

**GENE EXPRESSION PATTERNS OF CHEEK, GUM AND TONGUE
SQUAMOUS CELL CARCINOMA**

AMYZA MOHAMMED SALEH

THESIS SUBMITTED IN FULFILMENT

OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF

PHILOSOPHY

DEPARTMENT OF ORAL PATHOLOGY,

ORAL MEDICINE AND PERIODONTOLOGY

FACULTY OF DENTISTRY

UNIVERSITY OF MALAYA

KUALA LUMPUR

2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of candidate: Amyza Mohammed Saleh

Registration /Matrix Number: DHA060002

Name of Degree: Doctor of Philosophy

Title of Thesis ("this Work"): Gene Expression Patterns of Oral Squamous Cell Carcinoma from Different Sites of the Oral Cavity

Field of study: Oral Medicine/Oral Oncology

I do solemnly and sincerely declare that:

- 1) I am the sole author/writer of this Work;
- 2) This Work is original;
- 3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract form, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship has been acknowledged in this Work;
- 4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- 5) I hereby assign all and every rights in the copyright to this work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- 6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's signature

Date

Subscribed and solemnly declared before

Witness' signature

Name:

Date

Designation

ABSTRACT

Clinical evidence suggested that biological behavior as well as response to treatment is different in OSCC arising from the different anatomical locations of the mouth and could be influenced by the activation or/and inactivation of different genes and pathways. **Objective:** Gene expression analysis of OSCC from different sites of the oral cavity was conducted to determine if there were significant differences in the expression pattern that could be associated with clinical observations. **Materials and Methods:** Formalin-fixed paraffin-embedded tissues from OSCC from cheek, gum, tongue and non-cancerous oral mucosal from the matching sites were used in microarray experiments (DASL, Illumina) to determine the gene expression patterns. Microarray data were analysed using Genespring to identify differentially expressed genes and these changes were validated using quantitative polymerase chain reaction and immunohistochemistry. The role of specific genes in driving OSCC were determined using cell lines genetically modified to exogenously express these genes. **Results:** This study demonstrated that FFPE tissues can be used for microarray experiments. Differentially expressed genes in OSCC were identified and their expressions were validated in independent samples. Principal Component Analysis demonstrated that different sites of OSCC have distinct gene expression profiles. Genes that were commonly altered in all sites and those that were distinct to a particular site were identified. Focusing on a gene *FOLR1* that was found to be enriched in OSCC of the tongue, exogenous expression of this gene was shown to promote migration and invasion. **Conclusion:** This study suggests that the genetic progression of OSCC in the different sites is distinct, thus cautioning the generalization of OSCC when identifying biomarkers for diagnosis, prognosis and therapy. Furthermore, specific genes may confer different cancer traits that may explain the clinical differences seen in the different sites of OSCC.

ABSTRAK

Bukti klinikal mencadangkan tingkah laku biologi serta respons kepada rawatan berbeza di kalangan OSCC dari lokasi berbeza di mulut. Ia berkemungkinan dipengaruhi oleh pengaktifan atau/dan penyahaktifan gen serta tapak jalan gen yang berbeza. **Objektif:** Analisis ekspresi gen OSCC dari lokasi berbeza di mulut dijalankan untuk menentukan jika ada perbezaan signifikan dalam pola ekspresi berkaitan dengan pemerhatian klinikal. **Bahan dan Kaedah:** “Formalin-fixed paraffin-embedded tissue” dari OSCC daripada bahagian pipi, gusi, lidah serta juga tisu bukan-kanser oral mukosa dari bahagian yang sama digunakan di dalam eksperimen mikroatur gen (DASL, Illumina) untuk menentukan pola ekspresinya. Gen-gen yang mengalami perubahan ekspresi yang signifikan pada OSCC dikenalpasti dengan Genespring dan disahkan melalui tindak balas rantai polimerase kuantitatif dan immunohistokimia. Peranan gen yang spesifik dalam memacu OSCC ditentukan dengan mengubahsuai sel OSCC secara genetik untuk meluahkan gen secara eksogenus. **Keputusan:** Kajian ini menunjukkan yang tisu FFPE boleh digunakan untuk eksperimen mikroatur gen. Gen yang mengalami perubahan pengepresan yang signifikan di dalam OSCC dikenalpasti dan disahkan menggunakan sampel OSCC yang berlainan. “Principal Component Analysis” menunjukkan OSCC dari lokasi berbeza di mulut mempunyai pola pengepresan gen yang berbeza. Ekspresi gen yang berubah dalam ketiga-tiga lokasi dan yang hanya berubah di lokasi tertentu dikenalpasti. Bertumpu kepada gen *FOLR1* yang diperkayai dalam OSCC lidah, luahan eksogenus menunjukkan yang ianya menggalakkan ciri-ciri migrasi dan invasi. **Kesimpulan:** Kajian ini menunjukkan yang perkembangan genetik OSCC berbeza mengikut lokasi berlainan di mulut. Oleh itu, perhatian perlu diberi kepada faktor ini apabila menyamaratakan OSCC dalam pencarian biopenanda diagnosis, prognosis dan terapi. Gen tertentu juga boleh memberikan ciri kanser yang berlainan, seterusnya menjelaskan perbezaan klinikal yang dilihat pada OSCC dari lokasi yang berbeza.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for giving me the strength, patience and perseverance to conduct and complete this study. I am extremely grateful to my supervisors Prof. Rosnah Mohd. Zain & Prof Cheong Sok Ching for their generous guidance, constant encouragement and unflagging support. Many thanks also to the CEO of Cancer Research Initiatives Foundation (CARIF), Prof. Teo Soo-Hwang for supporting this project and for invaluable discussions.

I would like to thank Puan Rusnani Kamal, Mrs Khoo Guat Sim & the rest of the team at the Diagnostic Laboratory, Dental Faculty, University of Malaya; Dr. Haizal Hussaini & Ms Zeti Saiful Bahry from Oral Pathology Department, Universiti Kebangsaan Malaysia for their technical assistance, help in finding tissue specimens and tissue processing as well as helpful discussions. Thank you to the OCRCC team for their support, especially Nabillah Wan Abd Ghani & Karen Ng- despite their hectic schedule they still made time to provide valuable assistance. I am grateful to Dr. Vivek Tanavde & Ms Felicia Ng from Bioinformatics Institute, Singapore for the bioinformatics training session and invaluable input. I would like to acknowledge Malaysia Genome Institute for the use of the microarray data analysis tool for this project. I am indebted to my colleagues especially those in the Oral Cancer Research team (past and present)- Sharifah Hamid, Hor Seen Yii, Dr. Lim Kue Peng, Gan Chai Phei, Sharifah Nurain Zanaruddin, Sam Kin Kit, Yee Pei San & Kong Yink Heay for their generosity in their friendship and knowledge. My thanks also to the other members of the CARIF team.

My deepest gratitude and special heartfelt thanks to my husband Kamal Khalid & my wonderful kids Maryam & Imran for their love, encouragement and welcome (as well as unwelcome!) distractions that kept me sane through this roller coaster ride. This thesis is dedicated to them. My thanks to Mak, Abah, Mama, Baba & other family members for their understanding and support. Finally, I would like to express my gratitude to the following institutes. This study would not be possible without the financial support of CARIF, University of Malaya and Ministry of Science Technology & Innovation (E-Science #02-04-03-SF0001, E-Science #02-04-03-SF0002).

LIST OF CONTENTS

	Page
COVER PAGE	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
LIST OF CONTENTS	v
LIST OF FIGURES	xi
LIST OF TABLES	xiii
ABBREVIATIONS	xv
LIST OF APPENDICES	xvii
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 PURPOSE OF STUDY	3
CHAPTER 3 LITERATURE REVIEW	4
3.1. Oral Cancer	4
3.2. Epidemiology	4
3.2.1. World Wide	4
3.2.2. Malaysia	8
3.3. Risk Factors for Oral Cancer	8
3.3.1. Smoking	9
3.3.2. Alcohol Consumption	9
3.3.3. Betel Quid Use	10

3.3.4.	HPV Infection	12
3.3.5.	Others	12
3.4.	Geographical Variation in Risk Factors	13
3.5.	Geographical Variation in the Prevalence of Oral Cancer	
	Subsites	17
3.6.	Challenges in Oral Cancer Management	19
3.7.	Global Gene Expression Studies in Cancer	20
3.7.1.	Challenges in Conducting Microarray	
	Experiment	22
3.8.	Types of Biological Specimens for Gene Expression	
	Microarray Studies	25
3.9.	Gene Expression Studies in Oral Squamous Cell	
	Carcinoma (OSCC)	27
3.9.1.	Evidence of Molecular Differences in OSCC	
	Subsites	29
3.9.2.	Clinical Differences in OSCC Subsites	30
3.10.	Importance in Addressing the Heterogeneity in the	
	Different Oral Subsites in OSCC	32
CHAPTER 4 MATERIALS & METHODS		34
4.1.	Study Design	34
4.2.	Study Specimens	34
4.3.	Specimen Selection and Processing	36
4.3.1.	Formalin-Fixed Paraffin-Embedded (FFPE)	
	Specimens	36

4.3.2.	Fresh Frozen Specimens	39
4.3.3.	Cell Lines	39
4.4.	RNA Extraction	40
4.5.	RNA Quantitation and Quality Control	40
4.5.1.	RNA Quality Control for FFPE Specimens	40
4.6.	cDNA Synthesis from Extracted RNA	41
4.7.	Microarray Experiments	41
4.8.	Microarray Data Analysis	42
4.8.1.	Identification of Genes Differentially Expressed Between OSCC and Non-cancerous Oral Mucosa	43
4.8.2.	Identification of Differentially Expressed Genes in OSCC and Non-cancerous Oral Mucosa of Specific Sites	43
4.8.3.	Identification of Significantly Altered Pathways in OSCC	43
4.8.4.	Comparison of Significant Pathways Commonly Changed and Those That Are Site Specific to OSCC	44
4.9.	Validation of Microarray Results	44
4.9.1.	Quantitative Polymerase Chain Reaction (qPCR)	44
4.9.2.	Immunohistochemistry (IHC)	45
4.10.	Developing an <i>In Vitro</i> Model to Study the Function of the FOLR1 in OSCC	46

4.11.	Using an <i>In Vitro</i> Model to Perform Functional Assays to Determine the Role of FOLR1 in OSCC	47
4.11.1.	Cell Proliferation Assay	47
4.11.2.	Cell Migration Assay	47
4.11.1.	Cell Invasion Assay	48
CHAPTER 5 RESULTS		52
5.1.	Quantity and Quality of RNA extracted from FFPE Specimens	52
5.1.1.	RNA Extraction	52
5.1.2.	RNA Quality Control	52
5.1.3.	Factors Associated with the Quality of RNA Extracted from FFPE Specimens	53
5.2.	Quality of Microarray Data Generated from RNA Extracted from FFPE Specimens	54
5.2.1.	Gene Expression Pattern of OSCC and Non-cancerous Oral Mucosa	56
5.3.	Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC	58
5.3.1.	Validation of the Expression of Genes Implicated in OSCC	61
5.4.	Analysis of Gene Expression Variation in OSCC from the Different Sites of the Oral Cavity	64

5.5.	Identification of Differentially Expressed Pathways and Genes Important in OSCC from Distinct Sites	67
5.5.1.	Validation of Site Enriched Pathways and Genes Identified among the Different Sites of OSCC	83
5.6.	Development of Oral Cancer Cell Lines That Over-Expresses Putative Oral Cancer Gene <i>FOLR1</i>	86
5.7.	Determining the Role of Putative Oral Cancer Gene <i>FOLR1</i> in OSCC using Genetically Modified Cell Lines	87
5.7.1.	The Role of FOLR1 in Cell Proliferation	87
5.7.2.	The Role of FOLR1 in Cell Migration	89
5.7.3.	The Role of FOLR1 in Cell Invasion	89
CHAPTER 6 DISCUSSION		91
6.1.	Quantity of RNA Obtained from FFPE Specimens	91
6.2.	Quality of Microarray Data Generated from RNA Extracted from FFPE Specimens	92
6.3.	Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC	93
6.4.	Gene Expression Variation in OSCC from Cheek, Gum and Tongue	95
6.5.	Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC from Distinct Sites	101
6.5.1.	Common Pathways and Genes in OSCC Derived from Cheek, Gum and Tongue.	101

6.5.2.	Site Specific Enriched Pathways and Genes in OSCC Derived from the Cheek, Gum and Tongue	102
6.6.	Determining the Role of FOLR1 in OSCC using Genetically Modified Cell Line	109
6.6.1.	The Role of FOLR1 in Cell Proliferation	109
6.6.2.	The Role of FOLR1 in Cell Migration and Invasion	111
6.7.	Limitation of Study	113
CHAPTER 7 CONCLUSION AND RECOMMENDATIONS		117
7.1.	Conclusion	117
7.2.	Recommendations for Future Work	118
BIBLIOGRAPHY		121

LIST OF FIGURES

Page

Figure 3.1 Cancer incidence worldwide based on ASR (W) rate per 100,000 estimated by GLOBOCAN 2008 in (a) both gender and (b) among males and females respectively (data obtained from GLOBOCAN.iarc.fr).....	5
Figure 3.2 Number of oral cancer case by subsites reported in government hospital in Malaysia in 2003-2005.....	19
Figure 3.3 Kaplan Meier survival curve of oral cancer demonstrating the different survival rates of patients with OSCC from different sites.	31
Figure 4.1 Project workflow.....	35
Figure 4.2 Images depicting how a specific area in a specimen was selected.	38
Figure 5.1 Percentage of genes with significant intensity generated from microarray compared to a) Ct difference b) age of FFPE specimens.	55
Figure 5.2 Unsupervised clustering of OSCC and non-cancerous oral mucosa through (a)Principal Component Analysis and (b) Hierarchical Clustering.	57
Figure 5.3 Differentially expressed genes validated through qPCR using independent samples of fresh frozen OSCC..	61
Figure 5.4 Representative images of OSCC and non-cancerous oral mucosa immunohistochemically stained with BCL2A1, ITGB4 and MMP1.....	63
Figure 5.5 PCA demonstrating the clustering pattern of OSCC from different sites of the oral cavity.....	64

Figure 5.6 The percentage of OSCC from the different sites having the highest standard deviation across the 51 gene chosen at random.....	65
Figure 5.7 Using Venn Diagram to compare genes and pathways among the different sites of OSCC..	67
Figure 5.8 Venn diagram showing (a) number of genes and (b) number of de-regulated pathways which are site enriched and common in different sites of OSCC.....	74
Figure 5.9 Representative images of OSCC and non-cancerous oral mucosa immunohistochemically stained with CD3 and FOXP3.	84
Figure 5.10 Validation of FOLR1 expression was determined by qPCR in 31 fresh frozen OSCC samples from cheek, gum and tongue.	86
Figure 5.11 Development of OSCC cell line overexpressing FOLR1.....	87
Figure 5.12 Determining the role of FOLR1 in cell proliferation and migration..	88
Figure 5.13 Determining the role of FOLR1 in cell invasion.	90

LIST OF TABLES

Page

Table 3.1 Incidence and mortality in oral cancer including the lips in 2008 (adapted from GLOBOCAN.iarc.fr).....	6
Table 3.2 Estimated Age-Standardized Incidence Rates of Oral Cancer Per 100,000 by World Area based on GLOBOCAN 2008 (http://GLOBOCAN.iarc.fr/).....	7
Table 3.3 Composition of different chewing substance (IARC, 2004).....	11
Table 5.1 Factors associated with RNA concentration in FFPE specimens.	53
Table 5.2 Genes differentially expressed in OSCC compared to non-cancerous oral mucosa.	59
Table 5.3 Significant de-regulated pathways in OSCC generated by comparing the T vs. N list to the pathways database in DAVID.	60
Table 5.4 Differentially expressed genes validated at mRNA level through qPCR.	62
Table 5.5 IHC scoring for BCL2A1, ITGB4 and MMP1 in OSCC and non- cancerous oral mucosa.	62
Table 5.6 Standard deviation of 51 genes randomly chosen for OSCC from cheek (B), gum (G) and tongue (T).	66
Table 5.7 B vs. NB gene list. Differentially expressed genes in cheek OSCC compared to non-cancerous oral mucosal samples from the cheek.	68
Table 5.8 De-regulated pathways in cheek OSCC generated by comparing the B vs. NB list to the pathways database in DAVID.	69

Table 5.9 AM vs. NG gene list. Differentially expressed genes in gum OSCC compared to non-cancerous oral mucosal samples from gum.	70
Table 5.10 De-regulated pathways in gum OSCC, generated by comparing AM vs. NG list to the pathways database in DAVID.....	71
Table 5.11 T vs. NT gene list. Differentially expressed genes in tongue OSCC compared to non-cancerous oral mucosal samples.	73
Table 5.12 De-regulated pathways in tongue OSCC, generated by comparing the T vs. NT list to the pathways database in DAVID.....	73
Table 5.13 Differentially expressed genes and pathways which were enriched all 3 sites of OSCC – cheek, gum and tongue.....	75
Table 5.14 De-regulated genes and pathways that were enriched in cheek and gum OSCC.	76
Table 5.15 Differentially expressed genes which were enriched in a. OSCC from the cheek and tongue b. OSCC from gum and tongue.	79
Table 5.16 Differentially expressed genes and pathways which were enriched in OSCC from cheek only.	80
Table 5.17 Differentially expressed genes and pathways which were enriched in OSCC from the gum only.....	81
Table 5. 18 Differentially expressed gene which were enriched in OSCC from the tongue only.	82
Table 5.19 Mean per high power field of CD3, FOXP3 and percentage of FOXP3 in OSCC from cheek, gum and tongue and non-cancerous oral mucosa.....	85

ABBREVIATIONS

ALL	Acute Lymphocytic Leukemia
AM	Gum
AML	Acute Myelogenous Leukemia
APS	Ammonium Persulphate
ASR	Age Standardised Ratio
B	Cheek
BSA	Bovine Serum Albumin
cDNA	Complementary Deoxyribonucleic Acid
CO ₂	Carbon Dioxide
Ct	Cycle Threshold
DAB	Diaminobenzidine
DASL	cDNA Mediated Annealing, Selection, Extension And Ligation Assay
DAVID	Database for Annotation, Visualization And Integrated Discovery
dH ₂ O	Distilled Water
DMEM:F12	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediaminetetraacetic Acid
FBS	Fetal Bovine Serum
FC	Fold Change
FDR	False Discovery Rate
FFPE	Formalin Fixed Paraffin Embedded
FOM	OSCC originating from Floor of Mouth
G	OSCC Originating from Gum
g	Relative Centrifugal Force
g/L	Gram per Liter
GEO	Gene Expression Omnibus
gm	Gram
GO	Gene Ontology
HC	Hierarchical Clustering
HPF	High Power Field
HPV	Human Papilloma Virus
I8	Gingival Tissues Obtained During the Surgical Removal of Impacted Wisdom Tooth
IARC	International Agency for Research on Cancer
ICD	International Classification Of Disease
IHC	Immunohistochemistry
kD	Kilodalton
KEGG Pathways	Kyoto Encyclopedia of Genes and Genomes
M	Molar

mA	Milliamp
mg	Milligram
MgCl ₂	Magnesium Chloride
ml	Milliliter
ml/100 ml	Milliliter per 100 Mililiter of Solution
ml/cm ³	Milliliter per Centimeter Cubed
mM	Milli Molar
mm	Millimeter
mm ²	Millimeter Squared
μM	Micro Molar
μg/ml	Microgram mer Mililiter
μl	Microliter
μl/L	Microliter per Liter
μm	Micron
NB	Non-cancerous Oral Mucosa Originating from Cheek
NCBI	National Center For Biotechnology Information
NCOM	Non-cancerous Oral Mucosa
NG	Non-cancerous Oral Mucosa Originating from Gum
ng	Nanogram
ng/μl	Nanogram per Microliter
NK	Natural Killer Cells
nm	Nanometer
NT	Non-cancerous Oral Mucosa Originating From Tongue
OSCC	Oral Squamous Cell Carcinoma
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pg	Pictogram
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RT buffer	Reverse Transcriptase Buffer
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SOC	Super Optimal Broth with Cabolite Repression
Std dev	Standard Deviation
T	Tongue
TBS	Tris Buffered Saline
T-cell	T-Lymphocytes
TMaA	Tissue Macroarray
Tregs	T Regulatory Lymphocytes
WHO	World Health Organization

LIST OF APPENDICES

	Page
Appendix A Sociodemographics of OSCC patients where FFPE specimens were obtained and details of the RNA quality and quantity	142
Appendix B Sociodemographics of OSCC patients where the fresh frozen samples were obtained and qPCR results of the genes selected for validation.....	146
Appendix C Sociodemographics of OSCC patients where the TMaA samples were obtained and immunohistochemistry results of proteins selected for validation	149
Appendix D Experimental protocols used in the study.....	154
Appendix E Preparation of buffers and solutions	166
Appendix F List of primers and antibodies used in the study.....	172
Appendix G Formula for calculation of cell doubling and proliferation ratio	173
Appendix H OSCC risk habits data from OSCC patients included in the gene expression analysis	174