at ACTREC was made enjoyable in large part due to the many friends and my dinner group that became a part of my life. During my first year I met amazing friends at ACTREC, a special mention to Manish and Srikanta (Kanto) for being by my side in all good and bad times

and I will always cherish their company. My very special thanks to Manish who stood by me during tough times of my PhD, I am very glad to have such a fantastic friend with me. I am grateful for the time spent with roommates (Pallavi and Asmita) they have cherished with me every great moment and supported me whenever I needed it. I would like to thank Sonali, Abha, Sarika and all the 2nd floor PG Room members for keeping the jolly atmosphere, I have had a wonderful time with all of you. I thank the batch of 2009: Srikanta, Snehal, Rupa, Madhura, Pooja, Kedar, Mansa, Tanmoy, Shafqat, Abira, Rubina, Swati, Sushmita, Bhanu and Aparna, for their pleasant company and our memorable trips. I also had a great time with my seniors and juniors at ACTREC and would like to thank them all for being so helpful.

Lastly, I would like to thank my family for all their love and encouragement. A special thanks to my loving and encouraging Brother who supported me in every possible way. Last but not the least; I thank all the participants of the study without which this work would have been impossible.

❧ Dedicated to My Family and Friends :) ❧ ❧

Table of Contents

SYNOPSIS.....	14
LIST OF FIGURES.....	28
LIST OF TABLES.....	30
1. Introduction	32
1.1. Oral Cancers and Oral Potentially Malignant Disorders (OPMD's)	33
1.1.1. Epidemiology.....	33
1.1.2. Etiology	35
1.1.3. Predisposition to Malignant Transformation.....	37
1.1.4. Oral Squamous Cell Carcinoma (OSCC)	41
1.1.5. OSCC Diagnosis and Staging.....	42
1.1.6. OSCC Treatment and Prognosis	44
1.2. Molecular Pathogenesis of OSCC.....	47
1.3. Genomic Imbalances in OSCC	48
1.3.1. Copy Number Alterations (CNA)	49
1.3.2. The Biological Impact of CNA.....	50
1.4. Identification of CNA in OSCC	50
1.4.1. Array Comparative Genomic Hybridization (aCGH)	51
1.4.2. Integration of Genetic Events and Gene Expression Patterns	53
1.5. Microarray Profiling of Oral Potentially Malignant Disorder (OPL) and OSCC	54
2. Aims and Objectives	56
2.1 Study Hypothesis	57
2.2 Objectives	57
2.2.1 To identify the genomic alterations and transcriptomic changes in non-HPV associated oral precancerous lesions (OPL) and early stages of OSCC.	57
2.2.2 To validate results of aCGH and GE in OPL and OSCC.	57
3. Materials and Methods	58
3.1 Clinical Tissue Sample Collection	59
3.1.1 Leukoplakia and OSCC Tissue Sample Collection	59
3.1.2 Normal Oral Cavity Tissue and Whole Blood Collection	60
3.1.3 Sample Sets and Study Design	60
3.2 Histopathological Evaluation of Tissues	62

3.3	DNA Isolation from Whole Blood.....	62
3.4	Simultaneous DNA/RNA Isolation from Tissues	62
3.5	HPV Detection.....	63
3.5.1	p16 Immunohistochemistry (IHC)	63
3.5.2	Nested PCR for HPV Detection.....	64
3.5.3	RNA <i>In-Situ</i> Hybridization (RNA ISH).....	66
3.6	Microarray Experiments.....	67
3.6.1	Array CGH Labeling and Hybridization	67
3.6.2	Gene Expression Microarray (GE)	69
3.7	Real-Time PCR Analysis	71
3.7.1	cDNA Synthesis.....	71
3.7.2	Real-Time PCR Analysis.....	72
3.8	Fluorescent <i>In-Situ</i> Hybridization (FISH).....	73
3.8.1	Interphase FISH (I-FISH) using the BAC Clones.....	73
3.8.2	Interphase FISH Using SureFISH Probe	77
3.8.3	Metaphase FISH	79
3.8.4	Visualization and FISH Signal Enumeration.....	79
3.9	Immunohistochemistry (IHC).....	80
3.9.1	Pretreatment and Poly-L-Lysine Coating of Glass Slides	82
3.9.2	De-paraffinization of Tissue Sections	82
3.9.3	Antigen Retrieval.....	82
3.9.4	Immunohistochemistry.....	83
3.10	Immunofluorescence	84
3.10.1	Immunostaining of Tissue.....	84
3.10.2	Immunostaining of Cultured Cells.....	85
3.11	Cell Line, Plasmid Constructs and Transfections.....	86
3.11.1	Cell Line Maintenance	86
3.11.2	Plasmids and Constructs	86
3.11.3	Transfection	87
3.12	Western Blot Analysis	87
3.13	<i>In-vitro</i> Cell Based Assays	88
3.13.1	Puromycin Kill Curve.....	88

3.13.2	Scratch Wound Healing Assays	88
3.13.3	Matrigel Cell Invasion Assay.....	89
3.13.4	Cell Proliferation Assay	89
3.14	Statistical Analysis.....	90
3.14.1	Statistical Analysis of qRT-PCR Data	90
3.14.2	Statistical Analysis of IHC and FISH Data	90
3.14.3	Statistical Analysis of <i>in-vitro</i> Assays	91
4.	Results	92
4.1	The Clinicopathological and Demographic Characteristics	93
4.2	Histopathology of Leukoplakia and OSCC	93
4.3	DNA and RNA Extraction and Quality Control	95
4.4	HPV Detection.....	96
4.5	Genome Wide Analysis of Copy Number Alterations (CNA)	97
4.5.1	CNA Associated with Disease Progression.....	98
4.5.2	CNA Associated with Clinicopathological Parameters and Clinical Outcome	116
4.6	Gene Expression Profiling.....	121
4.7	Integrative Analysis of Gene Expression and Copy Number Alterations	132
4.8	Validation of Microarray Results	134
4.8.1	Validation of Differentially Expressed Genes Associated with OSCC Progression	134
4.8.2	Copy Number Alterations Associated with OSCC Progression.....	149
4.8.3	Copy Number Dependent Changes Associated with Clinical Outcome.....	151
5.	Discussion.....	169
5.1	A Conspectus of Genome-Wide Copy Number Alterations in Oral Cancers	170
5.1.2	Lymph Node Metastasis Associated CNA	172
5.1.3	CNA Associated with Risk Prediction.....	174
5.2	Genome-Wide Gene Expression and Integrative Analysis	176
5.2.1	Differential Expression of Genes in Oral Carcinogenesis.....	176
5.2.2	Integrative Analysis to Identify Drivers in Oral Carcinogenesis	178
5.2.3	Pathways Deregulated in OSCC.....	181
6.	Summary and Conclusions	184
7.	Appendix	187
7.1	Tables	188
7.1.1.	Clinicopathological Characteristics of Leukoplakia Patients.....	188

7.1.2.	Clinicopathological Characteristics of OSCC Patients	189
7.1.3.	Clinicopathological Characteristics of OSCC Patients from Previous Study	191
7.2	Recipes for Common Reagents	194
8.	References.....	208
9.	Publications	220



Homi Bhabha National Institute

SYNOPSIS OF Ph. D. THESIS

- | | |
|--|---|
| 1. Name of the Student: | Priyanka Bhosale |
| 2. Name of the Constituent Institution: | Tata Memorial Centre, Advanced Centre for
Treatment Research and Education in Cancer |
| 3. Enrolment No. : | LIFE09200904017 |
| 4. Title of the Thesis: | Genomic Profiling of Oral Precancerous Lesions and
Early Stages of Oral Cancer. |
| 5. Board of study: | Life Science |

SYNOPSIS

Introduction

Oral Squamous Cell Carcinoma (OSCC) is the sixth most prevalent malignancy worldwide, and the third most common cancer in developing nations, and the leading cause of cancer associated mortality in India [1]. Majority of OSCC are etiologically associated with abuse of smokeless/smoked tobacco in combination with alcohol consumption; additionally, there exist a group of no-habit patients that may potentially harbor high-risk HPV infection [2-5]. The development of squamous cell carcinoma is believed to be a multistep process, in which several mutations within one cell, accumulate to develop a tumor through a series of histopathological stages from hyperplasia, through various extent of dysplastic changes, carcinoma in situ (CIS), and finally breaking through the basement membrane at the invasive OSCC stage [6]. One of the key features in the pathogenesis of many solid cancers, especially head and neck cancer, is chromosomal instability, with gene gain or loss reflecting this genetic instability. Gains and losses of genomic regions may contain proto-oncogenes and tumor suppressor genes, which may lead to aberrant expression useful for malignant transformation. Low-level copy number changes involving large regions with many genes have been frequently observed in OSCC, but their effect on gene expression remains ambiguous [7, 8]. Focal high-level copy number change (amplification and deletion), are likely to represent alterations continuously under selection for tumor growth [9]. The advent of genome-wide screening methods such as comparative genomic hybridization (CGH), and, more recently, array CGH (aCGH), have opened up new possibilities to catalogue copy number alterations at high resolution, additionally, transcriptomic profiling would highlight the copy number dependent and independent deregulation of genes in the altered locus [10, 11]. Not much data for the early events of genetic degeneration, namely the transition

from normal mucosa to dysplasia and carcinoma in situ are available [12, 13]. The subjective nature of dysplasia assessment and uncertainty over the relative importance of the individual histological features assessed are limiting factors in providing an accurate prognosis [24]. Identifying the key chromosomal alterations/ genes that play an important role in the progression of the carcinogenesis process may have potential clinical implications .

Along with OSCC progression, metastasis to lymph node is the most important predictor of patient survival. Predicting the risk of cervical lymph node metastasis would greatly help in patient stratification, it would either avoid the potential morbidity of overtreatment or prevent further progression of disease [14]. Understanding the genetic signatures as well as genome wide expression profile would help identify the node metastasis verses nonmetastatic OSCC. Broadly, objective of this study was to identify the genomic and transcriptomic signatures associated with progression from low risk lesions to high risk invasive OSCC and prediction of patient's clinical outcome.

Objective

1. To identify the genomic alterations and transcriptomic changes in non-HPV associated oral precancerous lesions (OPL) and early stages of OSCC.
2. To validate results of aCGH and GE in OPL and OSCC, using appropriate validation methods.

Results and Discussion

Results-Objective 1: To identify the genomic alterations and transcriptomic changes in non-HPV associated oral precancerous lesions (OPL) and early stages of OSCC.

Paraffin blocks and frozen tissue samples of OPL (leukoplakia n=60), and neo-primary gingivobuccal OSCC tissues (n= 200), were recruited from the Nair Hospital Dental College and

Tata Memorial Hospital respectively. Whole blood (n=28) and gingivobuccal tissues (n=5) were obtained from healthy individuals that served as control for aCGH and GE analysis. Written informed consent was obtained from all the study participants. Patients received neither radiation nor chemotherapy before surgery. Tissues with more than 60 % tumor content and leukoplakia samples with dysplastic or hyperplastic epithelium, as graded by two independent pathologists were subjected to simultaneous DNA-RNA extraction.

Analysis of Human Papilloma Virus (HPV) status:

HPV-positive tumors show a distinct molecular-genetic signature and have a better clinical response than HPV-negative counterpart [15, 16]. Even though, the proportion of HPV-related tumors are/is higher in the oropharynx, we have screened leukoplakia and gingivobuccal OSCC patients for the presence of HPV using multiple methods. 1) Detection of p16 overexpression (surrogate marker for the presence of HPV), 2) Nested DNA PCR methods using MY09/MY11 and GP5+/GP6+ primers and 3) RNA *In-Situ* Hybridization for detection of 7 HR-HPV E6/E7 mRNA. All the samples were confirmed to be HPV negative, and were used for further analysis.

Array CGH and Gene expression profiling:

The HPV negative tumor samples were used for genomic and transcriptomic profiling. The aCGH analysis was carried out on early stage OSCC [T1N0, T2N0] (n=28) and advanced stage OSCC [T1N+, T2N+] (n=10) and Leukoplakia (n=24) using oligonucleotide 105K arrays (Agilent Technologies), pooled normal Male/Female DNA was used as reference. Agilent array platforms constitute of 60mer oligonucleotides arrays, which offer genome wide profiling with average probe spacing ~15kb.

RNA samples with RIN value more than 8, and obtained simultaneously from above early stage OSCC (n=24), advanced stage OSCC and (n=10) and Leukoplakia (n=15), were used for gene

expression profiling using Agilent 44K arrays. RNA extracted from the buccal epithelium of healthy control (n=5) was used as reference for microarray analysis.

aCGH data analysis: aCGH analysis was carried out using GISTIC 2.1 tool. The comparison was carried out using 24 leukoplakia, 28 samples early stage, 10 cases of advanced stage node positive (T1N+, T2N+) and 54 advanced stage gingivobuccal cancers (T1N+, T2N+, T3N0, T4N0, T3N+ and T4N+) reported by us earlier (Ambatipudi et al. 2011). Array CGH analysis revealed multiple aberrations across the genome in early stage OSCC and OPL. These alterations are summarized in Table 1, depicting alterations that are common in all groups and could be associated with OSCC progression; common and unique alterations associated with node metastasis and with patients disease specific survival (DSS) and recurrence free survival (RFS).

Table 1: Common aberrations in Leukoplakia, Early and Advanced stage OSCC

Common Gain/ Amplification	
Leukoplakia (n= 24)	4q13.2, 6p21.32, 14q11.2, 8p11.22, 16p11.2, 6p25.3, 8q24.3, 8p23.2, 15q11.1,
Early stage OSCC (n=28)	1p36.33 11q13.3, 6p21.32, 11q13.1, 13q21.33, 8q24.3, 4q13.2, 7q11.23, 5p15.33, 14q32.33, 22q11.21, 1q44, 7q22.1, 19p13.3, 22q11.1, 11p15.5, 9q34.3, 7q11.21, Xq26.3, 1p36.33, 16p11.2
T1N+,T2N+ (n=10), Advanced stage OSCC (n=54) Ambatipudi et al. 2011	1p36.33, 11q13.3, 7p11.2, 6p21.32, 11q13.1, 14q32.33, 7p11.2, 17q25.3, 15q11.1, 11q22.1, 9p24.1, 14q11.2, 3q27.2, 8q24.21, 8q24.3, 8p23.1, 16p13.3, 3q26.31, 7p22.3, 22q11.23
All N0 (n=54)	11q13.3, 6p21.32, 19p13.3, 8q24.3, 7p11.2, 17q25.3, 5p15.33, 7q11.23, 22q11.21, 14q32.33, 11q13.1, 9p24.1, 14q11.2, 15q11.1, 22q11.23, 1p36.33, 7q22.1, 3q27.2, 13q21.33, 11p15.5, 4q13.2, Xq28, 1q44, 9q34.3, 17p13.1, 11p11.2, Xq22.2, 16p13.3, 16q21
All N+ (n= 37)	8p23.1, 8q24.21, 11q22.1 11q13.3, 6p21.32, 5p15.33, 7p11.2, 7q11.23, 15q11.1, 11q13.1, 9p24.1, 19p13.3, 4q13.2, 8q24.3, 14q32.33, 17q25.3, 3q27.2
DSS and RFS	1p36.33, 1q23.2, 3q27.2 , 11q13.3, 16p13.3, 16p11.2, 16q12.2
Common Loss/ Deletions	
Leukoplakia (n= 24)	14q11.2, 15q11.1, 16p11.2, 19p13.3, 8p11.22, 1q44, 8p23.1, 2q21.2, 17q21.31

Early stage OSCC (n=28)	8p11.22, 15q11.1, 8p23.2, 14q11.2, 4q35.2 , 8p23.1, 3p14.2 , 18q21.33, 9p21.3, 2q22.1, 4q13.2, 7q35, 10q11.22
T1N+,T2N+(n=10), Advanced stage OSCC (n=54) Ambatipudi et al. 2011	8p11.22, 8p23.2, 3p14.2 , 8p23.1, 10p11.21, 4q35.2 , 14q11.2, 19p13.3, 9p23, 10p11.21, 3p11.1, 8p22, 17q21.31, 1q44, 18q23, 2q21.2, 9p12, Xq21.31, 2q34, 11q22.3 , 1q31.3, 16p11.2, 17p13.1 , 19p13.11, 3p26.3, 16q23.3, 21q11.2, 7q31.1, 5p14.3, Yp11.2, 3p22.3, 18q12.1, 3p21.1, 2q23.3, 4q21.3 , 11p11.12, 5q12.1, 14q21.1, 13q21.32, 20p12.1
All N0 (n=54)	8p11.22, 15q11.1, 8p23.2, 14q11.2, 8p23.1, 4q35.2, 3p14.2 , 4q13.2 , 9p21.3, 3p11.1, 1q44, 4q13.2 , Xp21.3, 7q35, 8p22, 18q23, 1q31.3, 2q21.2, 9p23, 21q21.3, 2q22.1, Yp11.2, 16p11.2, 17q21.31, 3p26.1, 18q21.33, 2q34, 13q21.32, 3p21.1, 19p13.3, 4q22.1 , Xq21.31, 3p24.1, 19p12, Xq22.3, 10p11.21, 2q21.2, 5p14.3
All N+ (n=37)	8p11.22 , 3p14.2 , 8p23.2, 2q22.1, 8p23.2, 19p13.3, 14q11.2, 15q11.1, 9p23, 4q35.2, 8p23.1, 9p21.3, 10p11.21, 17q21.31, 2q21.2, 9p12, 3p26.3, 8p22, 18q22.1, 2q23.3, 3p12.3, 11q22.3 , 7q31.1, 18q12.1, 8p23.2, Xp21.3, 4q21.3, Xq21.31, 16p11.2, 2q34, 18q12.3, 6q21, 19q13.33
DSS and RFS	2p11.2, 2q22.1, 2q34, 3p14.2 , 4q13.2 , 4q22. , 9p23 11q22.3

aCGH analysis showed overall changes across the entire genome, most of the amplifications / deletions observed in OPL were present in early / advanced stage OSCC. Even there was a notable change in OPL, early verses advanced stage OSCC; early stage tumors showed 50% of the CNA as compared to that of advanced stage tumors. In comparison between N0 versus N+ tumors, few unique alteration were observed that could possibly be associated with lymph node metastasis.

Gene expression analysis: Bioconductor R (limma) package was used to identify differentially expressed genes in OPL, early stage OSCC and advanced stage OSCC when compared to Normal (N). Short discussion of results is given below:

1) OPL vs Normal: p value of 0.01.	
Top Upregulated Total 395 and 22 with more than 2 Fold change	<i>MFAP5, CA2, HOXC9, NELL2, SERPINB7, LRP12, PTPRZ1, EGR2, NCAPG, MYO5A, CDK6, ATP2A2, CDCA2, LPP, HSP90AA1, HDAC4, GLI2, DPYSL2, DKC1, CES2, ECT2, MTAP, MRPL1, WNT3, HOXC13</i>

Top Downregulated Total 454 and 69 with less than -2 Fold change	<i>KRT19, DERL3, PSCA, KRT4, CLDN7, CXCL1, MAL, ELF3, CEACAM1, CLCA4, GPX3, COL1A1, ECM1, PPP1R3C, DNASE1L3, ENDOU, IL1RN, MUC1, KLK13, GPRC5A, MMP9, CEACAM7, LAMC2, SH3GL3, CTSB, TCL1A, CD38, COL1A2, DCTN3, PACSIN3, ARFGAP2, ERBB3, PCDH1, LLGL2, ARSA, PTPRF</i>
2) Early stage OSCC vs. Normal : p value of 0.01.	
Top Upregulated Total 805 and 107 with more than 2 Fold change	<i>MMP10, CXCL10, CXCL11, FST, BATF2, SH2D5, GBP5, NELL2, INHBA, CXCL9, IDO1, MMP3, SPOCD1, WARS, IFI44L, IFI6, MFAP5, SERPINE1, OASL, HOXC9, LY6K, HOXC13</i>
Top Downregulated Total 1008 and 213 with less than -2 Fold change	<i>CRNN, MAL, LOR, KRT19, FAM3B, KRT4, ALOX12, ENDOU, KRT13, FMO2, DAPL1, CD177, SPINK7, PPP1R3C, TMPRSS11A, GPX3, DNASE1L3, HPGD, SH3GL, HOPX, DERL3, CLCA4, PTGDS, SILV, KRT3, MFAP4, KLK12, COMP, KRT76 (Advanced OSCC vs Normal)</i>

3. **Advanced stage OSCC vs Normal:** Total upregulated 798 and 108 with more than 2 Fold change and Total downregulated 1126 and 281 with less than -2 Fold change.

4. **Total OSCC vs Normal:** Total upregulated 785 and 97 with more than 2 Fold change and Total downregulated 1056 and 245 with less than -2 Fold change.

5. All Node negative Vs All Node positive: No differentially expressed genes found in these comparisons.

Concluding remark: Gene expression analysis identifies a large number of genes differentially expressed between two or more groups. Combined aCGH and GE analysis showed overall changes across the entire genome, some alteration covered few or no genes while others spanned cluster of genes. We found significant copy number dependent / independent over and under expression of genes not reported earlier in oral cancer, these genes warrant their enlistment in the oral cancer related genes.

Results-Objective 2: To validate results of aCGH and GE in OPL and OSCC, using appropriate validation methods. Validation was performed on the test set as well as independent cohort. Florescent In situ Hybridization (FISH) was used to validate altered

chromosomal loci, and TaqMan based qRT-PCR assay was used to validate differentially expressed genes, *18S RNA* was used as an endogenous control for qRT-PCR comparisons and *SLC4A1AP*, an unaltered gene in our analysis, was used as the reference gene. For few targets immunohistochemistry (IHC) was performed to validate the concurrent protein expression. Validation was carried out in OPL, early stage OSCC and advanced stage OSCC.

Genomic loci and target genes for validation were selected based on following criteria:

- Copy number alterations and deregulated genes associated with disease progression
- Copy number independent changes associated with disease progression
- Alterations associated with clinical outcome

Validation of copy number alterations and deregulated genes associated with disease progression: Gain of 8q is one of the most frequent event in oral cancers [7]. We observed gain of region 8q24.3 to be associated with OSCC progression. Additionally, overexpression of *LY6K* was observed by integrative GE analysis in all stages of OSCC (no change in OPLs). In order to validate the findings of 8q24.3 gain, we carried out FISH analysis on interphase nuclei on paraffin embedded OPL, early stage OSCC and advanced stage OSCC tissues, and *LY6K* overexpression by qRT-PCR. The FISH and qRT-PCR analysis validated the findings of array CGH in OPL and OSCC. Additionally copy number dependent overexpression of *EIF5A2* and *ECT2* (3q26.22); *HOXC9* and *HOXC13* (12q13.3); and *FUS* (16p11.2) was observed in either OPL or early stage OSCC or both; validation of *EIF5A2*, *HOXC9* and *FUS* was done by qRT-PCR.

Validation of copy number independent changes associated with disease progression: Gene expression analysis revealed significant copy number independent overexpression of *MFAP5*, *NELL2*, *INHBA* and downregulation of *KRT76*. Overexpression of *MFAP5* and *NELL2* was

observed in OPL as well as in OSCC. *MFAP5* (microfibrillar associated protein 5) promotes tumor cell survival and stimulates endothelial cell motility and survival via the $\alpha V\beta 3$ integrin receptor. While *NELL2* (neural EGFL like 2), plays important role in the protection of cells from cell death-inducing environments. We have validated *MFAP5* and *NELL2* overexpression by qRT-PCR and IHC, while *INHBA* gene upregulation was confirmed by qRT-PCR in an independent cohort.

In addition to the above targets, copy number independent downregulation of *KRT76* was observed in OSCC. *KRT76*, a type II epithelial keratin (K76), is specifically expressed in the suprabasal cell layers of oral masticatory epithelium lining the gingiva and the hard palate. We validated our microarray finding in an independent cohort of gingivobuccal cancers using qRT-PCR and IHC, further extending our analysis to *in-vivo* model systems: the hamster model was used to demonstrate, K76 downregulation during sequential progression of oral cancer, and the Knockout mice model, to evaluate the effect of *KRT76* loss. The *KRT76*-KO mice data underlines the potential of K76 being an early event, although its loss is not sufficient to drive cancer progression.

Alterations associated with clinical outcome: Multiple aberrations at 11q locus were observed in OSCC samples, including amplification of 11q13.3, 11q22.1 and deletion of 11q22.3. When compared between node metastatic and nonmetastatic tumors, 11q22.1 amplification was strongly associated with node metastasis, same locus was also associated with shorter survival (Ambatipudi et al 2011). Amplification of 11q22 is evident in glioblastomas, OSCC, and in cancers of the pancreas, lung, ovary, and cervix [17-19]. The human amplicon include cluster of matrix metalloproteinase (*MMP*) genes, two members of the BIRC family (*BIRC2* and *BIRC3*, protein known as cIAP1 and cIAP2 respectively), [4, 8, 20-22]. The expression of cIAP1 is

correlated to invasion of esophageal cancers and cervical cancers, and supports that if *cIAP1* is overexpressed in some cervical cancers, it may be associated with an unfavorable prognosis after radiotherapy and/or chemotherapy [22, 23]. We validated 11q22.1 amplification in OSCC using FISH, while overexpression of genes (*BIRC2* and *BIRC3*) located on this amplicon was confirmed by qRT-PCR and concurrent IHC analysis revealed overexpression and differential localization of cIAP1 and cIAP2 in OSCC. Overall analysis demonstrated strong correlation between 11q22.1 amplification and node metastasis.

Summary and conclusion

Array CGH analysis revealed chromosomal alterations associated with OSCC progression, lymph node metastasis and patient survival in HPV negative gingivobuccal cancers. Additionally, gene expression profiling identified several deregulated copy number dependent and independent genes. Representative targets (chromosomal loci and genes) were successfully validated in independent cohort. Most of the genes upregulated were either signalling molecules, or associated with various biological process such as angiogenesis, tissue morphogenesis, cell proliferation and apoptosis regulator. Genes found downregulated included predominantly the ones which are responsible for structural integrity and metabolism in epithelial cells. This is the first report which identifies the potential genetic biomarkers associated with transformation of OPL to invasive OSCC as well as node metastatic OSCC. The thesis will present in detail the relevant supporting literature, material and methods used, results and discussions.

References

1. Dikshit, R., et al., *Cancer mortality in India: a nationally representative survey*. Lancet, 2012. **379**(9828): p. 1807-16.
2. Chen, Y.J., et al., *Genome-wide profiling of oral squamous cell carcinoma*. J Pathol, 2004. **204**(3): p. 326-32.

3. Pentenero, M., S. Gandolfo, and M. Carrozzo, *Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: a review of the literature*. Head Neck, 2005. **27**(12): p. 1080-91.
4. Forastiere, A., et al., *Head and neck cancer*. The New England journal of medicine, 2001. **345**(26): p. 1890-900.
5. Mork, J., et al., *Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck*. The New England journal of medicine, 2001. **344**(15): p. 1125-31.
6. Warnakulasuriya, S., et al., *Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement*. J Oral Pathol Med, 2008. **37**(3): p. 127-33.
7. Baldwin, C., et al., *Multiple microalterations detected at high frequency in oral cancer*. Cancer Res, 2005. **65**(17): p. 7561-7.
8. Smeets, S.J., et al., *Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus*. Oncogene, 2006. **25**(17): p. 2558-64.
9. O'Regan, E.M., et al., *Distinct array comparative genomic hybridization profiles in oral squamous cell carcinoma occurring in young patients*. Head Neck, 2006. **28**(4): p. 330-8.
10. Ambatipudi, S., et al., *Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome*. PLoS One, 2011. **6**(2): p. e17250.
11. Agrawal, N., et al., *Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1*. Science, 2011. **333**(6046): p. 1154-7.
12. Brieger, J., et al., *Chromosomal aberrations in premalignant and malignant squamous epithelium*. Cancer Genet Cytogenet, 2003. **144**(2): p. 148-55.
13. Salahshourifar, I., et al., *Genomic DNA copy number alterations from precursor oral lesions to oral squamous cell carcinoma*. Oral Oncol, 2014. **50**(5): p. 404-12.
14. Mendez, E., et al., *Can a metastatic gene expression profile outperform tumor size as a predictor of occult lymph node metastasis in oral cancer patients?* Clin Cancer Res, 2011. **17**(8): p. 2466-73.
15. Braakhuis, B.J., et al., *Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus*. J Natl Cancer Inst, 2004. **96**(13): p. 998-1006.
16. Klussmann, J.P., et al., *Genetic signatures of HPV-related and unrelated oropharyngeal carcinoma and their prognostic implications*. Clin Cancer Res, 2009. **15**(5): p. 1779-86.
17. Baldwin, C., et al., *Multiple microalterations detected at high frequency in oral cancer*. Cancer research, 2005. **65**(17): p. 7561-7.
18. Bashyam, M.D., et al., *Array-based comparative genomic hybridization identifies localized DNA amplifications and homozygous deletions in pancreatic cancer*. Neoplasia, 2005. **7**(6): p. 556-62.
19. Gollin, S.M., *Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease*. Head Neck, 2001. **23**(3): p. 238-53.
20. Fernandez, L.A., et al., *Oncogenic YAP promotes radioresistance and genomic instability in medulloblastoma through IGF2-mediated Akt activation*. Oncogene, 2012. **31**(15): p. 1923-37.
21. Ginos, M.A., et al., *Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck*. Cancer Res, 2004. **64**(1): p. 55-63.
22. Imoto, I., et al., *Identification of cIAP1 as a candidate target gene within an amplicon at 11q22 in esophageal squamous cell carcinomas*. Cancer research, 2001. **61**(18): p. 6629-34.
23. Imoto, I., et al., *Expression of cIAP1, a target for 11q22 amplification, correlates with resistance of cervical cancers to radiotherapy*. Cancer research, 2002. **62**(17): p. 4860-6.
24. Hunter, K.D., E.K. Parkinson, and P.R. Harrison, *Profiling early head and neck cancer*. Nat Rev Cancer, 2005. **5**(2): p. 127-35.

Publications in Refereed Journal:

a. Published

Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse. Ambatipudi S*, Bhosale PG*, Heath E, Pandey M, Kumar G, Kane S, Patil A, Maru GB, Desai RS, Watt FM, Mahimkar MB. (PLoS One. 2013 Jul PMID: 23936238). ***Equal contribution**

b. Communicated and currently Under revision

Low Prevalence of Transcriptionally Active HPV in Indian Head & Neck Squamous Cell Carcinoma and Leukoplakia. Priyanka G. Bhosale, Manishkumar Pandey, Rajiv S. Desai, Asawari Patil, Shubhada Kane, Kumar Prabhash, Manoj B. Mahimkar. (Second Revision in Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology).

C. Manuscript under preparation :

- 1) Genome-Wide Expression and Copy Number Analysis Identifies the genetic progression model in Gingivobuccal Cancers.
- 2) Recurring copy number gain at chromosome 11q22 plays an important role in lymph node metastasis and radioresistance in OSCC.

Conferences attended:**Oral Presentations:**


- Oral Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. At Carcinogenesis 2015, Molecular pathways to therapeutics: Paradigms and challenges in oncology, ACTREC, India.

Poster Presentations:

- Poster Presentation on topic **“Challenges of HPV detection in Head & Neck Squamous cell carcinoma: Redefining the algorithm”** at TMC Platinum Jubilee Celebrations, A Conference of new ideas in Cancer- Challenging dogmas 2016, Mumbai, India.
- Poster Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. EACR conference, Radiation Biology and cancer, From Molecular response to clinics, Essen, Germany.
- Poster Presentation entitled **“Validation of altered expression of genes on 11q22 amplicon and its association with clinical outcome in OSCC patients”**. At 5th International Conference on Stem Cells and Cancer (ICSCC-2014): Proliferation, Differentiation, and Apoptosis, New Delhi, India.
- Poster Presentation entitled **“Differential expression of Keratin 76 in Oral Squamous Cell Carcinoma progression”**. 31st Annual convention of Indian Association for Cancer

Research & International symposium on Cancer Genomics and its impact in the clinics”

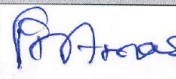
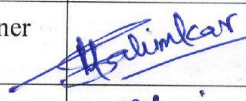


ACTREC, Navi Mumbai, India.



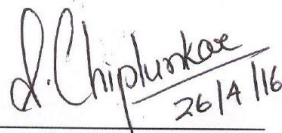
Priyanka Bhosale

Date: 25/04/2016

Doctoral Committee:

Sr. No.	Name	Designation	Signature	Date
1	Dr. Pratibha Kadam Amare	Chairman		25.4.2016
2	Dr. Manoj Mahimkar	Guide & Convener		25/4/2016
3	Dr. Tanuja Teni	Member		25/4/2016
4	Dr. Sanjeev Waghmare	Member		25/4/16

Forwarded through:



Dr. S.V. Chiplunkar,

Director,

ACTREC-TMC.

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research &
Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai 410210.



Dr. K. S. Sharma,

Director Academics,

PROF. K. S. SHARMA
DIRECTOR (ACADEMICS)
TATA MEMORIAL CENTRE,
PAREL, MUMBAI

List of Figures

Figure 1: Anatomy of the oral cavity.	33
Figure 2: Lip and Oral Cavity Related Incidence and Mortality	34
Figure 3: Lip and Oral Cavity Related Incidence and Mortality in Indian Population	35
Figure 4: Clinical and Histopathological Features of Leukoplakia	39
Figure 5: Treatment and Prognosis Overview of Oral Cavity Cancer	46
Figure 6: Tumor Field Cancerization: Visualization and Identification Methods	47
Figure 7: OSCC Disease Progression Model.	48
Figure 8: Genomic Imbalances in Cancers.	49
Figure 9: Overview of aCGH and Gene Expression Microarray Analysis Used in the Current Study.....	52
Figure 10: Venn Diagrams Representing Number of Samples Used for aCGH, GE and Validation Analysis.	61
Figure 11: Histopathology of Leukoplakia and OSCC Samples	93
Figure 12: A Representative Gel Picture of Genomic DNA Extracted from Study Samples.	95
Figure 13: Bioanalyzer Profile (electropherogram) of a Representative RNA Sample Depicting RNA Integrity.	95
Figure 14: Algorithm for HPV Detection in Patients with Leukoplakia and OSCC.	96
Figure 15: A Representative Gel Image of Nested GP5+/GP6+ PCR Results.....	96
Figure 16: p16 Overexpression in Study Samples.	96
Figure 17: Detection of High-Risk HPV RNA in Study Samples	97
Figure 18: The Amplification Plots And Venn Diagram Depicting Copy Number Gain.....	102
Figure 19: The Deletion Plots And Venn Diagram Depicting Copy Number Loss.	103
Figure 20: The Amplification/Deletion Plots And Venn Diagrams Depicting The CNA Associated With Lymph Node Metastasis	118
Figure 21: Kaplan-Meier Estimates of Patients Survival with Alterations on Chromosome 1p36.33 and 11q22.1	119
Figure 22: Overview of Gene Expression Changes in leukoplakia and OSCC.....	122
Figure 23: Heatmap of the top 20 differentially expressed genes in leukoplakia and OSCC	123
Figure 24: Biological Processes Altered in Leukoplakia and OSCC.....	126
Figure 25: Relative Quantitation (RQ) of Validation Targets	136
Figure 26: Log 2 Fold Change of Target Genes in Normal, Leukoplakia and OSCC	138
Figure 27: Immunohistochemical Analysis for Target Validation.	140
Figure 28: Correlation Between Protein Overexpression and Oral Cancer Progression	142
Figure 29: Validation of KRT76 Downregulation using Multiple Methods.	144
Figure 30: Correlation Between Loss of K76 Expression with Oral Cancer Development and Patient Survival	146
Figure 31: Sequential Downregulation of K76 Expression During Tumor Development in Hamster Buccal Epithelium.....	147

Figure 32. KRT76-KO Mice show Hyperplastic Lesions in Oral Epithelium	148
Figure 33: Interphase FISH for Validating Gain of 8q24.3 and 1p36.33.....	149
Figure 34: Correlation Between 8q24.3 Alteration and Oral Cancer Progression	150
Figure 35: Correlation Between 1p36.33 Gain with Clinical Outcome in Oral Cancer Patients	152
Figure 36: Interphase FISH for Detecting Gain of 11q13.3 and 11q22.2.....	154
Figure 37: Kaplan-Meier Estimates of Patients Survival with Alterations on Chromosome 11q22.2.....	157
Figure 38: Expression of MMP3 in Oral Cancers	159
Figure 39: Correlation of BIRC2 and BIRC3 Expression with Disease Advancement and Lymph Node Metastasis	160
Figure 40: Kaplan-Meier Estimates of Patient Survival with BIRC2 and BIRC3 overexpression	161
Figure 41: Immunohistochemical Analysis of cIAP1 and cIAP2 Expression in Normal Buccal Mucosa, Leukoplakia and Oral Cancers	163
Figure 42: Differential Subcellular Localization of cIAP1 and cIAP2.....	163
Figure 43: Correlation Between cIAP1 and cIAP2 Overexpression with Oral Cancer Development.....	164
Figure 44: Expression of cIAP1 and cIAP2 in SCC29B and Plasmid Backbone.	165
Figure 45: Affirmation of cIAP2 Knockdown by qRT-PCR and Western Blotting.....	166
Figure 46: cIAP2 Loss Leads to a Decrease in Cell Migration.	166
Figure 47. cIAP2 Loss Leads to a Decrease in Cell Invasion Without a Change in Cell Proliferation	167
Figure 48: Recurrent Amplifications and Deletions Among 12 aCGH Studies on Primary OSCC Tumors	171
Figure 49: Summary of the Genomic and Transcriptomic Signatures Associated with Gingivobuccal Cancer Progression from Pre-invasive Lesions	183

List of Tables

Table 1: Malignant Transformation of Leukoplakia.....	38
Table 2: Classification and Staging System for Oral Leukoplakias	39
Table 3: Staging of Oral Cavity Tumors	44
Table 4: A) Comparisons between different CGH methods B) Microarray platforms available with their technical details.	52
Table 5: List of 12 Studies that have Conducted Genome Wide Profiling of OSCC.....	55
Table 6: Primer sequences	66
Table 7: Taqman Assay IDs of Genes Validated by Real Time PCR.....	73
Table 8: SureFISH Probe Hybridization Mixture.	78
Table 9: Details of Antibodies Used in the Study	80
Table 10: The M10 media formulation.	86
Table 11: cIAP1 and cIAP2 shRNA.	87
Table 12: The clinicopathological characteristic of patients used for aCGH and GE study.	94
Table 13: Summary of Common and Unique Copy Number Gain/ Amplifications.	100
Table 14: Summary of Common and Unique Copy Number Loss/ Deletion.	100
Table 15: Gain and Amplification Observed in Leukoplakia.....	104
Table 16: Gain and Amplification Observed in OSCC.....	105
Table 17: Loss and Deletion Observed in Leukoplakia.	110
Table 18: Loss and Deletion Observed in OSCC	111
Table 19: Univariate Cox Proportional Hazards Regression Analysis of Single Predictors for Recurrence-Free and Disease Specific Survival.	120
Table 20: Top 50 Common Genes Up Regulated in Leukoplakia and OSCC.	127
Table 21: Top 50 Genes Down Regulated in Leukoplakia and OSCC.....	129
Table 22: Representative Biological Processes Significantly Altered in Leukoplakia.....	131
Table 23: Representative Biological Processes Significantly Altered in OSCC.....	131
Table 24: Copy Number Alterations and Direct Linkage Between Gene Expression Changes..	133
Table 25: Correlation of qRT-PCR validation targets with clinicopathological parameters.	136
Table 26: Demographic and Clinicopathological Characteristic of the Validation Study Group.	139
Table 27: Demographic and Clinicopathological Characteristic of the <i>KRT76</i> Study Group. ...	143
Table 28: The Effect of K76 Expression Loss with Development of Oral Lesions.	146
Table 29: The Demographic Details of 11q Validation Study Group.	153
Table 30: FISH Signal Enumeration and the Percentage of 11q13.3 and 11q22.2 Alteration in Validation Study Group.....	154
Table 31: Association of 11q13.3 and 11q22.2 Alteration with Clinicopathological Parameters	155
Table 32: Correlation of 11q13.3 and 11q22.2 Alterations with Lymph Node Metastasis.	156
Table 33: Association of 11q22 Amplification with Patients DSS and RFS.	157

Table 34: Multivariate Cox Regression Analysis to Identify the Association of Multiple Parameter on Clinical Outcome.	158
Table 35: Correlation of cIAP1 and cIAP2 Expression with Clinicopathological Parameters. ..	164
Table 36: The Effect of cIAP1 and cIAP2 Overexpression with OSCC Development.....	164

1. Introduction

1.1. Oral Cancers and Oral Potentially Malignant Disorders (OPMD's)

Oral cancer is a most common malignancy of the head and neck region. Based on the epidemiological and clinicopathological perspective, oral cancer is divided into following three categories: (1) Carcinoma of oral cavity proper, which includes the upper and lower gingiva, Buccal Mucosa (BM), Lip, Tongue, Retromolar Trigon (RMT), Alveolus, Floor of Mouth (FOM) and Hard palate, as represented in Figure 1,

(2) Carcinoma of lip vermilion, and

(3) Carcinoma arising in the oropharynx.

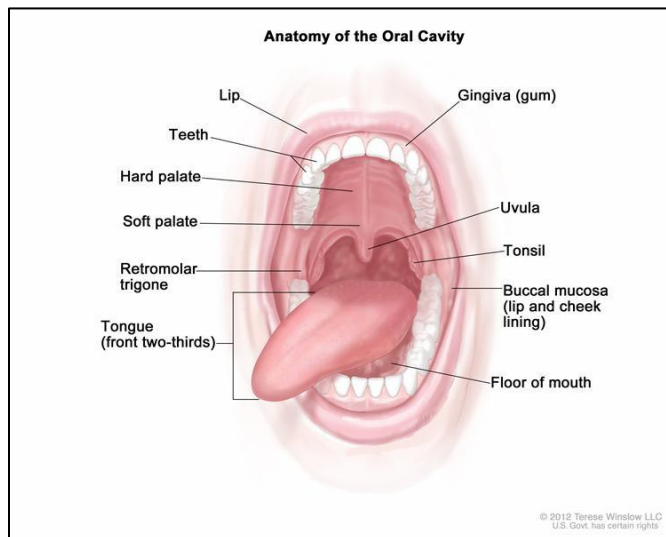


Figure 1: Anatomy of the oral cavity.

1.1.1. Epidemiology

Oral cancer mostly occurs in middle aged and older individuals (> 60 years), with the estimated incidence of around 198975 cases annually (GLOBOCAN 2012, IARC). Lip and oral cavity cancers are the leading cause of cancer related mortality worldwide (Figure 2A). India leads in terms of oral cavity cancer incidence (53842 cases annually) and mortality (36436 cases

annually), followed by the USA, China, Pakistan, Bangladesh, Brazil, Russia, Germany and Japan (Figure 2B) (GLOBOCAN, 2012, IARC).

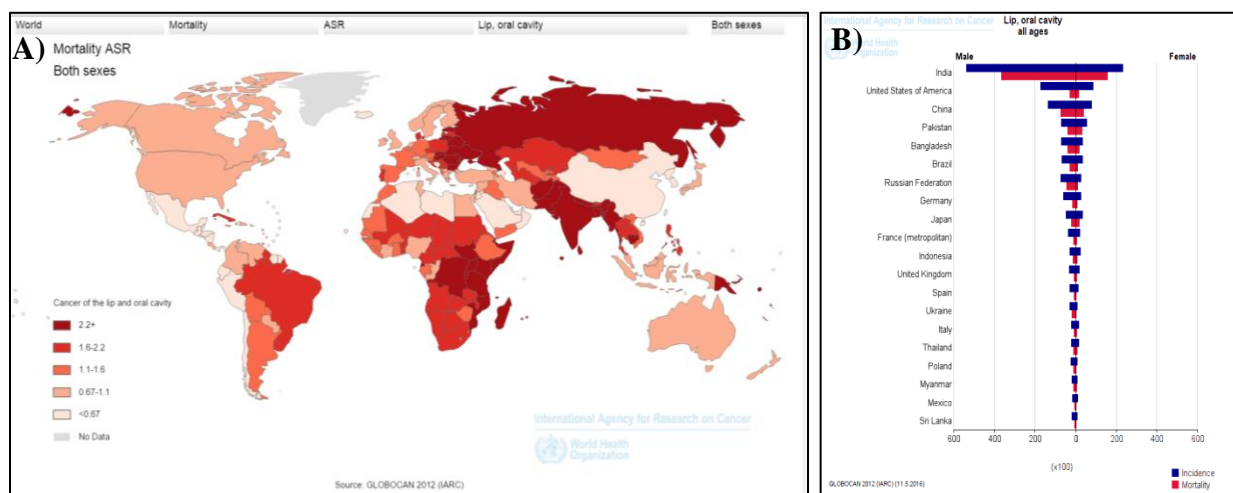


Figure 2: Lip and Oral Cavity Related Incidence and Mortality. A) Worldwide mortality ASR, for both sexes. **B)** Incidence and mortality in highest 20 populations across the globe. (Source: GLOBOCAN, 2012, IARC).

In Indian men, 53842 new cases of oral cancer (an age standardized rate (ASR) = 10.1) were reported in 2012, accounting for 36436 (an age standardized rate (ASR) = 6.7) deaths, making it one of the top most cancers in terms of incidence and the third most common cause of cancer related mortality (Figure 3). In females, the proportion of oral cavity cancer is lower as compared to men, with an estimated incidence and mortality of 23161 (ASR= 4.3) and 15361 (ASR= 3) cases respectively. As per GLOBOCAN, this number is predicted to increase to around 17900 new cancer cases (for both sexes) by 2020.

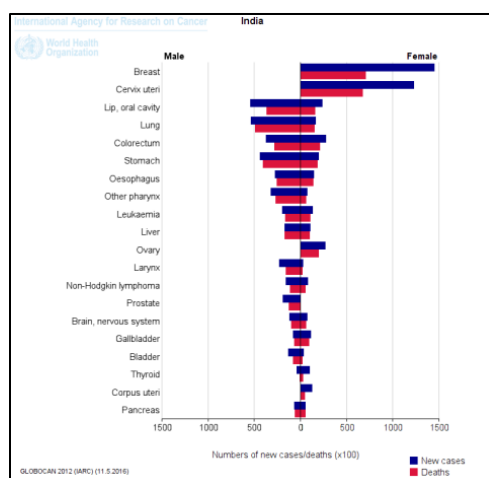


Figure 3: Lip and Oral Cavity Related Incidence and Mortality in Indian Population (Source: GLOBOCAN 2012, IARC).

1.1.2. Etiology

In India, a major proportion of oral cancers and OPMD's are etiologically associated with abuse of smokeless tobacco in various forms such as khaini, betel quid with tobacco in combination with smoking and alcohol consumption [1, 24, 25]. However, there exists a group of no-habit patients that may potentially harbor high-risk HPV infection [26]. The genetic differences in individuals for e.g polymorphisms in carcinogen metabolising enzymes, could greatly influence risk of cancer development [27-29]. In addition, other risk factors include poor oral hygiene, poor nutrition and chronic irritations by sharp tooth or loose dentures [30].

1.1.2.1. Smokeless and Smoking Tobacco

Smokeless and smoking tobacco in the form of khaini, mishri, betel quid, oral snuff, cigarette, and bidi are strongly associated with the development of oral cancers and OPMD's. Betel quid mostly consists of tobacco, areca nut, slaked lime, catechu and several condiments, wrapped in a betel leaf [31]. The risk of oral cancer development varies with the duration and frequency of tobacco usage. Various reports indicate that smoking or tobacco chewing in combination with

alcohol consumption, substantially increases the risk of oral cancer development [32]. To date 28 carcinogens have been identified in smokeless tobacco. The major and most abundant group of carcinogen are non volatile tobacco-specific *N*-nitrosamines (TSNA), *N*-nitrosoamino acids and volatile *N*-nitrosamines [33].

1.1.2.2. Alcohol

Alcohol (ethanol) consumption independently, and in combination with tobacco, is strongly associated with oral cancer development [32]. Alcohol impels its carcinogenic effect after getting metabolised to acetaldehyde or by modulating cell permeability. Additionally, alcohol influences systemic functions by causing immunogenic and nutritional deficiencies [30].

1.1.2.3. High Risk HPV

High-risk human papilloma virus (HPV) types 16, 18, 31, 33, 52, 58 have emerged as an important etiological factor in the development of a subset of HNSCC [34, 35]. Globally, the proportion of HPV-related cancers are higher in the oropharynx as opposed to other head and neck subsites [36-38]. HPV-positive tumors show a distinct molecular-genetic signature from its HPV negative counterpart, implying two different pathways for HNSCC development [15, 16, 26, 39-42]. HPV drives its carcinogenic effect by inactivating p53 and pRB via its oncoproteins E6 and E7, along with frequent loss of chromosomal loci 3p, 9p, and 17p, in contrast to HPV negative tumors, which show frequent p53 mutations and less frequent loss of the above loci [15, 16]. In addition to distinct pathogenesis, difference in clinical characteristics are evident in HPV

positive and negative HNSCC patients, this includes the early onset of disease and good prognosis in HPV positive HNSCC patients [26].

1.1.3. Predisposition to Malignant Transformation

Oral cancer has a well documented long preclinical phase observed as pre-cancerous lesion. These lesions include homogeneous leukoplakia, non-homogeneous leukoplakia, verrucous leukoplakia, erythroplakia, oral submucous fibrosis (OSMF), lichen planus and chronic traumatic ulcers [30]. In 2005, WHO recommended to use the term “potentially malignant” instead of “pre-malignant or pre-cancerous” and also insisted to use the term “potentially malignant disorder” instead of “pre malignant lesion/ potentially malignant condition”[43].

1.1.3.1. Leukoplakia

Leukoplakia is the most predominant potentially malignant disorder, clinically observed as white or opaque lesion/plaque. It is most commonly present on the buccal mucosa, alveolar mucosa, and lower lip. Lesions on the lateral tongue, lower lip or in the floor of mouth frequently show dysplastic changes or malignant transformation [44]. Worldwide estimate for the prevalence of leukoplakia is around 2%, and is six times more common in smokers, than amongst the nonsmokers [45]. However, the transformation rate of these lesions vary considerably from 0.13 to 17.5%, with observation periods ranging from 1 to 30 years (Table 1), it is also influenced by various factors such as the type of lesion, presence or absence of dysplasia, site of occurrence, observation period, oral habits, gender and age [46, 47].

Table 1: Malignant Transformation of Leukoplakia (Source: Amagasa, et al. 2011).

References	Country	Year	No. of patients	Malignant transformation (%)	Observation periods (mean) years
Silverman et al.	India	1976	4,762	0.13	2
Gupta et al.	India	1980	360	0.3	1–10 (7)
Mehta et al.	India	1972	117	0.8	10
Gupta et al.	India	1980	410	2.2	1–10 (8)
Roed-Peterson et al.	Denmark	1971	331	3.6	1≤ (4.3)
Einhorn et al.	Sweden	1967	782	4.0	1–20
Pindborg et al.	Denmark	1968	248	4.4	1–9
Kramer et al.	England	1969	187	4.8	1–16
Silverman et al.	USA	1968	117	6.0	1–11
Banocy et al.	Hungary	1977	670	6.0	1–30
Amagasa et al.	Japan	2006	444	7.9	1–29
Lind	Norway	1987	157	8.9	6≤
Gangadharan et al.	India	1971	626	10.0	1–19
Schepman et al.	Holland	1998	166	12.0	6 months–17 (2.7)
Silverman et al.	USA	1984	257	17.5	6 months–39 (7.2)

Early lesions are greyish white, which get thicker and whiter as the lesions progress, sometimes developing a leathery appearance with surface fissures (homogeneous or thick leukoplakia). Some leukoplakia develop surface irregularities and are referred to as “granular” or “nodular leukoplakia”, while other lesions develop a papillary surface and are known as “verrucous” or “verruciform leukoplakia” (Figure 4). Histologically, the epithelium undergoes distinct changes such as hyperplasia, hyperkeratosis, acanthosis and mild to severe dysplasia. The assessment and severity of dysplasia are based on the disruption of the epithelial architecture and cytological changes, e.g. irregular epithelial stratification, loss of polarity by basal cells, change in shape of rete ridges, increased keratinization, change in cell size and shape and increased nuclear size [43]. When alterations occurs in the basal or parabasal keratinocytes, it is referred to as mild dysplasia, the atypia found in the middle level is called moderate dysplasia. When changes are extended from base to the surface layer, it is termed as advanced/severe dysplasia [48].

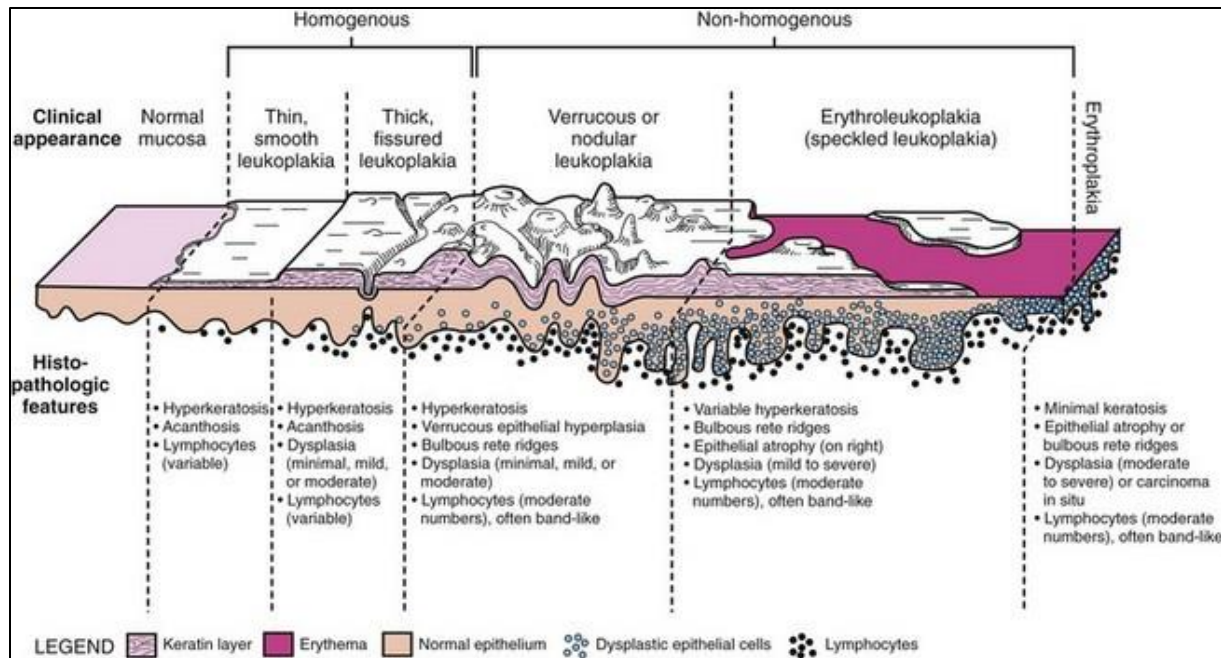


Figure 4: Clinical and Histopathological Features of Leukoplakia [49]

Oral leukoplakia are graded based on the size of the lesion as well as the histopathological characteristics as mentioned in Table 2, additional information regarding, gender, age of diagnosis, etiology is also recorded for further disease management [43]. The treatment modality for leukoplakia mostly consists of surgical excision or laser surgery (including evaporation). Time to time observational therapy along with administration of Retinoids, either topically or systemically and photodynamic therapy is also considered. Furthermore, the risk of local recurrence or development of new leukoplakia varies from zero to around 30% [43, 50, 51].

Table 2: Classification and Staging System for Oral Leukoplakias (OL- System) [43]

<i>L (size of the leukoplakia)</i>	
L ₁	Size of single or multiple leukoplakias together <2 cm
L ₂	Size of single or multiple leukoplakias together 2–4 cm
L ₃	Size of single or multiple leukoplakias together >4 cm
L _x	Size not specified
<i>P (pathology)</i>	

P ₀	No epithelial dysplasia (includes “no or perhaps mild epithelial dysplasia”)
P ₁	Mild or moderate epithelial dysplasia
P ₂	Severe epithelial dysplasia
P _x	Absence or presence of epithelial dysplasia not specified in the pathology
<i>OL-staging system</i>	
Stage I	L ₁ P ₀
Stage II	L ₂ P ₀
Stage III	L ₃ P ₀ or L ₁ L ₂ P ₁
Stage IV	L ₃ P ₁ , any L P ₂

1.1.3.2. Erythroplakia

Oral erythroplakia is the clinical term that refers to a red plaque with velvety texture, mostly observed on lateral tongue, retromolar pad, soft palate and floor of the mouth. Although the erythroplakia lesions are not very common, but these lesions are of higher risk, almost all these lesions show dysplasia, and more than 50% transform into squamous cell carcinoma [46]. Most recommended treatment for erythroplakia is surgery, either by cold knife or by laser, however, there are no data from the literature regarding the recurrence rate after excision of these lesions [43].

1.1.3.3. Oral Submucous Fibrosis

Oral submucous fibrosis (OSMF) is marked by rigidity of the mucosa, due to progressive fibrosis of the submucosal tissues and inflammation, thus the patient is unable to open the mouth. This disorder is frequently observed on the buccal mucosa and is usually associated with areca nut chewing, directly or in the form of betel quid. The pathogenesis of oral submucous fibrosis is not well established and believed to be multifactorial [52].

1.1.3.4. Other Lesions

Additional lesions observed in oral cavity include, oral lichen planus, nicotine stomatitis (related to pipe smoking), and tobacco pouch keratosis (associated with chewing tobacco). Although the frequency and transformation of these lesions varies considerably, and most of these lesions reverse back on habit discontinuation [46].

1.1.4. Oral Squamous Cell Carcinoma (OSCC)

More than 90% of the neoplasm of oral cavity are oral squamous cell carcinoma (OSCC) and usually presents an endophytic growth pattern that is characterized by a depressed, ulcerated surface with a raised, rolled border; or may have exophytic margins with fungating or papillary surface. OSCCs arise in the apparently normal mucosa or are preceded by clinically obvious potentially malignant disorder, especially erythroplakia and leukoplakia [53]. It is important for the clinician to be aware of this high-risk lesions/ulcers when examining the oral cavity, as early stage OSCC are mostly asymptomatic or may cause minor discomfort to patients, hence go unnoticed.

1.1.4.1. Metastasis

OSCC most frequently metastasize to cervical lymph nodes; either ipsilateral, contralateral or bilateral lymph nodes can be involved. In addition to locoregional metastasis, distant metastasis in lungs is also commonly observed [46]. Metastasis positive nodes, are usually enlarged and firm. If the tumor perforate the capsule of the involved node and invades the surrounding connective tissue (extracapsular spread), the node will feel fixed, and immovable. The metastatic

node, can be either palpable or occult (invisible) at the time of initial evaluation [46]. Lymph node metastasis is very frequent and an important factor in the prognosis and treatment of OSCC patients. Considering the risk of occult metastasis, elective neck surgery is performed in almost all the patients, even if the metastasis is not clinically apparent. The disadvantage of this policy is that the patient will be subjected to the morbidity of over treatment. The advancement of modern imaging techniques may have increased the sensitivity of detecting positive nodes, but still have many limitations, especially with micrometastasis [54-56]. Hence, it is important to have a reliable biomarker, which might help predict the risk of lymph node metastasis.

1.1.5. OSCC Diagnosis and Staging

“Early diagnosis gives the best prognosis”. Delay in identification and recognition of suspicious lesions contributes to disease advancement and thus reduce patient survival [46]. It is important that the general physician or dentist perform the oral examination specifically in patients with tobacco or alcohol habits. Any suspicious lesions or growth should be identified, assessed and treated immediately. Following the diagnosis, the staging assessment is completed and treatment is planned for the patient. Staging of OSCC is done according to The American Joint Committee on Cancer’s (AJCC) TNM system, based on three key components:

- **T** (Tumor size)
- **N** (Absence or presence of lymph Node metastasis)
- **M** (Distant Metastasis)

1) T: Primary Tumor assessment is done by physical examination and imaging, and categorized as below:

- TX. Primary tumour cannot be assessed
- T0. No evidence of primary tumour

Tis. Carcinoma in situ

T1 ≤ 2 cm, in greatest dimension

T2 >2 to 4 cm, in greatest dimension

T3 >4 cm

T4a. Invade through cortical bone, deep/extrinsic muscle of tongue, maxillary sinus, skin of the face

T4b. Involves, masticator space, pterygoid plates, skull base, or encases internal carotid artery

2) N: Cervical lymph Node assessment is done by physical/ histological examination, and/or by imaging, and categorized as below:

NX. Regional lymph nodes cannot be assessed

N0. No regional lymph node metastasis

N1. Metastasis in a single ipsilateral lymph node, <3 cm size

N2. Metastasis in a single or multiple lymph node (ipsilateral/ bilateral/ contralateral), > 3 cm and < 6 cm

N2a. Metastasis in a single ipsilateral lymph node, > 3 cm and < 6 cm

N2b. Metastasis in multiple ipsilateral lymph nodes, < 6 cm

N2c. Metastasis in bilateral or contralateral lymph nodes, < 6 cm

N3. Metastasis in a lymph node > 6 cm in greatest dimension

3) M: Distant Metastasis

MX. Distant metastasis cannot be assessed

M0. No distant metastasis

M1. Distant metastasis

Staging of OSCC is done based on the above classification, and further categorized as stage 1, stage 2, stage 3 or stage 4 tumors. Early stage OSCC belongs to stage 1 and stage 2 category, while all the stage 3 and stage 4 tumors are considered advanced stage OSCC. Detailed categorization is provided in Table 3.

Table 3: Staging of Oral Cavity Tumors (AJCC TNM System)

	Stage 0	Tis	N0	M0
Early Stage OSCC	Stage I	T1	N0	M0
	Stage II	T2	N0	M0
Advanced Stage OSCC	Stage III	T1, T2	N1	M0
		T3	N0, N1	M0
	Stage IVA	T1, T2, T3	N2	M0
		T4a	N0, N1, N2	M0
	Stage IVB	Any T	N3	M0
		T4b	Any N	M0
	Stage IVC	Any T	Any N	M1

1.1.6. OSCC Treatment and Prognosis

Treatment for OSCC patients is variable and depends on the stage/ location of the primary tumor, local and/or distant metastasis, and the patient's ability to tolerate the treatment. Early stage OSCC are treated with surgery and radiation therapy either as a single modality or in combination. Whereas locally advanced tumors are usually aggressive and treated with surgery followed by postoperative radiotherapy with or without chemotherapy. Elective neck dissection (partial or radical neck dissection) is performed in the majority of the cases, depending on the size and extent of node involvement, or if the possibility of node metastasis is predicted [30, 57].

1.1.6.1. Surgery

Depending on tumor stage, a surgical procedure is planned for the patient. For the potentially malignant disorders, micrographic surgery or laser surgery is recommended. While for patients

with OSCC, surgery is planned depending on the extent of disease spread, e.g. only tumor resection can be done in some cases, while for others, partial mandible (jawbone) resectioning or glossectomy or maxillectomy may be needed. Elective neck dissection is done in lymph node metastasis positive patients or in cases with predicated risk of nodal involvement. The neck dissection procedure can be partial/ selective, which involves removal of a few lymph nodes; or it can be radical neck dissection, that removes almost all lymph nodes on one side, with more extensive removal of muscle, nerves, and veins [58].

1.1.6.2. Radiotherapy

External Beam Radiation Therapy (EBRT) and/or brachytherapy is a type of radiation therapy given to OSCC patients and can be postoperative or primary modality of treatment. In EBRT, the patient receives a high-energy radiation dose of 2Gy daily for approximately 30 days. Postoperative radiation therapy has been shown to improve locoregional disease control. Primary radiation is provided to the patient in the following scenerios, a) early stage OSCC to avoid morbidity, b) unresectable tumors, c) high operative risk patients with poor performance status, d) recurrent disease when previous multiple surgeries have been undertaken, and e) patient's preference [59].

1.1.6.3. Chemotherapy and Targeted Therapy

Chemotherapy is used in combination with radiotherapy or targeted therapy. Currently no conclusive data are available regarding the use of chemotherapy as a single treatment modality. Most commonly used drugs in various combination are 5-fluorouracil (5-FU), Cisplatin,

Docetaxel, 5-chloro-2, 4-dihydroxypyridine (CDHP), and Potassium oxonate (Oxo) [60-62]. Targeted therapy is emerging as an alternative to the standard chemotherapy, due to its specificity and fewer side effects compared to that of standard chemotherapeutic drugs. EGFR and its signaling pathway components are the potential target in OSCC for which various clinical trials are ongoing [63, 64].

1.1.6.4. Prognosis

Lymph node involvement and tumor stage are the important prognostic factor, in OSCC patients. Most of the patients with OSCC succumb to the overtreatment, e.g., physical morbidity, or side effects of radiochemotherapy. Despite therapy advancement, no great improvement in the 5 year overall survival is observed in OSCC patients. Figure 5 represents an overview of the treatment management for OSCC patients at different stages of disease and their prognosis. Overall, early diagnosis and primary prevention remain the best modality for OSCC management.

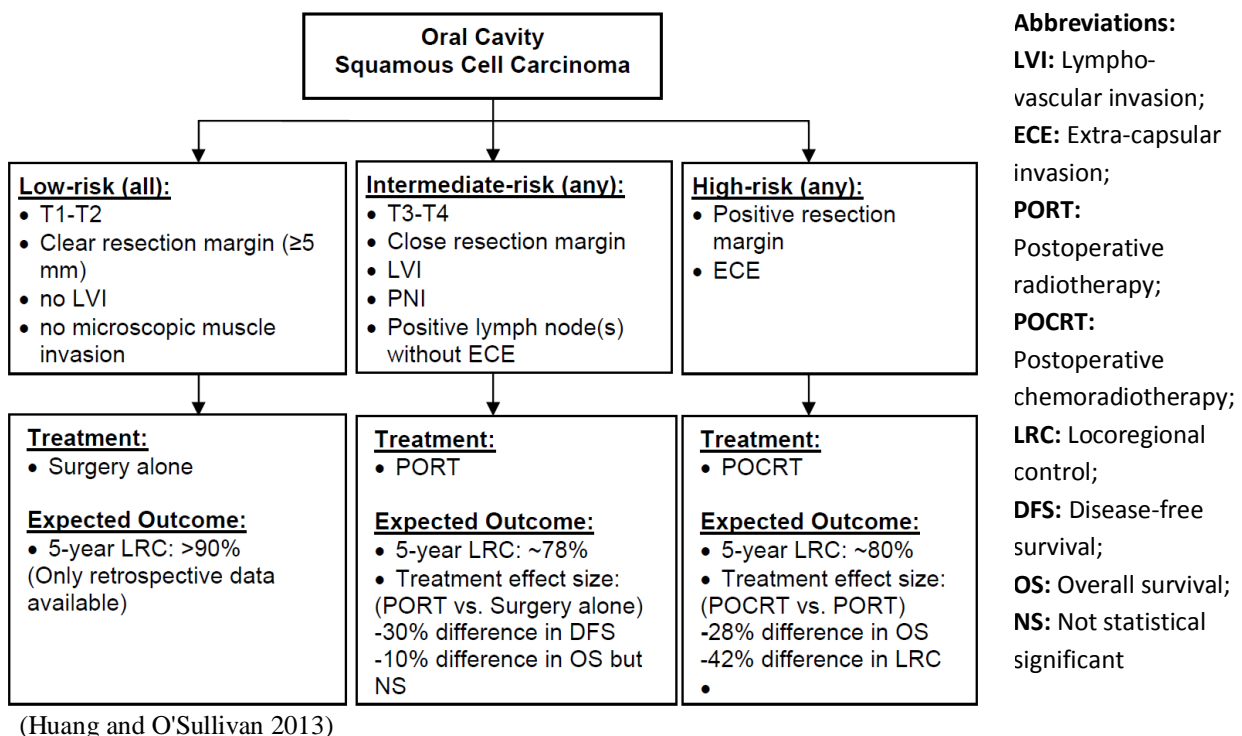


Figure 5: Treatment and Prognosis Overview of Oral Cavity Cancer [59].

1.2. Molecular Pathogenesis of OSCC

Oral carcinogenesis is a complex multistep process that involves accumulation of various genetic and epigenetic events. Earlier reports have demonstrated that these molecular changes (genetic/epigenetic) occurs well before the change is microscopically detectable. Additionally, molecular alterations are efficient in identifying the cancerized field (adjacent to tumor) which other wise may be considered as normal based on the microscopic evaluation (Figure 6). The number of acquired alterations progressively increases with disease advancement from squamous hyperplasia through dysplasia to invasive carcinoma (Figure 7) [48]. Chronic exposure to carcinogenic agent may lead to genomic imbalances either in the form of activating mutations, amplification of oncogenes or deletion of tumor suppressor genes. Most of the altered chromosomal loci, encodes for the critical components of the key molecular pathways regulating cell proliferation, migration, and stromal interactions. Epigenetic events do not alter the DNA sequence, however, they regulate the expression of genes by modifying promoter or histones (e.g. hyper- or hypo-methylation or acetylation) [53, 65].

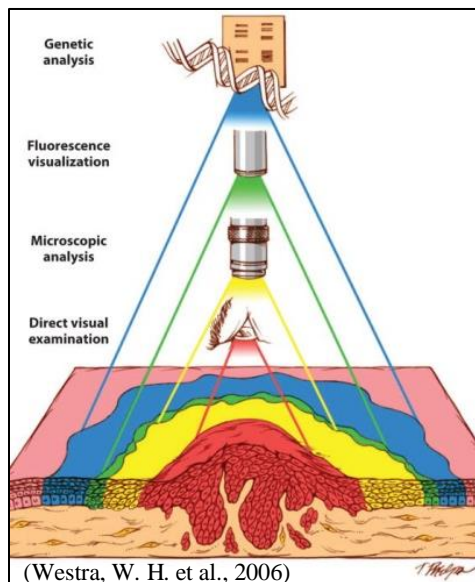


Figure 6: Tumor Field Cancerization: Visualization and Identification Methods [66].

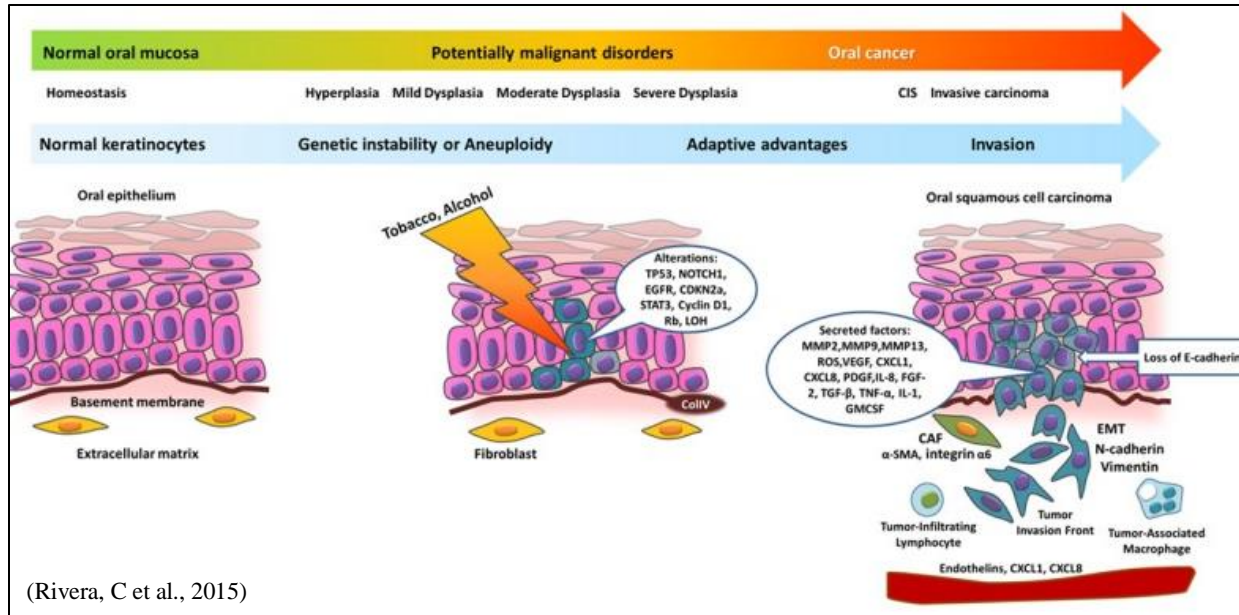


Figure 7: OSCC Disease Progression Model.

1.3. Genomic Imbalances in OSCC

OSCC arises through accumulation of genetic alterations, including chromosomal imbalances (structural and numerical chromosomal abnormalities) and/or DNA alterations (mutations, amplifications, deletions). Gain or losses of whole chromosome could change ploidy of the cells. Other important alterations include insertions, deletions (loss of small or large DNA segments), amplifications and translocations (balanced or unbalanced re-arrangement of chromosome segments) (Figure 8). Amplifications are thought to be generated as a result of ‘chromosomal breakage-fusion-bridge’ cycles, that give rise to multiple additional copies of a chromosomal region in the form of extrachromosomal double minutes [dmin] or intrachromosomal homogeneously staining regions [hsrs]), while deletions are predicted to occur as a result of chromosomal breaks at fragile sites [67-69].

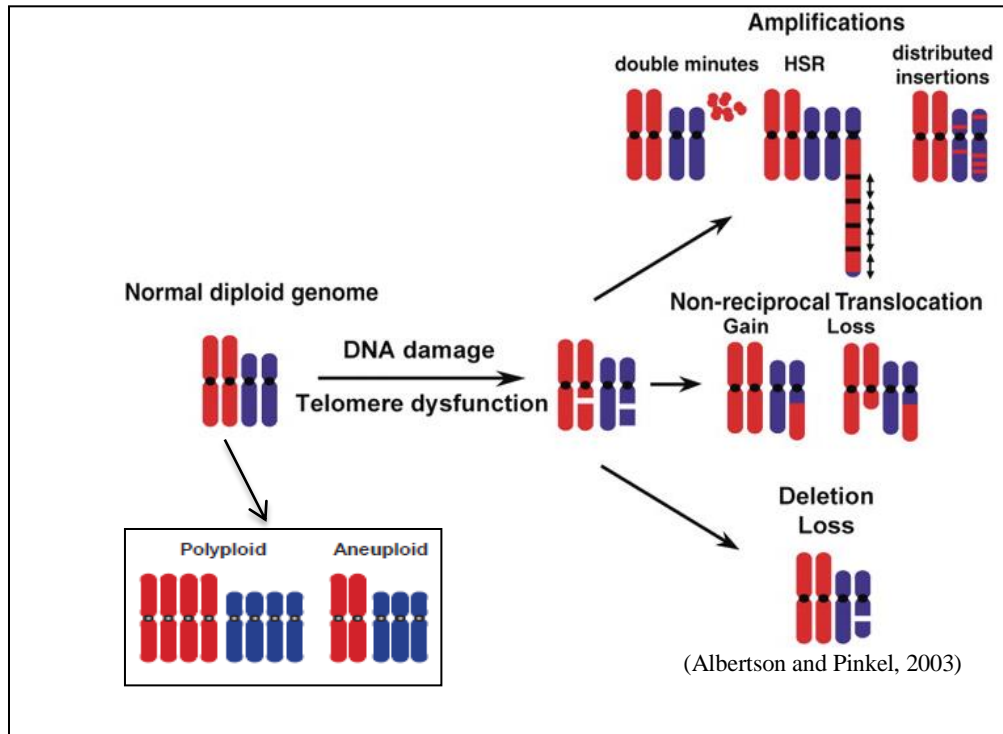


Figure 8: Genomic Imbalances in Cancers.

1.3.1. Copy Number Alterations (CNA)

Copy number change of either specific genes, or parts of a chromosome or an entire chromosomes have a dramatic impact on OSCC development and progression. DNA copy number alterations are categorised into the submicroscopic changes that are less than 500 kbp in size, and the microscopic changes for those which are greater than 500 Kbp. Depending on the size of alteration, the CNA are further categorised as copy number variation (1kbp and 1Mbp in length) specifically in germline cells, insertion or deletion (1 bp to 1 kbp in size), multiples of the entire genome (polyploidy), loss or gain of a single chromosome (aneuploidy) and small/large scale CNAs (segmental or arm level CNA) [67, 70].

1.3.2. The Biological Impact of CNA

The changes in the DNA copy (copy number gain/amplification or copy number loss/deletion) has a major impact on cellular homeostasis and it is presumed to alter the expression of the gene or gene cluster mapping within the region. The dramatic change in the gene copy (either overexpression or underexpression) can activate an oncogene or suppress a tumor suppressor gene and thus can have a major impact on various cellular pathways. Additionally, overexpression of many proteins at once in the absence of sufficient chaperones may lead to accumulation of misfolded proteins and this protein stoichiometric imbalance can also contribute to proteotoxicity. In addition, non-gene-specific effects are also observed, which includes: 1) Changes in the chromosomal copies (aneuploidy). 2) Alteration at centromeric or telomeric region have a major impact on chromosomal segregation. 3) The altered region may have regulatory molecules such as miRNAs, which may affect the expression of many genes under their regulation [68, 70].

1.4. Identification of CNA in OSCC

Large scale genomic, molecular genetic and cytogenetic techniques can be used for detection and identification of CNA. Few of the most widely used techniques includes: high-throughput analysis of loss of heterozygosity (LOH), comparative genomic hybridization (CGH)-aCGH, microarray, fluorescence *In Situ* hybridization (FISH), restriction landmark genome scanning (RLGS), SNP arrays, and more recent ones include next generation sequencing (NGS) and whole exome sequencing [67].

1.4.1. Array Comparative Genomic Hybridization (aCGH)

Comparative genomic hybridisation (CGH) is a molecular cytogenetic technique performed to study copy number gains and losses across the genome. This technique uses test and normal reference DNA labelled with different fluorochromes, which are hybridized to normal chromosome metaphase spreads. Differences in the tumor to normal fluorescence ratio along the metaphase chromosomes are then quantitated and reflect changes in the DNA sequence copy number in the tumor genome. However, this method does not detect specific transcripts within areas of gain and loss due to its poor resolution (20 Mb). In late 1990, aCGH was developed, which improved the resolution of copy number detection. In array CGH, test and reference DNA are hybridised to microarrays which have been printed with either genomic fragments such as bacterial artificial chromosome (BAC) clones from key genomic regions (aCGH) or microarrays containing cDNA clones which are used for gene expression analysis (Gene expression arrays (GE)). The brief layout of aCGH and GE process is depicted in Figure 9. With the rapid technological improvement the BAC arrays were soon replaced with high resolution and more specific oligonucleotide based arrays. Comparisons between different microarray platforms and their resolution is provided in Table 4 [71-73]. aCGH has wide applications in cancer research, including tumor classification, identification of prognostic markers, genome wide copy number screens, high resolution regional analysis, mapping of aberration boundaries, identification of tumor progression markers [74, 75].

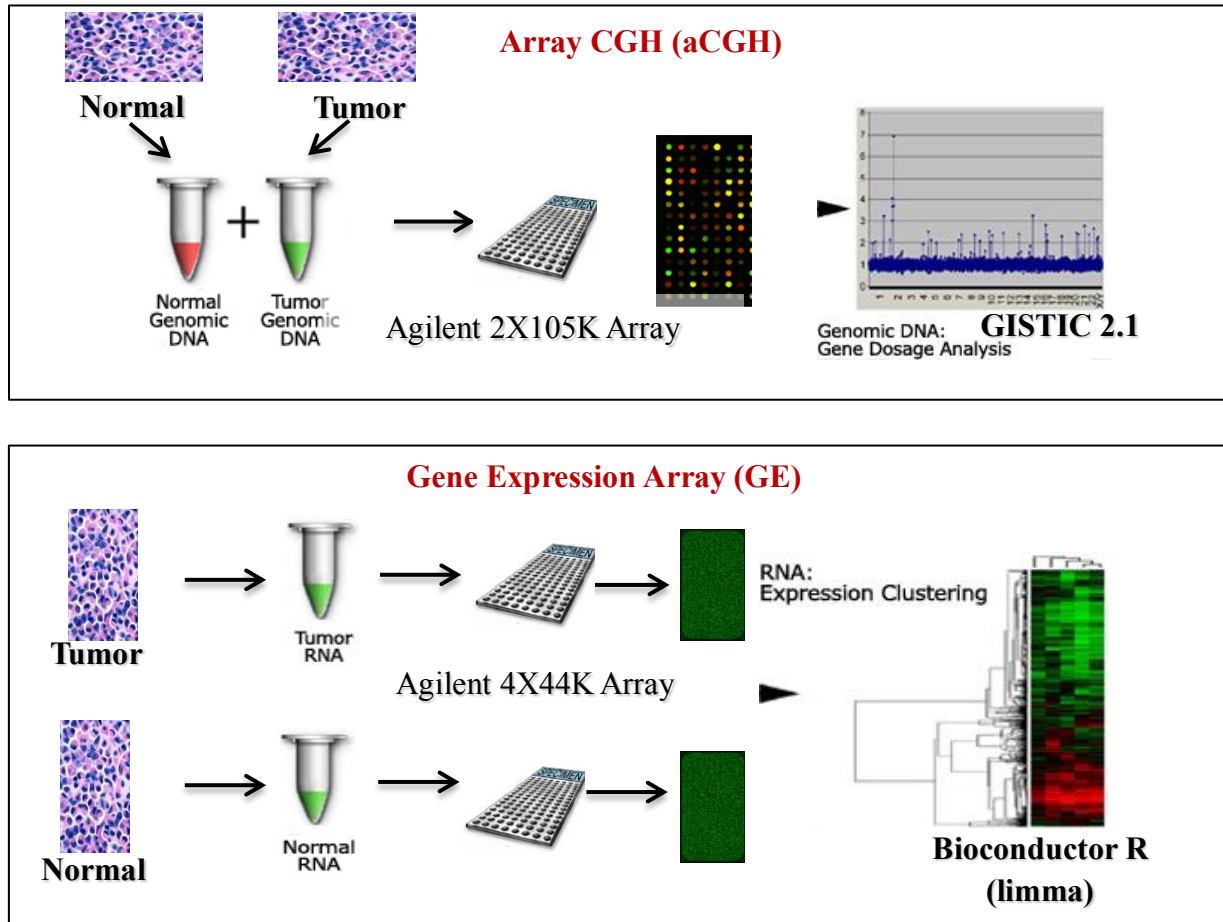


Figure 9: Overview of aCGH and Gene Expression Microarray Analysis Used in the Current Study.

Table 4: A) Comparisons between different CGH methods B) Microarray platforms available with their technical details.

Probe Type/ Platforms	Theoretical Resolution	Theoretical Coverage	Comments	A
Conventional CGH	>2Mb (cytoband)	Complete	•Technically demanding	
BAC	100Kb	Complete	• Tough to replicate , propagate, and maintain. •Specificity issues, •CNA that occur in gap between BAC will not be detected.	
cDNA	2Kb	Genes only	no intragenic regions	
Oligo (25-85 mer)	0.025 to 0.08 Kb	Complete	•High throughput, High resolution, •Select genes, inter-genic regions	

B	Vendor	Type of array	Resolution	Feature type
	College of Medicine, Houston, TX	Targeted array	Tiling path in regions of known medical significance. Resolution in these regions = BAC length (70–200 kb). Oligo array under development	BACs
	Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK	Targeted array	Tiling path in regions of known medical significance. Resolution in these regions = BAC length (70–200 kb)	BACs
		Whole-genome array	Tiling path across genome: resolution approximately 70–200 kb	
	Signature Genomics, Inc., Spokane, WA	Targeted arrays: SignatureChip, MarkerChip	Tiling path in regions of known medical significance. Resolution in these regions = BAC length (70–200 kb)	BACs
	Spectral Genomics, Inc., Houston, TX	Targeted array: Constitutional Chip	Tiling path in regions of known medical significance. Resolution in these regions = BAC length (70–200 kb)	BACs
		Whole-genome array: Spectral Chip 2600	Averaged resolution: 1-Mb resolution across entire genome	
	Agilent Technologies, Inc., Santa Clara, CA	All are whole-genome arrays:	Averaged resolution (gene centric):	60-mer oligo
		44B (43,000+ probes)	35 kb	
		105K (99,000+ probes)	15 kb	
		244K (236,000+ probes)	6.4 kb	
	Affymetrix, Inc., Santa Clara, CA	All are whole-genome and SNP-based ^a :	Averaged resolution:	25-mer oligo
		100K mapping array	26 kb	
		500K mapping array	5 kb	
		Genome-wide human SNP array 5.0	4 kb	
	Illumina, Inc., San Diego, CA	All are whole-genome and SNP based ^a :	Averaged resolution:	50-mer oligos on beads
		Infinium HumanHap300 (318,000 probes)	5 kb	
		Infinium HumanHap550 (555,000 probes)	2.8 kb	
		Infinium HumanHap650 (655,000 probes)	2 kb	
	Nimblegen, Inc., Madison, WI	All are whole-genome arrays (385,000 probes):	Averaged resolution (repeats masked):	50–75-mer oligo
		Single array	6.27 kb	
		4-set arrays	1.57 kb	
		8-set arrays	713 bp	
(Swaroop, A. et al., GIM 2007)				
CGH, comparative genomic hybridization; BAC, bacterial artificial chromosome; SNP, single nucleotide polymorphism.				
^a Require whole genome amplification before DNA labeling.				

1.4.2. Integration of Genetic Events and Gene Expression Patterns

Alteration in the genomic copies leads to deregulation of specific genes located within that region. The advent of array technology has made it feasible to study genomic alteration at the gene level resolution, allowing accurate integration of the two data types. Genomic alterations represent categorical data (fixed number of copies gained or lost) unlike the gene expression results, leading to fewer complexities in the analysis. Pure gene expression data measures the result of many complex gene interactions and the primary alterations may be lost in the noise [71]. Hence, by looking at genes altered as a result of copy number gain/loss, one can identify potentially meaningful interconnections in the pathways. Additionally, integration based analysis

may potentially allow the identification of predictive DNA based markers which are far easier to apply in clinical scenario, due to the ease of handling DNA samples that are more stable compared to the RNA [76]. A further benefit of integrative analysis is that the identification of genes controlled by copy number alterations can allow expansion of studies into archival samples where RNA may not be available but large sample sets can be acquired [71]. Akavia, U. D. et al., have postulated that driver mutations coincide with a “genomic footprint” in the form of a gene expression signature that can assist in distinguishing between driver and passenger mutations based on the following assumptions: 1) A driver mutation should occur in multiple tumors more often than would be expected by chance. 2) A driver mutation may be associated with the expression of a group of genes that form a ‘module’. 3) CNA often influence the expression of genes in the module via changes in expression of the driver genes [77].

1.5. Microarray Profiling of Oral Potentially Malignant Disorder (OPL) and OSCC

The genome-wide analysis of somatic CNAs is an important tool to distinguish the driver alterations responsible for tumor growth from the passenger events that may have accumulated during tumorigenesis. Most of the earlier studies have been carried out using either conventional CGH or BAC arrays [12, 78]. Baldwin, C. et al., have demonstrated few novel microalterations in 20 archival OSCC samples this includes, copy number gains at 3q23, 5p15.2, 7p12.3-13, 7q21.2 and 7q35 locus and copy number losses at 2p15, 4q34.3, and 16q23.2 locus. In addition, more frequent loss of 3p 8p, 9p, 10q, 11, 18q, and 21q and frequent gain of 8q, 3q, 9q, 11q, 14q, and 20q were also observed [7]. Garnis, C. et al., and Tsui, I. F. et al., highlighted the importance of segmental gain or loss compared to arm level alterations in 62 OPL and 24 invasive OSCC. In

addition, 12 more studies have been carried out on the multiple OSCC subsites which used different array platforms (Table 5) [10, 13, 79-84].

Table 5: List of 12 Studies that have Conducted Genome Wide Profiling of OSCC (Source: Salahshourifar, I. et al., 2014)

Studies	OSCC samples	Platform	Site
Vincent et al	46	SurePrint G3 Human 1 M aCGH, Agilent	Tongue, BM, FOM, HP, GUM
Chen et al.	60	GenoSensor Array 300, Vysis	BM and non BM
O'Regan et al.	20	GenoSensor Array 300, Vysis	Tongue, FOM, Soft palate, Gum
Sparano et al.	21	4134 BAC clones (Ultra GAPS; Corning, NY)	Site has not been mentioned
Freier et al.	40	DNA microarrays (matrix-CGH)	Soft palate, FOM, Mandible, maxilla
Ambatipudi et al.	60	105 K Oligo aCGH, Agilent	Tongue and gingivobuccal complex
Yoshioka et al	25	44 k oligo aCGH, Agilent	Tongue, BM, Gum, FOM
Snijders et al.	89	2 K aCGH, UCSF	BM, FOM, Gum, Tongue,
Noutomi et al.	35	CGH	Tongue, gingiva, buccal region, FOM, and palate
Sugahara et al	54	44 K Oligo aCGH, Agilent	Tongue, Gingiva, Palate, BM, FOM
Cha et al.	7	44 K Oligo aCGH, Agilent	Site has not been mentioned
Uchida et al.	50	MAC array Karyo 4 K, MacroGen	FOM, BM, Maxillary gingiva, Mandibular gingiva, Tongue

BM, Buccal mucosa; FOM, floor of mouth; HP, hard palate.

Very few studies have analysed the correlation between the gene expression changes in leukoplakia and OSCC with the clinical outcome [85-88]. Most of the recent studies focus on the integration of CNA and gene expression data for better understanding of significant driver alterations. Xu, C. et al., and Pickering, C. R. et al., independently identified the copy number dependent changes and their association with node metastasis, key pathways and drivers genes in OSCC, however, both these studies were carried out using SNP arrays [89, 90]. A previous report from our lab (Ambatipudi, S. et al.,) have focused on the CNA associated with advanced stage OSCC followed by integrative analysis to identify driver genes in oral tumorigenesis [10, 91]. However, till date, no study comprehensively reported genomic signatures associated with OSCC progression from pre invasive lesions (leukoplakia) to advanced stage OSCC using samples from a single subsite (gingivobuccal complex) making the study group more homogenous.

2. Aims and Objectives

2.1 Study Hypothesis

The development of OSCC is a multistep process involving the accumulation of both genetic and epigenetic changes. Systematic characterization of the oral cancer genome has revealed diverse aberrations that differ among individuals. Moreover, not much data is available for the early events of genetic degeneration, namely the transition from normal mucosa to dysplasia followed by OSCC development. Broadly, objective of this study was to identify the genomic and transcriptomic signatures associated with progression from low risk lesions (OPL-leukoplakia) to high risk invasive OSCC and the prediction of patient's clinical outcome.

2.2 Objectives

2.2.1 To identify the genomic alterations and transcriptomic changes in non-HPV associated oral precancerous lesions (OPL) and early stages of OSCC.

2.2.2 To validate results of aCGH and GE in OPL and OSCC.

3. Materials and Methods

3.1 Clinical Tissue Sample Collection

3.1.1 Leukoplakia and OSCC Tissue Sample Collection

The study was approved by The Institutional Review Board (IRB) and the Local Ethics Committee of Tata Memorial Hospital (TMH) and Nair Hospital Dental College. Written informed consent was obtained from all the study participants. Paraffin blocks and frozen tissue samples of leukoplakia and neo-primary treatment naive oral tumor patients were recruited from the Nair Hospital Dental College and Tata Memorial Hospital respectively. To ensure confidentiality, all cases were anonymized. Patients received neither radiation nor chemotherapy before surgery.

The cases that were recruited in this study were based on following inclusion criteria:

- Only primary leukoplakia lesions were included in the study. Any lesions adjacent to a tumor or reccured after removal of primary tumor were excluded from this study.
- Only primary OSCC was included. Tissue samples were collected at the time of surgery prior to chemotherapy or radiotherapy.
- Control tissues were obtained from healthy individuals with no previous personal history of cancer.

TNM classification and differentiation status were determined according to the AJCC classification 2002; and WHO International Histological Classification of tumors respectively. Patient details and follow up were obtained by reviewing the hospital records. The disease-specific survival (DSS) rate was defined as the time period, which usually begins at the time of diagnosis or at the start of treatment and ends at the time of death due to disease or last follow-up visit of the patient (<http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=44311>).

Recurrence free survival (RFS) was defined as the time period between diagnosis or at the start of treatment and recurrence of disease or last follow-up visit of the patient.

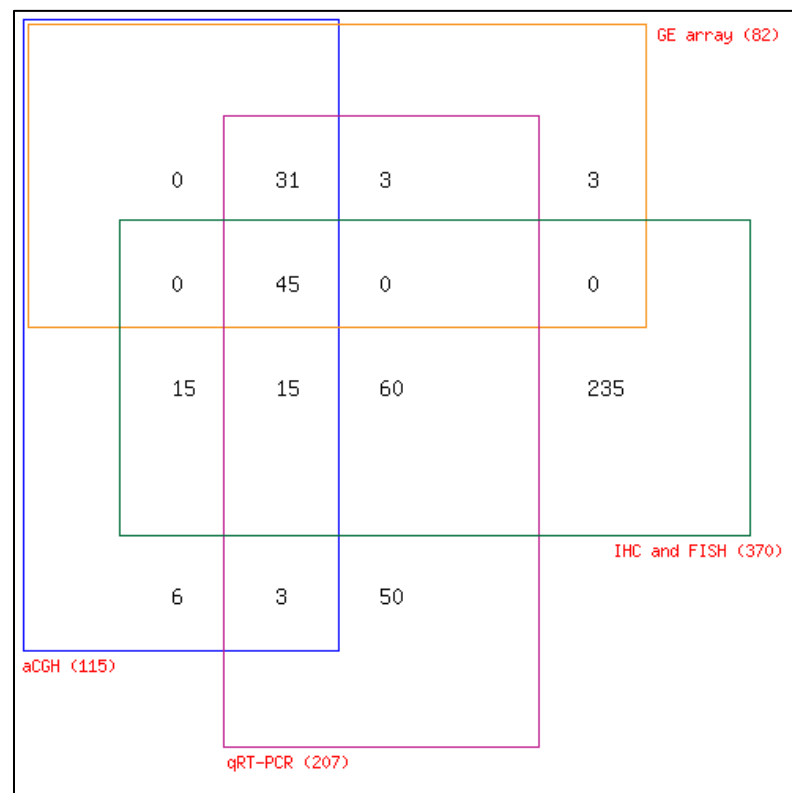
3.1.2 Normal Oral Cavity Tissue and Whole Blood Collection

Non-inflamed gingivobuccal mucosa tissue, from clinically healthy individuals with no previous personal history of cancer were collected from Nair Hospital Dental College, which was used as reference for GE, qRT-PCR and IHC analysis. 10 ml whole blood was collected from healthy individuals (men and women), which was used as gender-matched reference for aCGH analysis.

3.1.3 Sample Sets and Study Design

The number of samples overlapping between different experimental sets and those which form an independent validation set is represented in the Venn diagrams (Figure 10).

A) Total Sample Set (Normal, Leukoplakia and OSCC)



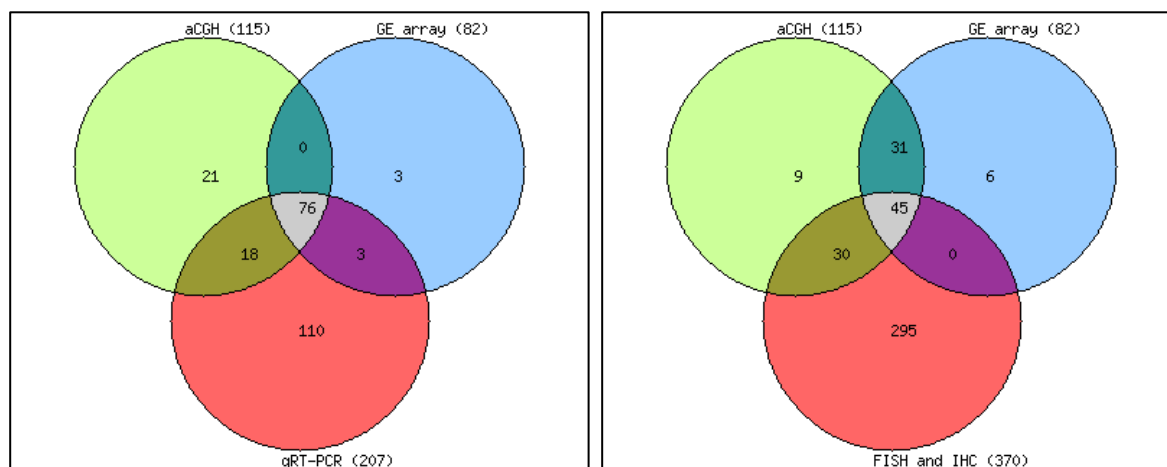
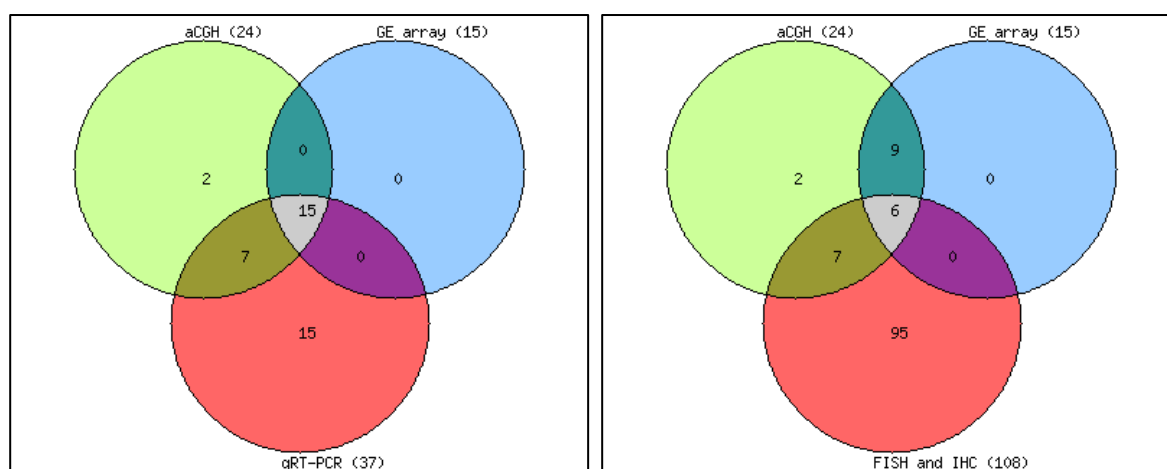
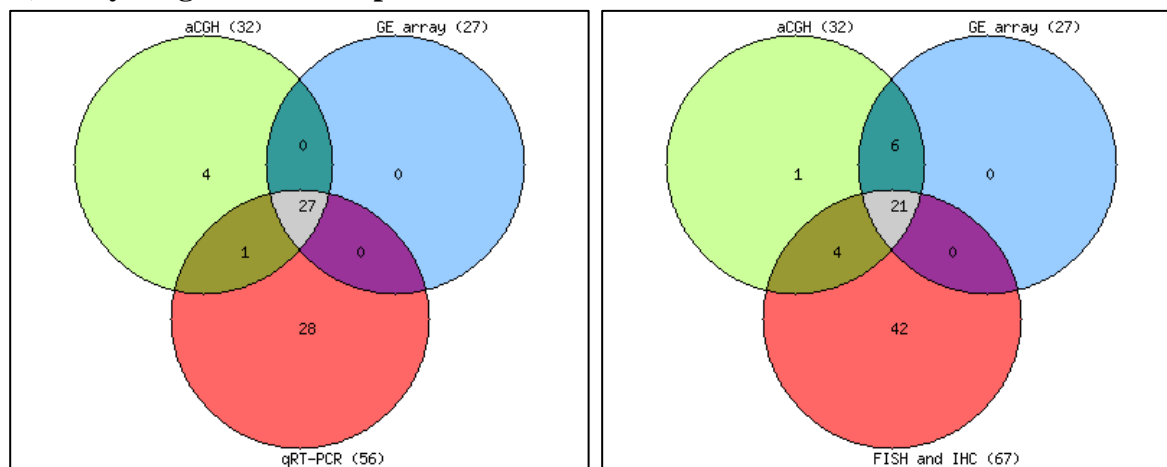
B) Total Sample Set (Normal, Leukoplakia and OSCC)**C) Leukoplakia****D) Early Stage OSCC Samples**

Figure 10: Venn Diagrams Representing Number of Samples Used for aCGH, GE and Validation Analysis. A) and B) 4-way and 3-way venn diagram representing the total number of samples (Normal, leukoplakia, early stage OSCC and Advanced stage OSCC) used in aCGH,

GE, qRT-PCR, IHC and FISH study; **C)** and **D)** 3-way venn diagram representing leukoplakia and early stage OSCC samples used in aCGH, GE and qRT-PCR or IHC/FISH analysis. (Demographic details of each study group are provided in the results section).

3.2 Histopathological Evaluation of Tissues

Fresh frozen leukoplakia and oral tumor tissues were sectioned using cryotome (Leica CM1100, USA). Sections of 5 μ m were mounted on the pre-cleaned slides and stained with Hematoxylin and Eosin (H&E). Histopathological evaluation of leukoplakia and tumor was determined by two pathologists independently. Only tumor tissues containing $\geq 70\%$ of tumor cells and leukoplakia with severe hyperplasia and/ or dysplasia were subjected to nucleic acid extraction.

3.3 DNA Isolation from Whole Blood

Genomic DNA was extracted from 10ml whole blood using QIAamp® DNA Blood Maxi Kit (Qiagen GmbH, Hilden) as per manufacturer's protocol, and eluted in nuclease free water (600 μ l). The quantity of DNA was checked using Nanodrop-1000 spectrophotometer (Thermo Scientific: 1000) and the quality was assessed by electrophoresis on 0.8% agarose gel. Lambda DNA (Invitrogen, Cat#25250-010) suitable for use as a molecular size standard for gel electrophoresis analysis was used to check the extent of genomic DNA degradation.

3.4 Simultaneous DNA/RNA Isolation from Tissues

Simultaneous purification of genomic DNA and total RNA was performed using AllPrep® DNA/RNA/miRNA Universal kit according to manufacturer's instructions (Qiagen, Hilden). Briefly, the 25-30 mg tumor/ leukoplakia/ normal tissues were pulverized by grinding with liquid

nitrogen in a pestle and mortar followed by addition of RLT plus buffer with β -mercaptoethanol (Sigma-Aldrich, St. Louis, MO, USA). The lysate was homogenized by passing through a 20-gauge needle fitted to an RNase-free syringe. The homogenate, so obtained was processed for column purification and isolation of DNA and total RNA as per manufacturer's instruction. Elution of DNA and RNA was done in 100 μ l and 50 μ l respectively with nuclease free water. DNA and RNA quantification was performed on Nanodrop-1000 spectrophotometer (Thermo Scientific: 1000). The Quality of DNA was assessed as described above and the RNA integrity was evaluated using Agilent RNA 6000 Nano Chip, Bioanalyzer (Agilent: 2100).

3.5 HPV Detection

We have screened leukoplakia and gingivobuccal OSCC's for the presence of HPV using multiple methods as described below:

3.5.1 p16 Immunohistochemistry (IHC)

p16 IHC was carried out using the CINtec® Histology Kit (MTM Laboratories, AG, Germany), as per manufacturer's protocol. Five-micron FFPE tumor sections were deparaffinized with xylene, rehydrated with sequential ethanol washes (100%, 90%, and 70%). After antigen unmasking for 20 min at 96 - 99 °C in epitope retrieval buffer, slides were allowed to cool down to room temperature. Endogenous peroxidase blocking was performed with hydrogen peroxide provided with the kit for 5 ± 1 min. Slides were then incubated with the primary antibody (provided with the kit) for 30 min at room temperature. Cervical cancer samples showing high expression of p16 were used as a positive control. Visualization was performed using reagent provided with the kit for 30 min at room temperature followed by development with a substrate—

chromogen solution for 5 - 10 min. The slides were finally counterstained with hematoxylin and examined under a microscope.

3.5.1.1 Immunohistochemical Assessment and Scoring

Two pathologists scored all the samples independently for IHC staining intensity as well as the percentage of cells stained and discordant cases were re-reviewed to arrive at a consensus score. For assessment of p16 protein expression, the cytoplasmic and nuclear staining intensity was categorized based on H- score (range 0 - 300), which was obtained as follows: $H\text{-score} = 1 \times (\% \text{ of cells staining } +1 \text{ intensity}) + 2 \times (\% \text{ of cells staining } +2 \text{ intensity}) + 3 \times (\% \text{ of cells staining } +3 \text{ intensity})$. p16 IHC was scored as positive if there was strong as well as diffuse nuclear and cytoplasmic staining present in greater than 70% of the malignant cells, according to the College of American Pathologist (CAP) criteria that is comparable to the H- score of more than 140. All other staining patterns were scored negative. However, for RNA ISH analysis, tumors with p16 staining in $\geq 5\%$ cells (H - score 50) were processed further.

3.5.2 Nested PCR for HPV Detection

HPV status was determined by nested PCR primers that target conserved HPV L1 region and detects a broad range of HPV subtypes. 300ng of DNA was subjected to the first round of PCR using MY09/MY11 primers to produce 450 bp product [92]. Amplification was performed in a thermal cycler for 40 cycles with following cycling conditions:

- Initial denaturation was for 5 minutes at 95°C,
- Followed by denaturation for 1 minute at 95°C,
- Annealing for 1 minute at 55°C,
- Extension for 1 minute at 72°C,
- Final extension step of 10 minutes at 72°C.

Subsequent PCR was performed using GP5+/GP6+ primers to obtain 150 bp product. MY09/MY11 PCR product was amplified using touchdown-cycling parameters [93]:

- Initial denaturation for 5 minutes at 95°C,
- Followed by denaturation for 1 minute at 95°C,
- Touchdown annealing cycles of 2 minutes with 0.5°C decrement/cycle from 50°C - 40°C (21 cycles),
- Additional annealing cycles 40 for 2 minutes (10 cycles),
- Extension for 1.5 minutes at 72°C and,
- Final extension of 4 minutes at 72°C.

To rule out chances of false positivity, no template control and negative control (C33A DNA) were included while performing all PCR reactions. To avoid cross-contamination of DNA samples, utmost precautions were taken which include strictly maintaining dedicated areas and equipments for DNA extraction and PCR setup.

Beta-Globin		MY09/MY11		GP5+/GP6+	
Components (Stock conc.)	Volume	Components (Stock conc.)	Volume	Components (Stock conc.)	Volume
5XPCR buffer	5.00 µl	5XPCR buffer	5 µl	5XPCR buffer	2.50 µl
10mM dNTP	2.50 µl	5mM dNTP	1 µl	10mM dNTP	1.25 µl
25mM MgCl ₂	1.00 µl	25mM MgCl ₂	4 µl	25mM MgCl ₂	1.75 µl
PC03 (10pM)	1.00 µl	MY09(100ng)	1 µl	GP5+(10pM)	1 µl
PC04 (10pM)	1.00 µl	MY11(100ng)	1 µl	GP6+(10pM)	1 µl
1U/ml Taq	0.25 µl	1U/ml Taq	0.25 µl	1U/ml Taq	0.15 µl
Nuclease free water	9.25 µl	Nuclease free water	4.75 µl	Nuclease free water	12.35 µl
DNA(60ng/µl)	5.00 µl	DNA(60ng/µl)	5.00 µl	DNA(60ng/µl)	5.00 µl
Total reaction volume	25.00 ml	Total reaction volume	25.00 ml	Total reaction volume	25.00 ml

Beta-Globin PCR was carried out with all samples using PC03/PC04 primer set, to verify the amplifiability and integrity of the extracted DNA. Amplification was performed in a thermal cycler for 40 cycles with following cycling conditions:

- Initial denaturation was for 5 minutes at 94°C,
- Followed by denaturation for 30 sec at 94°C,
- Annealing for 30 sec at 51°C,
- Extension for 30 sec at 72°C, and
- A final extension step of 5 minutes at 72°C.

Primers used for PCR amplifications are listed in Table 6. All the PCR products were electrophoresed on 2% agarose gel containing ethidium bromide and visualized under UV light.

Table 6: Primer sequences

Primers	Sequence (5'-3')	Amplicon (bp)
HPV Consensus Primers		
MY09	CGTCCMARRGGAWACTGATC	450
MY11	GCMCAGGGWCATAAYAATGG	
GP5+	TTTGTTACTGTGGTAGATACTAC	150
GP6+	GAAAAATAAACTGTAAATCATATTC	
Beta-Globin primers		
PC03	ACACAACGTGTGTTCACTAGC	110
PC04	CAACTTCATCCACGTTTACC	

3.5.3 RNA *In-Situ* Hybridization (RNA ISH)

Detection of seven high-risk HPV genotype (HPV16, 18, 31, 33, 35, 52 and 58) E6/E7 mRNA was manually performed using the RNAscope™ 2.5 HD assay - Brown HPV HR7 kit (Advanced Cell Diagnostics Inc. CA, USA) according to manufacturer's instructions. Cases that were HPV

DNA positive and /or expressed p16 (low or high) were processed for RNA ISH. Briefly, 5µm FFPE tissue sections were deparaffinized and pre-treated to allow access to target RNA, followed by hybridization for 2 h with target-specific probes for the E6 and E7 genes of high-risk HPV. Ubiquitin C (UBC, a constitutively expressed endogenous gene) was used as an internal positive control and the bacterial gene, dapB, was used as a negative control for background signal. HeLa cells, fixed with 10% formaldehyde/PBS (ACD Positive Control) were used during processing of each batch. Signal detection was done with diaminobenzidine (DAB) for 10 min; the slides were counterstained with hematoxylin and examined under a microscope (20X - 40X magnification). For each case, all 3 stained slides (HPV, UBC, and dapB) were examined simultaneously to determine the HPV status. The UBC test was used to assess the presence of hybridizable RNA; if UBC slide was negative, the sample was disqualified, presuming insufficient RNA quality. The dapB test was used to assess nonspecific staining; only those cases that were negative or had weak staining were considered for HPV scoring. A positive HPV test result was defined as punctate staining that co-localized to the cytoplasm and/or nucleus of the malignant cells [94].

3.6 Microarray Experiments

3.6.1 Array CGH Labeling and Hybridization

Whole-genome copy number profiling was performed on 2x105K CGH oligonucleotide arrays (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. Briefly, 1.5µg of tumor and pooled gender-matched reference DNA were labelled with fluorochromes Cy3 and Cy5 respectively, using Agilent Sure tag DNA labeling kit (Agilent Technologies, 5190-3400). Labelled samples were purified using the genomic DNA purification module

(Agilent Technologies). Briefly, 4.5 μ g of purified labelled samples was mixed with human Cot-1 DNA, and denatured at 95°C (Oligo aCGH hybridization kit, Agilent Technologies, 5188-5380). The mixture was applied to Agilent microarray slides and hybridization was performed at 65°C for 40 hours. After hybridization, the microarrays were washed with Oligo aCGH wash buffer followed by drying of slides. After drying, the arrays were scanned using an Agilent Scanner (Agilent Technologies, G2600D), and log₂-intensities were extracted from raw microarray image files using the Agilent feature extraction software version 11.5 (Agilent Technologies). The raw aCGH data will soon be submitted to Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>).

3.6.1.1 Genome Mapping and Human Structural Variation

Genomic coordinates were standardized to the NCBI build 36 (hg18) assembly of the human genome. Loci of structural and copy number variants were obtained from the Database of Genomic Variants (DGV) version 9 at The Centre for Applied Genomics (TCAG, <http://projects.tcag.ca/variation/>) [95].

3.6.1.2 Array CGH and Survival Analysis

The aCGH microarray data was pre-processed and analyzed with Bioconductor [96] packages implemented in the open-source programming language R [97]. The raw aCGH intensity values were background corrected and normalized using the package *limma* [98], and further segmented with the CBS algorithm, as implemented in the package *DNACopy* [99]. Recurring somatic copy number alterations (CNAs) across samples were called using the tool GISTIC 2.0 [100] with default parameters. GISTIC 2.0 is based on a probabilistic model for estimating the background occurrence rate of CNAs, as well as for defining the boundaries of the CNA regions of interest.

As suggested by the authors of the tool, CNAs were considered significant if their corrected p-value (q-value) was smaller than 0.25. The identified CNA profiles were further separated into arm-level (large) and focal alterations, the latter being the ones considered to drive cancer progression. In addition, GISTIC 2.0 distinguishes between strong amplifications (with a score equal to 2) and weak amplifications (with a score equal to 1), as well as between deletions of either both copies of a gene (with a score equal to -2) or of a single copy (with a score equal to -1), while a score of 0 signifies no change.

Survival analysis was performed by applying Cox proportional hazard regression models, as implemented in the Bioconductor package *survival*. BH-corrected (Benjamini – Hochberg) p-value < 0.25 were considered statistically significant [101]. Associations of CNAs with clinical covariates were tested with Fisher's exact test, and a p-value threshold of 0.05 was used to consider an association as significant.

3.6.2 Gene Expression Microarray (GE)

Whole Human Genome Microarray Kit, 4x44K (Agilent Technologies, Santa Clara, CA) with 43,376 oligonucleotide reporters were used for gene expression study. Briefly, 500ng of samples were labelled with fluorochromes Cy3 using Agilent Quick-Amp labeling Kit (Agilent Technologies, 5190-0442). Labelled cRNA was cleaned up using Qiagen RNeasy columns (Qiagen, Cat No: 74106) and quality assessed for yields and specific activity using the Nanodrop ND-1000. Briefly, 1.65 μ g of the labelled cRNA sample were fragmented at 60°C and hybridized on to microarray slides. Fragmentation of labelled cRNA and hybridization were done using the Gene Expression Hybridization kit (Agilent Technologies; 5190-0404). Hybridization was carried out in Agilent's Surehyb Chambers at 65° C for 16 hours. The hybridized slides were

washed using Agilent Gene Expression wash buffers (Agilent Technologies, 5188-5327) and scanned using the Agilent Microarray Scanner (Agilent Technologies, G2600D) with default parameters followed by feature extraction using FE software version 11.5 (Agilent Technologies). All data generated was MIAME compliant and the raw data will be deposited in the GEO database following the instructions on the MGED Society website (<http://www.mged.org/Workgroups/MIAME/miame.html>).

3.6.2.1 Gene Expression Data Preprocessing

Similar to the array CGH data, the gene expression microarrays were pre-processed and analyzed with Bioconductor packages [96]. Since the Agilent 44K array has many reporters whose expression cannot be uniquely mapped to the expression of a single gene, we restricted our analysis to the 25,505 unambiguously mapped reporters [102]. Following background correction and normalization performed with the package *marray* [103], 16,520 genes were kept for subsequent analyses. All absolute expression intensities were log₂ – transformed.

3.6.2.2 Analysis of Gene Expression Data

For detecting differentially expressed genes, a linear model with a single effect for each gene, as implemented in the package *limma* [98], was fitted to the preprocessed expression intensities. Consequently, each gene was tested for a significant difference in the mean log₂ intensity between two conditions of interest by using a moderated t-test. The resulting p-values were corrected for multiple testing with the Benjamini – Hochberg procedure [104], and the significance threshold on the corrected p-values (q-values) was set to 0.05. For selected sets of genes, gene set enrichment analyses based on hypergeometric testing were performed with the

package *GOstats* [105]. The resulting pathways were categorized into one or more of the following classes: biological processes, cellular components and molecular functions.

3.7 Real-Time PCR Analysis

3.7.1 cDNA Synthesis

Isolated RNA was subjected to cDNA synthesis using “The High Capacity cDNA Reverse Transcription Kit” (Applied Biosystems, Cat#4368814) according to procedure suitable for quantitative PCR amplification as described by the manufacturer. In brief, 2X RT master mix was prepared using kit components as follows:

Kit Component	Volume/Reaction (μl)
10X RT Buffer	2.0
25X dNTP mix (100mM)	0.8
10X RT Random Primers	2.0
Mutiscribe TM Reverse Transcriptase	1.0
RNase Inhibitor	1.0
Nuclease free water	3.2
Total reaction	10.0

The reaction mixture was placed on ice and mixed gently and 10 μl of 2X RT buffer was pipetted into the thin walled PCR reaction tubes. 10 μl of total RNA (1.5 μg) was pipetted in the tubes and was mixed with 2X RT buffer. The tubes were sealed and centrifuged briefly to remove air bubbles and placed on ice until loaded on thermal cycler. The following conditions were used for converting total RNA to cDNA:

	Step 1	Step 2	Step 3	Step 4
Temperature	25°C	37°C	85°C	4°C
Time	10 min	120 min	5 min	∞

The cDNA samples so obtained were stored at -20°C.

3.7.2 Real-Time PCR Analysis

Six ng of cDNA was used for TaqMan qRT-PCR analysis and experiments were performed in duplicates. Assays were done on a QuantStudio 12K Flex (Applied Biosystems, CA) instrument, using assays designed for selected candidate genes found significantly up regulated (*BIRC2*, *BIRC3*, *DVLI*, *EIF5A2*, *FUS*, *HOXC9*, *INHBA*, *LY6K* and *MFAP5*), down regulated (*MAL* and *DERL3*) or unchanged (*SLC4A1AP*). The fluorescent TaqMan probes were obtained from Applied Biosystems and TaqMan assay IDs are provided in Table 7. The results were analysed using QuantStudio 12K Flex software v1.2.2 and Gene Expression Suite software v1.0.4 (Applied Biosystems, CA) and the relative expression of messenger RNA (mRNA) was determined using *18S* ribosomal RNA as an endogenous control. The delta CT (DCt/ΔCT) values representing the difference in the threshold cycles of gene of interest and internal control gene represented by: (CT gene of interest – CT internal control) were compared between leukoplakia/ oral cancer and unrelated normal tissues from the same site. The numerical value of the ΔCT is inversely related to the amount of amplicon in the reaction (i.e., the lower the ΔCT, the greater the amount of amplicon). The gene expression of each sample was analysed using the comparative CT method (also known as the $2^{-\Delta\Delta CT}$ method) where:

$$\Delta\Delta CT = [(CT \text{ gene of interest} - CT \text{ internal control (18S)}) \text{ of the test sample}] - [(CT \text{ gene of interest} - CT \text{ internal control (18S)}) \text{ of the reference sample}].$$

Log fold change for qRT-PCR data was obtained by calculating log values of $2^{-\Delta\Delta CT}$ and was used for comparison with gene expression data.

Table 7: Taqman Assay IDs of Genes Validated by Real Time PCR.

Up-regulated genes		Down-regulated genes	
Assay ID	Gene	Assay ID	Gene
Hs01112284_m1	<i>BIRC2</i>	Hs00405320_g1	<i>DERL3</i>
Hs00985031_g1	<i>BIRC3</i>	Hs00210581_m1	<i>KRT76</i>
Hs00182896_m1	<i>DVL1</i>	Hs00242748_m1	<i>MAL</i>
Hs00702673_s1	<i>EIF5A2</i>	Unchanged gene/ Control	
Hs00192029_m1	<i>FUS</i>	Assay ID	Gene
Hs00396786_m1	<i>HOXC9</i>	Hs00250835_m1	<i>SLC4A1AP</i>
Hs00170103_m1	<i>INHBA</i>	Control gene	
Hs00382690_m1	<i>LY6K</i>	Assay ID	Gene
Hs00185803_m1	<i>MFAP5</i>	Hs99999901_s1	<i>18S rRNA</i>

3.8 Fluorescent *In-Situ* Hybridization (FISH)

Interphase FISH was performed using two different protocols 1) using Bacterial Artificial Chromosomes (BAC clones) for fresh, frozen oral cancer tissues and 2) using SureFISH probes (Agilent Technologies CA, US) for FFPE leukoplakia and oral cancer tissues.

3.8.1 Interphase FISH (FISH) using the BAC Clones

The BAC clones were obtained from the BACPAC resource centre, Children's Hospital Oakland Research Institute, California, USA. The BAC clones were selected based on the regions (11q13.3 and 11q22.2) specifically found altered by our aCGH analysis or were found associated with clinicopathological parameters.

BAC clone	Chromosomal region	Start	End
RP11-135H8	11-Centromere	56357375	56507039
RP11-300I6	11q13.2-q13.3	69162462	69323966
RP11-90M3	11q22.1-q22.2	101500590	101687323

FISH analysis comprised of the following steps:

3.8.1.1 Culturing of BAC Clones and Isolation of Plasmid DNA

Using a sterile toothpick, the single isolated BAC clone was inoculated in 2 ml Luria-Bertini medium (Appendix 7.2) containing 20 µg/ml (Appendix 7.2) chloramphenicol (USB, Cat#23660) in a 12- to 15-ml snap-cap polypropylene tube. The culture was allowed to grow overnight (≤ 16 h) at 37°C in an orbital shaker at 200 rpm. Snap cap was removed from the tube and the medium was centrifuged for 5 min at 1600 X g (4500 rpm) at room temperature. Supernatant was discarded and the pellet was re-suspended in 0.3 ml re-suspension solution (15 mM Tris.Cl, pH 8.0, 10 mM EDTA, 100 mg/ml DNase-free RNase A, sterilized using a 0.2 µm filter and stored at 4°C). Alkaline lysis solution (0.3 ml) was added and the tubes were shaken gently to mix the contents. Tubes were allowed to stand at room temperature ≤ 5 min. While gently shaking, 0.3 ml precipitation solution (3 M potassium acetate, pH 5.5, was autoclaved and stored at 4°C) was added slowly and tubes were allowed to stand on ice ≥ 5 min. Tubes were centrifuged for 15 min at 16,000 X g (14000 rpm) at 4°C to remove the white precipitate. The tubes were removed from the centrifuge and placed on ice. Using a disposable pipette, the supernatant was transferred to a 1.5 ml microcentrifuge tube containing 0.8 ml isopropanol. The contents were mixed by inverting a few times. Tubes were allowed to stand on ice ≥ 5 min and then centrifuged for 15 min at 22800 x g (18000 rpm) at room temperature. Supernatant was

aspirated followed by the addition of 0.5 ml of 70% ethanol. The DNA pellet was washed by inverting tube several times. Tubes were centrifuged for 5 min at 22800 x g (18000 rpm) at 4°C. Supernatant was aspirated and the DNA pellet was air dried at room temperature until the pellet became translucent. Quality and quantity of the plasmid DNA were ascertained by running on 1% agarose gel and spectrophotometric measurements (NanoDrop Technologies, Wilmington, Delaware). The specificity of the probes was ensured by hybridization on normal metaphase spreads (protocol described in 3.13).

3.8.1.2 Probe Labeling and Precipitation

Direct labeling of probe (plasmid DNA) was done using nick translation reaction. The reaction mixture consisted of a 5 µl dNTPs mix, 5 µl enzyme mix [DNase I & Pol (I)], 1.5 µl labelled dUTP, approximately, 2 µg of BAC DNA was used for labeling and volume was made to 50 µl with sterile deionized water. The reaction mixture was incubated at 15°C for 1-2 h. The incorporation of the labelled dUTP in DNA was checked by visualizing the unstained agarose gel with the UV transilluminator (UVP Bioimaging systems, Upland, CA). This gel was subsequently stained with ethidium bromide and the range of the labelled DNA was checked by comparing with λ Hind III digest marker. For interphase FISH, the size of the labelled probes ranged between 300 to 500 bp. After ascertaining the labels in this range, the reaction mixture was incubated at 70°C for 10 min to inactivate the enzymes and to stop the further digestion of plasmid DNA. The chromosome region-specific plasmid DNA were FITC-12-dUTP labelled while their respective centromeric plasmid DNA were labelled with Cy3-dUTP. The labeled probes (chromosome region and centromere specific) were mixed in equal proportion, along with human Cot-1 (Invitrogen, Cat# 15279011) and salmon sperm DNA solution (Invitrogen, Cat#15632011). This mixture was ethanol precipitated and kept at -80°C for 1 h. The labelled

DNA was precipitated by centrifugation at 16,000 X g (14000 rpm) for 30 min. The labelled DNA pellet so obtained was dried, and re-suspended in 12 µl of hybridization buffer (Appendix) and incubated at 37°C for 2-3 h.

3.8.1.3 Pre-treatment of Tissue Specimens

4 µm sections of fresh, frozen tissues were taken on poly-L-lysine (Sigma, Cat#P8920) slides. Slides were placed at -20°C until use. Slides were equilibrated in a coplin jar containing 2X SSC (Appendix) for 30 min at room temperature followed by treatment with a 120µl RNase solution (Appendix), diluted 1:200 from stock solution (Appendix) by incubating slides in a moist hybridization chamber at 37°C for 45 min. Slides were washed 3 times for 5 min each in 2X SSC solution at room temperature. Slides were treated for 5 min at 37°C with pepsin (Fluka, Cat#77152) working solution in pre-warmed 0.01 M HCl. After pepsin treatment slides were washed twice with 1X PBS (Appendix) followed by washing with 1X PBS/ MgCl₂ (0.05M MgCl₂ in 1XPBS) and incubation in 1% Formaldehyde/1XPBS/MgCl₂ for 10 min at room temperature. Slides were then washed with 1X PBS followed by dehydration in ethanol series: 70%, 90%, and 100% ethanol for 3 min each followed by air drying.

3.8.1.4 Denaturation and Hybridization

The slides were denatured for 5 min at 73°C in denaturing solution (Appendix) quenched in pre-chilled 70% ethanol and dehydrated in an ethanol series (70%, 90%, and 100% ethanol for 3 min each). The denatured slide was hybridized with dual color labelled probes which were also denatured at 73°C for 8 min followed by quenching on ice. Overnight hybridization was done in a humidified chamber (Neolab, India) at 39°C.

3.4.1.1 Post-Hybridization Washes

After overnight hybridization, the slides were washed with wash buffer (Appendix) 3 times for 10 min each at 42°C, followed by 0.1X SSC wash twice for 10 min each and detergent wash (0.03% Nonidet P-40 in 2XSSC) for 5 min. The slides were then rinsed with de-ionized water, air dried and counterstained with DAPI containing mounting medium (Vector Lab, Cat#H-1200). Slides were stored in dark at -20°C till visualization.

3.8.2 Interphase FISH Using SureFISH Probe

SureFISH probes (Agilent Technologies, USA) were used to validate alterations at 8q24.3 and 1p36.33 locus.

SureFISH probe	Chromosomal region	Probe length
G101197R-8	8q24.3	370.321 KB
G101034G-8	8-centromere	635.492 KB
G101182R-8	1p36.33	417.415 KB
G101064G-8	1-centromere	540.845 KB

FISH on leukoplakia and OSCC FFPE tissue analysis comprises of following steps:

3.8.2.1 Pre-treatment of Tissue Specimens

Five micron paraffin sections from oral tissue were collected on dry poly-L-lysine coated glass slides (method described in Section 3.11.1). Sections were kept at 56°C for 4-5 h followed by deparaffinization with xylene at 56°C (twice, for 10 minutes each), followed by treatment with Xylene at R.T, twice for 10 minutes each. Slides were dehydrate with 100% Ethanol (twice for 5 min each), followed by pretreatment with 2.5% Sodium Thiocyanate for 10 min with slight

agitation. Slides were air dried and incubate in 100% Ethanol for 2 min. Retrieval of tissues was done in 10mM Sodium Citrate buffer, in microwave at High power (900W) for 15 min.

3.8.2.2 Pepsin Treatment

Pepsin working solution (Appendix) was added on to the tissue sections and incubated in a humid chamber at 37°C for 80 min. Followed by dehydration of slides in series of chilled ethanol grades, 70%, 85%, 100% for 2 min each. Slides were dried on the slide warmer for 2-5min.

3.8.2.3 Probe Preparation and Hybridization (To be done in dark)

Probe hybridization mixture was prepared as described in Table 8, 1µl of each, centromere and locus specific probe used and volume made up to 10 µl, components were thoroughly mixed, and briefly spin it in a microcentrifuge.

Table 8: SureFISH Probe Hybridization Mixture.

Component	Volume per reaction
FISH hybridization buffer	7 µl
Labelled Probe 1	1 µl
Labelled Probe 1	1 µl
Milli-Q® autoclaved	1 µl (enough to bring the final volume to 10 µl)

Approximately 5 µl probe (depending on the size of the section) was added to tissues, followed by denaturation of probe DNA and genomic DNA in a humid chamber at 80°C for 8 min and later incubated in a humidified chamber (Neolab, India) at 39°C for overnight hybridization (16-20 h).

3.8.2.4 Post Hybridization Washes and Counterstaining (To be done in dark)

The slides were incubated in pre-warmed FISH Wash Buffer 1 (Agilent Technologies cat# G9401A) at 73°C for 3 minutes. Followed by treatment with FISH Wash Buffer 2 (Agilent

Technologies cat# G9402A) for 2 min at room temperature. Slides were air dried and counterstained with vectashield mounting media containing DAPI (Vector Lab, Cat#H-1200).

3.8.3 Metaphase FISH

FISH was performed on metaphase spreads to confirm the probe specificity. Metaphase slides (Vysis Inc, IL, USA cat# 32-806010) were equilibrated at RT in 2X SSC for 5 min and treated with 1:200 diluted RNase stock solution for 45 min at 37°C in a moist hybridization chamber. Slides were washed with mild agitation in 2X SSC at RT, followed by treatment with pre-warmed pepsin -0.01 M HCl solution, for 5 min at 37°C. Washing of slides was done with 1X PBS followed by 1X PBS/MgCl₂ for 5 min each at RT with shaking. Finally, slides were incubated in 1% Formaldehyde/1X PBS/MgCl₂ solution for 10 min. Slides were dehydrated in ethanol series: 70%, 90%, 100% ethanol, 3 min each, air dried and taken forward for probe treatment and hybridization as described in previous sections.

3.8.4 Visualization and FISH Signal Enumeration

For visualization of FISH signals, slides were scored using a Zeiss fluorescence microscope (Axioskop II, Germany) equipped with pinkel filter set 83000 or Metasystems D/G/O/GO/DGO filter set 89084 (cat# H-0650-010-CR) (Chroma technology corporation, VT, USA) along with 63X and 100X objectives. For enumeration, hybridization signals from 100 non-overlapping interphase cell nuclei captured at 63X from leukoplakia and tumor samples, were counted manually using a fluorescence microscope. Enumeration for each probe was done using a single filter. The relative number of chromosomal region signals to centromeric signals was evaluated visually. A copy number gain was scored if the average number of signals per nucleus was ≥ 3 to 8. Gene amplification was scored when number of signals per nucleus were > 8 or if tight

clusters of signals in multiple cells were observed per cell. For analysis the alteration (Gain) was categorised into two groups 1) Gain in $\geq 40\%$ cells of the tissue 2) No change (diploid) or gain in $< 40\%$ cells of the tissue (No change category was merged with $< 40\%$ as very few tissues were diploid for the analysed loci). Whereas, amplification was categorised into three groups 1) No change (diploid) 2) Amplification in 5- 20% cells of the tissue 3) Amplification in $\geq 20\%$ cells of the tissue. The cutoffs were set based on the frequency distribution of the percentage of altered cells across the samples (quartile analysis).

3.9 Immunohistochemistry (IHC)

Immunohistochemical analysis was performed to confirm the protein over expression of the validation (protein) targets listed in Table 9. Two different kit based methods were used for IHC analysis:

- 1) Vectastain Universal Elite ABC Kit (Vector labs, Cat#PK-6200) and
- 2) Dako EnVision™ FLEX Mini Kit, High pH (DAKO, Cat#K8023 Kit).

Table 9: Details of Antibodies Used in the Study.

	Target Protein & Antibody detail (Catalog)	Method (Kit)	Retrieval method	Retrieval Buffer	Retrieval time	IHC Antibody dilution and diluent	IF Antibody Dilution	Immuno-blot Antibody Dilution
1	ciAP1; Novus Biologicals (NB100-56128)	Dako EnVision™ FLEX Mini Kit, High pH	Waterbath at 100°C followed by microwave	EnVision™ FLEX Target Retrieval Solution, High pH (Cat#DM828)	20 min (15 min waterbath and 5 min microwave)	1:200, 1x PBS	(1:50)	(1:500)
2	ciAP2; Novus Biologicals (NBP1-	Vectastain Universal elite	Microwave	EDTA buffer pH 8	12 min (6 min twice)	1:1000, 1% serum+	(1:100)	(1:1000)

	27972)	ABC kit				1x PBST		
3	EIF5A2; Biorbyt (orb74027)	Vectasta in Univer sal elite ABC kit	Microwave	EDTA buffer pH 8	10 min (5 min twice)	1:350 , 1x PBS	NA	NA
4	ECT2; Biorbyt (orb5101)	Vectasta in Univer sal elite ABC kit	Microwave	EDTA buffer pH 8	10 min (5 min twice)	1:100 , 1x PBS	NA	NA
5	HOXC9; Biorbyt (orb2164)	Dako EnVision ™ FLEX Mini Kit, High pH	Waterbath at 100°C followed by microwave	EnVision™ FLEX Target Retrieval Solution, High pH (Cat#DM828)	20 min (15 min waterbath and 5 min microwave)	1:100 , 1x PBS	NA	NA
6	HOXC13; Biorbyt (orb15756 9)	Vectasta in Univer sal elite ABC kit	Microwave	EDTA buffer pH 8	10 min (5 min twice)	1:100 , 1x PBS	NA	NA
7	KRT76 (K76); Sigma- Aldrich (HPA0196 96)	Vectasta in Univer sal elite ABC kit	Microwave	Sodium citrate buffer pH 5.8	10 min (5 min twice)	1:225 1xPBS	1;225	NA
8	MFAP5; Sigma- Aldrich (HPA01055 3)	Vectasta in Univer sal elite ABC kit	Microwave	TE buffer pH 9	15 min (5 min- 3 times)	1:200 1xPBS	NA	NA
9	MMP3; Epitomics (1908-1)	Vectasta in Univer sal elite ABC kit	Microwave	Sodium citrate buffer pH 6.4	15 min (5 min- 3 times)	1:100 , 1x PBS	NA	NA
10	NELL2; Novus Biological s (NBP1- 82527)	Vectasta in Univer sal elite ABC kit	Microwave	EDTA buffer pH 8	18 min (6 min- 3 times)	1:75, 1x PBS	NA	NA

Following Steps were Employed IHC Analysis:**3.9.1 Pretreatment and Poly-L-Lysine Coating of Glass Slides**

Glass slides (Sunbeam, India) were dipped overnight in chromic acid solution, followed by washing under running water for 2 h to remove traces of chromic acid. Next the slides were boiled in soap solution for 30 min followed by washing the slides under running water for 1 h. Slides were then washed in deionized water for 30 min with a slow rocking each with three changes. The slides were dehydrated with 70% ethanol for 30 min followed by drying at 55°C for 30 min. Slides were coated in 0.01% (w/v) poly- L-lysine solution (Sigma, Cat#P8920) for 30 min with a slow rocking at room temperature. Slides were dried overnight at 37°C in an incubator.

3.9.2 De-paraffinization of Tissue Sections

Sections were kept at 56°C for 3-4 h followed by deparaffinization with xylene for 15 min twice at room temperature. Slides were transferred to xylene:alcohol (1:1) and were kept for 15 min at room temperature followed by transferring the sections to the coplin jar containing fresh absolute ethanol and subsequently rehydrated in 70% ethanol and 1X PBS for 15 min each.

3.9.3 Antigen Retrieval

Antigen retrieval was carried out using different buffers for respective time depending on the antigen (Table 9). Briefly, slides with tissue sections were boiled in microwave at high power or water bath at 100 °C. Slides were allowed to cool to room temperature followed by two washes with 1X PBS (Appendix) for 10 min each. Endogenous peroxidase activity was blocked by dipping the slides in methanol containing 0.03% H₂O₂ for 30- 45 min or with EnVision™ FLEX

Peroxidase-Blocking Reagent provided with the kit (Dako, Cat#SM801) for 15 min followed by two washes with 1X PBS for 10 min each.

3.9.4 Immunohistochemistry

Each slide had two sections of each tissue, of which one acted as negative control (without primary antibody) and the second one was used as a test. General protocol recommended by Vectastain Universal elite ABC Kit was followed in carrying out immunohistochemical analysis. Briefly, sections were incubated in the normal blocking serum at room temperature for 30 min. Excess serum was blotted from sections and were incubated with primary antibody diluted in 1X PBS. For Dako kit based method, serum blocking was not required hence sections were directly treated with respective primary antibody (details and dilutions provided in Table 9). Sections were incubated in a moist chamber overnight at 4°C. Next day slides were washed twice in 1X PBS or EnVision™ FLEX Wash Buffer (Dako, Cat#DM831) for 10 min each followed by incubating sections with a biotinylated secondary antibody solution or EnVision™ FLEX /HRP (Dako, Cat#SM802) for 30 -45 min at room temperature followed by washing slides twice in 1X PBS or EnVision™ FLEX Wash Buffer for 10 min each. Sections were incubated with Vectastain elite ABC (Vector Labs, Cat#PK6200) reagent for 45 min, ABC treatment was not required for Dako kit based methodology. The step was followed by washing slides twice in 1X PBS for 10 min each. For color development sections were incubated with peroxidase-DAB (Sigma, Cat#D5637) solution (0.8% H₂O₂ +0.8% (w/v) DAB) or EnVision™ FLEX DAB+ Chromogen (Dako, Cat#DM827) until desired stain intensity developed approximately 2- 10 min. Positive controls were always run along with the samples to check the proper color intensity development. DAB color development was stopped by rinsing the slides with deionized water for 10 min. Sections were counterstained with Mayer's hematoxylin followed by rinsing the slides in

deionized water. Tissue sections were dehydrated by passing them through grades of alcohol (70%, 90% and 100%) and xylene. The sections were then mounted with DPX (Merck, Cat#61803502501730). The slides were observed under a light microscope (AxioImager.Z, Carl Zeiss, GmBH) and images were captured using software Axiovision (Carl Zeiss, GmBH). Tissue sections were checked for staining in the cytoplasm, nucleus and cell membrane and compared between normal, leukoplakia and tumors. Based on the color intensity and percentage of tumor stained a score was given to each tissue verified by the pathologist, a mean score was then assigned to each tissue (as described in 3.5.1.1 section).

3.10 Immunofluorescence

Immunofluorescence analysis was performed on tissues and buccal mucosa derived OSCC cell line (SCC29B).

3.10.1 Immunostaining of Tissue

Deparaffinization and antigen retrieval steps were performed as described for IHC. Tissues were fixed in cold methanol for 10 min followed by blocking with 5% normal goat serum (prepared in 1XPBS with 0.3% (v/v) Triton X-100) for 1 h at room temperature. Tissues were next incubated at 4°C (overnight) with the respective antibody at a dilution provided in the Table 9. Next day slides were washed with 1X PBS and 1X PBS + 0.1% NP-40 four times alternately, followed by incubation with an Alexa Fluor 488 anti-rabbit secondary antibody (Life technologies, USA) at 1:200 dilution, in a humid chamber for 1 h at room temperature. The slides were then washed with 1X PBS and 1X PBS + 0.1% NP-40, six times alternately. The tissues were counterstained with 50µl of DAPI for exactly 5 minutes at room temperature, followed by four alternative

washes of 1X PBS and 1X PBS + 0.1% NP-40. The sections were mounted with vectasheild mounting media for florescence (Vector lab Cat# H1000) and viewed under an LSM 780 Carl Zeiss Confocal system with an Argon 488 nm, 561 nm and 405nm lasers (ZEISS, Germany).

3.10.2 Immunostaining of Cultured Cells

Immuno-fluorescence assays were performed to determine the intracellular localization of cIAP1 and cIAP2 in the SCC29B cells. The cells were cultured on glass coverslips at 50-70% confluency. The cells were washed carefully twice with 1X PBS followed by fixation in 4% paraformaldehyde for 20 minutes at room temperature or overnight at 4°C. Thereafter the cells were washed thrice or more with PBS. The cells were then permeabilized using 0.3% Triton-X100 in 1X PBS for 20 minutes at room temperature. Primary antibodies were prepared in 3% BSA in 1X PBS + 0.1% NP-40 solution. 50µl of the primary antibody solution was added on a parafilm and the coverslips were inverted onto the solution and incubated inside a humidified chamber for 16 h (overnight) at 4°C. The cells were then washed with 1X PBS and 1X PBS + 0.1% NP-40 four times alternately. Secondary antibodies were prepared in 3% BSA in 1X PBS + 0.1% NP-40 solution. Cell were treated with the Alexa 488 conjugated secondary antibody anti rabbit IgG (Invitrogen) at a dilution of 1:100 and incubated for an hour at room temperature in a humidifying container. The coverslips were then washed with 1X PBS and 1X PBS + 0.1% NP-40, six times alternately. Thereafter cell were treated with 50µl of DAPI for exactly 5 minutes at room temperature, followed by four alternative washes of 1X PBS and 1X PBS + 0.1% NP-40. The coverslips were then mounted on chromic acid treated, clean glass slides using 10-20 µl of Vectashield mounting agent (Vector Laboratories cat#1000). Confocal images were obtained by using an LSM 780 Carl Zeiss Confocal system with an Argon 488 nm, 561 nm and 405nm lasers (ZEISS, Germany).

3.11 Cell Line, Plasmid Constructs and Transfections

3.11.1 Cell Line Maintenance

Squamous Cell Carcinoma cell lines, SCC29B (UPCI:SCC029B- human buccal mucosa cancer derived) was obtained from The University of Pittsburgh Cancer Institute. The SCC29B cells were cultured in M10 media (Table 10). cIAP1 and cIAP2 knockdown stable cell lines and vector control cells were maintained in same media containing 0.75 μ g/ml of Puromycin for selection (selection media).

3.11.2 Plasmids and Constructs

The shRNAs against cIAP1 and cIAP2 were obtained from Addgene in a pLKO.1 puro vector with AgeI and EcoRI restriction sites. The sequences for the shRNA are in Table 11. The plasmids were purified on cesium gradient and the presence of the shRNA insert was confirmed by sequencing using PLKO.1 sequencing primer.

Table 10: The M10 media formulation.

Component	Complete Media (ml)	Selection Media (ml)
MEM (Minimum Essential Medium) with Earls salt and L-Glutamine (Gibco cat# 11095-080)	500	500
FBS (Gibco cat# 16140071)	55	55
L-Glutamine (Gibco Cat# 25030)	5.55	5.55
Antibiotic antimycotic solution (Sigma Cat# A5955)	0.55	0.55
MEM non essential amino acid solution (NEAA) (Gibco cat# 11140-050)	5	5
Puromycin (Sigma, cat# P8833)	NA	0.75 μ g/ml

Table 11: cIAP1 and cIAP2 shRNA.

Gene	Plasmid from Addgene	shRNA Sequence
<i>BIRC2</i> /cIAP1	pLKO-shcIAP1A (44129)	GCCGAATTGTCTTTGGTGCTT
	pLKO-shcIAP1B (44131)	CAGTTCGTACATTTCTTTCAT
<i>BIRC3</i> /cIAP2	pLKO-shcIAP2A (44130)	GCTGCGGCCAACATCTTCAAA
	pLKO-shcIAP2B (44132)	GCACTACAAACACAATATTC

3.11.3 Transfection

Transfection was done in 12 well plate using Lipofectamine™ 3000 Reagent (Invitrogen, USA). Transfection was carried out as per manufacturer's protocol, at 70-90% cell confluency, using the following components:

Culture Vessel	Vol. Growth Medium	Vol. Opti-MEM for complexing	Plasmid DNA (μg)	P3000™ (μL)	Lipofectamine® 3000 Reagent2 (μL)
12 well	1ml	100 μl	1	2	3

Stable clones for each of the shRNA were generated under puromycin selection (0.75 μg/ml).

3.12 Western Blot Analysis

For Western blots, cells were lysed in 1X sample buffer (Appendix) and protein concentration was quantitated using Folin-Lowry's method (Appendix), 75 μg of the lysate was resolved on 10% SDS-PAGE gel and transferred to PVDF membranes (Amersham Hybondtm-P, GE Healthcare) followed by western blotting with the indicated antibodies. The blots were

developed using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, NJ, USA) according to the manufacturer's instructions.

3.13 *In-vitro* Cell Based Assays

3.13.1 Puromycin Kill Curve

Mammalian cell sensitivity to antibiotics varies from one cell type to another. In order to generate a stable cell line expressing a transgene or shRNA of interest, it is important to determine the minimum concentration of antibiotic required to kill non-transfected (plasmid DNA) cells. The following protocol was used to determine the concentration of antibiotic needed to select SCC29B cells. Cells were grown to 40 to 50% confluency, complete growth media was replaced with selection media supplemented with a range of puromycin concentration (0 μ g, 0.25 μ g, 0.5 μ g, 0.75 μ g, 1 μ g). Cells were monitored daily under microscope for 7 days to get the minimum antibiotic concentration to be used for selection.

3.13.2 Scratch Wound Healing Assays

Scratch wound healing assays were performed in each well of a 6 well plate. Cells were grown to 90% confluency, followed by treatment with 10 μ g/ml of mitomycin C (Sigma) for three hours. Mitomycin C inhibits cell proliferation, thus the rate of wound healing will only depend on the migration rate of cells and not on cell proliferation. Three hours after the addition of mitomycin C, the cells were washed and a linear scratch wound was made on the bottom of plate per well. The wells of this plate were then visualized under an Axiovert 200M Inverted microscope (Carl Zeiss) fitted with a cell incubator stage maintained at 37°C and 5% CO₂. Cells were observed by time lapse microscopy and images taken every 10 minutes for 20 h using the AxioCamMRm

Camera (Zeiss) with a 10X phase I objective. Axiovision software version 4.8 (Zeiss) was used to measure the cell migration. Three independent experiments were performed in triplicates.

3.13.3 Matrigel Cell Invasion Assay

Matrigel invasion assays were performed in 24 well plate using 2×10^5 cIAP2 knockdown cells or vector control cells, resuspended in 200 μ l of serum free media. These cells were added to the upper chambers and total 600 μ l of conditioned media (1:1 ratio of conditioned and fresh serum containing media) was added in the lower chamber. The inner side of the insert with 0.8mm membrane (BD Falcon, NY USA) was pre-coated with 15 μ l of Matrigel (Corning, NY, USA). After 24 h, cell culture inserts were removed from the wells and the cells attached to the inner side of the insert were removed using cotton buds. The inserts with cells on the outer side of the membrane were fixed with 4% para-formaldehyde, stained with 1% crystal violet (Sigma cat# V5265) and mounted on slides using D.P.X mountant (MERCK). Images were taken using Olympus SZ61 stereo microscope using a 10X objective lens. Three independent experiments were performed.

3.13.4 Cell Proliferation Assay

Cell proliferation assay was performed in 96 well plates using 5000 cIAP2 Knockdown cells and the vector control cells. Cell proliferation was checked every 24 h using MTT based assay for 8 days. In dark, cell were treated with 20 μ l MTT (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide) (Sigma Aldrich) for 4 h at 37 °C in CO₂ Incubator. The reaction was stopped using 100 μ l stop solution (10% SDS in 0.01N HCL) plates were incubated

overnight at 37 °C. Absorbance was read at 530nm and at 690nm using microplate reader, SPECTROstar^{Nano} (BMG Lab).

3.14 Statistical Analysis

Statistical analysis was performed using IBM SPSS version 21.

3.14.1 Statistical Analysis of qRT-PCR Data

Shapiro-Wilk test was performed to check the normality distribution of expression levels of differentially expressed genes in normal and tumor samples. The genes which follow a normal distribution were compared using independent samples t-test while the genes not following normal distribution were compared using Mann-Whitney test. A p-value <0.05 was considered statistically significant.

3.14.2 Statistical Analysis of IHC and FISH Data

The Chi-square test or Spearman correlation test, were used to determine the correlation between protein expression levels or locus amplification with disease progression, as well as clinicopathological characteristics. Polytomous logistic regression was used to evaluate the relationship of protein expression scores to the risk of leukoplakia and OSCC development, with normal tissue as a reference; odds ratio (OR) were computed by adjusting for age and gender [106, 107]. Disease-specific survival (DSS) and recurrence free survival (RFS) was examined visually with Kaplan-Meier curves and analysed by log rank tests. Multivariate Cox regression

analysis was performed to know association of various parameters with clinical outcome. All p -values <0.05 were considered statistically significant.

3.14.3 Statistical Analysis of *in-vitro* Assays

All the *in vitro* assays were performed in triplicate. The paired t -test was performed to analyse the relation between vector control and knockdown clones. All p -values <0.05 were considered statistically significant.

4. Results

4.1 The Clinicopathological and Demographic Characteristics

The clinicopathological and demographic characteristics of all the leukoplakia and OSCC patients is summarized in Table 12. The patients included in this study were predominantly male (~ 80%), with a median age of ~45 years (inter quartile range of 33- 62). The major proportion of study patients were smokeless tobacco users (~75%) or had mixed habits (~30%) (chewing along with bidi/cigarette smoking and/or alcohol users). Twenty eight (78%) patients had early stage OSCC (T1N0 or T2N0) and 10 cases with advanced stage OSCC (T1N+ or T2N+) were also included. In addition, raw data (aCGH and GE) of 53 advanced stage OSCC profiled previously in the lab (Ambatipudi, S. et al.,) were used for analysis to get an overall perception of disease progression. In total, study analysis was performed on 24 leukoplakia and 91 OSCC (early and advanced stage OSCC) samples, the detailed clinicopathological characteristics of all the patients along with the follow-up data are provided in the Appendix 7.1.

4.2 Histopathology of Leukoplakia and OSCC

Histopathology of all the samples was confirmed independently by two pathologists (Dr. Anita Borges and Dr. Asawari Patil). All the leukoplakia samples were confirmed to be either hyperplastic (87%) or mild dysplastic (13%). The majority of tumor tissues were moderately differentiated OSCCs (65%). All the tumors used for the microarray study had more than 70% tumor content. Images of the H & E stained leukoplakia and OSCC are represented in Figure 11.

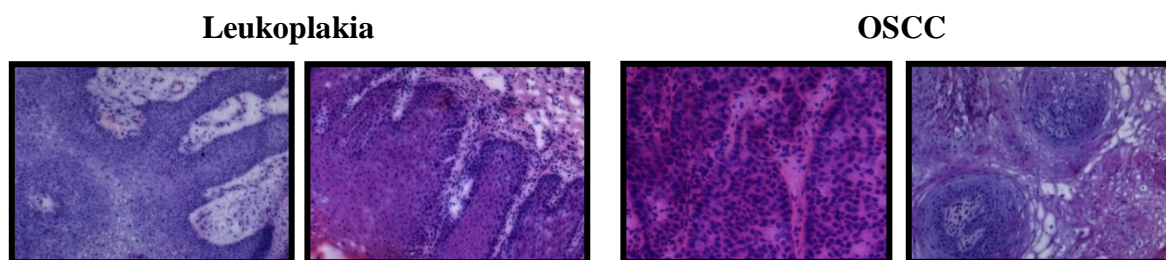


Figure 11: Histopathology of Leukoplakia and OSCC Samples.

Table 12: The clinicopathological characteristic of patients used for aCGH and GE study.

Clinicopathological characteristic	Leukoplakia	Early Stage OSCC	Advanced Stage OSCC (current study)	Samples from previous study used for analysis (Ambatipudi, S. et al.,)
Study (All the samples are from gingivobuccal complex (GBC))				
aCGH	24	28	10	53
Gene expression profiling	15	24	10	27
Age at Diagnosis				
Median	42	50	41	54
Range (IQR)*	38-50	43-62	33-50	44-62
Gender				
Male	21 (87.5%)	22 (78.6%)	7 (70%)	41 (77.4%)
Female	3 (12.5%)	6 (21.4%)	3 (30%)	12 (22.6%)
Pathological Stage				
Stage 1 and 2	NA	28	NA	4 (7.5%)
Stage 3 and 4	NA	NA	10	49 (92.5%)
Pathological T Classification				
T1	NA	3 (10.7%)	2 (20%)	2 (3.8%)
T2	NA	25 (89.3%)	8 (80%)	7 (13.2%)
T3	NA	NA	NA	4 (7.5%)
T4	NA	NA	NA	40 (75.5%)
Pathological Cervical Lymph Node Involvement				
Node Negative (N0)	NA	28	NA	27 (50.9%)
Node Positive (N+)	NA	NA	10	26 (49.1%)
Pathological Grade				
Well	NA	6 (21.4%)	NA	2 (3.8%)
Moderate	NA	14 (50%)	8 (80%)	33 (62.3%)
Poor	NA	8 (28.6%)	2 (20%)	18 (34%)
Hyperplasia	21 (87.5%)	NA	NA	NA
Mild Dysplasia	3 (12.5%)	NA	NA	NA
Moderate Dysplasia	NA	NA	NA	NA
Severe Dysplasia	NA	NA	NA	NA
Primary Treatment Modality				
Surgery	24 (100%)	14 (50%)	5 (50%)	6 (11.3%)
Surgery + RT [†]	NA	12 (42.9%)	5 (50%)	37 (69.8%)
Surgery+ CT ^{††}	NA	NA	NA	NA
Surgery + RT+ CT	NA	2 (7.1%)	NA	10 (18.9%)
Habit Profile				
No Habit	9 (45%)	NA	NA	NA
Exclusive Tobacco Users	3 (15%)	17 (89.5%)	6 (60%)	40 (75.5%)
Exclusive Smoker	NA	NA	NA	2 (3.8%)
Exclusive Alcohol Drinker	NA	NA	NA	NA
Mixed Habit	8 (40%)	2 (10.5%)	4 (40%)	11 (20.8%)
* IQR: Inter Quartile Range, ^{††} CT: Chemotherapy, [†] RT: Radiotherapy, NA: Not Applicable				

4.3 DNA and RNA Extraction and Quality Control

The concentration and length distribution of DNA samples and the quality and concentration of RNA samples were determined. Results of DNA sample separation on agarose gel are displayed on a gel- image (Figure 12), which provides a visualization of DNA quality. For RNA, integrity was assessed by a software algorithm that produces an RNA integrity number (RIN) as displayed on an electropherogram (Figure 13). Leukoplakia and OSCC samples with optimum DNA quality and RNA integrity value (RIN) of more than 7 were used for microarray analysis.

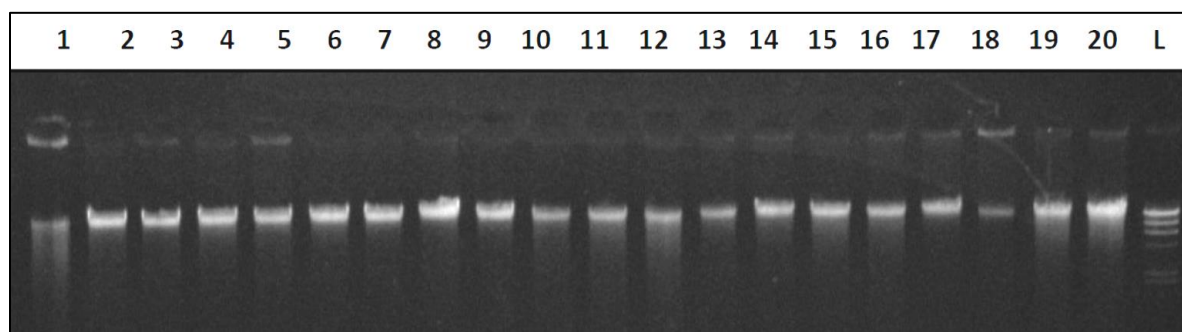


Figure 12: A Representative Gel Picture of Genomic DNA Extracted from Study Samples.

1- 20: genomic DNA from OSCC samples; L: Hind III digested Lambda DNA ladder.

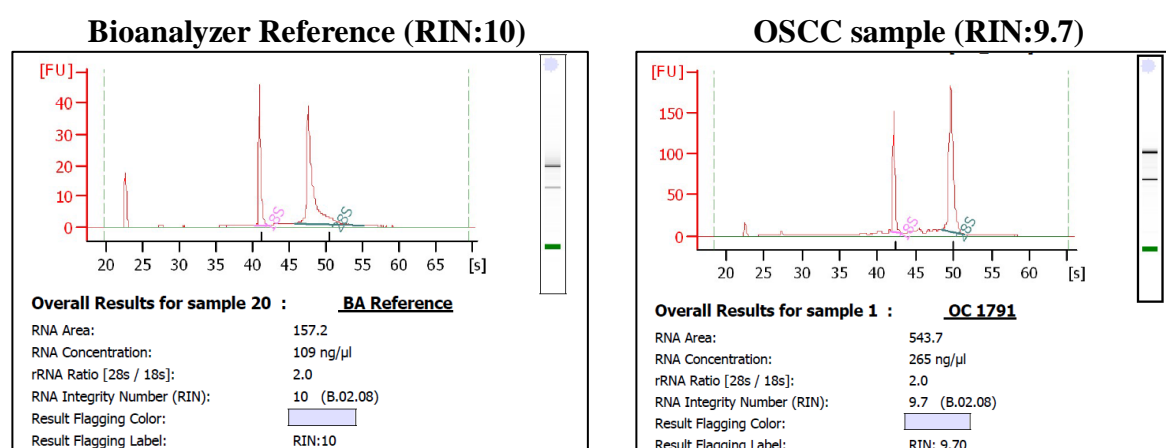


Figure 13: Bioanalyzer Profile (electropherogram) of a Representative RNA Sample Depicting RNA Integrity.

4.4 HPV Detection

We examined all the leukoplakia and OSCC specimens for the presence of high-risk HPV using an algorithm depicted in Figure 14. None of the study samples were positive for p16 expression, HPV DNA and HPV RNA (Figure 15-17).

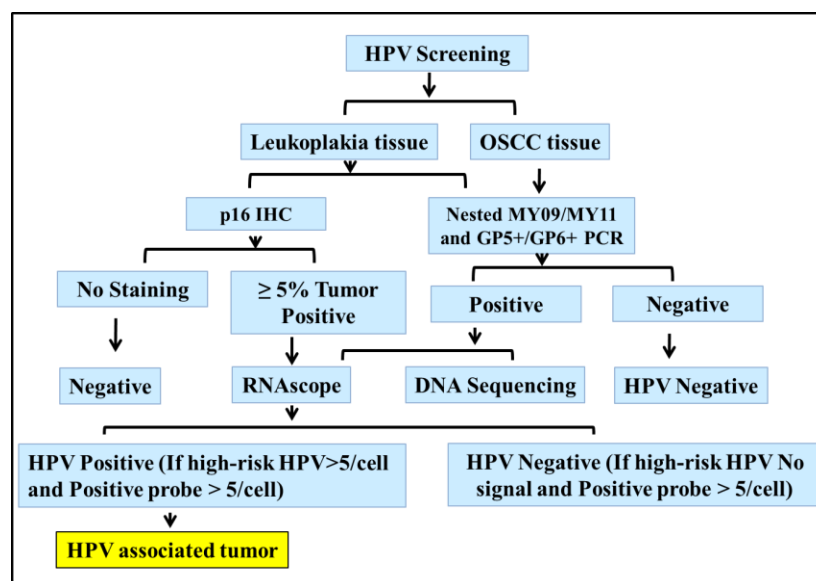


Figure 14: Algorithm for HPV Detection in Patients with Leukoplakia and OSCC.

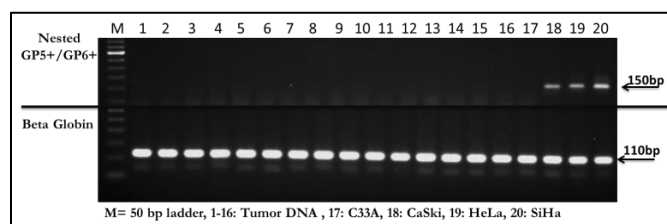


Figure 15: A Representative Gel Image of Nested GP5+/GP6+ PCR Results. β -globin PCR was done to check the genomic integrity in samples. 1-17: HPV DNA negative OSCC samples, 18-20: CaSki (HPV-16), SiHa (HPV-16) and HeLa (HPV-18) cervical cell lines, used as a positive control, M: GeneRuler 50 bp DNA Ladder (Thermo Scientific). Specific band indicated (arrow) at 110bp and 150bp for β -globin and HPV positivity.

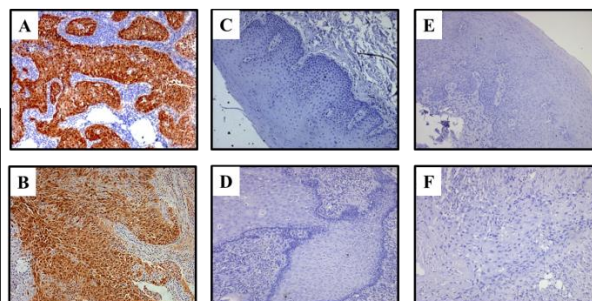


Figure 16: p16 Overexpression in Study Samples. IHC staining for p16 shows nuclear and cytoplasmic expression in positive control (A) and HPV positive oropharyngeal cancer (B); while no staining was observed in leukoplakia (C - Hyperplasia, and E - Hyperplasia with mild dysplasia) and gingivobuccal cancer (D - Well differentiated squamous cell carcinoma and F - Moderately differentiated squamous cell carcinoma). Magnification 100 \times

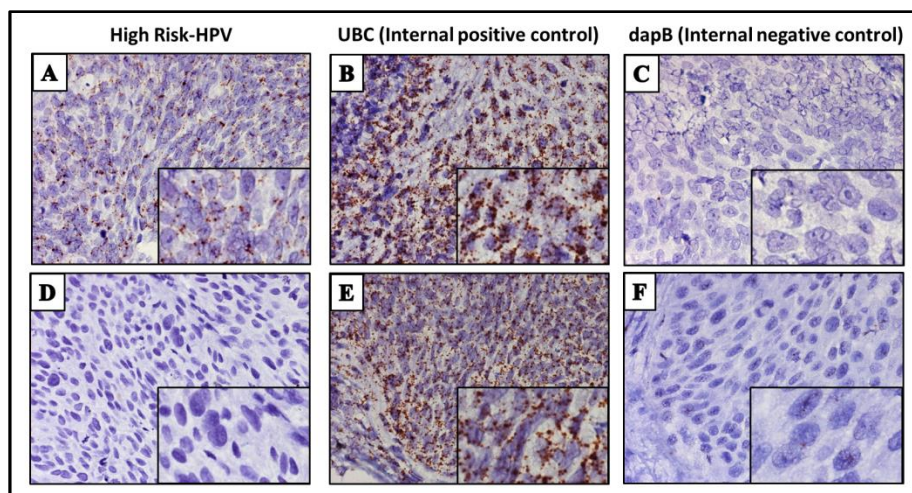


Figure 17. Detection of High-Risk HPV RNA in Study Samples. RNA ISH showing punctate brown signals for HR-HPV RNA in an HPV positive oropharyngeal cancer sample (A), and its respective internal positive control (B). HPV negative OSCC tumor tissue with no signal for HR-HPV RNA (D), and punctate brown signals for internal positive control (E). No signals were observed in internal negative controls (C and F) in either case.

4.5 Genome Wide Analysis of Copy Number Alterations (CNA)

The GISTIC 2.0 tool was used to carry out aCGH analysis. GISTIC (Genomic Identification of Significant Targets in Cancer) identifies the focal somatic copy-number alteration, i.e. regions of the genome that are significantly amplified or deleted across a set of samples. Each aberration was assigned a G-score that considers the amplitude of the aberration as well as the frequency of its occurrence across the samples. A q-value (false discovery rate) or adjusted p-value of 0.25 or less was considered significant. The comparison was carried out using 24 leukoplakia, 28 early stage OSCC (T1N0, T2N0), 10 advanced stage OSCC (T1N+, T2N+) and 53 gingivobuccal cancers reported by us earlier (Ambatipudi, S. et al.,) [these include 4 early stage OSCC (T1N0, T2N0) and 49 advanced stage OSCC (T1N+, T2N+, T3N0, T4N0, T3N+ and T4N+)]. Overall, the study analysis was done with 24 leukoplakia, 32 early stage OSCC and 59 advanced stage

OSCC and will be represented as a combined sample set henceforth. Firstly, the copy number profile of each leukoplakia and OSCC sample was determined, followed by comparisons within groups to identify common and unique alterations. The common alterations represent the altered loci that are identical between the two classes, but this does not necessarily need to span over the exact same coordinates. The unique alterations represent those altered loci which are exclusively seen in one class as compared to another.

The major objective of the aCGH analysis was to identify:

- 1) Complete CNA profile of leukoplakia (n=24) and OSCC (n=91) (represented in Table 15-18).
- 2) Common/unique copy number alterations that may be associated with **disease progression**, i.e. transformation from leukoplakia (n=24) to early stage OSCC (n=32) to advanced stage OSCC (n=59).
- 3) Identifying chromosomal alterations associated with **clinical outcome** (lymph node metastasis, recurrence and survival).

4.5.1 CNA Associated with Disease Progression

The common and unique alterations which may be associated with disease progression were identified by comparison between following groups 1) Leukoplakia v/s Early stage OSCC, 2) Leukoplakia v/s Advanced stage OSCC, and 3) Early stage OSCC v/s Advanced stage OSCC. The total number of CNAs identified were: 19 alterations in leukoplakia, 32 alterations in early stage OSCCs, 70 alterations in advanced stage OSCCs, and 100 alterations in the total OSCC (including both early and advanced OSCC). A very high similarity was observed between the

early stage and the advanced stage cancers, except for the number of alterations which were higher in the advanced stage OSCC.

4.5.1.1 Copy Number Gain/ Amplification

The most frequent gains in leukoplakia and/or OSCC were identified on chromosome regions 1p36.33 (50%), 3q26.31 (31%), 4q13.2 (10%), 6p21.32 (24%), 7p11.2 (18%), 8q24.21 (40%), 8q24.3 (52%), 9p24.1 (20%), 9q34.3 (48%), 11q13.1 (53%), 11q13.3 (30%), 11q22.1 (7%), 12q13.2 (27%), 16p11.2 (29%) as represented in Table 15 and 16. The number of copy number gains increased with disease progression from leukoplakia to advanced stage OSCC (Figure 18 A-C). In addition, the maximum overlap was observed between all the three categories (Figure 18 D). Copy number gains that were in common or unique to individual group are summarized in Table 13. Interestingly gain of 4q13.2, 6p21.32, 8q24.3, 14q11.2, 15q11.1, 16p11.2 was consistent in both leukoplakia and OSCC, suggesting its role in disease progression.

4.5.1.2 Copy Number Loss/ Deletion

The most frequent losses in leukoplakia and/or OSCC were identified on chromosome regions 1q44 (20%), 3p21.1 (41%), 3p14.2 (46%), 4q21.3 (16%), 8p23.2 (56%), 8p11.22 (52%), 9p23 (24%), 9p21.3 (26%), 11q22.3 (17%), 16p11.2 (12%), 17p13.1 (15%), 19p13.3 (21%) as represented in Table 17 and 18. The number of locus which were lost increased with disease progression from leukoplakia to advanced stage OSCC with maximum overlap between all the groups (Figure 19). Interestingly, number of loss/deletions were higher compared to number of gain/amplification in each of the comparisons. Loss/deletions that were common or unique to individual group are summarized in Table 14.

Table 13: Summary of Common and Unique Copy Number Gain/ Amplifications.

Common Gain/ Amplification	
Leukoplakia and Early stage OSCC /Advanced stage OSCC	4q13.2, 6p21.32, 8q24.3 , 14q11.2, 15q11.1, 16p11.2
Early stage OSCC and Advanced stage OSCC	1p36.33 , 4q13.2, 5p15.33, 6p21.32, 7q11.23, 7q22.1, 8q24.3 , 9q34.3, 11p15.5, 11q13.1 , 11q13.3 , 14q32.33, 19p13.3
N0 and N+	3q27.2, 4q13.2, 5p15.33, 6p21.32, 7p11.2, 7q11.23, 8q24.3, 9p24.1, 11q13.1, 11q13.3, 14q32.33, 15q11.1, 17q25.3, 19p13.3
Disease Specific Survival (DSS)	1p36.33 , 1q23.2, 3q27.2, 6p21.1, 11q13.3 , 11q22.1 , 6p13.3, 16p11.2, 16q12.2, 18p11.31, 22q11.21
Recurrence Free Survival (RFS)	1p36.33 , 1q23.2, 2q37.3, 3q27.2, 5p15.33, 11p15.5, 11q13.3 , 16p13.3, 16p11.2, 16q12.2, 16q21, 19p13.3, 22q11.23

Alterations represented in **Bold** font were selected for validation

Unique Gain/ Amplification	
Early stage OSCC compared to Leukoplakia	1p36.33, 1q44, 5p15.33, 7q11.21, 7q11.23, 7q22.1, 9q34.3, 11p15.5, 11q13.1, 11q13.3, 13q21.33, 14q32.33, 19p13.3, 22q11.1, 22q11.21
Advanced stage OSCC compared to Early stage OSCC	3q26.31, 3q27.2, 7p22.3, 7p11.2, 8p23.1, 8q24.21, 9p24.1, 11q22.1, 14q11.2, 15q11.1, 16p13.3, 17q25.3, 22q11.23
Unique to N0	1p36.33, 1q44, 7q22.1, 9q34.3, 11p15.5, 11p11.2, 13q21.33, 14q11.2, 16p13.3, 17p13.1, 22q11.21, 22q11.23
Unique to N+	8p23.1, 8q24.21, 11q22.1

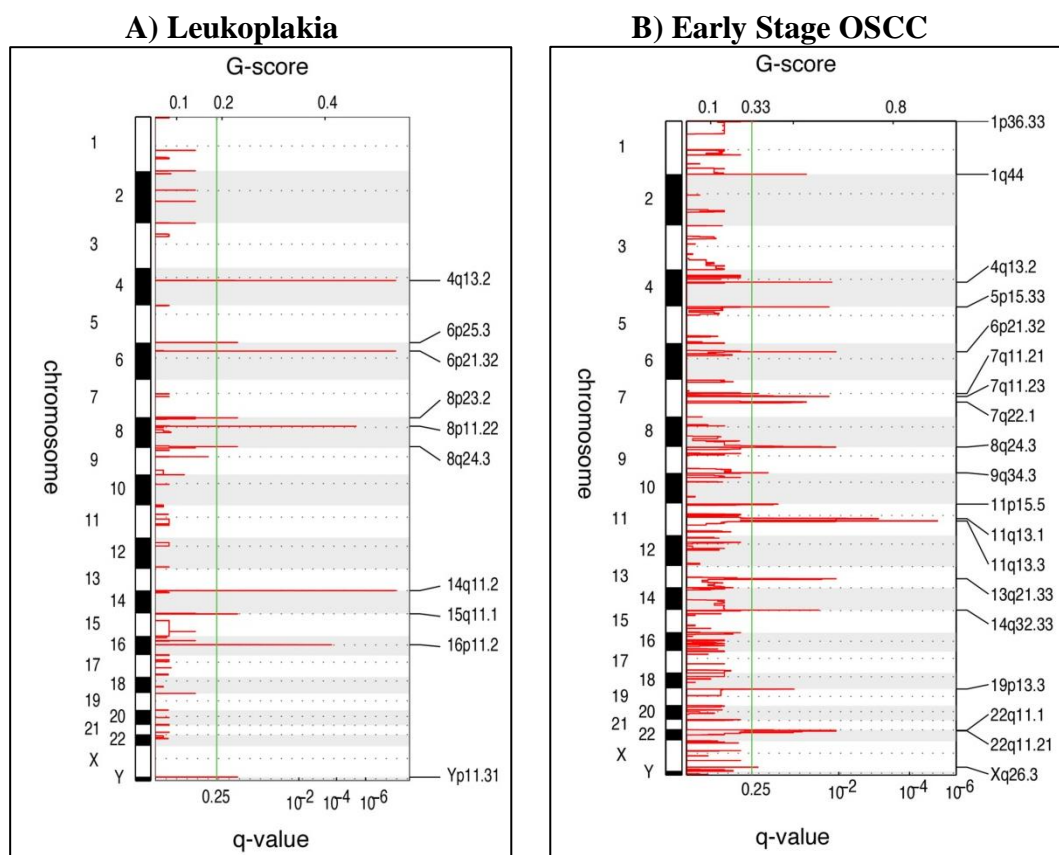
Alterations represented in **Bold** font were selected for validation

Table 14: Summary of Common and Unique Copy Number Loss/ Deletion.

Common Loss/ Deletions	
Leukoplakia and Early stage OSCC /Advanced stage OSCC	1q44, 2q21.2, 8p23.1, 8p11.22, 14q11.2, 15q11.1, 16p11.2, 17q21.31, 19p13.3
Early stage OSCC and Advanced stage OSCC	2q22.1, 3p14.2 , 3p11.1, 4q13.2, 4q35.2, 7q35, 8p23.2, 8p23.1, 8p11.22, 9p21.3 14q11.2, 15q11.1
N0 and N+	1q44, 2q21.2, 2q22.1, 2q34, 3p14.2, 4q35.2, 8p23.2, 8p23.1, 8p22, 8p11.22, 9p23, 9p21.3, 10p11.21, 14q11.2, 15q11.1, 16p11.2, 17q21.31, 18q22.1, 19p13.3
Disease Specific	1q31.3, 2p11.2, 2q22.1, 2q34, 3p14.2 , 3p11.1, 4q13.2, 4q13.2, 4q22.1,

Survival (DSS)	5p14.3, 7q31.1, 9p23, 11q22.3 , 13q21.32, 21q21.3
Recurrence Free Survival (RFS)	2p11.2, 2q22.1, 3p26.3, 3p14.2 , 4q13.2, 4q22.1, 9p23, 9p21.3, 11q22.3 , 14q11.2, 17p13.1
Unique Loss/ Deletions	
Early stage OSCC compared to Leukoplakia	2q22.1, 3p14.2, 3p11.1, 4q13.2, 4q35.2, 7q35, 8p23.2, 9p21.3, 10q11.22, 18q21.33
Advanced stage OSCC compared to Early stage OSCC	1q31.3, 1q44, 2q21.2, 2q23.3, 2q34, 3p26.3, 3p22.3, 3p21.1, 4q13.2, 4q21.3, 5p14.3, 7q31.1, 8p23.2, 8p22, 9p23, 9p12, 10p11.21, 11q22.3, 14q21.1, 16p11.2, 16q23.3, 17p13.1, 17q21.31, 18q12.1, 18q23, 19p13.3, 19p13.11, 21q21.2
Unique to N0	1q31.3, 2q21.2, 3p26.1, 3p22.3, 3p14.3, 3p11.1, 4q13.2, 4q13.2, 4q22.1, 5p14.3, 7q35, 13q21.32, 17p13.1, 18q23, 19p12, 21q21.3
Unique to N+	2q23.3, 3p26.3, 3p12.2, 4q21.3, 7q31.1, 8p23.2, 8p23.2, 9p12, 11q22.3, 18q12.1

Important alterations are represented in **Bold** font



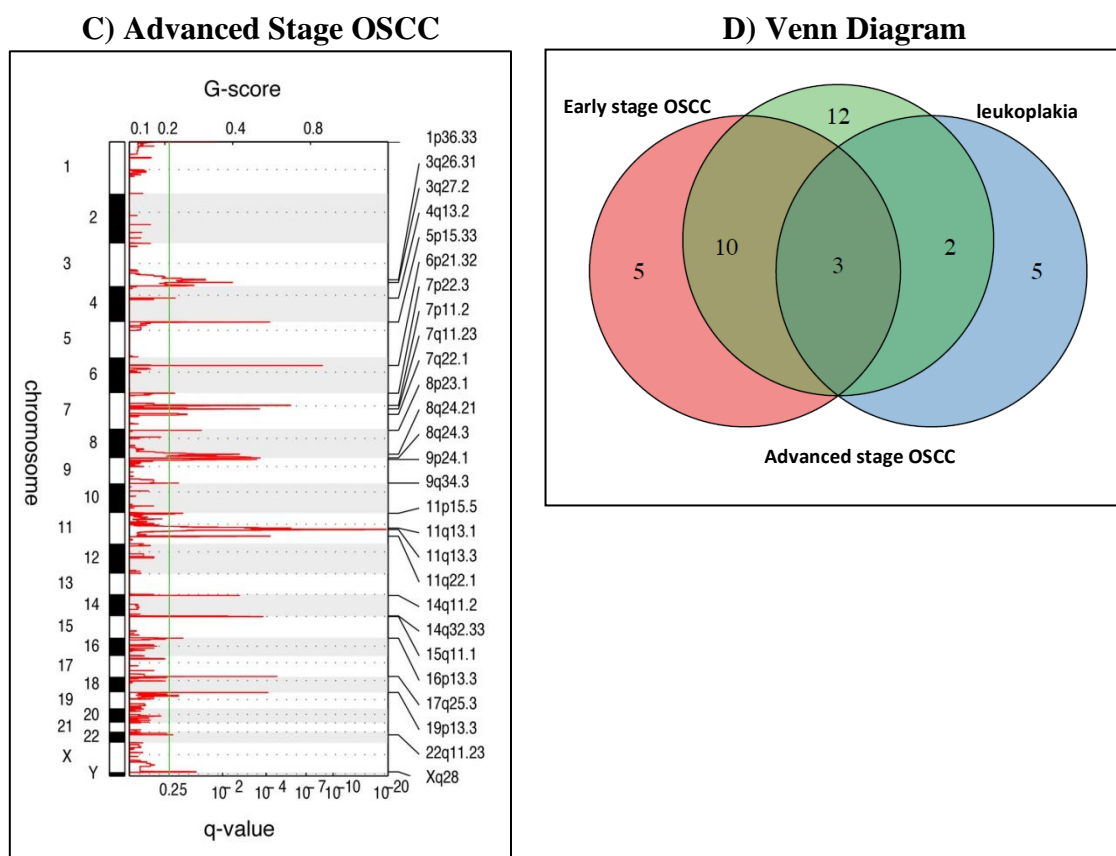


Figure 18: The Amplification Plots And Venn Diagram Depicting Copy Number Gain. The amplification plot depicting the G-scores (top) and q-values (bottom) with respect to amplifications for all markers over the entire region analysed in **A)** Leukoplakia with 10 alterations, **B)** Early stage OSCC with 18 alterations, **C)** Advanced stage OSCC with 27 alterations and **D)** Venn diagram representing a number of unique and overlapping copy number gain in leukoplakia (blue), early stage OSCC (pink) and advanced stage OSCC (green).

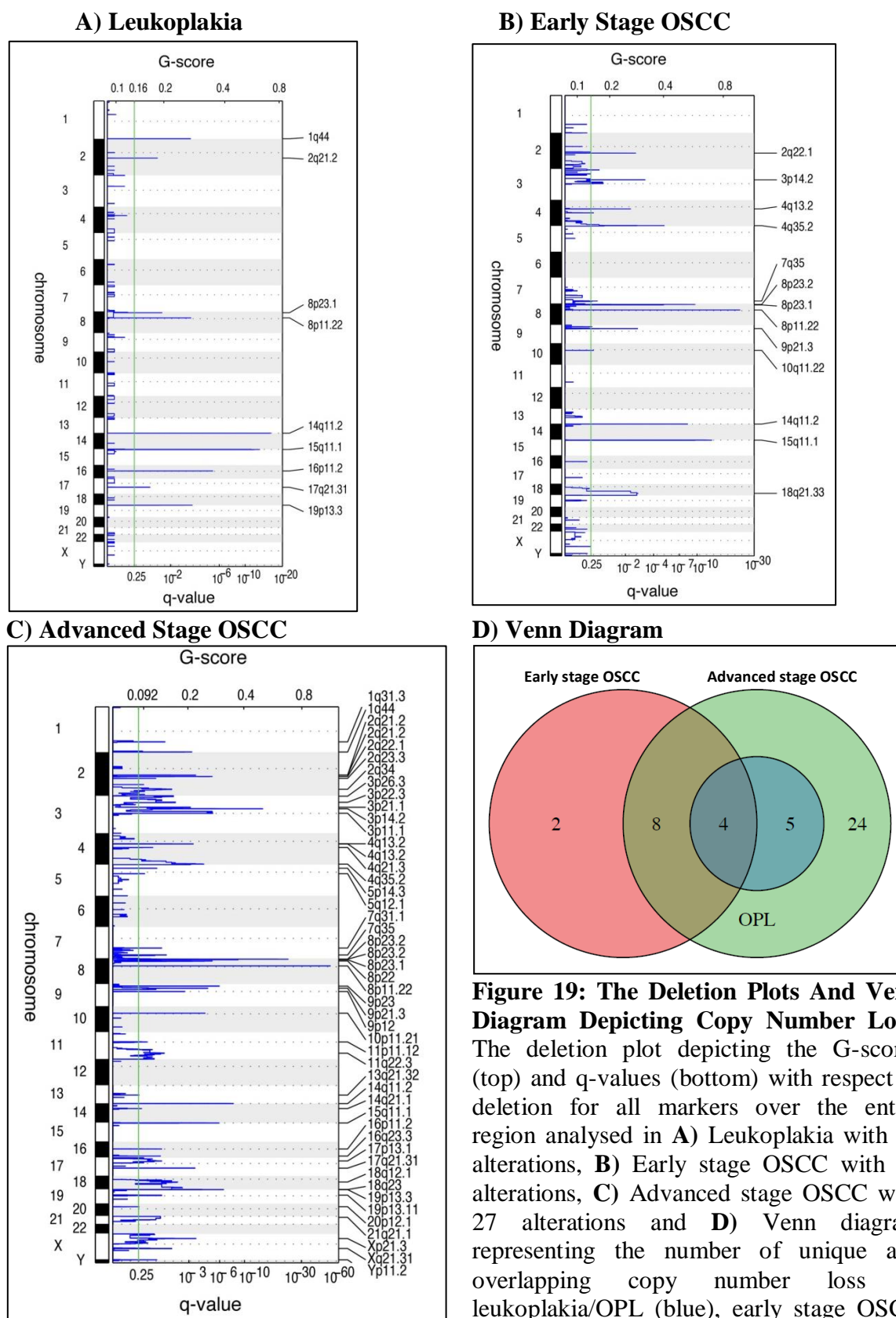


Figure 19: The Deletion Plots And Venn Diagram Depicting Copy Number Loss. The deletion plot depicting the G-scores (top) and q-values (bottom) with respect to deletion for all markers over the entire region analysed in **A) Leukoplakia** with 10 alterations, **B) Early stage OSCC** with 18 alterations, **C) Advanced stage OSCC** with 27 alterations and **D) Venn diagram** representing the number of unique and overlapping copy number loss in leukoplakia/OPL (blue), early stage OSCC (pink) and advanced stage OSCC (green).

Table 15: **Gain** and **Amplification** Observed in Leukoplakia.

Cytoband	q.- values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		Start	End		Gain	Amplification				
4q13.2	5.18E-06	70130976	70296171	0.165195	12.5	12.5			1	<i>UGT2B28</i>
6p25.3	0.13232	1	352462	0.352461	8.3	4.2			1	<i>DUSP22</i>
6p21.32	5.18E-06	32519936	32702994	0.183058	8.3	12.5			3	<i>HLA-DQA1, HLA-DQB1, HLA-DRB1</i>
8p23.2	0.13232	2158300	2511945	0.353645	29.2	4.2			1	<i>[MYOM2]</i>
8p11.22	0.000226	39314621	39535654	0.221033	16.7	12.5			2	<i>ADAM18, LOC100130964</i>
8q24.3*	0.13232	1.42E+08	1.43E+08	0.274854	58.3	0.0			2	<i>PTP4A3, FLJ43860</i>
14q11.2	5.18E-06	18624384	19287127	0.662743	25.0	8.3			1	<i>[OR11H12]</i>
15q11.1	0.13232	18848214	19805959	0.957745	4.2	4.2			1	<i>[CHEK2P2]</i>
16p11.2	0.001412	31835556	33875312	2.039756	33.3	0.0			6	<i>ZNF267, TP53TG3, LOC390705, TP53TG3C, LOC729264, TP53TG3B</i>

^aMb: mega base pair, * Targets selected for validation

Table 16: **Gain** and **Amplification** Observed in OSCC (Including Early stage and Advanced stage OSCC)

Cytoband	q-values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes (Genes represented in Bold font are known in OSCC)
		start	End		Gain	Amplification				
1p36.33*	0.000187	1179224	1488410	0.309186	50.4	0.9			22	<i>DVL1</i> , <i>SCNN1D</i> , <i>SSU72</i> , <i>MXRA8</i> , <i>CPSF3L</i> , <i>AURKAIP1</i> , <i>MRPL20</i> , <i>ATAD3A</i> , <i>VWA1</i> , <i>GLTPD1</i> , <i>CCNL2</i> , <i>TASIR3</i> , <i>ATAD3B</i> , <i>ACAP3</i> , <i>UBE2J2</i> , <i>PUSL1</i> , <i>LOC148413</i> , <i>ATAD3C</i> , <i>TMEM240</i> , <i>FAM132A</i> , <i>ANKRD65</i> , <i>TMEM88B</i>
1q23.2	0.201460	159718935	159768405	0.04947	24.3	0.9			1	<i>DUSP23</i>
1q44	0.014885	247071282	249250621	2.179339	28.7	0.0	1	hsa-mir-3124	61	<i>ZNF124</i> , <i>TRIM58</i> , <i>AHCTF1</i> , <i>OR1C1</i> , <i>OR2M4</i> , <i>OR2L2</i> , <i>OR2L1P</i> , <i>OR2T1</i> , <i>ZNF692</i> , <i>ZNF695</i> , <i>ZNF669</i> , <i>ZNF672</i> , <i>SH3BP5L</i> , <i>OR2G3</i> , <i>OR2G2</i> , <i>OR2C3</i> , <i>ZNF496</i> , <i>ZNF670</i> , <i>NLRP3</i> , <i>OR2M5</i> , <i>OR2M3</i> , <i>OR2T12</i> , <i>OR14C36</i> , <i>OR2T34</i> , <i>OR2T10</i> , <i>OR2T4</i> , <i>OR2T11</i> , <i>OR2B11</i> , <i>C1orf150</i> , <i>LOC148824</i> , <i>OR2T6</i> , <i>PGBD2</i> , <i>OR2L13</i> , <i>OR14A16</i> , <i>VNIR5</i> , <i>OR6F1</i> , <i>OR2W3</i> , <i>OR2T8</i> , <i>OR2T3</i> , <i>OR2T29</i> , <i>C1orf229</i> , <i>OR2M1P</i> , <i>OR11L1</i> , <i>OR2L8</i> , <i>OR2AK2</i> , <i>OR2L3</i> , <i>OR2M2</i> , <i>OR2T33</i> , <i>OR2M7</i> , <i>OR2G6</i> , <i>OR2T2</i> , <i>OR2T5</i> , <i>OR14I1</i> , <i>OR2T27</i> , <i>OR2T35</i> , <i>OR2W5</i> , <i>OR13G1</i> , <i>LOC646627</i> , <i>MIR3124</i> , <i>MIR3916</i> , <i>ZNF670-ZNF695</i>
2q37.3	0.078565	242514594	243199373	0.684779	23.5	2.6			10	<i>DTYMK</i> , <i>PDCD1</i> , <i>ATG4B</i> , <i>THAP4</i> , <i>GAL3ST2</i> , <i>ING5</i> , <i>NEU4</i> , <i>CXXC11</i> , <i>D2HGDH</i> , <i>LOC728323</i>
3q26.31	0.019042	164101777	185392160	21.290383	31.3	3.5	3	hsa-mir-1224, hsa-mir-569, hsa-mir-551b	107	<i>ACTL6A</i> , <i>BCHE</i> , <i>AP2M1</i> , <i>CLCN2</i> , <i>DVL3</i> , <i>ECT2</i> , <i>EHFADH</i> , <i>EIF4G1</i> , <i>EPHB3</i> , <i>MECOM</i> , <i>GHSR</i> , <i>NDUFB5</i> , <i>CLDN11</i> , <i>SERPINI1</i> , <i>SERPINI2</i> , <i>PIK3CA</i> , <i>PLD1</i> , <i>POLR2H</i> , <i>PRKCI</i> , <i>PSMD2</i> , <i>SI</i> , <i>SKIL</i> , <i>SLC2A2</i> , <i>SOX2</i> , <i>TERC</i> , <i>THPO</i> , <i>SEC62</i> , <i>FXR1</i> , <i>CHRD</i> , <i>TNFSF10</i> , <i>EIF2B5</i> , <i>USP13</i> , <i>MAP3K13</i> , <i>ECE2</i> , <i>ABCC5</i> , <i>ALG3</i> , <i>KCNMB2</i> , <i>IGF2BP2</i> , <i>PDCD10</i> , <i>SLITRK3</i> , <i>NLG1</i> , <i>TNIK</i> , <i>MCF2L2</i> , <i>ATP11B</i> , <i>VPS8</i> , <i>GPR160</i> , <i>LAMP3</i> , <i>KCNMB3</i> , <i>GOLIM4</i> , <i>ZNF639</i> , <i>PEX5L</i> , <i>DCUN1D1</i> , <i>KLHL24</i> , <i>ABCF3</i> , <i>PARL</i> , <i>MFN1</i> , <i>YEATS2</i> , <i>MYNN</i> , <i>EIF5A2</i> , <i>MCCC1</i> , <i>MRPL47</i> , <i>NCEH1</i> , <i>SLC7A14</i> , <i>SEN2</i> , <i>GNB4</i> , <i>MAGEF1</i> , <i>ZMAT3</i> , <i>FNDC3B</i> , <i>TBL1XR1</i> , <i>ZBBX</i> , <i>LRRC31</i> , <i>MAP6D1</i> , <i>PHC3</i> , <i>SPATA16</i> , <i>B3GNT5</i> , <i>ARPM1</i> , <i>KLHL6</i> , <i>VWA5B2</i> , <i>TMEM41A</i> , <i>CAMK2N2</i> , <i>DNAJC19</i> , <i>FAM131A</i> , <i>TTC14</i> , <i>WDR49</i> , <i>LRRC34</i> , <i>HTR3C</i> , <i>LIPH</i> , <i>HTR3D</i> , <i>RPL22L1</i> , <i>NAALADL2</i> , <i>HTR3E</i> , <i>C3orf70</i> , <i>CCDC39</i> , <i>LOC339926</i> , <i>LRRIQ4</i> , <i>SAMD7</i> , <i>SOX2-OT</i> , <i>TMEM212</i> , <i>FLJ46066</i> , <i>LOC646168</i> , <i>SNORD66</i> , <i>MIR551B</i> , <i>MIR569</i> , <i>LOC100128164</i> , <i>MIR1224</i> , <i>LOC100505687</i> , <i>MIR4789</i>

Cytoband	q-.values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		start	End		Gain	Amplification				
3q27.2	0.001845	185508349	185590254	0.081905	38.3	3.5			1	<i>IGF2BP2</i>
4q13.2	0.000056	68896317	69665978	0.769661	9.6	7.0			8	<i>UGT2B15</i> , <i>UGT2B17</i> , <i>TMPRSS11E</i> , <i>YTHDC1</i> , <i>TMPRSS11B</i> , <i>TMPRSS11F</i> , <i>SYT14L</i> , <i>LOC550113</i>
5p15.33	0.000000	763495	873415	0.10992	35.7	7.8			2	<i>BRD9</i> , <i>ZDHH11</i>
6p21.32	0.000000	32519936	32689774	0.169838	24.3	13.9			3	<i>HLA-DQA1</i> , <i>HLA-DQB1</i> , <i>HLA-DRB1</i>
6p21.1	0.107740	41328140	44375395	3.047255	33.0	0.9			71	<i>BYSL</i> , <i>CCND3</i> , <i>CDC5L</i> , <i>SLC29A1</i> , <i>GUCA1A</i> , <i>GUCA1B</i> , <i>HSP90AB1</i> , <i>MDFI</i> , <i>MEA1</i> , <i>NFKBIE</i> , <i>PEX6</i> , <i>PGC</i> , <i>POLH</i> , <i>PPP2R5D</i> , <i>PTK7</i> , <i>PRPH2</i> , <i>SRF</i> , <i>TBCC</i> , <i>VEGFA</i> , <i>TFEB</i> , <i>MED20</i> , <i>POLR1C</i> , <i>MAD2L1BP</i> , <i>CUL7</i> , <i>C6orf108</i> , <i>CNPY3</i> , <i>FRS3</i> , <i>SLC22A7</i> , <i>CAPN11</i> , <i>CUL9</i> , <i>UBR2</i> , <i>KIAA0240</i> , <i>ZNF318</i> , <i>YIPF3</i> , <i>USP49</i> , <i>GNMT</i> , <i>PRICKLE4</i> , <i>MRPL2</i> , <i>GTPBP2</i> , <i>MRPS18A</i> , <i>MRPS10</i> , <i>TMEM63B</i> , <i>TRERF1</i> , <i>AARS2</i> , <i>XPO5</i> , <i>MRPL14</i> , <i>DLK2</i> , <i>TBTK1</i> , <i>RRP36</i> , <i>ABCC10</i> , <i>KLC4</i> , <i>TJAP1</i> , <i>FOXP4</i> , <i>KLHDC3</i> , <i>TAF8</i> , <i>PTCRA</i> , <i>TCTE1</i> , <i>SPATS1</i> , <i>C6orf223</i> , <i>RSPH9</i> , <i>LRRC73</i> , <i>RPL7L1</i> , <i>SLC35B2</i> , <i>CRIP3</i> , <i>C6orf226</i> , <i>TMEM151B</i> , <i>C6orf132</i> , <i>LOC100132354</i> , <i>TOMM6</i> , <i>MIR4647</i> , <i>MIR4641</i>
7p22.3	0.039463	844146	962084	0.117938	33.9	3.5			3	<i>ADAP1</i> , <i>SUN1</i> , <i>GET4</i>
7p11.2	0.000000	54809237	55556229	0.746992	18.3	7.0			4	<i>EGFR</i> , <i>SEC61G</i> , <i>LANCL2</i> , <i>VOPPI</i>
7q11.23	0.000000	75900918	76028008	0.12709	38.3	5.2			5	<i>HSPB1</i> , <i>YWHAG</i> , <i>ZP3</i> , <i>SRCRB4D</i> , <i>SRRM3</i>
7q22.1	0.001274	99338841	100713947	1.375106	41.7	1.7	1	hsa-mir-106b	58	<i>ACHE</i> , <i>AZGP1</i> , <i>CYP3A4</i> , <i>EPHB4</i> , <i>EPO</i> , <i>GNB2</i> , <i>AGFG2</i> , <i>LRCH4</i> , <i>MCM7</i> , <i>PCOLCE</i> , <i>PMS2P1</i> , <i>TAF6</i> , <i>TFR2</i> , <i>TRIP6</i> , <i>ZAN</i> , <i>ZNF3</i> , <i>ZKSCAN1</i> , <i>ZSCAN21</i> , <i>AP4M1</i> , <i>MUC12</i> , <i>POP7</i> , <i>STAG3</i> , <i>COPS6</i> , <i>FBXO24</i> , <i>PILRB</i> , <i>PILRA</i> , <i>ACTL6B</i> , <i>SRRT</i> , <i>ZCWPW1</i> , <i>C7orf43</i> , <i>MEPCE</i> , <i>SLC12A9</i> , <i>MOSPD3</i> , <i>GIGYF1</i> , <i>CYP3A43</i> , <i>PVRIG</i> , <i>GAL3ST4</i> , <i>OR2AE1</i> , <i>TSC22D4</i> , <i>TRIM4</i> , <i>MUC17</i> , <i>PPP1R35</i> , <i>GPC2</i> , <i>NYAP1</i> , <i>CNPY4</i> , <i>MBLAC1</i> , <i>GJC3</i> , <i>GATS</i> , <i>C7orf59</i> , <i>C7orf61</i> , <i>UFSP1</i> , <i>MIR106B</i> , <i>MIR25</i> , <i>MIR93</i> , <i>SPDYE3</i> , <i>LOC100129845</i> , <i>SAP25</i> , <i>MIR4658</i>
8p23.1	0.011954	6929692	8117270	1.187578	9.6	4.3	1	hsa-mir-548i-3	21	<i>DEFB4A</i> , <i>SPAG11B</i> , <i>DEFB103B</i> , <i>DEFB104A</i> , <i>DEFB105A</i> , <i>DEFB106A</i> , <i>DEFB107A</i> , <i>FLJ10661</i> , <i>LOC349196</i> , <i>DEFB103A</i> , <i>DEFB107B</i> , <i>DEFB104B</i> , <i>DEFB106B</i> , <i>DEFB105B</i> , <i>SPAG11A</i> , <i>FAM66B</i> , <i>ZNF705G</i> , <i>FAM66E</i> , <i>LOC100132396</i> , <i>DEFB4B</i> , <i>MIR548I3</i>

Cytoband	q-values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		start	End		Gain	Amplification				
8p11.22	0.003901	39314621	39402432	0.087811	7.8	5.2			0	
8q24.21	0.003148	128199768	131486526	3.286758	40.0	10.4	4	hsa-mir-1208, hsa-mir-1207, hsa-mir-1205, hsa-mir-1204	15	<i>MYC, POU5F1B, PVT1, ASAP1-IT1, ASAP1, FAM49B, GSDMC, LOC727677, LOC728724, MIR1205, MIR1206, MIR1207, MIR1204, MIR1208, LOC100507117</i>
8q24.3*	0.000000	145568164	145718224	0.15006	52.2	13.9	2	hsa-mir-1234, hsa-mir-939	13	<i>TONSL, FOXH1, FBXL6, CPSF1, CYHR1, VPS28, SLC39A4, GPR172A, KIFC2, ADCK5, C8ORFK29, MIR939, MIR1234</i>
9p24.1	0.000049	5348183	5500694	0.152511	20.9	4.3			2	<i>CD274, C9orf46</i>
9q34.3	0.014246	139324937	139439577	0.11464	48.7	0.9			5	<i>NOTCH1, SEC16A, INPP5E, C9orf163, MIR4673</i>
11p15.5	0.002527	1	2246595	2.246594	47.0	1.7	4	hsa-mir-483, hsa-mir-675, hsa-mir-4298, hsa-mir-210	85	<i>AP2A2, CD151, CTSD, DRD4, DUSP8, HRAS, IGF2, INS, IRF7, LSP1, MUC2, MUC6, POLR2L, PSMD13, RNH1, MRPL23, RPLP2, SCT, TALDO1, TH, TSPAN4, TNNI2, TNNT3, RASSF7, IFITM1, BRSK2, IFITM3, DEAF1, IFITM2, PKP3, SIRT3, IGF2-AS1, BET1L, CEND1, CDHR5, TOLLIP, PIDD, PNPLA2, PHRF1, SIGIRR, RIC8A, EPS8L2, CHID1, SLC25A22, ATHL1, PTDSS2, MOB2, SYT8, ODF3, LRRC56, LOC143666, SCGB1C1, NLRP6, NS3BP, LOC255512, C11orf35, H19, EFCAB4A, TMEM80, ANO9, LOC338651, B4GALNT4, PDDC1, KRTAP5-1, KRTAP5-3, KRTAP5-4, IFITM5, FAM99A, IFITM10, MIR210, KRTAP5-5, KRTAP5-2, KRTAP5-6, MIR483, SNORA52, LOC653486, INS-IGF2, MUC5B, MIR675, FAM99B, LOC100133161, MRPL23-AS1, MIR4298, MIR210HG, MIR4686</i>
11p11.2	0.041452	47087936	47252031	0.164095	38.3	1.7			4	<i>DDB2, PACSIN3, C11orf49, ARFGAP2</i>
11q12.1	0.122930	56995495	57130488	0.134993	27.0	0.9			4	<i>APLNR, P2RX3, SSRP1, TNKS1BP1</i>
11q13.1	0.000000	65388768	65597429	0.208661	53.9	5.2			9	<i>OVOL1, RELA, SIPA1, KAT5, RNASEH2C, DKFZp761E198, PCNXL3, MIR4489, MIR4690</i>
11q13.3*	0.000000	69728479	69954920	0.226441	29.6	12.2			1	<i>ANO1</i>
11q22.1*	0.000123	101283818	101896670	0.612852	7.0	3.5			4	<i>TRPC6, KIAA1377, ANGPTL5, MIR3920</i>

12q13.2	0.162300	54264564	56621079	2.356515	27.8	0.0	3	hsa-mir-148b, hsa-mir-615, hsa-mir-196a-2	78	<i>CD63, CDK2, DGKA, ERBB3, BLOC1S1, NCKAP1L, HNRNPA1, HOXC4, HOXC5, HOXC6, HOXC8, HOXC9, HOXC10, HOXC11, HOXC12, HOXC13, ITGA5, ITGA7, MMP19, MYL6, NFE2, PA2G4, PDE1B, PPP1R1A, RAB5B, RDH5, RPL41, RPS26, PMEL, SMARCC2, SUOX, KIAA0748, RNF41, GDF11, COPZ1, ESYT1, CBX5, SMUG1, ZNF385A, ORMDL2, GPR84, NEUROD4, IKZF4, OBFC2B, WIBG, SARNP, ZC3H10, DNAJC14, LACRT, DCD, MUCL1, OR10P1, GTSF1, OR10A7, MYL6B, METTL7B, OR6C74, OR6C3, OR6C6, OR6C2, OR6C4, OR6C1, OR6C75, OR6C76, OR6C70, LOC400043, OR6C65, OR6C68, MIR196A2, FLJ12825, OR9K2, MIR148B, MIR615, HOTAIR, LOC100240734, LOC100240735, BLOC1S1-RDH5, MIR3198-2</i>
13q21.33	0.051647	71348128	75076910	3.728782	6.1	1.7			7	<i>KLF5, DACH1, PIBF1, KLF12, DIS3, BORA, MZT1</i>
14q11.2	0.000331	21739284	22034703	0.295419	33.0	3.5			9	<i>SALL2, TOX4, SUPT16H, METTL3, RPGRIPI, CHD8, RAB2B, SNORD8, SNORD9</i>
14q32.33	0.000007	105080341	105239423	0.159082	30.4	6.1			5	<i>AKT1, SIVA1, INF2, ADSSL1, MIR4710</i>
15p11.1	0.000005	18848214	19130747	0.282533	13.9	7.0			1	<i>[CHEK2P2]</i>
15q11.1	0.000000	19537036	20249885	0.712849	13.9	7.8			1	<i>[CHEK2P2]</i>
16p13.3	0.015655	672049	742357	0.070308	40.0	2.6			10	<i>RHBDL1, STUB1, RAB40C, WDR24, C16orf13, FAM195A, RHOT2, WFIKKN1, WDR90, JMJD8</i>
16p11.2	0.150340	28206302	31427794	3.221492	39.1	0.0	1	hsa-mir-762	115	<i>ALDOA, ATP2A1, CD19, CLN3, CTF1, SEPT1, FUS, ITGAD, ITGAL, ITGAM, ITGAX, KIF22, MAZ, PHKG2, PPP4C, MAPK3, PRSS8, SPN, SULT1A2, STX4, SULT1A1, SULT1A3, TBX6, TUFM, DOC2A, HIRIP3, EIF3C, BCL7C, TAOK2, ZNF646, SETD1A, RNF40, MVP, BCKDK, CD2BP2, CDIPT, SRCAP, CORO1A, ATXN2L, SEPHS2, XPO6, ZNF629, QPRT, SH2B1, TBC1D10B, SEZ6L2, NUPR1, LAT, PYCARD, MYLPF, ZNF771, FBXL19, APOBR, FBR5, ZNF747, PRR14, VKORC1, SLX1B, DCTPP1, GPD3, C16orf53, ZNF768, ZNF668, RABEP2, HSD3B7, YPEL3, FAM57B, SPNS1, KAT8, NFATC2IP, C16orf93, ZNF764, ORAI3, PRRT2, STX1B, CCDC101, ZNF689, TMEM219, C16orf92, ZNF785, ZNF688, PRSS36, ZNF48, IL27, KCTD13, ASPHD1, PYDC1, C16orf54, INO80E, FBXL19-AS1, PRSS53, SBK1, LOC388242, LOC440354, LOC440356, SULT1A4, TRIM72, SLX1A, BOLA2, LOC595101, LOC606724, LOC613037, LOC613038, ZG16, BOLA2B, SNORA30, EIF3CL, LOC100289092, MIR762, SLX1A-SULT1A3, SLX1B-SULT1A4, MIR4519, MIR4721, MIR4518, MIR4517</i>

Cytoband	q-.values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		start	End		Gain	Amplification				
16p11.2	0.072579	32558527	33395113	0.836586	24.3	1.7			4	TP53TG3, LOC390705, TP53TG3C, TP53TG3B
16q12.2	0.112060	55182079	55306392	0.124313	29.6	1.7			1	[IRX6]
16q21	0.186490	65443319	66440288	0.996969	32.2	0.9			2	CDH5, LOC283867
16q24.2	0.217470	87181439	90354753	3.173314	33.9	0.9			59	APRT, C16orf3, CA5A, CBFA2T3, CDH15, CYBA, DPEP1, FANCA, GALNS, GAS8, MC1R, MVD, CHMP1A, RPL13, SPG7, SLC7A5, CDK10, C16orf7, PIEZO1, TUBB3, PRDM7, TCF25, ZCCHC14, CPNE7, ILI7C, ANKRD11, TRAPPC2L, KLHDC4, DEF8, BANP, JPH3, DBNDD1, FBXO31, CDT1, MAP1LC3B, SPIRE2, ZNF469, CENPBD1, ZNF276, RNF166, SPATA2L, C16orf55, ZC3H18, SLC22A31, ZFPM1, MGC23284, ZNF778, ACSF3, LINC00304, SNAI3, CTU2, PABPN1L, LOC400558, SNORD68, LOC100128881, LOC100130015, LOC100287036, C16orf95, MIR4722
17p13.1	0.127250	6957178	7833837	0.876659	27.0	0.9	1	hsa-mir-324	62	ACADVL, ASGR1, ASGR2, ATP1B2, CD68, CHD3, CHRN1, CLDN7, DLG4, DVL2, EFN3, EIF4A1, EIF5A, FGF11, GPS2, POLR2A, SHBG, SLC2A4, SOX15, TP53, TNK1, TNFSF13, TNFSF12, KCNAB3, FXR2, MPDU1, ACAP1, CLEC10A, GABARAP, KDM6B, CTDNEP1, C17orf81, SENP3, SNORA67, YBX2, WRAP53, PLSCR3, NLGN2, ZBTB4, TRAPPC1, PHF23, LSMD1, NEURL4, TMEM88, SAT2, CYB5D1, DNAH2, KCTD11, C17orf74, C17orf61, LOC284023, TMEM102, TMEM95, SPEM1, TNFSF12-TNFSF13, MIR324, SLC35G6, SNORA48, SNORD10, SCARNA21, C17orf61-PLSCR3, SENP3-EIF4A1
17q25.3	0.000003	77422699	77546255	0.123556	37.4	5.2			1	RBFOX3
18p11.31	0.095700	2769189	4194789	1.4256	11.3	0.9			11	TGIF1, MYOM1, DLGAP1, LPIN2, MYL12A, SMCHD1, EMILIN2, MYL12B, LOC201477, FLJ35776, LOC727896
18q11.1	0.218900	18802816	18993432	0.190616	2.6	1.7			1	GREB1L
19p13.3	0.000000	1	306944	0.306943	31.3	8.7	1	hsa-mir-1302-11	6	PPAP2C, MIER2, OR4F17, FLJ45445, FAM138F, FAM138A
20q13.33	0.089631	60247697	60396261	0.148564	41.7	2.6			1	CDH4
22q11.21	0.000871	19482826	19685337	0.202511	21.7	4.3			3	CLDN5, CDC45, LOC150185
22q11.23	0.004285	23962240	24268561	0.306321	16.5	6.1			12	MIF, MMP11, SMARCB1, ZNF70, VPREB3, C22orf43, SLC2A11, DERL3, C22orf15, RGL4, LOC284889, CHCHD10

* Targets selected for validation

Table 17: Loss and Deletion Observed in Leukoplakia.

Cytoband	q- .values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes (Genes represented in Bold font are known in OSCC)
		Start	End		Loss	Deletion				
1q44	0.000561	2.47E+08	2.47E+08	0.159776	20.8	0.0			2	<i>SCCPDH, CNST</i>
2q21.2	0.040485	1.34E+08	1.34E+08	0.193917	8.3	0.0			1	<i>NCKAP5</i>
8p23.1	0.025689	6929692	8132138	1.202446	16.7	0.0	1	hsa-mir-548i-3	21	<i>DEFB4A, SPAG11B, DEFB103B, DEFB104A, DEFB105A, DEFB106A, DEFB107A, FLJ10661, LOC349196, DEFB103A, DEFB107B, DEFB104B, DEFB106B, DEFB105B, SPAG11A, FAM66B, ZNF705G, FAM66E, LOC100132396, DEFB4B, MIR548I3</i>
8p11.22	0.000524	39314621	39561021	0.2464	12.5	0.0			2	<i>ADAM18, LOC100130964</i>
14q11.2	8.42E-17	19154993	19497022	0.342029	54.2	0.0			2	<i>OR11H12, LOC642426</i>
15q11.1	7.62E-14	19418601	19819960	0.401359	50.0	0.0			1	<i>[CHEK2P2]</i>
16p11.2	5.61E-06	31321913	31446721	0.124808	25.0	0.0			4	<i>COX6A2, ITGAD, ITGAM, ITGAX</i>
17q21.31	0.079462	41500727	41983406	0.482679	16.7	0.0	1	hsa-mir-2117	10	<i>DHX8, DUSP3, ETV4, MEOX1, MPP2, MPP3, SOST, CD300LG, C17orf105, MIR2117</i>
19p13.3	0.00042	1	306944	0.306943	29.2	0.0	1	hsa-mir-1302-11	6	<i>PPAP2C, MIER2, OR4F17, FLJ45445, FAM138F, FAM138A</i>

^aMb: mega base pair

Table 18: Loss and Deletion Observed in OSCC (Including Early stage and Advanced stage OSCC).

Cytoband	q- .values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		Start	End		Loss	Deletion				
1q31.3	0.001520	1.95E+08	1.95E+08	0.081449	23.5	0.0			1	[<i>KCNT2</i>]
1q44	0.000000	2.47E+08	2.47E+08	0.087646	20.9	0.0			1	<i>CNST</i>
2p11.2	0.075644	89401839	90999011	1.597172	11.3	0.0			1	[<i>MIR4436A</i>]
2q21.2	0.000000	1.34E+08	1.34E+08	0.103608	20.9	0.0			1	<i>NCKAP5</i>
2q22.1	0.000000	1.42E+08	1.42E+08	0.207662	25.2	0.0			1	<i>LRP1B</i>
2q23.3	0.020653	1.55E+08	1.55E+08	0.099873	16.5	0.0			1	[<i>GALNT13</i>]
2q34	0.000958	2.09E+08	2.15E+08	5.64206	27.8	0.0	1	hsa-mir-548f-2	22	<i>ACADL, CPS1, CRYGA, ERBB4, IDH1, MAP2, MYL1, PTH2R, RPE, LANCL1, IKZF2, CPS1-IT1, SPAG16, C2orf67, PIKFYVE, UNC80, C2orf80, LOC100130451, MIR548F2, LOC100507443, MIR4776-1, MIR4776-2</i>
3p26.3	0.000462	1	8810684	8.810683	40.9	0.0			23	<i>CAV3, GRM7, IL5RA, ITPR1, OXTR, SETMAR, BHLHE40, EDEM1, CHL1, CNTN6, LINC00312, LMCD1, C3orf32, TRNT1, CRBN, ARL8B, LRRN1, CNTN4, SUMF1, EGOT, LOC100288428, LOC100507582, MIR4790</i>
3p22.3	0.000958	36166865	36664590	0.497725	40.9	0.0			1	<i>STAC</i>
3p21.1	0.000003	53848562	55517736	1.669174	41.7	0.0			8	<i>WNT5A, CHDH, IL17RB, CACNA2D3, LRTM1, SELK, ACTR8, ESRG</i>
3p14.2	0.000000	60462002	60541505	0.079503	46.1	0.0			1	<i>FHIT</i>
3p11.1	0.000000	89920002	90102303	0.182301	42.6	0.0			1	[<i>EPHA3</i>]
4q13.2	0.000000	68901211	69665978	0.764767	20.9	0.0			8	<i>UGT2B15, UGT2B17, TMPRSS11E, YTHDC1, TMPRSS11B, TMPRSS11F, SYT14L, LOC550113</i>
4q13.2	0.000015	70130976	70237777	0.106801	23.5	0.0			1	<i>UGT2B28</i>
4q21.3	0.031563	87334548	87494719	0.160171	16.5	0.0			1	<i>MAPK10</i>
4q22.1	0.054478	91894845	92154163	0.259318	22.6	0.0			1	<i>FAM190A</i>
4q35.2	0.000000	1.88E+08	1.88E+08	0.259721	30.4	0.0			1	[<i>FAT1</i>]
5p14.3	0.031563	21966990	22281768	0.314778	13.0	0.0			2	<i>CDH12, PMCHL1</i>
5q12.1	0.144450	59018370	59187585	0.169215	18.3	0.0			1	<i>PDE4D</i>
7q31.1	0.038277	1.11E+08	1.11E+08	0.114281	21.7	0.0			1	<i>IMMP2L</i>
7q35	0.000892	1.44E+08	1.44E+08	0.293871	24.3	0.0			9	<i>FAM115A, OR2F1, OR2A14, OR6B1, OR2F2, OR2A12, OR2A25, OR2A5, OR2A2</i>

Cytoband	q-values	Probe Position		Size(^a Mb)	% Frequency		Number of miRNA	miRNA	Number of Genes	Genes
		Start	End		Loss	Deletion				
8p23.2	0.000000	3698219	3793503	0.095284	56.5	0.0			1	<i>CSMD1</i>
8p23.2	0.000000	3929984	4042116	0.112132	51.3	0.0			1	<i>CSMD1</i>
8p23.1	0.000000	7729312	8132138	0.402826	43.5	0.0	1	hsa-mir-548i-3	7	<i>DEFB4A, DEFB103B, FLJ10661, DEFB103A, FAM66E, LOC100132396, MIR548I3</i>
8p22	0.000003	14106992	14413935	0.306943	43.5	0.0			1	<i>SGCZ</i>
8p11.22	0.000000	39341525	39561021	0.219496	52.2	0.0			2	<i>ADAM18, LOC100130964</i>
9p23	0.000000	9612218	9725747	0.113529	24.3	0.0			1	<i>PTPRD</i>
9p21.3	0.000000	21910519	21993650	0.083131	26.1	0.0			2	<i>CDKN2A, C9orf53</i>
9p12	0.002821	42709535	44199400	1.489865	9.6	0.0			9	<i>LOC286297, FOXD4L4, FAM75A6, ANKRD20A3, ANKRD20A2, LOC642929, CNTNAP3B, FOXD4L2, FAM95B1</i>
10p11.21	0.000027	34686003	34906112	0.220109	26.1	0.0			1	<i>PARD3</i>
11p11.12	0.188670	49594804	50156130	0.561326	18.3	0.0			3	<i>OR4C13, OR4C12, LOC440040</i>
11q22.3	0.069519	83351528	1.35E+08	51.65499	17.4	0.0	9	hsa-mir-3167, hsa-mir-100, hsa-mir-4301, hsa-mir-34c, hsa-mir-1260b, hsa-mir-548l, hsa-mir-1304, hsa-mir-1261, hsa-mir-3166	390	<i>ACAT1, ACRV1, BIRC2, BIRC3, APLP2, APOA1, APOA4, APOC3, ARCNI, ATM, FXYD2, CXCR5, CASP1, CASP4, CASP5, CBL, CD3D, CD3E, CD3G, CTSC, CHEK1, CRYAB, DDX6, DDX10, DLAT, DLG2, DPAGT1, DRD2, ETS1, FDX1, FLI1, FUT4, SLC37A4, GRIA4, GRIK4, GRM5, GUCY1A2, H2AFX, HMBS, HSPA8, HSPB2, HTR3A, IL10RA, IL18, STT3A, KCNJ1, KCNJ5, VWA5A, MCAM, MLL, MMP1, MMP3, MMP7, MMP8, MMP10, MMP12, MMP13, MRE11A, MTNR1B, NCAM1, NFRKB, NNMT, NPAT, NRG, OPCML, PAFAH1B2, PGR, TMEM123</i> <i>POU2AF1, PPP2R1B, PTS, PVRL1, RDX, RPS25, SC5DL, SCN2B, SCN4B, SDHD, ST3GAL4, SLN, SORL1, SRPR, ST14, TAGLN, TECTA, THY1, TRPC6, TYR, UPK2, ZBTB16, ZNF202, CUL5, PICALM, FZD4, BARX2, OR7E2P, JRKL, EED, ZNF259, MTMR2, USP2, PCSK7, HTR3B, ZW10, MMP20, UBE4A, MED17, EI24, FEZ1, CEP57, ARHGAP32, C2CD2L, NAALAD2, RBM7, MPZL2, YAP1, HYOU1, ATP5L, ME3, GPR83, SRSF8, ADAMTS8, PRSS23, TREH, CEP164, IGSF9B, ENDOD1, EXPH5, PHLDB1, SIK2, NCAPD3, ARHGEF12, SIK3, VSIG2, BACE1, TRIM29, RAB38, CADM1, PANX1, POU2F3, HINFP, REXO2, OR8G2, OR8B8, OR8G1, TIMM8B, OR8B2, CHORDC1, ACAD8, B3GAT1, DCPS, C11orf54, ZBTB44, THYN1, DDX25, NOX4, NTM, CDON, SIDT2, TRAPPC4, C11orf73, CWC15, SPA17, FXYD6, CNTN5, SIAE, C11orf71, ROBO4, SLC35F2, RAB39A, BTG4, FAM55D, SYTL2, ANKRD49, TTC12, C11orf57, ELMOD1, FOXRED1, KDM4D, SCN3B, VPS11, TMEM126B, TEX12, CRTAM,</i>

										<i>TMPRSS4, IFT46, C11orf75, PRDM10, TRIM49, DSCAML1, GRAMD1B, KIAA1377, ARHGAP20, USP28, CREBZF, CARD18, CCDC81, AASDHPPT, PKNX2, TP53AIP1, MMP27, ABCG4, ROBO3, C11orf1, TMEM135, TAF1D, RNF26, FAM118B, DYNC2H1, NLRX1, C11orf61, CCDC82, ALG9, CLMP, PDZD3, C11orf63, CCDC15, PDGFD, TMPRSS5, PUS3, MFRP, JAM3, BCO2, TMEM133, TMPRSS13, TMEM126A, DCUN1D5, KIAA1826, MAML2, KIRREL3, BUD13, TMEM25, RPUSD4, TBRG1, UBASH3B, C11orf70, DIXDC1, KIAA1731, ZC3H12C, GLB1L2, ESAM, ALKBH8, FDXACB1, C11orf52, VPS26B, GLB1L3, TIRAP, CARD16, C1QTNF5, TMEM123, PANX3, APOA5, SLC36A4, FAT3, TRIM64, TMEM45B, C11orf93, PIH1D2, FAM55A, FAM55B, AMICA1, FAM76B, SESN3, PIWIL4, ARHGAP42, KBTBD3, CWF19L2, KDELC2, LAYN, TTC36, AMOTL1, CCDC67, PATE1, C11orf65, ADAMTS15, MPZL3, FOLH1B, C11orf45, HYLS1, TMEM218, SLC37A2, OR8B12, OR8G5, OR10G8, OR10G9, OR10S1, OR6T1, OR4D5, TBCEL, TMEM136, SPATA19, CCDC83, HEPACAM, OAF, CCDC89, ANGPTL5, ANKK1, RNF214, LOC283143, BCL9L, FOXR1, CCDC153, OR8D1, OR8D2, OR8B4, KIRREL3-AS3, LOC283174, LOC283177, CCDC84, TMEM225, OR8D4, C11orf53, LOC341056, HEPHL1, C11orf34, VSTM5, TRIM77P, FOLR4, KDM4DL, BSX, OR6X1, OR6M1, OR10G4, OR10G7, OR8B3, OR8A1, LOC399939, LOC399940, C11orf87, C11orf92, C11orf88, MIR100HG, PATE2, PATE4, FLJ39051, SNX19, MIRLET7A2, MIR100, MIR125B1, MIR34B, MIR34C, DD11, BLID, CARD17, LINC00167, SCARNA9, HEPN1, TRIM64B, TRIM53P, TRIM49L2, UBTF1, LOC643037, LOC643733, LOC643923, CLDN25, LOC649133, SNORA8, SNORA1, SNORA18, SNORA40, SNORA25, SNORA32, SNORD5, SNORD6, TRIM49L1, LOC100128239, LOC100132078, PATE3, LOC100288077, LOC100288346, MIR1304, BACE1-AS, MIR4301, MIR3167, MIR1260B, LOC100499227, MIR3920, MIR3656, LOC100506368, CASP12, LOC100507392, LOC100526771, HSPB2-C11orf52, FXYD6-FXYD2, MIR4697, MIR4490, MIR4493, MIR4491, MIR4492, MIR4693, LOC100652768</i>
13q21.32	0.066347	66193890	66375686	0.181796	25.2	0.0			1	<i>[PCDH9]</i>
14q11.2	0.000000	19216370	19497022	0.280652	38.3	0.0			2	<i>OR11H12, LOC642426</i>
14q32.33	0.168600	1.06E+08	1.06E+08	0.164705	13.0	0.0			3	<i>BRF1, PACS2, BTBD6</i>
15q11.1	0.000000	19465360	19819960	0.3546	41.7	0.0			1	<i>[CHEK2P2]</i>
16p11.2	0.038277	31321913	31446721	0.124808	7.0	0.0			4	<i>COX6A2, ITGAD, ITGAM, ITGAX</i>
16p11.2	0.006609	33298900	33758081	0.459181	12.2	0.0			1	<i>[LOC390705]</i>
16q23.3	0.050249	82106387	82278755	0.172368	11.3	0.0			2	<i>HSD17B2, MPHOSPH6</i>
17p13.1	0.020653	3416599	15147858	11.73126	14.8	0.0	5	hsa-mir-548h-3,	217	<i>ACADVL, ALOX12, ALOX12B, ALOX15, ALOX15B, ARRB2, ASGR1,</i>

								hsa-mir-744, hsa-mir-4314, hsa-mir-324, hsa-mir-497		ASGR2, ATP1B2, ATP2A3, C1QBP, CD68, CHD3, CHRNBI, CHRNE, COX10, CLDN7, CTNS, DLG4, DNAH9, DVL2, EFNB3, EIF4A1, EIF5A, ENO3, FGF11, GPIBA, GPS2, GUCY2D, ITGAE, MYH1, MYH2, MYH3, MYH4, MYH8, MYH10, NUP88, P2RX1, P2RX5, PER1, PFAS, PFN1, PLD2, PMP22, POLR2A, PSMB6, RCVRN, RPL26, SCO1, MAP2K4, SHBG, SLC2A4, SOX15, VAMP2, TP53, UBE2G1, TRPV1, ZNF18, ZNF232, SLC25A11, GAS7, TNK1, MYH13, TNFSF13, TNFSF12, TM4SF5, USP6, RABEP1, KCNAB3, AURKB, GLP2R, NTN1, STX8, FXR2, MPDU1, SPAG7, ACAP1, KIAA0753, ARHGAP44, HS3ST3B1, HS3ST3A1, CLEC10A, MYBBP1A, KIF1C, GABARAP, NLRP1, ARHGEF15, CAMTA2, KDM6B, ZZEF1, WSCD1, CTDNEP1, PIK3R5, C17orf81, SHPK, AIPL1, RNF167, SENP3, SNORA67, PELP1, RANGRF, TAX1BP3, MINK1, MED31, DERL2, YBX2, ANKFY1, FAM64A, XAF1, C17orf59, GPR172B, WRAP53, C17orf85, DHX33, C17orf48, PLSCR3, NLGN2, ZBTB4, CXCL16, TRAPPC1, ALOXE3, ELAC2, MIS12, PHF23, CTC1, NDEL1, PITPNM3, TMEM93, TEK1, GSG2, ZMYND15, CAMKK1, RPAIN, TMEM107, LSM1D1, NEURL4, ZNF594, HES7, MGC12916, TXNDC17, TMEM88, MYOCD, CDRT7, SAT2, CNTROB, CYB5D1, USP43, KRBA2, CYB5D2, C17orf49, ZFP3, GGT6, SPNS2, DNAH2, CDRT15, WDR16, CCDC42, PIK3R6, ODF4, KCTD11, MFSD6L, TRPV3, SLC16A11, FBXO39, DHRS7C, SLC16A13, C17orf74, SPNS3, C17orf61, BCL6B, VMO1, LOC284023, LINC00324, FLJ34690, SLC13A5, TMEM102, LOC339166, TMEM95, SMTNL2, SPEM1, GLTPD2, INCA1, SCIMP, C17orf100, SPDYE4, TMEM220, SHISA6, SLC25A35, MED11, MIR195, TNFSF12-TNFSF13, RNASEK, MIR324, MIR497, SLC35G6, RNF222, PIRT, SNORA48, SNORD10, SCARNA21, LOC728392, MIR744, LOC100128288, C17orf107, LOC100130950, LOC100289255, MIR4314, MIR3676, LOC100506713, MIR497HG, RNASEK-C17ORF49, C17orf61-PLSCR3, SENP3-EIF4A1, P2RX5-TAX1BP3, MIR4520A, MIR4521, MIR4520B
17q21.31	0.000000	41500727	41706869	0.206142	15.7	0.0	1	hsa-mir-2117	3	DHX8, ETV4, MIR2117
18q12.1	0.000150	25713125	26719170	1.006045	25.2	0.0			1	CDH2
18q23	0.000000	75863392	75994876	0.131484	25.2	0.0			1	[SALL3]
19p13.3	0.000000	1	238072	0.238071	21.7	0.0	1	hsa-mir-1302-11	4	OR4F17, FLJ45445, FAM138F, FAM138A
19p12	0.011977	20474418	21483236	1.008818	18.3	0.0	2	hsa-mir-1270-2, hsa-mir-1270-1	9	ZNF708, ZNF85, ZNF430, ZNF714, ZNF431, ZNF626, ZNF737, MIR1270-1, MIR1270-2
21q21.3	0.017290	14307675	32051017	17.74334	23.5	0.0	7	hsa-mir-4327, hsa-mir-155, hsa-mir-548x, hsa-	85	APP, ATP5J, BACH1, CXADR, GABPA, GRIK1, NCAM2, TMPRSS15, HSPA13, NRIP1, CLDN8, ADAMTS1, RWDD2B, USP16, CCT8, BTG3, ADAMTS5, LTN1, CLDN17, N6AMT1, USP25, RBM11, LINC00158, CHODL-AS1, LINC00113, C21orf15, LINC00308, MRPL39, C21orf91,

								mir-125b-2, hsa-let-7c, hsa-mir-3118-5, hsa-mir-3156-3		<i>C21orf7, JAM2, SAMSNI, MIR155HG, CYYR1, LINC00161, KRTAP13-1, CHODL, LIPI, LINC00189, C21orf91-OT1, LINC00314, KRTAP15-1, D21S2088E, LINC00307, LINC00515, KRTAP13-4, POTES, KRTAP19-1, KRTAP13-2, KRTAP13-3, KRTAP23-1, KRTAP6-1, KRTAP6-2, KRTAP6-3, KRTAP19-2, KRTAP19-3, KRTAP19-4, KRTAP19-5, KRTAP19-6, KRTAP19-7, KRTAP20-1, KRTAP20-2, KRTAP22-1, KRTAP20-3, LOC339622, LINC00317, LINC00320, LOC388813, LINC00478, KRTAP26-1, MIRLET7C, MIR125B2, MIR155, MIR99A, GRIK1-AS1, KRTAP24-1, KRTAP27-1, KRTAP25-1, KRTAP20-4, KRTAP22-2, GRIK1-AS2, MIR4327, MIR3156-3, C21orf37, MIR4759</i>
--	--	--	--	--	--	--	--	--	--	--

^aMb: mega base pair, * Targets selected for validation

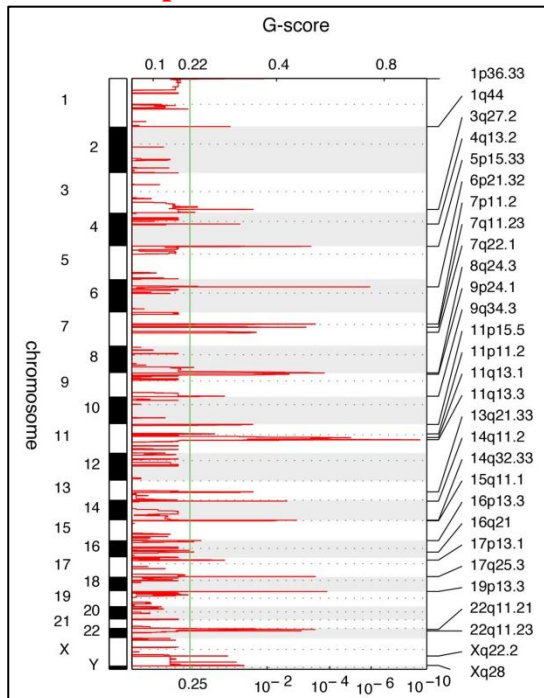
4.5.2 CNA Associated with Clinicopathological Parameters and Clinical Outcome

Chromosomal aberrations were analysed to understand their relevance and associations with clinicopathological parameters are nodal status, grade and clinical outcome. We did not find any significant association of chromosomal aberrations with tumor grade. Associations between CNA and lymph node metastasis as well as patient survival are described below.

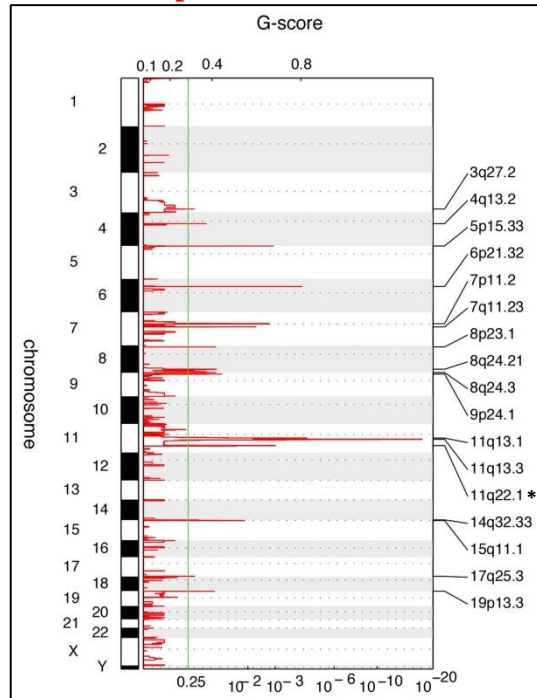
4.5.2.1 CNA Associated with Lymph Node Metastasis

The number of identified CNAs were: 48 alterations in lymph node metastasis positive OSCC (N+), 66 alterations in lymph node metastasis negative OSCCs (N0), 18 alterations in T1N+ and T2N+ OSCCs, 37 alterations in T3N0 and T4N0 OSCCs, 25 alterations in T3N+ and T4N+ OSCCs. Interestingly, the number of CNA in node negative cases were higher as compared to node positive cases (Figure 20). The unique alterations observed in N+ tumors were gain of 8p23.1, 8q24.21, 11q22.1 and losses at 3p26.3, 8p23.2, 9p12, 11q22.3, 18q12.1. A brief summary of common and unique CNA in N0 and N+ tumors is described in the Table 13 and 14.

**A) Node negative (N0) OSCC-
Gain/ Amplification**

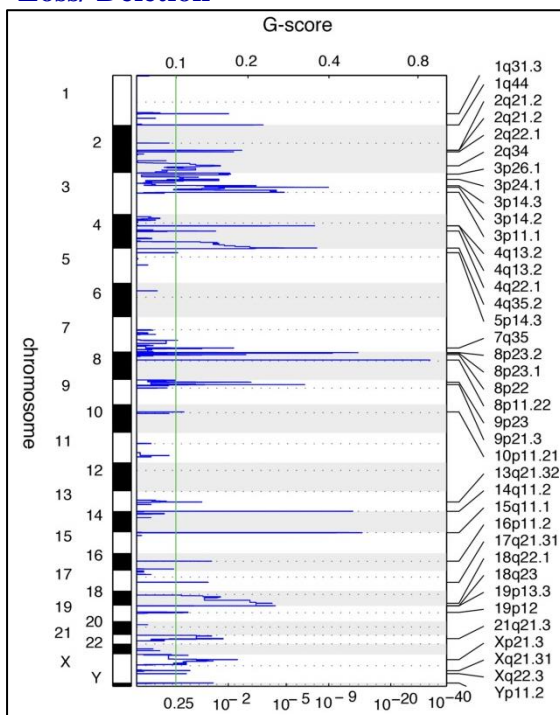


**B) Node positive (N+) OSCC-
Gain/ Amplification**

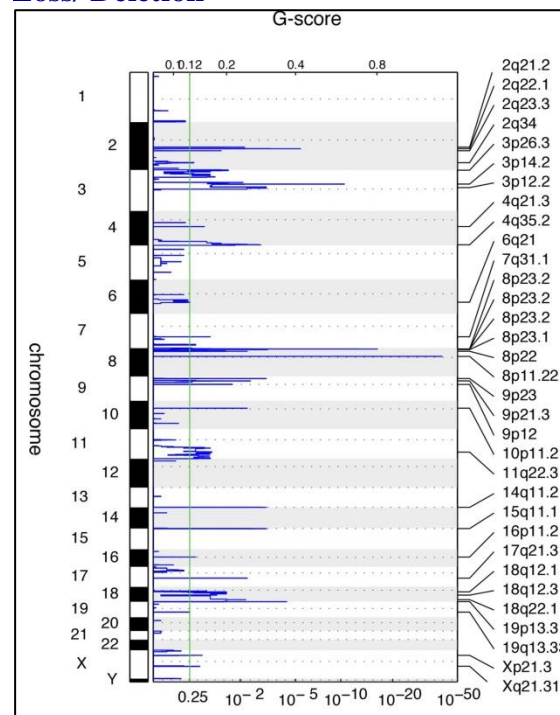


* Target selected for validation

**A) Node negative (N0) OSCC-
Loss/ Deletion**



**B) Node positive (N+) OSCC-
Loss/ Deletion**



C) Venn diagram- N0 v/s N+

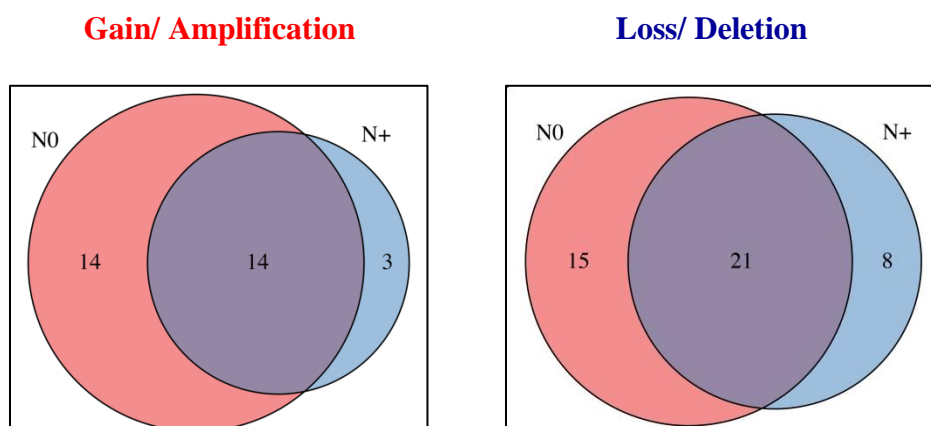


Figure 20: The Amplification/Deletion Plots And Venn Diagrams Depicting The CNA Associated With Lymph Node Metastasis. The amplification/ deletion plots depicting the G-scores (top) and q-values (bottom) with respect to amplification/ deletion for all markers over the entire region analysed in **A)** Lymph node metastasis negative (N0) OSCC, **B)** Lymph node metastasis positive (N+) OSCC and **C)** Venn diagram representing a number of unique and overlapping CNA between N0 and N+ OSCC.

4.5.2.2 CNA Associated with Patient Survival and Tumor Recurrence

The overall survival of OSCC patients is generally poor as the majority of the patients develop recurrent disease with chemo- and/or radio-resistance. Also, patients in similar stages of OSCC may not have an identical course of disease and often differ in their clinical outcome. The cox proportional hazards model was used to identify the CNA associated with the DSS (Disease-Specific Survival) and RFS (Recurrence Free Survival). We found that at BH-corrected p-value

< 0.25 , twenty five CNAs were associated with recurrence free survival while twenty six CNAs were associated with disease specific survival (Table 19). The gain of chromosomal regions 1p36.33 ($p = 0.032$), 11q13.3 ($p = 0.12$), 11q22.1 ($p = 0.15$) and 16p11.2 ($p = 0.049$) were associated with poor clinical outcome, whereas gain of 22q11.21 ($p = 0.18$) was associated with better survival. Chromosomal losses of 2p11.2 ($p = 0.047$), 3p14.2 ($p = 0.19$), 4q13.2 ($p = 0.05$), 9p23 ($p = 0.09$), and 11q22.3 ($p = 0.08$) were associated with poor clinical outcome (Table 19). Gain of 1p36.33 and 11q22.1 were the strongest predictors of poor clinical outcome. Kaplan-Meier survival curves for both these chromosomal aberrations are shown in Figure 21.

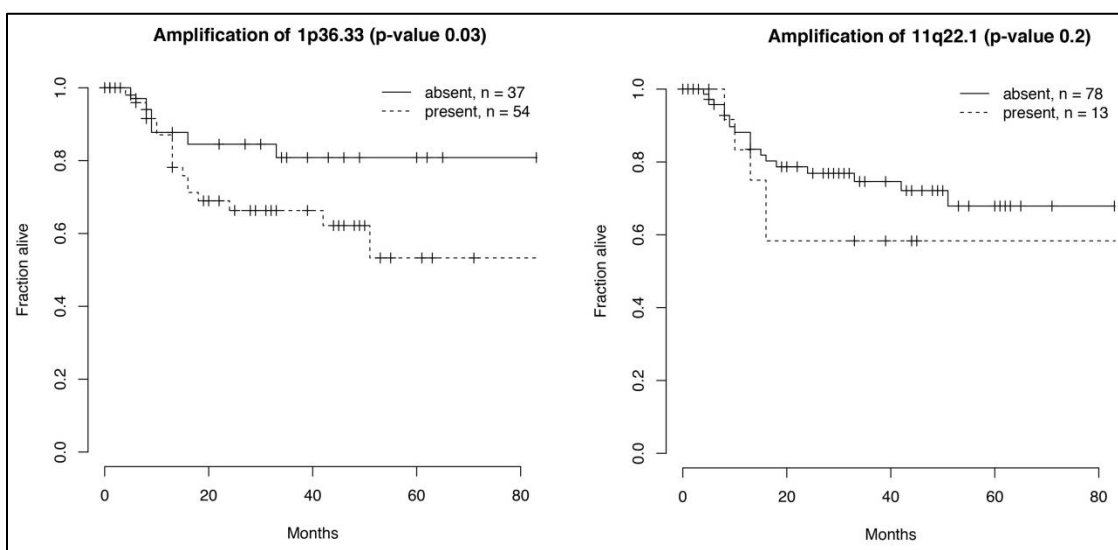


Figure 21: Kaplan-Meier Estimates of Patients Survival with Alterations on Chromosome 1p36.33 and 11q22.1. A) Patient groups with and without gain of 1p36.33 and B) Groups with and without gain of 11q22.1. Survival in months (x-axis) is plotted against the fraction of samples alive (y-axis).

Table 19: Univariate Cox Proportional Hazards Regression Analysis of Single Predictors for Recurrence-Free and Disease Specific Survival.

Cytoband	Aberration	Disease Specific Survival		Recurrence Free Survival	
		BH Corrected p-value [#]	CPH coef.	BH Corrected p-value [#]	CPH coef.
1p36.33*	Gain	0.0327	1.0292	0.0130	0.9185
1q23.2	Gain	0.0269	0.8763	0.0934	0.5740
3q27.2	Gain	0.2467	0.4360	0.2171	0.3823
11q13.3*	Gain	0.1232	0.4237	0.2142	0.2804
11q22.1*	Gain	0.1543	0.5204	-	-
16p13.3	Gain	0.0655	0.6641	0.0658	0.5375
16p11.2	Gain	0.0497	0.8665	0.0549	0.6726
16q12.2	Gain	0.0607	0.6662	0.0308	0.6497
2p11.2	Loss	0.0473	1.0051	0.0844	0.7770
2q22.1	Loss	0.0147	1.0214	0.0623	0.6329
2q34	Loss	0.0644	0.7772	0.0281	0.7450
3p14.2	Loss	0.1976	0.5665	0.2202	0.4236
4q13.2	Loss	0.0582	0.8036	0.1775	0.4836
4q22.1	Loss	0.0998	0.7098	0.2113	0.4591
9p23	Loss	0.0969	0.7110	0.1339	0.5327
11q22.3	Loss	0.0829	0.7902	0.0990	0.6467
6p21.1	Gain	0.2198	0.4606	-	-
18p11.31	Gain	0.2227	0.5128	-	-
22q11.21	Gain	0.1832	-0.6038	-	-
1q31.3	Loss	0.2371	0.5055	-	-
3p11.1	Loss	0.0809	0.7957	-	-
4q13.2	Loss	0.0900	0.7457	-	-
5p14.3	Loss	0.1120	0.7583	-	-
7q31.1	Loss	0.1624	0.5980	-	-
13q21.32	Loss	0.2166	0.5221	-	-
21q21.3	Loss	0.1385	0.6243	-	-
2q37.3	Gain	-	-	0.2185	0.3978
5p15.33	Gain	-	-	0.2230	0.3284
11p15.5	Gain	-	-	0.1793	0.4287
13q21.33	Gain	-	-	0.2436	-0.7795
16q21	Gain	-	-	0.1975	0.4048
19p13.3	Gain	-	-	0.1465	0.3602
22q11.23	Gain	-	-	0.1986	-0.4272
9p21.3	Loss	-	-	0.2007	0.4494
14q11.2	Loss	-	-	0.1116	0.5358
17p13.1	Loss	-	-	0.0346	0.7927

[#]Benjamini-Hochberg method of adjusting for multiple tests; CPH coef: Cox Proportional Hazard coefficient- A positive regression coefficient means that the hazard is higher, and thus the prognosis worse. * Targets selected for validation

4.6 Gene Expression Profiling

The GE analysis was performed using the R Bioconductor based statistical package *limma*. The comparisons were carried out using 6 Gingivobuccal normals, 15 Leukoplakia, 24 samples Early stage OSCC (T1N0, T2N0), 10 cases of Advanced stage OSCC (T1N+, T2N+) and 27 Gingivobuccal OSCC reported by us earlier (Ambatipudi, S. et al.,) [which includes 3 Early stage OSCC (T1N0, T2N0), 24 Advanced stage OSCC (T1N+, T2N+, T3N0, T4N0, T3N+ and T4N+)].

The major objective of the GE analysis was to identify:

- 1) Gene expression signatures/profiles in leukoplakia (n=15) and total OSCC (n=61) as compared to gingivobuccal normals (represented in Table 20 and Table 21).
- 2) Deregulated genes associated with disease progression, i.e. transformation from leukoplakia (n=15), to early stage OSCC (n=27), to advanced stage OSCC (n=34).
- 3) Integrative analysis to identify copy number dependent and independent expression change of genes at the altered chromosomal locus.

The PCA plot of all the 82 samples (normal, leukoplakia and OSCC) analysed based on the expression of 3805 genes were selected as the union of all genes differentially expressed, when comparing any two of the 4 categories of samples and is plotted in Figure 22 (A). The fold change threshold of 2 and corrected p-value of 0.01 was used for analysis. We identified a total of 849 genes differentially expressed (395 up-regulated and 454 down-regulated) in leukoplakia, 1813 genes differentially expressed (805 up-regulated and 1008 down-regulated) in early stage OSCC and 1924 genes differentially expressed (798 up-regulated and 1056 down-regulated) in the advanced stage OSCC (Figure 22 (B and C)).

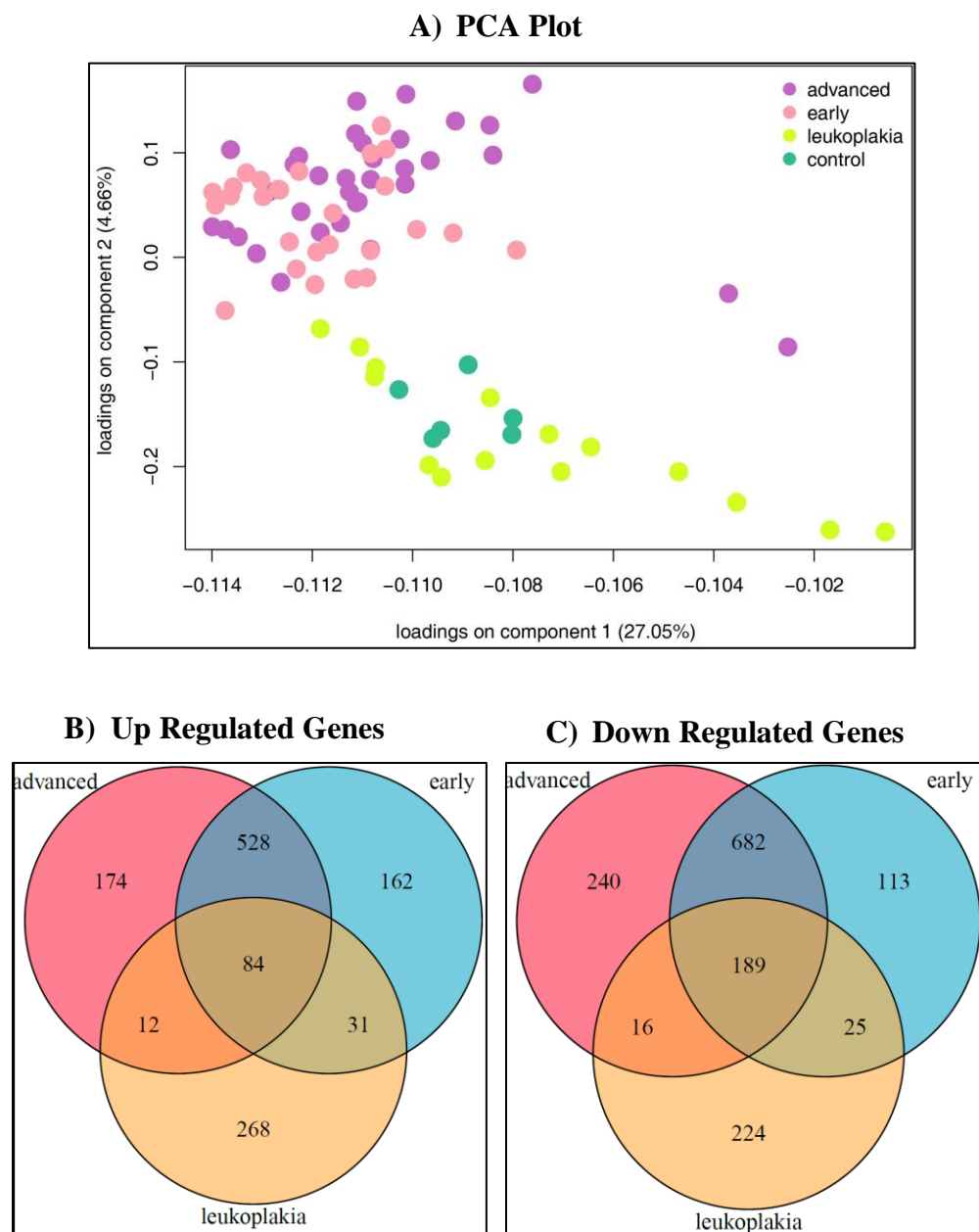


Figure 22: Overview of Gene Expression Changes in leukoplakia and OSCC. **A)** PCA plot of the 82 samples analysed, based on the union of all genes differentially expressed (3805 genes), when comparing any two of the 4 categories of samples plotted here (fold change thresholds of 2 and 0.01 for corrected p-value). The percent variance of each of the two components is represented in brackets. **B)** and **C)** Venn diagram depicting common and unique number of differentially expressed genes in leukoplakia (orange), early stage OSCC (blue), advanced stage OSCC (pink).

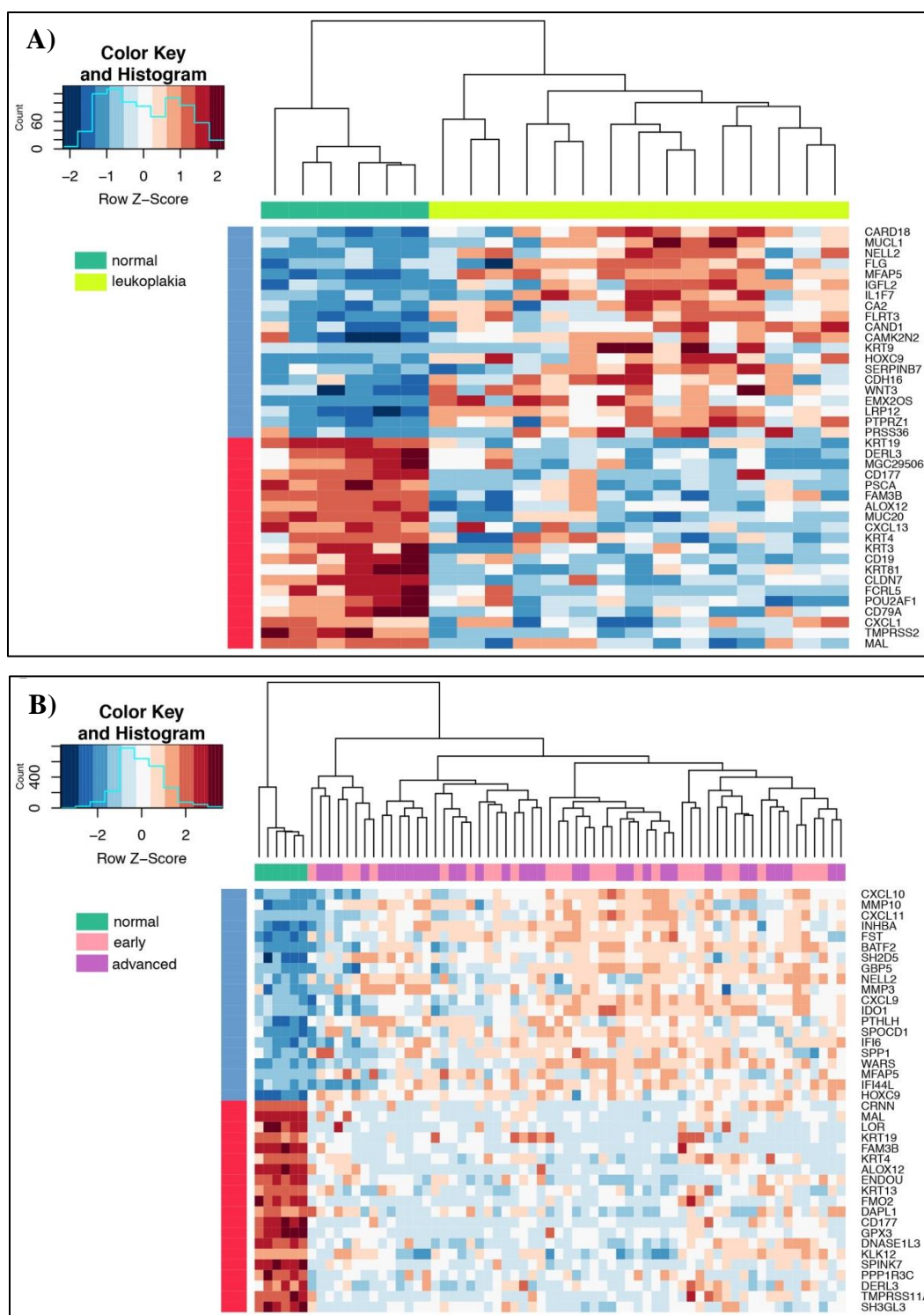


Figure 23: Heatmap of the top 20 differentially expressed genes in leukoplakia and OSCC.
 A) Differentially expressed genes in leukoplakia (chartreuse) when compared to normal (green)
 B) Differentially expressed genes in the early stage OSCC (pink), and advanced stage OSCC (magenta) when compared to normal (green). Genes represented in red were down regulated while up regulated ones are in blue. All expression values were scaled across samples.

The Cluster analysis of the top 20 up- and down- regulated genes with highest or lowest Log Fold Change values in leukoplakia and OSCC, classified and separated compared to the normals are represented in a heatmap (Figure 23). All expression values in the heatmap are represented as a Z score, which indicates standard deviations from the mean expression. Total 84 up-regulated genes and 189 down-regulated genes were common among all the 3 groups (leukoplakia, early stage OSCC and advanced stage OSCC), implying that these genes could be associated with disease progression (Figure 22 B and C). The top most 50 differentially expressed genes are listed in Table 20 and Table 21. Genes up regulated in all the three groups includes *NELL2*, *MFAP5*, *CA2*, *FLRT3*, *HOXC9*, *CDH16*, *LRP12*, *PTPRZ1*, *TNNT1*, *WDR66*, *NEXN*, *EGR2*, *HOXC13*, *E2F7*, *ECT2* and *EIF5A2*. In addition, a higher number of genes were found to be altered in early and advanced stage OSCC, few of these includes *CXCL10*, *MMP10*, *INHBA*, *GBP5*, *CXCL11*, *MMP3*, *FST*, *BATF2*, *SPP1*, *SH2D5*, *CXCL9*, *IFIT3*, *SERPINE1*, *GALNT6*, *FOXL2*, *PDPN*, *ITGA3*, *VEGFC*, *STAT1*, *LY6K*, *KLF7*, *SOX9* and *CD274*. Of the 189 common down- regulated genes amongst all the 3 groups, included *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* and *GPX3*. *KRT76* was the top most down regulated gene in the advanced stage OSCC. However, no specific altered genes were observed in node negative (N0) and node positive (N+) tumors comparisons.

Apart from the individual gene level analysis, we also performed functional annotation analysis to identify "Biological pathways" deregulated in leukoplakia and OSCC. The pathway analysis was performed using R package Gostats. We categorized the pathways into three large classes (biological processes; cellular components and molecular functions) and performed the

enrichment tests. For all three classes, the results demonstrate many divergent pathways involved forming a complex network. To simplify, a few representative biological processes altered are listed in Table 22 and Table 23 and a brief summary of various processes involved are listed below:

- **Biological process:** Apoptosis, biological adhesion, biological regulation, cellular component organization or biogenesis, developmental process, growth and immune system process.
- **Cellular components:** Cell junction, extracellular matrix, macromolecular complex, membrane and organelle.
- **Molecular Function:** Antioxidant activity, catalytic activity, enzyme regulator activity nucleic acid binding transcription factor activity, protein binding transcription factor activity, receptor activity, translation regulator activity and transporter activity.

Majority of pathways which were deregulated includes cell surface receptor signaling pathways, Cytokine-mediated signaling pathway, Integrin signaling pathway, Wnt signalling pathway, angiogenesis, EGFR receptor pathway, Ras signalling pathways and stem cell differentiation etc. To understand the association of deregulated genes with biological processes (i.e. The function of protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism), the PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system was used [108, 109]. The graphical representation of genes associated with various biological process is represented in Figure 24. Both leukoplakia and OSCC share a majority of processes, however the number of associated genes in OSCC is higher compared to pre-invasive lesions. 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*) were observed. Most of the genes shared overlapping processes.

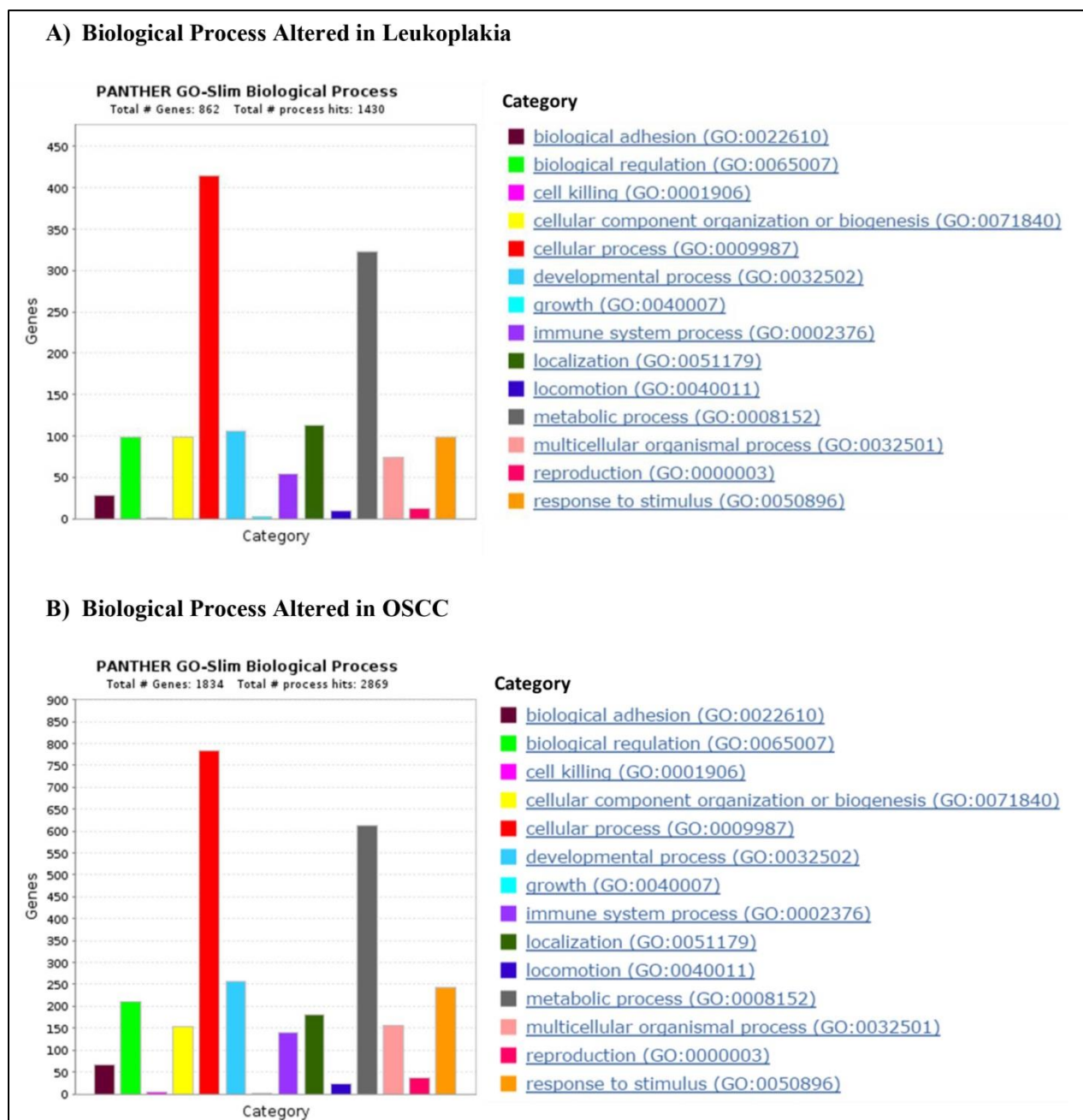


Figure 24. Biological Processes Altered in Leukoplakia and OSCC. The plots represent a number of genes associated with respective biological processes (built using an online tool “The PANTHER Classification System”). Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Process hits.

Table 20: Top 50 Common Genes Up Regulated in Leukoplakia and OSCC.

Up Regulated Genes in Leukoplakia					Up Regulated Genes in OSCC				
HUGO	Probe	LogFC	Adj.P.Val	Cytoband	HUGO	Probe	LogFC	Adj.P.Val	Cytoband
<i>MUCL1</i>	A_23_P150979	3.7366	2.8994E-02	12q	<i>CXCL10</i>	A_24_P303091	4.8235	7.4800E-05	4q21
<i>NELL2*</i>	A_23_P10025	3.7172	6.8259E-03	12q12	<i>MMP10</i>	A_23_P13094	4.7884	3.8000E-05	11q22.3
<i>MFAP5*</i>	A_23_P87700	3.4662	1.7462E-03	12p13.1-p12.3	<i>CXCL11</i>	A_24_P20607	4.2100	1.0900E-06	4q21.2
<i>IGFL2</i>	A_23_P153571	3.4335	5.8693E-03	19q13.32	<i>INHBA*</i>	A_23_P122924	4.1307	7.2800E-09	7p15-p13
<i>CA2</i>	A_23_P8913	3.0769	7.3602E-04	8q22	<i>FST</i>	A_23_P110531	4.0884	5.5100E-09	5q11.2
<i>FLRT3</i>	A_23_P166109	2.8271	3.3019E-03	20p11	<i>BATF2</i>	A_23_P370682	4.0424	3.9600E-06	11q13.1
<i>CAND1</i>	A_24_P47681	2.7596	1.2593E-02	12q14	<i>SH2D5</i>	A_24_P686965	3.9692	9.4300E-07	1p36.12
<i>HOXC9*</i>	A_23_P25150	2.4497	5.7500E-05	12q13.3	<i>GBP5</i>	A_23_P74290	3.9559	4.0000E-05	1p22.2
<i>CDH16</i>	A_23_P100240	2.2219	1.7015E-02	16q22.1	<i>NELL2*</i>	A_23_P10025	3.9185	3.3757E-04	12q12
<i>LRP12</i>	A_24_P415012	2.0482	3.4200E-09	8q22.2	<i>MMP3</i>	A_23_P161698	3.8340	1.0160E-03	11q22.3
<i>PTPRZ1</i>	A_23_P168761	2.0385	2.8800E-05	7q31.3	<i>CXCL9</i>	A_23_P18452	3.7629	1.1647E-03	4q21
<i>TNNT1</i>	A_23_P56050	2.0183	1.4291E-02	19q13.4	<i>IDO1</i>	A_23_P112026	3.5338	1.7049E-03	8p12-p11
<i>WDR66</i>	A_23_P363275	1.9863	2.5000E-07	12q24.31	<i>PTHLH</i>	A_23_P2271	3.4103	5.5600E-07	12p12.1-p11.2
<i>NEXN</i>	A_24_P409971	1.9275	2.9814E-02	1p31.1	<i>SPOCD1</i>	A_23_P431388	3.3819	1.8200E-06	1p35.2
<i>EGR2</i>	A_23_P46936	1.9041	3.2989E-02	10q21.1	<i>IFI6</i>	A_23_P201459	3.3098	9.0500E-06	1p35
<i>HOXC13*</i>	A_23_P64808	1.8665	2.3139E-03	12q13.3	<i>SPPI</i>	A_23_P7313	3.3056	3.3891E-04	4q22.1
<i>IRX2</i>	A_23_P156025	1.8405	4.9010E-04	5p15.33	<i>WARS</i>	A_23_P65651	3.2942	7.1200E-06	14q32.31
<i>CTSC</i>	A_24_P115762	1.7866	1.4684E-03	11q14.2	<i>MFAP5*</i>	A_23_P87700	3.2694	2.0443E-04	12p13.1-p12.3
<i>ARSJ</i>	A_23_P213298	1.7719	1.0044E-02	4q26	<i>IFI44L</i>	A_23_P45871	3.2463	2.3780E-04	1p31.1
<i>KIAA0101</i>	A_23_P117852	1.7223	2.2592E-03	15q22.31	<i>HOXC9*</i>	A_23_P25150	3.2133	4.8000E-10	12q13.3
<i>ASCC3</i>	A_23_P382654	1.6917	2.9274E-03	6q16	<i>IFIT3</i>	A_23_P35412	3.1899	6.2900E-06	10q24
<i>MICALCL</i>	A_23_P2041	1.6879	4.1405E-03	11p15.3	<i>TNFRSF12</i>	A_23_P49338	3.0517	9.3000E-10	16p13.3
<i>DUSP6</i>	A_23_P139704	1.6455	1.1381E-03	12q22-q23	<i>OASL</i>	A_23_P139786	3.0493	3.8700E-05	12q24.2
<i>PHLDB2</i>	A_24_P240166	1.6229	2.4606E-03	3q13.2	<i>DHRS2</i>	A_23_P48570	3.0262	2.5800E-08	14q11.2
<i>CAV2</i>	A_24_P925040	1.6217	8.9663E-03	7q31.1	<i>WDR66</i>	A_23_P363275	2.9925	1.5400E-21	12q24.31
<i>NFE2L1</i>	A_32_P163469	1.6179	1.5291E-03	17q21.3	<i>USP18</i>	A_32_P132206	2.9279	7.1800E-10	22q11.21
<i>E2F7</i>	A_32_P210202	1.6170	3.8969E-04	12q21.2	<i>GBP1</i>	A_23_P62890	2.8955	4.9400E-08	1p22.2
<i>LOC375295</i>	A_32_P16204	1.5945	3.9067E-02		<i>MUCL1</i>	A_23_P150979	2.8725	3.5215E-02	12q

Up Regulated Genes in Leukoplakia					Up Regulated Genes in OSCC				
HUGO	Probe	LogFC	Adj.P.Val	Cytoband	HUGO	Probe	LogFC	Adj.P.Val	Cytoband
<i>C6orf105</i>	A_23_P156826	1.5924	1.3843E-02	6p24.1	<i>C19orf21</i>	A_23_P390068	2.8311	2.2600E-06	19p13.3
<i>DLST</i>	A_23_P205697	1.5834	1.1226E-02	14q24.3	<i>SERPINE1</i>	A_24_P158089	2.8173	1.1700E-08	7q22.1
<i>SLC38A1</i>	A_24_P261734	1.5652	2.6800E-05	12q13.11	<i>ARSJ</i>	A_23_P213298	2.7547	1.0300E-06	4q26
<i>ZAK</i>	A_23_P366394	1.5566	7.9000E-06	2q24.2	<i>IFIT1</i>	A_23_P52266	2.7267	5.3043E-04	10q23.31
<i>TPM1</i>	A_24_P924721	1.5449	5.0000E-05	15q22.1	<i>FOXL2</i>	A_23_P110052	2.7205	5.3171E-04	3q23
<i>LTBP1</i>	A_23_P43810	1.5050	1.3645E-02	2p22-p21	<i>PDPN</i>	A_24_P299685	2.6871	1.2100E-09	1p36.21
<i>AMIGO2</i>	A_23_P14083	1.4723	1.7720E-02	12q13.11	<i>DFNA5</i>	A_23_P82449	2.6851	1.4400E-07	7p15
<i>KIF23</i>	A_23_P48835	1.4562	3.6687E-03	15q23	<i>TNC</i>	A_23_P157865	2.6771	2.1547E-04	9q33
<i>SEPT11</i>	A_23_P306964	1.4559	4.1572E-04	4q21.1	<i>HMGA2</i>	A_23_P95930	2.6760	8.2600E-05	12q15
<i>NCAPG</i>	A_23_P155815	1.4175	3.0417E-03	4p15.33	<i>ITGA3</i>	A_23_P55251	2.6592	4.1300E-09	17q21.33
<i>DEPDC1B</i>	A_23_P361419	1.4040	1.7524E-03	5q12.1	<i>GALNT6</i>	A_23_P204133	2.6526	4.2900E-07	12q13
<i>MYO5A</i>	A_24_P255218	1.4006	1.8100E-05	15q21	<i>FLRT3</i>	A_23_P166109	2.6093	6.7091E-04	20p11
<i>RACGAP1</i>	A_32_P186474	1.3945	7.3890E-04	12q13.12	<i>BST2</i>	A_23_P39465	2.5986	9.3300E-06	19p13.1
<i>MND1</i>	A_23_P133123	1.3780	7.3539E-03	4q31.3	<i>CYP27B1</i>	A_23_P36397	2.5873	4.3700E-05	12q14.1
<i>LRP8</i>	A_23_P200222	1.3637	7.9899E-03	1p34	<i>TREM2</i>	A_24_P12397	2.5796	6.7900E-08	6p21.1
<i>SCHIP1</i>	A_32_P62863	1.3584	4.9991E-03	3q25.32-q25.33	<i>RSAD2</i>	A_24_P316965	2.5756	7.2800E-07	2p25.2
<i>MRPS17</i>	A_23_P258321	1.3472	4.0735E-03	7p11	<i>VEGFC</i>	A_23_P167096	2.5658	2.2814E-04	4q34.3
<i>CDK6</i>	A_24_P166663	1.3364	4.3297E-03	7q21-q22	<i>CA2</i>	A_23_P8913	2.5646	3.5151E-04	8q22
<i>SNRPB2</i>	A_23_P40307	1.3148	4.4576E-04	20p12.1	<i>HOXC13*</i>	A_23_P64808	2.5178	6.0900E-07	12q13.3
<i>CDCA2</i>	A_23_P385861	1.2945	5.9745E-03	8p21.2	<i>HRASLS2</i>	A_24_P364263	2.4809	2.0400E-06	11q12.3

Footnote: LogFC: Log Fold Change, Adj.P.Val: Adjusted P values, * Targets selected for validation

Table 21: Top 50 Genes Down Regulated in Leukoplakia and OSCC.

Down Regulated Genes in Leukoplakia					Down Regulated Genes in OSCC				
HUGO	Probe	LogFC	Adj.P.Val	Cytoband	HUGO	Probe	LogFC	Adj.P.Val	Cytoband
<i>KRT19</i>	A_23_P66798	-6.2905	5.4035E-04	17q21.2	<i>CRNN</i>	A_23_P115202	-7.5956	2.6000E-09	1q21
<i>DERL3*</i>	A_23_P6362	-5.1339	1.5100E-05	22q11.23	<i>MAL*</i>	A_23_P17134	-7.5522	8.2300E-14	2q11.1
<i>MGC29506</i>	A_23_P84596	-4.4922	3.8895E-04	5q31.2	<i>LOR</i>	A_24_P171274	-6.9283	3.0100E-09	1q21
<i>CD177</i>	A_23_P259863	-4.2522	7.0200E-13	19q13.2	<i>KRT19</i>	A_23_P66798	-6.4009	1.1700E-05	17q21.2
<i>PSCA</i>	A_23_P71379	-4.0452	3.1554E-04	8q24.2	<i>FAM3B</i>	A_23_P166269	-6.3253	1.2400E-17	21q22.3
<i>FAM3B</i>	A_23_P166269	-3.8053	8.5000E-07	21q22.3	<i>KRT4</i>	A_23_P2674	-6.2520	7.2100E-07	12q13.13
<i>ALOX12</i>	A_23_P152909	-3.7317	1.9000E-10	17p13.1	<i>ALOX12</i>	A_23_P152909	-5.6707	8.7600E-28	17p13.1
<i>MUC20</i>	A_23_P92222	-3.7250	8.8243E-04	3q29	<i>ENDOU</i>	A_24_P347310	-5.4624	9.7400E-11	12q13.1
<i>CXCL13</i>	A_23_P121695	-3.6794	1.3361E-04	4q21	<i>KRT13</i>	A_24_P228149	-5.3280	8.2900E-06	17q21.2
<i>KRT4</i>	A_23_P2674	-3.5720	2.0152E-02	12q13.13	<i>FMO2</i>	A_23_P355295	-5.1128	2.4500E-12	1q24.3
<i>KRT3</i>	A_23_P33573	-3.5248	8.5000E-07	12q13.13	<i>DAPL1</i>	A_23_P165598	-5.1075	1.1200E-05	2q24.1
<i>CD19</i>	A_23_P113572	-3.4455	2.9200E-06	16p11.2	<i>CD177</i>	A_32_P143589	-4.7300	9.7900E-22	19q13.2
<i>KRT81</i>	A_23_P25187	-3.4325	9.7200E-05	12q13	<i>GPX3</i>	A_23_P133474	-4.6189	7.9500E-17	5q33.1
<i>CLDN7</i>	A_23_P164284	-3.4276	6.2400E-06	17p13.1	<i>DNASE1L3</i>	A_23_P257993	-4.6128	7.1500E-10	3p14.3
<i>FCRL5</i>	A_23_P201211	-3.4179	1.1800E-06	1q21	<i>KLK12</i>	A_23_P500010	-4.5891	3.3582E-04	19q13.33
<i>POU2AF1</i>	A_23_P312920	-3.4062	1.4878E-03	11q23.1	<i>SPINK7</i>	A_23_P213832	-4.5752	4.7600E-14	5q32
<i>CD79A</i>	A_23_P107735	-3.3388	1.7400E-06	19q13.2	<i>PPPIR3C</i>	A_23_P35414	-4.5528	1.2400E-09	10q23-q24
<i>TMPRSS2</i>	A_23_P29067	-3.2332	1.1300E-11	21q22.3	<i>DERL3*</i>	A_23_P6362	-4.3666	4.4200E-06	22q11.23
<i>MAL*</i>	A_23_P17134	-3.1510	4.7099E-03	2q11.1	<i>TMPRSS11A</i>	A_24_P109101	-4.3637	1.4500E-07	4q13.2
<i>KIAA0125</i>	A_23_P140190	-3.0880	4.8400E-06	14q32.33	<i>SH3GL3</i>	A_23_P48988	-4.3293	4.1600E-12	15q24
<i>TNFRSF17</i>	A_23_P37736	-2.9853	9.2280E-04	16p13.1	<i>PTGDS</i>	A_23_P146554	-4.2998	6.4100E-08	9q34.2-q34.3
<i>FCN1</i>	A_23_P157879	-2.9765	1.3200E-07	9q34	<i>KRT3</i>	A_23_P33573	-4.2872	3.9300E-12	12q13.13
<i>PNOC</i>	A_23_P253321	-2.9701	8.7500E-06	8p21	<i>HOPX</i>	A_23_P254507	-4.2047	1.9600E-12	4q12
<i>TMPRSS11A</i>	A_24_P109101	-2.9561	3.0417E-03	4q13.2	<i>HPGD</i>	A_23_P213050	-4.1722	5.1700E-09	4q34-q35
<i>CXCL17</i>	A_23_P50269	-2.9516	2.0473E-04	19q13.2	<i>MFAP4</i>	A_23_P164057	-4.1716	1.9400E-06	17p11.2
<i>CEACAM1</i>	A_23_P55738	-2.9431	5.8900E-15	19q13.2	<i>KLK13</i>	A_24_P333697	-4.1620	3.7600E-09	19q13.33

Down Regulated Genes in Leukoplakia					Down Regulated Genes in OSCC				
HUGO	Probe	LogFC	Adj.P.Val	Cytoband	HUGO	Probe	LogFC	Adj.P.Val	Cytoband
<i>FUT6</i>	A_23_P5092	-2.9366	2.6200E-11	19p13.3	<i>CLCA4</i>	A_23_P45751	-4.1462	4.0400E-09	1p22.3
<i>SMR3A</i>	A_23_P41365	-2.9263	9.5500E-05	4q13.3	<i>SILV</i>	A_23_P2233	-4.0967	1.1300E-21	12q13-q14
<i>NYX</i>	A_23_P137046	-2.8408	8.6400E-05	Xp11.4	<i>FAM3D</i>	A_23_P41145	-4.0754	3.7000E-05	3p14.2
<i>CLCA4</i>	A_23_P45751	-2.7688	9.2689E-04	1p22.3	<i>MGC29506</i>	A_23_P84596	-4.0311	5.8000E-05	5q31.2
<i>PITX1</i>	A_23_P58642	-2.6934	4.9900E-07	5q31.1	<i>CGNLI</i>	A_23_P163306	-3.9225	1.9300E-11	15q21.3
<i>DACT2</i>	A_24_P289260	-2.6646	3.9717E-04	6q27	<i>COMP</i>	A_24_P264943	-3.9135	2.0500E-05	19p13.1
<i>MEI1</i>	A_23_P211561	-2.6148	1.3047E-03	22q13.2	<i>PITX2</i>	A_23_P167367	-3.8988	6.2200E-13	4q25
<i>GPR110</i>	A_23_P214267	-2.5394	8.4171E-03	6p12.3	<i>GATM</i>	A_23_P129064	-3.8399	3.1600E-09	15q21.1
<i>GPX3</i>	A_23_P133474	-2.5261	2.0300E-05	5q33.1	<i>HBB</i>	A_23_P203558	-3.8275	1.1465E-04	11p15.5
<i>COLIA1</i>	A_23_P207520	-2.4706	2.7122E-04	17q21.33	<i>CXCL12</i>	A_24_P412156	-3.8150	1.8900E-12	10q11.1
<i>PIM2</i>	A_24_P379104	-2.4497	1.2432E-03	Xp11.23	<i>SH3BGRL2</i>	A_23_P111267	-3.8126	6.9100E-13	6q14.1
<i>FKBP11</i>	A_23_P25121	-2.4174	1.0682E-04	12q13.12	<i>FRZB</i>	A_23_P363778	-3.8037	2.0100E-18	2q32.1
<i>FAM20A</i>	A_24_P352952	-2.4138	2.5900E-07	17q24.2	<i>RDH12</i>	A_23_P37088	-3.7727	1.4819E-04	14q24.1
<i>FPR1</i>	A_23_P38795	-2.3979	3.3800E-05	19q13.4	<i>LY6G6C</i>	A_23_P8083	-3.7554	6.8000E-07	6p21.33
<i>ECM1</i>	A_23_P160559	-2.3914	1.6328E-03	1q21	<i>SPINK5</i>	A_23_P356494	-3.7517	4.1388E-04	5q32
<i>PRSS27</i>	A_23_P106806	-2.3392	2.4041E-03	16p13.3	<i>TMPRSS11D</i>	A_23_P144417	-3.7159	1.1981E-04	4q13.2
<i>FMO2</i>	A_23_P355295	-2.3368	4.3817E-03	1q24.3	<i>ALDH3A1</i>	A_23_P207213	-3.7061	7.7400E-07	17p11.2
<i>BAMBI</i>	A_23_P52207	-2.3266	1.2568E-02	10p12.3-p11.2	<i>DACT2</i>	A_24_P289260	-3.7008	4.4900E-09	6q27
<i>PPP1R3C</i>	A_23_P35414	-2.3256	7.6902E-03	10q23-q24	<i>CTSG</i>	A_23_P140384	-3.6199	8.7600E-06	14q11.2
<i>SPAG4</i>	A_23_P132027	-2.2869	4.5600E-05	20q11.21	<i>NMU</i>	A_23_P69537	-3.6171	7.3400E-08	4q12
<i>GLUL</i>	A_24_P53976	-2.2143	2.6783E-04	1q31	<i>SCEL</i>	A_23_P500000	-3.6086	6.6300E-05	13q22
<i>PCP4LI</i>	A_32_P214665	-2.1984	4.1419E-03	1q23.3	<i>SLURPI</i>	A_24_P269527	-3.5798	2.9398E-04	8q24.3
<i>EYA2</i>	A_23_P319859	-2.1980	6.6286E-04	20q13.1	<i>MUC20</i>	A_23_P92222	-3.5661	6.1900E-05	3q29

Footnote: LogFC: Log Fold Change, Adj.P.Val: Adjusted P values, * Targets selected for validation

Table 22: Representative Biological Processes Significantly Altered in Leukoplakia

GO:BP:ID	P value	Odds Ratio	Count	Size	Biological Process
GO:0048869	0.04	1.60	26	3943	Cellular developmental process
GO:0007166	0.00	2.27	18	2126	Cell surface receptor signaling
GO:0006955	0.04	1.84	13	1626	Immune response
GO:0042127	0.04	1.89	12	1458	Regulation of cell proliferation
GO:0051674	0.03	2.04	11	1237	Localization of cell
GO:0051094	0.03	2.06	10	1107	Positive regulation of developmental
GO:0001568	0.00	3.79	9	556	Blood vessel development
GO:0019221	0.00	3.35	8	551	Cytokine-mediated signaling pathway
GO:0030855	0.01	3.15	8	585	Epithelial cell differentiation
GO:0006954	0.03	2.57	7	616	Inflammatory response
GO:0001525	0.01	3.30	6	411	Angiogenesis
GO:0048863	0.01	4.65	5	244	Stem cell differentiation

Table 23: Representative Biological Processes Significantly Altered in OSCC

GO:BP:ID	P value	Odds Ratio	Count	Size	Biological Process
GO:0048869	0.00	1.57	95	3943	Cellular developmental process
GO:0010646	0.00	1.51	70	2905	Regulation of cell communication
GO:0032502	0.02	1.36	68	3486	Developmental process
GO:0051674	0.00	2.45	47	1237	Localization of cell
GO:0007166	0.00	1.86	42	1458	Cell surface receptor signaling pathway
GO:0060341	0.00	2.13	40	1176	Regulation of cellular localization
GO:0007155	0.00	2.02	38	1205	Cell adhesion
GO:1902531	0.04	1.40	37	1572	Regulation of intracellular signal
GO:0050776	0.01	1.71	28	986	Regulation of immune response
GO:0030855	0.00	2.64	25	585	Epithelial cell differentiation
GO:0009968	0.05	1.47	25	1008	Negative regulation of signal
GO:0008544	0.00	5.09	24	307	Epidermis development
GO:0008285	0.00	2.45	24	606	Negative regulation of cell proliferation
GO:0001944	0.00	2.41	22	560	Vasculature development
GO:0008284	0.01	1.79	20	680	Positive regulation of cell proliferation
GO:0048514	0.00	2.65	18	418	Blood vessel morphogenesis

GO: Gene ontology, BP: Biological processes, Count column: Lists the number of genes in our study mapped to have that particular GO term associated to them. The size column represents the number of genes in the original package that are annotated to that GO term.

4.7 Integrative Analysis of Gene Expression and Copy Number Alterations

To determine the genes whose expression are correlated with copy number status, we integrated the aCGH and gene expression data. Since genome wide profiling was done for both types of analysis, they are not biased by genomic position. Of all the chromosomal regions found altered by aCGH analysis we report 3q26.31, 6p21.1, 7p11.2, 8q24.21, 8q24.3, 9p24.1, 11q13.3, 12q13.2, 16q24.2, 17p13.1 as the chromosomal hotspots for copy number dependent 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 underexpression (adjusted p- value <0.25) as represented in Table 24. Chromosomal region 3q26.31 and 12q13.2, showed the strong correlation between copy number dependent over-expression spanning genes *ECT2*, *EIF5A2*, *HOXC9*, *HOXC13* and *MUCL1*. Deletion of 11q22.33 showed few genes with highly significant copy number dependent underexpression, including *CRYAB*, *POU2AF1*, *EXPH5*, *MPZL2* and *ARHGAP32*. The opposite direction expression change was observed for a few genes which includes *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 locus were down regulated, e.g. *CD19*, *GDPD3*, *NUPR1*, *SPN*, *CLDN7* and *DERL3*, suggesting epigenetic regulation of these genes.

Table 24: Copy Number Alterations and Direct Linkage Between Gene Expression Changes.

Cytoband	Alteration	q-values.	DE Genes in Leukoplakia v/s Normal	DE Genes in OSCC v/s Normal
1p36.33*	Amplification	0.00018695	<i>GLTPD1, LOC148413</i>	<i>MXRA8</i>
3q26.31	Amplification	0.17986	<i>CAMK2N2, GOLIM4, KLHL6, CLDN11, EIF4G1, ECT2*, GPR160, EIF5A2*, AP2M1</i>	<i>CHRD, ECT2*, EIF5A2*, KLHL6, NCEH1, GPR160</i>
6p21.1	Amplification	0.13067	<i>YIPF3, NFKBIE</i>	<i>MDFI, VEGFA, C6orf132, PTK7, YIPF3</i>
7p11.2	Amplification	3.25E-08		<i>EGFR</i>
7q11.23	Amplification	1.22E-07		<i>SRRM3, SRCRB4D</i>
7q22.1	Amplification	0.0012742	<i>PCOLCE, GNB2</i>	<i>PCOLCE</i>
8q24.21	Amplification	0.071112		<i>FAM49B</i>
8q24.3*	Amplification	1.12E-06		<i>FBXL6, GPR172A, ADCK5</i>
9p24.1	Amplification	4.88E-05		<i>CD274</i>
11p15.5	Amplification	0.0028178	<i>CHID1</i>	<i>SLC25A22, TMEM80, LSP1</i>
11p11.2	Amplification	0.045436	<i>PACIN3, ARFGAP2</i>	
11q13.1	Amplification	0.0001558	<i>RNASEH2C</i>	
11q13.3*	Amplification	1.59E-25		<i>ANO1</i>
12q13.2	Amplification	0.1623	<i>MUCL1, HOXC9*, HOXC13*, ERBB3, ZNF385A, HOXC10, IKZF4</i>	<i>HOXC9*, MUCL1, HOXC13*, ERBB3, ITGA7, ITGA5, ZNF385A</i>
14q11.2	Amplification	0.00053313		<i>SALL2</i>
14q32.33	Amplification	7.39E-06	<i>INF2</i>	
16p11.2	Amplification	0.19183	<i>CD19, PRSS36, IL27, SPN, SEZ6L2, FUS, CORO1A</i>	<i>CD19, GPD3, NUPR1, SPN, YPEL3, PRRT2</i>
16q24.2	Amplification	0.21747	<i>ANKRD11, CYBA, TCF25</i>	<i>TUBB3, CBFA2T3, SLC7A5, CDT1, GALNS</i>
17p13.1	Amplification	0.12725	<i>CLDN7, CD68, TMEM88, PHF23</i>	<i>CLDN7, CLEC10A, ATP1B2, EFNB3, TMEM88</i>
19p13.3	Amplification	8.44E-08		<i>PPAP2C</i>
22q11.21	Amplification	0.0033888		<i>CLDN5, CDC45</i>
22q11.23	Amplification	0.021241	<i>DERL3*, C22orf43</i>	<i>DERL3*, MMP11, C22orf43</i>
1q44	Deletion	2.19E-08		<i>CNST</i>
2q34	Deletion	0.0059182	<i>SPAG16</i>	<i>IKZF2, SPAG16</i>
3p26.3	Deletion	0.0082783	<i>BHLHE40, C3orf32</i>	<i>CRBN, CAV3</i>
3p21.1	Deletion	0.020653		<i>CACNA2D3, SELK, WNT5A</i>
10p11.21	Deletion	2.68E-05		<i>PARD3</i>
11q22.3	Deletion	0.069519	<i>CARD18, POU2AF1, CTSC, ST3GAL4, THY1, TRIM29, EI24, AMOTL1, PVRL1, ZC3H12C, CHEK1, DLAT</i>	<i>MMP10, MMP3, CRYAB, POU2AF1, EXPH5, FEZ1, CADM1, CTSC, TMEM25, VWA5A, CHEK1, HSPB2, PANX1, AMICA1, CARD17, ABCG4, ARHGAP32, UPK2, MPZL2, ETS1, FLJ1, SCN4B, TRIM29, ME3, C11orf52, C11orf54, THY1, FZD4, OAF, PAFAH1B2, NLRX1, LOC283143, AMOTL1</i>
17p13.1	Deletion	0.020653	<i>ALOX12, CLDN7, VMO1, HS3ST3A1, CD68, GPR172B, SLC25A11, C17orf59, TMEM88, ATP2A3, UBE2G1, PHF23</i>	<i>ALOX12, SPNS2, CLDN7, XAF1, GPR172B, GAS7, ALOX12B, UBE2G1, CLEC10A, PMP22, USP43, ATP1B2, C17orf59, EFNB3, AURKB, GGT6, TMEM88, ATP2A3</i>
17q21.31	Deletion	3.63E-07		<i>ETV4</i>
21q21.3	Deletion	0.01729	<i>ADAMTS5</i>	<i>JAM2, NRIP1, CYR1</i>

DE: Differential Expression, **Red font** indicates amplified locus or upregulated genes, **Blue font** indicates locus loss or downregulated genes. Few genes showed opposite direction expression changes and are coloured respectively (e.g genes in blue font were downregulated but located in the amplified region, while genes in red font were seen upregulated, but located in the deleted region) and genes listed in black font showed same direction expression change as that of locus). * Targets selected for validation

4.8 Validation of Microarray Results

Targets for validation were selected based on the following criteria:

- 1) Validation of differentially expressed genes associated with OSCC progression
 - Copy number independent expression change
 - Copy number dependent expression change
- 2) CNA associated with OSCC progression and clinical outcome (lymph node metastasis and patient survival).

Brief layout of the rationale for validation (targets and methods employed) is summarized below:

Copy Number Independent Changes/ Progression	<ul style="list-style-type: none"> •Down-regulation of <i>KRT76</i>, <i>MAL</i> and <i>DERL3</i> •Up-regulation of <i>MFAP5</i>, <i>NELL2</i> and <i>INHBA</i> (qRT-PCR and/ or IHC)
Copy Number Dependent Changes/ Progression	<ul style="list-style-type: none"> •8q24.3 Gain (FISH) •<i>ECT2</i>, <i>EIF5A2</i>, <i>HOXC9</i>, <i>HOXC13</i>, <i>LY6K</i>, <i>FUS</i> and <i>DVL1</i> (qRT-PCR and/ or IHC)
Clinical Outcome (Survival and Lymph Node Metastasis)	<ul style="list-style-type: none"> •1p36.33, 11q13.3 and 11q22.2 Gain (FISH) •<i>BIRC2</i> (cIAP1) and <i>BIRC3</i> (cIAP2) (qRT-PCR and IHC)

4.8.1 Validation of Differentially Expressed Genes Associated with OSCC Progression

From amongst the genes that were common either in leukoplakia and OSCC or common to early stage OSCC and advanced stage OSCC, we selected 11 genes for validation. The selection of targets was either based on the novelty of our findings or influenced by studies of these genes in the cancer related literature. To confirm the results of genome-wide gene expression analysis, quantitative RT-PCR (TaqMann assays) was done in both the test and the validation set for 7-up-regulated genes (*DVL1*, *EIF5A2*, *FUS*, *HOXC9*, *INHBA*, *LY6K* and *MFAP5*), 3-

downregulated genes (*DERL3*, *KRT76* and *MAL*), 1-unchanged gene (*SLC4A1AP*) and 1-endogenous control (*18S rRNA*). The detailed study of *KRT76* validation is provided in section 4.8.1.2.

The demographic details of the qRT-PCR study are provided in Table 26. Before initiating the experiment, PCR efficiency of the target and the internal control gene was checked by serial dilution with 5-log dilutions and were found to be similar. The slopes of all the standard curves were between -3.2 and -4.1 with a correlation coefficient of at least 0.99 and above.

The Log2FC (Fold Change = $2^{-\Delta\Delta CT}$) values were compared between leukoplakia and OSCC for the target genes using *18S* as an endogenous control, with normal cases as a reference (Figure 25). The detailed log 2FC difference for all the genes across the groups is represented as the box and whisker plots in Figure 26. There was no notable difference in the *SLC4A1AP* gene expression supporting our microarray findings. *HOXC9*, *MFAP5* and *INHBA* showed a very high fold expression change in all the three groups as compared to normal. Significant overexpression of *EIF5A2* and *LY6K* was observed only in early and advanced stage OSCC. However, for *DVL1* and *FUS* the log2FC was not very high and was consistent with the microarray data. Similarly, down-regulation of *MAL* and *DERL3* was confirmed by qRT-PCR. The association between all the target genes with clinicopathological parameters is provided in the Table 25. Most of the targets (*EIF5A2*, *HOXC9*, *MFAP5*, *LY6K*, *INHBA* and *DVL1*) demonstrated a positive correlation with OSCC progression, while *DERL3* and *MAL* had negative relation with respect to the progression. *EIF5A2*, *HOXC9*, *INHBA*, and *MFAP5* were also associated with disease advancement (early stage OSCC to advance stage OSCC). In addition *EIF5A2*, *HOXC9*, *INHBA*, *FUS* and *DVL1* were significantly correlated with lymph node metastasis (Table 25).

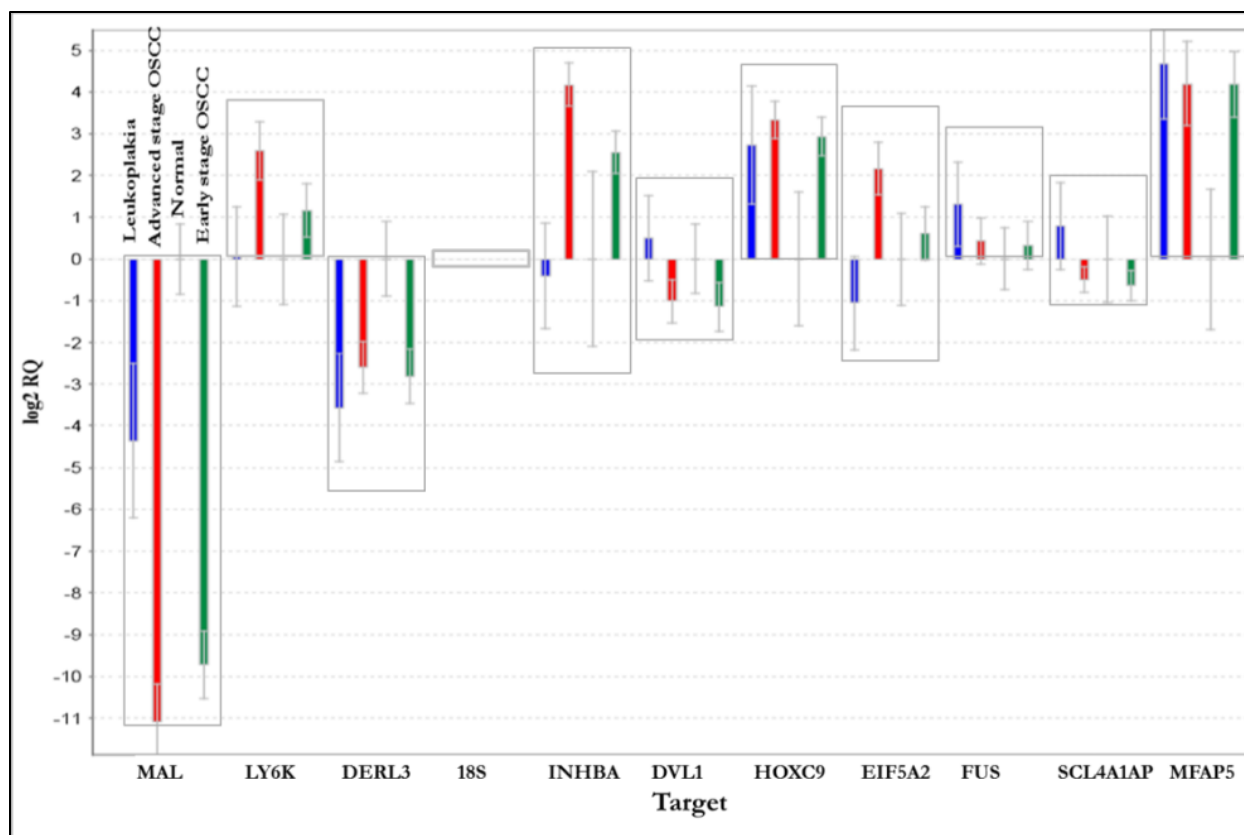


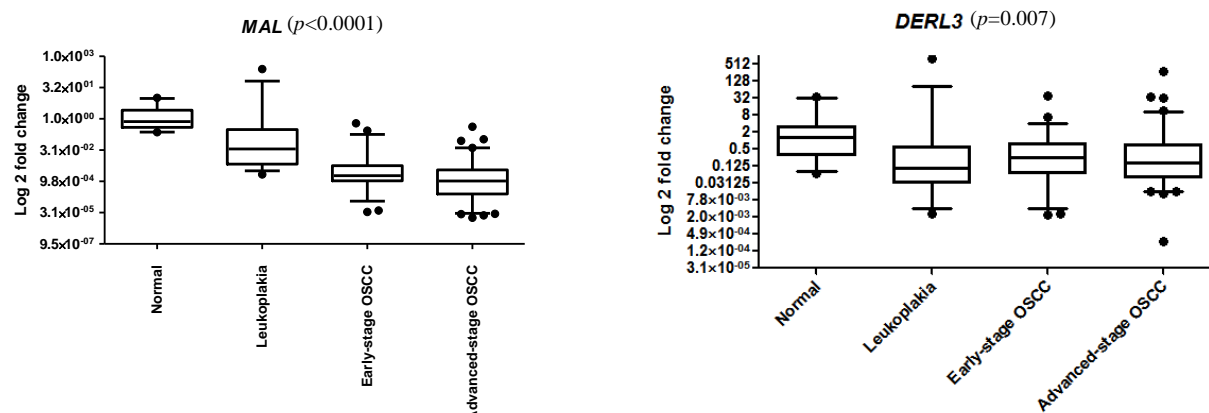
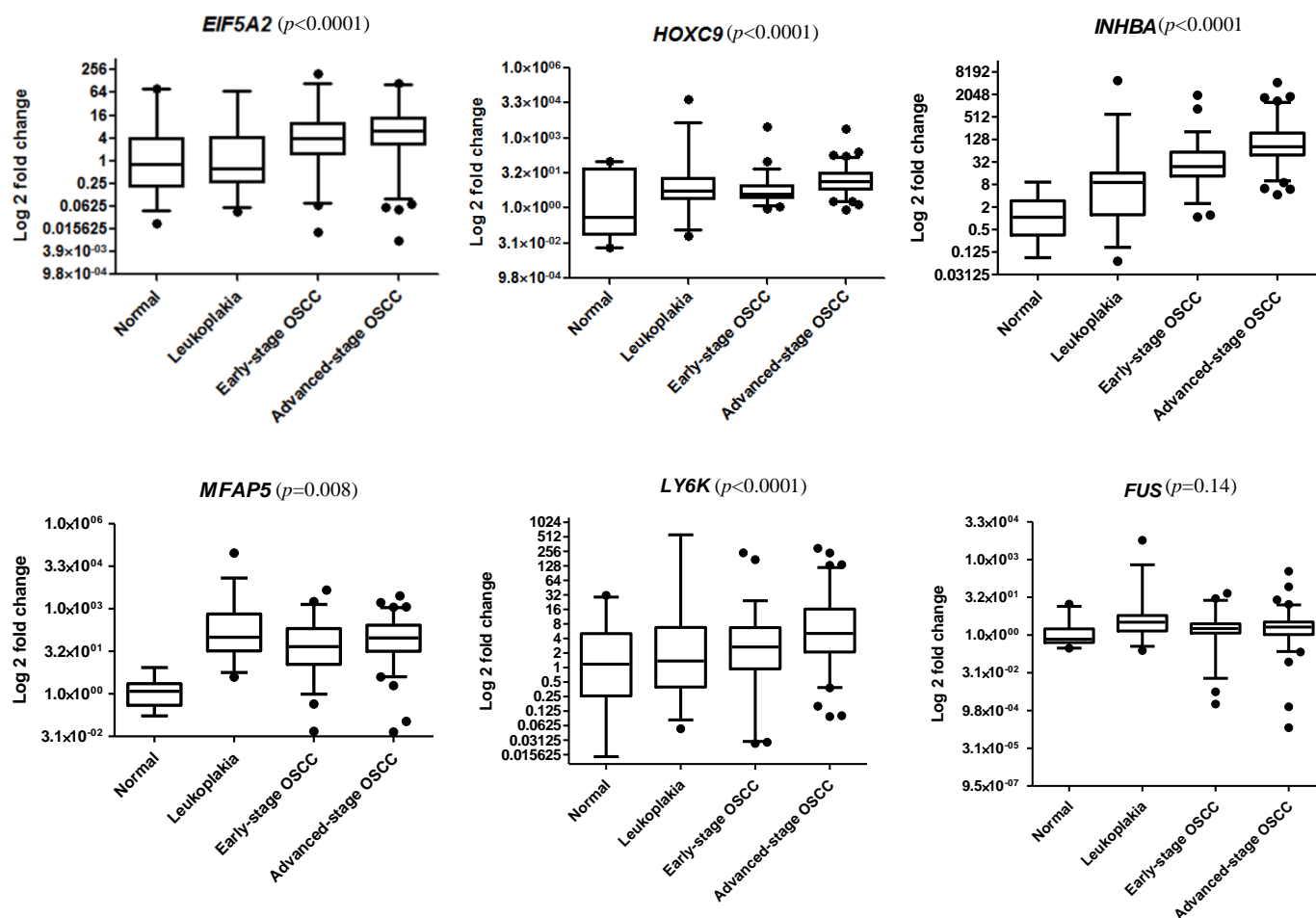
Figure 25: Relative Quantitation (RQ) of Validation Targets. Bars represent the log₂FC (+/- SD) in the gene expression of target genes in Leukoplakia (Blue), Early stage OSCC (Green) and Advanced stage OSCC (Red) with reference to Normal. 18S rRNA was used as an endogenous control.

Table 25: Correlation of qRT-PCR validation targets with clinicopathological parameters.

Targets (Genes) →	EIF5A2	HOXC9	INHBA	MFAP5	LY6K	FUS	DVL1	SLC4A1AP	DERL3	MAL
Clinicopathological Parameters ↓	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)	<i>p</i> value(cof*)
OSCC Progression	<0.001 (0.39)	<0.001 (0.35)	<0.001 (0.68)	0.008 (0.19)	<0.001 (0.36)	0.141 (0.1)	0.03 (-0.15)	0.540 (-0.04)	0.007 (-0.19)	<0.001 (-0.68)
Tumor Stage	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)	0.13 (0.21)	0.444 (-0.06)	0.007 (-0.23)
Node metastasis	0.018 (0.2)	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)
Tumor Grade	0.181 (0.12)	0.25 (-0.1)	0.062 (0.16)	0.62 (0.04)	0.5 (0.06)	0.815 (0.02)	0.144 (-0.12)	0.98 (-0.02)	0.418 (0.07)	0.629 (-0.04)

cof* : Spearman correlation coefficient (positive/ neagtive value represents, positive or negative correlation); significant changes (bold font)

Down-regulated genes

Up-regulated genes and unchanged gene (*SLC4A1AP*)

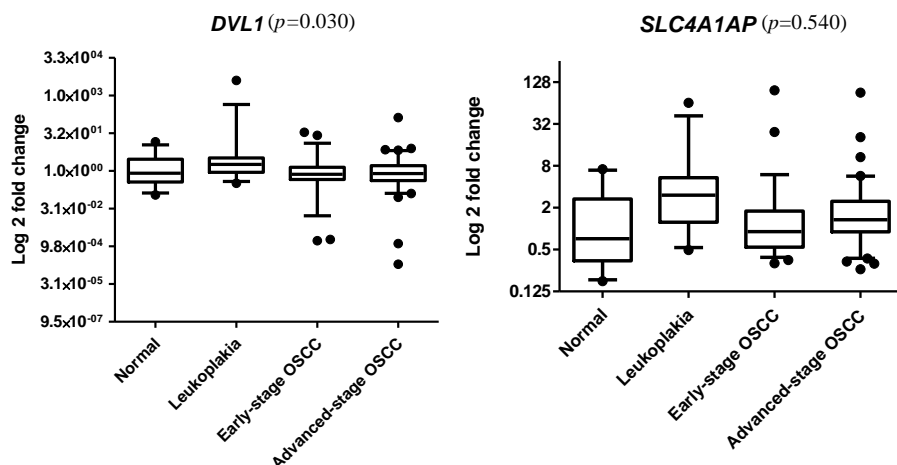


Figure 26: Log 2 Fold Change of Target Genes in Normal, Leukoplakia and OSCC. Box and whisker plot demonstrating the log 2FC in the gene expression (y-axis), with whiskers depicting 5-95 percentile. The p value was calculated using a Spearman correlation analysis.

4.8.1.1 Validation of Protein Overexpression of the Target Genes

Immunohistochemical analysis was performed to validate the protein overexpression (copy number dependent and independent) of the selected targets in normal, leukoplakia and OSCC. The demographic details of the study samples are provided in Table 26. The targets with copy number dependent up regulation were EIF5A2 and ECT2 on 3q26.31 locus, HOXC9 and HOXC13 on 12q13.2 locus. *MFAP5* and *NELL2* showed very high fold copy number independent gene expression change. All these targets were strongly associated with disease progression and single target from each locus was validated by qRT-PCR, as described in the previous section (4.8.1). The IHC analysis revealed strong protein expression of the targets in leukoplakia and OSCC as compared to the normal tissues Figure 27 and Figure 28.

Table 26: Demographic and Clinicopathological Characteristic of the Validation Study Group.

Patients Characteristic	qRT-PCR (n=207)		IHC & FISH (n=370)	
	Leukoplakia n=37	OSCC n=138	Leukoplakia n=108	OSCC n=185
Age at diagnosis				
Median age	41	50	44	50
Range (IQR)*	33-53	42-59	32-57	42-60
Gender				
Male	33 (89.2%)	102 (75.6%)	96 (90.6%)	140 (75.7%)
Female	4 (10.8%)	33 (24.4%)	10 (9.4%)	45 (24.3%)
Pathological Stage				
Stage 1 and 2 (Early stage OSCC)	NA	56 (41.5%)	NA	67 (36.2%)
Stage 3 and 4 (Advanced stage OSCC)	NA	79 (58.5%)	NA	118 (63.8%)
Pathological T classification				
T1	NA	25 (18.5%)	NA	24 (13%)
T2	NA	64 (47.4%)	NA	80 (43.2%)
T3	NA	4 (3%)	NA	8 (4.3%)
T4	NA	42 (31.1%)	NA	73 (39.5%)
Pathological Cervical Lymph Node involvement				
Node Negative (N0)	NA	79 (58.5%)	NA	112 (60.5%)
Node Positive (N+)	NA	56 (41.5%)	NA	73 (39.5%)
Pathological Grade				
Well	NA	12 (8.9%)	NA	23 (12.5%)
Moderate	NA	87 (64.4%)	NA	106 (57.6%)
Poor	NA	36 (26.7%)	NA	55 (29.9%)
Hyperplasia	31 (86.1%)	NA	80 (80.8%)	NA
Mild Dysplasia	3 (8.3%)	NA	9 (9.1%)	NA
Moderate Dysplasia	2 (5.6%)	NA	8 (8.1%)	NA
Severe Dysplasia	NA	NA	2 (2%)	NA
Habit Profile				
No habit	NA	3 (2.6%)	NA	8 (5.3%)
Exclusive tobacco users	13 (41.9%)	79 (70%)	30 (33%)	98 (64.5%)
Exclusive smoker	5 (16.2%)	2 (2%)	12 (13.2%)	4 (2.6%)
Exclusive alcohol drinker	NA	1 (0.8%)	NA	1 (0.7%)
Mixed Habit	13 (41.9%)	28 (24.7%)	49 (53.8%)	41 (27%)
Buccal mucosa Normals: n=32 (qRT-PCR) and n= 77 (IHC); * IQR: Inter Quartile Range; Details regarding number of samples overlapping with aCGH and GE study is provided in study design				
qRT-PCR: <i>DERL3, EIF5A2, FUS, HOXC9, INHBA, LY6K, MFAP5, SLC4A1AP, 18S, BIRC2</i> (excluding leukoplakia), <i>BIRC3</i> -(excluding leukoplakia),				
IHC: <i>EIF5A2, ECT2, HOXC9, HOXC13, MFAP5, NELL2</i>				
FISH: 8q24.3, 1p36.33 (excluding leukoplakia), NA: Not Applicable,				

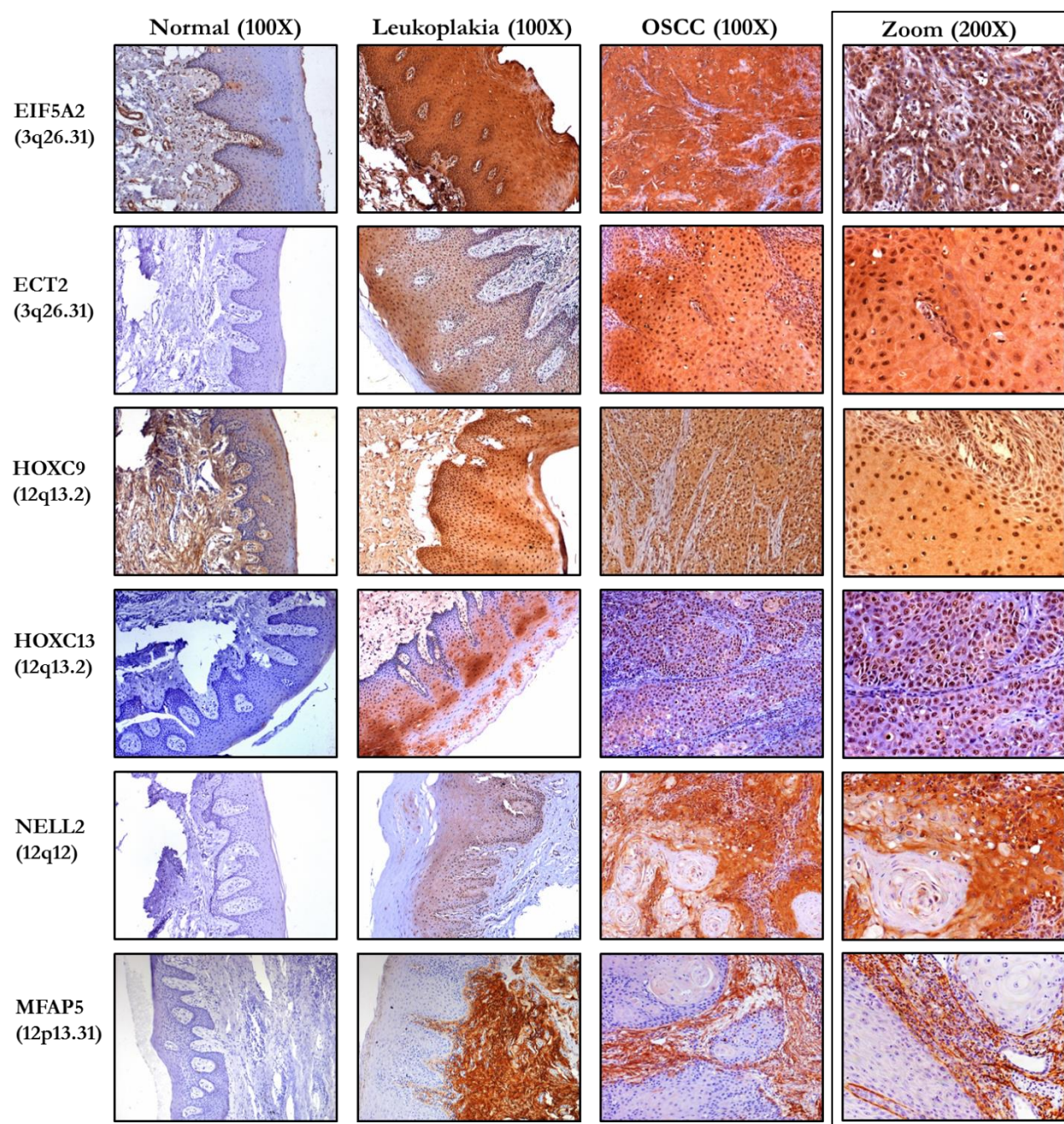
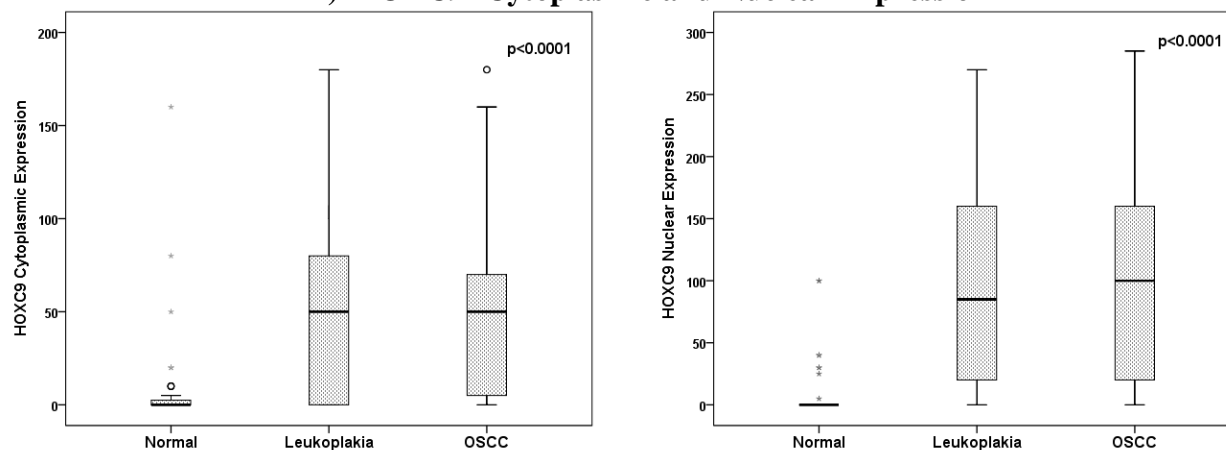
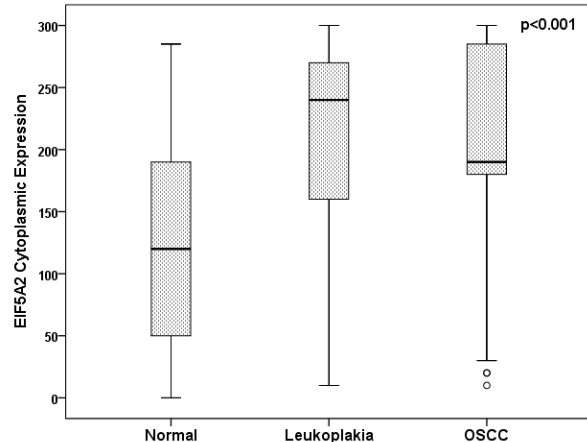
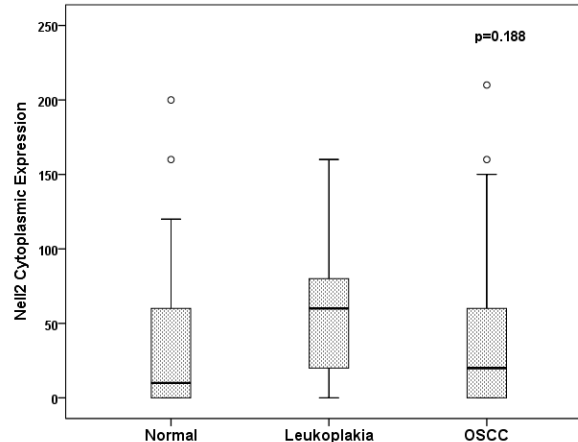
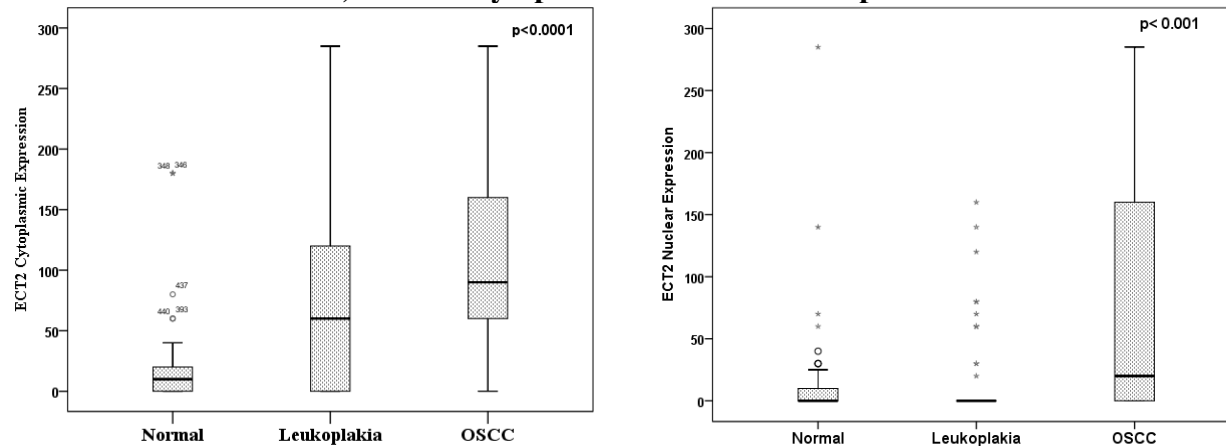


Figure 27: Immunohistochemical Analysis for Target Validation. Representative IHC staining of the targets in Normal, Leukoplakia, and OSCC (first three panels). Last panel showing differential localization of each marker. NELL2 showed cytoplasmic staining while EIF5A2, ECT2, HOXC9, HOXC13 showed cytoplasmic and nuclear expression and MFAP5 was localized predominantly into the matrix. Respective isotype controls for all the cases had no staining (images not shown). Magnification 100X (first three panels) and magnification 200X (last panel).

A) HOXC9- Cytoplasmic and Nuclear Expression**B) EIF5A2- Cytoplasmic Expression****C) NELL2- Cytoplasmic Expression****D) ECT2- Cytoplasmic and Nuclear Expression**

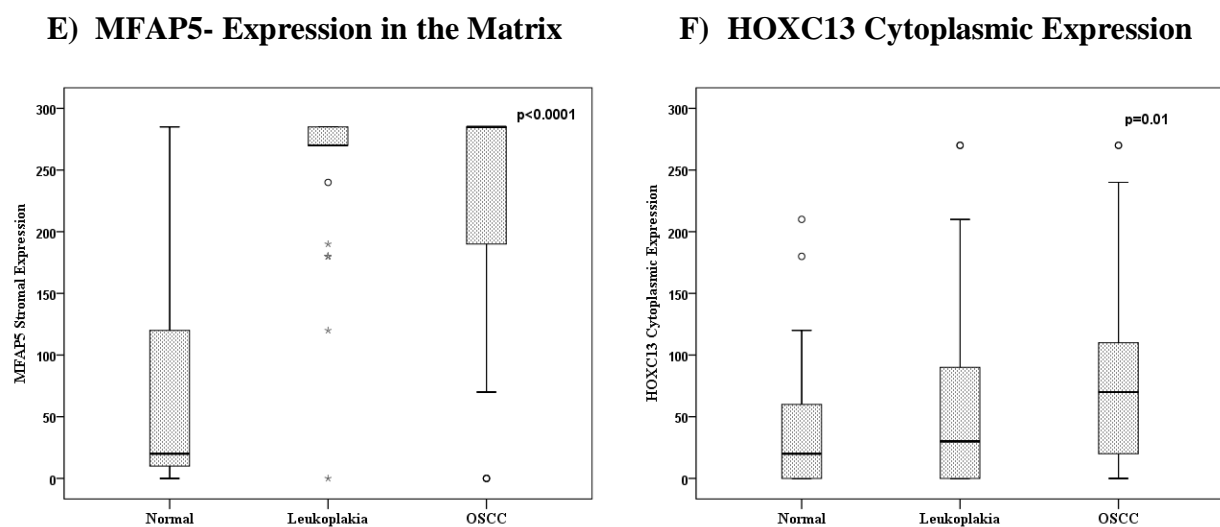


Figure 28: Correlation Between Protein Overexpression and Oral Cancer Progression. **A)** Significant increase in cytoplasmic and nuclear HOXC9 was observed in leukoplakia and OSCC compared to the normal buccal mucosa tissue. **B)** Cytoplasmic overexpression of EIF5A2 was more prominent in leukoplakia and OSCC tissues. **C)** Increase in expression of NELL2 was observed in leukoplakia and OSCC, however the difference was statistically not significant. **D)** ECT2 cytoplasmic overexpression observed in leukoplakia and OSCC and nuclear overexpression was pertinent to OSCC. **E)** Very high expression of MFAP5 in the matrix was observed in leukoplakia and OSCC. **F)** Increased cytoplasmic expression of HOXC13 was observed in leukoplakia and OSCC. (y-axis represents the H-score for protein expression calculated as described in materials and methods; the p value was calculated using a Spearman correlation analysis).

4.8.1.2 Differential Expression of *KRT76* (K76) in OSCC Progression

In addition to the targets described in previous section, copy number independent downregulation of *KRT76* was observed in advanced stage OSCC. *KRT76*, a type II epithelial keratin (K76), is specifically expressed in the suprabasal cell layers of oral masticatory epithelium lining the gingiva and the hard palate. We validated our microarray finding in an independent cohort of gingivobuccal cancers using qRT-PCR and IHC. The demographic and clinicopathological details of the study group are provided in Table 27.

Table 27: Demographic and Clinicopathological Characteristic of the *KRT76* Study Group.

Characteristics	qRT-PCR OSCC (n = 57) [†]	IHC(n = 163) [†]	
		OSCC(n = 102)	OPL(n = 61)
Gender			
Males	40 (70%)	80 (78.4%)	55 (90.2%)
Females	17 (30%)	22 (21.6%)	6 (9.8%)
Age			
Median (IQR) [#]	52 (43.5–57.5)	52 (41.7–64)	45(34.5–56.6)
Habit profile			
Exclusive Chewers	46 (80.7%)	34 (59.7%)	19 (32.7%)
Exclusive Smokers	3 (5.3%)	4 (7%)	12 (20.7%)
Chewing and Smoking	8(14%)	19 (33.3%)	27 (46.6%)
Grade			
Well	2 (3.5%)	12 (11.7%)	–
Moderate	39 (68.4%)	65 (63.7%)	–
Poor	16(28.1%)	25 (24.6%)	–
Nodal involvement			
Negative (N0)	29 (50.9%)	49 (48.0%)	–
Positive (N+)	28 (49.1%)	53 (52.0%)	–
Stage (pTNM)			
I & II	3 (5.3%)	13 (12.74%)	–
III & IV	54 (94.7%)	89 (87.26%)	–

[†]Shown is the number of cases, except for Age,

[#]IQR: Interquartile range.

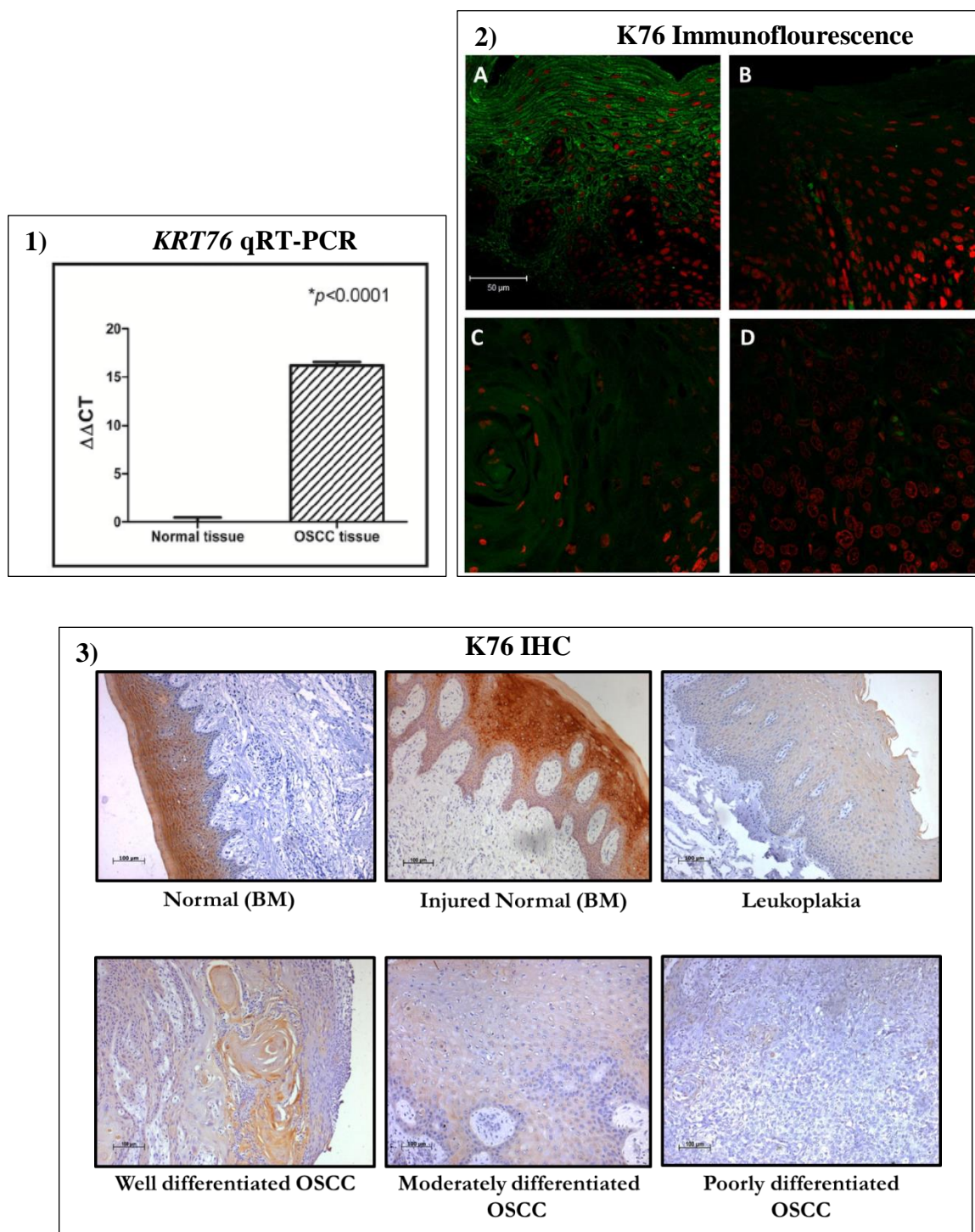


Figure 29: Validation of KRT76 Downregulation using Multiple Methods. 1) qRT-PCR analysis showed more than 15 fold downregulation of *KRT76* expression in tumors compared to normal oral tissue. 2) Representative Immunofluorescent staining of A) Normal oral tissue (Buccal mucosa (BM), B) OPL(leukoplakia), C) Well differentiated tumor, D) Poorly

differentiated tumor. K76 (Stained green, Alexa fluor 488), Nuclei stained with DAPI (pseudo red). Magnification 200X (Scale: 50 mm). **3)** Representative IHC staining of K76 in normal buccal mucosa, injured normal, leukoplakia and OSCC with different grades. Respective isotype controls for all the cases had no staining (images not shown). Magnification 100X (Scale: 100 mm).

qRT-PCR analysis revealed significant downregulation of *KRT76* RNA in tumor samples compared to the normal samples (Figure 29, panel 1). K76 protein expression was analyzed in 184 oral tissues, by immunohistochemistry (Figure 29 panel 3). All the IHC staining were graded by the pathologist. Normal gingivobuccal tissues expressed higher levels of K76 protein compared to leukoplakia and invasive OSCC. Distribution of K76 expression was confirmed by immunofluorescence (Figure 29, panel 2). Normal oral epithelium showed K76 expression confined to the suprabasal, differentiating cell layers while there was a gradual overall loss of K76 expression in leukoplakia and tumors. The frequency of K76 positive staining significantly decreased across the transition from normal tissue (100% positive) to leukoplakia (44%) to oral tumor (35%) (Figure 30 (A)). To examine whether *KRT76* downregulation was associated with benign epithelial hyperproliferation, IHC was performed on the injured normal tissue without any association with oral preinvasive and invasive lesions. Even though these epithelia histologically appeared hyperproliferative, K76 staining was consistent with that seen in normal buccal epithelium (Figure 29 panel 3). These results indicate that downregulation of K76 expression is not associated with injury related proliferation and acute inflammation.

Statistical analysis to determine the association of K76 expression and different clinical parameters, such as node, stage, grade, habit profile and outcome (recurrence and survival) was performed. Reduced expression of K76 showed a very weak association with survival ($p = 0.096$) (Figure 30 (B)), whereas other parameters analyzed did not show any association. Polytomous

logistic regression with normal as the reference group showed a significant correlation of K76 downregulation with risk of developing OPL ($p = 0.002$) and OSCC ($p < 0.0001$) (Table 28).

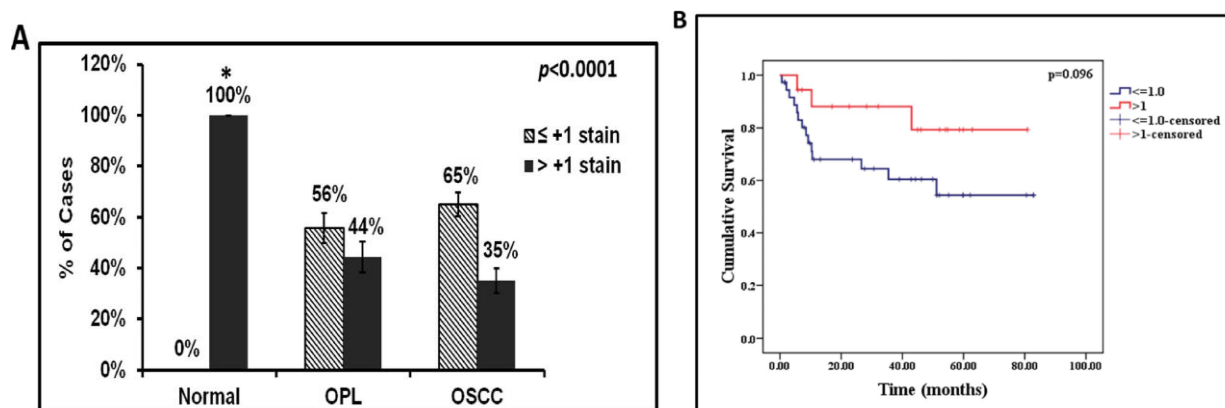


Figure 30: Correlation Between Loss of K76 Expression with Oral Cancer Development and Patient Survival. A) Significant downregulation of K76 was observed in OSCC and OPL(leukoplakia) compared to normal. B) Kaplan–Meier plot for DSS of gingivobuccal cancer patients with respect to K76 IHC staining intensity. *All normals showed more than +2 grade stain.

Table 28: The Effect of K76 Expression Loss with Development of Oral Lesions.

K76 staining	Normal (n = 21)	OPL (n = 61)	OR	95% CI	p value	OSCC (n = 102)	OR	95% CI	p value
High(>1)	21	27	1	3.4–216.7	0.002	36	1	5.1–307	<0.0001
low(≤1)	0	34	27			66	40		

Polytomous logistic regression performed using normal as reference group indicated significant increase in risk of developing OPL and OSCC with decrease in staining.

To understand association of K76 expression with disease progression, IHC analysis was performed on the buccal epithelium of 2, 4, 6, 8, 10, 12 and 16 weeks DMBA treated hamsters and control from each week (Tissues obtained from Dr. G. B. Maru, TMC, ACTREC). Irrespective of duration of treatment, the control group showed higher levels of K76, while

reduced expression was observed in premalignant lesions and oral tumors, which was similar to that seen in human hyperplastic lesions and OSCC (Figure 31).

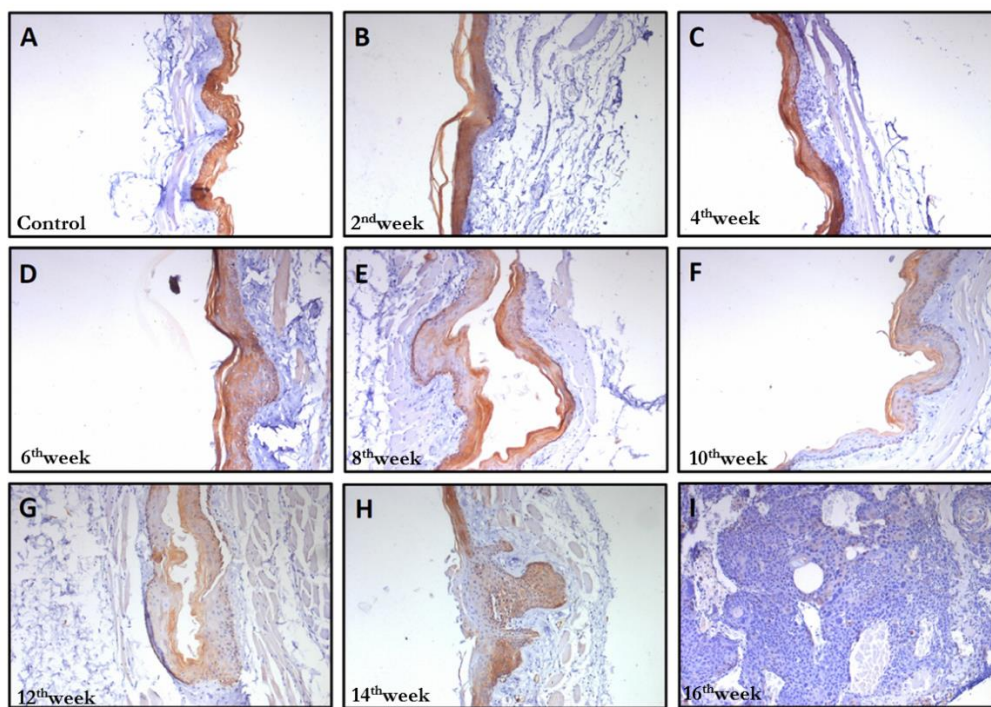


Figure 31: Sequential Downregulation of K76 Expression During Tumor Development in Hamster Buccal Epithelium. Gradual decrease in K76 IHC staining was observed in different weeks, (2nd week, 4th week, 6th week, 8th week, 10th week, 12th week, 16th week) of DMBA treated buccal epithelium (B-I); whereas the control of all weeks showed consistent higher staining (A). Respective isotype controls for all the cases had no staining (images not shown).

To determine whether loss of *KRT76* is sufficient to induce premalignant lesions in the oral cavity, we examined the oral epithelia of KRT76-KO and KRT76-WT mice (obtained from Dr. Fiona Watt, CRI, Cambridge). Immunohistochemical analysis showed specific K76 staining in buccal epithelium of WT mice, whereas no staining was observed in KRT76-KO buccal epithelium, confirming specificity of K76 antibody (Figure 32 A and B). Histological examination of the buccal mucosa of KRT76-KO mice showed development of hyperplastic

lesions along with increased keratinization across the epithelium, which was not observed in KRT76-WT mice (Figure 32 C and D). In contrast, the epithelium of the dorsal tongue, which is normally KRT76-negative, exhibited normal homeostasis in KRT76-KO mice, indicating that KRT76 loss associated abnormalities are highly sub-site specific in oral cavity (Figure 32 E and F). However, none of the KRT76-KO mice in the entire life span developed spontaneous oral tumors. Our current findings suggest that the loss of *KRT76* may not be a sole molecular event leading to oral cancer development. Thus, future studies to investigate the contributing role of KRT76 in light of other tumor driving events are warranted.

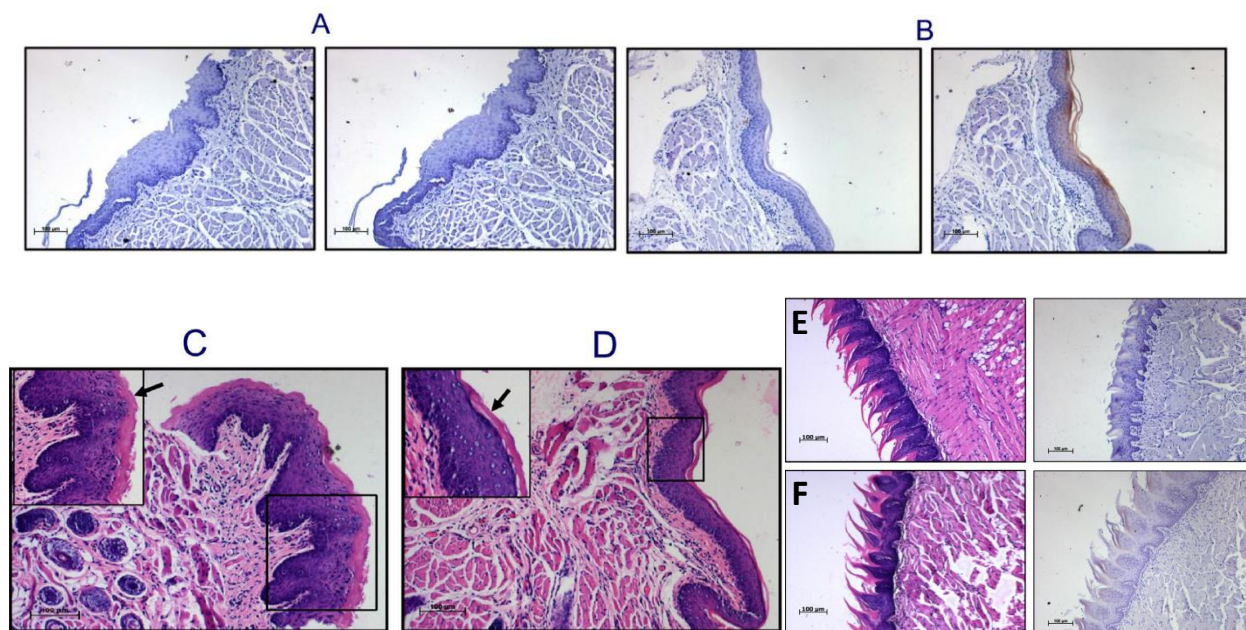


Figure 32. KRT76-KO Mice show Hyperplastic Lesions in Oral Epithelium. K76 antibody specificity was determined by IHC on Oral epithelium of KRT76-KO mice, which did not show any staining (A), whereas wild type mice of same strain showed moderate staining (B), with respective isotype control (left panel A & B). Histological observation of H & E stained buccal epithelium demonstrated hyperplastic changes and increased keratinization in KO (C), compared to WT (D). E and F: Histology of KO and WT mice dorsal tongue, along with respective K76 IHC staining. Magnification 100X (Scale: 100 mm); selected area under 200X magnification.

4.8.2 Copy Number Alterations Associated with OSCC Progression

Validation of 8q24.3 locus gain was carried out using interphase FISH in leukoplakia and OSCC (demographic details of study samples are provided in the Table 26). The centromere- (G101034G-8; Green) and region-specific (G101197R-8; Red) probes hybridized to their target loci and showed no cross reactivity (Figure 33 (A)). Signal enumeration was performed as described in the methods section 3.8.4, and tissues were categorized as normal, gained or amplified for 8q24.3 locus (Figure 33 C- G along with respective zoom image).

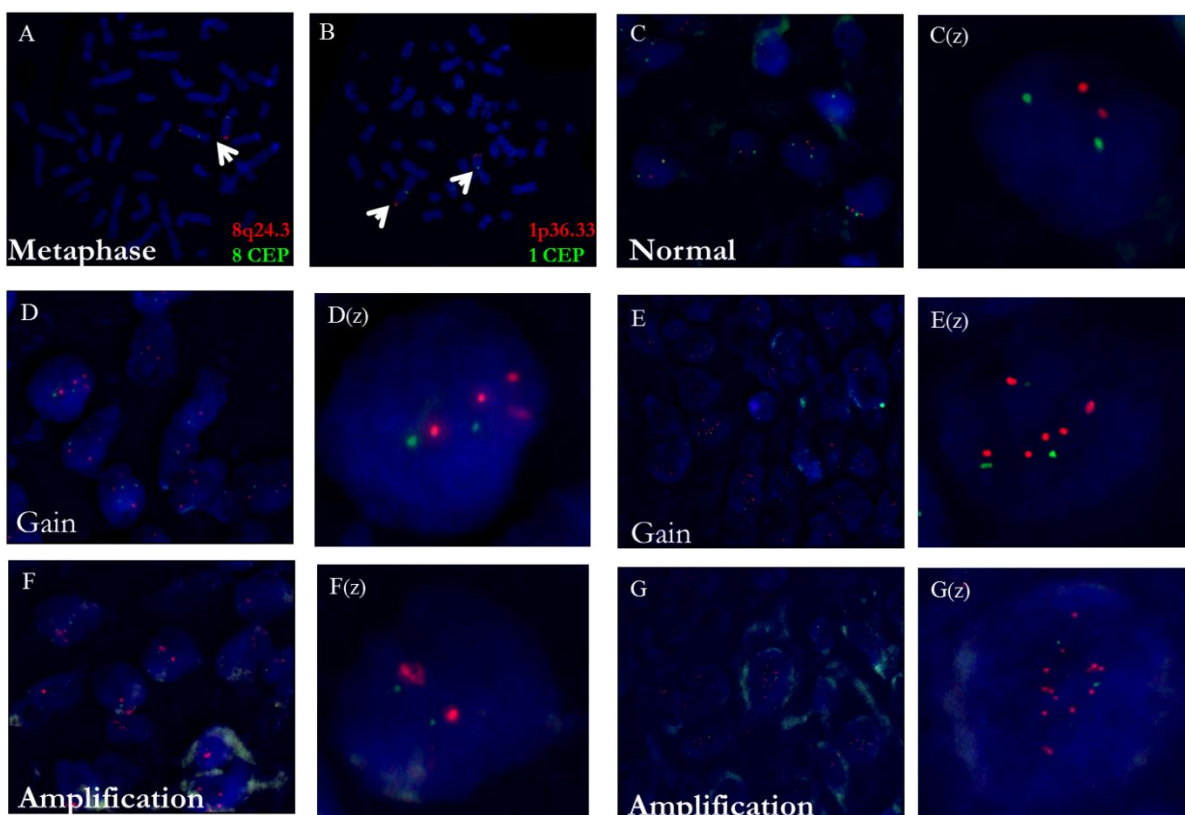


Figure 33: Interphase FISH for Validating Gain of 8q24.3 and 1p36.33. Specificity of Locus-probes (Red) and Centromere -CEP probe (Green) was confirmed by hybridizing to metaphase spreads (A and B). FISH signals were categorized as normal diploid nuclei (C), gain- 3 to 8 locus signal (D and E), amplification- cluster of signals that cannot be enumerated or more than 8 for locus signal (F and G) and represented along with respective zoom (z). Enumeration of

FISH signal for both locus and the centromere was performed using a single filter in at least 100 cells from different field per tissue. Magnification 630X (Single nuclei zoom).

We demonstrated that gain of 8q24.3 locus was associated with disease progression ($p < 0.001$).

The percentage of cells with 8q24.3 gain increased as the disease progresses from leukoplakia to OSCC, and almost 95% of the advanced stage OSCC had gain in more than 40% tumor cells (Figure 34). Additionally, we observed that moderate and poorly differentiated tumors had a higher percentage of 8q24.3 gain than well differentiated tumors ($p = 0.046$). Almost all the node positive tumor had more than 40% tumor cells with 8q24.3 gain ($p = 0.007$).

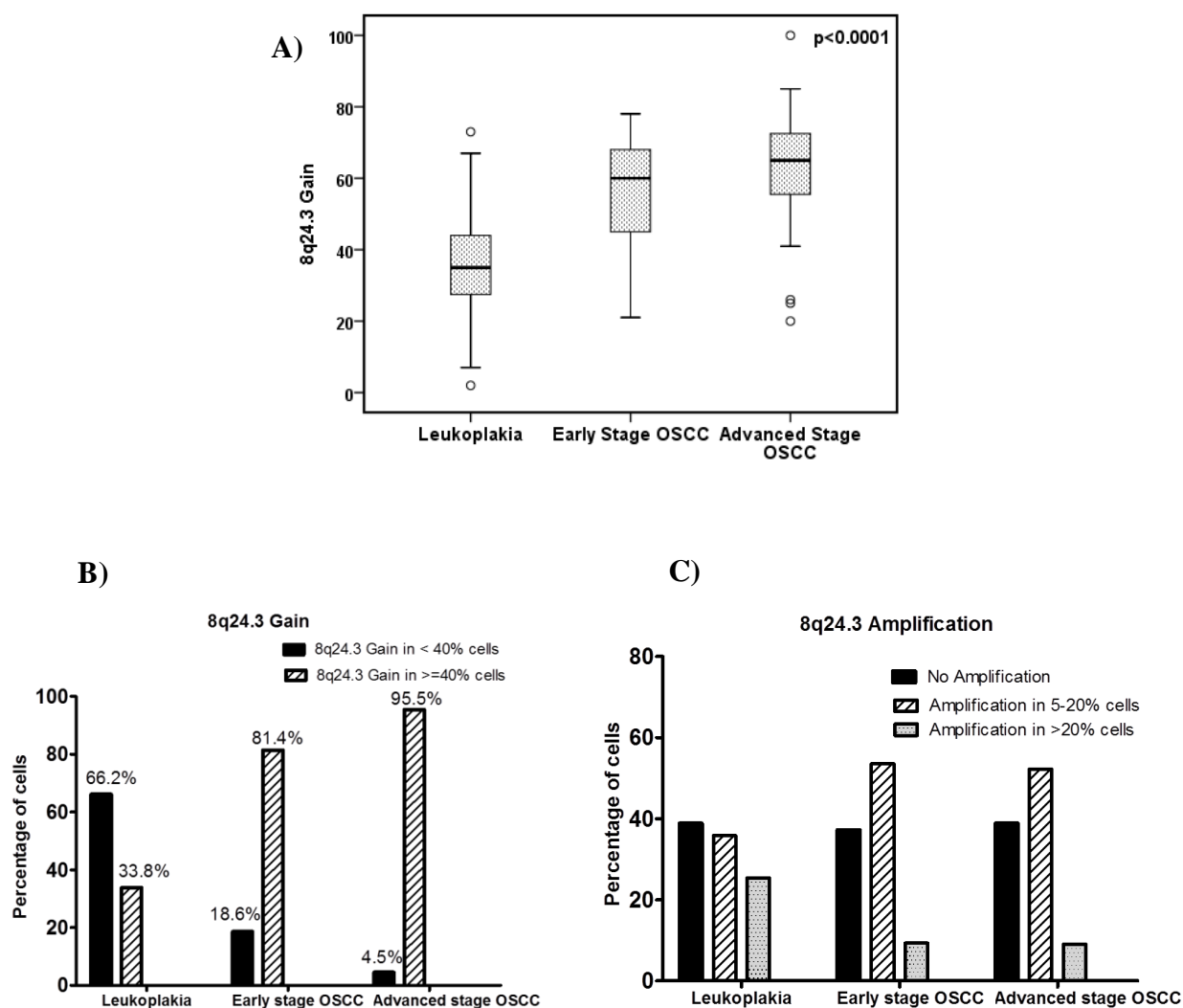


Figure 34: Correlation Between 8q24.3 Alteration and Oral Cancer Progression. A) Box plot representing the percentage of cells with 8q24.3 gain across different groups, the p value is

calculated using Spearman correlation analysis. B) Bar plot representing the increased percentage of tumor cells with 8q24.3 gain, as the disease progresses, C) Bar plot representing the percentage of cells with 8q24.3 amplification in leukoplakia and OSCC (Percentage cutoff described in methods 3.8.4). y axis represents the percentage of altered tumor cells per tissue.

We observed amplification of 8q24.3 in leukoplakia and OSCC tissues, however the percentage of cells with amplification was very low about 5 to 20% (Figure 34 (C)). Our validation results were consistent with the aCGH observation, indicating that 8q24.3 gain is a prominent event than its amplification and it is an important early event associated with OSCC progression.

4.8.3 Copy Number Dependent Changes Associated with Clinical Outcome

aCGH analysis revealed a strong association of 1p36.33, 11q13.3, 11q22.1-11q22.2 gain with clinical outcome (Table 19). These targets were validated by interphase FISH to understand their association with patients survival/ lymph node metastasis and are described in detail in the following sections.

4.8.3.1 Validation of 1p36.33 Gain and its Correlation with Patient Survival

Validation of 1p36.33 locus gain was carried out using interphase FISH in OSCC (demographic details of study samples are provided in the Table 26). The centromeric- (G101064G-8; Green) and region-specific (G101182R-8, Red) probes hybridized to their target loci and showed no cross reactivity (Figure 33 (B)). Signal enumeration was performed as described in the methods section 3.8.4, and tissues were categorized as normal, gained or amplified for 1p36.33 locus (Figure 33 (C- G) along with respective zoom image). We confirmed that 1p36.33 was associated with poor survival ($p=0.049$) in OSCC patients in an independent cohort (Figure 35

(A)). Although, 83.3% tumors with recurrence had 1p36.33 gain in more than 40% tumor cells, but we did not see a significant association with recurrence free survival (Figure 35 (B and C)). When compared with clinicopathological parameters, a strong association of 1p36.33 gain was observed with lymph node metastasis ($p=0.001$). Detailed study to understand the role of 1p36.33 gain with the OSCC patient outcome will further strengthen this observation.

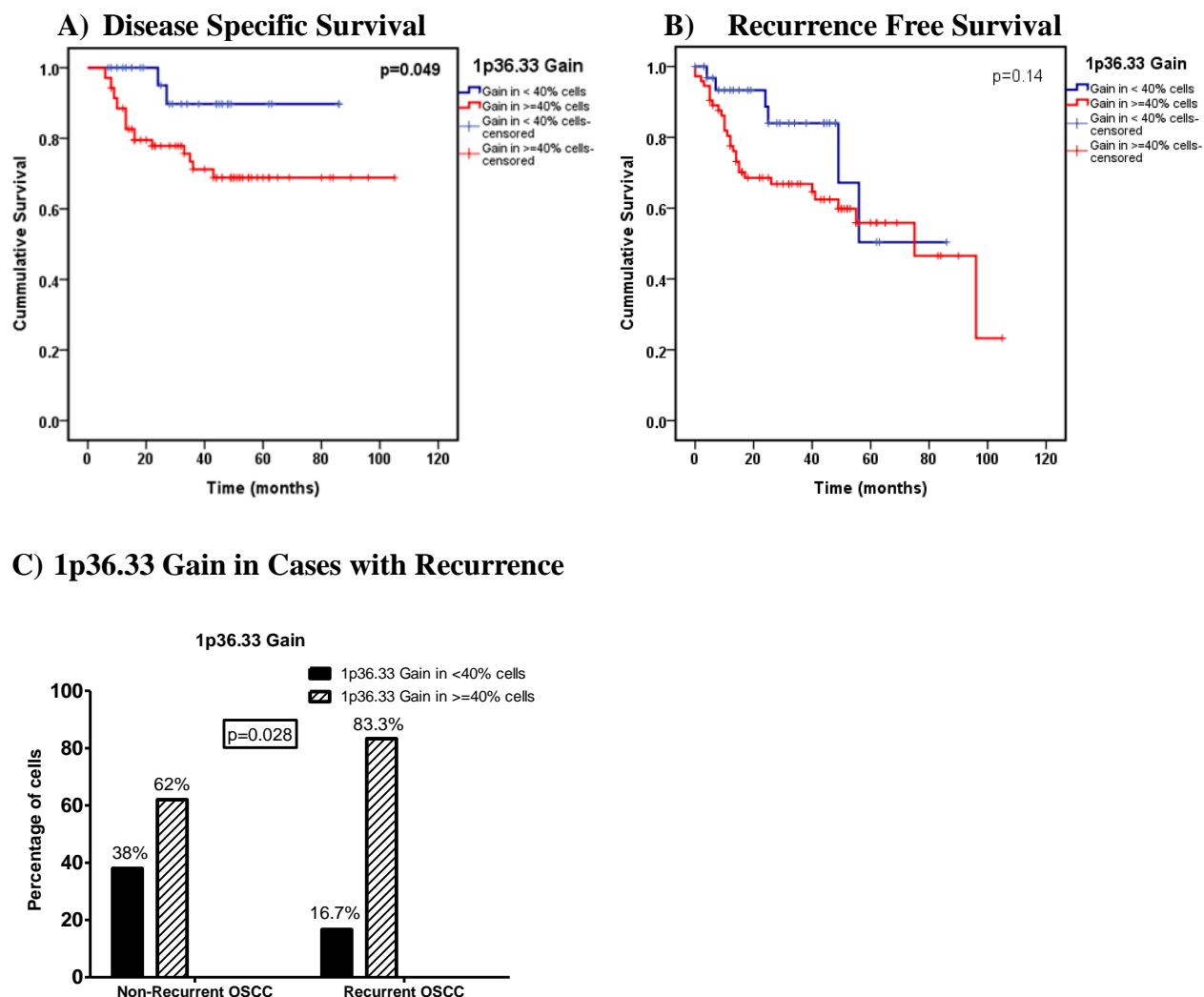


Figure 35: Correlation Between 1p36.33 Gain with Clinical Outcome in Oral Cancer Patients. A) and B) Kaplan-Meier estimates of the probability of survival and recurrence in patients with gain of chromosome locus 1p36.33. Survival in months (x-axis) is plotted against the fraction of samples alive (y-axis). C) Bar plot representing the percentage of cells with

1p36.33 gain in OSCC patients with or without recurrence. y axis represents the percentage of alter tumor cells per tissue.

4.8.3.2 Validation of Driver Genes on 11q Amplicon and Correlation with Node Metastasis and Patient Survival

aCGH data revealed multiple aberration at chromosome 11 of which 11q13.3 and 11q22.1-11q22.2 were significantly altered and associated with poor survival and lymph node metastasis in OSCC patients. We validated 11q alterations using interphase FISH (I-FISH) analysis, the demographic details of the study is provided in Table 29. The centromeric- (RP11-135H8; Red) and region-specific probes 11q13.2-q13.3 (RP11-300I6; Green); 11q22.1-q22.2 (RP11-90M3; Green), hybridized to their target loci and showed no cross reactivity (Figure 36 A and B).

Table 29: The Demographic Details of 11q Validation Study Group.

Characteristics	Total cases FISH n=182 (%)	Total cases IHC n=209 (%)	Characteristics	Total cases FISH n=182 (%)	Total cases IHC n=209 (%)
Patient cohort			Grade		
OSCC	182	132	Well	11 (6.1)	9 (6.8)
OPL	NA	57	Moderate	126 (69.2)	82 (62.1)
Healthy Normals	NA	20	Poor	45 (24.7)	41 (31.1)
Gender			Treatment		
Male	138 (75.8)	171 (81)	Surgery	52 (28.6)	5 (3.9)
Female	44 (24.2)	40 (19)	Surgery + RT [†]	78 (42.9)	79 (62.2)
Age			Surgery + CT [‡]	4 (2.2)	NA
Mean (IQR)	50.03 (42-58)	47.79 (39-57)	Surgery + RT + CT	48 (26.5)	43 (33.9)
Site of Primary			Radiation Compliant Cases		
GBC*	150 (82.4)	162 (85.7)	50-60 Gy	93 (51.1)	85 (64.9)
Tongue	32 (17.6)	27 (14.3)	<50 Gy	89 (48.9)	2 (1.5)
Nodal status			No Information	NA	44 (33.6)
N0	94 (51.6)	79 (59.8)	Habit Profile		
N+	88 (48.4)	53 (40.2)	Exclusive Chewers	63 (34.6)	66 (28.9)
Pathological Stage			Exclusive Smokers	3 (1.6)	16 (7)
I + II	33 (18.1)	25 (18.9)	Mixed Habitués	32 (17.6)	62 (27.2)
III + IV	149 (81.9)	107 (81.1)	No Habit	3 (1.6)	8 (3.5)
			No information	81 (44.5)	37 (16.2)

GBC: Gingivobuccal Cancer; IQR: Inter Quartile Range; RT: Radiotherapy; CT: Chemotherapy

†Mix habitués: Patients with at least two of the habits smoking, chewing, and drinking; ‡Gy: Gray.

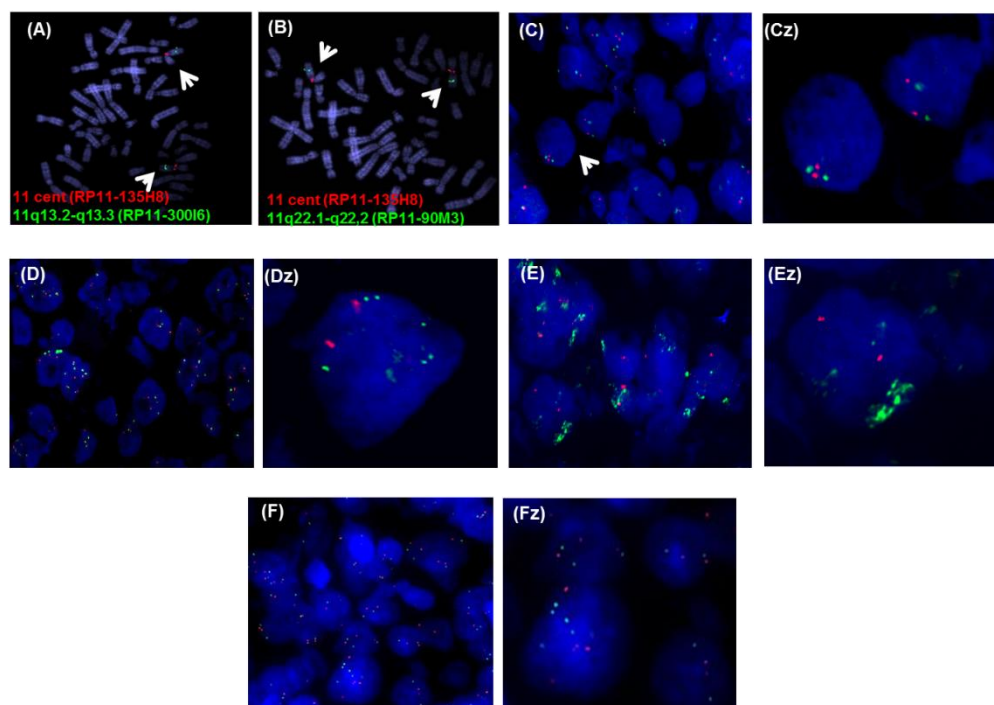


Figure 36: Interphase FISH for Detecting Gain of 11q13.3 and 11q22.2. A) and B) Metaphase plate with FISH probes confirming the specificity of clones for 11q centromere (red) and locus 11q13.3 / 11q22.2 (green). FISH signals were categorized as normal (diploid nuclei) (C), gain- 3 to 10 locus signal (D), amplification- cluster of locus signals that cannot be enumerated (E) and polysomy - multiple locus and centromeric signals (F). Respective single nuclei zoom (Cz, Dz, Ez and Fz) accompanying the original image. Enumeration of FISH signal for both locus and the centromere was performed using a single filter in at least 100 cells from different field per tissue. Magnification 630X.

Table 30: FISH Signal Enumeration and the Percentage of 11q13.3 and 11q22.2 Alteration in Validation Study Group.

Alteration	11q13 n (%)	11q22 n (%)
Change	72 (53)	60 (37.5)
-Gain	40 (55)	32 (53)
-Amplification	23 (31)	19 (32)
-Polysomy	9 (13)	9 (15)
No change	71 (47)	100 (62.5)

Change: Alteration of respective locus (Gain, Amplification or Polysomy); No change: No alteration at respective locus (Normal).

4.8.3.3 Association of 11q Alteration with Disease Progression and Lymph Node Metastasis

For analysis alteration for both 11q13.3 and 11q22.2 were categorised as either 1) No change (diploid nuclei) and change (gain, amplification and polysomy combined); 2) No change and amplification (Figure 36, (C to F) with respected zoom). The frequency of each of the alteration category is provided in the Table 30. Statistical analysis was performed between 11q13.3/11q22.2 locus no change and change (gain, amplification, polysomy) or amplification at both this locus. The analysis revealed a positive association of 11q13.3 change and amplification with disease advancement (Table 31 (A)). This observation was consistent with aCGH results. Further, both 11q13.3 and 11q22.2 alteration (change and amplification) demonstrated significant association with lymph nodal metastasis (Table 31 (B)). To understand the interdependence of 11q13.3 and 11q22.2 alteration on nodal metastasis, binary logistic regression was performed with three categories 1) Cases with both the locus altered, 2) Cases with only 11q13.3 altered and no change at 11q22.2 locus, 3) Cases with only 11q22.2 altered and no change at 11q13.3 locus. This analysis showed a strong association of 11q22.2 on node metastasis irrespective of 11q13.3 alteration status (Table 32).

Table 31: Association of 11q13.3 and 11q22.2 Alteration with Clinicopathological Parameters (A and B respectively).

A)	Clinicopathological Parameter	No Change	11q13 Change [‡]	<i>p</i> * Value	11q13 Amplification	<i>p</i> * Value
	Pathological Stage					
	I & II	18	6	0.006	0	0.003
	III & IV	53	66		23	
	Grade					
	Well	6	3	NS [#]	1	NS [#]
	Moderate	49	47		13	
	Poor	16	22		9	
	Nodal status					
	N0	45	27	0.002	7	0.006
	N+	26	45		16	

B)	Clinicopathological Parameter	No Change	11q22 Change [‡]	<i>p</i> * Value	11q22 Amplification	<i>p</i> * Value
	Pathological Stage					
	I & II	23	8	NS [#]	2	NS [#]
	III & IV	77	52		17	
	Grade					
	Well	9	2	NS [#]	0	0.018
	Moderate	72	39		10	
	Poor	19	19		9	
	Nodal status					
	N0	66	18	<0.0001	5	0.002
	N+	34	42		14	

[‡]: Change: Includes Gain, Amplification & Polysomy of respective locus; No change: No alteration of the respective locus

*Two tailed Fisher's exact test. #NS: Not Significant

Table 32: Correlation of 11q13.3 and 11q22.2 Alterations with Lymph Node Metastasis.

Variable	Total cases	N+ cases (%)	OR (95% CI)	<i>p</i> * Value
11q22 & 11q13 Both Change[‡]	88	36 (40.9)	7.05 (2.72-18.29)	<0.0001
11q22 change & No 11q13 change	63	22 (34.92)	7.23 (1.67-31.21)	0.008
11q13 change & No 11q22 change	78	27 (34.61)	2.71 (1.01-7.25)	0.047

*Binary logistic regression analysis; [‡]: Change: Includes Gain, Amplification & Polysomy; OR: Odds Ratio, N+: Node metastasis positive

1.8.3.3.1 Association of 11q Alterations with Patients Survival and Tumor Recurrence

Amplification of 11q22.2 was found as the strongest predictors of poor clinical outcome in terms of recurrence ($p=0.026$) and survival ($p=0.004$), specifically in radiation compliant patients (patients who have completed >50 Gy radiation treatment) (Table 33). Kaplan-Meier survival curves (Disease Specific Survival-DSS and Recurrence Free Survival-RFS) for 11q22.2 alterations are shown in Figure 37. Multivariate Cox proportional hazards models were used to assess the effect of multiple parameters in contribution of 11q22.2 alteration, and observed that in addition to nodal metastasis, 11q22.2 amplification was an independent predictor of patient survival (Table 34).

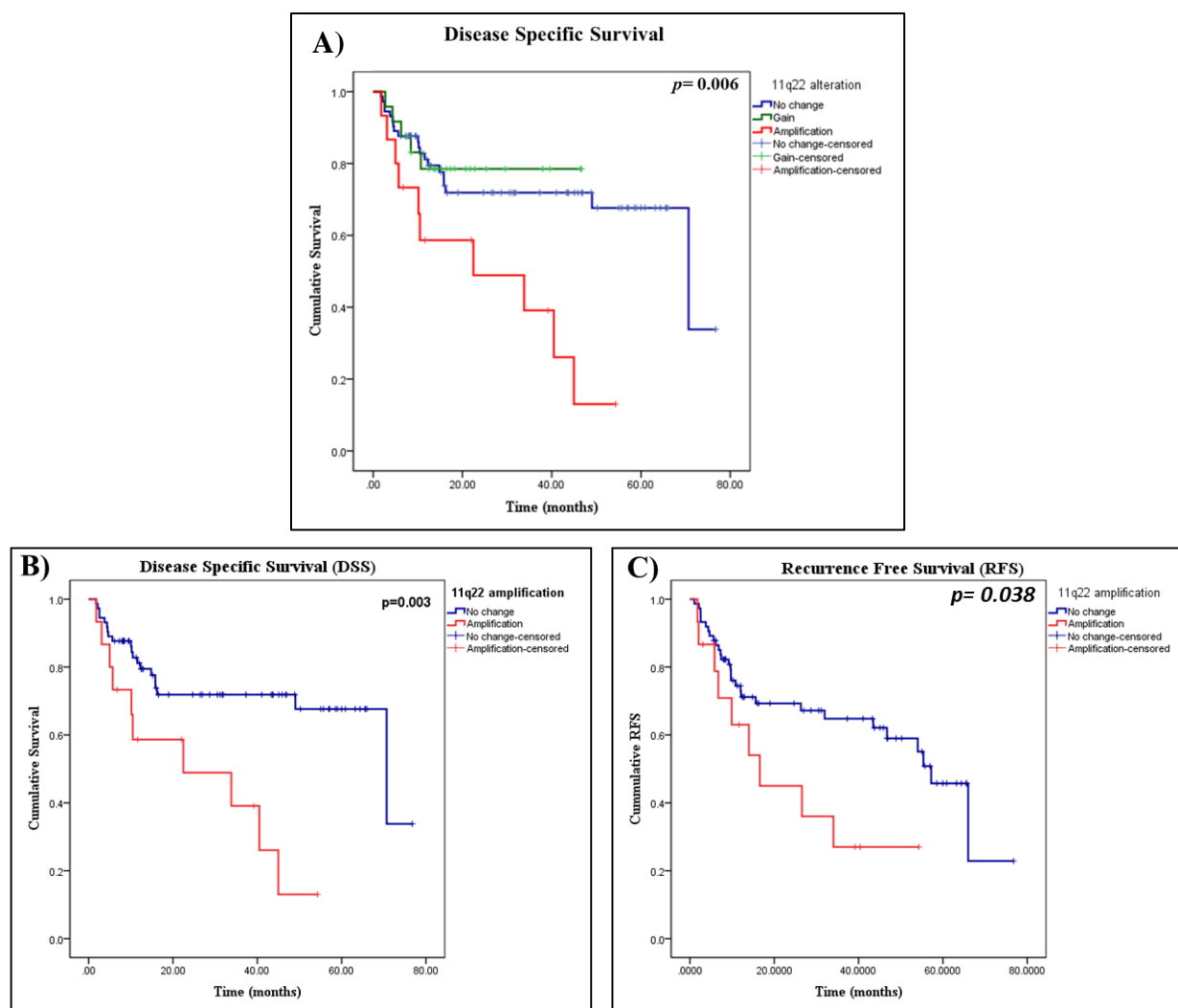


Figure 37: Kaplan-Meier Estimates of Patients Survival with Alterations on Chromosome 11q22.2. A). DSS of patient groups without 11q22.2 alteration or with 11q22.2 gain and amplification B). DSS of patient groups without 11q22.2 alteration or 11q22.2 amplification. C) RFS of patient without 11q22.2 alteration or with 11q22.2 amplification. Survival in months (x-axis) is plotted against the fraction of samples alive (y-axis).

Table 33: Association of 11q22 Amplification with Patients DSS and RFS.

	Disease Specific Survival			Recurrence Free Survival		
	Total/ Death (%)	HR (95% CI)	p^* value	Total/ Relapse (%)	HR (95% CI)	p^* value
Total cohort	115/ 35(30.4)	3.068 (1.42-6.62)	0.004	-	-	NS [†]
Radiation Compliant Patients	68/ 19(27.9)	4.944 (1.64-14.88)	0.004	79/ 32(40.2)	2.848 (1.13-7.15)	0.026

*Cox regression analysis; HR: Hazard Ratio

Table 34: Multivariate Cox Regression Analysis to Identify the Association of Multiple Parameter on Clinical Outcome.

Clinicopathological Parameter	Disease Specific Survival (n= 60)	
	HR (95% CI)	p* value
Nodal Status	3.897 (1.03-14.73)	0.045
Tumor Stage	0.671 (0.12-3.49)	NS [#]
11q13 Amplification	0.460 (0.08-2.59)	NS [#]
11q22 Amplification	7.126 (1.21-41.71)	0.029

* Multivariate Cox Regression Analysis; #NS: Not Significant; HR: Hazard Ratio

1.8.3.3.2 *BIRC2* and *BIRC3*: Targets on 11q22.2 Amplicon

The human 11q22.2 amplicon include cluster of matrix metalloproteinase (*MMP*) genes and two members of the BIRC family (*BIRC2* and *BIRC3*, a protein known as cIAP1 and cIAP2 respectively). *MMP3* gene overexpression is well reported in OSCC in addition, our GE data have also demonstrated their very high fold change in OSCC samples. Hence we validated this finding using semi quantitative RT-PCR and IHC. Although we observed a significant increase in *MMP3* expression in OSCC compared to normal (Figure 38), but no significant association with lymph node metastasis or patient survival was observed. Detailed analysis and statistical data not shown.

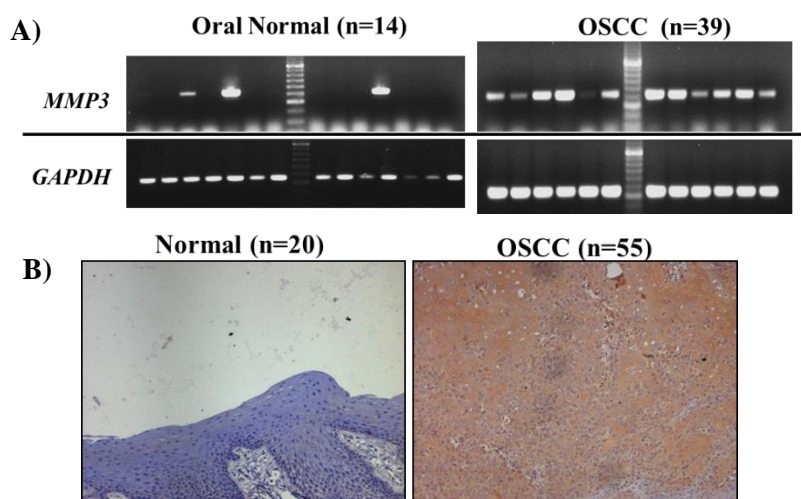
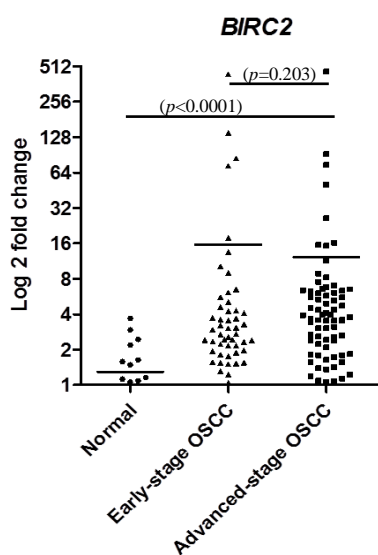


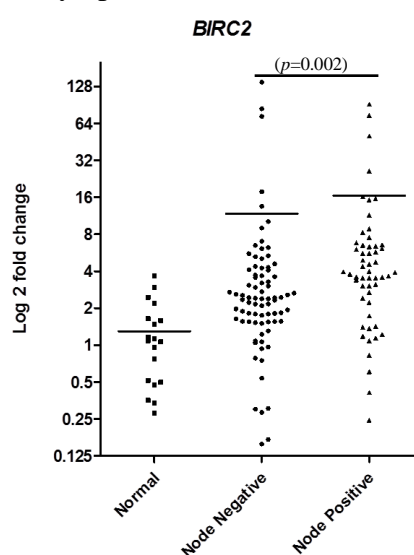
Figure 38: Expression of MMP3 in Oral Cancers. A) *MMP3* gene expression analysis using semiquantitative RT-PCR using *GAPDH* as reference in normal and OSCC. B) The representative IHC staining depicting MMP3 protein expression in normal and OSCC.

The next tentative targets on 11q22.2 amplicon were cIAP1 and cIAP2. These are multi-functional proteins, which not only regulates caspases and apoptosis, but also modulates inflammatory signaling and immunity, mitogenic kinase signaling, cell proliferation, invasion and metastasis [110-113]. The expression of cIAP1/cIAP2 is correlated with invasion in esophageal and cervical cancers [22, 23]. Reports support the fact that overexpression of these proteins may be associated with an unfavourable prognosis after radiotherapy and/or chemotherapy [114-117]. We performed qRT-PCR based analysis and IHC to confirm the gene and protein overexpression of both cIAP1 and cIAP2. *BIRC2* and *BIRC3* expression was significantly increased in OSCCs as compared to normals, and may increase the risk of OSCC development (*BIRC2*: $p=0.001$, OR=0.133, 95% CI=0.042-0.425; and *BIRC3*: $p=0.002$, OR=0.166, 95% CI=0.052-0.528). Additionally, both *BIRC2* and *BIRC3* up-regulation was associated with lymph node metastasis and node metastasis positive tumors and had a mean FC higher than the node metastasis negative tumors ($p<0.002$ and $p<0.007$) and represented in scatter plots (Figure 39). *BIRC2* and *BIRC3* up-regulation were observed as the strongest predictors of poor clinical outcome in terms of recurrence and overall survival in OSCC patients. Kaplan-Meier survival curves (Disease Specific Survival-DSS and recurrence free survival-RFS) for *BIRC2* and *BIRC3* gene expression are shown in Figure 40.

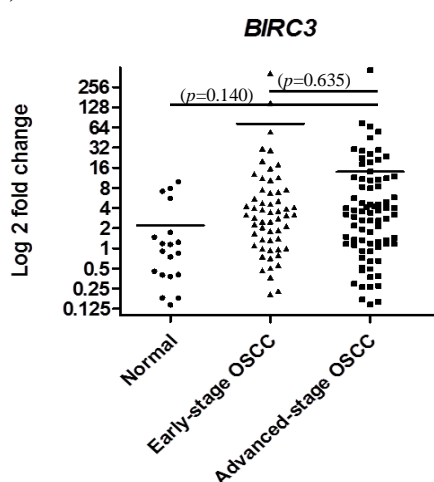
A) Disease advancement



B) Lymph node metastasis



C) Disease advancement



D) Lymph node metastasis

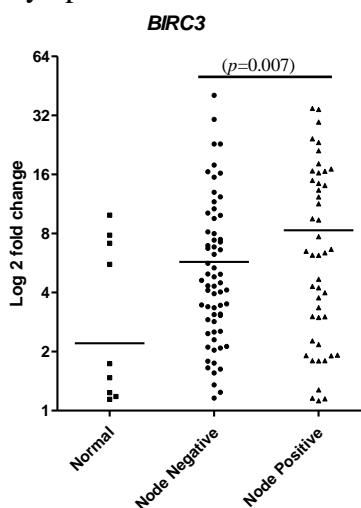


Figure 39: Correlation of BIRC2 and BIRC3 Expression with Disease Advancement and Lymph Node Metastasis. Scatter plot with mean and SD depicting the log 2 fold change (y-axis) of *BIRC2* and *BIRC3* expression and its association with disease advancement and lymph node metastasis.

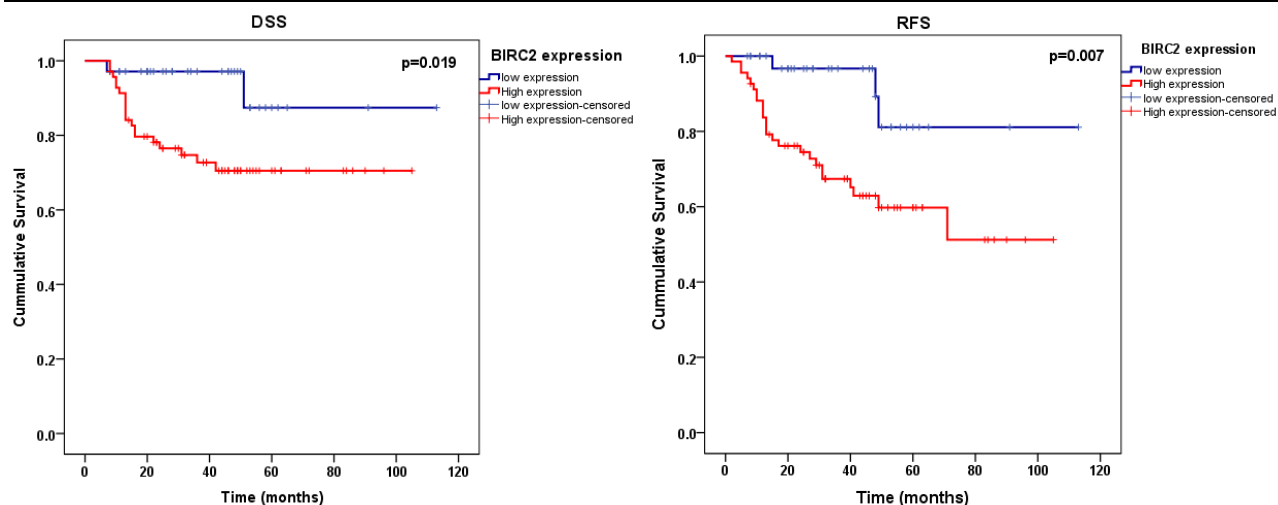
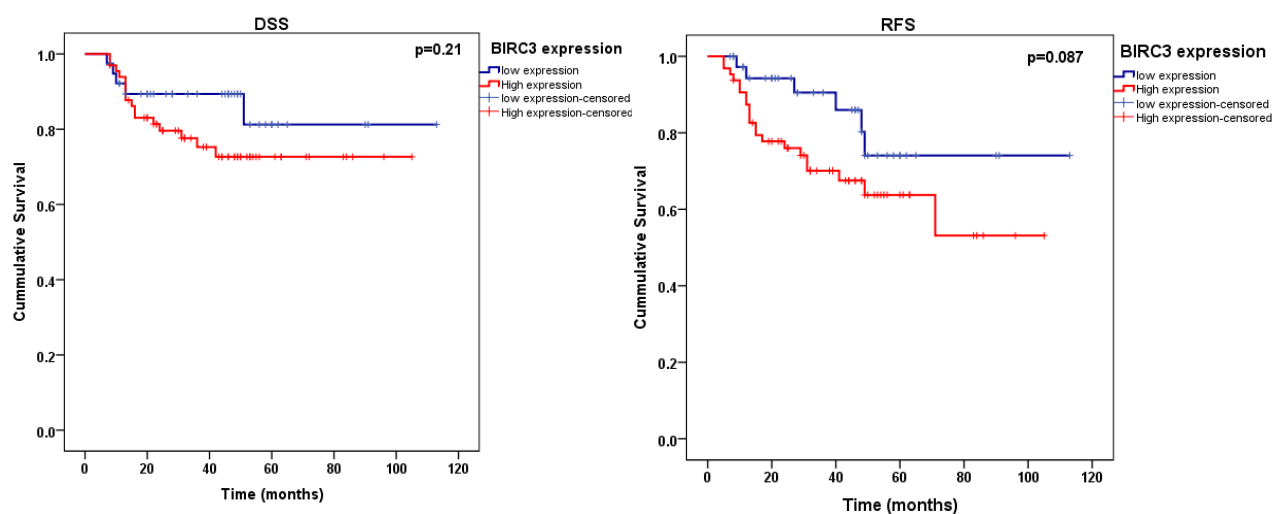
A) *BIRC2* expression with Disease Specific Survival (DSS) and Recurrence Free Survival (RFS)B) *BIRC3* expression with Disease Specific Survival (DSS) and Recurrence Free Survival (RFS)

Figure 40: Kaplan-Meier Estimates of Patient Survival with *BIRC2* and *BIRC3* overexpression. A) DSS and RFS of patient groups with *BIRC2* gene upregulation. **B)** DSS and RFS of patient groups with *BIRC3* gene upregulation. DSS/RFS in months (x-axis) is plotted against the fraction of samples alive/ without recurrence (y-axis).

The demographic details of the cIAP1 and cIAP2 analysis is provided in Table 29. IHC analysis demonstrated concurrent increased in expression of cIAP1 and cIAP2 across the transition from normal to OSCC as observed by qRT-PCR (Figure 41). Interestingly, differential localization of cIAP1 and cIAP2 was observed in normal, leukoplakia and OSCC. cIAP1 was localized to the nucleus and the cytoplasm of cells, while cIAP2 was seen in cytoplasm and at the cell membrane as depicted in Figure 42. (**Note:** aCGH and GE data demonstrated no apparent 11q22.2 locus change or *BIRC2* and *BIRC3* expression change in leukoplakia, hence we did not perform FISH and qRT-PCR in leukoplakia samples, but to understand the difference in localization, we have included leukoplakia in IHC analysis). Correlation analysis demonstrated a positive relation between cytoplasmic cIAP1 and cIAP2 with OSCC progression (normal to OSCC), while an inverse relation was observed between nuclear cIAP1 and membranous cIAP2 with OSCC progression (Figure 43). Additionally, higher membranous cIAP2 was observed in well differentiated tumors and staining decreased with increasing grade (Table 35). cIAP2 cytoplasmic expression was strongly associated with node metastasis. Polytomous Logistic regression with normal as the reference group showed a significant correlation of cIAP1 and cIAP2 cytoplasmic overexpression with risk of developing OSCC (Table 36).

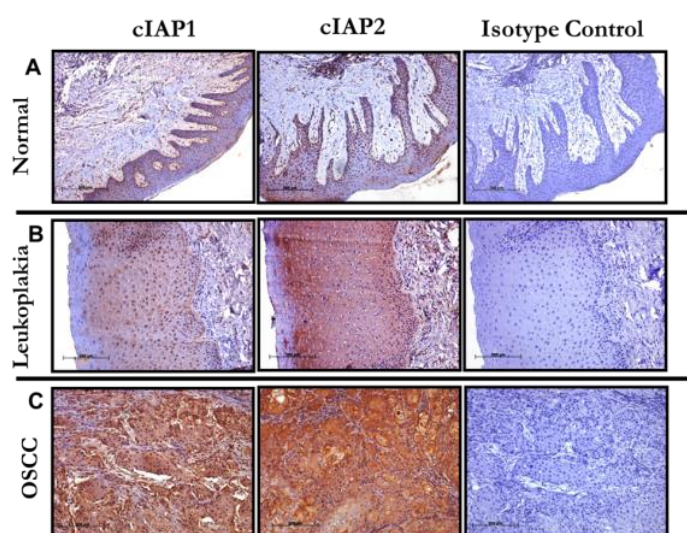


Figure 41: Immunohistochemical Analysis of cIAP1 and cIAP2 Expression in Normal Buccal Mucosa, Leukoplakia and Oral Cancers. Representative IHC staining on A) Normal buccal mucosa, B) Leukoplakia and C) OSCC along with the respective isotype control. Magnification 100X (Scale: 100 mm).

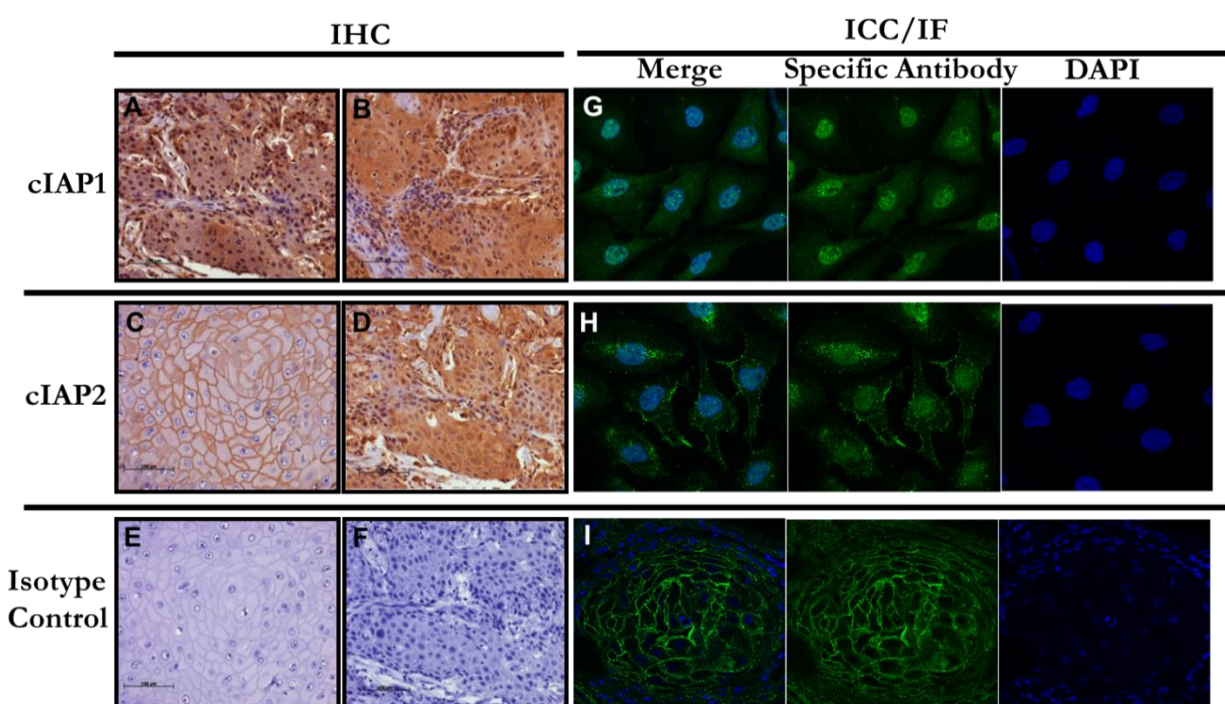


Figure 42: Differential Subcellular Localization of cIAP1 and cIAP2. Representative IHC staining demonstrated nuclear or cytoplasmic expression of cIAP1 (A, B); membranous or cytoplasmic expression of cIAP2 (C, D); the respective isotype control (E and F). Representative Immunofluorescent staining of cIAP2 membrane localization in OSCC tumour (I). The Immunocytochemistry staining in OSCC cell line, with nuclear/cytoplasmic localization of cIAP1 (G) and membranous and cytoplasmic staining of cIAP2 (H).

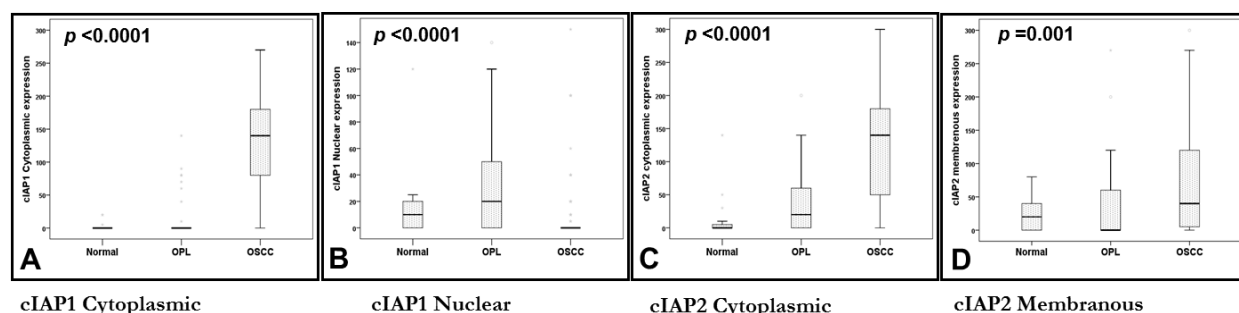


Figure 43: Correlation Between cIAP1 and cIAP2 Overexpression with Oral Cancer Development. A) and C) Significant increase in cytoplasmic cIAP1 and cIAP2 was observed in OSCC compared to normal and leukoplakia (OPL); B) Nuclear expression of cIAP1 was more prominent in normal and leukoplakia (OPL) than in OSCC; D) Membranous expression of cIAP2 in normal, leukoplakia and OSCC (y-axis represents the H-score for protein expression calculated as described in materials and methods).

Table 35: Correlation of cIAP1 and cIAP2 Expression with Clinicopathological Parameters.

	Clinicopathological parameters	
	Nodal metastasis	Grade
cIAP1 Cytoplasmic expression	NS	NS
cIAP1 Nuclear expression	NS	NS
cIAP2 Cytoplasmic expression	* $p=0.027$	NS
cIAP2 Membranous expression	** $p=0.049$	** $p=0.005$

NS: Not significant, * Spearman's Correlation, **: negative correlation

Table 36: The Effect of cIAP1 and cIAP2 Overexpression with OSCC Development.

	Normal (reference)	Leukoplakia		OSCC	
		p value	OR (95% CI)	p value	OR (95% CI)
cIAP1 Cytoplasmic expression	1	0.55	1.013 (0.971-1.056)	0.003	1.07 (1.02-1.13)
cIAP1 Nuclear expression	1	0.50	1.008 (0.985-1.033)	0.006	0.94 (0.89-0.98)
cIAP2 Cytoplasmic expression	1	0.52	1.007 (0.985-1.03)	0.04	1.02 (1.00-1.05)
cIAP2 Membranous expression	1	0.54	1.005 (0.989-1.022)	0.78	1.00 (0.98-1.03)

Polytymous logistic regression analysis was performed using normal as reference

4.8.3.2.4 Functional Validation of cIAP1 and cIAP2 in Gingivobuccal Oral Cancers

To understand the association of cIAP1 and cIAP2 overexpression in lymph node metastasis, we decided to perform shRNA mediated knockdown for cIAP1 and cIAP2 in buccal mucosa derived OSCC cell line (SCC29B). Expression and localization of cIAP1 and cIAP2 in SCC29B were confirmed using WB and ICC (Figure 44(A) and Figure 42) and all the shRNA sequences were confirmed by sequencing (sequencing data not shown in the thesis). SCC29B is a hard to transfect cell line (<5% transfection efficiency for most of the reagents - SAINT18, Xtreme gene and lipofectamine 3000), hence we could successfully generate stable line using pooled population (single clones did not survive) only for shcIAP2-1 and pLKO.1-EGFP-f-puro vector (Plasmid backbone-figure 44(B)). cIAP2 knockdown in the stable line at the mRNA and protein levels were validated using qRT-PCR (Figure 45(A)) and Western blotting (Figure 45 (B)) respectively.

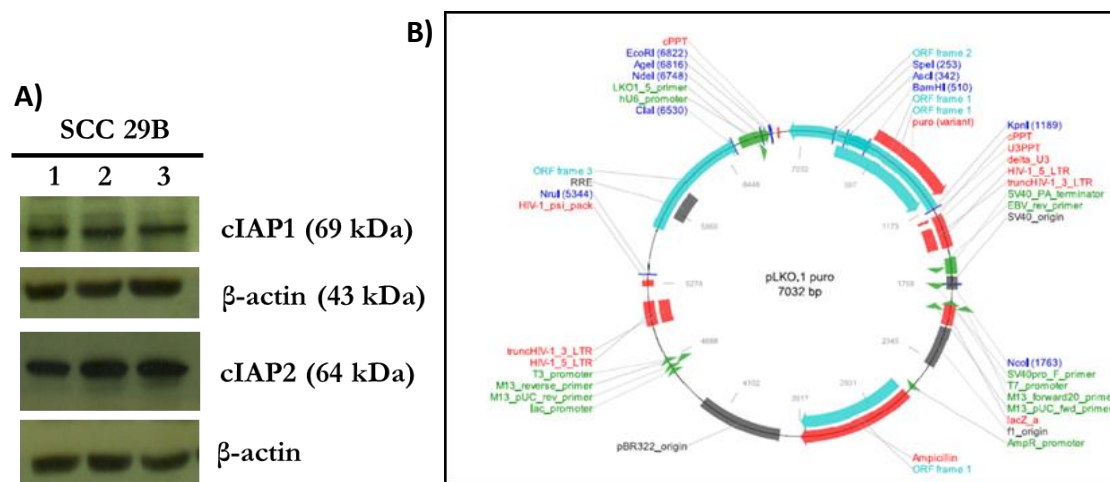


Figure 44: Expression of cIAP1 and cIAP2 in SCC29B and Plasmid Backbone. A) 75μg of whole cell extract (WCE) was resolved on 10% SDS PAGE gel followed by western blotting with antibodies specific to cIAP1, cIAP2 and β-actin (loading control). B) pLKO.1 puro plasmid backbone used in the study.

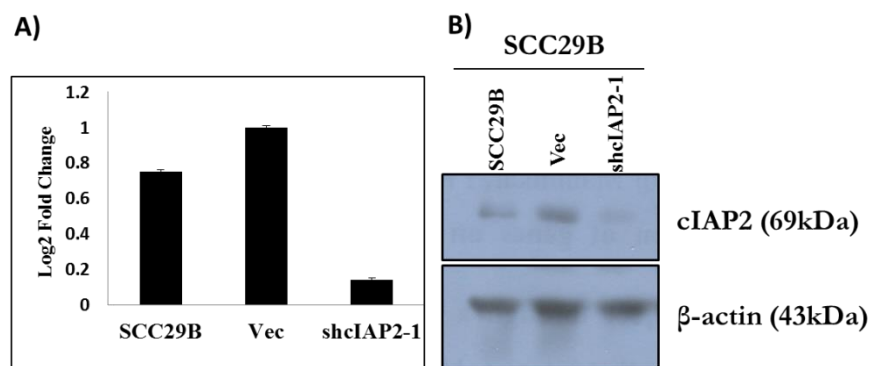


Figure 45: Affirmation of cIAP2 Knockdown by qRT-PCR and Western Blotting. **A)** qRT-PCRs were performed using oligonucleotides specific for *BIRC3* and *18S* rRNA in the SCC29B cells, the vector control (vec), the shcIAP2-1 containing SCC29B. All expressions were normalized to the levels of 18S. The standard errors are plotted and student's t- test was performed. **B)** 75μg of a whole cell extract (WCE) was resolved on 10% SDS PAGE gel followed by Western blotting with antibodies specific to cIAP2 and β-actin (loading control). Note that cIAP2 expression decreases in the knockdown clones both at mRNA and protein level.

To determine if cIAP2 knockdown in SCC29B can decrease cell migration, scratch wound healing assay was performed and it was observed that loss of cIAP2 indeed decreases cell migration (Figure 46(A and B)).

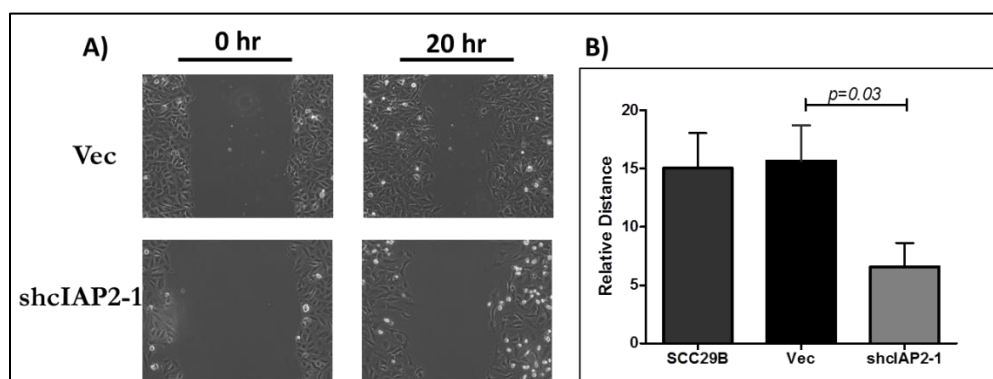


Figure 46: cIAP2 Loss Leads to a Decrease in Cell Migration. Scratch wound healing assays were performed on the SCC29B derived vector control (vec), cIAP2 knockdown (shcIAP2-1). Phase contrast images of wound healing at 0 hours (start) and 20 hours (end of the experiment)

have been shown (A). The mean and standard deviation of the relative distance migrated in 20 hours has been plotted (B). Note that migration is reduced in the cIAP2 knockdown cells. The p value was calculated using a student's t-test.

Matrigel invasion assays using Boyden chambers demonstrated decrease in invasion observed upon cIAP2 loss, without any change in the cell proliferation. (Figure 47). These results suggested that cIAP2 is required for the increased migration and invasion in SCC29B. However, further work is needed to validate cIAP2 findings in different OSCC cell line. Also, it will be interesting to see the effect of cIAP1 independently on invasion and metastasis and in combination with cIAP2.

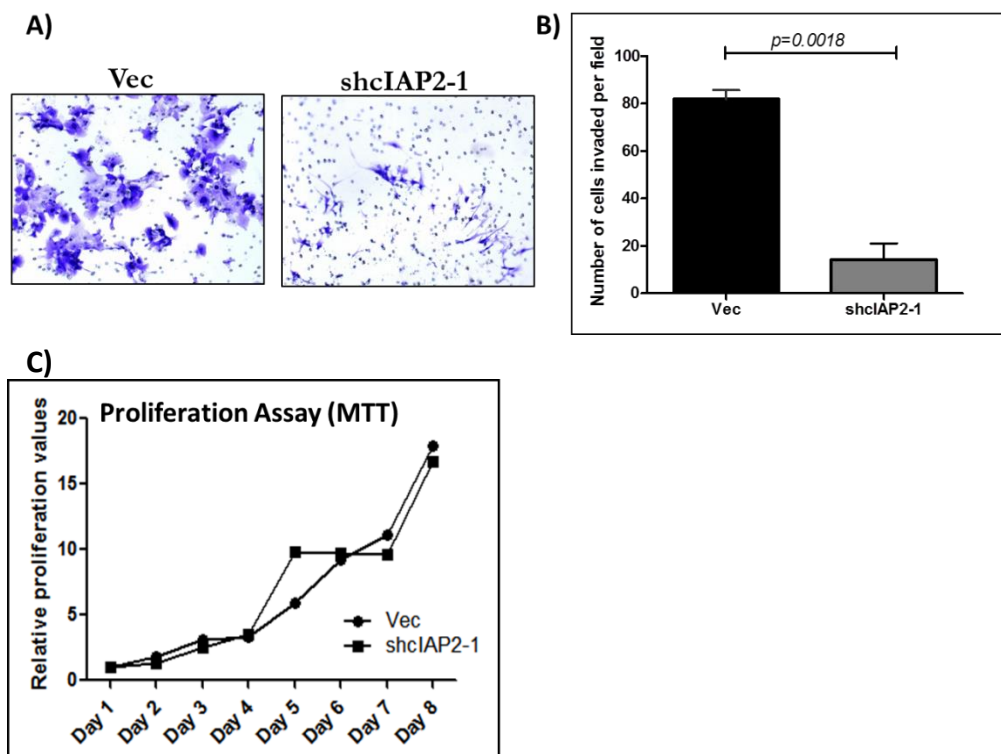


Figure 47. cIAP2 Loss Leads to a Decrease in Cell Invasion Without a Change in Cell Proliferation. Matrigel invasion assays were performed in Boyden's chambers for SCC29B derived vector control cells and cIAP2 knockdown. The number of cells observed in ten random

fields of the membrane for each clone was determined as described in materials and methods. Representative images of each clone are shown **A)** The mean and standard deviation of three independent experiments are plotted. **B)** MTT based cell proliferation assays were performed, and the relative difference in cell proliferation was determined for over a period of 8 days. **C)** The p value was calculated using a student's t-test. The loss of cIAP2 leads to a decrease in invasion as compared to the vector control.

5. Discussion

It is crucial to understand the early events associated with oral tumorigenesis and implicating the genomic profile to elucidate the pre invasive lesions that have a higher risk of malignant transformation. The results presented in the thesis provide a comprehensive and integrated genome wide analysis to identify important genomic alterations and the gene expression changes associated with progression, lymph node metastasis and patient survival in HPV-negative, gingivobuccal oral cancers. Our study revealed novel alterations/ gene expression changes, previously not described in leukoplakia and oral cancers as well as validated the previous reports.

5.1 A Conspectus of Genome-Wide Copy Number Alterations in Oral Cancers

DNA copy number changes are common in cancer and lead to altered expression and function of genes present within the affected regions of the genome. Identification of non random, recurring alterations that may be associated with progression or clinical outcome can offer a basis for better understanding of cancer development. Importantly, it may provide improved tools for clinical management of cancer, such as identification of new targets for prognosis and therapy [74, 82, 118-123].

We characterized copy number alterations in leukoplakia and early stage tobacco associated OSCC to identify the driver events and predict the markers related to progression and poor prognosis. The genomic panaroms of gingivobuccal cancer is predominated by gain of 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 loss of 3p21.1, 3p14.2, 4q21.3, 8p23.2, 8p11.22, 9p23, 9p21.3, 17p13.1. Study by Salahshourifar I. et al., systematically demonstrated the common genomic copy number alterations (CNAs) and their frequency in 12 studies that have been conducted on OSCCs using array comparative genomic hybridization (aCGH) [13]. The genomic alterations that were

reported >2 times among 12 aCGH studies on primary OSCC tumors are included in the histogram (Figure 48) [2, 9, 10, 78, 80, 81, 124-127].

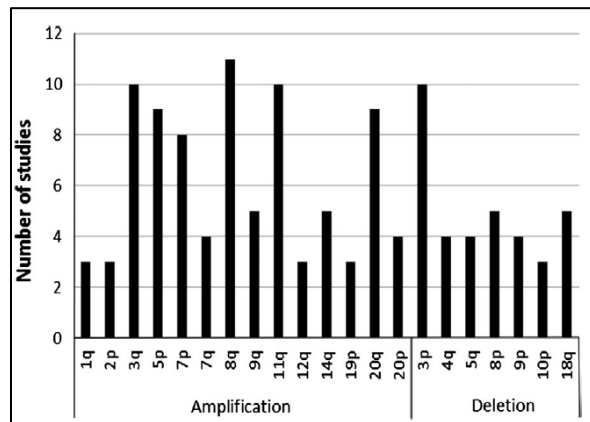


Figure 48: Recurrent Amplifications and Deletions Among 12 aCGH Studies on Primary OSCC Tumors (Source: Salahshourifa, I. et al., 2014).

5.1.1. Progression Associated CNA

Gain of 8q was the most common alteration associated with OSCC progression and could potentially harbour the genes driving oral tumorigenesis. 8q24.3 gain was the most frequent event observed in leukoplakia (58.3%), and OSCC (52.2%) while gain at 8q24.21 had the occurrence frequency of 40% in OSCC. *PTK2*, *LY6K*, *MYC* are the most reported candidate oncogenes on 8q locus [91, 128-130]. *MYC* is known to be one of the most frequently deregulated oncogenes which promotes selective gene expression amplification and increases cell growth and proliferation [131]. Additionally, *MYC* induces DNA damage at multiple levels, causes genomic instability, as well as changes the cell nuclear architecture and ultimately, promotes cell transformation. *MYC* dysregulation is one of the early events and is altered in more than 70% of cancers [132]. Copy number increase of 3q26.31 (31.3%) was strongly associated with disease advancement. *ECT2*, *PIK3CA* and *SOX2* located at 3q locus are reported to be involved in cell proliferation, metastasis and chemo/radioresistance in OSCC [133-138].

Additionally, we could identify *EIF5A2* a novel oncogene, which is being studied in ovarian, breast and esophageal cancer. However, till date, no study has shown its involvement in oral tumorigenesis [139-141]. Besides, the gain at 12q13.2 (27.8%) has been detected as a recurrent finding among OSCCs that subsequently altered the expression of genes located in the locus. Even though 12q gain was not reflected in leukoplakia, but a remarkable difference in *HOXC9* and *HOXC13* gene expression was evident in leukoplakia as well as in early stage OSCC.

Loss of multiple regions of 3p (3p26.3, 3p22.3, 3p21.1, 3p14.2, 3p11.1) and 8p (8p23.2, 8p23.1, 8p22, 8p11.22) were most frequently (> 40%) observed in OSCCs and can be considered as an important event associated with OSCC progression[83]. Interestingly, we observed more losses than locus gain across all the comparisons.

Our results were in concordance with a previous study from the lab using conventional CGH. This study estimated oncogenetic trees mixture models, describing the genetic development of oral cancers and observed 8q gain as an early event in cellular transformation of oral cancers. Following the occurrence of 8q (gain), the losses of 8p and 3p were predicted as subsequent early events during progression of OSCC [142].

5.1.2 Lymph Node Metastasis Associated CNA

In a current clinical scenario most of the patients undergo elective neck dissection even if not necessary. Predicting patients who may develop lymph node metastasis is a challenge, especially prior to surgical resectioning of tumor [56]. Hence, identifying the CNA unique to the node positive tumors and screening for these markers at the time of diagnosis can improve treatment decisions[143]. Pathare, S. et al., demonstrated that the process of oral carcinogenesis occur preferentially in certain orders that may define tumor subtypes, such as the node positive or node negative oral cancers [142]. Additionally, Bhattacharya, A. et al., identified that at least two

distinct routes of OSCC exist differing in genome instability and risk for cervical node metastasis [144]. We observed distinguishable CNAs in the lymph node metastasis positive primary tumors. This includes gain of 8p23.1, 8q24.21, 11q22.1 with 5- 37% frequency in lymph node positive tumors. 8q24.21 encoding *cMYC* had the highest frequency and it is postulated that *MYC* overexpression results in important tumor phenotypes such as metastasis, invasion and therapy resistance [145]. Few important chromosomal losses identified in lymph node metastasis positive primary tumors included loss at 3p26.3 (56%), 3p12.2 (56%), 8p23.2 (70%). Loss of 3p26.3 including the *CHL1* gene is reported to be an independent predictor of patients survival and lymph node metastasis in OSCC patients [125]. Focal loss of 8p23.2 encompassing the *CSMD1* (CUB and Sushi Multiple Domains 1) gene was the most frequent focal alteration in our study. Although loss of *CSMD1* has been reported in lung, head and neck, breast, and skin cancer specimens, its function as a tumor suppressor is less clear [146] and further studies are required to delineate its functional role in the pathogenesis of OSCC.

Our study demonstrates strong correlation between 11q22.2 alteration with concurrent overexpression of *BIRC2* and *BIRC3* to be associated with lymph node metastasis. *BIRC2* and *BIRC3* genes at 11q22.1-11q22.2 locus, which codes for cIAP1 and cIAP2 proteins, plays an important role in regulating proliferation and apoptosis. Few studies have also demonstrated that cIAP1 could be involved in lymph node metastasis in non-small-cell lung cancer, cervical cancer, tongue cancer and HNSCC [147-150]. However, its relevance in lymph node metastasis is not yet studied and till date association of cIAP2 in this context has not been explored. cIAP1 is a tumor necrosis factor receptor (TRAF2)-related proteins and localises either to the nucleus or cytosol. When in cytoplasm, it activates caspases in response to apoptotic signal while in the nucleus it regulates the cell cycle and cell proliferation [151]. Reports indicate that cIAP1

overexpression leads to genomic alterations due to defects in cell division and is a key regulator of cell proliferation and apoptosis [152]. cIAP2 is also involved in regulating apoptotic signal and knockout of cIAP1 has demonstrated to increase cIAP2 expression to compensate the cIAP1 loss, indicating that both cIAP1 and cIAP2 may be redundant [153]. cIAP2 is reported to localize to cytosol, nucleus and mitochondria and here in this study for the first time we report membranous localization of cIAP2 in OSCC. The mechanisms controlling the distribution of cIAPs into various subcellular location and the functional relevance in particular location require further clarification [152]. We further confirmed by BIRC3 knockdown that cIAP2 is associated with cell invasion and migration in buccal mucosa derived OSCC cell line SCC29B.

5.1.3 CNA Associated with Risk Prediction

The patients with OSCC generally develop recurrent disease, with chemo-/ radio- resistance that in turn impact patients overall survival. We analysed early and advanced stage OSCCs to delineate CNA that would identify subsets of tumors differing with respect to recurrence and survival. We observed that gain of 1p36.33, 11q13.3, 11q22.1 and 16p11.2 were strongly associated with disease specific survival of OSCC patients. 1p36.33 gain was significantly related to poor clinical outcome and this observation was confirmed by interphase FISH. Various lines of evidence indicate that genes located on chromosome 1 are closely related to the initiation and progression of several cancers, few of the genes include *c-Jun* (1p32-31), *p73* (1p36.3), *Casp-9* (1p36.21), and *N-ras* (1p13.2) [154]. Although we did not see the overexpression of *p73* in our study samples, but up-regulation of *MXRA8* (matrix remodeling associated 8/ Limitrin) was ascertained in gene expression analysis, and it strongly correlated with 1p36.33 gain. *MXRA8* is a tumor stroma-specific markers related to recovery of angiogenesis in capillaries by

enhancing perivascular ends in astrocytes. However the limited literature is available regarding its role in human malignancies [155, 156]. Another target on 1p36.33 is *DVL1* (Dishevelled-1), an important component of the Wnt signaling pathways shown to be involved in tumour growth, progression and metastasis in many tumours [157]. Although in the current study, we did not see a significant difference in *DVL1* expression in OSCC and leukoplakia, as confirmed by GE and qRT-PCR. It will be interesting to understand the role of 1p36.33 related targets that are associated with poor clinical outcome in OSCC.

Amplification at 11q13 is one of the most reported CNA in OSCCs [158, 159] and we did report the concurrent gain of 11q13 in the current study. The gain of the 11q13.3 region was validated using locus-specific FISH. Equivocal reports exist on the association between 11q13 alterations with patients clinical outcome [10, 89, 160, 161]. We observed the association of 11q13.3 gain with tumor stage and node metastasis, but no significant association was observed with patient survival. *CCND1* and *EMS1* amplification located at 11q13.3 are shown to be associated with increased cell proliferation in oral cancer [162, 163]. Another candidate on 11q13.3 is *ANO1*, we observed copy number dependent gene overexpression of *ANO1* in OSCC by data integration. A recent study by Shang, L. et al., demonstrated that *ANO1* expression was an independent prognostic biomarker for shorter survival of esophageal squamous cell carcinoma patients and further confirmed *ANO1* as a marker that can predict disease progression in individuals with esophageal precancerous lesions [164]. Further study is warranted to elucidate role of *ANO1* in lymph node metastasis and risk prediction in OSCC patients. Another predictor of poor survival in OSCC is the gain of 11q22.1–q22.2, the presence of this rare amplicon in 5% of OSCC cases and was successfully validated by Interphase FISH in this study as discussed in the previous section.

5.2 Genome-Wide Gene Expression and Integrative Analysis

The present study reports 849 and 1924 genes, differentially expressed in leukoplakia and OSCCs respectively. We validated the gene expression finding in an independent set of tumors, leukoplakia and unrelated normal oral tissues using qRT-PCR and IHC. For the first time we provide a comprehensive, integrative copy number and gene expression analysis in leukoplakia and gingivobuccal oral cancers, to delineate the important genomic alterations that may be driving the process of oral tumorigenesis. Further, we identified few novel genes with a potential role in OSCC progression.

5.2.1 Differential Expression of Genes in Oral Carcinogenesis

Of all the differentially expressed genes reported in this study, 61 genes from leukoplakia and 188 genes from OSCC have been previously known in oral and head and neck cancers, including *ECT2*, *INHBA*, *SERPINE1*, *GBP5*, *MMP10*, *MMP3*, *LY6K*, *SPP1*, *PDL1*, *PTHLH*, *KRT4*, *KRT76* and *MAL* [91, 133, 165-168]. In addition, we report deregulated novel genes such as *DERL3*, *EIF5A2*, *HOXC9*, *HOXC13*, *MFAP5*, *NELL2*, *CD274*, *DHRS2*, *FST* and *GPX3* in leukoplakia and/or OSCC, for which further investigation is warranted. The deregulated genes encoding for secretory proteins such as *MMPs*, *PTHLH*, *INHBA*, *LY6K* and *MFAP5* may be useful as a potential plasma/serum/saliva biomarkers to assess disease progression or therapy response in OSCC patients [169-171].

Deregulation of several keratins, namely *KRT4*, *KRT13*, *KRT19* and *KRT76* were observed in the current study, which are normally expressed in the oral cavity. Gene expression profiles of oral cancer obtained by other groups have also shown consistent downregulation of *KRT76* [21, 172].

We report, for the first time that differential expression of *KRT76* occurs in human and hamster oral precancerous and cancerous lesions. Also the loss of *KRT76* is sufficient to cause hyperplasia in the oral cavity of the mice. Our current findings suggest that the loss of *KRT76* may not be a sole molecular event leading to oral cancer development. However, the hyperplastic changes observed in *KRT76*-KO mice points to an indirect role of *KRT76* in regulating proliferation of the basal layers of buccal mucosa, similar to previous findings of *KRT10* loss [173]. We envision a number of possible ways in which *KRT76* loss contributes to cancer development. One is that it contributes to a barrier defect in the epithelium, which may render the tissue more susceptible to penetration by carcinogens [174]. Another possibility is that *KRT76* loss may lead to a disturbed inflammatory infiltrate; which is observed in human and mouse epidermis on the loss of structural proteins [174, 175]. Overall, our data implies that carcinogenesis is a multifactorial and multistep process. The potential role of *KRT76* could be as one of the contributing factor, but, it alone is not sufficient for cell transformation. However, its contribution in oral carcinogenesis cannot be ruled out.

Significant up-regulation of *INHBA*, *MFAP5* and *NELL2* was observed in leukoplakia and OSCC compared to unrelated normal buccal mucosa tissues. *INHBA* (Inhibin, Beta A) overexpression is associated with progression and metastasis in various cancers, including colorectal cancer, pancreatic cancer, urothelial carcinoma and HNSCC [176-179]. We validated this finding using qRT-PCR in an independent patient cohort and found a significant association between *INHBA* gene up-regulation with OSCC advancement and lymph node metastasis. Additionally, we validated genes previously unexplored in OSCC, including *MFAP5* and *NELL2*. *MFAP5* (microfibrillar-associated protein 5) is a stromal protein and overexpressed in leukoplakia and OSCC matrix as confirmed by qRT-PCR and IHC. Hence, we speculate its role in malignant

transformation as well as progression. Reports on ovarian cancer suggest that *MFAP5* promotes tumor cell survival and angiogenesis through $\alpha(V)\beta(3)$ integrin-mediated signalling [170, 180, 181]. Considering the fact that MFAP5 is a secretory protein, we hypothesize that it may be utilized as a potential serum biomarker to assess disease progression.

Moreover, contradictory reports exist regarding *NELL2* (Neural epidermal growth factor-like like 2) expression in cancers. In human renal cell carcinoma expression of *NELL2* is reported to inhibit cell migration, while in breast cancer and bladder cancer cells, it promotes cancer progression through the regulation of E2F1 factor [182, 183]. We observed overexpression of NELL2 protein by IHC in leukoplakia and OSCC. However, further study needs to be undertaken to understand association of NELL2 in OSCC development or maintenance.

5.2.2 Integrative Analysis to Identify Drivers in Oral Carcinogenesis

Integration of aCGH and GE microarray data revealed important chromosomal hotspots that results in the overexpression or underexpression of genes located within. This helps in distinguishing the driver genes that are essential for tumor development from the passenger events, accumulated during the transformation process. Integrative analysis in present study revealed that gene expression and copy number alterations were strongly correlated at 3q26.31 (*ECT2*, *EIF5A2*, *KLHL6*, *GPR160*) and 12q13.2 (*HOXC9*, *HOXC13*, *ERBB3*, *MUCL1*), in leukoplakia as well as in OSCC. 9p24.1 (*CD274*), 11q13.3 (*ANO1*), and 7p11.2 (*EGFR*) correlated in OSCCs only, indicating their role in disease advancement rather than at the pre-invasive stages. *ANO1* and *EGFR* are extensively studied in oral and HNSCCs [184-190]. *CD274* (*PD-L1*- Programmed death receptor ligand-1) is emerging as a novel therapeutic target in various cancers including OSCC [191, 192]. PD-L1 expression is found on antigen presenting

cells, as well as activated T-cells and when expressed on cancer cells, it is expected to regulate T cells negatively. It plays an important role in evading host immunity and tumor cells can escape host immune response [193]. Hence the blockade of tumor PD-L1 is expected to enhance the effects of immunotherapy [194].

3q26.31 is one of the most frequently altered locus in our sample set, with a syncronal overexpression of *ECT2* and *EIF5A2* genes located in this chromosomal region. *ECT2* (Epithelial cell transforming sequence 2) is a guanine nucleotide exchange factor (Rho family GTPase) which has been implicated in various human cancers (ovarian cancer, early-stage lung adenocarcinoma and gastric cancer) [195-197]. Iyoda, M. et al., for the first time demonstrated the role of *ECT2* in regulation of cell proliferation in oral cancers [133]. We validated *ECT2* overexpression in leukoplakia and OSCC by IHC in the current study, and observed a positive correlation between *ECT2* overexpression and disease progression. *EIF5A2* (eukaryotic initiation factor 5A2) was the next interesting target on 3q26.31 locus. For the first time we report overexpression of *EIF5A2* in leukoplakia and gingivobuccal oral cancers. A handful of literature is available indicating role of *EIF5A2* in metastasis and angiogenesis in human cancers, which includes esophageal squamous cell carcinoma, bladder cancer, and gastric cancer [140, 198-200]. Reports also confirmed *EIF5A2* (categorized as an oncogene) as a biomarker for the prognosis and may be a potential therapeutic target for many types of human tumors [197].

Gain at 12q13.2 is the next important hotspot for copy number dependent gene overexpression. We identified *HOXC9* and *HOXC13* as two interesting targets associated with disease progression in OSCC. Both these genes showed more than 2 fold expression change in leukoplakia and OSCC when compared to unrelated normals in microarray analysis. This observation was consistent when confirmed by IHC and/or qRT-PCR in an independent

validation cohort. HOX genes encode a family of transcriptional regulators that are involved in pattern formation and organogenesis during embryo development and in adult tissue they have a cell type specific expression [201]. However, the correlation of HOX gene expression patterns with cancer progression is now beginning to be understood, few HOX genes demonstrate the tumor suppressive role, while others have oncogenic effects [202-204]. HOX genes are also thought to be involved in maintenance and regulation of cancer stem cells [205]. *HOXC9* has been linked with cell cycle exit and promote cell invasion in breast cancer and neuroblastoma [201, 206-208]. *HOXC13* play an important role in maintaining skin homeostasis as well as it regulates the transcription of cytokeratins genes via highly conserved motif in the promoter of these genes [209, 210]. Study by Kasiri, S. et al., demonstrates that *HOXC13* is a key player in tumor cell growth and viability [211, 212].

Interestingly, we observed an opposite direction expression change for few genes with respect to the change in the locus (i.e reverse direction change of gene expression with respect to the locus). In the current study, we observed amplification of chromosomal locus (22q11.23). *DERL3* (Derlin-3) gene is located on this amplicon, however, we observed significant downregulation of *DERL3* both in the microarray analysis and in the validation study. *DERL3* accelerates degradation of the misfolded glycoproteins and imparts its tumour suppressor role in cancer development [213, 214]. Here we report downregulation of *DERL3* for the first time in leukoplakia and oral cancers. We hypothesized, that the inverse expression change could be as a result of epigenetic regulation of either the gene promoter or via gene specific miRNAs. Further study is needed to confirm the significance of the *DERL3* in oral tumorigenesis and to understand its gene regulation.

5.2.3 Pathways Deregulated in OSCC

The global gene expression profiling analysis has revealed various genes deregulated in leukoplakia and OSCC when compared to the normal buccal mucosa tissues. Current analysis provides the glimpse of the genes altered and its association with various cellular process, including cell signaling, gene transcription, cell cycle regulation, oncogenesis, tumor suppression, differentiation, motility and invasion in leukoplakia and OSCC. In our analysis, we observed that almost all of the pathways altered in OSCC were reflected in leukoplakia, except for the fold change and number of genes. The majority of the pathways deregulated were cell surface receptor signaling pathway (cytokine-mediated signaling pathway, integrin signaling pathway, Wnt signaling pathway, EGFR receptor pathway, Ras signaling pathways) or related to angiogenesis, stem cell differentiation etc.

A mitogenic signaling pathway that includes EGFR, TGF β , HRAS, PI3K and p38 MAPK pathway are altered in most of the human malignancies. EGFR is one of the most extensively studied receptor tyrosine kinase and overexpressed in over 95% HNSCC; with EGFR-PI3K-Akt-mTOR pathway as a most frequently altered axis that regulates cell growth, apoptosis, survival and differentiation [215, 216]. Although we observed overexpression of *EGFR* in advanced stage OSCC, but various studies do report its association with pre malignant conditions[188, 190, 217].

The Wnt- β -catenin pathway is most frequently altered in various cancers [218], however, there is little knowledge of its contributions and signaling mechanism in OSCC. Several components of the Wnt pathway are altered in oral cancers and our analysis also showed altered expression of various Wnt receptors, Frizzleds, Dishevelled, Homeobox protein engrailed-1, and Cadherins. Wnt signaling is mostly involved in organ development, maintenance of adult stem/progenitor

cell, and tumor development thus can be looked upon as a potential therapeutic target in OSCC [216].

As the tumor cells grow, the requirement of nutrient and oxygen need to be fulfilled, which is usually met by the formation of new blood vessels. Tumor cells modulate the angiogenic pathways depending on the cellular demand. We observed the up-regulation of angiogenic factors such as *FGF2* and *MFAP5*, in both leukoplakia and gingivobuccal OSCC tissues while *VEGFA*, *PTK2*, *PAK1*, *PREPL* and *SERPINE1* were enriched with disease advancement. Deregulation of angiogenic factors in leukoplakia suggests that the angiogenic process commences in the pre-invasive stages and may serve as a surrogate marker for tumour development, as the majority of the lesions do not progress to malignant disease [219, 220].

Additionally, pathways related to hypoxia, cell adhesion and migration play a crucial role in cancer progression, metastasis and therapy resistance [221]. In our study, we find the overexpression of putative hypoxia responsive genes like *SPPI* (*OPN*) and *CA9* which are reported as markers of poor clinical outcome in solid tumors including oral cancers [222-224]. Hypoxic regions of the tumor are speculated to harbor cancer stem cells. Okamoto, O. K. et al., reported possible up regulation of *HOXC9* in astrocytomas enriched in cancer stem cell like population [225]. We find overexpression of *HOXC9* in leukoplakia and oral cancers indicating the presence of the stem cell population in oral cancers although, the functional importance and clinical relevance of this overexpression needs to be investigated. Among the up-regulated genes observed in our study, the genes mediating apoptosis and anti-apoptotic function (*BAX*, *BAK*, *BIRC5*, *FADD*, *INHBA*) both were enriched, implying that both pathways operate in parallel, but the anti-apoptotic genes may override the effect of the apoptotic regulators or they may not be able to signal the downstream effectors [226].

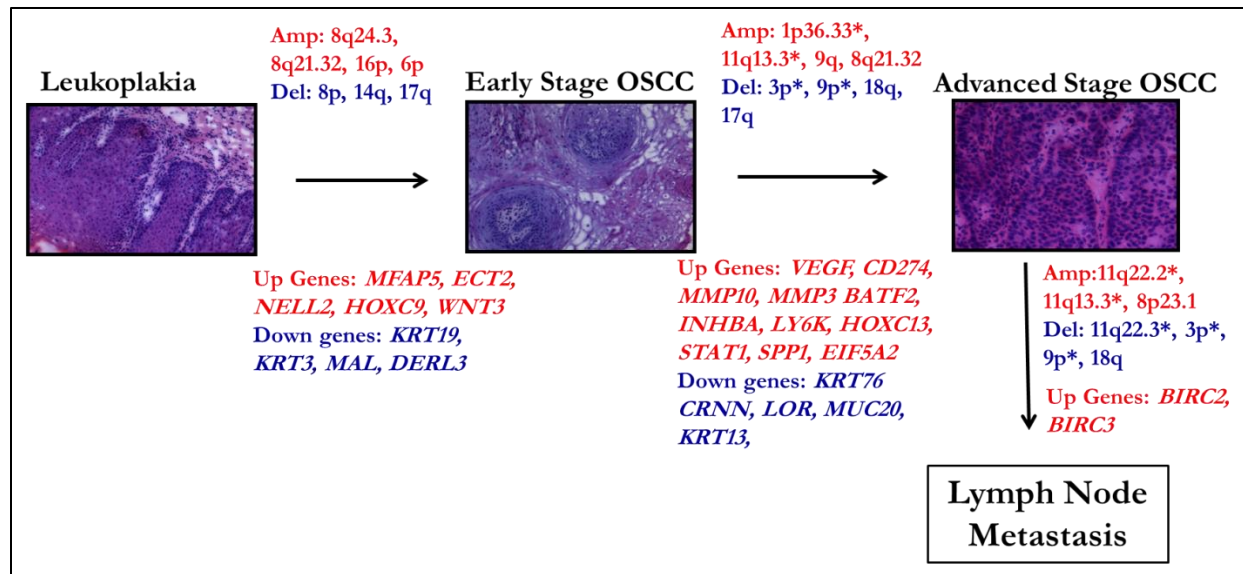


Figure 49: Summary of the Genomic and Transcriptomic Signatures Associated with Gingivobuccal Cancer Progression from Pre-invasive Lesions. Alterations and expression changes represented in leukoplakia were common to either early and/or advanced stage OSCC, however due to limited space they are listed on the first occurrence. Amp: Amplification/Gain (Red); Del: Deletion/Loss (Blue); Up: genes upregulated (Red), Down: genes downregulated (Blue). *Alterations associated with Disease Specific Survival in OSCC patients.

Overall, our study identifies oral cancer progression related copy number alterations and differentially expressed genes. Alterations that were in common between leukoplakia and OSCC can be considered as important early events that are essential for initial cell transformation and progression. The CNA data can detect only the most frequently occurring events, but with the integration of gene expression data, we could identify a novel and a tractable number of driver genes which may be biologically relevant to oral cancer pathogenesis. This study helps us to delineate the sequential genetic model associated with oral carcinogenesis (Figure 49).

6. Summary and Conclusions

Summary and Conclusions

To conclude, multiple recurrent chromosomal regions with frequent copy number alterations were detected in HPV negative leukoplakia and OSCC cases in our aCGH study. Using high-resolution genome-wide copy number analysis and gene expression profiling, we were able to identify the common CNA and several deregulated genes associated with transformation of pre-invasive lesions to invasive OSCC.

The Key observations of the study are enlisted below:

- The gains and losses at loci 6p21.32, 8q24.3, 16p11.2 and 1q44, 8p23.1, 8p11.22, 17q21.31 are identified in this study as CNA associated with OSCC progression. We validated 8q24.3 gain and confirmed it to be related to progression in an independent cohort.
- Our study revealed genomic lesions on chromosome arms 11q22.1 and 1p36.33 associated with a high risk of relapse and decreased patients survival. These chromosomal alterations are better predictors of survival than common clinicopathological markers. Additionally, we demonstrated that *BIRC2* (cIAP1) and *BIRC3* (cIAP2) genes on 11q22.2 could be associated with lymph node metastasis in OSCC.
- The genome-wide gene expression data from leukoplakia and OSCC, identified several known (*ECT2*, *INHBA*, *LY6K* and *MAL*) as well as novel genes (*EIF5A2*, *HOXC9*, *HOXC13*, *NELL2*, *DERL3* and *MFAP5*) that may prove to be important in the oral cancer progression. Validation of secretory factors confirmed them to be a novel biomarker for pre-invasive lesions and oral cancers, although its clinical relevance needs further investigation. For the first time we demonstrated that *KRT76* loss is one of the early event, but not essential for cell transformation.
- The impact of copy number alterations on gene expression is significant, we identified few novel targets with copy number dependent expression change, including gain of 3q26.31 (*ECT2*,

EIF5A2) and 12q13.2 (*HOXC9*, *HOXC13*) along with concomitant up-regulation of genes in the altered locus. We also report an opposite direction expression change for few regions including 22q11.23 (*DERL3*), in leukoplakia and OSCC.

- The candidate copy number dependent and independent genes were validated using I-FISH, real-time qRT-PCR and IHC. All of the validated genes showed statistically significant change in their expression in leukoplakia and oral cancers compared with normal buccal mucosa tissue.

OSCC progression and metastasis to lymph node is the most important predictor of patient survival. Identifying the risk of cervical lymph node metastasis would greatly help in patient stratification, either by avoiding the potential morbidity of over treatment or preventing further progression of disease. This is the first report that provides a comprehensive genomic and transcriptomic signature associated with progression of low risk lesions to high risk invasive OSCC and markers that could predict lymph node metastasis and patient's clinical outcome.

7. Appendix

7.1 Tables

7.1.1. Clinicopathological Characteristics of Leukoplakia Patients (aCGH and Gene Expression Study)

OC code	Age	Gender	Site of lesion	Grade	Chewing status	Smoking status	Drinking status	Recurrence of lesion	Status at last follow up	Tumor recurrence status
OC1555	40	Female	BM	Hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC2199	41	Female	BM	Hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC1566	22	Male	BM	Hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC1565	40	Male	BM	Hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC2204	44	Male	BM	Hyperplasia	Current	Current	Never	Recurrence	Recurrence of lesion	Recurrence
OC1560	45	Male	BM	Hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC2061	60	Male	BM	Mild Dysplasia	No Info	No Info	No Info	No Recurrence	Transformed to SCC	No Recurrence
OC2067	53	Male	BM	Mild Dysplasia	Current	Never	Never	Recurrence	Transformed to SCC	Recurrence
OC2209	38	Male	BM	Hyperplasia	current	Never	Current	No Recurrence	NED	No Recurrence
OC2293	33	Male	BM	Hyperplasia	current	Current	current	No Recurrence	NED	No Recurrence
OC2299	50	Male	BM	Hyperplasia	No Info	No Info	No Info	No Recurrence	NED	No Recurrence
OC2300	41	Male	BM	Mild Dysplasia	Current	Current	Never	No Recurrence	Transformed to SCC	No Recurrence
OC2303	42	Male	BM	hyperplasia	Current	Current	Never	No Recurrence	NED	No Recurrence
OC2202	44	Male	BM	hyperplasia	No Info	No Info	No Info	Recurrence	Recurrence of lesion	Recurrence
OC1669	20	Male	BM	hyperplasia	Current	Current	Current	No Recurrence	NED	No Recurrence
OC1348	26	Male	BM	hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC2203	32	Male	BM	hyperplasia	Current	Current	Never	Recurrence	Recurrence of lesion	Recurrence

OC2201	76	Male	BM	hyperplasia	Never	Current	Never	No Recurrence	NED	No Recurrence
OC1435	50	Male	BM	hyperplasia	Never	Current	Never	No Recurrence	NED	No Recurrence
OC1358	40	Male	BM	hyperplasia	Current	Current	Never	No Recurrence	NED	No Recurrence
OC1445	57	Male	BM	hyperplasia	Never	Current	Never	No Recurrence	NED	No Recurrence
OC1377	65	Female	BM	hyperplasia	Current	Never	Never	No Recurrence	NED	No Recurrence
OC1884	55	Male	BM	hyperplasia	No Info	No Info	No Info	Recurrence	Transformed to SCC	Recurrence
OC1952	42	Male	BM	hyperplasia	current	Never	Never	No Recurrence	NED	No Recurrence

Cases highlighted in **bold** were used for both aCGH and gene expression study, and for rest others cases only aCGH was done;

Site: BM= Buccal mucosa

Follow up: NED: Alive and No evidence of Disease, Transformed to SCC: OSCC development was observed in those patients

7.1.2. Clinicopathological Characteristics of OSCC Patients (aCGH and Gene expression study)

OC code	Age	Sex	Site	T	N	Stage	Grade	Treatment	C*	S*	D*	Recurrence	Follow-up	Survival
OC 1313	52	M	GBC	T2	N0	II	P	S	NA	NA	NA	No	NED	111
OC 1587	47	M	GBC	T2	N0	II	M	S + RT	C	N	N	Yes	AWD	31
OC 1698	61	M	GBC	T2	N0	II	W	S + RT	C	N	N	No	NED	6
OC 1791	43	M	GBC	T1	N0	I	P	S + RT	C	N	N	No	NED	105
OC 1843	53	M	GBC	T1	N0	I	M	S	NA	NA	NA	No	NED	60
OC 1844	38	M	GBC	T2	N0	II	P	S + RT	NA	NA	NA	No	NED	62
OC 1845	48	M	GBC	T2	N0	II	M	S+RT+CT	C	N	N	No	NED	8
OC 1846	44	M	GBC	T2	N0	II	M	S + RT	NA	NA	NA	No	NED	46
OC 1891	75	F	GBC	T2	N0	II	M	S	C	N	C	Yes	DOD	24

OC 1896	37	M	GBC	T2	N0	II	W	S		NA	NA	NA	No	NED	20
OC 1910	40	F	GBC	T2	N0	II	M	S + RT		C	N	N	No	NED	32
OC 1914	69	F	GBC	T2	N0	II	P	S + RT		C	N	N	No	NED	48
OC 1932	58	M	GBC	T2	N0	II	W	S		C	N	N	No	NED	22
OC 1947	50	F	GBC	T2	N0	II	W	S + RT		NA	NA	NA	No	NED	53
OC 1965	70	M	GBC	T2	N0	II	M	S + RT		C	N	N	No	NED	2
OC 1983	66	M	GBC	T2	N0	II	P	S		C	N	N	Yes	DOD	9
OC 1984	64	M	GBC	T2	N0	II	W	S		C	N	N	No	NED	61
OC 1986	47	M	GBC	T2	N0	II	P	S		NA	NA	NA	No	NA	0
OC 2048	62	M	GBC	T2	N0	II	P	S + RT		C	N	N	Yes	AWD	5
OC 2052	62	M	GBC	T2	N0	II	M	S		C	N	N	No	NED	45
OC 2122	58	F	GBC	T2	N0	II	M	S + RT		C	N	N	No	NED	25
OC 2109	32	M	GBC	T2	N+	III	M	S		C	N	C	No	NED	44
OC 2184	38	M	GBC	T1	N+	III	M	S		C	N	N	No	NED	46
OC 2185	45	F	GBC	T2	N+	III	M	S		NA	NA	NA	No	DOD	42
OC 1629	49	F	GBC	T2	N+	IV	M	S + RT		C	N	N	Yes	DOD	13
OC 1646	50	M	GBC	T2	N+	III	M	S + RT		C	N	N	Yes	AWD	6
OC 1716	37	M	GBC	T2	N+	III	M	S + RT		C	N	C	No	NED	55
OC 1760	43	M	GBC	T2	N0	II	M	S+RT+CT		NA	NA	NA	No	NED	63
OC 1823	33	M	GBC	T2	N+	III	P	S + RT		C	C	C	Yes	AWD	17
OC 1869	34	M	GBC	T2	N+	III	M	S + RT		C	N	N	No	NED	32
OC 2102	37	M	GBC	T2	N0	II	M	S		C	N	N	Yes	AWD	49
OC 2112	47	M	GBC	T1	N0	I	M	S		NA	NA	NA	No	NED	46
OC 2118	39	M	GBC	T2	N0	II	P	S		C	N	N	No	NED	20
OC 2176	42	M	GBC	T2	N0	II	M	S		C	N	N	No	NED	48

OC 2192	62	M	GBC	T2	N0	II	M	S + RT	C	N	C	No	NED	29
OC 2213	50	M	GBC	T1	N+	IV	P	S	C	N	N	No	NED	30
OC 2214	70	F	GBC	T2	N+	IV	M	S	C	N	N	Yes	DOD	12
OC 2215	50	F	GBC	T2	N0	II	W	S	C	N	N	No	NED	1

Cases highlighted in **bold** were used for both aCGH and gene expression study, and for rest other cases only aCGH was done

Site: GBC= Gingivo buccal complex (Lower gingivo buccal sulcus, Buccal mucosa, Lower alveolus & Retromolar trigone)

Sex: M: Male, F: Female

Node (N): N0: lymph node metastasis negative; N+ : lymph node metastasis positive.

Grade: W- Well differentiated, M- Moderately differentiated, P- Poorly differentiated.

Treatment Modality: Surgery=S, Surgery + Radiotherapy=S+RT, Surgery + Chemotherapy = S+CT, Surgery + Radiotherapy + Chemotherapy = S+RT+CT

Follow-up (Status at last follow up): Alive with disease=AWD, Alive and No Evidence of Disease=NED, Dead of disease=DOD, Dead of other cause= DOC, Lost to follow-up= LFU

For Chewing (C*), smoking (S*) and Drinking (D*): C-Current; N: Never

7.1.3. Clinicopathological Characteristics of OSCC Patients from Previous Study (aCGH and Gene Expression Study- Ambatipudi, S. et al.,)

OC Code	Age	Sex	Site	T	N	Grade	Stage	Treatment	C*	S*	D*	Recurrence	Follow-up	Survival
OC 1644	55	F	GBC	T4	N0	P	IV	S+CT+RT	C	N	N	Yes	AWD	13
OC 1651	46	M	GBC	T2	N0	P	II	S+CT+RT	C	N	N	Yes	DOD	13
OC 1647	48	M	GBC	T4	N0	M	IV	S+RT	C	C	N	Yes	DOD	18
OC 1653	53	M	GBC	T4	N+	M	IV	S+CT+RT	C	N	N	Yes	DOD	10
OC 1656	56	M	GBC	T4	N0	M	IV	S+RT	C	N	C	No	NED	39
OC 1645	48	M	GBC	T3	N+	M	IV	S+CT+RT	C	N	N	Yes	DOD	16
OC 1649	54	F	GBC	T4	N+	P	IV	S+CT+RT	C	N	N	Yes	DOD	8
OC 1652	45	F	GBC	T4	N0	M	IV	S+RT	C	C	N	Yes	AWD	46
OC	39	M	GBC	T2	N0	M	II	S+CT+RT	C	N	C	Yes	DOD	13

1658															
OC 0542	42	F	GBC	T2	N+	P	IV	S+RT	C	N	N	No	NED	50	
OC 1323	56	F	GBC	T4	N0	W	IV	S+RT	C	N	N	Yes	DOD	8	
OC 0547	42	M	GBC	T2	N+	M	III	S+RT	C	N	N	Yes	DOD	51	
OC 1464	59	M	GBC	T4	N0	M	IV	S+CT+RT	C	C	C	Yes	DOD	10	
OC 1462	49	M	GBC	T4	N0	M	IV	S+RT	C	C	C	No	NED	34	
OC 1440	67	M	GBC	T4	N0	M	IV	S+RT	C	N	N	No	NED	32	
OC 1718	78	F	GBC	T4	N+	P	III	S+RT	C	N	N	Yes	DOD	15	
OC 1487	42	M	GBC	T4	N0	M	IV	S+RT	C	N	N	No	NED	28	
OC 1294	45	M	GBC	T3	N0	P	III	S+RT	C	N	N	No	NED	19	
OC 1719	80	M	GBC	T4	N+	M	IV	S+RT	C	C	N	LFU	LFU	5	
OC 1721	56	M	GBC	T4	N0	P	IV	S+RT	N	C	N	No	NED	33	
OC 1722	73	M	GBC	T4	N+	M	IV	S+RT	C	N	N	No	DOC	27	
OC 1726	64	M	GBC	T4	N+	P	IV	S	C	N	N	No	NED	8	
OC 1750	68	M	GBC	T4	N+	M	IV	S+RT	C	N	N	Yes	DOD	4	
OC 1371	40	M	GBC	T4	N0	M	IV	S+CT+RT	C	N	N	Yes	DOD	6	
OC 1426	44	M	GBC	T4	N+	M	IV	S+CT+RT	C	N	N	LFU	LFU	0	
OC 1405	72	M	GBC	T4	N0	M	IV	S+RT	C	N	N	Yes	AWD	44	
OC 1662	60	M	GBC	T4	N+	P	IV	S+RT	C	N	N	Yes	DOD	16	
OC 1663	66	M	GBC	T2	N0	M	II	S+RT	C	N	N	No	NED	90	
OC 1664	31	M	GBC	T4	N+	P	IV	S+RT	C	N	N	No	NED	39	
OC 1665	49	M	GBC	T4	N+	M	IV	S+RT	N	C	N	No	NED	46	
OC 1496	37	M	GBC	T4	N0	M	IV	S+RT	C	N	N	Yes	AWD	49	
OC 1507	77	M	GBC	T4	N0	M	IV	S+RT	C	C	N	No	NED	91	
OC 1508	40	M	GBC	T4	N0	M	IV	S+CT+RT	C	N	N	No	NED	84	
OC 1501	50	F	GBC	T4	N0	M	IV	S	C	N	N	LFU	LFU	2	
OC 1418	54	M	GBC	T4	N+	M	IV	S+RT	C	N	N	Yes	AWD	71	

OC 1367	52	M	GBC	T4	N0	P	IV	S+RT	C	N	N	No	NED	86
OC 1331	59	F	GBC	T4	N+	M	IV	S	C	N	N	No	NED	43
OC 1666	43	M	GBC	T2	N+	M	III	S+RT	C	C	N	No	NED	113
OC 1667	54	F	GBC	T3	N0	M	III	S+RT	C	N	N	LFU	LFU	1
OC 105	39	M	GBC	T1	N0	P	I	S	C	N	N	Yes	DOD	16
OC 344	44	M	GBC	T4	N+	P	IV	S+RT	C	C	C	Yes	DOD	9
OC 672	64	F	GBC	T4	N0	P	IV	S+RT	C	N	N	No	NED	139
OC 739	42	M	GBC	T4	N+	P	IV	S+RT	C	N	N	Yes	DOD	8
OC 860	70	F	GBC	T2	N+	P	IV	S	C	N	N	Yes	DOD	33
OC 498	52	M	GBC	T4	N+	M	IV	S+RT	C	N	N	No	NED	83
OC 561	50	F	GBC	T3	N+	P	IV	S+RT	C	N	N	Yes	DOD	5
OC 588	66	M	GBC	T4	N0	W	IV	S+RT	C	N	N	No	NED	6
OC 811	60	M	GBC	T4	N+	P	IV	S+RT	C	N	N	Yes	AWD	13
OC 996	54	M	GBC	T1	N+	M	III	S+RT	C	N	N	Yes	AWD	35
OC 849	60	M	GBC	T4	N+	M	IV	S	C	N	N	LFU	LFU	0
OC 1025	64	M	GBC	T4	N0	M	IV	S+RT	C	N	N	Yes	AWD	96
OC 1004	56	M	GBC	T4	N+	M	IV	S+RT	C	C	N	No	NED	119
OC 1001	53	M	GBC	T4	N0	M	IV	S+RT	C	N	N	No	NED	65

Site: GBC= Gingivo buccal complex (Lower gingivo buccal sulcus, Buccal mucosa, Lower alveolus & Retromolar trigone).

Sex: M: Male, F: Female

Node (N): N0: lymph node metastasis negative; N+ : lymph node metastasis positive.

Grade: W- Well differentiated, M- Moderately differentiated, P- Poorly differentiated.

Treatment Modality: Surgery=S, Surgery + Radiotherapy=S+RT, Surgery + Chemotherapy = S+CT, Surgery + Radiotherapy + Chemotherapy = S+RT+CT

Follow-up (Status at last follow up): Alive with disease=AWD, Alive and No Evidence of Disease=NED, Dead of disease=DOD, Dead of other cause= DOC, Lost to follow-up= LFU

For Chewing (C*), smoking (S*) and Drinking (D*): C-Current; N: Never

7.2 Recipes for Common Reagents

10X PBS

80 g of sodium chloride, 2 g potassium chloride, 14.4 g disodium hydrogen phosphate, 2.4 g potassium dihydrogen phosphate were dissolved in 600 ml of Milli-Q®. pH of the buffer was adjusted to 7.4 with 1N NaOH and final volume was made to 1000 ml with Milli-Q®. The solution was autoclaved and stored at room temperature.

10X Tris-Borate EDTA (TBE)

108 g (890mM) of Tris base, 55 g (890mM) of boric acid and 40 ml (20mM) of EDTA (pH 8.0) were dissolved in 500 ml of Milli-Q® water and pH was set to 8.0 using 1N NaOH. The final volume was made to 1000 ml.

DNA Loading Dye

0.025 g of bromophenol blue and xylene cyanol were dissolved in 5 ml of Milli-Q® along with 1.5 g of ficoll to obtain the DNA loading dye. Make up to 10 ml.

EDTA (Ethylenediaminetetraacetic acid), 0.5 M (pH 8.0)

186.1 g of Na₂EDTA.2H₂O was dissolved in 700 ml of Milli-Q®. pH was adjusted to 8.0 with 10 M NaOH (~50 ml) and the final volume was made to 1000 ml with Milli-Q®.

Ethidium Bromide (10 mg/ml)

0.2 g of EtBr was dissolved in 20 ml Milli-Q®. It was mixed well and stored at 4°C in the dark.

4% Buffered Formalin

40% Formaldehyde 100ml

NaH₂PO₄.2H₂O 4.5g

Na₂HPO₄ 6.5g

the final volume was made to 1000 ml with Milli-Q®.

FISH Experiment Reagents

Cryosectioned Tissue Fixative: Methanol: Acetic Acid in 3:1 ratio

3M Sodium Acetate

408 g of sodium acetate.3H₂O was dissolved in 800 ml of Milli-Q®. pH was adjusted to 5.2 with 3 M acetic acid and final volume was made to 1000 ml.

20X Saline Sodium Citrate (SSC)

175.32 g of sodium chloride (3M) along with 88.23 g of sodium citrate (0.3M) was dissolved in 600 ml of Milli-Q®. pH of the resulting solution was adjusted to 7 and the volume was made up to 1000 ml. The solution was autoclaved and stored at room temperature.

Chloramphenicol Stock (10mg/ml)

10 mg of chloramphenicol was dissolved in 1 ml of methanol and stored at -20°C in small aliquots away from sunlight.

Luria-Bertini Broth

10 g Tryptone, 5 g yeast extract, and 10 g NaCl were dissolved in 950 ml Milli-Q®. The pH of the solution was adjusted to 7.0 with 1N NaOH and the volume was made up to 1000 ml. The broth was autoclaved on liquid cycle for 20 min at 15 psi it was allowed to cool to 55°C and appropriate antibiotic was added before culturing.

Pepsin Stock Solution

25 mg/ml of pepsin was dissolved in Milli-Q®. Aliquots of 50µl were made and stored at -20°C.

Pepsin Working Solution

Frozen tissue: 30- 40µl of the pepsin stock solution was added to 50 ml pre-warmed diluted hydrochloric acid (49.5 ml of sterile Milli-Q®+ 0.5 ml 1N HCl) for tissue digestion.

Paraffin tissue: 30µl of the pepsin stock solution was added to 970µl pre-warmed diluted hydrochloric acid (concentration same as above) for tissue digestion.

RNase A (Stock solution)

RNase A (20 mg/ml) was dissolved in sterile Milli-Q® and boiled for 15 min. Aliquots were prepared after cooling the stock solution and were stored at -20°C.

DNase I Stock (1 mg/ ml)

DNase I	10 mg
NaCl 1M	1.5 ml
Glycerol	5 ml
Sterile Milli-Q®	Make upto 10 ml

Aliquot and stored at -20°C

For dilution: 1:500 or 1:250 in buffer

1 ml Dilution Buffer:

NaCl 1M	150 µl
Glycerol	500 µl
Sterile Milli-Q®	350 µl

dNTP Mix (for reaction containing labelled dUTP)

Reagents	Volume	Final concentration
1M Tris.Cl (pH=7.5)	250µl	500mM
1M MgCl ₂	25µl	50mM
10µM dNTPs (dATP,dGTP,dCTP)	30µl (10 µl each)	0.2mM
10µM dTTP	2µl	0.04mM
Sterile Milli-Q®	193µl	
Total volume	500 µl	

Denaturation Buffer pH 7.0 (To be made fresh)

Reagents	Volume	Final concentration
Formamide	35ml	70%
20X SSC	5ml	2X
Sterile Milli-Q®	10ml	
Total volume	50ml	

Hybridization Buffer pH 7.0

Reagents	Amount	Final concentration
Formamide	5ml	50%
20X SSC	1ml	2X
Sterile Milli-Q®	4ml	
Dextran sulphate	1gm	15%
Total volume	10ml	

10 X PBS

NaCl	40 g
KCl	1 g
Na ₂ HPO ₄	7.2 g
KH ₂ PO ₄	1.2 g
Milli-Q®	400 ml

Adjust pH to 7.4 with 1 N NaOH make a final volume to 500 ml with Milli-Q® and autoclave.

FISH Washing Buffer pH 7.0

Reagents	Volume	Final concentration
Formamide	25 ml	50%
20X SSC	5 ml	2X
Sterile Milli-Q®	20 ml	
Total volume	50 ml	

1X PBS/ MgCl ₂	Volume
2M MgCl ₂	25ml
1X PBS	950ml
Sterile Milli-Q®	25ml

Formaldehyde/1X PBS/ MgCl ₂	Volume
Formaldehyde	2.7ml
1X PBS/ MgCl ₂	100ml

0.8 % Agarose : 200 ml 1X TBE + 1.6 g Agarose

EtBr Solution: 100 ml Milli-Q® + 10 µl ETBr stock.

Formamide:

	50 % Formamide	70 % Formamide
Formamide	25 ml	35 ml
20 X SSC	5 ml	5 ml
Final volume was made up to 50 ml with Milli-Q®		

NP 40 Solution

2X SSC	50 ml
NP40	15 µl

Reagents For BAC Clones Culture And Plasmid DNA Extraction**Luria-Bertani Broth**

Tryptone	10 g
Yeast extract	5 g
NaCl	5 g

Final volume was made up to 1 L, autoclaved and stored at 4°C.

Low Salt Luria-Bertani Broth (for Cesium Chloride plasmid purification)

Tryptone	10 g
Yeast extract	5 g
NaCl	5 g
1M Tris pH 7.5	10 ml

Final volume was made up to 1 L, autoclaved and stored at 4°C.

Low Salt Luria-Bertani Plates (for Cesium Chloride plasmid purification)

Tryptone	10 g
Yeast extract	5 g
NaCl	5 g
1M Tris pH 7.5	10 ml
Agar	20 g

Final volume was made up to 1 L, autoclaved and stored at 4°C.

30 % Glycerol

30ml glycerol was dissolved in 70ml of Milli-Q®, autoclaved and stored at RT.

Chloramphenicol (Stock 1mg/ml)

1mg chloramphenicol in 50% ethanol was prepared and filtered through 45µm filter. Working stock of 25µg/ml was prepared and stored at -20°C.

Solution for Plasmid extraction

Solution 1: Resuspension buffer consists of GTE (50mM)

Glucose (50mM)	4.5g
Tris HCl (pH-8) (25mM)	1.5 g
EDTA (pH-8) (10mM)	1.9 g

Final volume was made up to 500ml, autoclaved and stored at 4°C.

Solution 2: Lysis solution consists of 0.2NaOH/1% SDS

10N NaOH was diluted to 0.2 N NaOH in a total volume of 10ml and 100mg SDS was added. The solution was prepared fresh.

Solution 3: Neutralizing solution

60 ml - potassium acetate ($\text{CH}_3\text{CO}_2\text{K}$, 5M),

11.5 ml - glacial acetic acid

Volume was made up to 50 ml with Milli-Q®.

Lysozyme (50mg/ml) was freshly prepared in 1M Tris HCl.

IHC and IF Experiment Reagents

10 X Phosphate Buffer Saline (pH-7.4) for 500 ml

Na_2HPO_4	5.45 g
$\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$	1.6 g
NaCl	45 g
pH was adjusted to 7.4 and final volume was made up to 500 ml Milli-Q®	

1 X Sodium Citrate Buffer (pH-6) for 500 ml-(freshly prepared)

1.47 g Tri Sodium Citrate was dissolved in 450 ml Milli-Q®. pH 6 was adjusted and final volume was made up to 500ml. Finally 250 µl of Tween 20 was added.

1 X TE Buffer (pH- 9) for 500 ml-(freshly prepared)

0.605 g Tris Base and 0.185 g EDTA were dissolved in 450 ml Milli-Q®. pH 9 was adjusted and final volume was made up to 500 ml. Finally 250 µl of Tween 20 was added.

1X EDTA buffer (pH-8) for 500ml-(freshly prepared)

0.186g EDTA was dissolved in 450 ml Milli-Q®. pH 8 was adjusted and final volume was made up to 500ml. Finally 250 µl of Tween 20 was added.

Blocking Solution for IF: 5% normal goat serum + 0.3% v/v Triton X-100 in 1X PBS.

Primary/Secondary Antibody Diluent for IF: 3% BSA+0.1% NP-40 in 1X PBS.

Primary Antibody Diluent IHC: 1X PBS

IF Wash Buffer

- 1) 1X PBS
- 2) 1X PBS + 0.1% NP-40

Chromic Acid Solution

K ₂ Cr ₂ O ₇	40 g
H ₂ SO ₄	80 ml
Make up volume to 1000ml	

DAPI: DAPI stock of 5µg/ml was prepared in Milli-Q® and stored at -20°C in the dark. Working solution of DAPI (0.5µg/ml) was prepared and stored at 4°C in the dark.

DAB Substrate Solution: 8mg DAB was dissolved in 10 ml 1XPBS; add 10µl H₂O₂ was added just before use.

Vectashield ABC Kit Reagents

Blocking Serum (100 µl)*	
Normal Horse Serum	1 µl
1X PBS	99 µl
Secondary Antibody (100 µl)*	
Normal Horse Serum	2 µl
Secondary Antibody	2 µl
1X PBS	96 µl
* Stored on ice	
ABC Reagent (100 µl)	
Reagent A	2 µl
Reagent B	2 µl
1X PBS	96 µl
Was prepare at least 30- 45 min before use	

4% Paraformaldehyde

4 g of paraformaldehyde (Sigma) was prepared in 100ml 1X PBS. This mixture was heated at 70°C on heating magnetic stirrer until the paraformaldehyde was dissolved completely (few drops of NaOH were added to get clear solution). The paraformaldehyde solution was filtered through a filter paper if white precipitate was observed). This solution can be stored at 4°C for about a month.

0.3% Triton X-100: 0.3ml Triton-X-100 (Sigma) was dissolved in 100ml PBS.

Endogenous Peroxidase Blocking (3%) 50 ml

Methanol	35 ml
1X PBS	15 ml
30 % H ₂ O ₂	5 ml

Haematoxylin (100 ml)

Haematoxylin crystal	5 g
Absolute alcohol	50 ml
Ammonium or potassium aluminium	100 mg
Mercuric oxide	2.5 mg
Milli-Q®	100 ml
Add 2ml acetic acid before use	

Preparation: Haematoxylin was dissolved in absolute alcohol. Potassium aluminium sulphate was added to heated Milli-Q®. The solution was cooled. Mix both the solutions and boiled rapidly. After cooling mercuric oxide was added slowly. The mixture was heated till dark purple colour was developed. The solution was cooled by placing in cold water. Before use 2 ml acetic acid was added to 100 ml of filtered stain.

Poly-L-Lysine (0.01%) Solution

Poly-L-Lysine (0.1%)	10 ml
Final volume was made up to 100 ml with Milli-Q®	

Cell Culture Reagents

Mitomycin C (stock 0.5 mg/ml)

To the Mitomycin C powder (in bottle) sterile Milli-Q® was added to obtain a final concentration of 0.5 mg/ml (in the culture hood). The stock solution was syringe filter sterilized. Aliquots were stored in liquid Nitrogen. A working concentration of 10 µg/ml was used for migration assay (20µl/well from stock was used).

MTT Solution (stock 5mg/ml)

100 mg MTT was dissolved in 20 ml 1X PBS. Filter sterilized and stored at 4°C, for long term storage MTT solution was kept at -20°C.

MTT Stop Solution (10% SDS in 0.01 N HCl)

100 ml stop solution was prepared by dissolving 10% SDS in Milli-Q®, followed by addition of 100µl concentrated HCl. The solution was stored at room temperature.

Puromycin

Puromycin powder (in bottle) was dissolved in sterile Milli-Q® in culture hood to obtain a final concentration of 1mg/ml. Stock solution was syringe filter sterilized and aliquots were stored at -20°C.

1X PBS: 1 L PBS was prepared by dissolving; NaCl-8 g; KH₂PO₄-0.2 g; KCL-0.2 g; Na₂HPO₄-2.18g; in Milli-Q®.

1 X Trypsin EDTA: 0.5g Trypsin and 1.06 ml of (0.5 EDTA stock) was dissolved in 1XPBS 1000ml, filter sterilized and stored at 4°C.

Freezing Media: 90%FBS+ 10% DMSO

Mini Prep and Cesium chloride method for Plasmid Purification

Solution 1

Glucose	0.9 g
1M Tris HCl (pH-8)	2.5 ml
0.5 M EDTA (pH-8)	2 ml

Final volume was made up to 100ml with Milli-Q®. The solution was autoclaved and stored at 4°C.

Solution 2

2M NaOH 1ml
10% SDS 1ml

Final volume was made up to 10 ml with Milli-Q®. The solution was freshly prepared.

Solution 3

60 ml of potassium acetate (5M) and 11.5 ml of glacial acetic acid were dissolved and final volume was made up to 100 ml with Milli-Q®. The solution was stored in the freezer.

Phenol Chloroform Mix: Phenol+chloroform+isoamyl alcohol: in 25:24:1 ratio was prepared. The solution was kept undisturbed for a few minutes and stored in refrigerator.

Lysozyme Stock Solution (for Cesium Prep): Stock 50mg/ml

Working Solution (For 8ml)

Lysozyme: 1.6 ml
Tris pH8: 80µl
Sterile Milli-Q®: 6.32ml

Cesium Chloride Gradient Solution

Exactly 1g of CsCl per 1ml of DNA solution and 0.5 ml of EtBr were added (use mask and gloves while handling CsCl).

Western Blotting Reagents

Cell Lysis Buffer

	Concentration	Stock	For 100 ml
Tris pH 6.8	50 mM	1 M	5 ml
Glycerol	10%		10 ml
SDS	2%	10%	20 ml
Final volume was made up to 100 ml with Milli-Q®			

3X Sample Buffer

	Concentration	Stock	For 100 ml
Tris pH 6.8	150 mM	1 M	15 ml
Glycerol	30%		30 ml
SDS	6%		6 g
BPB	0.3%		0.3 g

Final volume was made up to 100 ml with Milli-Q®. 150 µl of BME to 850 µl of 3X Sample Buffer was added before use.

10X Running Buffer (10X Electrode buffer)

Component	Final Concentration	Volume
Tris base (Sigma)	250mM	30g
Glycine (Sigma)	2.5M	187.7g
SDS (Sigma)	10%	10g
Heated Milli-Q®		750ml

After all the components were mixed, the volume of the solution was increased to 1 L by adding heated Milli-Q®.

Tris buffered saline with Tween-20 (TBS-T)

Component	Final concentration	Volume
1M Tris pH8.0 (Sigma)	10mM	10ml
2.5M NaCl	150mM	60ml
Tween-20		1ml
Milli-Q®		930ml

The components were mixed by vigorous stirring either manually or on a magnetic stirrer.

2.5M NaCl: 146.1 g of NaCl was added to 950ml of Milli-Q®.

SDS- Page Gel

Resolving Gel: 5% resolves 60 -200 kDa; 10% resolves 16-70 kDa; 12.5% resolves 14-60 kDa; 15% resolves 12- 45kDa (TEMED in µl, rest all volumes in ml,)

For 10 ml	6%	7.5%	10%	12%	15%
Milli-Q®	5.3	4.8	4.0	3.3	2.3
30% Acrylamide	2	2.5	3.3	4	5
1.5 M Tris pH	2.5	2.5	2.5	2.5	2.5
10% SDS	0.1	0.1	0.1	0.1	0.1
10% APS	0.1	0.1	0.1	0.1	0.1
TEMED	8	7 µl	4 µl	4 µl	4 µl

Stacking Gels

Comp.	2ml	3ml	4ml	5ml	6ml	8ml	10ml
Milli-Q®	1.4	2.1	2.7	3.4	4.1	5.5	6.8
30% Acrylamide	0.33	0.5	0.67	0.83	1	1.3	1.7
1 M Tris pH 6.8	0.25	0.38	0.5	0.63	0.75	1	1.25
10% SDS	20 µl	30 µl	40 µl	50 µl	60 µl	80 µl	100 µl
10% APS	20 µl	30 µl	40 µl	50 µl	60 µl	80 µl	100 µl
TEMED	2 µl	3 µl	4 µl	5 µl	6 µl	8 µl	10 µl

1.5 M Tris pH 8.8: 181.7 g Tris was dissolved in 700ml Milli-Q®, pH 8.8 was adjusted and final volume was made up to 1 L. (Solution was autoclaved).

1M Tris pH 6.8: 121.1g Tris was dissolved in 950ml Milli-Q®, pH was set to 6.8 and final volume was made up to 1 L with Milli-Q® (Solution was autoclaved).

1 M Tris, pH 8.0: 121.1 g Tris was dissolved in 950ml Milli-Q®, pH was set to 8 and final volume was made up to 1 L with Milli-Q® (Solution was autoclaved).

0.5 M EDTA (pH 8): 18.61 g EDTA was dissolved in 60ml Milli-Q® (NaOH pellets were added till solution gets clear, pH was set to 8). Final volume was made up to 100 ml (Solution was autoclaved).

2.5M NaCl (500 ml): 73.05g NaCl was dissolved in 300 ml Milli-Q® and final volume was made up (Solution was autoclaved).

Water Saturated n-Butanol: 50 ml n-Butanol + 5 ml Milli-Q® was combined in a bottle. The solution was mixed by inverting the bottle upside down and top phase was used.

Sealing Agar: 2% agar was prepared in Milli-Q®

30% Acrylamide: 30 g Acrylamide was dissolved in 100ml Milli-Q®.

0.8% Bis-acrylamide: 0.8 g *N,N'*-Methylenebis(acrylamide) was dissolved in 100ml Milli-Q®.

AB Solution: 30 g acrylamide + 1 g Bis-acrylamide was added in 100 ml Milli-Q®. Ascertain pH less than 7. The solution was filtered and stored at 4°C in a dark bottle.

1 X Transfer Buffer

COMP.	3 L	4 L
Tris Base	9.1 g	12.1 g
Glycine	43.2 g	57.6 g
Methanol	600 ml	800 ml
10 % SDS	3 ml	4 ml
Milli-Q®	Make up to 3 L	Make up to 4 L

10% APS: 10g APS+ 100ml Milli-Q®

1° Antibody Diluent: 1% BSA in TBST + 0.02% sodium azide (for 20ml: 0.2 gm BSA+ 200µl Sodium azide (2% stock, wear gloves while handling)); Note: antibody prepared in azide were reused 2-3 times and stable for at least a month at 4°C.

2° Antibody Diluent: 2.5% milk in TBST.

Blocking Solution: 5% Milk or 5% BSA in TBST.

2% sodium azide (100X stock): 2g sodium azide in 100ml Milli-Q®.

Fast Green Stain: 0.1% Fast Green in 10% acetic acid and 20% methanol (0.2 g Fast Green+ 20 ml acetic acid+ 40 ml methanol+ 140 ml Milli-Q®).

Folin- Lowry Protein Estimation Reagents

1. 1mg/ml BSA (standard).

2. **CTC (Copper Tartarate Carbonate) Solution:** I] 20% Na₂CO₃ [20 g Na₂CO₃ in 100 ml of Milli-Q®]; II] 0.4 % CuSO₄ [0.4 g CuSO₄ in 100 ml of Milli-Q®]; III] 0.4% sodium potassium tartarate [0.4 g Na-k-tartarate in 40ml Milli-Q®]. Solution II and III were mixed with constant stirring and filtered, followed by the addition of 50 ml of this mixture to 50 ml solution I (with constant stirring). The solution was filtered and stored at room temperature in a dark bottle (CTC solution is stable for 8 week at room temperature).

3. 10 % SDS

4. 0.8 % NaOH: 3.2 g NaOH was dissolved in 100 ml Milli-Q®.

5. FC (Folin Ciocalteu) Reagent: Commercially available FC reagent was diluted in 1:5 ratio (1ml FC + 4 ml Milli-Q®).

6. Working reagents

Solution 1: [CTC:SDS: NaOH : Milli-Q®]: Milli-Q®

Therefore 1ml CTC + 1ml SDS+ 1ml NaOH + 1ml Milli-Q®= 4 ml + 4ml Milli-Q®= 8 ml solution 1.

Solution 2: 1ml FC + 4ml of Milli-Q® = 5ml FC reagent.

BSA (std) conc.	Milli-Q® (µl)	BSA (µl)	Samples (µl)	Solution 1		Solution 2		Absorbance at 750 nm
Blank	1000	-	-	1ml	Vortex And incubate for 10 min in dark	0.5 ml	Vortex and incubate for 30 min in dark	Vortex and record Absorbance at 750 nm
20 µg	980	20	-					
40 µg	960	40	-					
60µg	940	60	-					
80 µg	920	80	-					
100 µg	900	100	-					
Sample1	995	-	5					

8. References

References

1. Dikshit, R., et al., *Cancer mortality in India: a nationally representative survey*. Lancet, 2012. **379**(9828): p. 1807-16.
2. Chen, Y.J., et al., *Genome-wide profiling of oral squamous cell carcinoma*. J Pathol, 2004. **204**(3): p. 326-32.
3. Pentenero, M., S. Gandolfo, and M. Carrozzo, *Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: a review of the literature*. Head Neck, 2005. **27**(12): p. 1080-91.
4. Forastiere, A., et al., *Head and neck cancer*. The New England journal of medicine, 2001. **345**(26): p. 1890-900.
5. Mork, J., et al., *Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck*. The New England journal of medicine, 2001. **344**(15): p. 1125-31.
6. Warnakulasuriya, S., et al., *Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement*. J Oral Pathol Med, 2008. **37**(3): p. 127-33.
7. Baldwin, C., et al., *Multiple microalterations detected at high frequency in oral cancer*. Cancer Res, 2005. **65**(17): p. 7561-7.
8. Smeets, S.J., et al., *Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus*. Oncogene, 2006. **25**(17): p. 2558-64.
9. O'Regan, E.M., et al., *Distinct array comparative genomic hybridization profiles in oral squamous cell carcinoma occurring in young patients*. Head Neck, 2006. **28**(4): p. 330-8.
10. Ambatipudi, S., et al., *Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome*. PLoS One, 2011. **6**(2): p. e17250.
11. Agrawal, N., et al., *Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1*. Science, 2011. **333**(6046): p. 1154-7.
12. Brieger, J., et al., *Chromosomal aberrations in premalignant and malignant squamous epithelium*. Cancer Genet Cytogenet, 2003. **144**(2): p. 148-55.
13. Salahshourifar, I., et al., *Genomic DNA copy number alterations from precursor oral lesions to oral squamous cell carcinoma*. Oral Oncol, 2014. **50**(5): p. 404-12.
14. Mendez, E., et al., *Can a metastatic gene expression profile outperform tumor size as a predictor of occult lymph node metastasis in oral cancer patients?* Clin Cancer Res, 2011. **17**(8): p. 2466-73.
15. Braakhuis, B.J., et al., *Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus*. J Natl Cancer Inst, 2004. **96**(13): p. 998-1006.
16. Klussmann, J.P., et al., *Genetic signatures of HPV-related and unrelated oropharyngeal carcinoma and their prognostic implications*. Clin Cancer Res, 2009. **15**(5): p. 1779-86.
17. Baldwin, C., et al., *Multiple microalterations detected at high frequency in oral cancer*. Cancer research, 2005. **65**(17): p. 7561-7.
18. Bashyam, M.D., et al., *Array-based comparative genomic hybridization identifies localized DNA amplifications and homozygous deletions in pancreatic cancer*. Neoplasia, 2005. **7**(6): p. 556-62.
19. Gollin, S.M., *Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease*. Head Neck, 2001. **23**(3): p. 238-53.
20. Fernandez, L.A., et al., *Oncogenic YAP promotes radioresistance and genomic instability in medulloblastoma through IGF2-mediated Akt activation*. Oncogene, 2012. **31**(15): p. 1923-37.
21. Ginos, M.A., et al., *Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck*. Cancer Res, 2004. **64**(1): p. 55-63.

22. Imoto, I., et al., *Identification of clAP1 as a candidate target gene within an amplicon at 11q22 in esophageal squamous cell carcinomas*. Cancer research, 2001. **61**(18): p. 6629-34.
23. Imoto, I., et al., *Expression of clAP1, a target for 11q22 amplification, correlates with resistance of cervical cancers to radiotherapy*. Cancer research, 2002. **62**(17): p. 4860-6.
24. Boffetta, P., et al., *Smokeless tobacco and cancer*. Lancet Oncol, 2008. **9**(7): p. 667-75.
25. Gupta, P.C. and C.S. Ray, *Smokeless tobacco and health in India and South Asia*. Respiriology, 2003. **8**(4): p. 419-31.
26. Leemans, C.R., B.J. Braakhuis, and R.H. Brakenhoff, *The molecular biology of head and neck cancer*. Nat Rev Cancer, 2011. **11**(1): p. 9-22.
27. Lu, D., X. Yu, and Y. Du, *Meta-analyses of the effect of cytochrome P450 2E1 gene polymorphism on the risk of head and neck cancer*. Mol Biol Rep, 2011. **38**(4): p. 2409-16.
28. Chuang, S.C., et al., *Sequence Variants and the Risk of Head and Neck Cancer: Pooled Analysis in the INHANCE Consortium*. Front Oncol, 2011. **1**: p. 13.
29. Winn, D.M., et al., *The INHANCE consortium: toward a better understanding of the causes and mechanisms of head and neck cancer*. Oral Dis, 2015. **21**(6): p. 685-93.
30. Sankaranarayanan, R., et al., *Oral Cancer: Prevention, Early Detection, and Treatment*, in *Cancer: Disease Control Priorities, Third Edition (Volume 3)*, H. Gelband, et al., Editors. 2015: Washington (DC).
31. Humans, I.W.G.o.t.E.o.C.R.t., *Betel-quid and areca-nut chewing and some areca-nut derived nitrosamines*. IARC Monogr Eval Carcinog Risks Hum, 2004. **85**: p. 1-334.
32. Humans, I.W.G.o.t.E.o.C.R.t., *Alcohol consumption and ethyl carbamate*. IARC Monogr Eval Carcinog Risks Hum, 2010. **96**: p. 3-1383.
33. Humans, I.W.G.o.t.E.o.C.R.t., *Smokeless tobacco and some tobacco-specific N-nitrosamines*. IARC Monogr Eval Carcinog Risks Hum, 2007. **89**: p. 1-592.
34. Ragin, C.C. and E. Taioli, *Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: review and meta-analysis*. Int J Cancer, 2007. **121**(8): p. 1813-20.
35. Michaud, D.S., et al., *High-risk HPV types and head and neck cancer*. Int J Cancer, 2014. **135**(7): p. 1653-61.
36. Horewicz, V.V., et al., *Human papillomavirus-16 prevalence in gingival tissue and its association with periodontal destruction: a case-control study*. J Periodontol, 2010. **81**(4): p. 562-8.
37. Anantharaman, D., et al., *Human papillomavirus infections and upper aero-digestive tract cancers: the ARCAGE study*. J Natl Cancer Inst, 2013. **105**(8): p. 536-45.
38. Zaravinos, A., *An updated overview of HPV-associated head and neck carcinomas*. Oncotarget, 2014. **5**(12): p. 3956-69.
39. Ang, K.K., et al., *Human papillomavirus and survival of patients with oropharyngeal cancer*. N Engl J Med, 2010. **363**(1): p. 24-35.
40. Duray, A., et al., *Human papillomavirus predicts the outcome following concomitant chemoradiotherapy in patients with head and neck squamous cell carcinomas*. Oncol Rep, 2013. **30**(1): p. 371-6.
41. Gillison, M.L., *Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity*. Semin Oncol, 2004. **31**(6): p. 744-54.
42. Fakhry, C., et al., *Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial*. J Natl Cancer Inst, 2008. **100**(4): p. 261-9.
43. van der Waal, I., *Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management*. Oral Oncol, 2009. **45**(4-5): p. 317-23.

44. Waldron, C.A. and W.G. Shafer, *Leukoplakia revisited. A clinicopathologic study 3256 oral leukoplakias*. Cancer, 1975. **36**(4): p. 1386-92.
45. Petti, S., *Pooled estimate of world leukoplakia prevalence: a systematic review*. Oral Oncol, 2003. **39**(8): p. 770-80.
46. Neville, B.W. and T.A. Day, *Oral cancer and precancerous lesions*. CA Cancer J Clin, 2002. **52**(4): p. 195-215.
47. Amagasa, T., M. Yamashiro, and N. Uzawa, *Oral premalignant lesions: from a clinical perspective*. Int J Clin Oncol, 2011. **16**(1): p. 5-14.
48. Rivera, C., *Essentials of oral cancer*. Int J Clin Exp Pathol, 2015. **8**(9): p. 11884-94.
49. mrzezo. 11: *Leukoplakia, Oral Dysplasia, and Squamous Cell Carcinoma*. 2015 [cited 2015 Jan 12]; Available from: <http://pocketdentistry.com/11-leukoplakia-oral-dysplasia-and-squamous-cell-carcinoma/>.
50. Chiesa, F., et al., *Randomized trial of fenretinide (4-HPR) to prevent recurrences, new localizations and carcinomas in patients operated on for oral leukoplakia: long-term results*. Int J Cancer, 2005. **115**(4): p. 625-9.
51. Chandu, A. and A.C. Smith, *The use of CO2 laser in the treatment of oral white patches: outcomes and factors affecting recurrence*. Int J Oral Maxillofac Surg, 2005. **34**(4): p. 396-400.
52. Messadi, D.V., *Diagnostic aids for detection of oral precancerous conditions*. Int J Oral Sci, 2013. **5**(2): p. 59-65.
53. Markopoulos, A.K., *Current aspects on oral squamous cell carcinoma*. Open Dent J, 2012. **6**: p. 126-30.
54. Castelijns, J.A. and M.W. van den Brekel, *Detection of lymph node metastases in the neck: radiologic criteria*. AJNR Am J Neuroradiol, 2001. **22**(1): p. 3-4.
55. Mehanna, H., et al., *PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer*. N Engl J Med, 2016. **374**(15): p. 1444-54.
56. D'Cruz, A.K., et al., *Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer*. N Engl J Med, 2015. **373**(6): p. 521-9.
57. D'Cruz, A., et al., *Consensus recommendations for management of head and neck cancer in Asian countries: a review of international guidelines*. Oral Oncol, 2013. **49**(9): p. 872-7.
58. 2015; Available from: <http://www.cancercenter.com/oral-cancer/surgery/>.
59. Huang, S.H. and B. O'Sullivan, *Oral cancer: Current role of radiotherapy and chemotherapy*. Med Oral Patol Oral Cir Bucal, 2013. **18**(2): p. e233-40.
60. Hidetaka, Y., et al., *Neoadjuvant Chemotherapy with S-1 for Patients with Oral Squamous Cell Carcinoma*. Journal of Cancer Science & Therapy, 2010.
61. Takacs-Nagy, Z., et al., *Docetaxel, cisplatin and 5-fluorouracil induction chemotherapy followed by chemoradiotherapy or chemoradiotherapy alone in stage III-IV unresectable head and neck cancer: Results of a randomized phase II study*. Strahlenther Onkol, 2015. **191**(8): p. 635-41.
62. Posner, M.R., et al., *Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer*. N Engl J Med, 2007. **357**(17): p. 1705-15.
63. Dassonville, O., et al., *EGFR targeting therapies: monoclonal antibodies versus tyrosine kinase inhibitors. Similarities and differences*. Crit Rev Oncol Hematol, 2007. **62**(1): p. 53-61.
64. Sacco, A.G. and F.P. Worden, *Molecularly targeted therapy for the treatment of head and neck cancer: a review of the ErbB family inhibitors*. Onco Targets Ther, 2016. **9**: p. 1927-43.
65. Pai, S.I. and W.H. Westra, *Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment*. Annu Rev Pathol, 2009. **4**: p. 49-70.
66. Westra, W.H. and D. Sidransky, *Fluorescence visualization in oral neoplasia: shedding light on an old problem*. Clin Cancer Res, 2006. **12**(22): p. 6594-7.

-
67. Albertson, D.G., et al., *Chromosome aberrations in solid tumors*. Nat Genet, 2003. **34**(4): p. 369-76.
 68. Reshmi, S.C. and S.M. Gollin, *Chromosomal instability in oral cancer cells*. J Dent Res, 2005. **84**(2): p. 107-17.
 69. Albertson, D.G. and D. Pinkel, *Genomic microarrays in human genetic disease and cancer*. Hum Mol Genet, 2003. **12 Spec No 2**: p. R145-52.
 70. Tang, Y.C. and A. Amon, *Gene copy-number alterations: a cost-benefit analysis*. Cell, 2013. **152**(3): p. 394-405.
 71. Coe, B.P., et al., *Evolving strategies for global gene expression analysis of cancer*. J Cell Physiol, 2008. **217**(3): p. 590-7.
 72. Waddell, N., *Microarray-based DNA profiling to study genomic aberrations*. IUBMB Life, 2008. **60**(7): p. 437-40.
 73. Aradhya, S. and A.M. Cherry, *Array-based comparative genomic hybridization: clinical contexts for targeted and whole-genome designs*. Genet Med, 2007. **9**(9): p. 553-9.
 74. Kallioniemi, A., *CGH microarrays and cancer*. Curr Opin Biotechnol, 2008. **19**(1): p. 36-40.
 75. Cowell, J.K. and L. Hawthorn, *The application of microarray technology to the analysis of the cancer genome*. Curr Mol Med, 2007. **7**(1): p. 103-20.
 76. Costa, J.L., et al., *Array comparative genomic hybridization copy number profiling: a new tool for translational research in solid malignancies*. Semin Radiat Oncol, 2008. **18**(2): p. 98-104.
 77. Akavia, U.D., et al., *An integrated approach to uncover drivers of cancer*. Cell, 2010. **143**(6): p. 1005-17.
 78. Freier, K., et al., *Recurrent copy number gain of transcription factor SOX2 and corresponding high protein expression in oral squamous cell carcinoma*. Genes Chromosomes Cancer, 2010. **49**(1): p. 9-16.
 79. Cervigne, N.K., et al., *Recurrent genomic alterations in sequential progressive leukoplakia and oral cancer: drivers of oral tumorigenesis?* Hum Mol Genet, 2014. **23**(10): p. 2618-28.
 80. Yoshioka, S., et al., *Genomic profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization*. PLoS One, 2013. **8**(2): p. e56165.
 81. Sparano, A., et al., *Genome-wide profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization*. Laryngoscope, 2006. **116**(5): p. 735-41.
 82. Garnis, C., et al., *Genomic imbalances in precancerous tissues signal oral cancer risk*. Mol Cancer, 2009. **8**: p. 50.
 83. Tsui, I.F., et al., *Multiple aberrations of chromosome 3p detected in oral premalignant lesions*. Cancer Prev Res (Phila), 2008. **1**(6): p. 424-9.
 84. Liu, C.J., et al., *Array-comparative genomic hybridization to detect genomewide changes in microdissected primary and metastatic oral squamous cell carcinomas*. Mol Carcinog, 2006. **45**(10): p. 721-31.
 85. Roepman, P., et al., *An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas*. Nat Genet, 2005. **37**(2): p. 182-6.
 86. Kondoh, N., et al., *Gene expression signatures that can discriminate oral leukoplakia subtypes and squamous cell carcinoma*. Oral Oncol, 2007. **43**(5): p. 455-62.
 87. Sumino, J., et al., *Gene expression changes in initiation and progression of oral squamous cell carcinomas revealed by laser microdissection and oligonucleotide microarray analysis*. Int J Cancer, 2013. **132**(3): p. 540-8.
 88. Odani, T., et al., *Gene expression profiles of oral leukoplakia and carcinoma: genome-wide comparison analysis using oligonucleotide microarray technology*. Int J Oncol, 2006. **28**(3): p. 619-24.

89. Xu, C., et al., *Integrative analysis of DNA copy number and gene expression in metastatic oral squamous cell carcinoma identifies genes associated with poor survival*. Mol Cancer, 2010. **9**: p. 143.
90. Pickering, C.R., et al., *Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers*. Cancer Discov, 2013. **3**(7): p. 770-81.
91. Ambatipudi, S., et al., *Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers*. Genes Chromosomes Cancer, 2012. **51**(2): p. 161-73.
92. Winder, D.M., et al., *Sensitive HPV detection in oropharyngeal cancers*. BMC Cancer, 2009. **9**: p. 440.
93. Evans, M.F., et al., *Touchdown General Primer (GP5+/GP6+) PCR and optimized sample DNA concentration support the sensitive detection of human papillomavirus*. BMC Clin Pathol, 2005. **5**: p. 10.
94. Schache, A.G., et al., *Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma*. Br J Cancer, 2013. **108**(6): p. 1332-9.
95. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome*. Nat Genet, 2004. **36**(9): p. 949-51.
96. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biol, 2004. **5**(10): p. R80.
97. R Development Core Team, *R: A Language and Environment for Statistical Computing*. 2013, R Foundation for Statistical Computing: Vienna, Austria.
98. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
99. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data*. Bioinformatics, 2007. **23**(6): p. 657-63.
100. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers*. Genome biology, 2011. **12**(4): p. 1.
101. Therneau, T.M. and P.M. Grambsch, *Modeling survival data: extending the Cox model*. 2000: Springer Science & Business Media.
102. Gertz, E.M., et al., *Evaluating annotations of an Agilent expression chip suggests that many features cannot be interpreted*. BMC Genomics, 2009. **10**: p. 566.
103. Hwa, Y., *Yang with contributions from Agnes Paquet and Sandrine Dudoit: marray: Exploratory analysis for two-color spotted microarray data*. R package version, 2007. **1**(0).
104. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society Series B (Methodological), 1995. **57**: p. 289--300.
105. Falcon, S. and R. Gentleman, *Using GOstats to test gene lists for GO term association*. Bioinformatics, 2007. **23**(2): p. 257-258.
106. Takikita, M., et al., *Fascin and CK4 as biomarkers for esophageal squamous cell carcinoma*. Anticancer Res, 2011. **31**(3): p. 945-52.
107. Biesheuvel, C.J., et al., *Polytomous logistic regression analysis could be applied more often in diagnostic research*. J Clin Epidemiol, 2008. **61**(2): p. 125-34.
108. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function*. Genome Res, 2003. **13**(9): p. 2129-41.
109. Mi, H., et al., *The PANTHER database of protein families, subfamilies, functions and pathways*. Nucleic Acids Res, 2005. **33**(Database issue): p. D284-8.
110. Dai, Z., et al., *A comprehensive search for DNA amplification in lung cancer identifies inhibitors of apoptosis clAP1 and clAP2 as candidate oncogenes*. Hum Mol Genet, 2003. **12**(7): p. 791-801.

111. Zhou, A.Y., et al., *IKKepsilon-mediated tumorigenesis requires K63-linked polyubiquitination by a cIAP1/cIAP2/TRAFF2 E3 ubiquitin ligase complex*. Cell Rep, 2013. **3**(3): p. 724-33.
112. Beug, S.T., et al., *Combinatorial cancer immunotherapy strategies with proapoptotic small-molecule IAP antagonists*. Int J Dev Biol, 2015. **59**(1-3): p. 141-7.
113. Sethi, G., et al., *Pinitol targets nuclear factor-kappaB activation pathway leading to inhibition of gene products associated with proliferation, apoptosis, invasion, and angiogenesis*. Mol Cancer Ther, 2008. **7**(6): p. 1604-14.
114. Che, X., et al., *Nuclear cIAP1 overexpression is a tumor stage- and grade-independent predictor of poor prognosis in human bladder cancer patients*. Urol Oncol, 2012. **30**(4): p. 450-6.
115. Nagata, M., et al., *Overexpression of cIAP2 contributes to 5-FU resistance and a poor prognosis in oral squamous cell carcinoma*. Br J Cancer, 2011. **105**(9): p. 1322-30.
116. Lu, J., et al., *Therapeutic potential and molecular mechanism of a novel, potent, nonpeptide, Smac mimetic SM-164 in combination with TRAIL for cancer treatment*. Mol Cancer Ther, 2011. **10**(5): p. 902-14.
117. Miura, K., et al., *Inhibitor of apoptosis protein family as diagnostic markers and therapeutic targets of colorectal cancer*. Surg Today, 2011. **41**(2): p. 175-82.
118. Peng, C.H., et al., *Somatic copy number alterations detected by ultra-deep targeted sequencing predict prognosis in oral cavity squamous cell carcinoma*. Oncotarget, 2015. **6**(23): p. 19891-906.
119. Wood, H.M., et al., *The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing*. J Pathol, 2015. **237**(3): p. 296-306.
120. Bergshoeff, V.E., et al., *Chromosome instability predicts progression of premalignant lesions of the larynx*. Pathology, 2014. **46**(3): p. 216-24.
121. Giaretti, W., et al., *Chromosomal instability, DNA index, dysplasia, and subsite in oral premalignancy as intermediate endpoints of risk of cancer*. Cancer Epidemiol Biomarkers Prev, 2013. **22**(6): p. 1133-41.
122. Poh, C.F., et al., *Unique FISH patterns associated with cancer progression of oral dysplasia*. J Dent Res, 2012. **91**(1): p. 52-7.
123. Castagnola, P., et al., *Genomic DNA Copy Number Aberrations, Histological Diagnosis, Oral Subsite and Aneuploidy in OPMDs/OSCCs*. PLoS One, 2015. **10**(11): p. e0142294.
124. Sugahara, K., et al., *Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma*. Int J Oncol, 2011. **39**(4): p. 761-9.
125. Uchida, K., et al., *Loss of 3p26.3 is an independent prognostic factor in patients with oral squamous cell carcinoma*. Oncol Rep, 2011. **26**(2): p. 463-9.
126. Vincent-Chong, V.K., et al., *Genome wide analysis of chromosomal alterations in oral squamous cell carcinomas revealed over expression of MGAM and ADAM9*. PLoS One, 2013. **8**(2): p. e54705.
127. Noutomi, Y., et al., *Comparative genomic hybridization reveals genetic progression of oral squamous cell carcinoma from dysplasia via two different tumourigenic pathways*. J Pathol, 2006. **210**(1): p. 67-74.
128. Skinner, H.D., et al., *Proteomic profiling identifies PTK2/FAK as a driver of radioresistance in HPV negative head and neck cancer*. Clin Cancer Res, 2016.
129. Roy-Luzarraga, M. and K. Hodivala-Dilke, *Molecular Pathways: Endothelial Cell FAK-A Target for Cancer Treatment*. Clin Cancer Res, 2016.
130. Matsuda, R., et al., *LY6K is a novel molecular target in bladder cancer on basis of integrate genome-wide profiling*. Br J Cancer, 2011. **104**(2): p. 376-86.
131. Stine, Z.E., et al., *MYC, Metabolism, and Cancer*. Cancer Discov, 2015. **5**(10): p. 1024-39.

132. Kuzyk, A. and S. Mai, *c-MYC-induced genomic instability*. Cold Spring Harb Perspect Med, 2014. **4**(4): p. a014373.
133. Iyoda, M., et al., *Epithelial cell transforming sequence 2 in human oral cancer*. PLoS One, 2010. **5**(11): p. e14082.
134. Ren, Z.H., C.P. Zhang, and T. Ji, *Expression of SOX2 in oral squamous cell carcinoma and the association with lymph node metastasis*. Oncol Lett, 2016. **11**(3): p. 1973-1979.
135. Fu, T.Y., et al., *Association of OCT4, SOX2, and NANOG expression with oral squamous cell carcinoma progression*. J Oral Pathol Med, 2016. **45**(2): p. 89-95.
136. Attramadal, C.G., et al., *High nuclear SOX2 expression is associated with radiotherapy response in small (T1/T2) oral squamous cell carcinoma*. J Oral Pathol Med, 2015. **44**(7): p. 515-22.
137. Shah, S., et al., *Genetic alterations of the PIK3CA oncogene in human oral squamous cell carcinoma in an Indian population*. Oral Surg Oral Med Oral Pathol Oral Radiol, 2015. **120**(5): p. 628-35.
138. Chen, Y., et al., *PIK3CA is critical for the proliferation, invasiveness, and drug resistance of human tongue carcinoma cells*. Oncol Res, 2011. **19**(12): p. 563-71.
139. Guan, X.Y., et al., *Isolation of a novel candidate oncogene within a frequently amplified region at 3q26 in ovarian cancer*. Cancer Res, 2001. **61**(9): p. 3806-9.
140. Li, Y., et al., *Increased expression of EIF5A2, via hypoxia or gene amplification, contributes to metastasis and angiogenesis of esophageal squamous cell carcinoma*. Gastroenterology, 2014. **146**(7): p. 1701-13 e9.
141. Liu, Y., et al., *EIF5A2 is a novel chemoresistance gene in breast cancer*. Breast Cancer, 2015. **22**(6): p. 602-7.
142. Pathare, S., et al., *Construction of oncogenetic tree models reveals multiple pathways of oral cancer progression*. Int J Cancer, 2009. **124**(12): p. 2864-71.
143. Kuriakose, M.A., *Commentary on Elective versus therapeutic neck dissection in node-negative oral cancer*. Indian J Med Paediatr Oncol, 2015. **36**(3): p. 137-9.
144. Bhattacharya, A., et al., *Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis*. Clin Cancer Res, 2011. **17**(22): p. 7024-34.
145. Prochownik, E.V., *c-Myc: linking transformation and genomic instability*. Curr Mol Med, 2008. **8**(6): p. 446-58.
146. Ma, C., et al., *Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues*. Cancer Biol Ther, 2009. **8**(10): p. 907-16.
147. Qi, S., et al., *Expression of cIAP-1 correlates with nodal metastasis in squamous cell carcinoma of the tongue*. Int J Oral Maxillofac Surg, 2008. **37**(11): p. 1047-53.
148. Tanimoto, T., et al., *Nuclear expression of cIAP-1, an apoptosis inhibiting protein, predicts lymph node metastasis and poor patient prognosis in head and neck squamous cell carcinomas*. Cancer Lett, 2005. **224**(1): p. 141-51.
149. Imoto, I., et al., *Expression of cIAP1, a target for 11q22 amplification, correlates with resistance of cervical cancers to radiotherapy*. Cancer Res, 2002. **62**(17): p. 4860-6.
150. Ferreira, C.G., et al., *Assessment of IAP (inhibitor of apoptosis) proteins as predictors of response to chemotherapy in advanced non-small-cell lung cancer patients*. Ann Oncol, 2001. **12**(6): p. 799-805.
151. Chen, X., et al., *Expression of the IAP protein family acts cooperatively to predict prognosis in human bladder cancer patients*. Oncol Lett, 2013. **5**(4): p. 1278-1284.
152. Samuel, T., et al., *cIAP1 Localizes to the nuclear compartment and modulates the cell cycle*. Cancer Res, 2005. **65**(1): p. 210-8.
153. Bertrand, M.J., et al., *cIAP1 and cIAP2 facilitate cancer cell survival by functioning as E3 ligases that promote RIP1 ubiquitination*. Mol Cell, 2008. **30**(6): p. 689-700.

154. Ma, Y., et al., *Expression of c-Jun, p73, Casp9, and N-ras in thymic epithelial tumors: relationship with the current WHO classification systems*. Diagn Pathol, 2012. **7**: p. 120.
155. Kiflemariam, S., et al., *Tumor vessel up-regulation of INSR revealed by single-cell expression analysis of the tyrosine kinome and phosphatome in human cancers*. Am J Pathol, 2015. **185**(6): p. 1600-9.
156. Yonezawa, T., et al., *Limitrin, a novel immunoglobulin superfamily protein localized to glia limitans formed by astrocyte endfeet*. Glia, 2003. **44**(3): p. 190-204.
157. Zhang, K., et al., *Silencing dishevelled-1 sensitizes paclitaxel-resistant human ovarian cancer cells via AKT/GSK-3beta/beta-catenin signalling*. Cell Prolif, 2015. **48**(2): p. 249-58.
158. Jin, C., et al., *Molecular cytogenetic characterization of the 11q13 amplicon in head and neck squamous cell carcinoma*. Cytogenet Genome Res, 2006. **115**(2): p. 99-106.
159. Jin, Y., et al., *FISH characterization of head and neck carcinomas reveals that amplification of band 11q13 is associated with deletion of distal 11q*. Genes Chromosomes Cancer, 1998. **22**(4): p. 312-20.
160. Akervall, J., et al., *Overexpression of cyclin D1 correlates with sensitivity to cisplatin in squamous cell carcinoma cell lines of the head and neck*. Acta Otolaryngol, 2004. **124**(7): p. 851-7.
161. Huang, X., et al., *High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells*. Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11369-74.
162. Kyomoto, R., et al., *Cyclin-D1-gene amplification is a more potent prognostic factor than its protein over-expression in human head-and-neck squamous-cell carcinoma*. Int J Cancer, 1997. **74**(6): p. 576-81.
163. Timpson, P., et al., *Aberrant expression of cortactin in head and neck squamous cell carcinoma cells is associated with enhanced cell proliferation and resistance to the epidermal growth factor receptor inhibitor gefitinib*. Cancer Res, 2007. **67**(19): p. 9304-14.
164. Shang, L., et al., *ANO1 protein as a potential biomarker for esophageal cancer prognosis and precancerous lesion development prediction*. Oncotarget, 2016.
165. Chung, C.H., et al., *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, 2006. **66**(16): p. 8210-8.
166. Mendez, E., et al., *A genetic expression profile associated with oral cancer identifies a group of patients at high risk of poor survival*. Clin Cancer Res, 2009. **15**(4): p. 1353-61.
167. Chen, C., et al., *Gene expression profiling identifies genes predictive of oral squamous cell carcinoma*. Cancer Epidemiol Biomarkers Prev, 2008. **17**(8): p. 2152-62.
168. Ziober, A.F., et al., *Identification of a gene signature for rapid screening of oral squamous cell carcinoma*. Clin Cancer Res, 2006. **12**(20 Pt 1): p. 5960-71.
169. Seder, C.W., et al., *Upregulated INHBA expression may promote cell proliferation and is associated with poor survival in lung adenocarcinoma*. Neoplasia, 2009. **11**(4): p. 388-96.
170. Leung, C.S., et al., *Calcium-dependent FAK/CREB/TNNC1 signalling mediates the effect of stromal MFAP5 on ovarian cancer metastatic potential*. Nat Commun, 2014. **5**: p. 5092.
171. Yao, M., et al., *Tumor signatures of PTHLH overexpression, high serum calcium, and poor prognosis were observed exclusively in clear cell but not non clear cell renal carcinomas*. Cancer Med, 2014. **3**(4): p. 845-54.
172. Toruner, G.A., et al., *Association between gene expression profile and tumor invasion in oral squamous cell carcinoma*. Cancer Genet Cytogenet, 2004. **154**(1): p. 27-35.
173. Reichelt, J. and T.M. Magin, *Hyperproliferation, induction of c-Myc and 14-3-3sigma, but no cell fragility in keratin-10-null mice*. J Cell Sci, 2002. **115**(Pt 13): p. 2639-50.

174. Pan, X., R.P. Hobbs, and P.A. Coulombe, *The expanding significance of keratin intermediate filaments in normal and diseased epithelia*. *Curr Opin Cell Biol*, 2013. **25**(1): p. 47-56.
175. Sevilla, L.M., et al., *Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier*. *J Cell Biol*, 2007. **179**(7): p. 1599-612.
176. Khammanivong, A., et al., *Involvement of calprotectin (S100A8/A9) in molecular pathways associated with HNSCC*. *Oncotarget*, 2016. **7**(12): p. 14029-47.
177. Lee, H.Y., et al., *INHBA overexpression indicates poor prognosis in urothelial carcinoma of urinary bladder and upper tract*. *J Surg Oncol*, 2015. **111**(4): p. 414-22.
178. Togashi, Y., et al., *Activin signal promotes cancer progression and is involved in cachexia in a subset of pancreatic cancer*. *Cancer Lett*, 2015. **356**(2 Pt B): p. 819-27.
179. Okano, M., et al., *Significance of INHBA expression in human colorectal cancer*. *Oncol Rep*, 2013. **30**(6): p. 2903-8.
180. Spivey, K.A. and J. Banyard, *A prognostic gene signature in advanced ovarian cancer reveals a microfibril-associated protein (MAGP2) as a promoter of tumor cell survival and angiogenesis*. *Cell Adh Migr*, 2010. **4**(2): p. 169-71.
181. Milwid, J.M., et al., *Enriched protein screening of human bone marrow mesenchymal stromal cell secretions reveals MFAP5 and PENK as novel IL-10 modulators*. *Mol Ther*, 2014. **22**(5): p. 999-1007.
182. Nakamura, R., et al., *Expression and regulatory effects on cancer cell behavior of NELL1 and NELL2 in human renal cell carcinoma*. *Cancer Sci*, 2015. **106**(5): p. 656-64.
183. Kim, D.H., et al., *The E2F1 oncogene transcriptionally regulates NELL2 in cancer cells*. *DNA Cell Biol*, 2013. **32**(9): p. 517-23.
184. Li, Y., J. Zhang, and S. Hong, *ANO1 as a marker of oral squamous cell carcinoma and silencing ANO1 suppresses migration of human SCC-25 cells*. *Med Oral Patol Oral Cir Bucal*, 2014. **19**(4): p. e313-9.
185. Reddy, R.B., et al., *Meta-Analyses of Microarray Datasets Identifies ANO1 and FADD as Prognostic Markers of Head and Neck Cancer*. *PLoS One*, 2016. **11**(1): p. e0147409.
186. Ayoub, C., et al., *ANO1 amplification and expression in HNSCC with a high propensity for future distant metastasis and its functions in HNSCC cell lines*. *Br J Cancer*, 2010. **103**(5): p. 715-26.
187. Rodrigo, J.P., et al., *Clinical significance of Anoctamin-1 gene at 11q13 in the development and progression of head and neck squamous cell carcinomas*. *Sci Rep*, 2015. **5**: p. 15698.
188. Taoudi Benchekroun, M., et al., *Epidermal growth factor receptor expression and gene copy number in the risk of oral cancer*. *Cancer Prev Res (Phila)*, 2010. **3**(7): p. 800-9.
189. Kong, D.C., et al., *The effect of epiregulin on epidermal growth factor receptor expression and proliferation of oral squamous cell carcinoma cell lines*. *Cancer Cell Int*, 2014. **14**: p. 65.
190. Ries, J., et al., *The relevance of EGFR overexpression for the prediction of the malignant transformation of oral leukoplakia*. *Oncol Rep*, 2013. **30**(3): p. 1149-56.
191. Cho, Y.A., et al., *Relationship between the expressions of PD-L1 and tumor-infiltrating lymphocytes in oral squamous cell carcinoma*. *Oral Oncol*, 2011. **47**(12): p. 1148-53.
192. Lyford-Pike, S., et al., *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*, 2013. **73**(6): p. 1733-41.
193. Ritprajak, P. and M. Azuma, *Intrinsic and extrinsic control of expression of the immunoregulatory molecule PD-L1 in epithelial cells and squamous cell carcinoma*. *Oral Oncol*, 2015. **51**(3): p. 221-8.
194. Zandberg, D.P. and S.E. Strome, *The role of the PD-L1:PD-1 pathway in squamous cell carcinoma of the head and neck*. *Oral Oncol*, 2014. **50**(7): p. 627-32.
195. Jin, Y., et al., *Up-regulation of ECT2 is associated with poor prognosis in gastric cancer patients*. *Int J Clin Exp Pathol*, 2014. **7**(12): p. 8724-31.

196. Murata, Y., et al., *ECT2 amplification and overexpression as a new prognostic biomarker for early-stage lung adenocarcinoma*. *Cancer Sci*, 2014. **105**(4): p. 490-7.
197. Huff, L.P., et al., *The Role of Ect2 Nuclear RhoGEF Activity in Ovarian Cancer Cell Transformation*. *Genes Cancer*, 2013. **4**(11-12): p. 460-75.
198. Wang, F.W., et al., *Ablation of EIF5A2 induces tumor vasculature remodeling and improves tumor response to chemotherapy via regulation of matrix metalloproteinase 2 expression*. *Oncotarget*, 2014. **5**(16): p. 6716-33.
199. Meng, Q.B., et al., *Overexpression of eukaryotic translation initiation factor 5A2 (EIF5A2) correlates with cell aggressiveness and poor survival in gastric cancer*. *PLoS One*, 2015. **10**(3): p. e0119229.
200. Wei, J.H., et al., *EIF5A2 predicts outcome in localised invasive bladder cancer and promotes bladder cancer cell aggressiveness in vitro and in vivo*. *Br J Cancer*, 2014. **110**(7): p. 1767-77.
201. Hur, H., et al., *HOXC9 Induces Phenotypic Switching between Proliferation and Invasion in Breast Cancer Cells*. *J Cancer*, 2016. **7**(7): p. 768-73.
202. Shah, N. and S. Sukumar, *The Hox genes and their roles in oncogenesis*. *Nat Rev Cancer*, 2010. **10**(5): p. 361-71.
203. Gu, Z.D., et al., *[Expression of 39 HOX genes in esophageal cancer cell lines]*. *Zhonghua Wei Chang Wai Ke Za Zhi*, 2007. **10**(4): p. 365-7.
204. Hur, H., et al., *Analysis of HOX gene expression patterns in human breast cancer*. *Mol Biotechnol*, 2014. **56**(1): p. 64-71.
205. Bhatlekar, S., J.Z. Fields, and B.M. Boman, *HOX genes and their role in the development of human cancers*. *J Mol Med (Berl)*, 2014. **92**(8): p. 811-23.
206. Wang, X., et al., *HOXC9 directly regulates distinct sets of genes to coordinate diverse cellular processes during neuronal differentiation*. *BMC Genomics*, 2013. **14**: p. 830.
207. Mao, L., et al., *HOXC9 links cell-cycle exit and neuronal differentiation and is a prognostic marker in neuroblastoma*. *Cancer Res*, 2011. **71**(12): p. 4314-24.
208. Hassan, N.M., et al., *Aberrant expression of HOX genes in oral dysplasia and squamous cell carcinoma tissues*. *Oncol Res*, 2006. **16**(5): p. 217-24.
209. Jave-Suarez, L.F., et al., *HOXC13 is involved in the regulation of human hair keratin gene expression*. *J Biol Chem*, 2002. **277**(5): p. 3718-26.
210. Awgulewitsch, A., *Hox in hair growth and development*. *Naturwissenschaften*, 2003. **90**(5): p. 193-211.
211. Cantile, M., et al., *Increased HOX C13 expression in metastatic melanoma progression*. *J Transl Med*, 2012. **10**: p. 91.
212. Kasiri, S., et al., *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*, 2013. **3**(10): p. 3260-3269.
213. Oda, Y., et al., *Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation*. *J Cell Biol*, 2006. **172**(3): p. 383-93.
214. Lopez-Serra, P., et al., *A DERL3-associated defect in the degradation of SLC2A1 mediates the Warburg effect*. *Nat Commun*, 2014. **5**: p. 3608.
215. Freudlsperger, C., et al., *EGFR-PI3K-AKT-mTOR signaling in head and neck squamous cell carcinomas: attractive targets for molecular-oriented therapy*. *Expert Opin Ther Targets*, 2011. **15**(1): p. 63-74.
216. Molinolo, A.A., et al., *Dysregulated molecular networks in head and neck carcinogenesis*. *Oral Oncol*, 2009. **45**(4-5): p. 324-34.
217. Bates, T., et al., *Changes in Epidermal Growth Factor Receptor Gene Copy Number during Oral Carcinogenesis*. *Cancer Epidemiol Biomarkers Prev*, 2016. **25**(6): p. 927-35.

- 218. Polakis, P., *Wnt signaling and cancer*. Genes Dev, 2000. **14**(15): p. 1837-51.
- 219. Raica, M., A.M. Cimpean, and D. Ribatti, *Angiogenesis in pre-malignant conditions*. Eur J Cancer, 2009. **45**(11): p. 1924-34.
- 220. Menakuru, S.R., et al., *Angiogenesis in pre-malignant conditions*. Br J Cancer, 2008. **99**(12): p. 1961-6.
- 221. Gupta, G.P. and J. Massague, *Cancer metastasis: building a framework*. Cell, 2006. **127**(4): p. 679-95.
- 222. Ilie, M., et al., *High levels of carbonic anhydrase IX in tumour tissue and plasma are biomarkers of poor prognostic in patients with non-small cell lung cancer*. Br J Cancer, 2010. **102**(11): p. 1627-35.
- 223. Choi, S.W., et al., *Expression of carbonic anhydrase IX is associated with postoperative recurrence and poor prognosis in surgically treated oral squamous cell carcinoma*. Hum Pathol, 2008. **39**(9): p. 1317-22.
- 224. Weber, G.F., G.S. Lett, and N.C. Haubein, *Osteopontin is a marker for cancer aggressiveness and patient survival*. Br J Cancer, 2010. **103**(6): p. 861-9.
- 225. Okamoto, O.K., et al., *Expression of HOXC9 and E2F2 are up-regulated in CD133(+) cells isolated from human astrocytomas and associate with transformation of human astrocytes*. Biochim Biophys Acta, 2007. **1769**(7-8): p. 437-42.
- 226. Sajnani, M.R., et al., *Identification of novel transcripts deregulated in buccal cancer by RNA-seq*. Gene, 2012. **507**(2): p. 152-8.

9. Publications

Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse

Srikant Ambatipudi^{1,2,3}, Priyanka G. Bhosale^{1,3}, Emma Heath², Manishkumar Pandey¹, Gaurav Kumar¹, Shubhada Kane³, Asawari Patil³, Girish B. Maru¹, Rajiv S. Desai⁴, Fiona M. Watt², Manoj B. Mahimkar^{1*}

1 Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, India, **2** King's College London Centre for Stem Cells and Regenerative Medicine, London, United Kingdom, **3** Department of Pathology, Tata Memorial Hospital, Tata Memorial Centre, Parel, Mumbai, India, **4** Department of Oral Pathology and Microbiology, Nair Hospital Dental College, Mumbai, India

Abstract

Background: Keratins are structural marker proteins with tissue specific expression; however, recent reports indicate their involvement in cancer progression. Previous study from our lab revealed deregulation of many genes related to structural molecular integrity including KRT76. Here we evaluate the role of KRT76 downregulation in oral precancer and cancer development.

Methods: We evaluated KRT76 expression by qRT-PCR in normal and tumor tissues of the oral cavity. We also analyzed K76 expression by immunohistochemistry in normal, oral precancerous lesion (OPL), oral squamous cell carcinoma (OSCC) and in hamster model of oral carcinogenesis. Further, functional implication of KRT76 loss was confirmed using KRT76-knockout (KO) mice.

Results: We observed a strong association of reduced K76 expression with increased risk of OPL and OSCC development. The buccal epithelium of DMBA treated hamsters showed a similar trend. Oral cavity of KRT76-KO mice showed preneoplastic changes in the gingivobuccal epithelium while no pathological changes were observed in KRT76 negative tissues such as tongue.

Conclusion: The present study demonstrates loss of KRT76 in oral carcinogenesis. The KRT76-KO mice data underlines the potential of KRT76 being an early event although this loss is not sufficient to drive the development of oral cancers. Thus, future studies to investigate the contributing role of KRT76 in light of other tumor driving events are warranted.

Citation: Ambatipudi S, Bhosale PG, Heath E, Pandey M, Kumar G, et al. (2013) Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse. PLoS ONE 8(7): e70688. doi:10.1371/journal.pone.0070688

Editor: Takashi Tokino, Sapporo Medical University, Japan

Received: March 20, 2013; **Accepted:** June 21, 2013; **Published:** July 30, 2013

Copyright: © 2013 Ambatipudi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors thankfully acknowledge Lady Tata Memorial Trust, Mumbai for generously funding MBM for the gene expression profiling study; Advanced Centre for Treatment Research and Education in Cancer (ACTREC) for core funds to MBM and GBM; and the UK Medical Research Council and the Wellcome Trust for the financial support to FMW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: All the authors have approved the submission and have declared no conflict of interest.

* E-mail: mmahimkar@actrec.gov.in

† Current address: Epigenetics Group, Section of Mechanisms of Carcinogenesis, International Agency for Research on Cancer, Lyon, France

‡ These authors contributed equally to this work.

Introduction

Keratins are filament forming proteins of epithelial cells and are essential for normal tissue structure and function [1]. In contrast to actin filaments and microtubules, keratins are encoded by a large family of genes clustered at two divergent chromosomal sites: 17q21.2 (type I keratins, except K18) and 12q13.13 (type II keratins, including K18). These are also expressed in tissue and differentiation state-specific manner and play an important role in protecting epithelial cells from mechanical and non-mechanical stress and injury [2,3,4,5].

Epithelial tumors continue to express keratins that are characteristic of their site of origin and therefore keratins are extensively used as immunohistochemical markers in diagnostic tumor pathology [3,4]. Accumulating evidence points to the importance of keratins as prognostic markers and, more interestingly, as active regulators of epithelial tumorigenesis and treatment

responsiveness [3]. Previous studies have reported alterations in keratin expression during oral carcinogenesis [6,7,8,9]. Further, many keratins are recognized as independent markers of prognosis in OSCC [10,11].

Within the oral cavity there is a complex pattern of keratin expression, reflecting both the type of epithelium and stage of differentiation specific expression. The basal proliferative layer of all oral epithelia expresses K5/K14 and K19. The suprabasal, differentiating layers of keratinized (cornified) epithelia express K1 and K10, while the differentiating layers of non-keratinized epithelia such as buccal mucosa and esophagus synthesize predominantly K4 and K13. Suprabasal epithelial cells of the hard palate and gingiva express K6, K16, and K76 [5,12,13,14,15]. Previous studies have reported altered terminal differentiation and keratin expression patterns in oral tumors, such as downregulation of K4, K5, K13 and K19 [11,16,17,18,19,20,21,22,23,24,25,26,27,28]. Conversely, increased expression

of K8/K18, K17 and K14 is reported in oral tumor tissues compared to the normal counterparts [6,7,8,9,10,17,18,23,29]. Various studies using *in-vitro* system have elucidated mechanistic role of keratins (K8/18, K19) in tumor invasion and metastasis [30,31,32]. However, *in-vitro* data may not fully reflect the *in-vivo* condition [33]. Interestingly, alterations of keratin expression pattern marks the common signature in human oral cancers and experimental oral tumors developed in animal models [34,35]. Hence, we selected *in-vivo* model systems: the hamster model to demonstrate K76 downregulation during sequential progression of oral cancer, and the KO mice model to evaluate the effect of *KRT76* loss.

Gene expression analysis from our laboratory has revealed downregulation of *KRT76* in tumors of the oral cavity [26]. *KRT76*, a type II epithelial keratin (previously designated as K2p), is specifically expressed in the suprabasal cell layers of oral masticatory epithelium (the slightly orthokeratinized stratified squamous epithelium lining the gingiva and the hard palate) [13]. We now present data indicating that *KRT76* is downregulated prior to tumor development and its potential association with hyperproliferation in the formation of preneoplastic lesions.

Materials and Methods

Human Tissue Specimen Collection

The Institutional Review Board and the Local Ethics Committee of Tata Memorial Hospital (TMH) and Nair Hospital Dental College, approved the study. Written informed consents were

obtained from all the study participants. Treatment naive neoplastic frozen tissues (n = 57) and paraffin embedded tissue blocks (n = 102) of different cohort of patients with gingivobuccal cancer (GBC) were obtained from the ICMR National Tumor Tissue Repository and Department of Pathology TMH, Mumbai respectively. Precancerous lesions (incident leukoplakia cases which are histopathologically hyperplastic lesions with focal mild to moderate dysplasia n = 61), independent normal tissues (n = 35), and inflamed tissues not associated with oral malignancy or pre-malignant conditions (n = 7) were collected from the Department of Oral Pathology, Nair Hospital Dental College, Mumbai; all these tissues were from gingivobuccal region. Tumor tissues with more than 70% tumor content were subjected to RNA extraction.

Animal Models

The study on hamsters was conducted after approval from the Institutional Animal Ethics Committee (IAEC) of ACTREC, endorsed by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India guidelines. Inbred male Syrian hamsters (6–8 weeks old; Animal house, ACTREC, India) were randomized (10 animals per group) and maintained under standard conditions: $22 \pm 2^\circ\text{C}$, 45% $\pm 10\%$ relative humidity, and 12-h light/dark cycle (7:00 to 19:00 light; 19:00 to 7:00 dark). The animals received an autoclaved standard pellet diet and plain drinking water *ad libitum*. Hamsters (3–5) were housed in the polypropylene cages provided with autoclaved rice husk bedding material available locally. The hamsters were topically treated with 7,12-dimethylbenz[α]anthracene (DMBA) (0.5%) in corn oil using a Gilson pipette (80 μl \approx 0.4 mg) on their right buccal pouch, thrice a week for 16 weeks. The ‘corn oil’ was used for the treatment in vehicle control group. Animals in all groups were observed for apparent signs of toxicity such as weight loss or mortality during the entire study period. Following 1, 2, 4, 6, 8, 10, 12 and 16 weeks of DMBA applications, hamsters were euthanized (by CO₂ chamber) 24 h after the last DMBA dose. Their buccal pouches were excised and fixed in 10% buffered formalin [36,37].

The animal research ethical review committees of the Cancer Research UK Cambridge Research Institute and Cambridge University approved all the studies involving mice. *KRT76*-KO mice were obtained from the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/mouseportal/search?query=KRT76>), and were maintained under the terms of a UK Government Home Office license 80/2378 (license holder Fiona M. Watt).

RNA Isolation from Tissues

RNA was isolated from human tumor and normal tissues using the RNeasy mini kit (Qiagen, Germany) according to the manufacturer’s protocol. Briefly, 15–20 mg tissue was pulverized by grinding with liquid nitrogen, followed by addition of RLT buffer with β -mercaptoethanol (Sigma-Aldrich, USA). The homogenate was processed for column purification and isolation of RNA. DNA contamination was avoided by treating the column with RNase free DNase I (Ambion, USA). The quantity and quality of RNA was determined using Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) and RNA 6000 Nano LabChip Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, CA) respectively.

Quantitative Reverse Transcriptase-Polymerase Chain Reaction (qRT-PCR)

For complementary DNA (cDNA) synthesis, 1.5 μg of total RNA was reverse-transcribed with the High-Capacity cDNA

Table 1. Demographic and clinicopathological characteristic of the study group.

Characteristics	qRT-PCR OSCC (n = 57) [†]	IHC(n = 163) [†]	
		OSCC(n = 102)	OPL(n = 61)
Gender			
Males	40 (70%)	80 (78.4%)	55 (90.2%)
Females	17 (30%)	22 (21.6%)	6 (9.8%)
Age			
Median (IQR) [#]	52 (43.5–57.5)	52 (41.7–64)	45(34.5–56.6)
Habit profile			
Exclusive Chewers	46 (80.7%)	34 (59.7%)	19 (32.7%)
Exclusive Smokers	3 (5.3%)	4 (7%)	12 (20.7%)
Chewing and Smoking	8(14%)	19 (33.3%)	27 (46.6%)
Grade			
Well	2 (3.5%)	12 (11.7%)	–
Moderate	39 (68.4%)	65 (63.7%)	–
Poor	16(28.1%)	25 (24.6%)	–
Nodal involvement			
Negative (N0)	29 (50.9%)	49 (48.0%)	–
Positive (N+)	28 (49.1%)	53 (52.0%)	–
Stage (pTNM)			
I & II	3 (5.3%)	13 (12.74%)	–
III & IV	54 (94.7%)	89 (87.26%)	–

[†]Shown is the number of cases, except for Age,

[#]IQR: Interquartile range.

doi:10.1371/journal.pone.0070688.t001

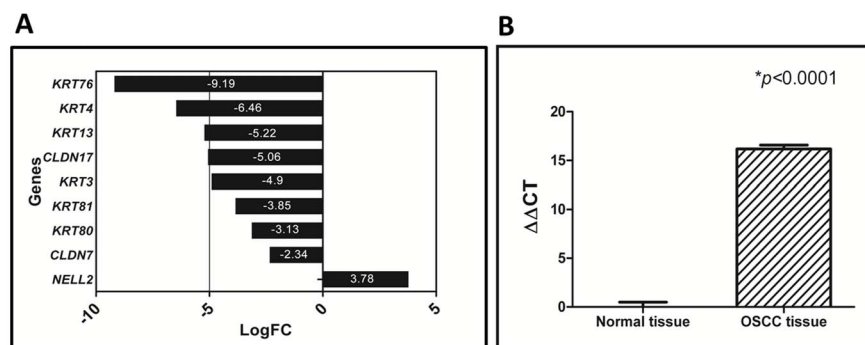


Figure 1. Downregulation of *KRT76* in GBCs. **A:** Data analyzed using GEO accession: GSE23558 demonstrate genes associated with structural molecular activity in GBCs. **B:** qRT-PCR analysis showed more than 15 fold downregulation of *KRT76* expression in tumors compared to normal oral tissue.

doi:10.1371/journal.pone.0070688.g001

Reverse Transcription Kit (Applied Biosystems, USA) following the manufacturer's protocol. Twenty ng of cDNA were used for TaqMan qRT-PCR analysis and experiments were performed in duplicate (*KRT76* Assay Id: Hs00210581_m1, 18S RNA Assay Id: Hs99999901). Results were analyzed using SDS 2.3 and RQ manager software (Applied Biosystems). The relative expression of *KRT76* messenger RNA (mRNA) was determined using 18S ribosomal RNA as an endogenous control. These were compared between GBC cancers and unrelated normal tissues from the same site. The expression of *KRT76* in each sample was analyzed using the comparative CT method (also known as the $2^{-\Delta\Delta CT}$ method) where $\Delta\Delta CT = [CT \text{ gene of interest} - CT \text{ internal control (18S)}]$ of test sample $- [CT \text{ gene of interest} - CT \text{ internal control (18S)}]$ of reference sample. Fold change values for qRT-PCR data were calculated as $2^{-\Delta\Delta CT}$ [38].

Immunostaining of K76 in Human Oral Tissues

Formalin-fixed, paraffin-embedded GBC tissues (n=102), OPLs (n=61) and normal oral tissues (n=21) were used for immunohistochemical (IHC) analysis. Five micron tissue sections were deparaffinized with xylene, rehydrated with sequential ethanol washes (100%, 90% and 70%). To quench the endogenous peroxidase activity, sections were incubated with 3% hydrogen peroxide in methanol for 30 min in dark. After heat based antigen retrieval with sodium citrate buffer (pH=5.8), sections were incubated with normal horse serum. The sections were incubated overnight with rabbit polyclonal anti-human K76 antibody (1:225, HPA019696, Sigma-Aldrich) at 4°C. For negative or isotype control, the primary antibody was replaced with rabbit serum used at respective antibody concentration. Sections were then incubated with biotinylated universal secondary antibody solution for 30 min followed by incubation with VectastainVR elite ABC reagent for the same time. The immunoreaction in tissue sections was visualized using 3,3'-diaminobenzidine tetrahydrochloridehydrate (Sigma-Aldrich). The slides were finally counterstained with hematoxylin and examined under microscope.

For immunofluorescence, deparaffinization and antigen retrieval steps were similar to those for IHC. Tissues were fixed in cold methanol for 10 min followed by blocking with 5% normal goat serum, 0.3% (v/v) Triton X-100 in PBS for 1 hr at room temperature. Tissues were next incubated with K76 antibody at a dilution of 1:250 overnight at 4°C, followed by incubation with an Alexa Fluor 488 anti-rabbit antibody (Life technologies, USA) at 1:200 dilution, for 1 hr at room temperature. Cells were

counterstained with DAPI and viewed under a fluorescence microscope (Zeiss; LSM-510 Meta Germany).

Immunostaining of K76 in Animal Models

Formalin fixed hamster buccal pouch tissues were used from the following experimental groups for IHC analysis: 1) Control group: 1st, 2nd, 4th, 6th, 8th, 10th, 12th and 16th week hamsters buccal pouch topically treated with vehicle (no DMBA); 2) DMBA treated group: 1st, 2nd, 4th, 6th, 8th, 10th, 12th and 16th week hamsters buccal pouch topically treated with DMBA. Formalin fixed tissues from *KRT76*-Wild type (WT) and *KRT76*-KO mice were used for immunostaining and histopathological analysis. For experimental models, the IHC staining procedure was similar to that described earlier with minor changes in blocking, which was performed with 3% BSA and 2% goat serum; while secondary antibody was biotin conjugated anti-rabbit secondary raised in goat (Santa Cruz Biotechnology, USA).

Immunohistochemical Assessment and Scoring

For assessment of K76 protein expression, the cytoplasmic staining intensity was categorized as 0 (absence of staining in any cell), +1 (weak staining in less than 10% of cells), +2 (moderate staining and/or 10 to 50% of positive cells), or +3 (strong staining in more than 50% cells) by pathologist (AP) (Figure S1). For further statistical analysis, the stained tissues were categorized in two groups: 0 and +1 as mild to no expression, while +2 and +3 as moderate to strong expression.

Statistical Analysis

All statistical analyses were performed using IBM SPSS version 21. The Mann Whitney test was performed to analyze the difference between ΔCT values of tumor and normal samples obtained by qRT-PCR. The Chi-square test was used to determine the correlation between expression levels of K76 protein and tissue type, as well as clinicopathological characteristics. Polytomous logistic regression was used to evaluate the relationship of protein expression scores to the risk of OPL and OSCC development, with normal tissue as a reference; odds ratio (OR) were computed by adjusting for age and gender [22,39]. Disease-specific survival (DSS) was calculated as the time from surgical diagnosis to the date of death due to cancer or to the last clinical follow-up prior to death. DSS was examined visually with Kaplan-Meier curves and analyzed by log rank tests. All p-values <0.05 were considered statistically significant.

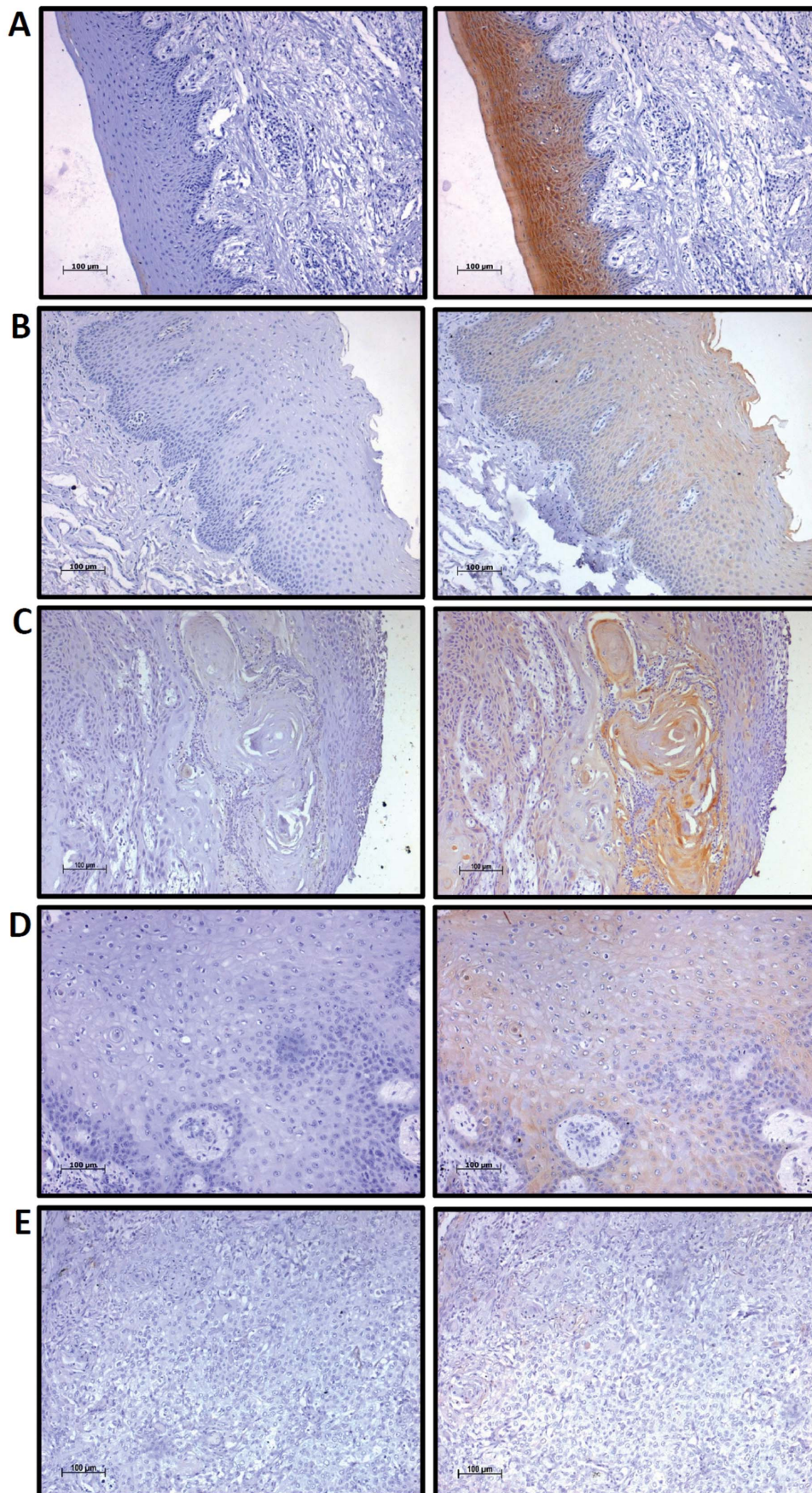


Figure 2. Immunohistochemical analysis of K76 expression in normal buccal mucosa, oral premalignant lesions and oral cancers. Representative IHC staining on **A:** Normal buccal mucosa, **B:** Oral Premalignant Lesions and OSCC (**C:** well differentiated, **D:** Moderately differentiated, **E:** poorly differentiated), with respective isotype control. Magnification 100X (Scale: 100 μ m).
doi:10.1371/journal.pone.0070688.g002

Results

Patient Characteristics

The clinicopathological and demographic characteristics of all OPLs and tumor samples are summarized in Table 1. The patients in this study cohort were predominantly male tobacco habitués and tobacco chewing was the most prevalent habit. Most of the tumor samples were of moderate or poor grade, and mainly of pTNM stages III or IV. Approximately 50% of the cases showed lymph node invasion. Majority of OPLs had mild to severe hyperplasia and few showed presence of focal mild to moderate dysplasia.

Validation of Microarray Results by qRT-PCR

Microarray analysis of 27 GBC cases showed a significant downregulation of *KRT76*, as reported previously [26]. We observed downregulation of many genes associated with structural molecule activity Gene Ontology: 0005198 of which *KRT76* showed the highest fold change (Figure 1A). The Oncomine data source illustrated two more studies reporting consistent downregulation of *KRT76* in OSCC (Figure S2) [40,41,42]. To confirm the findings of the microarray analysis, we performed qRT-PCR using

primers specific for *KRT76* in 57 OSCC and 14 normal tissues. qRT-PCR analysis revealed significant downregulation of *KRT76* RNA in tumor samples compared to normal samples (Figure 1B).

Sequential Downregulation of K76 in Oral Carcinogenesis

K76 expression was analyzed in 184 oral tissues by immunohistochemistry (Figure 2). Normal gingivobuccal tissues expressed higher levels of K76 protein compared to OPL and invasive OSCC. Distribution of K76 expression was confirmed by immunofluorescence as illustrated in Figure 3. Normal oral epithelium showed K76 expression confined to the suprabasal, differentiating cell layers while, there was a gradual overall loss of K76 expression in OPLs and tumors. The frequency of K76 positive staining significantly decreased across the transition from normal tissue (100% positive) to OPL (44%) to oral tumor (35%) (Figure 4A).

To examine whether *KRT76* downregulation was associated with benign epithelial hyperproliferation (injured normal tissue without any association with oral preinvasive and invasive lesions), we performed IHC on inflamed buccal mucosa ($n = 7$). Even though these epithelia histologically appeared hyperproliferative, K76 staining was consistent with that seen in normal buccal

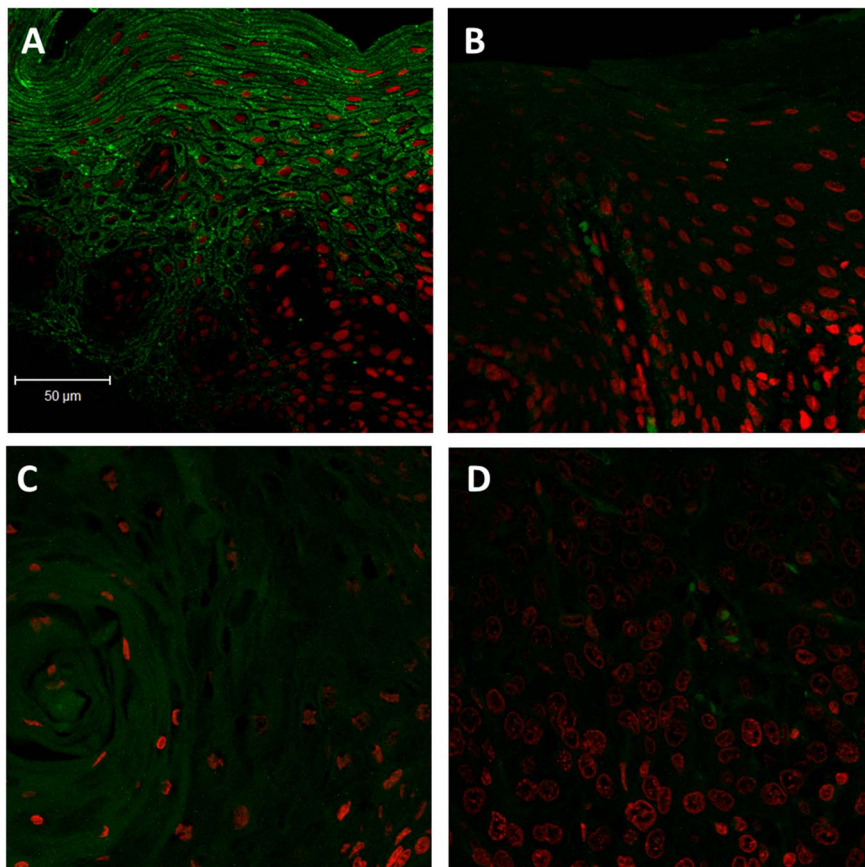


Figure 3. Immunofluorescence staining of K76 on human oral tissues. Representative Immunofluorescent staining of **A:** Normal oral tissue, **B:** OPL, **C:** Well differentiated tumor, **D:** Poorly differentiated tumor. K76 (Stained green, Alexa fluor 488), Nuclei stained with DAPI (pseudo red). Magnification 200X (Scale: 50 μ m).
doi:10.1371/journal.pone.0070688.g003

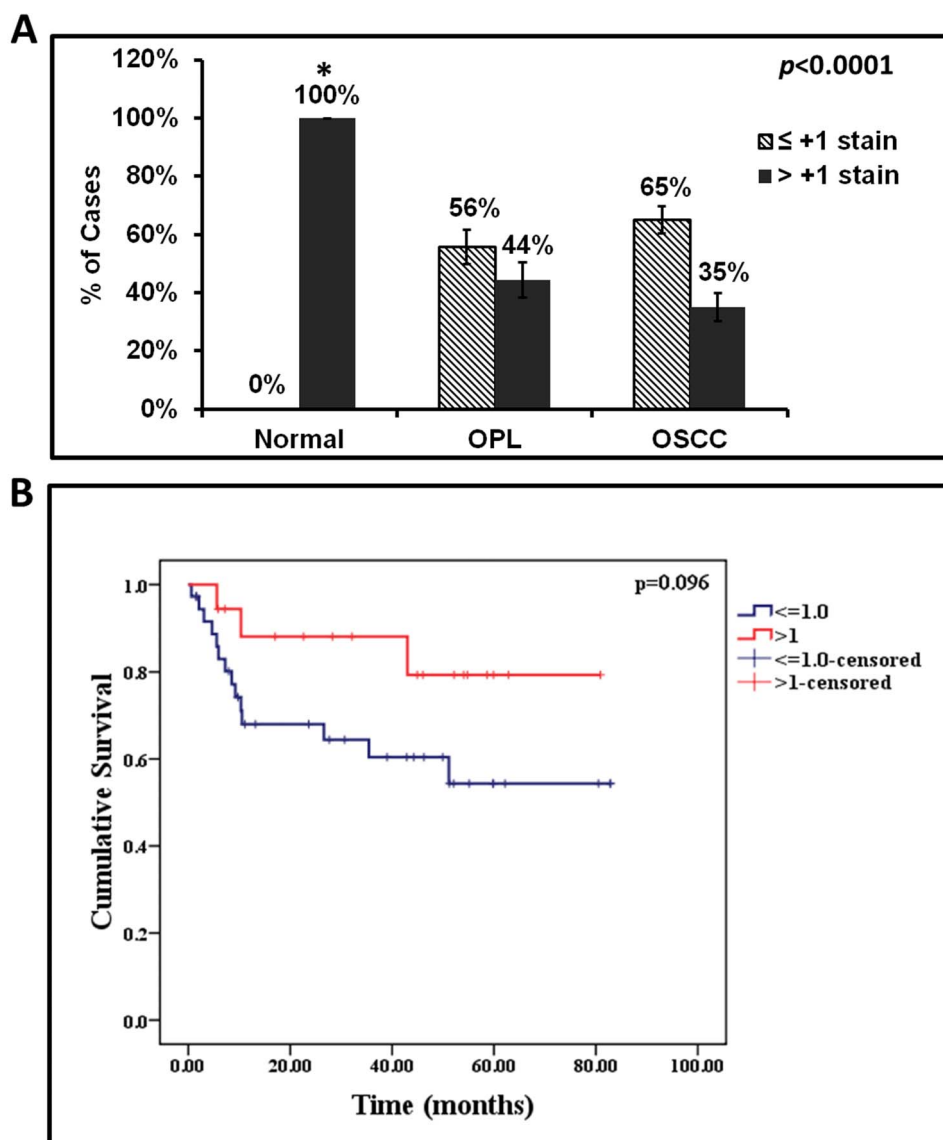


Figure 4. Correlation between loss of K76 expression with oral cancer development and patient survival. **A:** Significant downregulation of K76 was observed in OSCC and OPL compared to normal. **B:** Kaplan–Meier plot for DSS of gingivobuccal cancer patients with respect to K76 IHC staining intensity. Footnote: *All normals showed more than +2 grade stain. doi:10.1371/journal.pone.0070688.g004

epithelium (Figure S3). These results indicate that downregulation of *KRT76* expression is not associated with injury related proliferation and acute inflammation.

Correlation of K76 Expression with Clinicopathological Parameters

Statistical analysis to determine the association of K76 expression and different clinical parameters, such as node, stage, grade, habit profile and outcome (recurrence and survival) was performed. Reduced expression of K76 showed a very weak

Table 2. The effect of K76 expression loss with development of oral lesions.

K76 staining	Normal (n = 21)	OPL (n = 61)	OR	95% CI	p value	OSCC (n = 102)	OR	95% CI	p value
High(>1)	21	27	1	3.4–216.7	0.002	36	1	5.1–307	<0.0001
low(≤1)	0	34	27			66	40		

Polytomous logistic regression performed using normal as reference group indicated significant increase in risk of developing OPL and OSCC with decrease in staining. doi:10.1371/journal.pone.0070688.t002

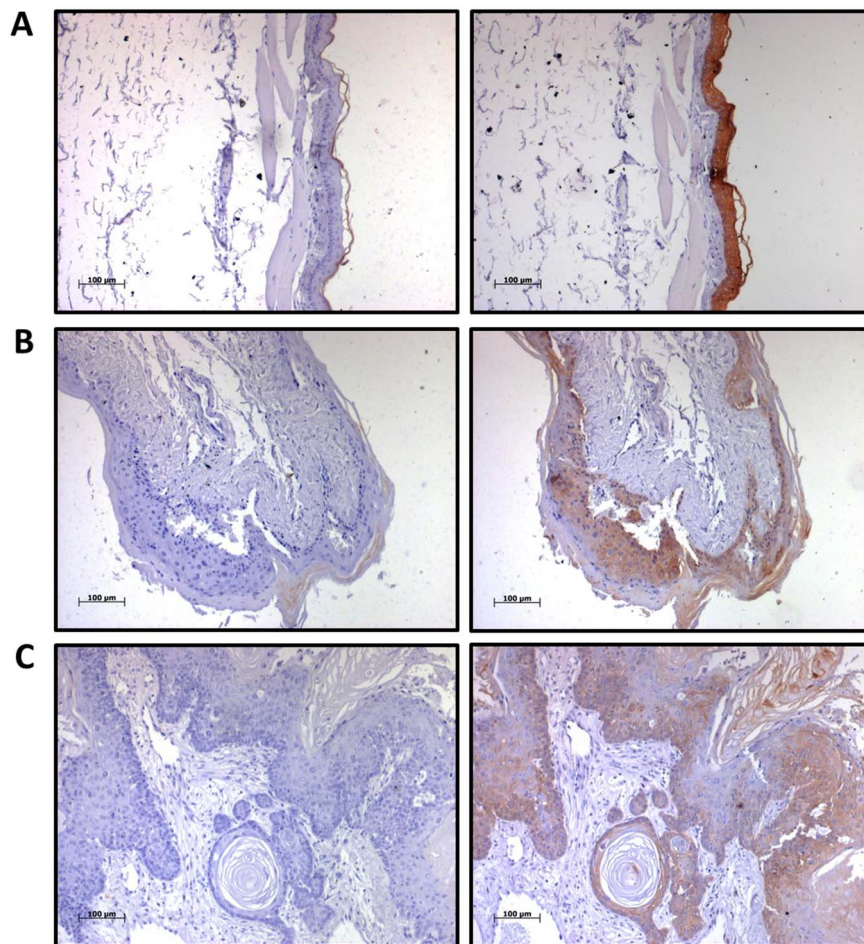


Figure 5. Expression of K76 in hamster model of oral carcinogenesis. IHC staining for K76 expression in hamster oral epithelium of **A:** Control group **B:** Hyperplastic lesion, **C:** Tumor with respective isotype controls. Magnification 100X (Scale: 100 µm). doi:10.1371/journal.pone.0070688.g005

association with survival ($p = 0.096$) (Figure 4B), whereas other parameters analyzed did not show any association. Polytomous Logistic regression with normal as the reference group showed a significant correlation of K76 downregulation with risk of developing OPL ($p = 0.002$) and OSCC ($p \leq 0.0001$) (Table 2).

Loss of K76 Expression in an Experimental Model of Oral Carcinogenesis

K76 expression was analyzed by IHC in the buccal epithelium of DMBA treated hamsters (group details described in methods). Interestingly gradual decrease in staining intensity was observed with disease progression in hamster buccal epithelium (Figure S4). Irrespective of duration of treatment, control group showed higher levels of K76, while reduced expression was observed in premalignant lesions and oral tumors, which was similar to that seen in human hyperplastic lesions and OSCC (Figure 5).

Mice Lacking *KRT76* Develop Hyperplastic Oral Lesions

To determine whether loss of *KRT76* is sufficient to induce premalignant lesions in the oral cavity, we examined the oral epithelia of *KRT76*-KO and *KRT76*-WT mice. Immunohistochemical analysis showed specific K76 staining in buccal epithelium of WT mice, whereas no staining was observed in *KRT76*-KO buccal epithelium, confirming specificity of K76

antibody (Figure 6 A, B). Histological examination of the buccal mucosa of *KRT76*-KO mice showed development of hyperplastic lesions along with increased keratinization across the epithelium, which was not observed in *KRT76*-WT mice (Figure 6 C, D). In contrast, the epithelium of the dorsal tongue, which is normally *KRT76*-negative, exhibited normal homeostasis in *KRT76*-KO mice indicating that *KRT76* loss associated abnormalities are highly sub-site specific in oral cavity (Figure S5). However, none of the *KRT76*-KO mice in the entire life span developed spontaneous oral tumors.

Discussion

Deregulated keratin expression is associated with impaired epithelial differentiation and organization during OSCC progression [4,15,21,24,41,43,44]. Our microarray based gene expression profile of 27 advanced stage gingivobuccal cancers previously revealed deregulation of several keratins, namely *KRT4*, *KRT13*, *KRT19*, *KRT76*, which are normally expressed in the oral cavity. *KRT76* was found to be the topmost downregulated gene amongst all differentially expressed genes [26]. Gene expression profiles of oral cancer obtained by other groups have also shown consistent downregulation of *KRT76* [21,40,41]. We now report, for the first time, differential expression of *KRT76* in human and hamster oral

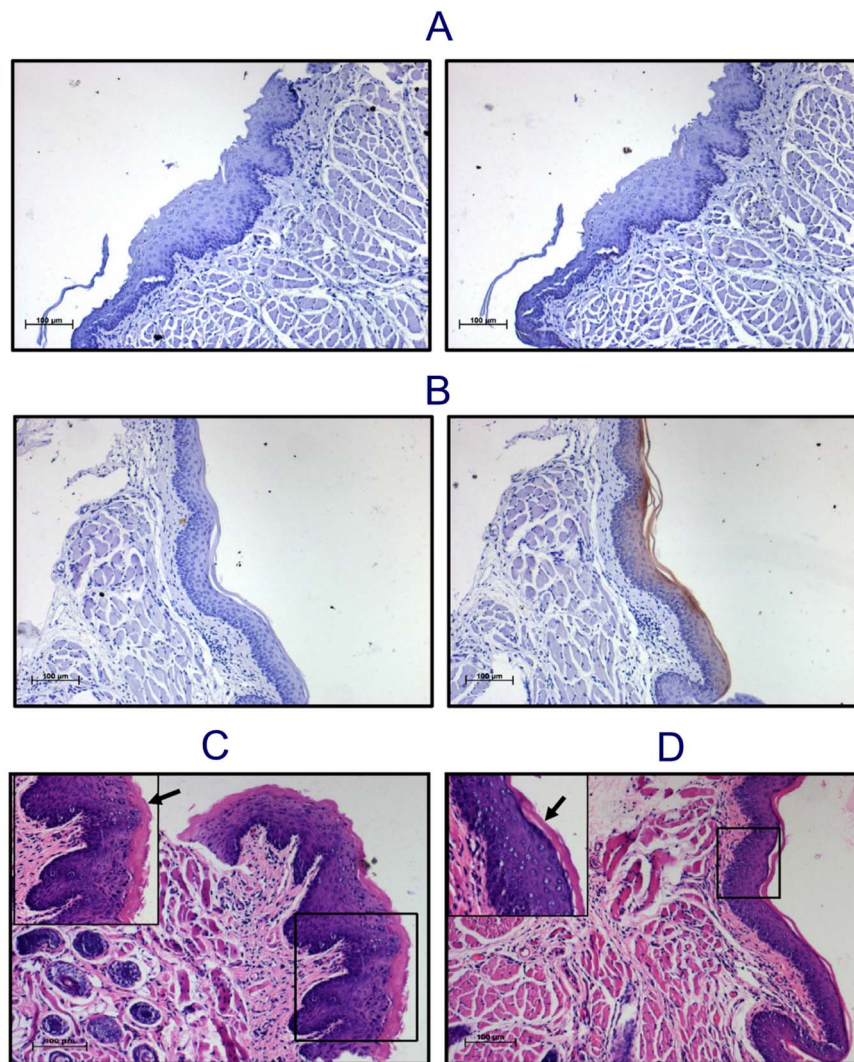


Figure 6. *KRT76*-KO mice show hyperplastic lesions in oral epithelium. K76 antibody specificity was determined by IHC on Oral epithelium of *KRT76*-KO mice which did not show any staining (A), whereas wild type mice of same strain showed moderate staining (B), with respective isotype control (left panel A & B). Histological observation of H & E stained buccal epithelium demonstrated hyperplastic changes and increased keratinization in KO (C), compared to WT (D). Magnification 100X (Scale: 100 μm); selected area under 200×magnification. doi:10.1371/journal.pone.0070688.g006

precancerous and cancerous lesions, and show that loss of *KRT76* is sufficient to cause hyperplasia in the oral cavity of the mice.

We validated our previous microarray findings in an independent patient cohort by qRT-PCR and IHC; both these techniques showed reduced expression of *KRT76*. While previous reports have demonstrated changes in keratin gene expression associated with severe dysplasia and poorly differentiated SCC, reflecting gross changes in epithelial differentiation and maturation [43,44], our studies are the first to indicate that loss of a specific keratin is sufficient to initiate preneoplastic changes. We did not find association of K76 downregulation with clinicopathological parameters such as node, grade, clinical outcome; nor with benign inflammation-associated hyperproliferation. Although, the fact that K76 downregulation is observed in leukoplakia, a preinvasive oral lesion and is sustained during the development of frank malignancy, indicates its association with the early stages of oral carcinogenesis.

Interestingly, we observed gradual decrease in K76 expression during the sequential process of tumor development in DMBA

treated buccal epithelium of hamster (Figure S4). The K76 downregulation was consistent with human OPL and OSCC. Although hamster cheek pouch model has several areas of uniqueness, it also lacks lymphatic drainage as observed in humans, mice, or rats, which makes it immunoprotected [33,45,46]. However none of the existing animal models in studies on oral cancer are fully satisfactory and simulate tobacco chewing [33,47,48]. Hamster is one of the extensively used models, as the oral epithelium has similar histological and genetic events involved in the development of premalignant lesions and tumors as in humans [34,49,50,51].

In order to investigate the effect of *KRT76* loss, we used *KRT76*-KO mice. The transgenic and knockout mouse models provide unique advantage of genetic manipulation of specific target gene/s, it also has similar intracellular signaling pathways as of humans [52]. In-vivo systems over comes the weakness of in-vitro experiments which fails to replicate the complex cellular and tissue interaction in an organism; hence, better suited for observing the overall effects of a target gene in a living system.

KRT76-KO mice displayed hyperplastic changes in buccal epithelium, however they do not spontaneously develop tumors similar to previous reports on other keratin knockout mice models [53,54,55]. Our current findings suggest that the loss of *KRT76* may not be a sole molecular event leading to oral cancer development. However, the hyperplastic changes observed in *KRT76*-KO mice points to an indirect role of *KRT76* in regulating proliferation of the basal layers of buccal mucosa similar to previous findings of *KRT10* loss [54]. Overall, our data implies the fact that carcinogenesis being multifactorial and multistep process, potential role of *KRT76* as one of the factor, which alone is not sufficient for cell transformation; however, its contribution in oral carcinogenesis cannot be ruled out.

We envision a number of possible ways in which *KRT76* loss contributes to cancer development. One is that it contributes to a barrier defect in the epithelium, which may render the tissue more susceptible to penetration by carcinogens [56]. Another is that *KRT76* loss may lead to a disturbed inflammatory infiltrate; which is observed in human and mouse epidermis on loss of structural proteins [57,58]. We did not see loss of *KRT76* in benign hyperproliferative oral epithelium, with associated inflammation, nevertheless, altered immune infiltrates are a hallmark of OSCC [59,60].

Future investigations are needed to assess the impact of *KRT76* loss in predicting high-risk precancerous lesions of oral cavity. We observed *KRT76* downregulation in patients with gingivobuccal cancers – a sub site of oral cancer, which is etiologically associated with peculiar tobacco and betel quid chewing habit common in India. These results have to be generalized with caution to other etiologies associated with development of oral tumors. Although, *KRT76* loss is characteristic of gingivobuccal tumors it is not associated with cell transformation, our results warrant future studies to understand other key players driving the process of oral carcinogenesis.

Supporting Information

Figure S1 Representative images of IHC grades. Manual grading of IHC staining was done as 0, +1, +2, +3 depending on staining intensity. (TIF)

References

- Schweizer J, Bowden PE, Coulombe PA, Langbein L, Lane EB, et al. (2006) New consensus nomenclature for mammalian keratins. The Journal of cell biology 174: 169–174.
- Coulombe PA, Omary MB (2002) ‘Hard’ and ‘soft’ principles defining the structure, function and regulation of keratin intermediate filaments. Current opinion in cell biology 14: 110–122.
- Karantza V (2011) Keratins in health and cancer: more than mere epithelial cell markers. Oncogene 30: 127–138.
- Moll R, Divo M, Langbein L (2008) The human keratins: biology and pathology. Histochem Cell Biol 129: 705–733.
- Bragulla HH, Homberger DG (2009) Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. Journal of anatomy 214: 516–559.
- Gires O, Mack B, Rauch J, Matthias C (2006) CK8 correlates with malignancy in leukoplakia and carcinomas of the head and neck. Biochemical and biophysical research communications 343: 252–259.
- Matthias C, Mack B, Berghaus A, Gires O (2008) Keratin 8 expression in head and neck epithelia. BMC cancer 8: 267.
- Wei KJ, Zhang L, Yang X, Zhong LP, Zhou XJ, et al. (2009) Overexpression of cytokeratin 17 protein in oral squamous cell carcinoma in vitro and in vivo. Oral diseases 15: 111–117.
- Xu XC, Lee JS, Lippman SM, Ro JY, Hong WK, et al. (1995) Increased expression of cytokeratins CK8 and CK19 is associated with head and neck carcinogenesis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 4: 871–876.
- Fillies T, Werkmeister R, Packeisen J, Brandt B, Morin P, et al. (2006) Cytokeratin 8/18 expression indicates a poor prognosis in squamous cell carcinomas of the oral cavity. BMC cancer 6: 10.
- Yanagawa T, Yoshida H, Yamagata K, Onizawa K, Tabuchi K, et al. (2007) Loss of cytokeratin 13 expression in squamous cell carcinoma of the tongue is a possible sign for local recurrence. Journal of experimental & clinical cancer research : CR 26: 215–220.
- Dale BA, Salonen J, Jones AH (1990) New approaches and concepts in the study of differentiation of oral epithelia. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists 1: 167–190.
- Collin C, Ouhayoun JP, Grund C, Franke WW (1992) Suprabasal marker proteins distinguishing keratinizing squamous epithelia: cytokeratin 2 polypeptides of oral masticatory epithelium and epidermis are different. Differentiation 51: 137–148.
- Presland RB, Dale BA (2000) Epithelial structural proteins of the skin and oral cavity: function in health and disease. Crit Rev Oral Biol Med 11: 383–408.
- Chu PG, Weiss LM (2002) Keratin expression in human tissues and neoplasms. Histopathology 40: 403–439.
- Bloor BK, Seddon SV, Morgan PR (2000) Gene expression of differentiation-specific keratins (K4, K13, K1 and K10) in oral non-dysplastic keratoses and lichen planus. Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology 29: 376–384.
- Ohkura S, Kondoh N, Hada A, Arai M, Yamazaki Y, et al. (2005) Differential expression of the keratin-4, -13, -14, -17 and transglutaminase 3 genes during the development of oral squamous cell carcinoma from leukoplakia. Oral oncology 41: 607–613.

Figure S2 Oncomine data search for *KRT76* expression in Oral tissues. Data search showed two studies reporting *KRT76* downregulation; A: Ginos et.al Cancer Res. 2004 Jan 1;64(1): 55–63; Observed fold change of about –24.55, and it ranked in top 10% of under expressed genes. B: Toruner GA et.al Cancer Genet Cytogenet. 2004 Oct 1;154(1): 27–35; Observed fold change of about –69.55, and it ranked in top 14% of under expressed genes. (TIF)

Figure S3 Expression of K76 in inflamed buccal mucosa. IHC staining of inflamed buccal epithelium showed higher expression of K76 with respective isotype control. (TIF)

Figure S4 Sequential downregulation of K76 expression during tumor development in hamster buccal epithelium. Gradual decrease in K76 IHC staining was observed in different weeks, [1st week (B), 2nd week (C), 4th week (D), 6th week (E), 8th week (F), 10th week (G), 12th week (H), 16th week (I)], of DMBA treated buccal epithelium; whereas controls of all weeks showed consistent staining.(A). (TIF)

Figure S5 Histology of KO (A) and WT (B) mice dorsal tongue, along with respective K76 IHC staining. (TIF)

Acknowledgments

The authors thank all participants of the study. Mrs. Sadhana Kannan is acknowledged for her help in statistical analysis. ICMR National Tumor Tissue Repository, Tata Memorial Centre, Mumbai is acknowledged for providing tumor tissues. The authors sincerely acknowledge Dr. Miriam Rosin and Dr. Hector Hernandez-Vargas for their critical suggestions in improving the manuscript.

Author Contributions

Conceived and designed the experiments: SA PGB MM. Performed the experiments: SA PGB MP. Analyzed the data: SA PGB AP. Contributed reagents/materials/analysis tools: EH GK SK GBM RSD FMW MM. Wrote the paper: SA PGB GBM MM.

18. Boldrup L, Coates PJ, Gu X, Nylander K (2007) DeltaNp63 isoforms regulate CD44 and keratins 4, 6, 14 and 19 in squamous cell carcinoma of head and neck. *The Journal of pathology* 213: 384–391.
19. Schaaij-Visser TB, Bremmer JF, Braakhuis BJ, Heck AJ, Slijper M, et al. (2010) Evaluation of cornulin, keratin 4, keratin 13 expression and grade of dysplasia for predicting malignant progression of oral leukoplakia. *Oral oncology* 46: 123–127.
20. Mikami T, Cheng J, Maruyama S, Kobayashi T, Funayama A, et al. (2011) Emergence of keratin 17 vs. loss of keratin 13: their reciprocal immunohistochemical profiles in oral carcinoma in situ. *Oral oncology* 47: 497–503.
21. Sakamoto K, Aragaki T, Morita K, Kawachi H, Kayamori K, et al. (2011) Down-regulation of keratin 4 and keratin 13 expression in oral squamous cell carcinoma and epithelial dysplasia: a clue for histopathogenesis. *Histopathology* 58: 531–542.
22. Takikita M, Hu N, Shou JZ, Giffen C, Wang QH, et al. (2011) Fascin and CK4 as biomarkers for esophageal squamous cell carcinoma. *Anticancer Res* 31: 945–952.
23. Su L, Morgan PR, Lane EB (1996) Keratin 14 and 19 expression in normal, dysplastic and malignant oral epithelia. A study using in situ hybridization and immunohistochemistry. *Journal of oral pathology & medicine* : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology 25: 293–301.
24. Crowe DL, Milo GE, Shuler CF (1999) Keratin 19 downregulation by oral squamous cell carcinoma lines increases invasive potential. *J Dent Res* 78: 1256–1263.
25. Khanom R, Sakamoto K, Pal SK, Shimada Y, Morita K, et al. (2012) Expression of basal cell keratin 15 and keratin 19 in oral squamous neoplasms represents diverse pathophysiologies. *Histology and histopathology* 27: 949–959.
26. Ambatipudi S, Gerstung M, Pandey M, Samant T, Patil A, et al. (2012) Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Genes Chromosomes Cancer* 51: 161–173.
27. Vaidya MM, Sawant SS, Borges AM, Ogale SB, Bhisey AN (1998) Cytokeratin expression in precancerous lesions of the human oral cavity. *Oral oncology* 34: 261–264.
28. Vaidya MM, Borges AM, Pradhan SA, Bhisey AN (1996) Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa. *European journal of cancer Part B, Oral oncology* 32B: 333–336.
29. Toyoshima T, Vairaktaris E, Nkenke E, Schlegel KA, Neukam FW, et al. (2008) Cytokeratin 17 mRNA expression has potential for diagnostic marker of oral squamous cell carcinoma. *Journal of cancer research and clinical oncology* 134: 515–521.
30. Raul U, Sawant S, Dange P, Kalraiya R, Ingle A, et al. (2004) Implications of cytokeratin 8/18 filament formation in stratified epithelial cells: induction of transformed phenotype. *International journal of cancer Journal international du cancer* 111: 662–668.
31. Fortier AM, Asselin E, Cadrin M (2013) Keratin 8 and 18 Loss in Epithelial Cancer Cells Increases Collective Cell Migration and Cisplatin Sensitivity through Claudin1 Up-regulation. *J Biol Chem* 288: 11555–11571.
32. Crowe DL, Milo GE, Shuler CF (1999) Keratin 19 downregulation by oral squamous cell carcinoma lines increases invasive potential. *Journal of dental research* 78: 1256–1263.
33. Lu SL, Herrington H, Wang XJ (2006) Mouse models for human head and neck squamous cell carcinomas. *Head & neck* 28: 945–954.
34. Gimenez-Conti IB, Shin DM, Bianchi AB, Roop DR, Hong WK, et al. (1990) Changes in keratin expression during 7,12-dimethylbenz[a]anthracene-induced hamster cheek pouch carcinogenesis. *Cancer Res* 50: 4441–4445.
35. Boyd NM, Reade PC (1991) Temporal alterations in cytokeratin expression during experimental oral mucosal carcinogenesis. *Carcinogenesis* 12: 1767–1771.
36. Kumar G, Tajpara P, Maru G (2012) Dietary turmeric post-treatment decreases DMBA-induced hamster buccal pouch tumor growth by altering cell proliferation and apoptosis-related markers. *Journal of Environmental Pathology, Toxicology and Oncology* 31: 295–312.
37. Salley JJ (1954) Experimental carcinogenesis in the cheek pouch of the Syrian hamster. *Journal of dental research* 33: 253–262.
38. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402–408.
39. Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KG (2008) Polytomous logistic regression analysis could be applied more often in diagnostic research. *Journal of clinical epidemiology* 61: 125–134.
40. Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, et al. (2004) Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res* 64: 55–63.
41. Toruner GA, Ulger C, Alkan M, Galante AT, Rinaggio J, et al. (2004) Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer Genet Cytogenet* 154: 27–35.
42. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, et al. (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9: 166–180.
43. Clausen H, Vedtofte P, Moe D, Dabelsteen E, Sun TT, et al. (1986) Differentiation-dependent expression of keratins in human oral epithelia. *J Invest Dermatol* 86: 249–254.
44. Bloor BK, Seddon SV, Morgan PR (2001) Gene expression of differentiation-specific keratins in oral epithelial dysplasia and squamous cell carcinoma. *Oral oncology* 37: 251–261.
45. Tanaka T, Ishigamori R (2011) Understanding carcinogenesis for fighting oral cancer. *Journal of oncology* 2011: 603740.
46. Schwartz JL, Sloane D, Shklar G (1989) Prevention and inhibition of oral cancer in the hamster buccal pouch model associated with carotenoid immune enhancement. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 10: 297–309.
47. Moggetti B, Di Carlo F, Berta GN (2006) Animal models in oral cancer research. *Oral oncology* 42: 448–460.
48. Kanojia D, Vaidya MM (2006) 4-nitroquinoline-1-oxide induced experimental oral carcinogenesis. *Oral oncology* 42: 655–667.
49. Santis H, Shklar G, Chauncey HH (1964) Histochemistry of Experimentally Induced Leukoplakia and Carcinoma of the Hamster Buccal Pouch. *Oral surgery, oral medicine, and oral pathology* 17: 207–218.
50. Vairaktaris E, Spyridonidou S, Papakosta V, Vylliotis A, Lazaris A, et al. (2008) The hamster model of sequential oral oncogenesis. *Oral oncology* 44: 315–324.
51. Gimenez-Conti IB, Slaga TJ (1993) The hamster cheek pouch carcinogenesis model. *Journal of cellular biochemistry Supplement* 17F: 83–90.
52. Taneja P, Zhu S, Maglic D, Fry EA, Kendig RD, et al. (2011) Transgenic and knockout mice models to reveal the functions of tumor suppressor genes. *Clinical Medicine Insights Oncology* 5: 235–257.
53. Kroger C, Vijayaraj P, Reuter U, Windoffer R, Simmons D, et al. (2011) Placental vasculogenesis is regulated by keratin-mediated hyperoxia in murine decidua tissues. *The American journal of pathology* 178: 1578–1590.
54. Reichelt J, Magin TM (2002) Hyperproliferation, induction of c-Myc and 14-3-3sigma, but no cell fragility in keratin-10-null mice. *Journal of cell science* 115: 2639–2650.
55. Konig K, Meder L, Kroger C, Diehl L, Florin A, et al. (2013) Loss of the keratin cytoskeleton is not sufficient to induce epithelial mesenchymal transition in a novel KRAS driven sporadic lung cancer mouse model. *PLoS One* 8: e57996.
56. Pan X, Hobbs RP, Coulombe PA (2012) The expanding significance of keratin intermediate filaments in normal and diseased epithelia. *Current opinion in cell biology*.
57. Brown SJ, McLean WH (2012) One remarkable molecule: filaggrin. *The Journal of investigative dermatology* 132: 751–762.
58. Sevilla LM, Nachat R, Groot KR, Klement JF, Uitto J, et al. (2007) Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *The Journal of cell biology* 179: 1599–1612.
59. Szczepanski MJ, Czysowska M, Szajnik M, Harasymczuk M, Boyiadzis M, et al. (2009) Triggering of Toll-like receptor 4 expressed on human head and neck squamous cell carcinoma promotes tumor development and protects the tumor from immune attack. *Cancer Res* 69: 3105–3113.
60. Mignogna MD, Fedele S, Lo Russo L, Lo Muzio L, Bucci E (2004) Immune activation and chronic inflammation as the cause of malignancy in oral lichen planus: is there any evidence ? *Oral oncology* 40: 120–130.

Low prevalence of transcriptionally active human papilloma virus in Indian patients with HNSCC and leukoplakia



Priyanka G. Bhosale, MSc,^a Manishkumar Pandey, MSc,^a Rajiv S. Desai, MDS,^b Asawari Patil, MD,^c Shubhada Kane, MD,^c Kumar Prabhaskar, MD, DM, ECMO, PDCR,^d and Manoj B. Mahimkar, MSc, PhD^a

Objectives. In the present study, we comprehensively analyzed the prevalence of transcriptionally active human papilloma virus (HPV) in tissue samples of Indian patients with leukoplakia, predominantly hyperplastic lesions and head and neck squamous cell carcinoma (HNSCC). In addition, saliva samples from patients with HNSCC were screened for HPV detection.

Study Design. P16 overexpression was analyzed by immunohistochemistry. Tissue samples of leukoplakia (n = 121) and HNSCC (n = 427) and saliva from patients with HNSCC (n = 215) were tested for HPV using nested polymerase chain reaction. Positive samples were sequenced for subtyping. The presence of HPV E6/E7 mRNA was confirmed by RNA in situ hybridization.

Results. P16 expression and HPV DNA were not detected in any of the leukoplakia specimens. Of the 427 HNSCC tumors, 9 showed p16 overexpression and 7/427 cases were positive for HPV16 DNA, in saliva or tissue. E6/E7 mRNA positivity was observed in 8 HNSCC samples, primarily from patients with no habit of tobacco consumption. The prevalence of high-risk HPV was restricted to oropharynx and larynx, with very little concordance between p16 overexpression and HPV positivity. All patients with HPV-positive saliva samples had transcriptionally active HPV present in their tumors.

Conclusion. The presence of HPV DNA does not necessarily reflect transcriptionally active virus in tumors; hence, it is important to consider this fact while categorizing HPV-associated tumors. (Oral Surg Oral Med Oral Pathol Oral Radiol 2016; 122:609-618)

Tobacco and alcohol abuse are known risk factors for the development of head and neck squamous cell carcinoma (HNSCC). Over the last 2 decades, high-risk human papilloma virus (HPV) types 16, 18, 31, 33, 52, and 58 have emerged as important etiologic factors in the development of a subset of HNSCCs.^{1,2} HPV-positive tumor samples show a distinct molecular genetic signature and have a better clinical response than HPV-negative samples, implying that HPV-positive patients may not require aggressive therapy.³⁻⁹ These considerations provide a strong rationale to consider HPV association as one of the stratification factors and call for a reliable biomarker for HPV detection.¹⁰⁻¹³

Methods to detect HPV vary considerably across laboratories, not just in their design, but also in their targets, such as HPV-specific nucleic acids,

oncoproteins, cellular proteins, or serum antibodies.¹⁴⁻²¹ However, for widespread implementation in the clinical field, detection methods must be accurate, cost-effective, and readily transferable to the routine diagnostic laboratory. Considering the differing sensitivities and specificities of each of the HPV detection methods, establishing a consensus approach for HPV detection is a challenge. Multimodality detection strategies utilize the strengths of individual assays to optimize the overall reliability of HPV detection and would thus determine HPV status more accurately.^{11,22,23} Detection of p16 INK4 A (p16) overexpression followed by screening of HPV DNA is a widely used method at present.²⁴ In addition, E6 and E7 viral oncoproteins contribute to malignant transformation of cells; hence, detection of E6/E7 transcripts is regarded as the gold standard for the detection of HPV infection, as it also verifies the viral transcription.^{11,25,26}

Wang et al. described a novel RNA in situ hybridization (ISH) technology, RNAscope, that enables analysis of RNA in formalin-fixed, paraffin-embedded

The authors are thankful to the Council of Scientific and Industrial Research and Tata Memorial Center — SIA Intramural Research Grant for funding the project, and to Tata Memorial Center and the Department of Atomic Energy for providing fellowships to P.B. and M.P. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

^aCancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, India.

^bDepartment of Oral Pathology and Microbiology, Nair Hospital Dental College, Mumbai, India.

^cDepartment of Pathology, Tata Memorial Hospital, Tata Memorial Centre, Parel, Mumbai, India.

^dDepartment of Medical Oncology, Tata Memorial Hospital, Tata Memorial Centre, Parel, Mumbai, India.

Received for publication Nov 3, 2015; returned for revision Jun 3, 2016; accepted for publication Jun 5, 2016.

© 2016 Elsevier Inc. All rights reserved.

2212-4403/\$ - see front matter

<http://dx.doi.org/10.1016/j.oooo.2016.06.006>

Statement of Clinical Relevance

The presence of transcriptionally active human papilloma virus (HPV) in head and neck squamous cell carcinoma tumors can form a better stratifying factor. Additionally, saliva can be used as a surrogate biological sample for HPV detection.

(FFPE) tissues along with visualization of mRNAs, while preserving tissue morphology.²⁷ RNAscope HR-HPV Assay (Advanced Cell Diagnosis, Hayward, CA, USA) is an explicit and sensitive method that utilizes oligonucleotide probes specific to each subtype of E6/E7 mRNA and facilitates detection of viral transcripts in FFPE tumor specimens.^{11,28-30}

Globally, the proportion of HPV-related tumors is higher in the oropharynx than in other head and neck cancer subsites.³¹⁻³³ Indian patients with HNSCC are etiologically associated with the use of smokeless tobacco, although a proportion of tumors that may harbor HPV infections cannot be neglected. Overall, 18 Indian studies have reported the prevalence of HPV infections in HNSCC; however, none of these studies comprehensively evaluated the role of HPV in leukoplakia and HNSCC subsites using multiple methods (Supplemental Table SI). The aim of the present study was to evaluate the prevalence of transcriptionally active HPV along with its sitewise distribution in patients with leukoplakia, a predominant and potentially malignant oral disorder, and HNSCC. In addition, we screened the saliva samples of patients with HNSCC for the presence of HPV.

MATERIALS AND METHODS

Human tissue and saliva sample collection

The study was approved by the Institutional Review Board and the Local Ethics Committee of Tata Memorial Hospital and Nair Hospital Dental College. Written informed consent was obtained from all patients who participated in this study. Pathologically confirmed paraffin blocks and frozen tissue samples of leukoplakia ($n = 121$) and tissue samples of HNSCC ($n = 427$) were obtained from Nair Hospital Dental College and Tata Memorial Hospital, respectively. In addition, saliva samples from patients with HNSCC ($n = 215$; 134 cases with concurrent tissue and saliva) were collected using the Oragene DNA Self-Collection Kit OG-510 (DNA Genotek Inc., ON, Canada). HNSCC subsites were classified as per AJCC 7th Edition TNM Staging 2010 guidelines, with 297 (58.5%) tumor samples from the oral cavity, 43 (8.5%) from the hypopharynx, 71 (14%) from the larynx, and 97 (19.1%) from the oropharynx.

p16 immunohistochemistry

p16 immunohistochemistry (IHC) was performed using the CINtec Histology Kit (MTM Laboratories, AG, Heidelberg, Germany), per the manufacturer's protocol. Five-micron FFPE tumor sections were deparaffinized with xylene and rehydrated with sequential ethanol washes (100%, 90%, and 70%). After antigen unmasking for 20 min at 96°C to 99°C in epitope

retrieval buffer, slides were allowed to cool down to room temperature. Endogenous peroxidase blocking was performed for 5 ± 1 min with the hydrogen peroxide provided with the kit. Slides were then incubated for 30 min at room temperature with the primary antibody provided with the kit. Cervical cancer samples showing high expression of p16 were used as a positive control. Visualization was performed for 30 min at room temperature using the reagent provided with the kit, followed by development with a substrate-chromogen solution for 5-10 min. Finally, the slides were counterstained with hematoxylin and examined under upright microscope (AxioCam MRcS Carl Zeiss Imager, Z1, GmBH) at 10 \times magnification.

Immunohistochemical assessment and scoring

Two pathologists (A.P., S.K.) scored all samples independently for IHC staining intensity as well as the percentage of cells stained, and discordant cases were reviewed again to arrive at a consensus score. For assessment of p16 protein expression, cytoplasmic and nuclear staining intensity was categorized based on H-score (range, 0-300), which was obtained as follows: $H\text{-score} = 1 \times (\% \text{ of cells staining} + 1 \text{ intensity}) + 2 \times (\% \text{ of cells staining} + 2 \text{ intensity}) + 3 \times (\% \text{ of cells staining} + 3 \text{ intensity})$. p16 IHC was scored positive if there was strong as well as diffuse nuclear and cytoplasmic staining present in $>70\%$ of the malignant cells, according to College of American Pathologist criteria, comparable to an H-score of more than 140. All other staining patterns were scored negative. However, for RNA ISH analysis, tumor samples with p16 staining in $\geq 5\%$ cells ($H\text{-score} = 50$) were processed further.

DNA extraction from tumor tissue and saliva samples

Genomic DNA was isolated from the frozen tumor tissue samples using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, 25-30 mg of tissue was pulverized by grinding with liquid nitrogen, followed by the addition of RLT (Qiagen) plus buffer with β -mercaptoethanol (Sigma-Aldrich, USA). The homogenate was processed for column purification and isolation of DNA as well as RNA. QIAamp DNA FFPE Tissue Kit (Qiagen) was used for DNA extraction from paraffin tissue. Ten sections of 10 μ m were processed for DNA extraction per manufacturer's protocol. Samples were deparaffinized with xylene followed by overnight lysis in proteinase K. Subsequently, column purification of DNA was performed. In addition, genomic DNA from saliva was extracted by heating the saliva sample at 56°C for 6 h. Further purification and extraction were performed per manufacturer's protocol. The quantity and quality of

DNA/RNA were determined using NanoDrop ND-1000 (NanoDrop Technologies, DE, USA).

Nested polymerase chain reaction for HPV detection

HPV was determined by nested polymerase chain reaction (PCR) primers that targeted conserved HPV L1 region and detected a broad range of HPV subtypes. A total of 300 ng of DNA was subjected to the first round of PCR using MY09/MY11 primers to produce 450 bp product.³⁴ Amplification was performed in a Peltier thermal cycler (PTC-100, MA, USA) for 40 cycles with the following cycling conditions: initial denaturation for 5 min at 95°C, followed by denaturation for 1 min at 95°C, annealing for 1 min at 55°C, extension for 1 min at 72°C, and a final extension step for 10 min at 72°C. Subsequently, PCR was performed using GP5+/GP6+ primers to obtain 150 bp product. MY09/MY11 PCR product was amplified using touchdown-cycling parameters¹⁹: initial denaturation for 5 min at 95°C, followed by denaturation for 1 min at 95°C, touchdown annealing cycles of 2 min with 0.5°C decrement/cycle from 50°C to 40°C (21 cycles), additional annealing cycles at 40°C for 2 min (10 cycles), extension for 1.5 min at 72°C, and final extension of 4 min at 72°C. To rule out the chance of false positives, no template control or negative control (C33 A DNA) was included while performing PCR reactions. To avoid cross-contamination of DNA samples, utmost precautions were taken, including strictly maintaining dedicated areas and equipment for DNA extraction and PCR setup. β -Globin PCR was performed on all samples using PC03/PC04 primer set to verify the amplifiability and integrity of the extracted DNA. Primers used for PCR amplification are listed in [Supplemental Table SII](#).

HPV PCR sensitivity

The sensitivity of the PCR system used in the present study was checked by performing MY09/MY11 PCR at different concentrations (100, 50, 25, 10, 5, 0.5, 0.05, and 0.005 ng) of standard HPV controls CaSki (HPV16), SiHa (HPV16), and HeLa (HPV18) DNA. For subsequent GP5+/GP6+ PCR, 5 μ L of 1:100 dilution of an MY09/MY11 PCR product was used ([Supplemental Figure S1](#)).

DNA sequencing

Nested PCR-positive products were purified using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Duren, Germany) and sequenced with GP5+ primers (1.5 pmol); the sequences were compared with the HPV genome available at the National Center for Biotechnology

Information GenBank using the basic local alignment search tool (BLAST). A complementary analysis of the sequence was performed using a papilloma viral genome database, the Papillomavirus Episteme (PAVE).³⁵

RNA ISH

Detection of 7 high-risk HPV genotype (HPV16, 18, 31, 33, 35, 52, and 58) E6/E7 mRNAs was manually performed using the RNAscope 2.5 HD assay with Brown HPV HR7 Kit (Advanced Cell Diagnostics Inc., CA, USA) according to the manufacturer's instructions. Cases that were HPV DNA-positive or expressed p16 (low or high) were processed for RNA ISH. Briefly, 5- μ m FFPE tissue sections were deparaffinized and pretreated to allow access to target RNA, followed by hybridization for 2 h with target-specific probes for the E6 and E7 genes of high-risk HPV. Ubiquitin C (UBC, a constitutively expressed endogenous gene) was used as an internal positive control, and the bacterial gene dapB was used as a negative control for background signal. HeLa cells, fixed with 10% formaldehyde/phosphate-buffered saline (ACD positive control) were used during the processing of each batch. Signal detection was performed with diaminobenzidine for 10 min; the slides were counterstained with hematoxylin and examined under upright microscope (20 \times and 40 \times magnification).

For each case, all 3 stained slides (HPV, UBC, and dapB) were examined simultaneously to determine the HPV status. The UBC test was used to assess the presence of hybridizable RNA; if the UBC slide was negative, the sample was disqualified, presuming insufficient RNA quality. The dapB test was used to assess nonspecific staining; only those cases that were negative or had weak staining were considered for HPV scoring. A positive HPV test result was defined as punctate staining that colocalized to the cytoplasm or nucleus of the malignant cells.³⁶

Statistical analysis

All statistical analysis was performed using IBM SPSS version 21. Chi-square test was used to determine the correlation between p16 overexpression, DNA PCR, RNA ISH, and HPV status in HNSCC tissue samples. Multinomial logistic regression analysis was performed to evaluate the prognostic value of each of patient's characteristic and HPV RNA status. All *P* values were 2-sided, and those <.05 were considered statistically significant.

RESULTS

Patient characteristics

The clinicopathological and demographic characteristics of all leukoplakia and HNSCC tissue samples are

summarized in Table I. The patients included in this study were predominantly male, and a majority of these patients were smokeless tobacco users or had mixed habits (chewing along with bidi/cigarette smoking and/or alcohol use). Smokeless tobacco products commonly used were khaini (tobacco with lime) and betel quid (tobacco with betel leaf and lime). Very few patients were exclusively smokers or exclusively alcohol users, and ~10% patients did not have any tobacco or alcohol habits. Most of the tumor samples were in TNM stages III or IV, and approximately 50% of the cases had lymph node metastasis. With regard to leukoplakia, a majority of the lesions had hyperplasia (82%), while mild, moderate, and severe dysplasia was observed in 10%, 7%, and 1% of these lesions, respectively.

p16 protein expression

p16 expression was analyzed by IHC in leukoplakia (n = 121) and HNSCC (n = 427) tissue samples. IHC staining was graded as high p16 expression or low/no p16 expression, as described in Methods (Supplemental Figure S2). All leukoplakia tissue samples were negative for p16 expression, whereas 9 HNSCC tissue samples showed p16 overexpression and were considered positive as per CAP criteria (Table II).

Detection of HPV DNA

Nested PCR methods were used to detect HPV DNA in all samples irrespective of their p16 protein expression status. PCR analysis was performed in leukoplakia (n = 121) and HNSCC (n = 347) tissue samples and in saliva samples from patients with HNSCC (n = 215). Because of the limited tissue quantity, DNA could not be extracted from the remaining 80 HNSCC tumor samples. In 134 HNSCC cases, we performed concurrent analysis of HPV DNA in both tumor tissue and saliva samples (Table I). Samples that showed saliva positivity were screened for the presence of HPV DNA in their respective tumor tissue. PCR analysis revealed that 5 of the 347 HNSCC tissue samples (1.4%) and 4 of the 202 saliva samples (2%) were positive for the presence of HPV, of which 2 cases were positive in both tissue and saliva samples (Table III). Furthermore, the presence of HPV16 subtype in PCR-positive cases was confirmed by sequencing (Table II).

Detection of transcriptionally active HPV

All the cases that were positive for HPV DNA or expressed p16 (low or high) were screened for expression of E6/E7 mRNA using RNAscope 2.5 HD assay with Brown HPV HR7 Kit. Tumor samples were

Table I. Demographic details of study cohort

Characteristic	Patients with leukoplakia, n = 121 (%)	Patients with HNSCC, n = 508 (%)
Sample type		
Tissue	121 (100)	293 (57.7)
Saliva	NA	81 (15.9)
Tissue + saliva	NA	134 (26.4)
Site of origin		
Oral cavity	120 (99.2)	297 (58.5)
Gingiva, buccal mucosa, and lower lip	115 (95)	239 (47)
Retromolar trigone and alveolar ridges	1 (0.8)	16 (3.1)
Floor of mouth	0 (0)	4 (0.8)
Hard palate	0 (0)	0 (0)
Mobile tongue	4 (3.3)	38 (7.5)
Hypopharynx	0 (0)	43 (8.5)
Postcricoid region	0 (0)	6 (1.2)
Pyriform sinus	0 (0)	37 (7.3)
Larynx	0 (0)	71 (14)
Glottic	0 (0)	17 (3.3)
Supraglottic	0 (0)	54 (10.6)
Transglottic	0 (0)	0 (0)
Subglottic	0 (0)	0 (0)
Oropharynx	1 (0.8)	97 (19.1)
Base of tongue	0 (0)	38 (7.5)
Posterior wall	0 (0)	8 (1.6)
Soft palate	1 (0.8)	9 (1.8)
Tonsil	0 (0)	42 (8.3)
Age at diagnosis		
Median (range)	45 (17-80)	52 (18-89)
Gender		
Male	108 (91.5)	396 (78.1)
Female	10 (8.5)	111 (21.9)
Stage		
Stage I and II	NA	115 (22.7)
Stage III and IV	NA	392 (77.3)
T classification		
T1	NA	74 (14.6)
T2	NA	160 (31.6)
T3	NA	111 (21.9)
T4	NA	162 (32)
N classification		
Lymph node metastasis negative (N0)	NA	215 (42.3)
Lymph node metastasis positive (N+)	NA	293 (57.7)
Patient habit profile		
No habit	1 (0.8)	50 (9.8)
Exclusive tobacco chewer	28 (23.1)	155 (30.5)
Exclusive smoker	13 (10.7)	44 (8.7)
Exclusive alcohol user	0 (0)	11 (2.2)
Mixed user	52 (43)	133 (26.2)
No info	27 (22.3)	115 (22.6)

HNSCC, head and neck squamous cell carcinoma; NA, not applicable.

considered HR-HPV RNA—positive if punctate brown staining was observed in the cytoplasm or the nucleus of the malignant cells (Figure 1). In total, 8 of the 427 (1.9%) tumor tissue samples showed HR-HPV RNA positivity; however, the HPV status of 5 tissue samples was undetermined due to poor RNA quality (Table II).

Table II. HPV positivity in leukoplakia and HNSCC, and its correlation with patient's tobacco habit

	<i>Leukoplakia n (%)</i>	<i>HNSCC n (%)</i>	<i>Clinical parameter</i>	
			<i>No habit</i>	<i>Tobacco habit*</i>
p16 immunostaining	(n = 121)	(n = 427)		
Negative/p16 No expression	121 (100)	418 (97.9)	40 (93)	254 (97.7)
Positive/p16 overexpressing	0 (0)	9 (2.1)	3 (7)	6 (2.3)
HPV DNA				
Tissue sample	(n = 121)	(n = 347)		
Negative	121 (100)	342 (98.6)	35 (92.1)	279 (99.3)
High-risk HPV [†]	0 (0)	5 (1.4)	3 (7.9)	2 (0.7)
Saliva sample	(n = 0)	(n = 202)		
Negative	NA	198 (98)	19 (86.4)	176 (99.4)
High-risk HPV [†]	NA	4 (2)	3 (13.6)	1 (0.6)
High-risk HPV RNA (7 Subtypes)	(n = 121)	(n = 427)		
HPV RNA Negative	121 (100)	414 (97)	39 (88.6)	347 (98)
HPV RNA Positive	0 (0)	8 (1.9)	5 (11.4)	3 (0.8)
Insufficient RNA for Assessment	0 (0)	5 (1.2)	0 (0)	4 (1.1)

HPV, human papilloma virus; HNSCC, head and neck squamous cell carcinoma.

*All cases excluding "No info" and exclusive alcohol users.

[†]Confirmed by sequencing.

All the HR-HPV RNA—positive tumor samples were of laryngeal and oropharyngeal origin, with approximately 20% positivity in soft palate and tonsillar tumors (Table IV). In addition, in a single case, we observed the presence of HPV DNA with negative E6/E7 mRNA expression, implying the presence of latent HPV that may not have any role in tumor development. HPV activity was observed in 11.4% of patients with no tobacco habit, compared with 0.8% of those with a history of tobacco consumption (Table II). Multinomial logistic regression analysis showed a high probability of the presence of active HPV in cases with no tobacco consumption ($P = .001$, odds ratio 11.1, and 95% confidence interval 2.5–48.5). All tumor samples that harbored transcriptionally active HPV were in an advanced stage and mostly positive for lymph node metastasis. Complete demographic characteristics of HPV RNA—positive cases are represented in Table III.

p16 overexpression and its correlation with HPV DNA and E6/E7 mRNA expression

The current analysis revealed that 3 out of 8 samples overexpressing p16 had HPV DNA in either tumor tissue or saliva samples (single case) and also expressed E6/E7 mRNA in concurrent tumor tissues. On the other hand, 5 out of 8 samples positive for HR-HPV RNA had weak nuclear and cytoplasmic p16 staining ($P < .0001$), demonstrating a poor correlation between p16 expression and the presence of transcriptionally active HPV infection (Table III).

The cases that had HPV DNA in saliva samples showed positivity for E6/E7 mRNA ($P < .0001$) in respective tumor samples, implying a positive

correlation between the presence of HPV DNA in saliva samples and viral transcription activity. In addition, 80% of the tumor tissue samples that were positive for HPV DNA harbored biologically active HPV, and 60% of these had confirmed presence of HPV in saliva ($P = .04$; Table III). Overall, our results indicate that relying on the single detection method may lead to misinterpretation of HPV status. Hence, we used the combinational algorithm for detection of high-risk HPV in the study group (Figure 2).

DISCUSSION

High-risk HPV has emerged as an important etiologic agent associated with the development of a subset of HNSCC, and its prognostic value has stimulated an increasing number of studies in the last 2 decades.^{1,13,37} In India, a major proportion of HNSCCs and precancerous lesions are etiologically associated with abuse of smokeless tobacco in various forms, such as khaini and betel quid with tobacco, along with smoking and alcohol consumption.^{38–40} However, there exists a group of patients with no habit of tobacco consumption who may potentially harbor high-risk HPV infection.

Few studies have reported on HPV screening in the Indian population, and most of these studies focused on the use of a single marker for HPV detection (Supplemental Table SI). The present study was conducted on a large number of patients, detailing the prevalence of HPV and its site-wise distribution in Indian patients with leukoplakia and HNSCC using a combinational detection algorithm. Our analysis reveals no leukoplakia and gingivobuccal tumor samples positive for HPV, and this was consistent with our previous report on gingivobuccal cancers.⁴¹ However,

Table III. Demographic details of HNSCC patients positive for transcriptionally active HPV

Case	Age (y)	Gender	Site of primary	Subsite of primary	T	Node	Stage	Habit	TC	S	D	P16 (*N)	P16 (*C)	PCR (tumor)	PCR (saliva)	HPV subtype	RNA ISH
1	45	Male	Larynx	Supraglottis	T3	N+	IV	Yes	Yes	No	Yes	0	0	+	-	HPV 16	+
2	31	Female	Larynx	Supraglottis	T4	N+	IV	No	No	No	No	0	0	+	+	HPV 16	+
3	51	Male	Larynx	Supraglottis	T3	N0	III	Yes	No	Yes	No	15	15	-	-	Negative	+
4	26	Female	Larynx	Supraglottis	T1	N+	III	No	No	No	No	210	140	+	-	HPV 16	+
5	49	Male	Oropharynx	Tonsil	T4	N+	IV	No	No	No	No	0	0	+	+	HPV 16	+
6	49	Female	Oropharynx	Tonsil	T3	N+	IV	No	No	No	No	140	100	-	+	HPV 16	+
7	20	Male	Oropharynx	Tonsil	T4	N+	IV	No	No	No	No	120	160	-	-	Negative	+
8	61	Female	Oropharynx	Soft palate	T4	N+	IV	Yes	Yes	No	No	60	50	-	+	HPV 16	+
9*	42	Male	Oropharynx	Tonsil	T1	N+	IV	Yes	No	Yes	No	180	180	+	-	Unevaluated	+

HNSCC, head and neck squamous cell carcinoma; HPV, human papilloma virus; TC, tobacco chewing; S, cigarette/bidi smoking; D, alcohol drinking; *N, nuclear expression H-score; *C, cytoplasmic expression H-score; N0, no lymph node metastasis; N+, lymph node metastasis.

*Unevaluated: unable to sequence.

transcriptionally active HPV16 was detected in 6.1% of laryngeal and 9.1% of oropharyngeal tumors.

Our results are in line with most of the studies conducted globally on different HNSCC cohorts, indicating a site-specific association of HPV.⁴²⁻⁴⁵ Most of the previous studies on Indian HNSCCs reported >25% HPV-positive samples, which may be due to smaller study size (Supplemental Table SI). Additional variability could be attributed to discrepancies in tumor subsite categorization and methods used for HPV detection, along with biological sample types, such as tumor tissue, exfoliated oral cells, saliva, or serum used for screening.

Recently, the use of saliva for detection of HPV and various HNSCC markers has gained importance (Supplemental Table SIII). Detection of HPV DNA in saliva samples is possible if the virus particles are actively produced by infected cells or released after the infected cells die. Furthermore, saliva can be obtained by a noninvasive method, which improves compliance. In addition, the simple DNA extraction protocol makes it a promising biological fluid for larger population-based studies. Reports have shown that the probability of detecting HPV in salivary rinses increases with the viral load, which in turn may be associated with a high risk of developing HPV-related oral lesions, including malignancy, along with an increased risk of recurrence and distant metastasis.⁴⁶⁻⁴⁸ In the present study, 2% of samples were positive for salivary HPV, and all of these cases had a biologically active virus, independently confirming the lower prevalence of HPV in the Indian population.

Active HPV infection is marked with an expression of viral E6 and E7 oncoproteins, which further abrogates p53 and pRB functions. Repression of pRB leads to overexpression of cellular p16, which is widely being used as a surrogate marker for HPV detection.^{17,49} The rationale for using p16 is largely driven by its simplicity and low cost and the feasibility of IHC analysis.⁵⁰ Larsen et al., in their extensive review, suggested the use of CAP criteria to achieve the highest correlation between p16 IHC and HPV results, wherein tumor samples showing p16 overexpression in >70% of cells (nuclear and cytoplasmic) are considered HPV positive.⁵¹

Substantial differences exist in the definition of p16 overexpression and means of HPV diagnostics among the studies; hence, several studies question the reliability of this marker. In this context, several factors need to be considered, including the subjective nature of IHC evaluation, the varied mechanisms of p16 expression in HNSCC, and the lack of uniform scoring and interpretive criteria.^{9,24,50} In addition, it must be noted that deletion of and hypermethylation in the p16 (*CDKN2 A*) gene is an early event in the development

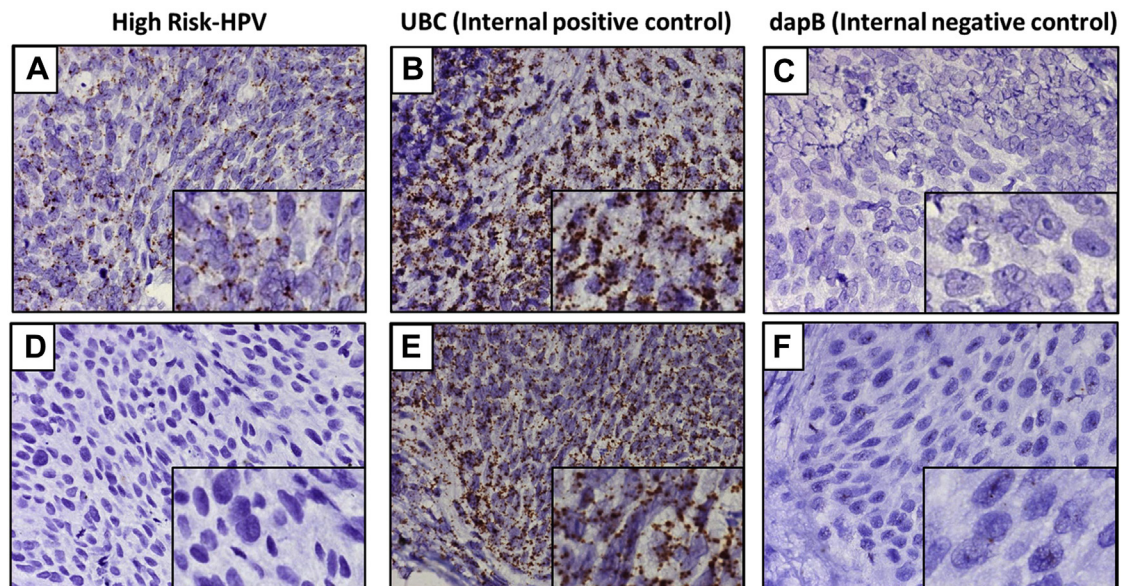


Fig. 1. Detection of high-risk human papilloma virus (HPV) RNA in head and neck squamous cell carcinoma (HNSCC) tissue samples. RNA in situ hybridization (ISH) showing punctate brown signals for HR-HPV RNA in an HNSCC tumor tissue sample (A), and its respective internal positive control (B). HPV-negative HNSCC tumor tissue with no signal for HR-HPV RNA (D), and punctate brown signals for internal positive control (E). No signals were observed in internal negative controls (C and F) in either case.

Table IV. Sitewise distribution of HPV in HNSCC cohort

	<i>p16 immunostaining n (%)</i>	<i>High-risk HPV DNA n (%)</i>		<i>High-risk HPV RNA n (%)</i>
	<i>Tumor tissue</i>	<i>Saliva</i>	<i>Tumor tissue</i>	<i>Tumor tissue</i>
Oral cavity	4/296 (1.4)	0/5 (0)	0/294 (0)	0/296 (0)
Gingiva, buccal mucosa, and lower lip	1/239 (0.4)	NA	0/239 (0)	0/239 (0)
Retromolar trigone and alveolar ridges	2/15 (13.3)	0/3 (0)	0/14 (0)	0/15 (0)
Floor of mouth	0/4 (0)	NA	0/4 (0)	0/4 (0)
Mobile tongue	1/38 (2.6)	0/2 (0)	0/37 (0)	0/38 (0)
Hypopharynx	0/27 (0)	0/42 (0)	0/11 (0)	0/27 (0)
Posterior region	0/4 (0)	0/6 (0)	0/2 (0)	0/4 (0)
Pyramidal sinus	0/23 (0)	0/36 (0)	0/9 (0)	0/23 (0)
Larynx	1/49 (2)	1/68 (1.5)	3/16 (18.8)	3/49 (6.1)
Glottic	0/9 (0)	0/16 (0)	0/1 (0)	0/9 (0)
Supraglottic	1/40 (2.5)	1/52 (1.9)	3/15 (20)	3/40 (7.5)
Oropharynx	4/54 (7.4)	3/87 (3.4)	2/26 (7.7)	5/55 (9.1)
Base of tongue	0/23 (0)	0/35 (0)	0/9 (0)	0/24 (0)
Posterior wall	1/5 (20)	0/8 (0)	0/4 (0)	0/5 (0)
Soft palate	0/5 (0)	1/9 (11.1)	0/4 (0)	1/5 (20)
Tonsil	3/21 (14.3)	2/35 (5.7)	2/9 (22.2)	4/21 (19)

HPV, human papilloma virus; NA, not applicable.

of HNSCC,^{52,53} which could further affect p16 protein expression and the interpretation of HPV status. Current data also demonstrate the presence of HPV in tumor samples that were negative for p16 expression; furthermore, tumor samples with strong nuclear and cytoplasmic p16 expression were HPV-negative. In addition to screening for p16 overexpression, PCR-based methods are universally used to detect HPV L1 region and E6/E7 genes. Nevertheless, for PCR-based

analysis, the quality of DNA and type of starting tissue material could be a limitation; furthermore, over-sensitivity may lead to false positive results. Due to the lack of a defined algorithm for HPV detection, use of multiple methods is recommended.^{23,54-56}

The HPV-positive HNSCC subset in the absence of tobacco abuse comprises a distinct molecular, clinical, and pathologic disease entity, and it has a markedly improved prognosis.^{6,25,57} In contrast, HPV-positive

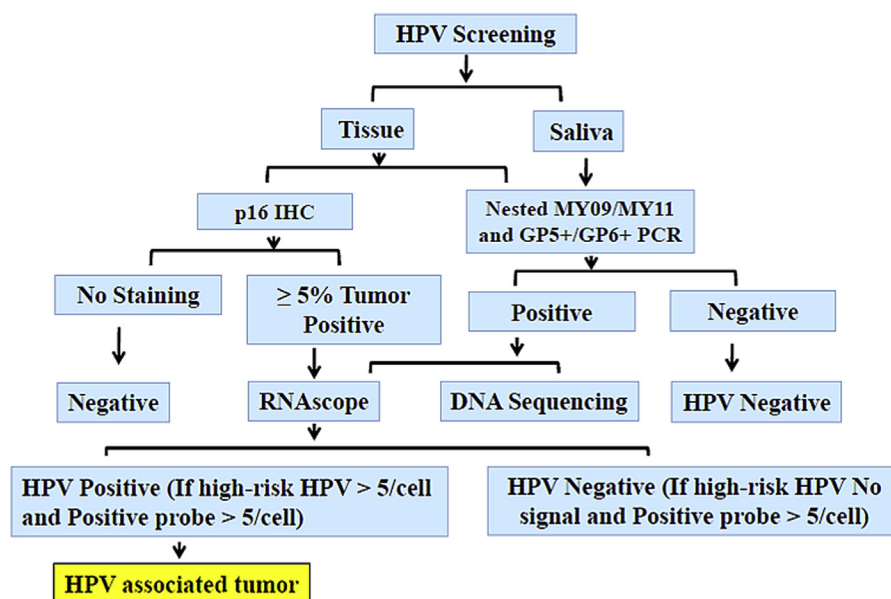


Fig. 2. Algorithm for human papilloma virus detection in patients with leukoplakia and head and neck squamous cell carcinoma.

HNSCC tumor samples associated with abuse of tobacco and/or alcohol are more aggressive and behave similar to conventional HNSCC.⁵⁸ Hence, it is crucial to understand the causal association of HPV in the development of HNSCC, because the mere presence of latent HPV would not necessarily imply its involvement in tumorigenesis. Thus it is essential to check for transcriptional activity of HPV that could potentially help in planning the treatment of patients with HNSCC.

In this study, we have screened for biologically active HPV using RNA ISH. Histology-based approaches have an advantage in that the viral distribution in tumor and non-neoplastic surrounding areas can be distinguished and it is easy to detect E6/E7 mRNA using small tissue biopsies. PCR-based approaches have various other limitations, such as the challenge of extracting a desirable quality and quantity of RNA, especially from paraffin blocks. Eight of the tumor samples analyzed in this study had transcriptionally active HPV. In addition, a case was detected as positive for HPV16 DNA but did not have E6/E7 mRNA expression, implying that in such tumors, HPV may not have any role in tumor development. Thus, to improve the accuracy of categorizing HPV-positive HNSCC, it is necessary to follow an algorithm that utilizes the strengths of various detection assays in combination, which reduces the chance of false detection.

Despite all efforts, our study has a few limitations. First, the number of HPV-positive cases was very low; hence, we could not comment on the correlation between the presence of HPV in saliva and concurrent tumor tissue samples. Second, the PCR-based methods

that we employed efficiently detected multiple HPV subtypes but may favor the amplification of more abundant HPV subtypes. Furthermore, we could not perform quantitative PCR-based analysis because of poor quality and yield of extracted DNA and RNA; this would have provided insight about the tumor viral load.

In summary, our analysis revealed the absence of HPV in leukoplakia and gingivobuccal tumors. Only 10% of laryngeal and oropharyngeal tumors were seen to harbor transcriptionally active HPV16. Screening E6/E7 transcripts by RNA ISH confirmed the presence of integrated, transcriptionally active virus and its histologic distribution in tissues. Overall, this study highlights the importance of HPV RNA screening as a gold standard for detecting biologically active infection. In addition, our analysis emphasizes the use of saliva for HPV detection in patients with HNSCC. Further study is warranted before saliva can be used as a surrogate for tumor tissue for HPV screening in Indian cohorts, where the majority of patients with leukoplakia and HNSCC have a tobacco habit.

The authors thank all study participants. The Indian Council of Medical Research National Tumor Tissue Repository and the Department of Pathology, Tata Memorial Centre, are acknowledged for providing tumor tissue samples. We also thank Mr. P. Chavan and Mr. V. Sakpal from the Histology Section, Advanced Center for Treatment, Research and Education in Cancer, Tata Memorial Centre, for their help. The authors sincerely acknowledge Dr. N. Joshi, Dr. D. Anantharaman, and Dr. S. Ambatipudi for their critical suggestions in improving the manuscript. The authors would like to thank SciEditage (<http://www.scieditage.in>) for editing the manuscript for the English language.

REFERENCES

1. Ragin CC, Taioli E. Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: Review and meta-analysis. *Int J Cancer*. 2007;121:1813-1820.
2. Michaud DS, Langevin SM, Eliot M, et al. High-risk HPV types and head and neck cancer. *Int J Cancer*. 2014;135:1653-1661.
3. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363:24-35.
4. Duray A, Descamps G, Decaestecker C, et al. Human papillomavirus predicts the outcome following concomitant chemoradiotherapy in patients with head and neck squamous cell carcinomas. *Oncol Rep*. 2013;30:371-376.
5. Braakhuis BJ, Snijders PJ, Keune WJ, et al. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *J Natl Cancer Inst*. 2004;96:998-1006.
6. Gillison ML. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. *Semin Oncol*. 2004;31:744-754.
7. Fakhry C, Westra WH, Li S, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst*. 2008;100:261-269.
8. Klussmann JP, Mooren JJ, Lehnen M, et al. Genetic signatures of HPV-related and unrelated oropharyngeal carcinoma and their prognostic implications. *Clin Cancer Res*. 2009;15:1779-1786.
9. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer*. 2011;11:9-22.
10. Lassen P. The role of human papillomavirus in head and neck cancer and the impact on radiotherapy outcome. *Radiother Oncol*. 2010;95:371-380.
11. Mirghani H, Amen F, Moreau F, et al. Human papilloma virus testing in oropharyngeal squamous cell carcinoma: What the clinician should know. *Oral Oncol*. 2014;50:1-9.
12. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: A virus-related cancer epidemic. *Lancet Oncol*. 2010;11:781-789.
13. Ang KK, Sturgis EM. Human papillomavirus as a marker of the natural history and response to therapy of head and neck squamous cell carcinoma. *Semin Radiat Oncol*. 2012;22:128-142.
14. Agoston ES, Robinson SJ, Mehra KK, et al. Polymerase chain reaction detection of HPV in squamous carcinoma of the oropharynx. *Am J Clin Pathol*. 2010;134:36-41.
15. Argyris PP, Kademani D, Pambuccian SE, Nguyen R, Tosios KI, Koutlas IG. Comparison between p16 INK4 A immunohistochemistry and human papillomavirus polymerase chain reaction assay in oral papillary squamous cell carcinoma. *J Oral Maxillofac Surg*. 2013;71:1676-1682.
16. Deng Z, Hasegawa M, Kiyuna A, et al. Viral load, physical status, and E6/E7 mRNA expression of human papillomavirus in head and neck squamous cell carcinoma. *Head Neck*. 2013;35:800-808.
17. Lewis JS Jr. p16 Immunohistochemistry as a standalone test for risk stratification in oropharyngeal squamous cell carcinoma. *Head Neck Pathol*. 2012;6(suppl 1):S75-S82.
18. Lewis JS Jr, Chernock RD, Ma XJ, et al. Partial p16 staining in oropharyngeal squamous cell carcinoma: Extent and pattern correlate with human papillomavirus RNA status. *Mod Pathol*. 2012;25:1212-1220.
19. Evans MF, Adamson CS, Simmons-Arnold L, Cooper K. Touchdown General Primer (GP5+/GP6+) PCR and optimized sample DNA concentration support the sensitive detection of human papillomavirus. *BMC Clin Pathol*. 2005;5:10.
20. Liang C, Marsit CJ, McClean MD, et al. Biomarkers of HPV in head and neck squamous cell carcinoma. *Cancer Res*. 2012;72:5004-5013.
21. Wang H, Wang MX, Su N, et al. RNAscope for in situ detection of transcriptionally active human papillomavirus in head and neck squamous cell carcinoma. *J Vis Exp*. 2014;11(85).
22. Braakhuis BJ, Brakenhoff RH, Meijer CJ, Snijders PJ, Leemans CR. Human papilloma virus in head and neck cancer: The need for a standardised assay to assess the full clinical importance. *Eur J Cancer*. 2009;45:2935-2939.
23. Westra WH. Detection of human papillomavirus (HPV) in clinical samples: Evolving methods and strategies for the accurate determination of HPV status of head and neck carcinomas. *Oral Oncol*. 2014;50:771-779.
24. Hoffmann M, Tribius S, Quabius ES, et al. HPV DNA, E6*I-mRNA expression and p16 INK4 A immunohistochemistry in head and neck cancer: how valid is p16 INK4 A as surrogate marker? *Cancer Lett*. 2012;323:88-96.
25. Jung AC, Briolat J, Millon R, et al. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. *Int J Cancer*. 2010;126:1882-1894.
26. van Houten VM, Snijders PJ, van den Brekel MW, et al. Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. *Int J Cancer*. 2001;93:232-235.
27. Wang F, Flanagan J, Su N, et al. RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn*. 2012;14:22-29.
28. Bishop JA, Ma XJ, Wang H, et al. Detection of transcriptionally active high-risk HPV in patients with head and neck squamous cell carcinoma as visualized by a novel E6/E7 mRNA in situ hybridization method. *Am J Surg Pathol*. 2012;36:1874-1882.
29. Ukpo OC, Flanagan JJ, Ma XJ, Luo Y, Thorstad WL, Lewis JS Jr. High-risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. *Am J Surg Pathol*. 2011;35:1343-1350.
30. Holzinger D, Schmitt M, Dyckhoff G, Benner A, Pawlita M, Bosch FX. Viral RNA patterns and high viral load reliably define oropharynx carcinomas with active HPV16 involvement. *Cancer Res*. 2012;72:4993-5003.
31. Horewicz VV, Feres M, Rapp GE, Yasuda V, Cury PR. Human papillomavirus-16 prevalence in gingival tissue and its association with periodontal destruction: A case-control study. *J Periodontol*. 2010;81:562-568.
32. Anantharaman D, Gheit T, Waterboer T, et al. Human papillomavirus infections and upper aero-digestive tract cancers: The ARCA study. *J Natl Cancer Inst*. 2013;105:536-545.
33. Zaravinos A. An updated overview of HPV-associated head and neck carcinomas. *Oncotarget*. 2014;5:3956-3969.
34. Winder DM, Ball SL, Vaughan K, et al. Sensitive HPV detection in oropharyngeal cancers. *BMC Cancer*. 2009;9:440.
35. Van Doorslaer K, Tan Q, Xirasagar S, et al. The Papillomavirus Episteme: A central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res*. 2013;41:D571-D578.
36. Schache AG, Liloglou T, Risk JM, et al. Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma. *Br J Cancer*. 2013;108:1332-1339.
37. Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol*. 2014;50:565-574.
38. Boffetta P, Hecht S, Gray N, Gupta P, Straif K. Smokeless tobacco and cancer. *Lancet Oncol*. 2008;9:667-675.
39. Dikshit R, Gupta PC, Ramasundarahettige C, et al. Cancer mortality in India: A nationally representative survey. *Lancet*. 2012;379:1807-1816.

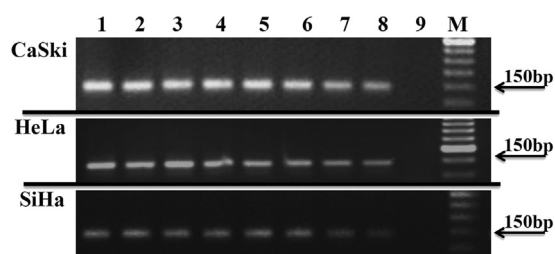
40. Gupta PC, Ray CS. Smokeless tobacco and health in India and South Asia. *Respirology*. 2003;8:419-431.
41. Ambatipudi S, Gerstung M, Gowda R, et al. Genomic profiling of advanced-stage oral cancers reveals chromosome 11 q alterations as markers of poor clinical outcome. *PLoS One*. 2011;6:e17250.
42. Lopez RV, Levi JE, Eluf-Neto J, et al. Human papillomavirus (HPV) 16 and the prognosis of head and neck cancer in a geographical region with a low prevalence of HPV infection. *Cancer Causes Control*. 2014;25:461-471.
43. Machado J, Reis PP, Zhang T, et al. Low prevalence of human papillomavirus in oral cavity carcinomas. *Head Neck Oncol*. 2010;2:6.
44. Ha PK, Pai SI, Westra WH, et al. Real-time quantitative PCR demonstrates low prevalence of human papillomavirus type 16 in premalignant and malignant lesions of the oral cavity. *Clin Cancer Res*. 2002;8:1203-1209.
45. Lingen MW, Xiao W, Schmitt A, et al. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Oncol*. 2013;49:1-8.
46. Adamopoulou M, Vairaktaris E, Panis V, Nkenke E, Neukam FW, Yapijakis C. HPV detection rate in saliva may depend on the immune system efficiency. *In Vivo*. 2008;22:599-602.
47. Zhao M, Rosenbaum E, Carvalho AL, et al. Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. *Int J Cancer*. 2005;117:605-610.
48. Chuang AY, Chuang TC, Chang S, et al. Presence of HPV DNA in convalescent salivary rinses is an adverse prognostic marker in head and neck squamous cell carcinoma. *Oral Oncol*. 2008;44:915-919.
49. Hoffmann M, Ihloff AS, Gorogh T, et al. p16 (INK4 A) overexpression predicts translational active human papillomavirus infection in tonsillar cancer. *Int J Cancer*. 2010;127:1595-1602.
50. El-Naggar AK, Westra WH. p16 expression as a surrogate marker for HPV-related oropharyngeal carcinoma: A guide for interpretative relevance and consistency. *Head Neck*. 2012;34:459-461.
51. Gronhøj Larsen C, Gyldenlove M, Jensen DH, et al. Correlation between human papillomavirus and p16 overexpression in oropharyngeal tumours: A systematic review. *Br J Cancer*. 2014;110:1587-1594.
52. Kresty LA, Mallery SR, Knobloch TJ, et al. Alterations of p16 (INK4 A) and p14 (ARF) in patients with severe oral epithelial dysplasia. *Cancer Res*. 2002;62:5295-5300.
53. Sailasree R, Abhilash A, Sathyan KM, Nalinakumari KR, Thomas S, Kannan S. Differential roles of p16 INK4 A and p14 ARF genes in prognosis of oral carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2008;17:414-420.
54. Mirghani H, Ugolin N, Ory C, et al. A predictive transcriptomic signature of oropharyngeal cancer according to HPV16 status exclusively. *Oral Oncol*. 2014;50:1025-1034.
55. Robinson M, Schache A, Sloan P, Thavaraj S. HPV specific testing: A requirement for oropharyngeal squamous cell carcinoma patients. *Head Neck Pathol*. 2012;6(Suppl 1):S83-S90.
56. Smeets SJ, Hesselink AT, Speel EJ, et al. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. *Int J Cancer*. 2007;121:2465-2472.
57. D'Souza G, Kreimer AR, Viscidi R, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med*. 2007;356:1944-1956.
58. Maxwell JH, Kumar B, Feng FY, et al. Tobacco use in human papillomavirus-positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. *Clin Cancer Res*. 2010;16:1226-1235.

Reprint requests:

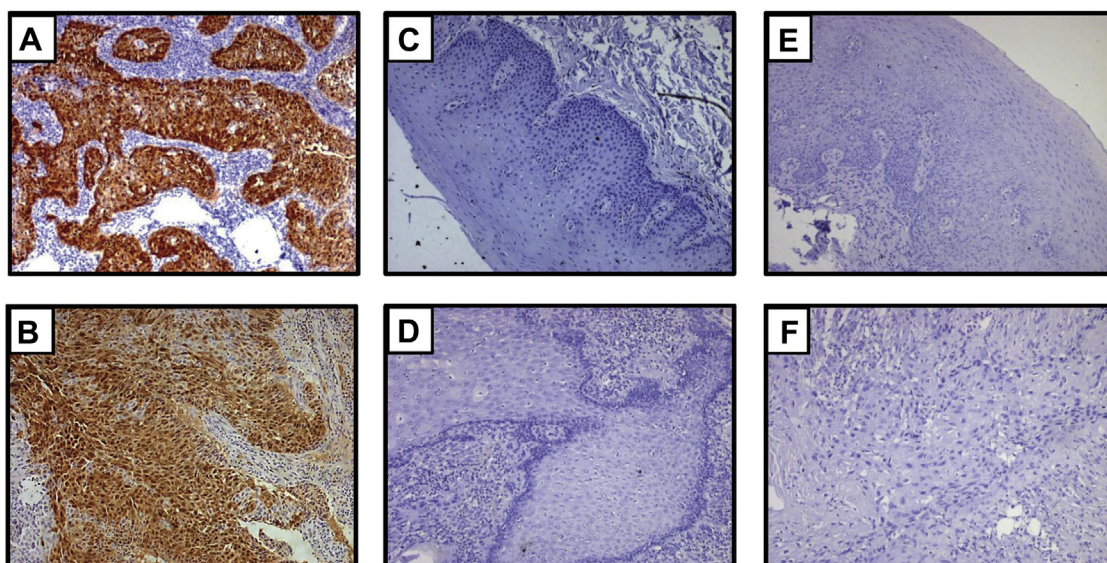
Manoj B. Mahimkar
Cancer Research Institute
Advanced Centre for Treatment
Research and Education in Cancer
Kharghar
Navi Mumbai 410210
India
mmahimkar@actrec.gov.in

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.oooo.2016.06.006>.



Supplemental Fig. S1. The sensitivity of nested PCR assay for HPV detection. M is Gene Ruler 50 bp DNA Ladder (Thermo Scientific, Massachusetts, USA); 1–8 correspond to 100, 50, 25, 10, 5, 0.5, 0.05, and 0.005 ng per reaction of CasKi, HeLa, and SiHa DNA; 9 is no template control. Specific band indicated (arrow) at 150 bp for GP5+/GP6+.



Supplemental Fig. S2. Overexpression of p16 in HNSCC tissue samples. IHC staining shows nuclear and cytoplasmic expression for p16 in the positive control (A), and HPV-positive oropharyngeal tumors (B - Moderately differentiated squamous cell carcinoma); whereas, no staining was observed in leukoplakia (C - Hyperplasia, and E - Hyperplasia with mild dysplasia) and gingivobuccal tumors (D - Well differentiated squamous cell carcinoma and F - Moderately differentiated squamous cell carcinoma). Magnification 100 \times .

Supplemental Table SI. Prevalence of HPV in Indian oral precancerous lesions and HNSCC

Author and year	Pubmed I.D.	Tumor site and sample size (n)	Detection method			Sequencing and Genotyping array	ISH (DNA/RNA)	HPV subtype detected	HPV positivity (percentage)
			ELISA	IHC	PCR				
Current study		Leukoplakia (121), Patient Saliva (217), and HNSCC (484)	Unevaluated	p16 IHC using CINtec Kit	Nested PCR for L1 gene	Sequencing	RNA in situ hybridization (RNAscope)	HPV 16	Leukoplakia 0/121 (0%), Oral cavity 0/296 (0%), Hypopharynx 0/27 (0%), Larynx 3/49 (6.1%), Oropharynx 5/55 (9.1%)
Talukdar et al, 2014 ¹	25213493	UADT SCC (219)	Unevaluated	Unevaluated	PCR for L1 gene	Unevaluated	Unevaluated	HPV 16 and HPV 18	UADT SCC cases 81/219 (36.96%)
Pattanshetty et al, 2014 ²	24588599	Normal oral mucosa (60)	Unevaluated	Unevaluated	PCR for E6 gene of HPV 16 and 18	Unevaluated	Unevaluated	HPV 16 and HPV 18	Healthy subjects 39/60 (65%)
Prakash et al, 2013 ³	24551639	Leukoplakia (21) and OSCC (69)	Unevaluated	p16 IHC (BioGenex)	Unevaluated	Unevaluated	Unevaluated	Unevaluated	p16 positivity Leukoplakia 12/21(57.14%), OSCC 49/69 (71.01%)
Maitra et.al, 2013 ⁴	24292195	OSCC (110)	Unevaluated	Unevaluated	Nested PCR for L1 gene	Sequencing	Unevaluated	HPV 16 and HPV 18	OSCC cases 13/50 (26%)
Jamaly et al, 2012 ⁵	22231431	HNSCC (110)	Unevaluated	Unevaluated	Unevaluated	Unevaluated	DNA ISH	HPV 16 and HPV 18	HNSCC cases 110 (45%) numbers not specified
Chaudhary et al, 2010 ⁸	20863370	OSMF (208) and OSCC (222)	Unevaluated	Unevaluated	PCR for L1 gene and Hybrid capture II assay	Unevaluated	Unevaluated	HPV 16	OSMF 57/208 (27.4%), OSCC 70/222 (31.53%)
Khanna et al, 2009 ⁹	23133118	Healthy controls (45), Leukoplakia (30), and OSCC (45)	Unevaluated	Unevaluated	Unevaluated	Unevaluated	Southern blot	HPV 16 and HPV 18	Healthy controls 9/45 (20%), Leukoplakia 12/30 (40%), oral cancer 29/45 (64.5%)
Bhattacharya et al, 2009 ¹⁰	19453846	Healthy controls (236 including tissue or blood) and HNSCC (236)	Unevaluated	Unevaluated	PCR for L1 gene	Unevaluated	Southern blot	HPV 16 and HPV 18	HNSCC 145/229 (63%)
Ghosh et al, 2009 ¹¹	19023882	HNSCC (155)	Unevaluated	p16 IHC (Santa Cruz)	PCR for L1 gene	Genotyping array	Unevaluated	HPV 16 and HPV 18	Head and neck lesions 87/155 (56%)
Majumder et al, 2009 ¹²	18764855	Healthy controls (100); Leukoplakia (91) and OSCC (83)	Unevaluated	Unevaluated	PCR for L1 gene	Sequencing	Southern blot	HPV 16 and HPV 18	Healthy controls 23/100 (23%), Leukoplakia 37/87 (42.52%), SCC 38/83 (45.78%)
Koppikar et al, 2005 ¹⁴	15514945	OSCC (83) and HNSCC (19)	Unevaluated	Unevaluated	PCR for L1 gene	Sequencing	Unevaluated	HPV 16,18 and HPV 8	HNSCC 32/102 (31%)
Katiyar et al, 2003 ¹⁵	14577584	Healthy controls (20) and OSCC (44)	Unevaluated	Unevaluated	Southern Blot PCR for HPV 16 and 18 and PCR for L1 gene	Unevaluated	Southern blot	HPV 16 and HPV 18	OSCC 13/44 (29.54%)

(continued on next page)

Supplemental Table SI. Continued

Author and year	Pubmed I.D.	Tumor site and sample size (n)	Detection method				Sequencing and Genotyping array	ISH (DNA/RNA)	HPV subtype detected	HPV positivity (percentage)
			ELISA	IHC	PCR					
Nagpal et al, 2002 ¹⁶	11807792	OSCC (110)	Unevaluated	Unevaluated	PCR for L1 gene	Unevaluated	Unevaluated	HPV 16 and HPV 18	OSCC cases 37/110 (33.6%)	
D'Costa et al, 1998 ¹⁷	9861351	Healthy controls (48), Malignant lesions (80), OSCC (100)	Unevaluated	Unevaluated	PCR for L1 gene	Unevaluated	Southern blot	HPV 16 and HPV 18	Normal Mucosa 15/48 (31%), Malignant Lesions 27/80 (34%), Oral cancer 15/100 (15%)	
Balaram et al, 1995 ¹⁸	7759149	Pre-cancerous lesions (15), OSCC (91)	Unevaluated	Unevaluated	Specific PCR for HPV 6/11	Sequencing	Unevaluated	HPV 6,11,16 and 18	Pre-cancerous lesions 13/15 (86.66%), Oral cancer 67/91 (73.62%)	

HNSCC, head and neck squamous cell carcinoma; *OSCC*, oral squamous cell carcinoma; *UADT*, upper aerodigestive tract; *OSMF*, oral submucous fibrosis; *ELISA*, enzyme-linked immunosorbent assay; *IHC*, immunohistochemistry; *PCR*, polymerase chain reaction; *HPV*, human papilloma virus; *ISH*, in situ-hybridization.

References:

1. Talukdar FR, Ghosh SK, Laskar RS, et al. Epigenetic pathogenesis of human papillomavirus in upper aerodigestive tract cancers. *Mol Carcinog* 2015;54:1387-1396.
2. Pattanshetty S, Kotrashetti VS, Nayak R, et al. PCR based detection of HPV 16 and 18 genotypes in normal oral mucosa of tobacco users and non-users. *Biotech Histochem* 2014;89:433-439.
3. Prakash P, Khandare M, Kumar M, et al. Immunohistochemical detection of p16(INK4a) in leukoplakia and oral squamous cell carcinoma. *J Clin Diagn Res* 2013;7:2793-2795.
4. Maitra A, Biswas NK, Amin K, et al. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat Commun* 2013;4:2873.
5. Jamaly S, Khanehkenari MR, Rao R, et al. Relationship between p53 overexpression, human papillomavirus infection, and lifestyle in Indian patients with head and neck cancers. *Tumour Biol* 2012;33:543-550.
6. Mathew A, Mody RN, Patait MR, et al. Prevalence and relationship of human papilloma virus type 16 and type 18 with oral squamous cell carcinoma and oral leukoplakia in fresh scrappings: a PCR study. *Indian J Med Sci* 2011;65:212-221.
7. Kulkarni SS, Kulkarni SS, Vastrad PP, et al. Prevalence and distribution of high risk human papillomavirus (HPV) Types 16 and 18 in Carcinoma of cervix, saliva of patients with oral squamous cell carcinoma and in the general population in Karnataka, India. *Asian Pac J Cancer Prev* 2011;12:645-648.
8. Chaudhary A, Pandya S, Singh M, et al. Identification of high-risk human papillomavirus-16 and -18 infections by multiplex PCR and their expression in oral submucous fibrosis and oral squamous cell carcinoma. *Head Neck Oncol* 2013;5:4.
9. Khanna R, Rao GR, Tiwary SK, et al. Detection of human papilloma virus 16 and 18 DNA sequences by southern blot hybridization in oral leukoplakia and squamous cell carcinoma. *Indian J Surg* 2009;71:69-72.
10. Bhattacharya N, Roy A, Roy B, et al. MYC gene amplification reveals clinical association with head and neck squamous cell carcinoma in Indian patients. *J Oral Pathol Med* 2009;38:759-763.
11. Ghosh A, Ghosh S, Maiti GP, et al. SH3GL2 and CDKN2A/2B loci are independently altered in early dysplastic lesions of head and neck: correlation with HPV infection and tobacco habit. *J Pathol* 2009;217:408-419.
12. Majumder M, Indra D, Roy PD, et al. Variant haplotypes at XRCC1 and risk of oral leukoplakia in HPV non-infected samples. *J Oral Pathol Med* 2009;38:174-180.
13. Mitra S, Banerjee S, Misra C, et al. Interplay between human papilloma virus infection and p53 gene alterations in head and neck squamous cell carcinoma of an Indian patient population. *J Clin Pathol* 2007;60:1040-1047.
14. Koppikar P, deVilliers EM, Mulherkar R, et al. Identification of human papillomaviruses in tumors of the oral cavity in an Indian community. *Int J Cancer* 2005;113:946-950.
15. Katiyar S, Thelma BK, Murthy NS, et al. Polymorphism of the p53 codon 72 Arg/Pro and the risk of HPV type 16/18-associated cervical and oral cancer in India. *Mol Cell Biochem* 2003;252:117-124.
16. Nagpal JK, Patnaik S, Das BR, et al. Prevalence of high-risk human papilloma virus types and its association with P53 codon 72 polymorphism in tobacco addicted oral squamous cell carcinoma (OSCC) patients of Eastern India. *Int J Cancer* 2002;97:649-653.
17. D'Costa J, Saranath D, Dedhia P, et al. Detection of HPV-16 genome in human oral cancers and potentially malignant lesions from India. *Oral Oncol* 1998;34:413-420.
18. Balaram P, Nalinakumari KR, Abraham E, et al. Human papillomaviruses in 91 oral cancers from Indian betel quid chewers—high prevalence and multiplicity of infections. *Int J Cancer* 1995;61:450-454.

Supplemental Table SII. PCR primer sequences

<i>Primers</i>	<i>Sequence (5'-3')</i>	<i>Amplicon (bp)</i>
Consensus primers		
MY09	CGTCCMARRGGAWACTGATC	450
MY11	GCMCAGGGWCATAAAYAATGG	
GP5+	TTTGTTACTGTGGTAGATACTAC	150
GP6+	GAAAAATAAACTGTAAATCATATTC	
Beta-Globin primers		
PC03	ACACAACTGTGTTCCTACTAGC	110
PC04	CAACTTCATCCACGTTCCACC	

Supplemental Table SIII. Global scenario of HPV detection in saliva

Author and year	Pubmed I.D.	Sample type	Tumor site and sample size (N)	Detection method					HPV subtype detected
				ELISA	IHC	PCR	Sequencing and Genotyping array	ISH (DNA/RNA)	
Current Study		Saliva and HNSCC tissue	Leukoplakia (121), Saliva (217) and HNSCC (484)	Unevaluated	p16 IHC using CINtec Kit	Nested PCR for L1 gene	Sequencing	RNA in situ hybridisation (RNAscope)	HPV16
Ahn et al, 2014 ¹	25078109	Saliva and OPSCC	OPSCC (93)	Unevaluated	Unevaluated	QPCR for E6 and E7	Unevaluated	Unevaluated	HPV16
D'Souza et al, 2014 ²	24778397	Oral rinse	OPSCC (164) and their partners (93)	ELISA for HPV16 L1, E6 and E7	p16 IHC	PCR for L1 region and Real Time PCR	Genotyping Array	Line-blot hybridization	HPV16
Adamopoulou et al, 2013 ³	23428459	Saliva and Cervical scraping	Cervical cancer (43)	Unevaluated	Unevaluated	Nested PCR for L1 region	Unevaluated	Unevaluated	HPV6, HPV16, HPV31, HPV53, HPV73, HPV77, HPV81 and HPV84
Chen et al, 2013 ⁴	23300223	Saliva	Healthy control (19), HNSCC (156)	Unevaluated	Unevaluated	QPCR for E6 gene of HPV16	Unevaluated	Unevaluated	HPV16
Goot-Heah et al, 2012 ⁶	23464414	Saliva	Healthy controls (30) and OPLs (16),OSCC (14)	Unevaluated	Unevaluated	Nested PCR for HPV18	Unevaluated	Unevaluated	HPV18
Flake et al, 2012 ⁷	23088565	Saliva	Healthy Controls (118)	Unevaluated	Unevaluated	QPCR for HPV16 and 18	Unevaluated	Unevaluated	HPV16 and HPV18
Sun et al, 2012 ⁸	22438973	Saliva	HNSCC (197)	Unevaluated	Unevaluated	Quantitative methylation specific PCR and E6,E7 PCR for HPV	Unevaluated	Unevaluated	Unevaluated
Asvadi Kermani et al, 2012 ⁹	http://troca.tbzmed.ac.ir/uploads/82/CMS/user/file/145/Data1/Human%20Papilloma%20Virus%20in%20Head%20and%20Neck%20Squamous%20Cell.pdf	Saliva and HNSCC tissue	Healthy Controls (94) and HNSCC (14)	Unevaluated	Unevaluated	PCR for L1 region	Unevaluated	Unevaluated	HPV16 and HPV18
Peixoto, et al, 2011 ¹¹	22167030	Saliva and Oral scrap	Healthy Controls (100)	Unevaluated	Unevaluated	PCR for L1 region	Unevaluated	Unevaluated	Unevaluated
Turner et al, 2011 ¹²	21985030	Saliva	Healthy Controls (151)	Unevaluated	Unevaluated	HPV16 & 18 PCR and qRT-PCR	Unevaluated	Unevaluated	HPV16

(continued on next page)

Supplemental Table SIII. Continued

Author and year	Pubmed I.D.	Sample type	Tumor site and sample size (N)	Detection method					HPV subtype detected
				ELISA	IHC	PCR	Sequencing and Genotyping array	ISH (DNA/RNA)	
Carvalho, et al, 2011 ¹³	21628494	Saliva, Serum and HNSCC tissue	HNSCC (61)	Unevaluated	Unevaluated	Quantitative methylation specific PCR and Real Time PCR for E6 and E7 of HPV16	Unevaluated	Unevaluated	HPV16
Kulkarni et al, 2011 ¹⁴	21627358	Saliva and OSCC, cervical tissue	Healthy Controls (396), OSCC (34) and Cervical cancer(60)	Unevaluated	Unevaluated	HPV16 & 18 PCR	Unevaluated	Unevaluated	HPV16 and HPV18
Park et al, 2011 ¹⁵	21292634	Saliva and HNSCC tissue	HNSCC (22)	Unevaluated	Unevaluated	QPCR for HPV16, E6/E7 DNA	Unevaluated	Unevaluated	HPV16
Saheb Jamee et al, 2009 ¹⁶	19680210	Saliva	Healthy Control (20) and OSCC (22)	Unevaluated	Unevaluated	PCR for L1 region	Unevaluated	Unevaluated	HPV6, HPV11, HPV16 and HPV18
Chuang et al, 2008 ¹⁸	18329326	Salivary rinse	HNSCC (59)	Unevaluated	Unevaluated	QPCR for HPV16, E6 and E7 DNA	Unevaluated	Unevaluated	HPV16
Montaldo et al, 2007 ¹⁹	17686007	Saliva	Healthy Controls (164)	Unevaluated	Unevaluated	Seminested PCR for L1 region	Sequencing	Unevaluated	HPV16 and HPV 31
Zhao et al, 2005 ²⁰	15929076	Saliva and HNSCC tissues	Healthy Controls (604) and HNSCC (92)	Unevaluated	Unevaluated	QPCR for HPV16, E6 and E7 DNA	Unevaluated	Unevaluated	HPV16
Tominaga et al, 1996 ²¹	9001342	Saliva, Oral swab and Normal tonsillar mucosa	Oral Pappiloma (8)	Unevaluated	Unevaluated	PCR for HPV 6, E5 gene	Unevaluated	Southern Hybridization	HPV6

HNSCC, head and neck squamous cell carcinoma; OSCC, oral squamous cell carcinoma; OPSCC, oropharyngeal squamous cell carcinoma; OPL, oral premalignant lesions; UADT, upper aerodigestive tract; OSMF, oral submucous fibrosis; ELISA, enzyme-linked immunosorbent assay; IHC, immunohistochemistry; PCR, polymerase chain reaction; ISH, in-situ hybridization; QPCR, quantitative real time PCR; qRT PCR, quantitative reverse transcriptase PCR; HPV, human papilloma virus; HIV, human immunodeficiency virus.

References

- Ahn SM, Chan JY, Zhang Z, et al. Saliva and plasma quantitative polymerase chain reaction-based detection and surveillance of human papillomavirus-related head and neck cancer. JAMA Otolaryngol Head Neck Surg 2014;140:846-854.
- D'Souza G, Gross ND, Pai SI, et al. Oral human papillomavirus (HPV) infection in HPV-positive patients with oropharyngeal cancer and their partners. J Clin Oncol 2014;32:2408-2415.
- Adamopoulou M, Vairaktaris E, Nkenke E, et al. Prevalence of human papillomavirus in saliva and cervix of sexually active women. Gynecol Oncol 2013;129:395-400.
- Chen KM, Stephen JK, Ghanem T, et al. Human papilloma virus prevalence in a multiethnic screening population. Otolaryngol Head Neck Surg 2013;148:436-442.
- Stephen JK, Chen KM, Mahan M, et al. HPV and methylation indicators in paired tumor and saliva in HNSCC. Cancer Clin Oncol 2013;2:42 -54.
- Goot-Heah K, Kwai-Lin T, Froemming GR, et al. Human papilloma virus 18 detection in oral squamous cell carcinoma and potentially malignant lesions using saliva samples. Asian Pac J Cancer Prev 2012;13:6109-6113.
- Flake C, Arafa J, Hall A, et al. Screening and detection of human papillomavirus (HPV) high-risk strains HPV16 and HPV18 in saliva samples from subjects under 18 years old in Nevada: a pilot study. BMC Oral Health 2012;12:43.
- Sun W, Zaboli D, Liu Y, et al. Comparison of promoter hypermethylation pattern in salivary rinses collected with and without an exfoliating brush from patients with HNSCC. PLoS One 2012;7:e33642.
- Kermani A, Seifi S, Dolatkhah R, et al. Human papilloma virus in head and neck squamous cell cancer. Iranian J Cancer Prev 2012;5:21-26.
- Gillison ML, Broutian T, Pickard RK, et al. Prevalence of oral HPV infection in the United States, 2009-2010. JAMA 2012;307:693-703.
- Peixoto AP, Campos GS, Queiroz LB, et al. Asymptomatic oral human papillomavirus (HPV) infection in women with a histopathologic diagnosis of genital HPV. J Oral Sci 2011;53:451-459.

12. Turner DO, Williams-Cocks SJ, Bullen R, et al. High-risk human papillomavirus (HPV) screening and detection in healthy patient saliva samples: a pilot study. *BMC Oral Health* 2011;11:28.
13. Carvalho AL, Henrique R, Jeronimo C, et al. Detection of promoter hypermethylation in salivary rinses as a biomarker for head and neck squamous cell carcinoma surveillance. *Clin Cancer Res* 2011;17:4782-4789.
14. Kulkarni SS, Kulkarni SS, Vastrad PP, et al. Prevalence and distribution of high risk human papillomavirus (HPV) types 16 and 18 in carcinoma of cervix, saliva of patients with oral squamous cell carcinoma and in the general population in Karnataka, India. *Asian Pac J Cancer Prev* 2011;12:645-648.
15. Park IS, Chang X, Loyo M, et al. Characterization of the methylation patterns in human papillomavirus type 16 viral DNA in head and neck cancers. *Cancer Prev Res (Phila)* 2011;4:207-217.
16. SahebJamee M, Boorghani M, Ghaffari SR, et al. Human papillomavirus in saliva of patients with oral squamous cell carcinoma. *Med Oral Patol Oral Cir Bucal* 2009;14:e525-528.
17. Adamopoulou M, Vairaktaris E, Panis V, et al. HPV detection rate in saliva may depend on the immune system efficiency. *In Vivo* 2008;22:599-602.
18. Chuang AY, Chuang TC, Chang S, et al. Presence of HPV DNA in convalescent salivary rinses is an adverse prognostic marker in head and neck squamous cell carcinoma. *Oral Oncol* 2008;44:915-919.
19. Montaldo C, Mastinu A, Quartuccio M, et al. Detection and genotyping of human papillomavirus DNA in samples from healthy Sardinian patients: a preliminary study. *J Oral Pathol Med* 2007;36:482-487.
20. Zhao M, Rosenbaum E, Carvalho AL, et al. Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. *Int J Cancer* 2005;117:605-610.
21. Tominaga S, Fukushima K, Nishizaki K, et al. Presence of human papillomavirus type 6f in tonsillar condyloma acuminatum and clinically normal tonsillar mucosa. *Jpn J Clin Oncol* 1996;26:393-397.