THE UNIVERSITY
of EDINBURGH

The use of multiple platform “omics” datasets to define new biomarkers in oral cancer and to determine biological processes underpinning heterogeneity of the disease

Anas A M Saeed

Thesis submitted for the degree of Doctor of Philosophy

Edinburgh Dental Institute
College of Medicine and Veterinary Medicine
University of Edinburgh

April 2013

Contents

Table of Contents

Declaration ..... **xi**

Acknowledgements ..... **xii**

List of Abbreviations ..... **xiii**

Abstract ..... **xix**

Chapter 1: Introduction **1**

1.1 Oral cancer: Disease and diagnosis 1

1.2 Oral cancer pathology (development and metastasis) 4

1.3 Aetiology and risk factors of oral cancer 6

1.4 Multiplicity of gene expression in human 8

1.5 Tumour heterogeneity of the oral cavity 9

1.5.1 Oral cancer in the UK and western countries 9

1.5.2 Oral cancer in Sri Lanka and Indo-Asian sub-continent 9

1.6 Cancer as a metabolic disease 10

1.7 Biomarkers as diagnostic tool for HNSCC and OSCC 13

1.8 Metabolomics as tumour biomarkers 14

1.9 Microarray technology 17

1.10 Transcriptomic analysis 18

1.11 Selection of bioinformatics approaches 19

1.11.1 Bioconductor 21

1.11.2 Data filtering and normalization 21

1.11.3 Corrections of batch effects 22

1.11.4 Differential gene expression analyses 22

1.11.5 Biological annotations of significant genes and metabolites 23

1.12 Metabolomics and metabonomics 24

1.12.1 NMR metabolomics/metabonomics techniques 29

1.12.2 Points to be considered using NMR spectroscopy for metabolomic analysis 33

1.12.3 Data analysis and interpretation of metabolomics 35

1.13 Integrative Analysis of transcriptomics and metabolomics 36

Chapter 2: Common gene expression profiling reveals biological alterations involved in early tumourigenesis of Oral Squamous Cell carcinomas **42**

2.1 Introduction 42
2.2 Aim of the study 43
2.3 Materials and methods 43
2.3.1 Selection of study samples 43
2.3.2 Construction of the list of common genes 44
2.3.3 Exploration of the common genes expression in the raw data 45
2.3.4 Biological Annotations of significant common genes 45
2.4 Results 46
2.4.1 Gene expression profile of OSCC versus normal oral mucosa 49
2.4.2 Exploring the set of common genes in a raw data 53
2.4.3 Biological interpretation of the common expressed genes 55
2.5 Discussion 63
Chapter 3: Detection of a robust gene signature for oral squamous cell carcinoma by integrating multiple microarray datasets 67
3.1 Introduction 67
3.2 Aims and objectives of the study 68
3.3 Materials and Methods 68
3.3.1 Selection of Data Sets 69
3.3.2 Data processing and analysis 71
3.3.3 Integration of datasets and batch correction 71
3.3.4 Differential gene expression analyses 72
3.3.5 Biological annotations of significant genes 73
3.4 Results 73
3.4.1 Filtering and processing of the datasets 73
3.4.2 Integration of datasets and correction of batch effect 76
3.4.3 Verification of data integration 77
3.4.4 Gene expression profile of OSCC versus normal oral mucosa 77
3.4.5 Prediction models 86
3.4.6 Biological interpretation of the differentially expressed meta-genes 96
3.4.7 Comparison between meta-genes and common genes of the review study 106
3.5 Discussion 109

Chapter 4: Gene expression profiling reveals biological pathways responsible for phenotypic heterogeneity between UK and Sri Lankan oral squamous cell carcinomas 115

4.1 Introduction 115

4.2 Aims and objectives of the study 115

4.3 Materials and Methods 116

4.3.1 Sample characteristics and biopsy specimens 116

4.3.2 RNA extraction / cDNA synthesis, processing, labeling, and hybridization to microarray chip 120

4.3.3 Data quality control, pre-processing and analysis 120

4.3.4 Quality checks of microarray GeneChip results (QC) 121

4.3.5 Differential gene expression analyses 123

4.3.6 Biological interpretation of the expressed genes 124

4.3.7 Quantitative real-time PCR analysis (validation of gene expression level) 125

4.4 Results 125

4.4.1 Study populations 125

4.4.2 Report of microarray quality control 127

4.4.3 Filtration and pre-processing data 132

4.4.4 Gene expression profile of OSCC versus normal oral mucosa from both population groups 132

4.4.5 Gene expression profile of tumour versus normal samples from UK 135

4.4.6 Gene expression profile of tumour versus normal samples from Sri Lanka 148

4.4.7 Common gene expression profile of tumour versus normal samples from both UK and Sri Lanka 156

4.4.8 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK 163

4.4.9 Gene expression profile of high-stage versus low- stage tumour samples from both UK and Sri Lankan tumours 164

4.4.10 Gene expression profile between UK and Sri Lankan tumour samples 166

4.4.11 Gene expression profile between UK and Sri Lankan normal mucosa 172

4.4.12 qPCR validation of Affymetrix GeneChip microarray gene expression data 179

4.5 Discussion 181

4.5.1 Choice of normal comparator 181

4.5.2 Data uploading, Functional, Networks, Canonical pathways analysis using IPA system 182
4.5.3 Gene expression profile of tumour versus normal samples from UK..... 182
4.5.4 Gene expression profiling of Sri Lankan OSCCs compared to normal oral tissues 183
4.5.5 Common gene expression profile of tumour versus normal samples from both UK and Sri Lanka 184
4.5.6 Gene expression differences between UK and Sri Lankan tumour samples 186
4.5.7 Gene expression differences between UK and Sri Lankan normal mucosa and its impact on the observed phenotypic differences of the two tumour populations 187
4.5.8 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK..... 189
4.5.9 Gene expression profile of high-stage versus low- stage tumour samples from both UK and Sri Lankan tumours 189
4.6 Conclusion..... 191
Chapter 5: Biological pathway integration of significant metabolomic and transcriptomic signatures of OSCC 193
5.1 Introduction 193
5.2 Aims and objectives of the study 194
5.3 Materials and Methods 195
5.3.1 Construction of common metabolites list 195
5.3.2 Biological annotations of the metabolic profile 195
5.3.3 Integrating metabolomic and transcriptomic profiles 196
5.4 Results 196
5.4.1 Construction of the metabolic profile set 196
5.4.2 Biological interpretation of the expressed metabolites 199
5.4.3 Integrating metabolomic and transcriptomic profiles 204
5.5 Discussion 208
5.5.1 Review study of metabolomics 208
5.5.2 Metabolomics analysis of both meta-genes and the 45 metabolites 212
5.5.3 Direct integration between the 45 metabolomic profile and the 32 transcriptomic biomarkers of OSCC and HNSCC..... 212
5.6 Summary 220
Chapter 6: Summary and Discussion..... 223

6.1 Rational and objectives of the project 223
6.2 Summary of the findings 224
6.2.1 Common gene expression profiling reveals biological alterations involved in early tumourigenesis of oral squamous cell carcinomas (Chapter 2) 224
6.2.2 Detection of a robust gene signature for oral squamous cell carcinoma by integrating multiple microarray datasets (Chapter 3) 226
6.2.3 Gene expression profiling reveals biological pathways responsible for phenotypic heterogeneity between UK and Sri Lankan oral squamous cell carcinomas (Chapter 4) 228
6.2.4 Evidences support theory that cancer is a metabolic disease 229
6.2.5 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK 230
6.2.6 Gene expression profile of high-stage versus early and low- stage tumour samples from both UK and Sri Lankan tumours 231
6.2.7 Review study of metabolic profile 232
6.2.8 IPA metabolomics analysis of both meta-genes and the 45 metabolites 233
6.2.9 Direct integration between the 45 metabolomic profile and the 32 OSCC and HNSCC transcriptomic biomarkers 234
6.3 Suggestion for the future work 236
Supplementary Materials 237
References 238

List of Figures

Figure 1.1: Numerous components must be integrated to study the molecular basis of human cancer 5
Figure 1.2 : A schematic view of metabolomics 26
Figure 1.3: Three major steps of metabolomics analysis 35
Figure 1.4: Plot of CYP2D6 genotype vs. log10 (DXT MR) for 229 individuals 37
Figure 1.5: Two methods for cross correlating data from different ‘omic’ approaches 39
Figure 2.1: Hierarchical clustering displaying common genes overlapped over raw data of UK tumour versus normal samples. 54

Figure 2.2: Molecular cellular functions of common genes related to OSCC using IPA functional annotation analysis. 55

Figure 2.3: Central area of the comprehensive molecular network of the common genes related to cancer. 57

Figure 2.4: Master molecular network of the common genes related to OSCC and head and neck cancer. 58

Figure 2.5: Highly involved biological pathway in OSCC, The Role of Tissue Factor in Cancer Pathway; B- Intracellular TF in metastasis..... 61

Figure 2.6: Common genes biomarkers related to OSCC and HNSCC..... 62

Figure 3.1: Filtration of probe sets of both tumour and normal samples..... 74

Figure 3.2: Density and Q-Q plot of normalized genes before filtration method. 75

Figure 3.3: Density and Q-Q plot of normalized genes after filtration method. 76

Figure 3.4: Unsupervised Pearson correlation hierarchical clustering of merged 18394 normalized genes before ComBat integration, discriminating normal oral mucosal samples (N=95) from OSCC samples (N=270). 78

Figure 3.5: Unsupervised Pearson correlation hierarchical clustering of 12092 normalized genes after ComBat integration, discriminating normal oral mucosal samples (N=95) from OSCC samples (N=270). 79

Figure 3.6: Hierarchical clustering of differentially expressed genes separating OSCC tumour (n=236) and both pure oral normal (N=95) and marginal normal oral samples (n=34) using SAM supervised analysis at FDR=0. 81

Figure 3.7: Hierarchical clustering of differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis at FDR=0. 85

Figure 3.8: Hierarchical clustering of top 38 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis. 88

Figure 3.9: Hierarchical clustering of top 8 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis. 89

Figure 3.10: Hierarchical clustering of top 3 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis. 90

Figure 3.11: Box Plots displaying expression of 6 genes (MMP1, LAMC2, PTHLH, TPBG, GPD1L, and MAL) chosen for validation with an external dataset from GEO (GSE6631). 93

Figure 3.12: Molecular cellular functions of Meta-genes related to cancer with the greatest numbers of differentially expressed genes using IPA functional annotation analysis. 97

Figure 3.13: Central area of the comprehensive molecular network of meta-genes related to cancer. 99

Figure 3.14: Master molecular network of meta-genes related to OSCC and head and neck cancer. 100

Figure 3.15: Biomarker network of meta-genes involved in OSCC and head and neck cancer 104

Figure 3.16: Highly expressed meta-genes biomarkers related to OSCC, HNSCC, and Oesophageal cancer. 105

Figure 4.1:Simpleaffy plots displaying the QC of 51 and 48 tumour samples from UK and Sri Lankan population groups before (Left) and after (Right) removal of bad quality microarray chips respectively. 129

Figure 4.2: Simpleaffy plot displaying the QC of 11 and 8 normal samples from UK and Sri Lankan population groups before (top) and after (bottom) removal of bad quality microarray chips respectively. 131

Figure 4.3: Unsupervised Pearson correlation hierarchical clustering of all normalized genes clearly discriminating normal oral mucosal samples (N=8) from OSCC samples (N=48) of both UK and Sri Lankan populations. 133

Figure 4.4: Hierarchical clustering of the 1222 most differentially expressed genes separating UK and Sri Lankan normal oral mucosa (N=8) and patient OSCC specimens (N=48) using SAM supervised analysis at FDR≤0.001. 134

Figure 4.5: Hierarchical clustering of the 896 most differentially expressed genes separating 5 UK normal oral mucosal and 21 patient OSCC specimens using SAM supervised analysis at FDR≤0.001. 137

Figure 4.6: Hierarchical clustering of the 185 most differentially expressed genes separating 5 UK normal oral mucosal and 21 OSCC specimens using SAM supervised analysis at FDR=0.00 138

Figure 4.7: Master molecular network of significant genes of tumour samples from UK. 141

Figure 4.8: Hierarchical clustering of the1065 most differentially expressed genes separating 3 Sri Lankan normal oral mucosal and 27 patient-paired OSCC specimens using SAM supervised analysis at FDR≤0.001. 149

Figure 4.9: Master molecular network of significant genes of tumour samples from Sri Lanka. 155

Figure 4.10: Master molecular network of common genes existed in both UK and Sri Lankan tumour samples. 160

Figure 4.11: Hierarchical clustering of three differentially expressed genes separating 6 perineural invasive OSCC and 15 non-invasive OSCC samples using SAM supervised analysis at FDR≤0.05 163

Figure 4.12: Hierarchical clustering of 12 differentially expressed genes discriminating high stage (N=15) from low stage tumours (N=33) of UK and Sri Lankan OSCCs using SAM supervised analysis at FDR≤0.05 165

Figure 4.13: Hierarchical clustering of 21 differentially expressed genes separating Sri Lankan (N=27) and UK (N=21) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05. 167

Figure 4.14: Hierarchical clustering of 12 differentially expressed genes separating Sri Lankan (N=27) and UK (N=21) OSCC samples using SAM supervised analysis at FDR≤0.01 168

Figure 4.15: Hierarchical clustering of 11 differentially expressed genes separating "High Stage" Sri Lankan (N=3) and UK (N=12) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05 170

Figure 4.16: Hierarchical clustering of 11 differentially expressed genes separating "Low Stage" Sri Lankan (N=24) and UK (N=9) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05 171

Figure 4.17: A. Unsupervised Pearson correlation hierarchical clustering of all normalized genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples. B. Supervised Pearson correlation hierarchical clustering of the 205 most differentially expressed genes using SAM supervised analysis at FDR≤0.01. 173

Figure 4.18: Hierarchical clustering of 1439 differentially expressed genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples using SAM supervised analysis at FDR≤0.01. 176

Figure 4.19: Hierarchical clustering of 16 differentially expressed genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples using SAM supervised analysis at FDR≤0.005. 177

Figure 4.20: Box plots displaying results of qPCR analysis for the genes, CLU, MAL, and PTPN13 chosen for qPCR validation. 180

Figure 5.1: Top molecular network of the altered metabolites in OSCC. 201

Figure 5.2: Master molecular network of the altered metabolites in OSCC. 202

Figure 5.3: Biomarker analysis of common metabolites involved in cancer. 203

Figure 5.4: Top networks of the interactive meta-genes and metabolites associated with significant molecular and cellular functions. 204

Figure 5.5: Master molecular network of meta-genes and metabolites using IPA Metabolomics analysis. 205

Figure 5.6: Interactive network between metabolomic profile and the meta-genes biomarkers of OSCC and HNSCC. 207

List of Tables

Table 2.1: Selected studies using DNA Microarray technology for transcriptional analysis of OSCC and HNSCC. 47

Table 2.2: Common genes significantly higher in OSCC tumours compared to normal oral tissue samples**. 50

Table 2.3: Common genes significantly lower in OSCC tumours compared to normal oral tissue samples.** 51

Table 3.1: Characteristics of the selected datasets samples*. 70

Table 3.2: Upregulated differentially expressed meta-genes in OSCC compared to normal oral tissue samples.* 82

Table 3.3: Downregulated differentially expressed meta-genes in OSCC compared to normal oral tissue samples.* 83

Table 3.4: The percentages of misclassified tumour and normal samples among the prediction models.* 87

Table 3.5: Summary and biological annotations of the 8 genes' model. 94

Table 3.6: Oral, oropharyngeal, and head and neck cancer biomarkers of differentially expressed meta-genes. 102

Table 3.7: Significant genes commonly expressed between meta-genes set and common genes from the review study (chapter 2).* 107

Table 3.8: Distribution and expression of the 8 genes model among external datasets. 110

Table 4.1: Factors contributing to the clinical staging of OSCC. 116

Table 4.2: Patient and tumour related factors in the UK cohort of oral cancers. 117

Table 4.3: Patient and tumour related factors in the Sri Lankan cohort of oral cancers 118

Table 4.4: Site of origin of normal oral mucosa samples from both population groups 119

Table 4.5: Significantly differentially expressed genes discriminating UK tumour from normal samples. * 136

Table 4.6: Significantly differentially expressed genes discriminating Sri Lankan tumour from normal samples.* 150

Table 4.7: The common significantly differentially expressed genes discriminating tumour from normal samples of both UK and Sri Lankan population groups. * 157

Table 4.8: List of the top upregulated and downregulated genes differentially expressed between normal samples of both UK and Sri Lankan groups at FDR level ≤ 0.01.* 174

Table 4.9: List of the top upregulated and downregulated genes differentially expressed between normal samples of both UK and Sri Lankan groups at FDR level ≤ 0.005.* 175

Table 5.1: Characteristics of metabolomic studies. 197

Table 5.2: 45 metabolites with altered expression in OSCC and HNSCC.* 199

Declaration

I declare that this thesis is of my own composition, except where otherwise stated in the text, and that it contains no material previously submitted for any other degree or personal qualification.

Anas A M Saeed

Acknowledgements

First and foremost, I am greatly indebted to my Lord in helping me to complete this study. A considerable number of people helped me during this project. First of all I would like to thank my supervisor Prof Victor Lopes for his guidance all along the difficult ways of my PhD. All thanks and respect to Dr Howard Moody, my previous principal supervisor.

Special thanks to Andy Sims who has taken a considerable amount of his time to supervise the whole bioinformatics work throughout this PhD. My thanks also go to all members of Edinburgh Cancer Research UK Centre for their kind welcoming and cooperation during my access and work.

Many thanks to Prof Peter Ghazal, Dr Douglas Roy, and Dr Kevin Robertson at DPM (Division of Pathway Medicine, University of Edinburgh Medical School), who gave me the facility to use IPA software for many years under their license and for providing me the scientific hospitality for this work section of the thesis.

I would like to thank my wife, Dr Abeer, who helped me with the corrections of the manuscript and for all the good and bad times that we went through together.

I'm so grateful to all friends, colleagues, and members at Edinburgh Dental Institute, especially Jackie McGurk, Gillian O'Brien, and all members of the surgery department. I want to thank College of Dentistry / University of Mosul / Ministry of Higher Education and Scientific Research for providing me the scholarship and fund for this project. Particular thanks to the good friends and colleagues; Qudamah Quboa, Ziad Al-Sarraf, and Nouf Laqtom For their help, advice, and support related to statistical analysis of the results.

Finally, I dedicate this thesis to:

My parents, the source of love, support, and strength; **My wife**, Abeer, for her love, encouraging, patience, and work check of the thesis; **My children**, Ahmed, Zahraa, and Noor who have enriched my life.

List of Abbreviations

2D	Two-Dimensional
AGE	Advanced Glycated End Products
CE-TOF-MS	Capillary Electrophoresis Time-Of-Flight Mass Spectrometry
FGFR2	Fibroblast Growth Factor Receptor-2
iNOS	Inducible Nitric Oxide Synthase
PRR	Pro-Renin Receptor
RAGE	AGE Receptor
RT-PCR	Reverse Transcription-PCR
tCho	Total Choline-Containing Compounds
VEGF	Vascular Endothelial Growth Factor
%	Percentage
^{1}H and ^{31}P MRS	Magnetic Resonance Spectroscopy
^{1}H NMR	^{1}H Nuclear Magnetic Resonance
AA	Arachidonic Acid
ADH1B/ALDH2	Alcohol Dehydrogenase 1B / Aldehyde Dehydrogenase 2
ADP-ribose	Adenosine Diphosphate Ribose
AJCC/UICC TNM	American Joint Committee on Cancer / International Union Against Cancer. Tumour-Node-Metastasis classification.
AKR1A1	Aldo-Keto Reductase Family 1
Akt	Serine/Threonine Protein Kinase
ApoE	Apo Lipoprotein E
ATF-2	Activating Transcription Factor-2
Bp	Base Pairs
cAMP	Cyclic Adenosine Monophosphate
Cdc42-MRCK MRCK Beta	Cell Division Control Protein 42 / Serine/Threonine-Protein Kinase
cDNA	Complementary DNA

The ICD9 codes;

Code140	Lip
Code141	Tongue
Code142	Major Salivary Glands
Code143	Gum
Code144	Floor of Mouth
Code145	Other Non-Specific Sites
Code146	Oropharynx
Code147	Nasopharynx
Code148	Hypopharynx
CRABP2	Cellular Retinoic Acid Binding Protein 2
cRNA	Complementary RNA
CYP2E1	Cytochrome P450 2E1
D2O	Heavy Water 2H2O
DCA	Deoxycholic Acid
DHHS/USDA	U.S. Department Of Health And Human Services
DNA	Deoxyribonucleic Acid
DWT	Dyadic Discrete Wavelet Transforms
EBV	Epstein-Barr Virus
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EIF2 Signalling	Eukaryotic Initiation Factor 2
ELISA	Enzyme-Linked Immuno Sorbent Assay
EMT	Epithelial Mesenchymal Transition
ENO2	Enolase2
ENSEMBL IDs	European Molecular Biology Laboratory Identification Numbers
EOC	Epithelial Ovarian Cancer

ER	Oestrogen Receptor
ERK1/2	Extracellular Signal-Regulated Kinase
ERK-MAPK	Extracellular Signal-Regulated Kinases / (Mitogen-Activated Protein Kinases)
ESTs	Expressed Sequence Tags
FAK	Focal Adhesion Kinase
FC	Fold Change
FDR	False Discovery Rate
FID	Free Induction Decay
FOS	Part of AP-1 Complex
GABA	Gamma-Amino Butyric Acid
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
gcRMA	GeneChip-Robust Multi-Array Average Algorithm
GEO	Gene Expression Omnibus
GO	Gene Ontology
GPD1L	Glycerol-3-Phosphate Dehydrogenase 1-Like
GPX3	Glutathione Peroxidase-3
GSH	Glutathione
GST	Glutathione-S-Transferase
h5T4	Human Oncofetal Antigen
HG Focus platform.	Human Genome Focus Platform.
HIF1α	Hypoxia-Inducible Factor 1 Alpha
HMDB	Human Metabolome Database
HNSCC	Head And Neck Squamous Cell Carcinomas
HPLC-MS	High performance Liquid Chromatography–Mass Spectrometry
HSV	Herpes Simplex Virus
HPV	Human Papilloma Virus
HR-MAS NMR	High performance Magic Angle Spinning Techniques Nuclear

	Magnetic Resonance
HSCs	Hepatic Stellate Cells
IFN	Interferon
IFNγ	Interferon-Gamma
IL	Interleukin
ILK	Integrin-Linked Kinase
IPA	Ingenuity Pathway Analysis
JNK	C-Jun N-Terminal Kinase
KEGG	Kyoto Encyclopaedia Of Genes And Genomes
KHSRP	Kh-Type Splicing Regulatory Protein
LCM	Laser Capture Micro Dissection
LOD	Logarithm Of The Odds
LXR/c	The Liver X Receptor/ Retinoic Acid Receptor
M2-PK	M2 Pyruvate Kinase
MAPK	Mitogen-activated protein kinases
MM	Mismatch
MRA	Multi-Resolution Analysis
MRCK complex	Myotonic Dystrophy Kinase-Related Cdc42-Binding Kinase
mRNA	Messenger RNA
MRSI	Magnetic Resonance Spectroscopy Imaging
MS	Mass Spectrometry
mtDNA	Mitochondrial DNA
MYC	V-Myc Myelocytomatosis Viral Oncogene Homolog
N+	Lymph Node Positive
N0	Lymph Node Negative
NAD(P)H	Nicotinamide Adenine Dinucleotide
NCBI GEO Omnibus	National Centre Biotechnology Information Gene Expression
NCBI RefSeq	National Centre For Biotechnology Information Reference

	Sequence
NFAT5	Nuclear Factor Of Activated T-Cells 5
NFκB	Nuclear Factor-Kappa Beta
NMR	Nuclear Magnetic Resonance Spectroscopy
NRF2	Nuclear Factor (Erythroid-Derived 2)-Like 2
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OSCC	Oral Squamous Cell Carcinoma
OSMF	Oral Sub Mucous Fibrosis
OSM	Oncostatin M
OXPHOS	Oxidative Phosphorylation Pathway
p300 HAT	Histone Acetyltransferase
p38 MAPK	P38 Mitogen-Activated Protein Kinases
PARP	Poly (ADP-Ribose) Polymerase
PGM	Phosphoglycerate Mutase
PI3K–AKT	Phosphatidylinositol-3-Kinase- Serine/Threonine Protein Kinase
PM	Perfect Match
PTP313/FAP1	Protein Tyrosine Phosphatase, Non-Receptor-Type, 13/ Familial Adenomatous Polyposis
Qc	Quality Check
qPCR	Quantitative Polymerase Chain Reaction
RARα	Retinoic Acid Receptor Alpha
RGS5	Regulator of G-Protein Signalling 5
RHOA	Ras Homolog Family Member A
RNA	Ribonucleic Acid
RNase	Ribonuclease
ROS	Reactive Oxygen Species
rRNA	Ribosomal RNA
RTG	Retrograde Response

SAM	Significance Analysis For Microarrays
SCC	Squamous Cell Carcinoma
SNP	Single Nucleotide Polymorphism
STAT1/3	Janus Kinase and (3) Signal Transducer and Activator of Transcription
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TGF-β	Transforming Growth Factor, Beta
TGF-β1	Transforming Growth Factor Beta-1
TIMPs	Tissue Inhibitors of Metalloproteinase
TLR	Toll-Like Receptor
TNF	Tumour Necrosis Factor
tRNAs	Transfer RNAs
TrxR	Thioredoxin Reductase
UPLCQTOFMS	Ultra-Performance Liquid Chromatography Coupled With Quadrupole/Time- Of-Flight Mass Spectrometry
β-Carotene	Beta Carotene

Abstract

Oral cancer in early stages (I and II) may be curable by surgery or radiation therapy alone but advanced stage disease (III and IV) has a relatively low survival rate. The pathogenic pathways that contribute to Oral Squamous Cell Carcinoma (OSCC) remain poorly characterised and the critical factor in the lack of prognostic improvement is that a significant proportion of cancers initially are asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. Hence, a clinically applicable gene expression signature is in high demand and improved characterization of the OSCC gene expression profile would constitute substantial progress. For OSCC, possible themes that might be addressed using microarray data include distinguishing the disease from normal at the molecular level; determining whether specific biomarkers or profiles are predictive for tumour behaviour; and identifying biologic pathways necessarily altered in tumourigenesis, potentially illuminating novel therapeutic targets. However, OSCC is a heterogeneous disease, making diagnostic biomarker development difficult. Although this phenotypic variation is striking when one compares OSCC from different geographic locales, little is known about the underpinning biological mechanisms.

Cancer may be accompanied by the production and release of a substantial number of proteins, metabolites and/or hormones into the blood, saliva, and other body fluids that could also serve as useful markers for assessing prognosis, metastasis, monitoring treatment, and detecting malignant disease at an early stage.

The primary aim of this thesis is to investigate metabolomic and transcriptomic profiles using multiple bioinformatics approaches and biological annotation tools in an attempt to identify specific biomarkers and prediction models for OSCC from each profile as well as from the interface outcomes of integrating the two platforms. Additional aims of the thesis go further to identify the mechanisms underlying the biological changes during tumorigenic transformation of OSCC, as well as to determine biological processes underpinning the heterogeneity of the disease among populations.

Two review studies were carried out in this thesis. The review study of published transcriptomic profiles of OSCC specified several genes and pathways exhibiting substantially altered expression in cancerous versus noncancerous states across studies. However, the result of the review suggests not relying on the final set of genes published by the individual studies, but to access the raw data of each study and start subsequent analysis from that stage using unified bioinformatics approaches to acquire useful and complete understanding of the systems biology. The other review study focused on the metabolic profiles of OSCC and revealed a systemic metabolic response to cancer, which bears great potential for biomarker development and diagnosis of oral cancer. However, the metabolic signature still needs to improve specificity for OSCC from other types of cancer.

In an attempt to detect a robust gene signature of OSCC overcoming the limitation of the transcriptomic review in accessing the raw data from the previous works, four public microarray raw datasets (comprising 365 tumour and normal samples) of OSCC were successfully integrated using ComBat data integration method in R software, determining the common set of genes, biomarkers, and the relative regulatory pathways possibly accountable for tumour transformation and growth in OSCC. Examination of the meta-analysis datasets showed several discriminating gene expression signatures for OSCC relative to normal oral mucosa; with a signature of 8 genes (MMP1, LAMC2, PTHLH, TPBG, GPD1L, MAL, TMPRSS11B, and SLC27A6) exhibiting the best discriminating performance and show potential as a diagnostic biomarker set. In addition, 32 biomarkers specific to OSCC and HNSCC were identified with the majority involved in extracellular matrix (ECM), interleukins, and peptidase activity where around 2/3 of them are located in the extracellular space and plasma membrane.

Additionally, investigation of the interactive network created by merging metabolic and transcriptomic profiles highlighted the significant molecular and cellular biofunctions, pathways, and biomarkers distinguishing OSCC from normal oral mucosa. The results highlighted interactions of significantly altered expression of D-glucose, ethanol, glutathione, GABA, taurine, choline, creatinine, and pyruvate metabolites with the expressed PTGS2, IL1B, IL8, IL6, MMP1, MMP3, MMP9,

SERPINE1, COL1A1, COL4A1, LAMC2, POSTN, ADAM12, CDKN2A, PDPN, TGM3, SPINK5, TIMP4, KRT19, and CRYAB biomarkers of OSCC. Such a pattern may represent a clinically useful surrogate for the presence of OSCC which might help in deciphering some of the obscure multifaceted mechanisms underlying carcinogenesis of OSCC which emerged from dysregulated genetic and metabolic system of the body.

In an attempt to define pathways of importance in two phenotypically different forms of OSCC, transcriptomic analysis of OSCC from UK and Sri Lankan patients was undertaken. The development of OSCCs in UK and Sri Lankan populations appears largely mediated by similar biological pathways despite the differences related to race, ethnicity, lifestyle, and/or exposure to environmental carcinogens. However, results revealed a highly activated "Cell-mediated Immune Response" in Sri Lankan tumour and normal samples relative to UK cohorts which may reflects a role in resistance of patients to invasiveness, metastasis, and mortality observed in Sri Lankan relative to UK patients.

In conclusion, multiple molecular profiles were able to identify a unique transcriptomic profile for OSCC and could further discriminate the tumour from normal oral mucosa on the basis of 8 genes. Altered expression of several metabolic and transcriptomic biomarkers specific for OSCC were identified, along with several dysregulated pathways and molecular processes found common in patient with oral cancer. Integrating both metabolomic and transcriptomic signatures revealed a promising strategy in analysing the concurrent perturbation in both genetic and metabolic systems of the body. Additional results revealed possible impact of specific supplementary dietary components in boosting the immune system of the body against invasion, progression, and metastasis of the disease. Further clinical studies are required to confirm and validate the current results.

Chapter 1: Introduction

1.1 Oral cancer: Disease and diagnosis

Oral squamous cell carcinoma (OSCC) is a sub-site of HNSCC. Tumour of the oral cavity consists 30% of all head and neck cancers. It includes carcinoma arising from the soft mucosal lining of the lips and all intra-oral sites corresponding to the ICD9 codes 140 (lip), 141 (tongue), 143 (gum), 144 (floor of mouth) and 145 (other non-specific sites), but excludes sites 142 (major salivary glands), 146 (oropharynx), 147 (nasopharynx), 148 (hypopharynx) and 149 (ill-defined oral/oropharynx) (Who, 1995). Approximately 90% of oral cancers are primary squamous cell carcinomas arising from the lining mucosa of the mouth, most commonly the tongue and the floor of the mouth (Severino et al., 2008). Tongue cancer is the most frequently reported in the oral cavity with no strong barrier to stop the tumour from spreading. In addition, comparing to another head and neck anatomic sites, the vascular and lymphatic systems of the tongue are well-developed and thus, the rate of cervical lymph node metastasis is high (Ho et al., 1992, Lopes, 2007).

Early stage (I and II) oral cancer may be curable by surgery or radiation therapy alone but advanced cancers (stage III and IV) are generally treated by surgery followed by radiation therapy (Harrison et al., 1999). Using multimodal protocols that combine surgery with pre-operative or post-operative radiotherapy and/or adjuvant chemotherapy, the 2-year and 5-year survival rates for advanced cancers were as low as 20% and 12%, respectively (Reichard et al., 1993). The survival of progressive–stage patients rarely exceed 30 months, even for those that initially achieve complete clinical remission (Hill and Price, 1994).

A critical factor in the lack of prognostic improvement is that significant proportions of cancers initially are asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. Early detection of cancer is the most effective means to reduce death from this disease.

While oral cancer is quite often discovered in a patient during dental treatment, no specific tumour markers that enable the effective and simple detection and diagnosis

of oral cancer are available. In many cases of suspected cancer, a prompt diagnosis at the primary medical institution is not made, and patients are referred for examination to specialists only after the disease has developed to an advanced stage. The result of this delay in diagnosis is a missed opportunity for early treatment that forces the patient to undergo extended treatment (Katakura et al., 2007).

The World Health Organization has clearly identified prevention and preliminary detection as major objectives in the control of oral cancer burden worldwide. Currently, screening of oral cancer and its pre-invasive intra-epithelial stages, as well as its early detection, is still largely based on visual examination of the mouth. A strong available evidence suggest that sight inspection of the oral mucosa is effective in reducing mortality from oral cancer in individuals exposed to risk factors (Fedele, 2009).

Simple visual examination, however, is limited by subjective interpretation and by the potential, albeit rare, occurrence of dysplasia and early OSCC within areas of normal-looking oral mucosa. As a consequence, adjunctive techniques have been suggested to increase ability to differentiate between benign abnormalities and dysplastic/malignant changes as well as to identify areas of dysplasia/preliminary OSCC invisible to naked eye. These include exfoliative cytology, the use of toluidine blue, brush biopsy, chemiluminescence and tissue auto fluorescence (Fedele, 2009). However, intrinsic differences in the optical properties of tissue can be due to changes in absorbance(Piazza et al., 2008) , scatter (Sharwani et al., 2006), or fluorescence (Fedele, 2009), (Roblyer et al., 2008, Pavlova et al., 2008, Schwarz et al., 2008, Mallia et al., 2008), and these changes can result from inflammation, increased cellularity, alter cellular anatomy (e.g., cell shape and nuclear volume), appearance of stromal, and biochemical differences associated with cellular proliferation and tissue reorganization. Even in auto fluorescence screening method, imaging tools based on tissue auto fluorescence cannot reveal the molecular basis of neoplasia with the specificity needed for guiding molecularly targeted drug interventions (Kelloff et al., 2009). Saliva containing exfoliated cells combined with cytospin may be a good approach; however, no specific marker can be used universally for the detection of oral cancer (Kao et al., 2009).

Histopathological examination itself shows weak points in cancer detection and classification, which require searching for more subtle and accurate method for early detection and stage evaluation of oral cancer. The anticipated biologic behaviour of squamous cell carcinoma of the head and neck region has been traditionally estimated by a qualitative system of grading initiated by Broders (Broders, 1926), based on the extent to which the neoplasm resembles the tissue of origin. Now, additional parameters, such as nuclear pleomorphism and the number of mitosis, enter this subjective determination of differentiation. This classification is usually simplified into three categories: poorly differentiated, moderately differentiated, and well-differentiated tumours. The association between the preceding classification and prognosis has proven little significance in squamous cell carcinoma of the head and neck region. A semi quantitative assessment of several histologic parameters would render a higher point value than would result from the degree of differentiation alone (Anneroth et al., 1987, Bryne et al., 1989). Leemans group have shown that about 15% to 30% of stage III and IV patients with histopathologically tumour-free resection margins have recurrence (Leemans et al., 1994). The results of Chen group demonstrated that differentiation, invasion depth, perineural invasion, and lymph vascular permeation significantly affect nodal metastasis in tongue cancer. Therefore, patients in early stages I or II with such pathologic covariates may need more appropriate early neck treatment (Chen et al., 2008b). Moreover, the invasive nature of a biopsy makes histopathology unsuitable for cancer screening in high-risk populations (Li et al., 2004a).

Given the recent progress in the field of molecular biology, the potential exists for finding molecular markers that aid in the early diagnosis and the prediction of prognosis. However, because of the diversity of anatomic sites, as well as genetic, ethnic, and geographic factors, it is unlikely that any universal biological marker will be appointed for squamous cell carcinoma of the head and neck region (Rose et al., 2000). Exploration will be more efficient and successful if appropriate rules of evidence are applied to direct that exploration and determine when results can support strong claims and high expectations (Ransohoff, 2005).

1.2 Oral cancer pathology (development and metastasis)

Oral cancer is considered the result of a multi-hit process, which involves a number of aberrant genetic events. Multiple oncogenes, regulatory factors, and tumour suppressor genes play a key role in the development and progression of this cancer (Vora et al., 2003). Deregulation of many genes like oncogenes and tumour suppressor genes has been associated with oral carcinogenesis (Spandidos et al., 1985). The exact affected nature of each molecule in oral carcinogenesis remains largely elusive (Tsai et al., 2004).

Some genetic alterations can be detected in tissue that appears normal by histological assessment (Jonason et al., 1996, Deng et al., 1996). Conversely extensive genetic changes can be seen at premalignant stages of human tumour development in many tissues (Rehman et al., 1994, Gong et al., 2001, Shaaban et al., 2002). The crucial determinant of progression may be the accumulation of specific combinations of genetic alterations or the occurrence of the mutations in a particular subset of target cells that has higher propensity for malignant progression (Brown et al., 1998). Metastasis is the process by which cancers spread to distinct sites in the body. It is the principal cause of death in individuals suffering from cancer. For some types of cancer, early detection of metastasis at lymph nodes close to the site of the primary tumour is pivotal for appropriate treatment. Because it can be difficult to detect lymph node metastases reliably, many individuals currently receive inappropriate treatment (Roepman et al., 2005). Improving our understanding of cancer and developing theoretical models will require an increased knowledge of the contributions of and interactions between the numerous components contributed to tumour formation and progression (Figure 1.1).

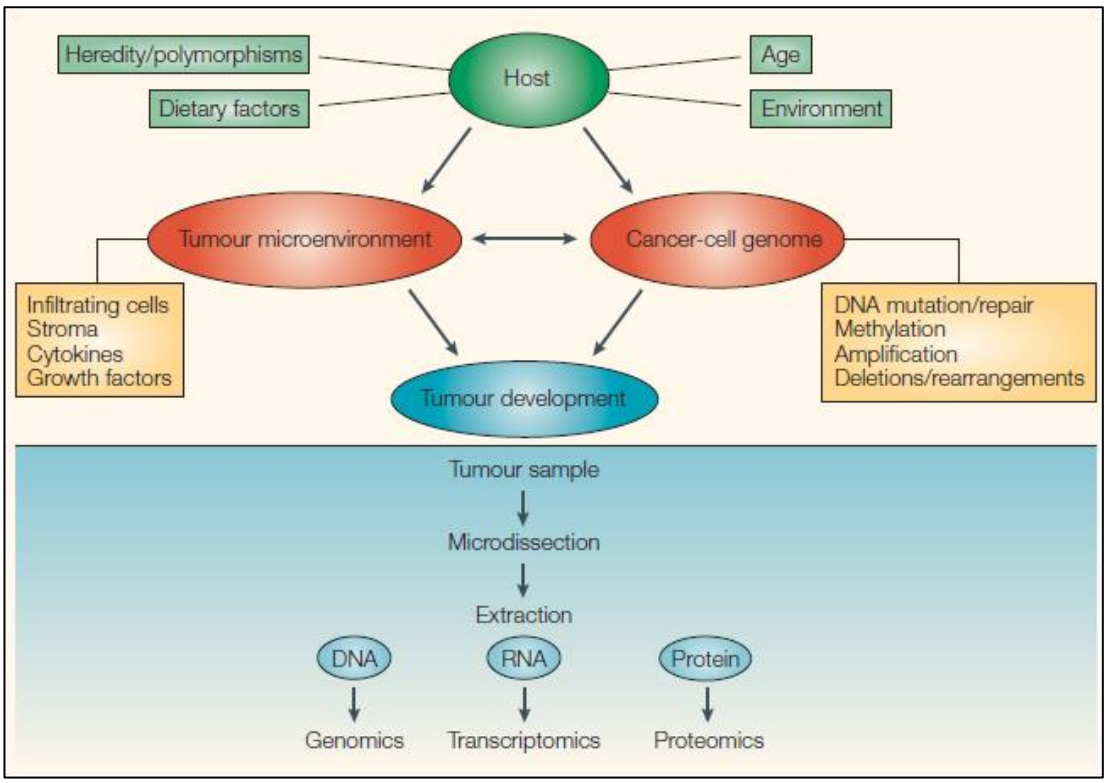


Figure 1.1: Numerous components must be integrated to study the molecular basis of human cancer

Several host factors contribute to tumourigenesis in humans, including diet, environmental factors, polymorphisms and mutations in susceptibility genes, age and immunity. Cells undergo genomic changes (DNA mutations and repair, methylation, amplification, deletions and rearrangements), leading to tumourigenesis. Tumour development also depends on factors in the microenvironment — some of these are produced locally, whereas others are produced systemically (growth factors, infiltrating cells and cytokines). Reciprocal interactions between the premalignant and malignant cells, stromal cells, extracellular-matrix components, various inflammatory cells and a range of soluble mediators therefore contribute to tumour development and progression. Once tumour samples are obtained, genomic, transcriptomic and proteomic tools can be used to profile specific compartments (Adapted from Samir Hanash 2004).

DNA microarray gene expression profiling can detect lymph node metastases for primary head and neck squamous cell carcinomas that arise in the oral cavity and oropharynx (Roepman et al., 2005). Regional lymph node metastasis is an important prognostic factor in head and neck squamous cell carcinoma (HNSCC) and plays a decisive role in the choice of treatment (Hensen et al., 2008).

A study showed that pathways differentially expressed between metastasized and non-metastasized HNSCC are involved in the processes of extracellular matrix remodelling, hypoxia and angiogenesis (Hensen et al., 2008).

Focusing on pathways and functional gene sets instead of individual probes in the analysis of microarray data making biological context of the results more interpretable. Furthermore, a supervised pathway-based analysis reduces multiple testing problems associated with microarray analysis by focusing on a limited number of pathways instead of analysing all the probes available on the microarray. The comparability of results from distinct microarray studies is greatly improved. A supervised, pathway based analysis can reveal biologically relevant similarity between results of various gene-expression studies, even if studies have used distinct microarray platforms with different probes and probe content, where the insight into the biological steps that lead from normal and dysplastic tissue to carcinogenesis in OSCC and HNSCC is improved (Hensen et al., 2008).

Furthermore, removal of the probes that are not relevant to the biological processes of interest reduces the statistical noise and multiple testing problems associated with microarray analysis. The significant advantage of this strategy is increased comparability of data from different microarray studies. Microarray analysis based on individual genes is highly dependent on the exact gene content of the employed microarray in the study and thus on the chosen microarray platform. In a pathway-based analysis, gene expression does not have to be measured from every single gene involved in a particular pathway, as long as a representative subset of genes is assessed. These illustrative subsets of genes involved in a specific pathway may vary between studies. A pathway-based analysis thus can reveal biologically relevant similarity between results of different microarray studies even though the gene contents of the used microarray platforms do not match exactly (Hensen et al., 2008).

1.3 Aetiology and risk factors of oral cancer

Tobacco consumption in all its forms is the common worldwide aetiological risk factor for the development of oral cancer (Blot et al., 1988, Morse et al., 2007, Lambert et al., 2011). In developing countries the use of tobacco and/or areca (betel) nut produces chronic potentially malignant lesions (leukoplakia, erythroplakia and submucous fibrosis) from which most oral cancers arise. The use of tobacco chewing products combined with areca nut, betel leaf palm and lime is a strong cultural

tradition in south India and Sri Lanka. This combination leads to the production of a potent mix of carcinogens as well as mild stimulant substances (Muir and Kirk, 1960). The mixture is formed into a small pouch or "betel quid" and is held in the cheek area of the mouth from where it is slowly absorbed. Conversely, in western developed countries, oral cancers are recognized in only a minority of cases and the majority of lesions developed from clinically normal mucosa. Furthermore, these lesions are more aggressive and followed by a poorer prognosis than those derived from an area of tobacco induced leukoplakia (Lopes, 2007).

In the Western world cigarette smoking is accountable for most tobacco related oral cancers. Strength of tobacco usage is directly related to the risk of developing OSCC (La Vecchia et al., 2004). Alcohol is another significant risk factor influencing development of OSCC and acts synergistically with tobacco. Heavy drinkers and heavy smokers have a 24 fold risk of developing OSCC than non-drinkers and non-smokers (McCoy and Wynder, 1979). Dietary deficiencies or imbalances considered as additional risk factor which can contributed to approximately 15% of OSCC (LaVecchia et al., 1997), particularly β-Carotene and vitamin E deficiency as well as diets with high nitrites and nitrosamines which reported to increase the lifetime risk of developing OSCC.

Many viruses have been aetiologically linked with OSCC including, retroviruses, adenoviruses, Epstein-Barr virus (EBV), Herpes Simplex virus and Human Papillomavirus (HPV). The role of EBV and HPV in OSCC has convincing evidence (Cruz et al., 2000, Ha and Califano, 2004, Syrjanen, 2005).

A small genetic risk associated with OSCC was also reported with first degree relatives of persons with OSCC have a 3.8 fold risk for developing the same disease (Foulkes et al., 1996).

1.4 Multiplicity of gene expression in human

According to Darwin's mechanism, cancer is a consequence of continuous, gradual, and accidental acquirement of the changes in the population of pre-malignant individual cells of genomes. These gradual modifications progressively change the phenotype of the normal cells making them more malignant through loss of their overall stability of genome.

Cancers of diverse organ and tissue systems are regularly considered as different diseases because of the pragmatic dissimilarities in their cell phenotypes and specific prognoses. However, the differences between two types of tumours might be explained by the underlying differences in the tissues of their respective origins. Each normal human tissue has a discrete gene expression pattern from other tissue type, with an exception of constitutively expressing housekeeping genes shared between all the tissues (Patti et al., 2003). Further reports revealed that even the housekeeping genes are not necessarily expressed at the same level across all tissues; rather, each tissue seems to have a specific expression profile of housekeeping genes. In fact, housekeeping genes are less compact and developed than tissue-specific genes, and they evolve more slowly for both coding and core promoter sequences (Zhang and Li, 2004).

Generally, the unique patterns of gene expression in numerous tissue types are elucidated by the dissimilarities in the functions of the resulting proteins in a specific tissue. Furthermore, alternative splicing of many human genes resulted in multiple expression of mRNA transcripts with different sets of exons joined together (Modrek and Lee, 2002). A wide variation in the alternative transcript expression patterns across tissues of human body was reported (Landry and Mager, 2002, Yeo et al., 2004), though their mechanism remains poorly understood (Matlin et al., 2005, Singh and Valcarcel, 2005).

1.5 Tumour heterogeneity of the oral cavity

Oral cancer has a wide heterogeneity in the biological behaviour, however few biological indicators are available to help understanding such patterns of behaviour (Williams, 2000, Woolgar, 2006). Considering tumour heterogeneity in and between patients is required in selecting and interpreting biomarkers to optimize personalized cancer therapy for head and neck cancer patients (Williams, 2010).

OSCC in patients from UK has been described as highly aggressive and metastasising, with early recurrence (less than 12 months) and poor survival. However, a less invasive, non-metastasising form of OSCC is also seen with a better survival. Sri Lankan OSCC is similar to the less aggressive UK phenotype and is exophytic (outward growing) in its growth pattern unlike that in UK, which shows endophytic (inward growing) and invasive behaviour (Lopes, 2007).

1.5.1 Oral cancer in the UK and western countries

The incidence of OSCC in England, Europe, and Scandinavia is reported rising (Bonifazi et al., 2011) especially amongst young women; a population not previously associated with a high incidence of OSCC. In general, UK OSCC behaves as an aggressive, endophytic and deeply invasive tumour that frequently metastasizes to the loco-regional lymph nodes in the neck but rarely to distant organs (Lopes, 2007).

1.5.2 Oral cancer in Sri Lanka and Indo-Asian sub-continent

Oral cancer behaves very differently in patients from the Indo-Asian sub-continent and is typified by that seen in Sri Lanka. Oral cancer is the most common cancer in Sri Lanka (Moore et al., 2010, Moore et al., 2000). In Sri Lanka, oral cancer is [aetiologically] associated with chewing the Betel leaf quid containing Areca nut (Muir and Kirk, 1960, Ranasinghe et al., 1993). Such cancers are phenotypically different from those seen in Western countries and clinically display an exophytic "cauliflower" appearance. The clinically significant feature of OSCC in Sri Lanka that it infrequently metastasizes to the loco-regional lymph nodes of the neck and so most patients do not undergo the procedure of neck dissection (Wardbooth et al., 1982) and

so it is suggested that OSCC in Sri Lanka is less aggressive than that seen in Western countries including the UK (Lopes, 2007).

1.6 Cancer as a metabolic disease

Emerging evidences indicate that impaired cellular energy metabolism is the defining characteristic of nearly all cancers regardless of cellular or tissue origin. In contrast to normal cells, which derive most their usable energy from oxidative phosphorylation, most cancer cells become heavily dependent on substrate level phosphorylation to meet energy demands. Reviewed evidence supporting a general hypothesis that genomic instability and essentially all hallmarks of cancer, including aerobic glycolysis (Warburg effect), can be linked to impaired mitochondrial function and energy metabolism (Seyfried and Shelton, 2010).

According to Warburg (Warburg, 1956), modifications of respiratory machinery should result in compensatory increase in glycolytic ATP production leading to carcinogenesis. In fact, malignant cells meet their energy (ATP) requirements by glycolysis rather than through oxidative phosphorylation, possibly an adaptation to hypoxia development during growing of tumour (Assaily and Benchimol, 2006).

Mitochondrial mutations may contribute to the development of a malignant phenotype by direct effects from increased reactive oxygen species production as well as induction of aerobic glycolysis and growth promotion (Zhou et al., 2007). While normal mitochondrial function can reduce tumourigenesis, various studies show that tumour mitochondria are structurally and functionally abnormal and unable to generate normal levels of energy (Seyfried and Shelton, 2010).

As retrograde (RTG) response is “off” in healthy cells with normal mitochondrial function, It turned “on” following impairment in mitochondrial energy production where transcription and signalling commences for multiple energy and anti-apoptotic related genes and proteins to include MYC, TOR, p53, RAS, CREB, NFκB, and CHOP (Wolfman et al., 2006, Kulawiec et al., 2009, Kulawiec et al., 2008, Singh et al., 2005, Miceli and Jazwinski, 2005, Butow and Avadhani, 2004). The RTG response also takes a part in the participation of multiple negative and positive

regulators, which facilitate the bioenergetics' transition from respiration to substrate level Phosphorylation (Butow and Avadhani, 2004). Most cancer types are similar in expressing mitochondrial dysfunction and elevated substrate level Phosphorylation. When loss of respiration leads to glycolysis, dedifferentiation, and unbridled proliferation (Szentgyorgyi, 1977), it is expected that presence of normal mitochondria in tumour cells would restore the cellular redox status, turn off the RTG response, and reduce or eliminate the need for glycolysis (Warburg effect) and glutaminolysis to maintain viability. Basically, normal mitochondrial function would facilitate expression of the differentiated state thereby suppressing the tumorigenic or undifferentiated state (Seyfried and Shelton, 2010).

Moreover, mitochondria and mitochondrial functions are reported to be involved in apoptotic cell death (Jacobson et al., 1994, Nguyen et al., 1993, Hockenbery et al., 1990, Hennet et al., 1993, Newmeyer et al., 1994, Schulzeosthoff et al., 1992). Recent studies suggested mitochondrial function may play a substantial role during the commitment of cells undergoing apoptosis (Heerdt et al., 1997, Krippner et al., 1996, Liu et al., 1996, Zhang et al., 1996, Marchetti et al., 1996, Zamzami et al., 1996, Vayssiere et al., 1994, Hennet et al., 1993, Schulzeosthoff et al., 1992).

Alternative oxidations of fatty acids, ketone bodies, short-chain carboxylic acids, propionate, acetate and butyrate in tumour cells may contribute to ATP supply by oxidative phosphorylation (OXPHOS)(Ralph et al., 2010). Tumour cells maintain a significant level of OXPHOS capacity to rapidly switch from glycolysis to OXPHOS during carcinogenesis. The more realistic theory considers that OXPHOS and glycolysis cooperate to enhance energy demands along tumourigenesis (SmolkovÃ¡ et al., 2011).

According to previous studies, glycolysis appeared a feature of several tumours associated with faster growth in high glucose environment; however, active OXPHOS is also a significant feature of other tumours occurred at a particular stage of carcinogenesis which might be more beneficial than a "glycolysis-only" type of metabolism in conditions of intermittent shortage in glucose supply. In addition, the metabolism of cancer cells is not constant during carcinogenesis and might depend

on both the activated oncogenes and the tumour cell microenvironment (Jose et al., 2011). Moreover, several studies concluded aerobic glycolysis as a consequence of hypoxia rather inherent to cancer (Zu and Guppy, 2004).

In fact, recent studies considered Mitochondria as therapeutic targets for cancer chemotherapy had been suggested combined strategies involving modulation of both glycolytic and mitochondrial pathways which might be required for more efficient elimination of malignant cells (Rodriguez-Enriquez et al., 2009, Galluzzi et al., 2006). As tumour cells shift among different patterns of energy metabolism (Pedersen, 1978, SmolkovÃ¡ et al., 2011, Nadege et al., 2009, Rodriguez-Enriquez et al., 2009), variant pharmacological trials could provide drug specificity through variable interference with distinctive steps of cancer energy production pathways.

Furthermore, any unspecific condition that damages cell's oxidative phosphorylation, but not severe enough to induce apoptosis, can potentially initiate a path to malignant tumour. Impaired mitochondrial function can also induce abnormalities in tumour suppressor genes and oncogenes. The accumulation of mitochondrial damage over time is what ultimately leads to malignant tumour formation. An increased dependency on substrate level Phosphorylation for survival would follow each round of metabolic and genetic damage thus initiating uncontrolled cell growth and eventual formation of a malignant neoplasm. Basically, the well-documented tumour-associated abnormalities in oncogenes, tumour suppressor genes, and chromosomal imbalances can arise as a consequence of the progressive impairment of mitochondrial function (Seyfried and Shelton, 2010).

Considering tumour heterogeneity, several reports suggested that distinct OSCC sub-sites varied on their glycolytic phenotype. Genes related to KEGG Oxidative Phosphorylation Pathway, GO (Gene Ontology, www.geneontology.org/) term, Organelle ATP Synthesis Coupled Electron Transport, several cytochrome oxidases and ATPases showed differences in expression when sub-sites were compared (Severino et al., 2008).

1.7 Biomarkers as diagnostic tool for HNSCC and OSCC

Biomarkers are defined as biological molecules that; correlate with the presence or absence of a disease state; prognostic correlating with a disease course; predictive of a tumour's response to a specific therapy (Williams, 2010). Global approaches for tumour assessment via DNA and gene expression microarrays have provided new information into altered molecules and pathways for further investigations in squamous cell carcinoma; though remain limited in their benefit for individual tumour assessment as a biomarker (Williams, 2000).

In an era in which health economics will dictate the availability of new and expensive agents, robust biomarkers for dose, schedule, and patient selection must remain as important priority (Banerji et al., 2008). Recent progressions in basic research of molecular biology have developed the understanding of the molecular process of HNSCC development and resulted in identification and classification of numerous biomarkers. Biomarkers of HNSCC are expected to help the early detection of primary and relapsed tumours (Lee et al., 2011).

A distinct pattern of gene expression, with progressive up- or down-regulation of expression, is frequently reported during the progression from histologically normal tissue to primary carcinoma and to nodal metastasis (Belbin et al., 2005, Roepman and Holstege, 2005, Roepman et al., 2006, Roepman et al., 2005, Odani et al., 2006, Yan et al., 2008), whilst alternatively expressed genes are potentially useful biomarkers for the prediction of malignant transformation (Odani et al., 2006).

HNSCC comprise a wide spectrum of heterogeneous neoplasms for which biomarkers are needed to aid in earlier diagnosis, risk assessment, and therapeutic response. The search for biomarkers includes evaluation of tumour tissues and surrogate materials by molecular, genomic and phenotypic means. Ideal biomarkers should be accurate and easy to perform, highly specific, objective, quantitative, and cost effective. Because of the heterogeneity of head and neck tumours, the integration of multiple selected markers in association with the histopathological features is supported for risk assessment (Williams, 2010).

Circulating mRNA biomarkers in serum or plasma have been targets for reverse transcription-PCR (RT-PCR)-based detection strategies in patients with cancers (Kopreski et al., 2001, Bunn, 2003). Parallel to the increasing number of such biomarkers in human body fluids, is the growing availability of technologies with more powerful and cost-efficient methods that enable mass screening for genetic alterations. Recent discovery by microarray technology revealed a large panel of human mRNA exists in saliva (Li et al., 2004b) suggesting a novel clinical approach, salivary transcriptome, for diagnostic applications of the disease and normal health surveillance.

Overexpression of EGFR in HNSCC has been associated with poorer overall survival and recurrence, while up to 90% of HNSCC patients express EGFR (Kalyankrishna and Grandis, 2006, Psyrri et al., 2005, Ang et al., 2002, Grandis et al., 1998). With implication of elevated EGFR in poor prognosis, it was one of the first biomarkers targeted as a potential therapy for HNSCC. Cetuximab, a monoclonal antibody directed against the extracellular receptor domain of EGFR, blocks ligand binding and subsequent downstream signalling, in addition to its role in the long-term downregulation of the receptor expression (Sunada et al., 1986, Jaramillo et al., 2006, Prewett et al., 1996, Mendelsohn, 1997, Lilenbaum, 2006). It has been the most effective targeted therapy applied to HNSCC to date (Yavrouian and Sinha, 2012).

1.8 Metabolomics as tumour biomarkers

Recent years, tumour metabolome is beginning to be characterized. Using standard metabolomic methods, tumours display elevated phospholipids' levels. Characterized by an elevation of total choline-containing compounds (tCho) and Phosphocholine), increased glycolytic capacity, including increased utilization of glucose carbons to drive synthetic processes, high glutaminolytic function, and over-expression of the glycolytic isoenzyme, pyruvate kinase type M2 (Glunde and Serkova, 2006, Mazurek and Eigenbrodt, 2003, Ackerstaff et al., 2003).

Interestingly, lipid metabolic profiles have been documented 83% accurate at discriminating between cancer patients and controls, using NMR-based

metabolomics of blood samples (Bathen et al., 2000). Importantly, *in vivo*, tCho determination via MRSI has detected breast, prostate, and brain tumours and correlates well with diagnosis via dynamic contrast enhanced-MRI (Howe et al., 2003, Jacobs et al., 2004, Yeung et al., 2001, Bolan et al., 2003, Stanwell et al., 2005).

Regarding the use of ^{1}H NMR metabonomics methods to detect cancer in serum, a study by Odunsi group (Odunsi et al., 2005) illustrated that principal component analysis of NMR spectra obtained for serum from patients could not only detect epithelial ovarian cancer (EOC) but could distinguish EOC patient serum from that obtained from patients with benign ovarian cysts as well (Whitehead and Kieber-Emmons, 2005).

Nuclear magnetic resonance (NMR) spectroscopy is a commonly used analytical method to analyse the small- molecule composition of body fluids such as urine and blood serum. Variations in metabolite concentrations have been associated with the biochemical status of organisms and reflect changes in metabolism arising from biologic conditions, including disease and response to chemical treatment. Recent studies demonstrate the applicability of NMR-based metabolomics using serum samples for the diagnosis and prognosis of disease (Beckwith-Hall et al., 2003, Brindle et al., 2003, Carraro et al., 2007, Duarte et al., 2007, Makinen et al., 2006, Odunsi et al., 2005, Stubbs et al., 2002, Tsang et al., 2006, Whitehead and Kieber-Emmons, 2005).

Both ^{1}H and ^{31}P magnetic resonance spectroscopy have previously been used to identify metabolic signatures of squamous cell carcinoma compared with benign tumours and normal tissues (Adalsteinsson et al., 1998, Arias-Mendoza et al., 2006, Bezabeh et al., 2005, El-Sayed et al., 2002, Maheshwari et al., 2000a, Maheshwari et al., 2000b, Mukherji et al., 1997, Star-Lack et al., 2000, Yokota et al., 2007). Both in vivo and in vitro clinical studies have shown that the choline-creatine ratio is significantly higher in OSCC than in normal tissue (Maheshwari et al., 2000b, Mukherji et al., 1997, Star-Lack et al., 2000). Analysis of squamous cell carcinoma cell line cultures was suggestive of rapid membrane biosynthesis during increased

cell proliferation (Mukherji et al., 1997). Moreover, in vitro studies using two-dimensional (2D) correlated spectroscopy revealed that a variety of metabolites, such as alanine, glutathione, histamine, isoleucine, and leucine, were showed a higher concentration in tumour compared to normal tissue (Mukherji et al., 1997). An NMR study of ex vivo tumour tissue identified elevated concentrations of taurine, choline, glutamic acid, lactic acid, and lipids in squamous cell tissue compared with normal tissue (El-Sayed et al., 2002).

Many solid tumours showed rising in glycolytic metabolism, as in case of OSCC, associated with the over-expression of glucose transporters, especially of Glut-1 (Kunkel et al., 2003). This metabolic change is commonly used to locate primary tumours and associated metastases using positron emission tomography by monitoring [^{18}F]-2-fluoro-2-deoxy-D-glucose uptake (Kunkel et al., 2003).Additional studies examined the role of advanced glycated end products (AGE) and increased levels of the AGE receptor (RAGE) in patients with diabetes mellitus type 2 and primary gingival carcinoma showing that RAGE expression is closely associated with the Invasiveness of OSCC (Bhawal et al., 2005, Ujpal et al., 2007).

Recent study of cancer patients samples, showed that the levels of valine, ethanol, lactate, alanine, acetate, citrate, phenylalanine, tyrosine, methanol, formaldehyde, and formic acid were reduced compared with those of healthy controls, whereas signals arising from glucose, pyruvate, acetone, acetoacetate, 3-hydroxybutyrate and 2-hydroxybutyrate, choline, betaine, and to a lesser degree, dimethyl glycine, sarcosine, asparagine, and ornithine showed enhanced loadings (Tiziani et al., 2009).

Further uses of metabolomics are to determine the maximum tolerated dose of new drugs or drug combinations while also collecting information on drug tolerability, pharmacokinetics, and pharmacodynamics (Spratlin et al., 2009).

1.9 Microarray technology

Microarray technology is based on hybridization of complementary nucleotide strands (DNA or RNA). Microarray chips consist of thousands of DNA molecules that are immobilized and gridded onto a support such as glass, silicon or nylon membrane. Because mRNA is susceptible to degradation it is converted to complementary DNA (cDNA) to form the target signal and for each type the probes are composed of DNA.

Each spot on the chip is representative for a certain gene or transcript. Fluorescently or radioactively labelled nucleotides (targets) that are complementary to the isolated mRNA are prepared and hybridized to the immobilized molecules (probes). Targets that did not bind to probes during the hybridization process are washed away. The amount of hybridized target molecules is proportional to the amount of isolated mRNA. The relative abundance of hybridized molecules on a defined array spot can be determined by measuring the fluorescent or radioactive signal. This method provides the advantage that it can interrogate the level of transcription of thousands of different genes from one sample in one experiment. Several competing technologies for microarray probe implementation have emerged, including the use of full-length cDNAs, or pre-synthesized or in situ synthesized oligonucleotides as probes. This platform is cDNA based and the probes are double stranded PCR products amplified from expressed sequence tags (ESTs) which range from 100-1000 base pairs in length. This has increased sensitivity but reduced specificity compared to oligonucleotide-based systems.

One of the "gold standard" technologies is the GeneChip distributed by Affymetrix. The GeneChips are constructed using a combination of two techniques, photolithography and solid-phase DNA synthesis. In Affymetrix GeneChip Microarray the probes are 25 nucleotide bases long and there are multiple oligonucleotide probes per gene which leads to a high specificity of the application compared to other Microarray platforms. Only a single mRNA population can be studied per GeneChip providing an absolute measurement of gene expression level. The Affymetrix GeneChip Microarray platform has been widely adopted as a

standard within the research community. The endorsed advantage of the Affymetrix Microarray platform is the ability to compare any single Microarray with any other and in any combination. Other distributors of DNA microarrays include; Applied Biosystems, Beckman Coulter, Eppendorf Biochip Systems, Agilent and Illumina.

1.10 Transcriptomic analysis

Transcriptomic is the functional analysis at the level of the genome (large-scale DNA sequencing that provided insight to the heterogeneity in coding regions of genes that lead to polymorphisms of gene expression). The genetic basis of the development of oral cancer has been considerably made in recent years. It is known that an accumulation of genetic alterations is the basis for the progression from a normal cell to a cancer cell (Califano et al., 1996). Progression is enabled by the more aberrant function of genes that positively or negatively regulate aspects of proliferation, apoptosis, genome stability, angiogenesis, invasion and metastasis (Hanahan and Weinberg, 2000). However, several studies suggest that early genetic changes do not necessarily correlate with altered morphology.

Increasing gene dosage through DNA amplification is a common feature of many tumours and results in upregulation of tumour- promoting genes (Oh and Mao, 1997). Although many studies contributed greatly to current understanding, they did not explain the complexities of such malignancies (Mendez et al., 2002).

With the advent of DNA microarrays, genomic-scale differential gene expression profiles can be obtained. Microarray analysis was influential in discovering clinically relevant knowledge by connecting changes in gene expression patterns with specific pathological conditions where mRNA expression profiles are generated for thousands of genes once across a collection of samples that belong to either one of two classes, for example, pathological specimen versus healthy tissue controls or early stage versus late stage of a specific disease. Transcriptional characteristics of tumours can be assessed with unprecedented completeness using large-scale microarray technologies (Schena et al., 1995) and/ or high-throughput quantitative PCR (Bernard and Wittwer, 2002).

Because phenotypic changes are heralded by genotypic changes, it also may be possible to detect differentially expressed genes that may serve as markers for the early detection of pre-malignant lesions, as well as, cancer invasion and metastasis. More than just focusing on the expression of a handful of genes, genomic-scale expression profiles allow the investigator to look at genetic expression variability in proliferation, invasion, carcinogenic metabolism, apoptosis and other genetic themes and pathways (Mendez et al., 2002).

A comparison of tumour profiling studies concluded that the composition of microarray gene expression signatures is highly unstable and dependent on the selection of patients used for determining the signature (Michiels et al., 2005). Furthermore, due to insufficient sample numbers and improper validation, some authors concluded that the results of such studies are overoptimistic. This is not surprising since many genes are simply interchangeable without affecting predictive accuracy because many prognosticative genes have similar, if not identical expression patterns across tumour samples (Ein-Dor et al., 2005). HNSCCs including OSCCs are complex heterogeneous tumours and microarray technology allows quite rapid, large scale screening of genes with the potential to identify candidate genes and pathways involved in such complex characteristic disease.

1.11 Selection of bioinformatics approaches

Bioinformatics is a specialized scientific discipline which emerged and developed as a consequence to the increase of the volume of biotechnological datasets and the need for the preservation and analysis of such data. It can be generally described as the application of a combination of machine learning and statistical approaches to explore biological processes by analysis of the large-scale and multi-dimensional data generated from numerous sources. It incorporates organized storage of the data, development of tools to analyse the data and the actual analysis of data.

Bioinformatics uses both informatics and statistical tools implemented to extract and analyse information (Wu, 2001). The type and number of employed analytical tools in each session of bioinformatics analysis depend on the data type and nature of biological question addressed during analysis of a specific dataset. Generally,

bioinformatics is a management information system for molecular biology and has many practical applications (Umar, 2004), and so it became an inevitable research element of the molecular biology. As a consequence of hastened generation of large-scale datasets by high throughput research platforms, including gene expression experiments, cancer biology commonly linked to informatics. Bioinformatics is mainly used to identify cancer biomarkers, their function and molecular mechanisms underlying cancer progression. A considerable number of novel bioinformatics approaches were already recognized with a sophisticated level allowing to perform the functional analysis in semi-automated mode, for example, gene set enrichment and pathway-based analysis, (Subramanian et al., 2005, Yi et al., 2006).

Functional analysis of biomarkers identified in large-scale datasets can be performed using a wide variety of bioinformatics tools. Gene annotation is the principal requirement for the initiation of functional analysis. The annotative information is provided by various biotechnological database management organizations like NCBI (Gene), EBI (Ensembl), and caArray database, along with independent groups like GeneCards (Safran et al., 2002, Curwen et al., 2004, Maglott et al., 2007, Maglott et al., 2011). Pathway information systems, for example, Kyoto encyclopaedia of genes and genomes (KEGG), and Biocarta provide information concerning the molecular interactions of gene products in an organized fashion (Aoki-Kinoshita and Kanehisa, 2007). Association of genes with their corresponding ontological terms is another way of enhancing functional analysis. The gene ontology (GO) database was developed to standardize the gene and gene product attributes across the unending list of biological databases (Ashburner et al., 2000). Database for annotation, visualization, and integrated discovery (DAVID) is a hybrid functional analysis tool combining all these systems in a single web interface (Huang et al., 2009b). Such integrative tools enhanced the functional analysis of the genes and/or proteins while providing a holistic perspective of large-scale datasets (Manyam, 2009).

Considering metabolomic profile, the human metabolome database (HMDB) (http://www.hmdb.ca/), a freely available electronic database containing detailed information about small molecule metabolites existed in the human body was employed.

1.11.1 Bioconductor

Bioconductor is open source software for bioinformatics (www.bioconductor.org/index.html) which provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source, has 671 software packages until April 2013, and continuing development. Bioconductor has advanced facilities for analysis of microarray platforms including Affymetrix, Illumina, Nimblegen, Agilent, and other one- and two-colour technologies. Bioconductor includes extensive support for analysis of expression arrays, and well-developed support for exon, copy number, SNP, methylation, and other assays. Major workflows in Bioconductor include pre-processing, quality assessment, differential expression, clustering and classification, gene set enrichment analysis, and genetical genomics. Bioconductor offers extensive interfaces to community resources, including GEO, ArrayExpress, Biomart, genome browsers, GO, KEGG, and diverse annotation sources.

1.11.2 Data filtering and normalization

Normalisation is essential to explicate and minimise systematic and experimental variations in the calculation of gene expression data, hence identifying biologically relevant data and removing the non-biological effect on the data (Schuchhardt et al., 2000). The preferred method used in research work of the thesis was the GeneChip Robust Multichip Average method (gcRMA) (Wilson and Miller, 2005).

After normalisation, the data are filtered to reduce variation before machine learning or statistical algorithms are used. Genes that have a low overall variance across all samples can be removed because they are of little value. Filtration is normally incorporated into the first stage of many analysis packages.

1.11.3 Corrections of batch effects

Integration of microarray datasets is advantageous for researchers to increase statistical power in detecting biological phenomena from studies where logistical considerations restrict sample size (Turnbull et al., 2012). Generally, it is inappropriate to combine datasets without adjusting for batch effects (Johnson et al., 2007a, Lazar et al., 2012).

Batch effects can be caused by laboratory conditions, reagents, and personnel diversity. This becomes a major problem when batch effects correlated with an outcome of biological interest and result in incorrect conclusions (Leek et al., 2010). In this study, the ComBat batch correction method was utilised to minimise effects between studies. ComBat package in R software (Johnson et al., 2007a) implicated parametric and non-parametric empirical Bayes frameworks for adjusting data of batch effects. According to recent published work (Turnbull et al., 2012), ComBat expressed the best performance.

1.11.4 Differential gene expression analyses

Samples are relatively compared to examine differential gene expression and to extract meaningful biological results from this comparison. Generally, data analysis takes two main forms. The first is 'unsupervised' analysis where data is sorted and gene expression differences are derived based on inherent differences within the dataset. Unsupervised clustering is achieved and heatmaps generated using Pearson correlation and hierarchical clustering in R. Supervised analysis is carried out using the Significance Analysis for Microarrays (SAM) (Tusher et al., 2001) in *siggenes* package available from Bioconductor. SAM is one of the more potent and effective ways of identifying particular gene products involved in differentiating transcriptional profiles between two groups (Tusher et al., 2001). SAM gives good all-round discriminatory ability based on average signal intensity differences within the supervised groups and between the same groups at pre-determined fold-change levels and gives an estimate of the false discovery rate for a significantly changed gene (Jeffery et al., 2006). Hierarchical clustering heatmaps are generated using *gplots* package (Gentleman et al., 2004). Fold change values for relative

“upregulated/over-expressed” and “downregulated/ under-expressed” genes were obtained using Wilcoxon Rank Statistics function in siggenes R software.

Lists of genes are further refined by trimming out low magnitude fold changes and / or reducing the false discovery rate (FDR) value of the significant genes by adjustment of delta value. The value of delta is chosen that provides the best balance between the number of identified genes and the estimated FDR, i.e. that allows to simultaneously attain the two competing goals “As many genes as possible" and “As low FDR as possible".

1.11.5 Biological annotations of significant genes and metabolites

Differentially expressed genes and/or metabolites in different comparison analyses are subsequently uploaded and examined using tools available from IPA (Ingenuity Systems, version 9.0, www.Ingenuity.com) to define the relative cellular functions, canonical pathways, upstream regulators, potential biomarkers, and molecular processes that are most commonly perturbed in correlation with a set of altered genes or metabolites of a specific comparison, as well as, to inspect the most significant pre-generated networks associated with the uploaded group of genes or metabolites. IPA knowledge base recruits the distinguished genes (and/or metabolites) that involved in and regulated a specific disease or disorder to generate and highlight the above functions. Canonical pathways analysis identified pathways from IPA library of canonical pathways, which are most significant to the input data set. The significance of the association between the data set and the canonical pathway was determined based on two parameters: (1) A ratio of the number of genes from the data set that map to the pathway divided by the total number of genes that map to the canonical pathway and (2) a P value calculated using Fischer's exact test determining the probability of the association between the genes in the data set and the canonical pathway is due to chance alone.

IPA analysis generates networks connecting the differentially regulated genes (and/or metabolites) from published interactions. This approach has the advantage of building pathways independently of the traditional canonical pathways database.

Therefore, this approach potentially highlights unknown relationships and connections between genes and their products including metabolites. The relationship between the genes (nodes in the pathway) is indicated by a line, which corresponds to at least one reference from the literature. The significance of a canonical pathway is controlled by p-value (right-tailed Fisher's exact test).

The identification of regulatory networks gives insight into the mechanisms of a specific disease, while the use of pathway analysis to further test the validity of candidate genes and their likely contribution to a specific disease is certainly a valuable tool for gene expression analysis.

1.12 Metabolomics and metabonomics

The metabolism reflects an abnormal change in an organism (Griffin and Shockcor, 2004). Both the theory and the application have been well documented (Nicholson et al., 1999, Jurs, 1986, Eriksson et al., 2006). Metabolomics is the systematic identification and quantitation of all metabolites in a given organism or biological sample. It's resided at the end of the biological road that commences with DNA. It is the place where the variation in genomes, both static and expressed, becomes integrated. Its goal is to relate the physiological consequences of disease, i.e. the metabolites associated with pathological processes, to those diseases' genomic origins.

The enhanced resolution provided by nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS), along with powerful chemo metric software, allows the simultaneous determination and comparison of thousands of chemical entities, which has led to an expansion of small-molecule biochemistry studies in bacteria, plants, and mammals. Continued development of these analytical platforms will accelerate the widespread use of metabolomics and allow further integration of small molecules into systems biology (Idle and Gonzalez, 2007). Metabolomics is an analytic tool used with pattern recognition approaches and bioinformatics to detect metabolites and follow their changes in bio fluids or tissue (Fiehn, 2002, Griffin and Shockcor, 2004, Kell and Mendes, 2000). Precise number of human metabolites is unknown, with estimation ranging from the thousands to

tens of thousands. Metabolomics is a term that encompasses several types of analyses, including (*a*) metabolic fingerprinting, which measures a subset of the whole profile with little differentiation or quantitation of metabolites (Ryan and Robards, 2006); (*b*) metabolic profiling, the quantitative study of a group of metabolites, known or unfamiliar, within or associated with a particular metabolic pathway (Dunn et al., 2005, Roessner et al., 2001); and (*c*) target isotope-based analysis, which focuses on a precise segment of the metabolome by analysing only a few selected metabolites that comprise a specific biochemical pathway (Boros et al., 2005).

Metabolomics allows for a global assessment of a cellular state within the context of the immediate environment, taking into account genetic regulation, altered kinetic activity of enzymes, and changes in metabolic reactions (Griffin and Shockcor, 2004, Mendes et al., 1996, Mendes et al., 1992). Thus, compared with genomics or proteomics, metabolomics reflects changes in phenotype and therefore, function. The "omic" sciences are, however, complementary as "upstream" changes in genes and proteins are measured "downstream" as changes in cellular metabolism (Griffin and Shockcor, 2004, Boros et al., 2005). Conversely, metabolomics is a terminal view of the biological system, not allowing for representation of the increased or decreased genes and proteins (Figure 1.2). Other features of metabolomics are similar to those of proteomics and transcriptomics, in their ability to assay bio fluids or tumour samples and the relatively inexpensive, rapid, and automated techniques once start-up costs are taken into account (Spratlin et al., 2009).

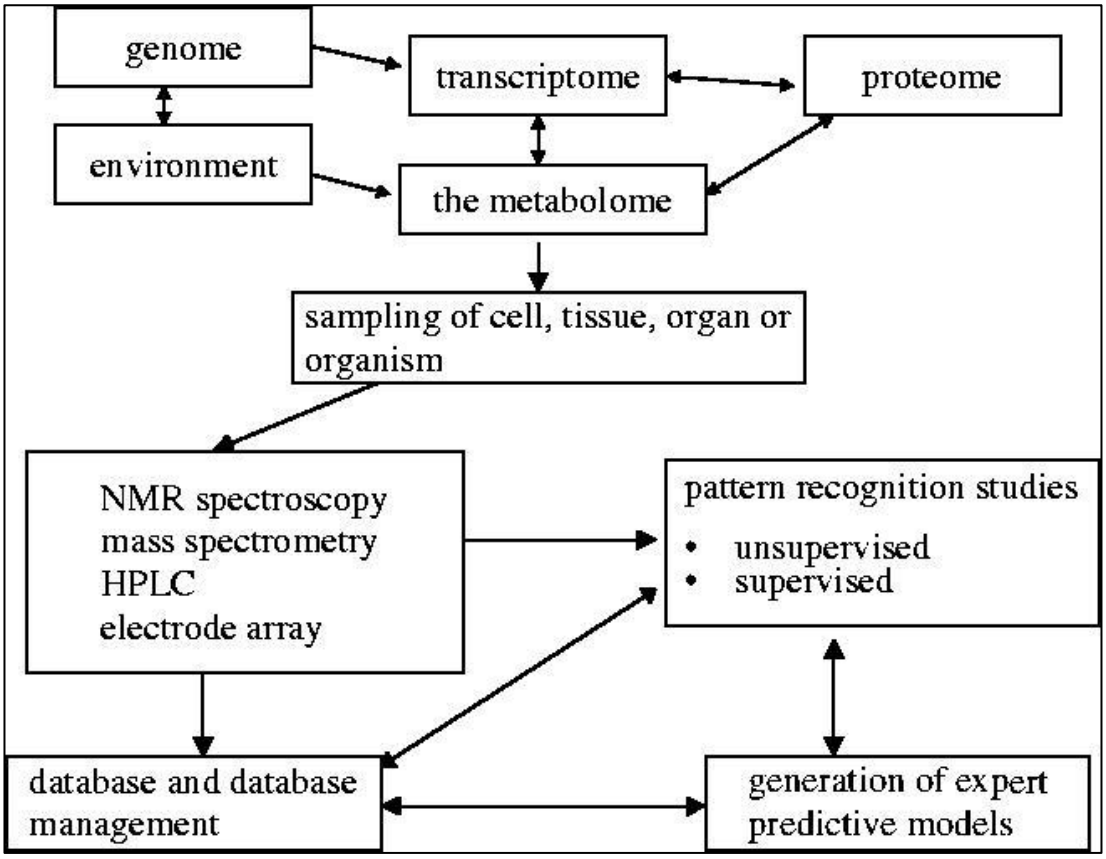


Figure 1.2 : A schematic view of metabolomics

The position of metabolomics is shown in respect to the other omic approaches and the influence of the environment. In addition a scheme is shown for producing pattern recognition models capable of deducing metabolic profiles. The initial phase usually involves the acquisition of a large dataset in terms of both the variables (metabolites) and observations (subjects), commonly using either ^{1}H NMR spectroscopy or mass spectrometry. However, the use of pattern recognition techniques is an integral part of the approach if the innate metabolic variation associated with different individuals is to be separated from that induced by the insult or physiological stimulus. Following the generation of a database from metabolic responses, this can then be used to build a predictive pattern recognition model that is capable of predicting class membership (e.g. clustering according to the gene deleted) (Adapted from Griffin et al. 2006).

Metabonomics is defined as ` the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification'. This concept has arisen from work on the application of "H-NMR spectroscopy to study the multicomponent metabolic composition of biofluids, cells and tissues over the past two decades (Nicholson et al., 1983, Bales et al., 1984, Nicholson et al., 1985, Gartland et al., 1989, Foxall et al., 1992, Moka et al., 1998). There is also a related concept of the ` Metabolome' that represents the total small molecule complement of a cell. However, metabonomics deals with

detecting, identifying, quantitating and cataloguing the history of time-related metabolic changes in an integrated biological system rather than the individual cell. Such multidimensional metabolic trajectories are then related to the biological events in an ingoing pathophysiological process (Nicholson et al., 1999).

Metabonomic analysis has been increasingly used to monitor metabolic abnormalities in cells and their microenvironment for detection of cancer markers (Zhou et al., 2009). Results of Zhou group indicated that ^{1}H NMR-based Metabonomic approach is a feasible and efficient method for differentiating the OSCC patient's plasma from the healthy controls. As a potential novel strategy and convenient technique, it deserves a further evaluation for an early detection of oral cancer (Zhou et al., 2009).

As a useful tool, ^{1}H nuclear magnetic resonance (^{1}H NMR) spectroscopy of blood from patients ,were diagnosed with cancers such as breast, ovarian, prostate, and brain cancer, has produced better results lately (Whitehead and Kieber-Emmons, 2005, Odunsi et al., 2005, Whelehan et al., 2006, Lindon et al., 2003).

Because only a little biofluid contains a complete metabolic profile, it will be useful for assessing overall disease with less stress for the patient. Application of automated data reduction algorithms and chemo metric analysis, which called pattern recognition analysis, can be competent for the description and recognition of the dynamic multivariate metabolism (Zhou et al., 2009).

The study of proteins at a global level, proteomics, has not yet achieved such widespread use (Idle and Gonzalez, 2007), possibly because of the weak points in the chain of analysis (such as sample handling, protein separation and digestion) and the greatest challenge with regard to protein detection comprised in the fact that there is no protein stain currently available that possesses all desired properties; a routine and reproducible lowest limit of detection with optimal signal-to-noise ratio; a wide dynamic range, with a linear relationship between protein quantity and staining intensity; compatibility with downstream microchemical characterization techniques; ease of use; and inexpensive, high throughput rates of use (Gauci et al., 2011).

In contrast, metabolomics is emerging as a field with a tremendous promise in extending "omics" from the gene to the small molecule (Idle and Gonzalez, 2007). Via the fast development of metabolomics technology platforms and incorporating metabolomics into research efforts, it is now possible to answer key questions that could not be fully addressed by the other omics alone. From bacteria to humans, examples of this principle are accruing at a rapid pace that has been made promising by remarkable recent developments in analytical chemistry, such as high-field nuclear magnetic resonance (NMR) and mass spectrometry (MS) platforms for small- molecule separation, detection, and characterization, together with the availability of relatively user-friendly multivariate data analysis software packages that can de-convolute the huge data matrices generated in a metabolomic experiment. Such analysis of the metabolome is applied to an ever-increasing number of biological and medical research topics, from bacteria to human (Idle and Gonzalez, 2007).

In mammals, the potential of metabolomics is enormous, particularly regarding to studying the interface between chemical universe and human biology. Humans are probably exposed to some 1–3 million discrete chemicals in the lifetimes, and most of these xenobiotics are metabolized in the body to render them less poisonous. However, in a significant number of cases, the detoxification apparatus renders a xenobiotic more chemically reactive, potentially more toxic, and, in some cases, carcinogenic.

Understanding the role of human genetic differences in xenobiotic metabolism and its relationship to inter-individual susceptibility to chemical toxicities and carcinogenesis is an area of intense interest (Wogan et al., 2004). A convenient, cost-effective means to determine metabolic potential would be of great value in understanding the molecular epidemiology of cancer (Idle and Gonzalez, 2007).

This task is quite daunting because of the genetic differences, dietary diversity, and other problems inherent in the human population that led to mark inter-individual differences in the human serum metabolome, while these difficulties can be overcome by studying large populations of cases and controls with powerful

statistical multivariate data analysis software that can mine huge data sets. Biomarkers that can be used to diagnose pre-diabetes, cancer, and coronary artery disease would be of great value in clinical medicine (Idle and Gonzalez, 2007).

All authors and investigators look forward to a day when metabolomic analysis that can capture the whole metabolome is truly accessible and widely available in the way that PCR, SNP mapping, and microarray analysis are now (Idle and Gonzalez, 2007, Vora et al., 2003).

1.12.1 NMR metabolomics/metabonomics techniques

For testing and monitoring procedures, Metabonomic methods are generally non-invasive, and based upon instrumentation availability they can be conducted either in vivo or in vitro. For in vitro applications, samples such as urine, plasma, serum, bile, and cerebrospinal fluid can be analysed using the same basic techniques. Whole tissue samples may also be analysed if the appropriate instrumentation is available (HR-MAS NMR). Depending upon observable metabolites and their patterns, one can describe, using the changing concentrations of hydrophilic and lipophilic metabolites, a disease in various stages of malignancy as in case of using metabolic changes to establish self-organizing maps for breast cancer tissue by Beckonert group where breast-cancer tissue was successfully classified as unaffected, of intermediate malignancy, or of high malignancy (Beckonert et al., 2003). In addition, multiple metabolites were monitored simultaneously and suggested useful for classifying tissue type based upon their concentration (Whitehead and Kieber-Emmons, 2005).

Generally, techniques are highly quantifiable and reproducible, experimental time takes only a matter of minutes, and sample preparation is minimal, usually involving the addition of D2O (Heavy Water, 2H2O) directly to the sample and presenting the sample to the instrument for analysis. The greatest advantage to the technique is that samples are not destroyed during analysis. This allows the investigator to pursue a wide range of studies on any one sample, including kinetic analysis, protein–ligand binding studies, structural analysis of metabolites, and ‘separation’ methods using the magnetic field of the instrument. In addition, by monitoring the same sample over

time (kinetically), changes in metabolite concentration or composition can be observed (Whitehead and Kieber-Emmons, 2005).

Recent technological progress in NMR spectroscopy and mass spectrometry (MS), the two most accepted methods used in the measurement of metabolites, has improved the sensitivity and spectral resolution of analytic assays on metabolomic samples in attempts to achieve a comprehensive biochemical assessment. Because cancer cells are known to possess a highly unique metabolic phenotype, development of specific biomarkers in oncology is possible and might be used in identifying fingerprints, profiles, or signatures to detect the presence of cancer, determine prognosis, and/or assess the pharmacodynamic effects of therapy (Glunde and Serkova, 2006, Griffin and Shockcor, 2004, Serkova et al., 2007, Serkova and Niemann, 2006, Denkert et al., 2006, Odunsi et al., 2005, Howe et al., 2003, Bathen et al., 2007, Kline et al., 2006, ElDeredy et al., 1997, Morvan and Demidem, 2007). MS analysis requires more labour-intensive and destructive tissue preparation than NMR, but has greater sensitivity for metabolite detection (Dunn et al., 2005).

Compared with MS, NMR is less sensitive (on the order of 10 µmol/L), and requires more expensive instrumentation. In addition, ^{1}H-NMR spectra is sensitive to pH, ionic content, and temperature, and may require solvent suppression. The major advantages of ^{1}H-NMR include its nonbiased metabolite detection, quantitative nature, and reproducibility. ^{1}H-NMR can also be used for liquid or solid samples, using magic angle spinning (HR-MAS) techniques, with minimal sample preparation (Serkova and Niemann, 2006). Another significant advantage of NMR is that metabolic markers discovered and analysed *in vitro* can be measured *in vivo*, assuming sufficient tissue abundance, using localized magnetic resonance spectroscopy imaging (MRSI). MRSI is an additional technology related to magnetic resonance imaging (MRI) whereby metabolites instead of anatomy are imaged. Essentially , MRSI is a composite of traditional NMR spectroscopy and MRI that allows for non-invasive *in vivo* visualization and determination of spatial distribution of a specific metabolite in patients without exposure to ionizing radiation (Spratlin et al., 2009).

The major disadvantage of NMR technique involves the complexity of the data obtained. The presence of thousands of resonances arising from hundreds of metabolites that may be observed in complex biofluids makes spectral interpretation difficult, forcing the researcher to rely on statistical algorithms for such analysis. Generally, the largest problem in using spectroscopy to analyse such mixtures is the presence of water. Water poses a problem for signal-to noise ratios and spectral overlap, particularly when using magnetic resonance techniques. Unfortunately, all biological fluids and, specifically, the human body, involve ample amounts of water. Thus, suppression of the water resonance which requires sophisticated instrumentation is a major hurdle that must be overcome. In addition, the human body contains other sources of interference (such as lipids), which must be suppressed as well (Whitehead and Kieber-Emmons, 2005). However, several studies were published to manage this weak point; one of them is a "WAVEWAT study" a new processing algorithm to suppress the on resonance water signal in NMR spectra. It is based on a Multi-resolution analysis (MRA) of the free induction decay (FID) using a dyadic discrete wavelet transforms (DWT). The width of the suppressed signal can be adjusted so that signals close to water are recovered without distortion of the signal shape and intensity. Computational efficiency is comparable to that of convolution filters employing a Fourier transform (Gunther et al., 2002).

NMR spectroscopic methods focus on the analysis of magnetically susceptible atomic nuclei (e.g. 1H, 13C, 15N, 31P) that are present in molecules which are introduced to the instrument in small sample volumes, in either the solid or liquid state, allowing for the observation of chemical structures and processes at the molecular level. In terms of experimental acquisition time, effective spectroscopic studies require a relatively short period of time (30–45 min) (Morrison et al., 2007).

While acquisition times and the presence of interfering biological molecules limit clinical MR techniques in terms of spectroscopy, studies can be performed ex vivo (intact tissue analysis) and in vitro (soluble tissue extracts and biofluids) using classical NMR spectroscopic methods at larger magnetic field strengths and utilizing complex pulse sequences which will allow for enhanced suppression of unwanted resonances, ultimately leading to greater sensitivity and spectral resolution. In order

to perform ex vivo studies, sampling from the body must occur prior to analysis, usually involving the clinical excision of tissue from a tumour or the surrounding region. Fortunately, only small sample amounts are generally required for such studies. In vitro analyses are less invasive, usually focusing on samples such as plasma, serum, urine, saliva, and to some extent cerebrospinal fluid which are considered ideal for biomarker monitoring. Ex vivo spectroscopy provides detailed information that is usually lost or altered during in vitro extraction processes by instead focusing on the intact tissue itself (Whitehead and Kieber-Emmons, 2005).

In vitro analyses can be performed using NMR spectrometers equipped with standard probes designed for solution-state samples, with the major limitations for quality spectra being that of spectrometer field strength and solvent-suppression capabilities. However, extraction techniques that are required for the sample preparation phase often alter metabolite concentrations making it difficult to quantitatively determine metabolite levels in vivo from in vitro extracts (Whitehead and Kieber-Emmons, 2005). Despite this fact, a great deal of information has been obtained from in vitro extraction studies, with the presence of various metabolites and other small organic molecules being correlated to various stages of oral cancer.

NMR (1H-NMR) advantages include the theoretical possibility to detect and structurally elucidate any molecule containing hydrogen and also being very quantitative in its nature. However, disadvantages include low sensitivity, typical detection limits are in the range milli- to high micromolar (Pendyala et al., 2007), and also the fact that one-dimensional spectra gets very dense with spectral information resulting in more abundant metabolites masking signal of less abundant ones.

Compared with MS, NMR requires more expensive instrumentation. In addition, ^{1}H-NMR spectra is sensitive to pH, ionic content, and temperature, and may require solvent suppression. The major advantages of ^{1}H-NMR include its nonbiased metabolite detection, quantitative nature, and reproducibility (Serkova and Niemann, 2006).

In mass spectrometry, mass separation is based on the fact that molecules become charged in a process called ionization. Mass spectrometry advantages include high sensitivity (milli- to nanomolar) and being easily interfaced with chromatography resulting in the capability to temporally resolve the analysis of thousands of components. Disadvantages include potential difficulties in quantitation without proper internal standards due to variations in ionization efficiency and discrimination against metabolites due to their elemental composition. MS analysis requires more labour-intensive and destructive tissue preparation than NMR (Dunn et al., 2005).

Generally, techniques are highly quantifiable and reproducible, experimental time takes only a matter of minutes, and sample preparation is minimal, usually involving the addition of D2O (Heavy Water, 2H2O) directly to the sample and presenting the sample to the instrument for analysis. In both NMR and MS techniques, samples must be frozen as quickly as possible after collection, preferably into -20 °C, -80 °C, or dry ice. Authors conclude that storage of human urine samples at or below -25 °C do not require a preservative. Aliquoting to 3-4 vials before freezing in order to avoid freezing-thaw cycles, which can affect the sample. Both techniques can deal with tissue and biofluids. However, MS can't be used in vivo (Morrison et al., 2007).

1.12.2 Points to be considered using NMR spectroscopy for metabolomic analysis

1-The aspects (nutritional/physiological status, sample type, cohort size and heterogeneity, and analytical sensitivity) should be carefully assessed before initiation of a metabolomics study aimed at finding clinically relevant biomarkers (Nordstrom and Lewensohn, 2010).

2-Extraction techniques required in the sample preparation phase often alter metabolite concentrations making it difficult to quantitatively determine metabolite levels in vivo from in vitro extracts. In addition, extract samples frequently contain additional resonances not present in vivo. Despite this fact, a great deal of information has been obtained from in vitro extraction studies, with the presence of various metabolites and other small organic molecules correlated to various stages of breast cancer (Whitehead and Kieber-Emmons, 2005).

3-A number of potential metabolomics pitfalls lurk in these fluids. Firstly, factors like age, sex, nutritional status, and time of sampling might be reflected in the metabolite composition of chosen body fluid (Nordstrom and Lewensohn, 2010).

4-In the metabolomics studies involving urine, plasma, and saliva samples with healthy humans; it was demonstrated that standardizing the diet could reduce the physiological variation in the urine samples but not in the plasma and saliva samples (Walsh et al., 2006).

5-Another aspect to consider when performing discovery metabolomics in body fluids is the dilution of potential markers in the relatively large volumes body fluids represent (Lutz et al., 2007).

6-The heterogeneity of many diseases also poses a problem (Nordstrom and Lewensohn, 2010).

7-High-resolution 1H NMR spectroscopy is a relatively insensitive technique which only measures the high-concentration metabolites present in a tissue extract, bio fluid or intact tissue. Typically, using a simple one-dimensional pulse sequence 30–100 metabolites are observable in urine, 20–30 metabolites in blood plasma or serum and 10–30 metabolites in tissue extracts. Thus, a perturbation at one point in the metabolic network can rapidly be transferred to other metabolic pathways through these highly connected metabolites. Accordingly, the high-concentration metabolites will be greatly variable during a perturbation, and hence 1HNMR spectroscopy is sensitive to a wide range of genetic modifications, toxicology insults and physiological stimuli (Griffin, 2006). If these metabolites vary across a wide range of modifications, it then becomes difficult to use it as unique biomarkers for disease processes or for assuming which pathways have been stimulated during a modification. This can be further confounded by the detection of non-specific effects, which accompany a genetic modification or toxicological insult (Griffin, 2006). According to this, sensitivity is accompanied with low specificity, which could reach the level of baffling and misleading.

1.12.3 Data analysis and interpretation of metabolomics

The guiding principle of metabolomics is the global assessment of hundreds of endogenous metabolites in a biological sample simultaneously. Statistical analyses are then applied to provide meaningful information about the metabolic profile of the sample. The three major steps in multivariate metabolomic analyses are depicted in Figure 1.3. Because ^{1}H-NMR or MS spectra from bio fluids or tumour tissue contain hundreds of signals from endogenous metabolites and are highly redundant, spectral data sets, reduced to 100 to 500 spectral segments, and their respective signal intensities are directly entered into statistical programs (Dunn et al., 2005, Dettmer et al., 2007, Buchholz et al., 2002).

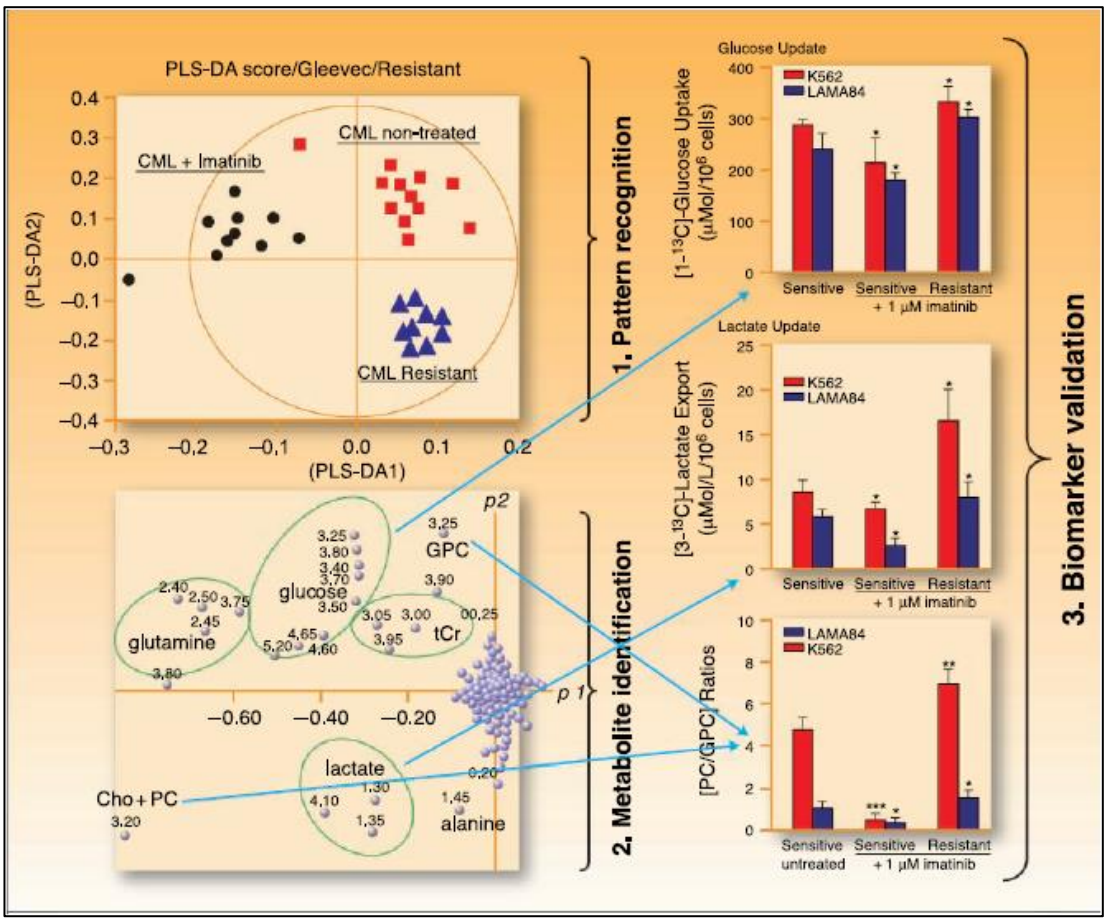


Figure 1.3: Three major steps of metabolomics analysis

The example is given for imatinib treatment in chronic myeloid leukemia cells using (1) 1H-NMR spectra of cell extracts followed by principal component analysis for pattern recognition (2) metabolite identification resulting in a biomarker (3) metabolite quantification and validation (Adapted from (Spratlin et al., 2009)).

Most of metabolic profiling studies have used computer-aided statistical interpretation for the data. This improves refining and distilling of the complex raw data. Similar to gene array analysis, multivariate statistics have been designed for large data sets, with two major types of pattern recognition processes, unsupervised and supervised. Unsupervised data analysis, such as hierarchical cluster analysis and principal component analysis, measures the innate variation in data sets, whereas the supervised approach, including principal component regression and neural networks, uses prior information to generate the clusters of patterns (Lee et al., 2008). Although many other statistical approaches exist, including cluster analysis, linear discriminant analysis, Bayesian spectral decomposition, and several other chemometric methods (Holmes and Antti, 2002).

1.13 Integrative Analysis of transcriptomics and metabolomics

Transcriptomics provides the tool for deciphering gene expression networks, and proteomics links these networks to protein products. The third crucial partner is metabolomics, which defines the metabolic network(s) linked to gene expression. NMR and mass spectrometry enable the broad screen analysis of the metabolome and its transformation pathways, transcending classical targeted metabolic studies. Fan et al. combined these tools to investigate the anticancer mechanisms of different selenium forms in human lung cancer cells. Using 2D NMR and tandem-MS, they mapped perturbations of 13C labelling patterns in numerous metabolites induced by selenite and selenomethionine. This information was used to interpret selenite-induced changes in gene expression networks. Linking metabolic dysfunctions to altered gene expression profiles provided new insights into the regulatory network underlying the metabolic dysfunctions, enabled the assembly of discrete gene expression events into functional pathways, and revealed protein targets for proteomic analysis (Fan et al., 2006).

One advantage of the genotype lies in the realization that it is unaffected by the clinical environment. Whereas enzyme activity may vary for one clinical reason or another, the genotype provides a point of reference from which to understand

differences in enzyme expression of that gene among patients, impossible with the more descriptive phenotype assessment.

Chou group used combined genotype-phenotype relationship of CYP2D6. Allele-specific PCR and CYP450 GeneChip were used for genotypic analysis. DXT-MR was used for metabolic ratio assessment. The study recruits the CYP2D6 genotype to provide a point of reference from which to understand differences in CYP2D6 enzyme expression among patients that are impossible with the more descriptive phenotypic assessment (Figure 1.4)(Chou et al., 2003).

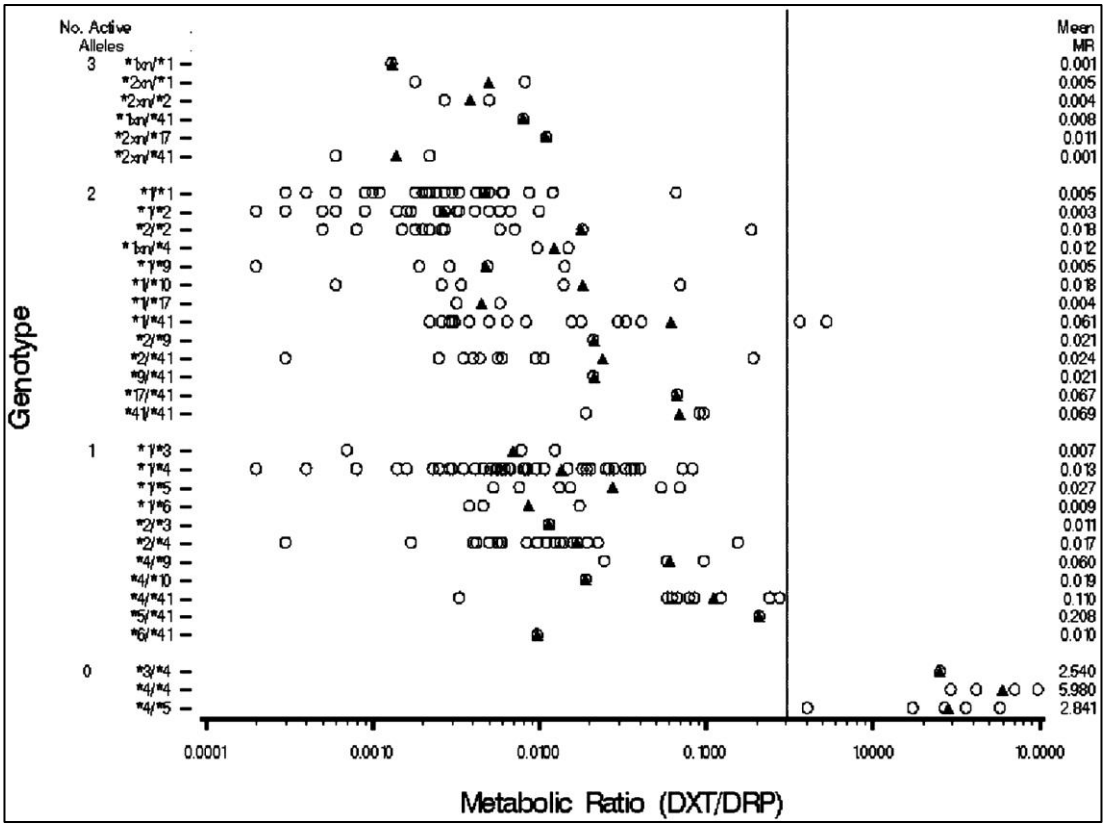


Figure 1.4: Plot of CYP2D6 genotype vs. log10 (DXT MR) for 229 individuals

Seven samples that could not be retested for the _1584G/C promoter polymorphism were excluded from the plot. The mean MR for each CYP2D6 genotype group is listed on the right-hand y axis and plotted as OE. Genotype groups are divided into four separate categories based on 3, 2, 1, or 0 active (fully or partially active) gene copies. The anti-mode separating the PM from the rest of the population is shown as a dotted vertical line (Adapted from Chou at al.2003).

Metabolic profiling can work as a functional genomic tool when the mutation or modification produces a severe phenotype such as for many of the inborn errors of metabolism. However, mutations of non-rate-determining enzymes, or proteins that

do not produce an extreme pathological response, can still influence metabolic profiles by perturbing the many small modules that make up the network of metabolic pathways. A perturbation at one point will be rapidly transferred to many of the key nodes, some of which will be NMR detectable (Ravasz et al., 2002). This makes for a powerful analytical tool (Griffin, 2003).

A considerable number of further studies were carried out using approaches integration to correlate transcriptomic and metabolomic data sets. Griffin group applied a combined genomic, transcriptomic and metabolomic approach to distinguish the fatty liver disease associated with supplementation of orotic acid to normal food intake (as the exact mechanism from orotic acid exposure to disrupted lipid transport is unknown) (Griffin, 2006).

To compare the solution and solid-state spectra from the liver with the transcriptional profiles measured using DNA microarrays the pattern recognition process (prediction to latent structures through partial least squares) was used to cross correlate the metabolite changes (X matrix) to the transcriptional dataset (Y matrix) (Figure 1.5).

This combined analysis identified metabolic changes involving uridine metabolism and choline turnover, and transcriptional changes such as the expression of stress proteins (Griffin et al., 2004). Coen et al. used combined transcriptional and metabolomic approach to investigate acetaminophen induced hepatotoxicity (In this study, the transcriptional profile of liver tissue was supplemented by metabolic profiles of liver tissue (both extracts and intact tissue); blood plasma and urine obtained using 1H NMR spectroscopy. These metabolic profiles identified decreases in glucose and glycogen and increases in lipids in liver tissue, alongside glucose, pyruvate, alanine and lactate increases in blood plasma. These metabolic changes were indicative of an increased glycolytic flux, and this was also confirmed at the transcriptional level demonstrating that the two approaches provided complementary information about the drug safety assessment process (Coen et al., 2003).

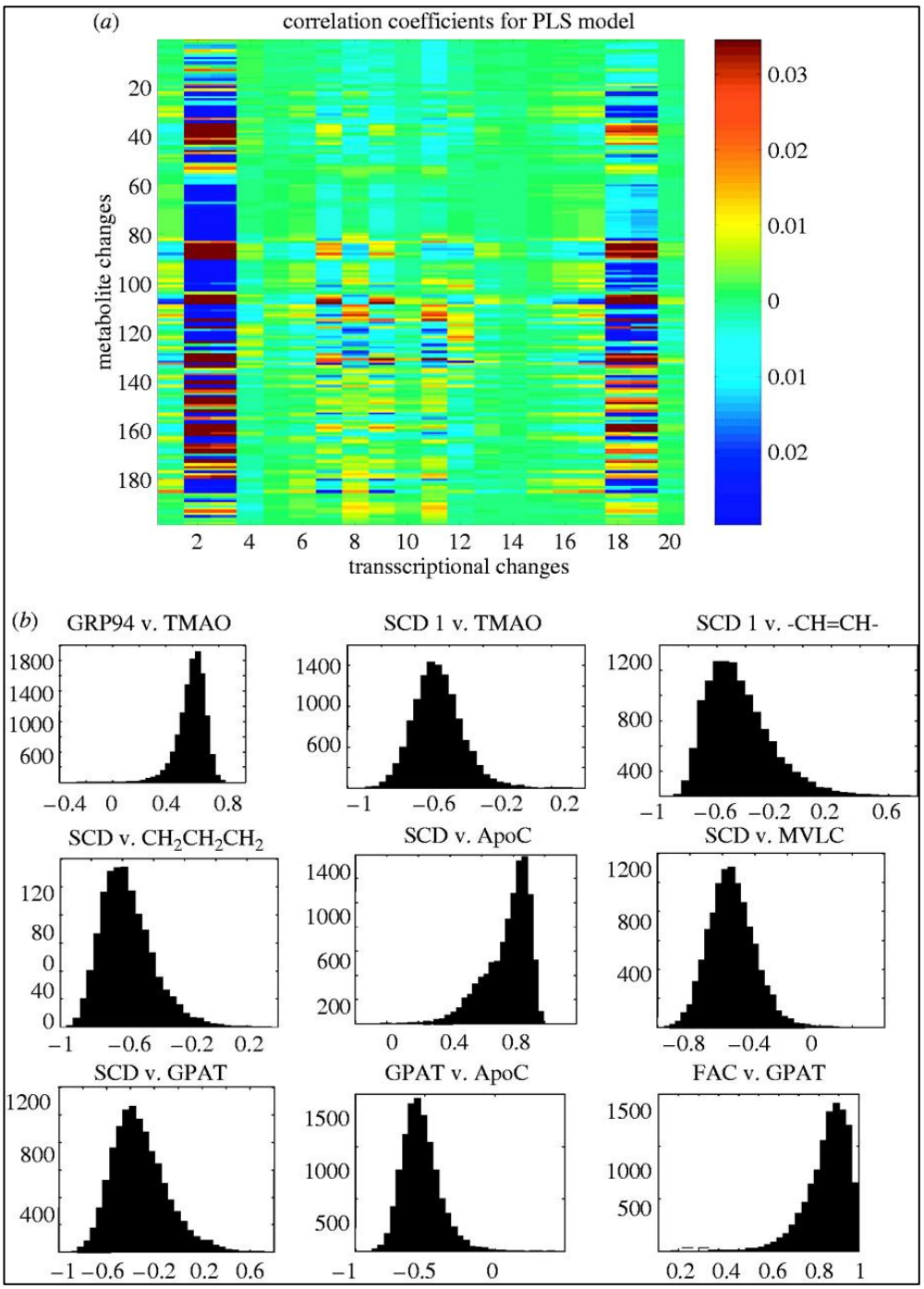


Figure 1.5: Two methods for cross correlating data from different 'omic' approaches

(a) shows a heat map plot of correlation coefficients from a partial least squares model of 20 transcripts (along the x-axis) and 200 metabolites (along the y-axis). Regions with an intense red colour are positively correlated. (b) Histograms of bootstraps for correlation coefficients between key metabolite regions and transcripts. The x-axis represents the correlation coefficients while the y-axis represents the number of times this correlation was returned during 10 000 iterations. Key: SCD, stearoylCoA desaturase 1; ApoC, apolipoprotein C III; MVLC, mitochondrial very long-chain acyl CoA thioesterase; GPAT, glycerol 3 phosphate acyltransferase; FAC, fatty acid CoA ligase 4; CHaCH unsaturated lipid resonance, CH2CH2CH2 saturated lipid resonance (Adapted from Griffin et al. 2004b).

Clish group investigated the alterations in the liver at transcriptomic, proteomic and metabolomic levels. They identified a range of transcripts, proteins and metabolites changed in the ApoE knock-out mouse. Finally, these changes were used to produce a correlation network, highlighting which metabolic pathways were most perturbed (Clish et al., 2004).

Martin et al. studied insulin resistance, obesity, dyslipidaemia, and high blood pressure metabolic syndrome. The factors extracted using the phenotypic, genetic, and genome wide LOD score correlation matrices followed different patterns, which may suggest a distinct effect. The results implied that different methods of multivariate data reduction provide unique clues on the clustering of this complex syndrome (Martin et al., 2003).

The U.S. Department of Health and Human Services (DHHS)/USDA Dietary Guidelines for Americans used similar approach to analyse the nutritional genotype-phenotype relationship which can lead to future individualized dietary recommendations to diminish cancer risk. One of the strategies used, is the systems biology approach: genes, proteins, and metabolite nutrients or nutritionary regimens is catalogued and integrated into a functional metabolism assessment tool to develop biomarkers of early metabolic dysregulation and susceptibility influenced by diet. Through ‘omic’ technologies, all the DNA sequences, mRNAs, all the proteins, a large multiple feedback loop from metabolites to proteins or transcripts involved in a particular cellular process are measured (Go et al., 2005).

This massive database is then integrated and analysed by advanced computer modules and bioinformatics techniques that potentially will lead to the genotypic-phenotypic characterization of different cellular network systems responding to a specific diet (Romero et al., 2005). However, functional analysis is incomplete unless quantification of metabolic fluxes is determined by tracer-based metabolomics (Nielsen, 2003, Lee and Go, 2005). Numerous databases have been created and employed in the development of various biological networks to understand the relationship of diet and cancer during exploration of genotype-phenotype relationship (Chou et al., 2003, Martin et al., 2003, Eerligh et al., 2004).

These approaches will lead to a systems biology understanding as the interplay between genotype, environment, and nutrition in health (Desiere, 2004). Eventually, the future target is to be able to examine personal variations responding to specific diet and provide knowledge of how diet influences metabolic regulation in health and prevention of chronic diseases at the personal level to achieve the goal of personal dietary guideline recommendations throughout an individual's life cycle (Go et al., 2005).

In the context of systems biology, few sparse approaches have been proposed so far to integrate several data sets. It is, however, an important and fundamental issue that will be widely encountered in post genomic studies, when simultaneously analysing transcriptomics, proteomics, and metabolomics data using different platforms, to understand the mutual interactions between the distinct data sets (Le Cao et al., 2009).

Chapter 2: Common gene expression profiling reveals biological alterations involved in early tumourigenesis of Oral Squamous Cell carcinomas

2.1 Introduction

Early stage (I and II) oral cancer may be curable by surgery or radiation therapy alone but advanced cancers (stage III and IV) are generally treated by surgery followed by radiation therapy (Harrison et al., 1999, Idle and Gonzalez, 2007). In fact, survival of advanced –stage patients rarely exceeds 30 months, even for those that initially achieve complete clinical remission (Hill and Price, 1994).

A critical factor in the lack of prognostic improvement is that a significant proportion of cancers initially are asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. Early detection of cancer is the most effective means to reduce death from this disease (Li et al., 2004a). Moreover, most patients are recommended to have radiotherapy and chemotherapy after surgery to reduce the possibility of recurrence. Tumour molecular profiles including transcriptomic and or metabolomic abnormalities are likely to find increasing application in predicting which individuals will benefit from additional treatment and which individuals do not need it.

Cancer may be accompanied by the production and release of a substantial number of proteins, metabolites and/or hormones into the blood, saliva, and other body fluids that could serve as useful markers for assessing prognosis, metastasis, monitoring treatment, and detecting malignant disease at an early stage.

In the perspective of OSCC and HNSCC, possible themes can be addressed with the help of microarray data include distinguishing the disease from normal at the molecular level; determining whether specific biomarkers or profiles are predictive for cancer; and identifying biologic pathways necessarily altered in the tumourigenesis, potentially illuminating novel therapeutic targets (Choi and Chen, 2005).

2.2 Aim of the study

In this study, an attempt was carried out to find common genes that can be used as biomarkers and would allow the development of a screening test having the potential to increase the early discovery and treatment of OSCC, as well as to explore the underlying biological pathways, functions, and molecular processes.

2.3 Materials and methods

Published reports of gene expression profiles of OSCC and HNSCC were identified using Medline database search. A review of the reports was performed to identify genes repeatedly exhibiting substantially altered expression in OSCC and HNSCC compared to normal oral mucosa.

2.3.1 Selection of study samples

2.3.1.1 Sample's site of studies

The selected studies consist of pure samples of oral cavity while studies with other origins of tissue samples (e.g., nasopharyngeal, maxillary sinus, oropharynx, and larynx (Belbin et al., 2002, Sok et al., 2003, Ginos et al., 2004, Reis et al., 2005, Kuriakose et al., 2004)) were excluded from the review as these samples already originated from hetero-origins and can reflect different molecular gene expression profiles regardless of the disease status, whilst some of them are associated with HPV infection.

2.3.1.2 Technology used in the analyses

Current work comprised only studies using DNA microarray or cDNA library technology in measuring gene expression of OSCCs. Studies using other analytical methods (such as; PCR, immunohistochemical, ELISA, RNA sequencing, flow cytometry, optical sensors, proteomic, and metabolomic signature (Leethanakul et al., 2000b, Kuo et al., 2002, Vora et al., 2003, Kuo et al., 2003, Kitajima et al., 2004, Kobayashi et al., 2004, Schneider-Stock et al., 2004, St John et al., 2004, Kim et al., 2006, Madan et al., 2006, Sartini et al., 2007, Shpitzer et al., 2007, Sasahira et al., 2007, Katakura et al., 2007, Bergmann et al., 2008, Zhao et al., 2008, Bagchi and Mills, 2008, Choi et al., 2008, Tan et al., 2008b, Shigeishi et al., 2009, Yuan et al.,

2006, Vairaktaris et al., 2008)) were removed from the initial selection of the studies. Other studies used in vitro cell-lines culture methodology (Moriya et al., 2003, Zhang et al., 2002, Huang et al., 2002, Banerjee et al., 2003, Vigneswaran et al., 2005, Diamond et al., 2008, Sinpitaksakul et al., 2008, Pettus et al., 2009, Al Moustafa et al., 2002, Jeon et al., 2004, Shibata et al., 2005) and those targeting gene expression signature of distant and lymph nodes metastasis of OSCC were also excluded.

2.3.1.3 Status of the collected OSCCs

Studies targeting staging evaluation of oral cancer disease and its metastasis (Zhong et al., 2008, Odani et al., 2006, Nagata et al., 2003, Irie et al., 2004, Marcus et al., 2004, Schmalbach et al., 2004, Warner et al., 2004, Carinci et al., 2005, Chin et al., 2005, Roepman et al., 2006, Belbin et al., 2005, Kato et al., 2006, Tomioka et al., 2006, Sartini et al., 2007, Hensen et al., 2008, Gu et al., 2008, Chiang et al., 2008) were exempted from the current review. However, some studies harvested 2 signatures (Signatures for both early detection and metastasis) were included in this review by implementing their specific signature only related to early detection.

2.3.2 Construction of the list of common genes

Searching for duplicate genes across the 28 studies had difficulty in matching the gene symbols. Since most of genes have more than one synonym (aliases) and not all studies used the same gene synonym in their final list of genes, there was a demand to construct a gene library that contains all reported synonyms of all genes identified in the 28 studies of the review. NCBI's repository for gene-specific information, Entrez Gene, (http://jura.wi.mit.edu/entrez_gene/) was used to collect the information essential for constructing the library. The library can be searched without having access to the internet service (Supplementary file 1b). However, to provide quick and precise list of common genes, all provided genes from the 28 studies were unified to ENSEMBL gene ID using David conversion tool (Huang et al., 2009a, Huang et al., 2009b).

2.3.3 Exploration of the common genes expression in the raw data

"Merge" function in R statistical programming language was used to assess the availability of the common gene expression among the genes in the raw data.

2.3.4 Biological Annotations of significant common genes

Differentially expressed genes in tumours were subsequently examined using tools available from IPA (Ingenuity Systems, version 9.0, www.Ingenuity.com) to define the relative canonical pathways, upstream regulators, and molecular processes that are most commonly perturbed in the early tumorigenic transformation of OSCC and HNSCC, as well as, to inspect the most significant pre-generated networks associated with the uploaded group of genes including those related to OSCC development. IPA knowledge base recruits the distinguished genes that regulated OSCC and HNSCC to generate and highlight the above functions. Canonical pathways analysis identified pathways from IPA library of canonical pathways, which are most significant to the input data set. The significance of the association between the data set and the canonical pathway was determined based on two parameters: (1) A ratio of the number of genes from the data set that map to the pathway divided by the total number of genes that map to the canonical pathway and (2) a P value calculated using Fischer's exact test determining the probability of the association between the genes in the data set and the canonical pathway is due to chance alone.

IPA cannot map pseudogenes, hypothetical genes, ESTs and ambiguous (mapping to multiple) genes as well as duplicates of multiple IDs mapped in one molecule or multiple molecules snapped in one ID. However, the unmapped genes were converted using David conversion tool (Huang et al., 2009a, Huang et al., 2009b) to include them in the subsequent analysis, some were substantial key players in the master networks of the analysis and contributing to the tumorigenic process.

2.4 Results

From 2000 to 2009, more than 60 studies recruiting DNA microarray analyses were reviewed to examine genome wide genetic expression changes associated with the development of OSCC and HNSCC were reviewed. 28 studies were coincided with the criteria previously specified in the methods section and selected as samples for the current study (Table 2.1) (Ziober et al., 2006, Villaret et al., 2000, Leethanakul et al., 2000a, Alevizos et al., 2001, El-Naggar et al., 2002, Mendez et al., 2002, Nagata et al., 2003, Hwang et al., 2003, Tsai et al., 2004, Leethanakul et al., 2003, Gonzalez et al., 2003, Ha et al., 2003, Toruner et al., 2004, Irie et al., 2004, Li et al., 2004a, Whipple et al., 2004, Carinci et al., 2005, Shimada et al., 2005, Chin et al., 2005, Belbin et al., 2005, Dysvik et al., 2006, Tomioka et al., 2006, Lopes, 2007, Li et al., 2006, Suhr et al., 2007, Ye et al., 2008, Chen et al., 2008a, Estilo et al., 2009). Although the selected studies are similar in many features (criteria of study's selection, material and methods), there is a considerable heterogeneity among these studies in study design, number of samples, proportion of tumour-to-stromal cells analysed, variety of microarray platform, and validation methods of the results. Some studies employed split-up of tumour cells from stromal cells by LCM, while others estimated the percentage of tumour cells within tumour specimens before RNA extraction and subsequent analyses.

Table 2.1: Selected studies using DNA Microarray technology for transcriptional analysis of OSCC and HNSCC.

Studies*	Year published	Sample types	Array platform	LCM	Validation
Villaret et al.	2000	16 HNSCC from multiple sites, 22 unmatched controls from multiple different organs (e.g., tonsil, soft palate, esophagus, skin, lung, small intestine, stomach, heart, brain, kidney)	Synteni cDNA (~985)	N	NONE
Leethanakul et al.	2000	5 HNSCC, 5 matched normal tissue controls	Clontech cDNA array (~588)	Y	NONE
Alevizos et al.	2001	5 OSCC, 5 matched normal tissue controls	Affymetrix HuGeneFL (~7000)	Y	qRT-PCR
El-Naggar et al.	2002	12 OSCC, 12 matched normal tissue controls	Research Genetics cDNA arrays GF200 (~ 5184) or GF211 (~4048)	N (90%)	IHC, qRT-PCR
Mendez et al.	2002	26 OSCC, 18 matched and unmatched normal tissue controls	Affymetrix Test-1 and HuGeneFL (~7000 genes)	N (60%)	qRT-PCR
Nagata et al.	2003	15 OSCC, Normal control: pooled mRNA from oral cavity of 58 patients without cancer	Takara IntelliGene Human Cancer CHIP 2.1 cDNA array (~557)	N	IHC, qRT-PCR
Hwang et al.	2003	5 OSCC, 5 matched normal tissue controls	Affymetrix HuGeneFL (~7000)	Y	qRT-PCR
Leethanakul et al.	2003	5 OSCC, 3 cDNA libraries from normal tissue controls	Custom oral cancer-specific cDNA microarray (~384)	Y	NONE
Gonzalez et al.	2003	3 OSCC, 3 matched normal tissue controls	1) Differential display RT-PCR 2) Incyte Genomics UniGEM V cDNA array (~9350) 3) Research Genetics cDNA array GF204 (~5184)	N	IHC, qRT-PCR
Ha et al.	2003	7 primary OSCC, 6 unmatched normal tissue controls	Affymetrix HG-U95A.v2 (~12,650)	N (85%)	qRT-PCR
Tsai et al.	2004	22 OSCC, 22 pair-wise matched normal tissue controls	MillenniaChip II (~1177)	N	Northern blotting, qRT-PCR
Toruner et al.	2004	16 OSCC, 4 matched normal tissue controls	Affymetrix Hu133A (over 39,000 transcripts)	Y	qRT-PCR
Irie et al.	2004	11 OSCC, 11 normal tissue controls	Human Cancer 1.2 Atlas cDNA array (Clontech)	Y	NONE
Li et al.	2004	32 salivary sample from OSCC patients, 32 salivary	Affymetrix HG U133A (over 39,000	N	qRT-PCR

Continue

Studies*	Year published	Sample types	Array platform	LCM	Validation
		samples from individuals without OSCC	transcripts)		
Whipple et al.	2004	29 OSCC, 19 unmatched normal tissue controls	Affymetrix TM oligonucleotide (~7070)	N	NONE
Carinci et al.	2005	8 OSCC, 11 unmatched normal tissue controls	custom array-Ontario Cancer Institute (~19,200)	N	NONE
Shimada et al.	2005	58 OSCC, 58 matched normal tissue controls	in-house cDNA microarray (~2304)	N (80%)	qRT-PCR
Chin et al.	2005	7 OSCC, 7 matched normal tissue controls	custom array-Ontario Cancer Institute (~19,200)	N	IHC
Belbin et al.	2005	9 OSCC, 9 matched normal tissue controls	Custom cDNA clones (~17840)	N (70%)	IHC
Dysvik et al.	2006	11OSCC, pool of 8 matched normal tissue controls	human cDNA DNR (~15,000)	N (70%)	IHC, qRT-PCR
Tomioka et al.	2006	14 OSCC, 14 Pair-wised control tissue samples	16 617 human cDNA fragments	N	qRT-PCR
Ziober et al.	2006	13 OSCC, 13 Pair-wised control tissue samples	Affymetrix HG U133A (over 39,000 transcripts)	N (80%)	IHC, qRT-PCR
Li et al.	2006	32 serum sample from OSCC patients, 32 serum samples from individuals without OSCC	Affymetrix HG U133A (over 39,000 transcripts)	N	qRT-PCR
Suhr et al.	2007	15 OSCC, 15 Pair-wised control tissue samples	Operon human oligonucleotide microarrays (~34,580)	N (70%)	qRT-PCR
Lopes et al.	2007	48 primary OSCC, 8 unmatched normal tissue controls	Affymetrix Human Genome U133 Plus 2.0 (over 47,000transcript+ FOCUS Array (over 8,500)	N (65%)	qRT-PCR
Ye et al.	2008	26 Tongue OSCC, 12 matched tissue controls	Affymetrix Human Genome U133 Plus 2.0 (over 47,000 transcripts)	N (80%)	IHC, qRT-PCR
Chen et al.	2008	167 OSCC, 45 unmatched normal tissue controls	Affymetrix Human Genome U133 Plus 2.0 (over 47,000 transcripts)	N	qRT-PCR
Estilo et al.	2009	16 Tongue OSCC, 22 matched tissue controls	Affymetrix HG_U95Av2 (~12,625)	N	qRT-PCR

LCM: laser capture microdissection; Y: yes; N: no; qRT-PCR: quantitative real-time polymerase chain reaction; IHC: immunohistochemistry; * Summary of head and neck carcinoma genetic profiling literature. Studies were identified by Medline database search, and are listed according to the year published. Microarray platforms used, array data validation methodologies, and whether laser capture microdissection was employed are listed for each study. Minimum percentage of tumour cell content of samples in non-LCM studies is shown in parentheses.

2.4.1 Gene expression profile of OSCC versus normal oral mucosa

Reviewing the reports identified genes that constantly differentially expressed among the studies with significantly altered expression in OSCC and HNSCC relative to normal oral mucosa (Table 2.2 and Table 2.3). A total of 368 genes were common among at least three studies of the total 28 review studies, while 63.3% (233) downregulated, only 36.6% (135) were upregulated genes in tumours. Among the genes identified commonly altered in these studies, overexpression of several genes encoding for matrix metalloproteinases (MMP1, MMP3, MMP7, MMP9, MMP10, MMP12, and MMP13), collagens (COL1A2, COL3A1, COL4A1, COL5A2, COL4A2, COL5A1, COL6A3, COL7A1, and COL11A1), C-X-C chemokines (CXCL1, CXCL2, CXCL6, CXCL9, CXCL10, and CXCL11), interferons (IFI6, IFI27, IFI44, and IFI44L), transforming growth factors (TGFB1, TGFBR3, and TGFBI), tumour necrosis factors (TNFAIP1, TNFAIP6, and TNFSF10), laminins, fibronectins, tenascin C, along with perturbation of interleukins, integrins, serpin peptidase inhibitors. These accompanied with repression of DNA binding inhibitors, protein phosphatase 1 regulatory (inhibitor) subunits, several genes encoding for keratins, especially those frequently involved in OSCC and HNSCC including aerodigestive tract cancers (KRT4, 13, 15, and 19).

Table 2.2: Common genes significantly higher in OSCC tumours compared to normal oral tissue samples.**

GENE SYMBOLE	NO.*	GENE SYMBOLE	NO.*	GENE SYMBOLE	NO.*	GENE SYMBOLE	NO.*
MMP1	11	COL5A1	4	CXCL11	3	SERPINH1	3
COL1A2	9	COL6A3	4	CXCL6	3	SFN	3
KRT17	9	COX4I1	4	CXCL9	3	SLC39A14	3
PLAU	9	CTNNA1	4	DFNA5	3	SLC7A11	3
SOD2	9	CXCL1	4	EBNA1BP2	3	TEAD4	3
POSTN	8	CXCL2	4	FAP	3	THBS1	3
SPARC	8	EXT1	4	FLNA	3	TNFAIP1	3
CDH3	7	FAT1	4	FSCN1	3	TNFAIP6	3
COL1A1	7	HIF1A	4	FTH1	3	TNFSF10	3
COL4A1	7	IFI44	4	FTL	3		
FN1	7	IL1B	4	GBP1	3		
IFI6	7	ITM2A	4	GJA1	3		
LAMC2	7	KRT14	4	GSK3B	3		
MMP12	7	LAMA3	4	GSTO1	3		
MMP3	7	LAMB1	4	H3F3A//H3F3B	3		
STAT1	7	LGALS1	4	HMGA2	3		
COL3A1	6	LOXL2	4	HOMER3	3		
IL8	6	MMP7	4	IFI27	3		
ISG15	6	PDPN	4	IFI44L	3		
LUM	6	PRNP	4	INHBA	3		
MMP10	6	PTHLH	4	IGFBP7	3		
MYO1B	6	RHOA	4	LAMB3	3		
SERPINE1	6	SLC16A1	4	LY6E	3		
SPP1	6	SOX4	4	MICAL2	3		
TGFBI	6	TGFB1	4	MMP13	3		
TPBG	6	THBS2	4	MMP9	3		
ADH1A/B/C	5	THY1	4	MSN	3		
BST2	5	TMSB10	4	MYO10	3		
COL5A2	5	TPM1	4	NRG1	3		
CTSL1	5	VEGFA	4	OASL	3		
CXCL10	5	APOL2	3	PKM2	3		
ITGA6	5	BIRC3	3	PLA2G7	3		
KRT16	5	C5ORF13	3	PLAUR	3		
KRT5	5	CALD1	3	PPIAL3	3		
RAB31	5	CASP1	3	PSMB2	3		
TNC	5	CD44	3	PXDN	3		
ACTN1	4	CD47	3	RAN	3		
AIM2	4	CFL1	3	RBP1	3		
APOL1	4	CHST2	3	REL	3		
CAV1	4	COL11A1	3	RHOC	3		
CAV2	4	COL7A1	3	S100A2	3		
COL4A2	4	CTSC	3	SCRN1	3		

* Number of studies where the gene exists. ** Full names and genes IDs presented in Supplementary file 2.

Table 2.3: Common genes significantly lower in OSCC tumours compared to normal oral tissue samples.**

GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*
KRT4	11	CCNG2	4	ANXA1	3	DUSP1	3	MALL	3	SREBF2	3
KRT13	10	CD24	4	ANXA11	3	DUSP6	3	MAOA	3	SRPX	3
MAL	10	CEACAM5	4	AP2S1	3	EHD3	3	MAPK9	3	STK39	3
TGM3	8	CEACAM7	4	AP3D1	3	EHF	3	MC1R//TUBB3	3	THBD	3
EMP1	7	CKMT1A// CKMT1B	4	APP	3	EIF3B	3	MFGE8	3	TJP1	3
KRT15	7	CLU	4	APPBP2	3	ENO3	3	MICA	3	TM7SF2	3
RPL10 //4	7	CRCT1	4	ASAH1	3	ERBB2	3	MXD1	3	TPM2	3
ALDH3A2	6	CRIP1	4	ATP1A2	3	ERBB3	3	MYH11	3	TPM4	3
ALDH9A1	6	CRNN	4	ATP5A1	3	ERCC5	3	NCOA1	3	TRIM13	3
CES2	6	CSTB	4	BACH1	3	ERO1L	3	NET1	3	TRIO	3
CLEC3B	6	EVPL	4	BID	3	ESPL1	3	NFE2L2	3	TRPS1	3
CRISP3	6	GNAI1	4	BPGM	3	FAM107A	3	P4HB	3	TSPAN3	3
ECM1	6	HSPB8	4	C10ORF116	3	FOXN3	3	PCOLCE	3	TSPAN6	3
GPX3	6	IGFBP5	4	C2ORF54	3	GLTSCR2	3	PEBP1	3	TUBA1A	3
HPGD	6	IK	4	CAPG	3	GPD1L	3	PGD	3	TYRP1	3
KRT1	6	KRT10	4	CAPN3	3	HBA1//HBA2	3	PKP4	3	UBE2C	3
MAOB	6	KRT76	4	CAST	3	HBB	3	PLAGL1	3	UBE2L3	3
ADH7	5	MEIS1	4	CCDC6	3	HRAS	3	PMM1	3	UCK2	3
CRYAB	5	NDRG2	4	CDKN1A	3	HS3ST1	3	PPP1R7	3	UGP2	3
DNASE1L3	5	NT5C2	4	CEACAM1	3	ID2	3	PSCA	3	VAV3	3
HLF	5	NUCB2	4	CEACAM6	3	ID3	3	PTP4A1	3	VCAN	3
ID4	5	PDCD4	4	CEBPD	3	IDH3G	3	QARS	3	WNK1	3
IL1RN	5	PHLDA2	4	CETN3	3	IL18	3	RCAN2	3	XBP1	3
KRT19	5	PPL	4	CLCA4	3	IL1F6	3	RHCG	3		
LEPR	5	PPP1R1A	4	CLIC3	3	IRS1	3	RNASE4	3		
MGLL	5	PRSS3	4	COL18A1	3	ITGA2	3	RORA	3		
MGST2	5	RPL10A	4	COX5B	3	ITGB1	3	RRAGD	3	***Continue***	
NEBL	5	SCNN1A	4	COX7A1	3	IVL	3	S100A11	3		

GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*	GENE SYMBOL	NO.*
NMU	5	SP100	4	CSNK1G2	3	JUP	3	S100A14	3		
PPP1R3C	5	TGFBR3	4	CUL3	3	KAT2B	3	SCNN1B	3		
PTK6	5	TGM1	4	CYB5A	3	KLK11	3	SCP2	3		
SCEL	5	TSPAN8	4	CYP2C18	3	KLK13	3	SERPINB13	3		
SERPINB1	5	TYRO3	4	CYP3A5	3	KRT7	3	SERPINB2	3		
SPINK5	5	UBL3	4	CYP51A1	3	LCN2	3	SFRS5	3		
TNXA//TNXB	5	UGT1A7	4	DBI	3	LGALS3	3	SKP1	3		
ZNF185	5	ABCD3	3	DEFB1	3	LGALS7	3	SLC26A2	3		
ABCA8	4	ACAA2	3	DEGS1	3	LIMK2	3	SLC9A3R1	3		
ACPP	4	ACYP1	3	DHCR24	3	LMO4	3	SLURP1	3		
ALOX12	4	AIM1	3	DMD	3	LOX	3	SMAD4	3		
BLNK	4	ALDH3A1	3	DPT	3	LPIN1	3	SPARCL1	3		
BTF3P11	4	ALOX12B	3	DR1	3	LTA4H	3	SPINT1	3		
CCL14//CCL15	4	ALOX5	3	DSG1	3	LYN	3	SPRR2C	3		

* Number of studies where the gene expressed. ** Full names and genes IDs presented in Supplementary file2.

2.4.2 Exploring the set of common genes in a raw data

Because of the lack of consistency of the expressed common genes across the 28 studies, an attempt was carried out to search for the common genes expression in raw data. The raw data selected was belong tumour versus normal from UK population which will be analysed separately in chapter 4. A significant overlapping of the common genes over the genes from the raw data was observed while the expression (whether up- or down-regulated) of all overlapped genes were constant between both groups of common and raw data genes (Figure 2.1).

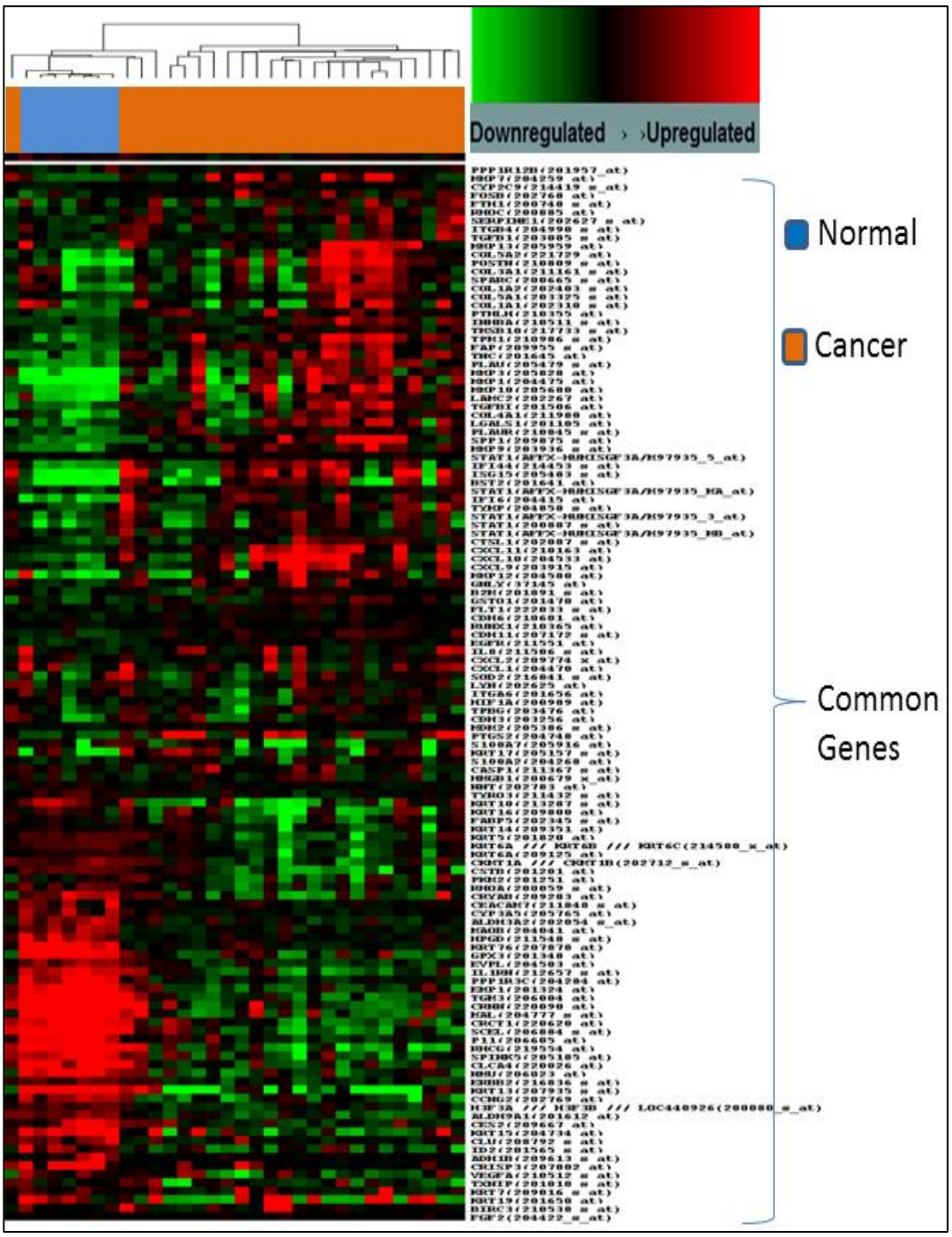


Figure 2.1: Hierarchical clustering displaying common genes overlapped over raw data of UK tumour versus normal samples.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents absolute expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on absolute expression scale.

2.4.3 Biological interpretation of the common expressed genes

To obtain more biological understanding of why expressed genes in OSCCs are different to normal oral tissue samples, the Ingenuity pathway analysis was used to examine gene molecular function, biological process, canonical pathways, and upstream/transcriptional regulators analysis. IPA system depends on the knowledge database of selected functional and regulatory interactions extracted from the literature and provides integrated graphical representation of the biological relationships between genes and gene products considering both up- and down-regulated genes from the comparison analysis.

2.4.3.1 Functional analysis

Cellular Movement, Growth, Proliferation, Development, Death and Survival, and Immune cell trafficking functions displayed at the top of the perturbed molecular cellular functions in relation to cancer (Figure 2.2).

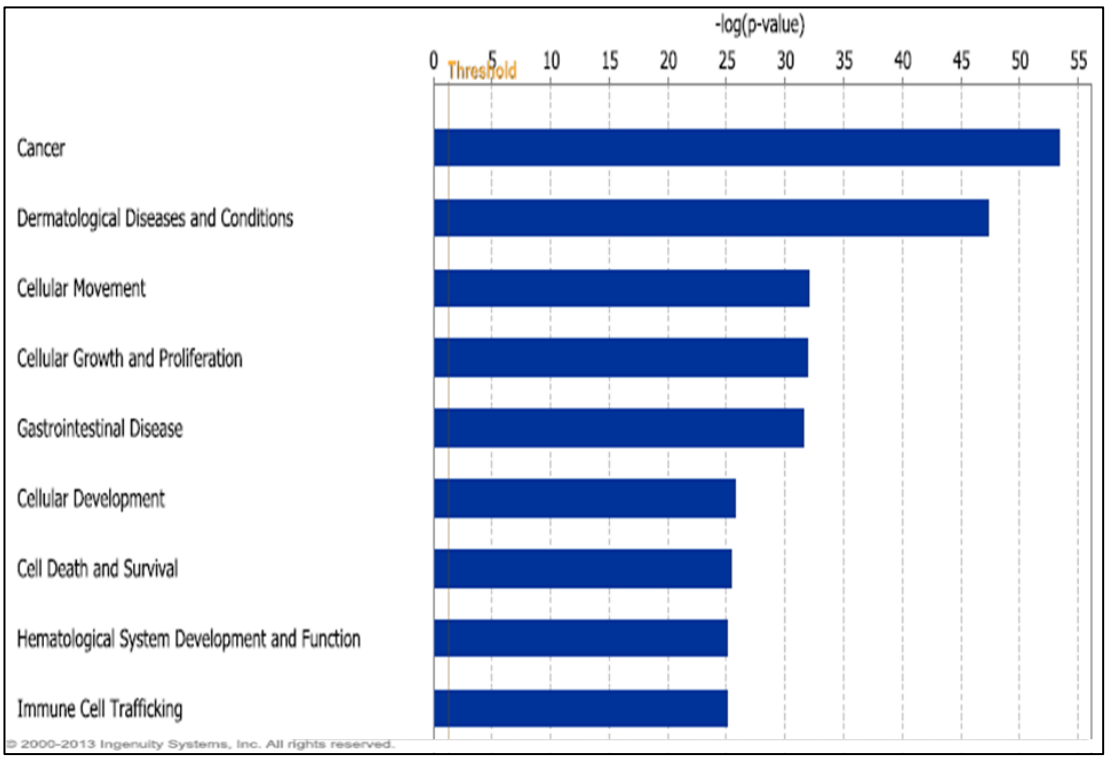


Figure 2.2: Molecular cellular functions of common genes related to OSCC using IPA functional annotation analysis.
Higher bars denote greater significance and the numbers of genes above the yellow line, which signifies a p value of 0.05, represent those that are significantly differentially expressed than by chance alone

2.4.3.2 Associated network analysis

Investigation of the biological relationships between genes and gene products was carried out by performing a network analysis of the common genes. A master network was constructed by connecting the selected genes involved in cancer using Fisher's exact test. The highly connected nodes are likely to reflect their ability to regulate a large number of genes in the master network and potentially to control the gene expression pattern identified tumour signature of the comparison. Over-expression of VEGFA, HIF1A, TGFB1, IL8, IL1B, FN1, SPP1, STAT1, SERPINE1, MMP1, MMP3, MMP7, MMP9, MMP13, TNC, RHOA, PLAU, CXCL10, CXCL2, THBS1, COL1A1, COL1A2, COL3A1, and PTHLH along with under-expression of CDKN1A, ERBB2, ERBB3, IGFBP5, ITGB1, LGALS3, SMAD4, HRAS, NFE2L2, MAPK9, and DUSP1 gene nodes were identified at the centre of the network (Figure 2.3).

Additional, more specific master network was constructed by highlighting the only selected genes involved in OSCC and HNSCC using Fisher's exact test (Figure 2.4).

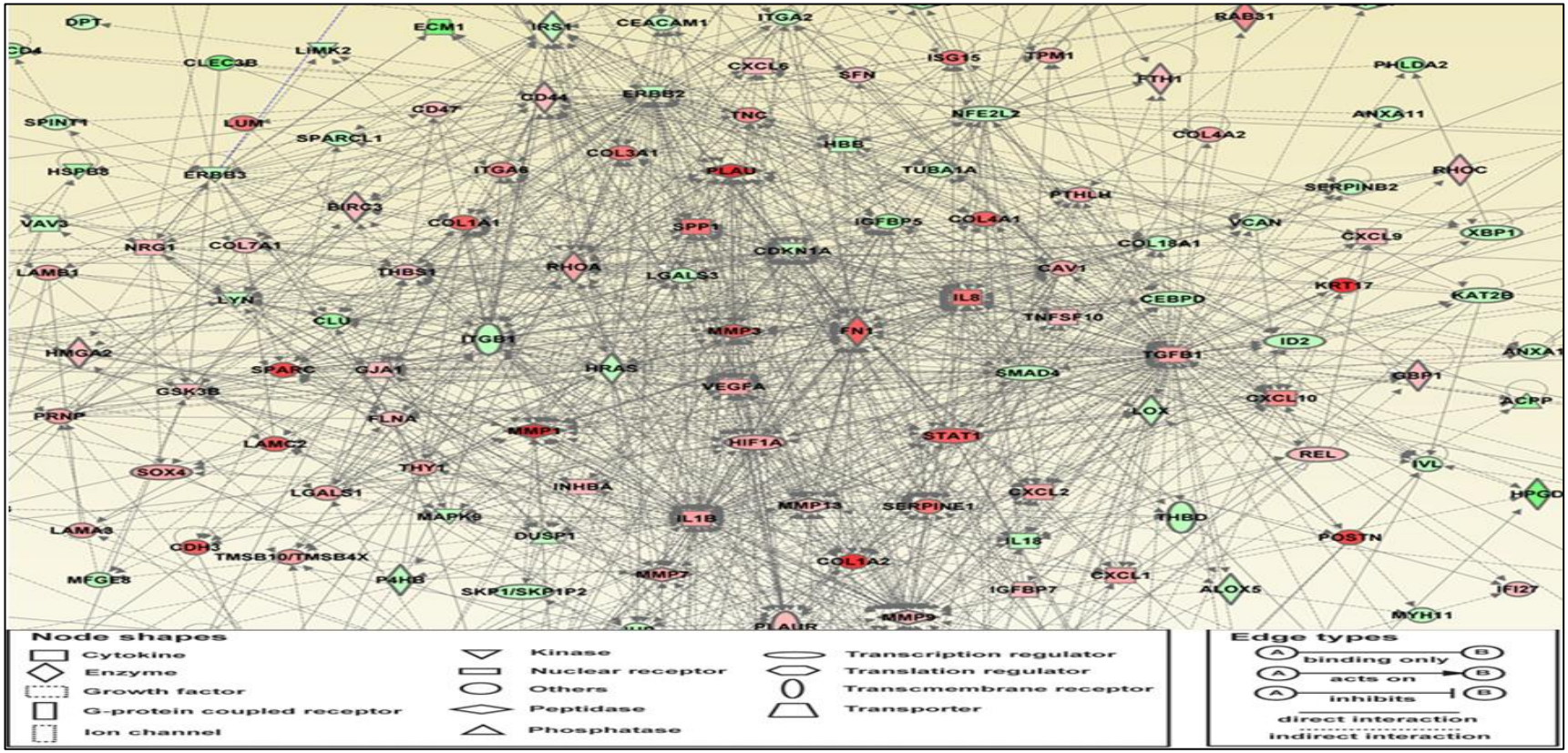


Figure 2.3: Central area of the comprehensive molecular network of the common genes related to cancer.
Master network built by connecting significant genes involved in cancer disease, identified by IPA tool (version 9.0) from up- and downregulated genes using overlap function of core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The network represents significant genes common among the review studies in at least 3 studies. For full view of the network, see supplementary file 3.

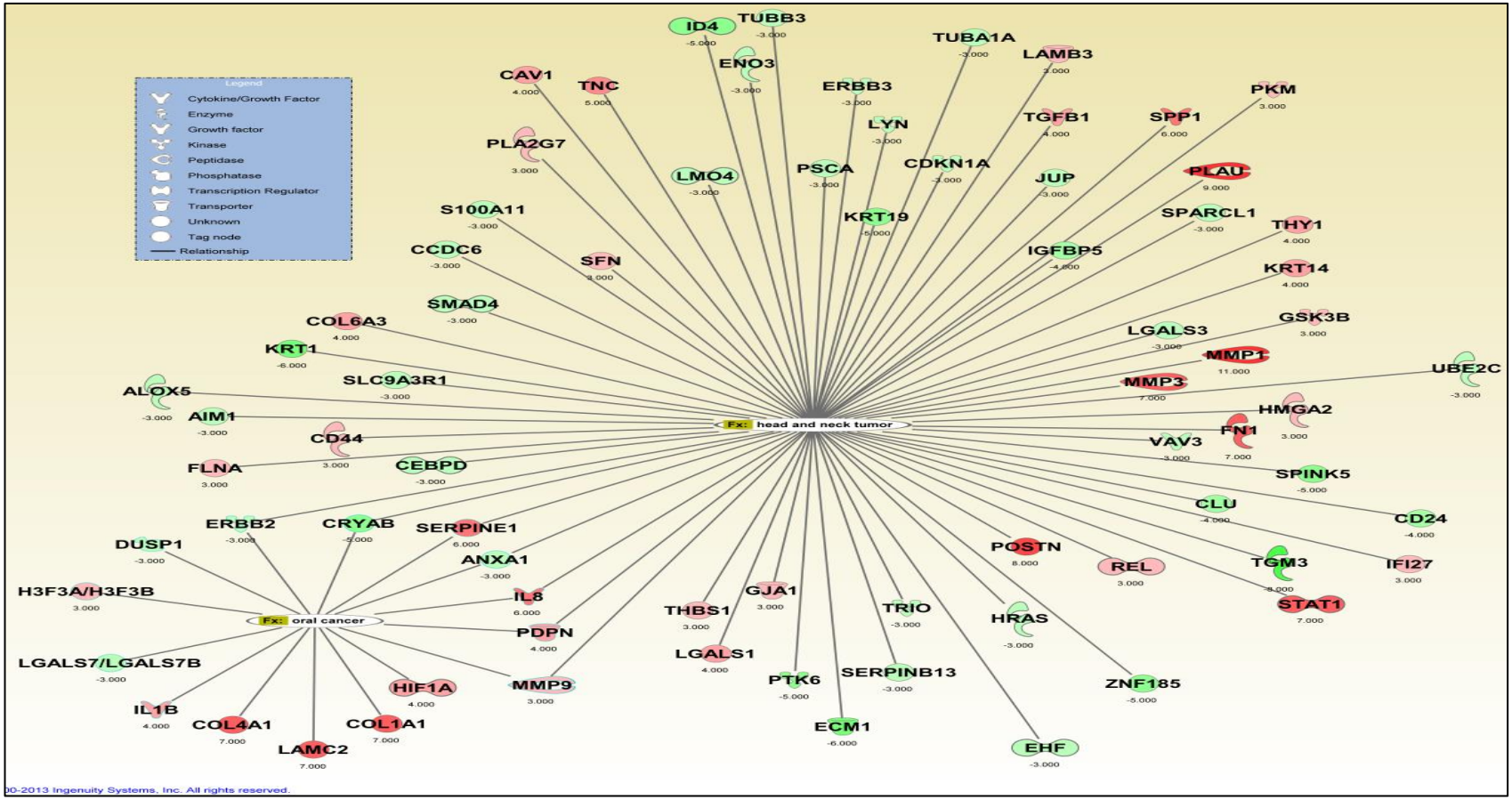


Figure 2.4: Master molecular network of the common genes related to OSCC and head and neck cancer.
Master network built by identifying significant genes involved in OSCC and HNSCC by IPA tool (version 9.0) from up- and downregulated genes using overlap function of core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship events between the nodes and the tumour types). The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The numbers displayed under the nodes represent the number of studies where the gene exists. The nodes are represented using various shapes that represent the functional class of the gene products.

2.4.3.3 Upstream regulators

It's well known that regulation of human genes depends on sets of transcription factors rather one individual factor. For this reason "Upstream and Transcriptional Regulator Analysis" tool available from IPA system was used to identify several key players in the biological context. These include activation of CTNNB1, JUN, NFKBIA, and STAT3, along with inhibition of TP53 and MYC transcriptional regulators. In addition, activation of TGFB1, EGF, and EGFR growth factors and growth factor's receptor was accompanied with activation of TNF and IFNG cytokines along with repression of ERBB2 cytokine.

2.4.3.4 Canonical pathway analysis

By utilizing canonical pathway analysis tool (IPA), Role of Tissue Factor in Cancer, Leukocyte Extravasation Signalling, Inhibition of matrix metalloproteases, HIF1a Signalling, ILK Signalling, and Xenobiotic Metabolism signalling displayed as the most common perturbed pathways involved in tumour transformation. Previous works extensively described the role of these pathways during tumour transformation. The determination of the significant dysregulated pathways is accomplished by overlapping the input common genes over the pre-generated pathways. Figure 2.5 displays the top pathway dysregulated in this study (Role of Tissue Factor in Cancer).

2.4.3.5 Biomarker analysis

In an attempt to find potential biomarkers that reported an involvement in OSCC and HNSCC among the common genes set, running biomarker filter analysis (IPA) identified 27 specific biomarkers related to OSCC and HNSCC (Figure 2.6). The figure describes clearly the particular function of each biomarker related to specific tumour type.

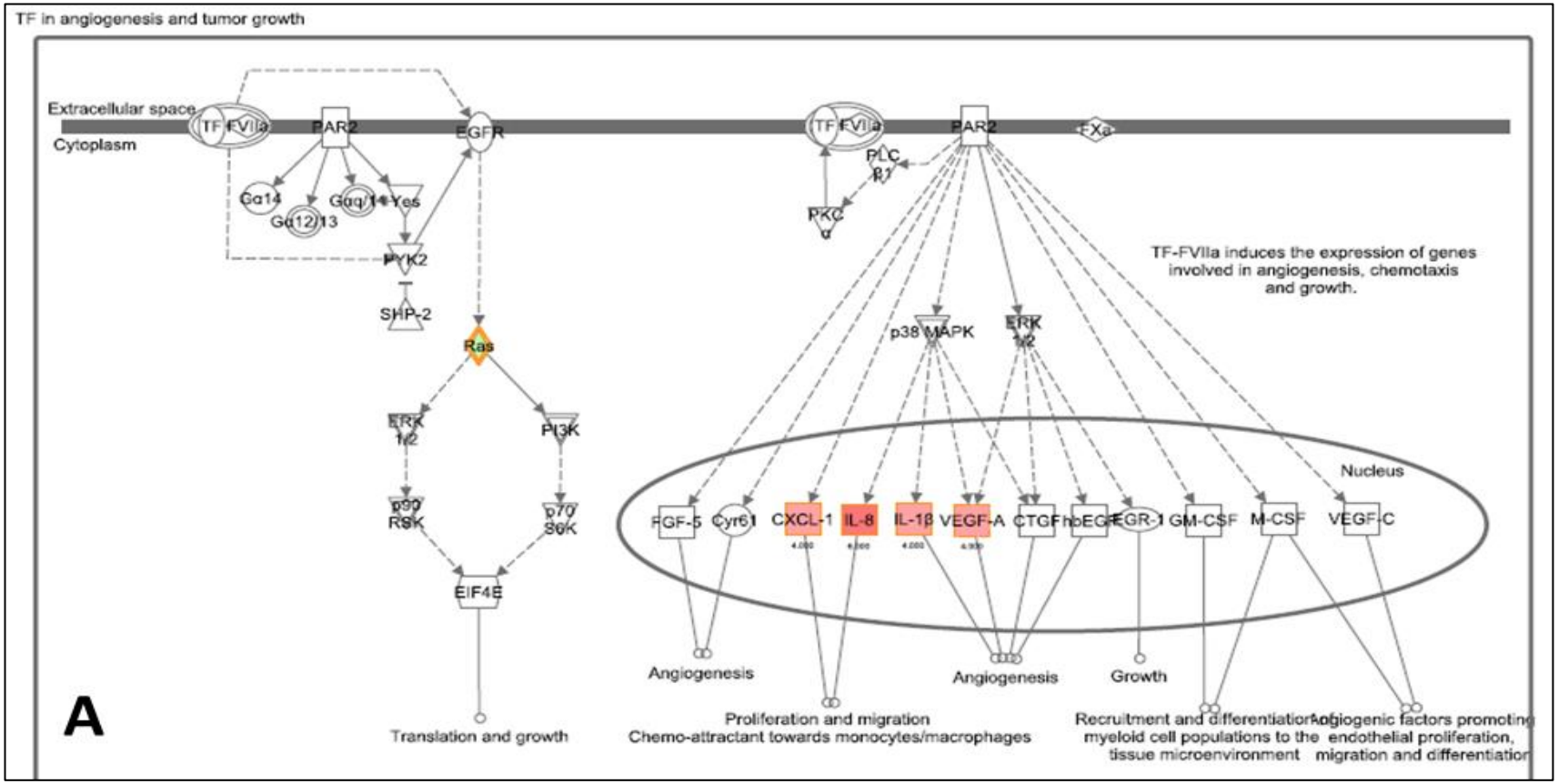


Figure 3.4.: Highly involved biological pathway in OSCC, The Role of Tissue Factor in Cancer Pathway; A- TF in angiogenesis and tumour growth.

Red denotes upregulation and green denotes downregulation of the gene. Ingenuity Systems, (version 9.0). Full comprehensive view of the pathway is available in supplementary file 4.

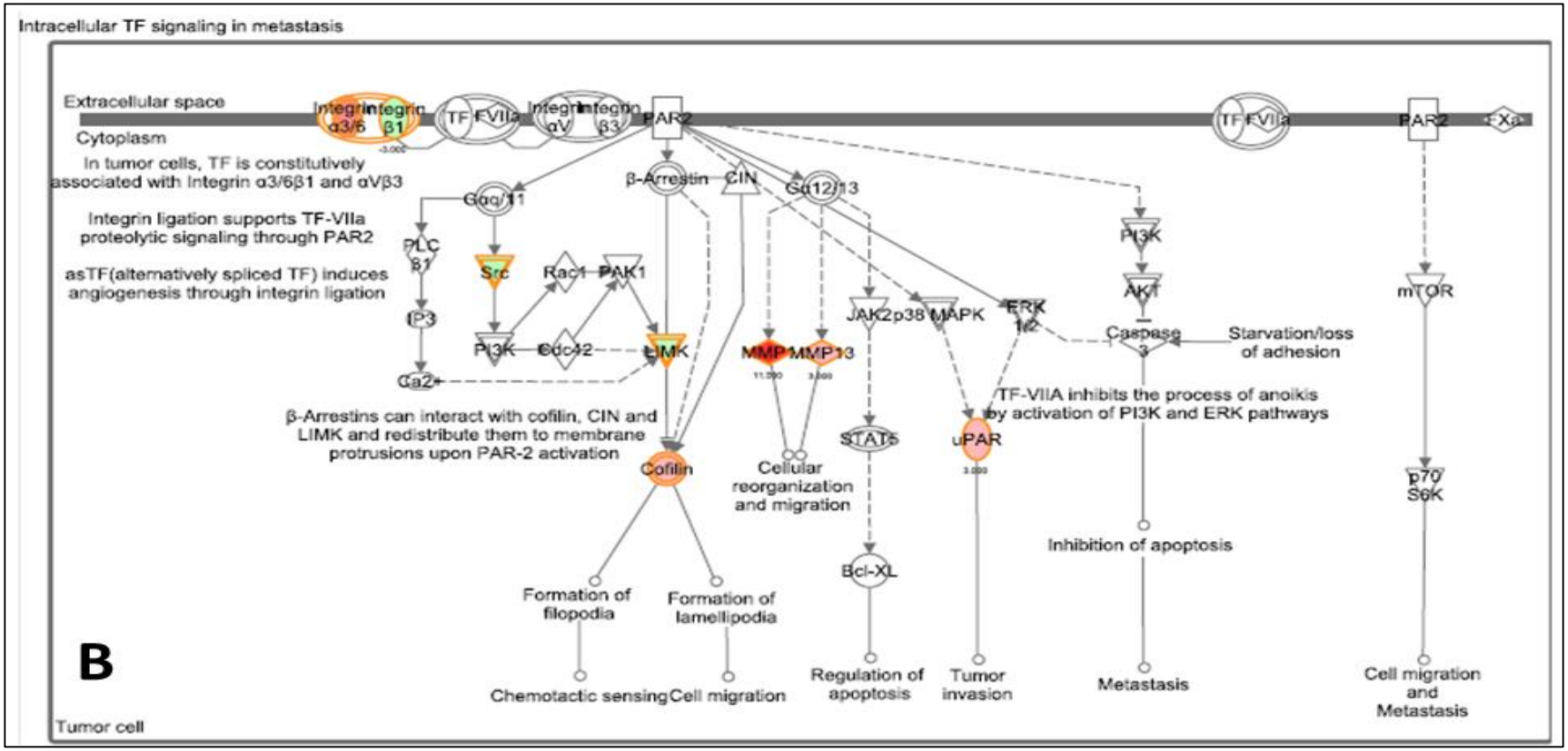


Figure 2.5: Highly involved biological pathway in OSCC, The Role of Tissue Factor in Cancer Pathway; B- Intracellular TF in metastasis.

Red denotes upregulation and green denotes downregulation of the gene. Ingenuity Systems, (version 9.0). Full comprehensive view of the pathway is available in supplementary file 4.

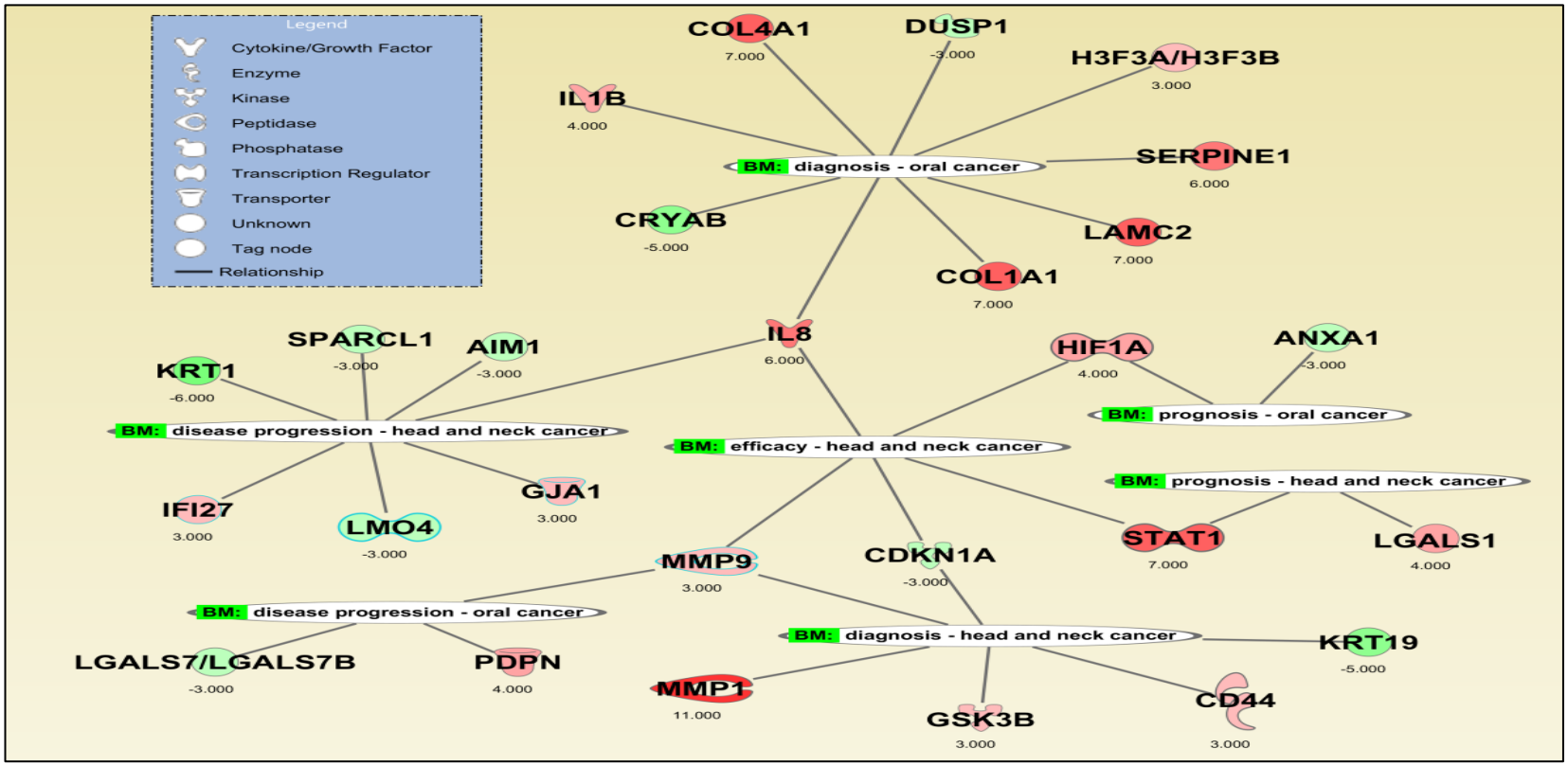


Figure 2.6: Common genes biomarkers related to OSCC and HNSCC.
Biomarker network built by connecting significant genes involved in OSCC and HNSCC, identified by biomarker filter tool (IPA, version 9.0) from up- and down-regulated common genes using overlap function of the biomarker analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship events between the nodes and the tumour types). The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. The numbers displayed under the nodes represent the number of studies where the gene exists.

2.5 Discussion

The review study specified several genes and pathways exhibiting substantially altered expression in cancerous versus noncancerous states across studies. Among the genes identified commonly altered in these studies, included overexpression of several genes encoding for matrix metalloproteinases, collagens, C-X-C chemokines, interferons, transforming growth factors, tumour necrosis factors, laminins, fibronectins, tenascin C, along with perturbation of interleukins, integrins, serpin peptidase inhibitors. These accompanied with repression of DNA binding inhibitors, protein phosphatase 1 regulatory (inhibitor) subunits, several genes encoding for keratins, especially those frequently associated with OSCC and HNSCC including aerodigestive tract cancers (KRT4, 13, 15, and 19). Furthermore, both activation and inhibition of the upstream and transcriptional regulators detected in the result section have been reported to have a major contribution in the tumorigenic process. In fact, according to the knowledge base of IPA, Figure 2.4 displayed 76 (over 20% of the total common genes) genes associated with OSCC and HNSCC where 27 of them are verified biomarkers (Figure 2.6).

Over-expression of chemokines in tumours has been ascribed frequently to constitutive stimulus of NFκB, the signalling pathway which plays a key role in anti-apoptosis and is widely proposed now as potential molecular tumorigenic marker in various types of tumours including colorectal, bladder, pancreatic, breast, ovarian, as well as oral cancer (He et al., 2009). Elevated level of several members of CXC family in tumours was another finding which enhances angiogenesis (Vandercappellen et al., 2008, Vinader and Afarinkia, 2012, Zhu et al., 2012). These include CXCL8 (IL8), CXCL1, CXCL2, CXCL3, CXCL5, CXCL6, and CXCL7. Besides their functions in the immune system, they also play a critical role in tumour initiation, promotion and progression (Vandercappellen et al., 2008). C-X-C chemokines 2, 3, 9, and genes encoding for extracellular matrix proteins and collagens have previously been involved in the motility and invasion of tumour (Chen et al., 2008a).

Notably, some of these changes, although identified in multiple reports, may also present in other sample studies which are not identified, basically due to the absence of probes for these genes in the platforms used for these studies. Arrays used by different research groups contained probes for as few as 384 genes to as many as over 47,000 transcripts (Table 2.1). These results were encouraging considering the heterogeneity of the studies where the samples are collected. Although, the criteria of studies' selection in this review reduced the impact of anatomic specific genetic expression differences might have on comparison of OSCCs by selecting only tumours from oral cavity, considerable heterogeneity among these studies in study design, number of samples, ratio of tumour-to-stromal cells analysed, and choice of microarray platform were still existed. Some studies employed separation of tumour cells from stromal cells by LCM, whereas others estimated the percentage of tumour cells within tumour samples before RNA extraction and subsequent analyses. The impact of interaction between tumour cells and their surrounding stromal and non-tumour constituents (e.g. immune response cells) is well recognized in tumours of human epithelium and other tissue of the body (Moinfar et al., 2000, Matrisian et al., 2001, Mueller and Fusenig, 2002, Chung et al., 2005). Comprehensive genetic profiling of OSCC and HNSCC, therefore, must not calculate only cancer cell-specific gene expression, but the surrounding stromal cell expression as well, to provide complete understanding of altered gene expression. Furthermore, the most significant source of heterogeneity among microarray tumour profiling studies was derived from the various methods of data generation and analysis (Choi and Chen, 2005). Although these findings may serve focusing in selecting markers for further analysis to improve understanding and management of OSCC and HNSCC, limitations to the function of the list of genes and pathways were noticed in the current study. Considering the top differentially expressed gene, only 11 out of 28 studies found common, suggesting doubts about the reasons making that gene absent in the other remaining studies. Other genes from the list of common genes were much poorer in their common existence (Table 2.2 and Table 2.3). Perhaps the largest downside of microarray based examination of HNSCC is the lack of well defined standards for their use, interpretation, and validation. Variability in tissue procurement, tumour cell isolation, RNA extraction, choice of array platform, and

data normalization and analysis preventing rigorous comparison of results from one laboratory with those from others. Furthermore, estimation of tumour-to-stromal cell ratios in tumour samples by histological evaluation of adjacent tissue sections is imprecise. This is mainly due to the nature of HNSCCs, which are not simple, homogeneous masses with discrete edges, but rather complex, three-dimensional entities with varying degrees of specular projections and interlaced stromal components, not to mention the potential clonal heterogeneity among tumour cells themselves. Estimation of percentage tumour cell content within one or a few histologic sections of a tissue sample does not ensure the same percentage tumour cell content within the entire sample. Nevertheless, estimated tumour-to-stromal cell ratios can be useful in establishing the simple presence or absence of cancer cells, and repeating the pathologic diagnoses of experimental samples (Choi and Chen, 2005). The current review pointed out to a popular significant inaccuracy among the studies using microarray technologies in underestimating the tumour origin of HNSCC during sample collection. A study used microarray design demonstrated that HNSCCs originating from different anatomical locations can exhibit varying behaviours and concluded that HNSCC tumours originating from different anatomic sites share consistent changes in gene expression when comparing primary tumours to normal adjacent mucosa, where these common changes most likely reflect alterations required for tumour development. In contrast, once a tumour has developed, tumour-host interactions at different anatomic sites are likely responsible for the site-specific signatures associated with aggressive versus non-aggressive disease. Hence, predictions of outcome based on gene expression profiling are therefore heavily influenced by the anatomical site of the primary tumour (Belbin et al., 2008).

The top differentially expressed gene in 11 out of 28 studies is considered a poor outcome in finding unique identifier/s for OSCC. The weakness of tumour profiling of the list of common genes originates primarily from limited sample sizes and heterogeneity in the experimental design and execution, as well as the differences in the size of the final set of genes published by the individual studies. One study provided only 9 genes from the whole set of microarray genes without revealing the

other significant genes harvested before the final filter of their statistical analysis (Gonzalez et al., 2003). Published data from such studies should not viewed as endpoints of research accomplishments, but rather as screening tools of identifying potential genes for validation and further investigation. Nevertheless, the capability to measure the expression of thousands of genes simultaneously gives researchers a powerful method to analyse global genetic events responsible for HNSCC progression (Choi and Chen, 2005).

The attempt of searching common genes among raw data of UK tumour versus normal samples displayed clearly the satisfactory overlapping between the two gene sets (common genes vs. genes of raw data). This finding confirms the impact of using different post-lab analytical approaches in modifying and shaping of the procurement final list of genes.

This review did not access raw data expression of most of the studies, and therefore did not consider how different fold changes in gene expression might impact different clinical outcomes or biological pathways. Ideally, all published studies containing DNA microarray analyses should deposit their raw expression data in publically accessible repositories (e.g., NCBI GEO, Array express, others), to permit accomplishment of systematic and comprehensive meta-analysis. For this reason, additional attempt was carried out (chapter 3) to perform meta-analysis profiling OSCCs relative to normal mucosa by recruiting similar bioinformatics approach on multiple raw data sets.

Conclusion of the current review suggests not only relying on the final set of genes published by the individual studies, but to access the raw data of each study and start subsequent analysis from that stage using unified bioinformatics approaches to acquire useful and complete understanding of the systems biology and identifying the subsequent dysregulated biological pathways, molecular functions, and processes.

Chapter 3: Detection of a robust gene signature for oral squamous cell carcinoma by integrating multiple microarray datasets

3.1 Introduction

The pathogenic pathways that contribute to OSCC remain poorly characterised (Liu et al., 2012); hence, a clinically applicable gene expression signature is in high demand and improved characterization of the OSCC gene expression profiles would constitute substantial progress. Analysing the publicly available gene expression datasets has already proven new promising insights on biological processes.

The review study (chapter two) demonstrated that whilst there is a number of promising microarray studies on HNSCC and OSCC in the literature, many of these lack the acceptable characteristics to be clinically useful in practice.

There are many limiting factors to perform meta-analysis of the various profiles, these include; differences in numbers of laboratory samples and processing, microarray platform (e.g., Affymetrix, Illumina, Agilent chips), gene chip model (e.g., Affymetrix HU-U133A, HU-U133 Plus2, HU-U95A) and diverse bioinformatics approaches used for data analysis (Sims et al., 2008, Kitchen et al., 2010). The limited sample size of the individual studies along with different bioinformatics approaches used for post-laboratory analysis are considered the most influencing of the above mentioned factors resulting in only a small overlap between the various gene signatures of OSCC. Large studies and frequently repetitive experimentation for an individual research study is undesirable due to the high costs. The problems of low accuracy and efficiency in a single analysis can be, to some extent, overcome by merging different single datasets to generate a larger sample size study accompanied by applying the same subsequent bioinformatics analysis approaches (Ein-Dor et al., 2006, Tseng et al., 2012).

With the availability of considerable publicly accessible microarray datasets, it is of substantial concern to understand how to merge microarray data across various individual studies of OSCC to enlarge sample size, which may improve the establishment of more reliable and solid gene signature of OSCC.

Integration of microarray datasets is advantageous for researchers to increase statistical power in detecting biological phenomena from studies where logistical considerations restrict sample size (Turnbull et al., 2012). Generally, it is inappropriate to combine datasets without adjusting for batch effects (Johnson et al., 2007a, Lazar et al., 2012).

Batch effects can be caused by laboratory conditions, reagents, and personnel diversity. This becomes a major problem when batch effects correlated with an outcome of biological interest and result in incorrect conclusions (Leek et al., 2010).

In this study, the ComBat batch correction method was utilised to minimise batch effects across studies.

3.2 Aims and objectives of the study

1-Using multiple bioinformatics approaches to integrate and process large and widely variable samples in an attempt to increase statistical power in detecting and identifying specific biomarkers and prediction models for OSCC.

2-To investigate the harvested transcriptomic profile using multiple biological annotation tools in an attempt to identify potential mechanisms underlying the biological changes during tumorigenic transformation of OSCC. Furthermore, to detect biological pathways necessarily altered in tumourigenesis which can potentially illuminate novel therapeutic targets.

3.3 Materials and Methods

Four public microarray datasets of OSCC were integrated to determine the common set of genes, biomarkers, and the relative regulatory pathways possibly accountable for tumour transformation and growth in OSCC.

Raw Affymetrix .CEL files from published datasets were downloaded from public datasets repositories (explained below in details), summarised with Ensembl alternative CDF (Dai et al., 2005) and normalised with RMA (Irizarry et al., 2003), before integrating using ComBat (Johnson et al., 2007b) to remove dataset-specific bias as previously described (Sims et al., 2008).
The pre-processing steps were performed as a submitted job on the Edinburgh Compute and Data Facility high performance compute cluster 'Eddie', due to lack a of memory on a standalone PC.

3.3.1 Selection of datasets

A search for OSCC and HNSCC microarray datasets among NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), ArrayExpress Archive (http://www.ebi.ac.uk/arrayexpress/), and caARRAY array data management system (https://array.nci.nih.gov/caarray/home.action) was carried out. Only microarray gene expression datasets from published studies of oral and head and neck cancer performed on frozen material extracted before chemotherapy, radiotherapy or adjuvant treatment were selected. Studies using Laser Capture microdissection technologies were excluded as it reflects non-comprehensive gene expression profile of just tumour cells, neglecting the tumour-stromal interaction (the signature of each tumour sample is a reflection of the combined tumour, stromal and inflammatory components of the original heterogeneous tumour sample (Ross et al., 2000)).
The study comprises pure samples of oral cavity in which other origins of tissue samples (e.g., nasopharyngeal, maxillary sinus, posterior wall of oropharynx, and larynx) were excluded from the analysis as these samples already originated from different origins which can reflect different molecular gene expression profiles regardless of the disease status, whereas some of them associated with HPV infection (oropharyngeal tumour). HPV positive samples were also excluded. Four datasets (Table 3.1) were selected that used the same platform (Affymetrix U133 Plus2, www.affymetrix.com) for the gene expression analysis of tumour versus normal tissue samples (Ye et al., 2008, Chen et al., 2008a, Pyeon et al., 2007, Reis et al., 2011). Harvested samples consist of 270 tumour and 95 normal samples comprising a total of 365 samples.

Table 3.1: Characteristics of the selected datasets samples*.

Authors	Country of origin	No. of OSCC samples	No. Of Normal samples	Sex ratio (male %)	Tumours stage	Gene chips platform	Array series No.
Ye et al, 2008	US, Los Angeles, Chicago	26	12	76.3	I-IV	HG- U133 Plus2.0	GSE9844
Chen et al, 2008	US, Washington	167	45	70	I-IV	HG- U133 Plus2.0	GSE26549, GSE30784, GSE31056
Pyeon et al, 2007	US, Iowa	22	14	52.7	I-IV	HG- U133 Plus2.0	GSE6791
Reis et al, 2011	Canada, Toronto	55	24	77.2	I-IV	HG- U133 Plus2.0	E-GEOD31056

* All samples are harvested from whole tumour specimens.

3.3.2 Data processing and analysis

Data analysis was performed using Bioconductor packages (Gentleman et al., 2004), http://www.bioconductor.org/) and the statistical programming language, R. The *simpleaffy* package was used to assess average background, Scale factors, Number of genes called present (% Present), 3' to 5' ratios. Normalization was performed using *RMA* algorithm to extract gene expression values (Irizarry et al., 2003). Subsequently, *Nonspecific Filtering* was performed using the function *rowSds (*Row variance and standard deviation of a numeric array*)/ genefilter* in R statistical software package available on Bioconductor.

3.3.2.1 Number of genes called present (% Present)

Present/Marginal/Absent calls are generated by looking at the difference between Perfect Match (PM) and Mismatch (MM) values for each probe pair in a probeset. As the Mismatch (MM) probe intensity increases and becomes equal to or greater than the Perfect Match (PM) intensity, the Discrimination score decreases. More specifically, as the intensity of the Mismatch (MM) increases, the ability to discriminate between the PM and MM decreases.

Probesets are flagged "Marginal" or "Absent" when the PM values for that probeset are not considered to be significantly above the MM probes. The '% Present' call simply represents the percentage of probesets called Present on an array.

3.3.3 Integration of datasets and batch correction

Integration of all normalized single datasets was carried out using *ComBat* package in R software (Johnson et al., 2007a), while parametric and non-parametric empirical Bayes frameworks were implicated for adjusting data of batch effects.

3.3.4 Differential gene expression analyses

Tumour samples were compared to normal samples to examine differential gene expression and to extract meaningful biological results from this comparison. In this comparison, data analysis takes two main forms. The first is 'unsupervised' analysis where data is sorted and gene expression differences are derived based on inherent differences within the dataset. Unsupervised clustering was achieved and heatmaps generated using Pearson correlation and hierarchical clustering in R. Supervised analysis was carried out using the Significance Analysis for Microarrays (SAM) (Tusher et al., 2001) in *siggenes* package available from Bioconductor. SAM is one of the more potent and effective ways of identifying particular gene products involved in differentiating transcriptional profiles between two groups (Tusher et al., 2001). SAM gives good all-round discriminatory ability based on average signal intensity differences within the supervised groups and between the same groups at pre-determined fold-change levels and gives an estimate of the false discovery rate for a significantly changed gene (Jeffery et al., 2006). Hierarchical clustering heatmaps were generated using *gplots* package (Gentleman et al., 2004). Fold change values for relative "upregulated/overexpressed" and "downregulated / under-expressed" genes were obtained using Wilcoxon Rank Statistics function in siggenes R software.

Lists of genes were further refined by trimming out low magnitude fold changes and / or reducing the false discovery rate (FDR) value of the significant genes by adjustment of delta value. The value of delta is chosen that provides the best balance between the number of identified genes and the estimated FDR, i.e. that allows to simultaneously attain the two competing goals "As many genes as possible" and "As low FDR as possible".

3.3.5 Biological annotations of significant genes

Ingenuity Pathway Analysis 9.0, IPA, was used to highlight the most dysregulated molecular and cellular bio-functions, canonical pathways, diseases and disorders, transcriptional regulators, and most significant pre-generated networks associated with the uploaded group of genes (8995 meta-genes at FDR ≤ 0.05 with no fold change cut -off) including those related to OSCC development. IPA knowledge base recruits the distinguished genes that regulated OSCC and HNSCC to generate and highlight the above functions.

As mentioned in the previous chapter, IPA cannot map pseudogenes, hypothetical genes, ESTs and ambiguous (mapping to multiple) genes as well as duplicates of multiple IDs mapped in one molecule or multiple molecules snapped in one ID. However, the unmapped genes were converted using David conversion tool (Huang et al., 2009a, Huang et al., 2009b) to include them in the subsequent analysis, some were substantial key players in the master networks of the analysis and contributing to the tumorigenic process.

3.4 Results

3.4.1 Filtering and processing of the datasets

After normalization, the huge number of probesets was reduced by *Nonspecific Filtering* using the function *rowSds/ genefilter* in R statistical software package available on Bioconductor (www.bioconductor.org/packages/release/bioc/vignettes/genefilter/inst/doc/howtogenefilter.pdf). By applying this filtering method (Figure 3.1, Figure 3.2, and Figure 3.3), not only a reduction in the non-useful and non-diagnostic probe sets found, but also an improvement in the sample discrimination ability of the subsequent analysis. Whilst all probe sets whose standard deviation below 0.242 were discarded as they have very low variability and can safely assumed that it will not be able to infer differential expression for their target genes.

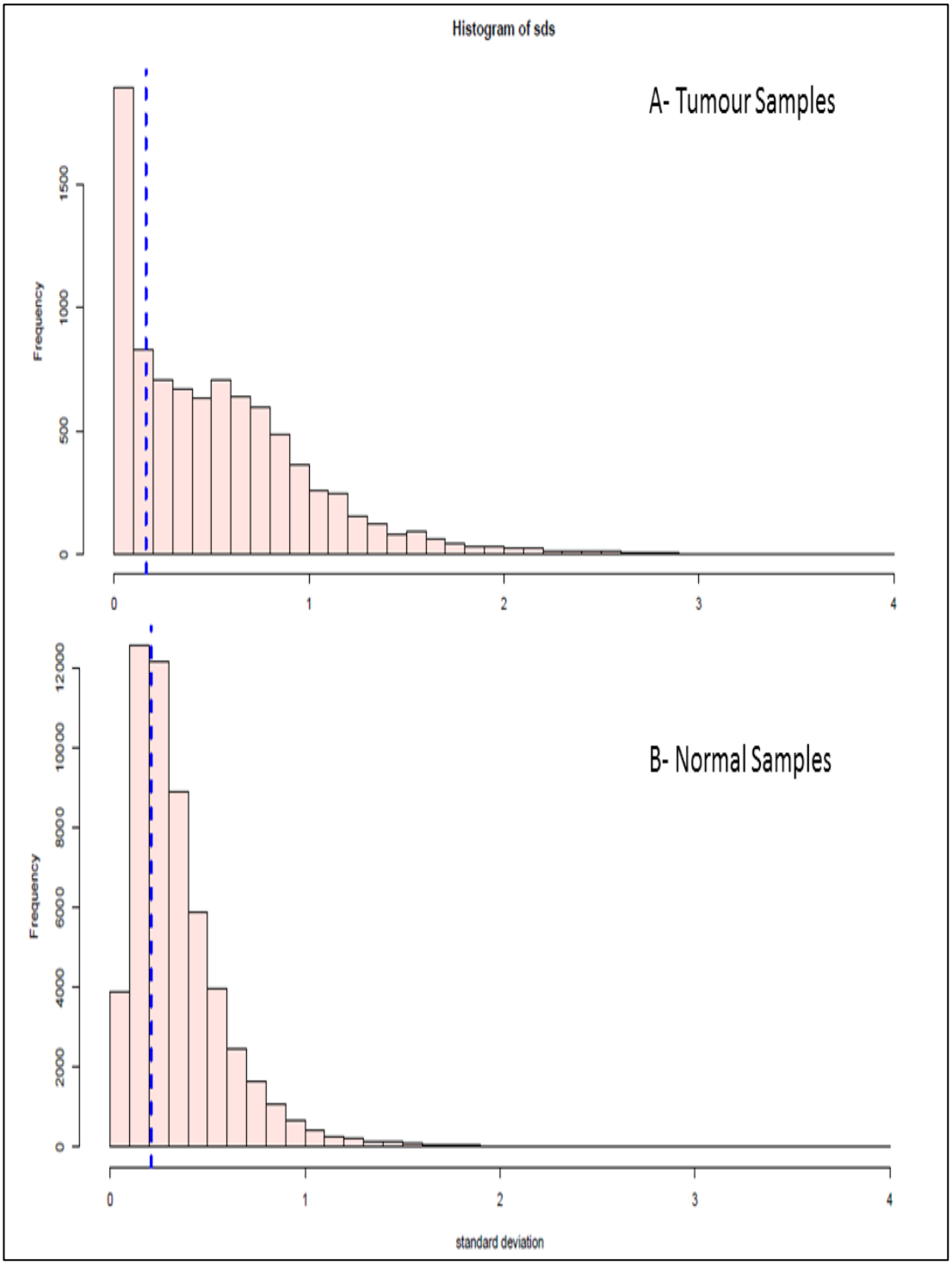


Figure 3.1: Filtration of probe sets of both tumour and normal samples

The filtration carried out using non-specific filtering method from *genefilter* in R statistical software package available on Bioconductor. The blue dotted vertical lines represent the cut-off value of rowSds =0.242.

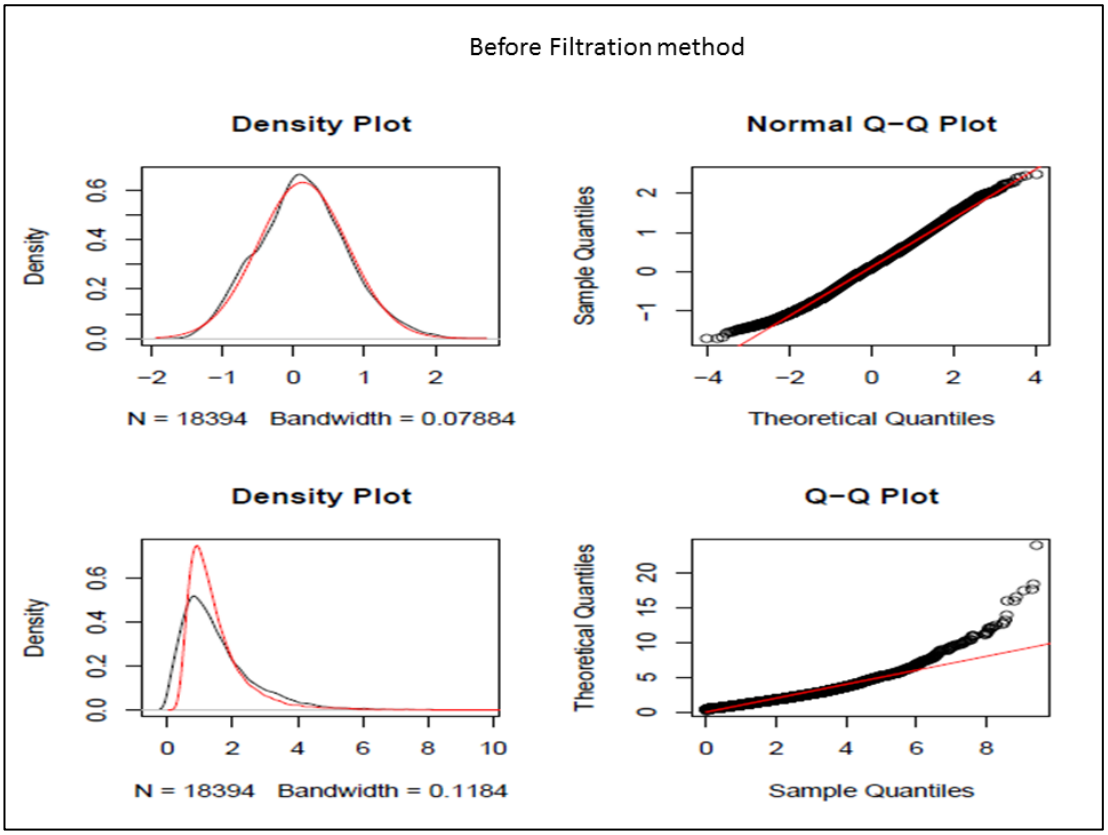


Figure 3.2: Density and Q-Q plot of normalized genes before filtration method.

The density plots (density vs. log2 intensity (1/2(log2(cy5*cy3)))) (produced using the R statistical language with the LIMMA package), illustrating the distribution of single-channel intensities. The plots allow visually comparing different normalization methods. The Q-Q (Quantile-Quantile) plots (produced using the R statistical language with the Bioconductor package) displaying the magnitude of differentially gene expression within the sample tested based on the Students t-test relative to a theoretical t-statistic. The Q-Q plot indicate the degree a sample diverges from a normal distribution where points which deviate markedly from a linear relationship to a theoretical t-statistic could be considered suspect genes exhibiting differentially expression.

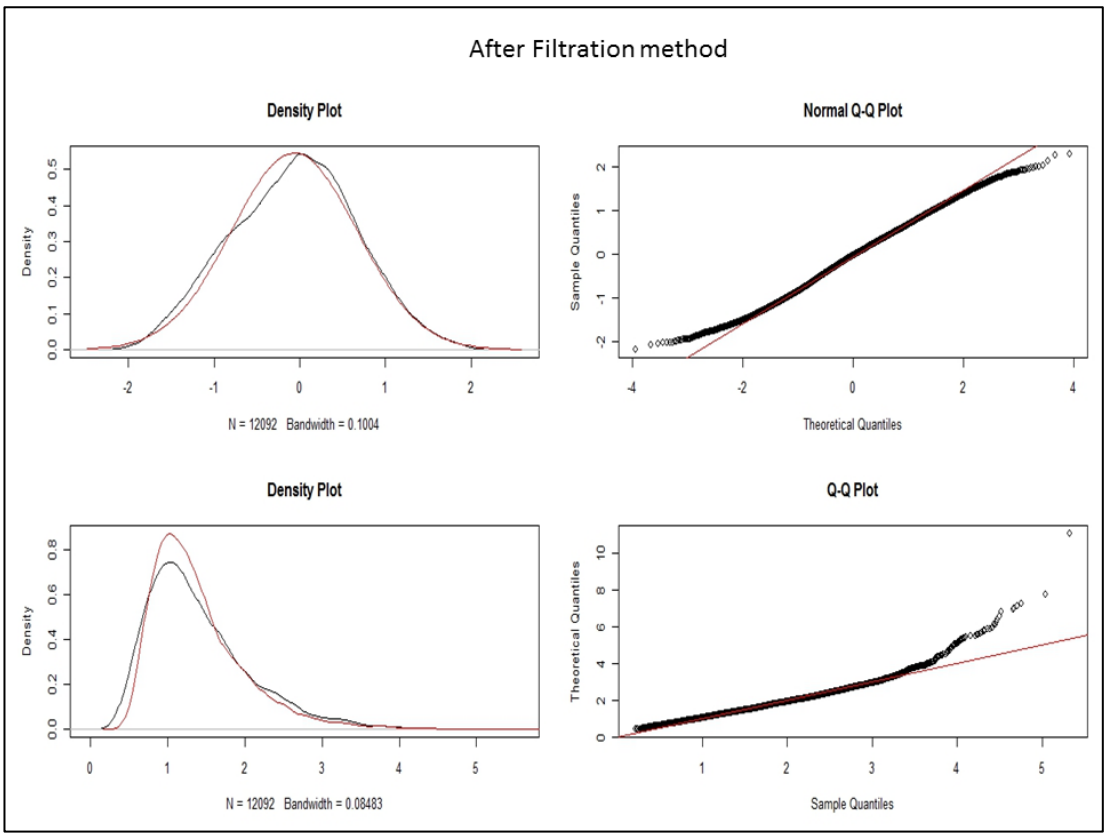


Figure 3.3: Density and Q-Q plot of normalized genes after filtration method.

The density plots (density vs. log2 intensity (1/2(log2(cy5*cy3)))) (produced using the R statistical language with the LIMMA package), illustrating the distribution of single-channel intensities. The plots allow visually comparing different normalization methods. The Q-Q (Quantile-Quantile) plots (produced using the R statistical language with the Bioconductor package) displaying the magnitude of differentially gene expression within the sample tested based on the Students t-test relative to a theoretical t-statistic. The Q-Q plot indicate the degree a sample diverges from a normal distribution where points which deviate markedly from a linear relationship to a theoretical t-statistic could be considered suspect genes exhibiting differentially expression.

3.4.2 Integration of datasets and correction of batch effect

The four datasets were merged using ComBat data integration method in R software (Figure 3.2 and Figure 3.3). Ensembl IDs were used to form the comparison of the common genes.

3.4.3 Verification of data integration

To assess the removal of microarray bias effect across datasets, Pearson correlation and hierarchical clustering were applied to the datasets after the application of data integration method (ComBat). The results of these tests are shown in Figure 3.4 and Figure 3.5. As can be seen from those figures, ComBat was successful in removing dataset-specific biases (Figure 3.5) where samples clustered independent to the dataset after correction compared to those merged randomly before the use of ComBat (Figure 3.4).

3.4.4 Gene expression profile of OSCC versus normal oral mucosa

Following data normalization, probeset filtering, and batch correction for dataset integration, unsupervised analysis of tumour and normal samples was performed using Pearson correlation comparison in which normal samples clustered together away from tumour samples (Figure 3.5). The samples did not cluster by the datasets from which they were derived, suggesting that intra- and inter-platform batch effects have been minimised, whilst samples from the same individual dataset still cluster independently, signifying that the true biological differences (disease status) are maintained. This indicates ComBat correction significantly reduced the bias between samples from different datasets.

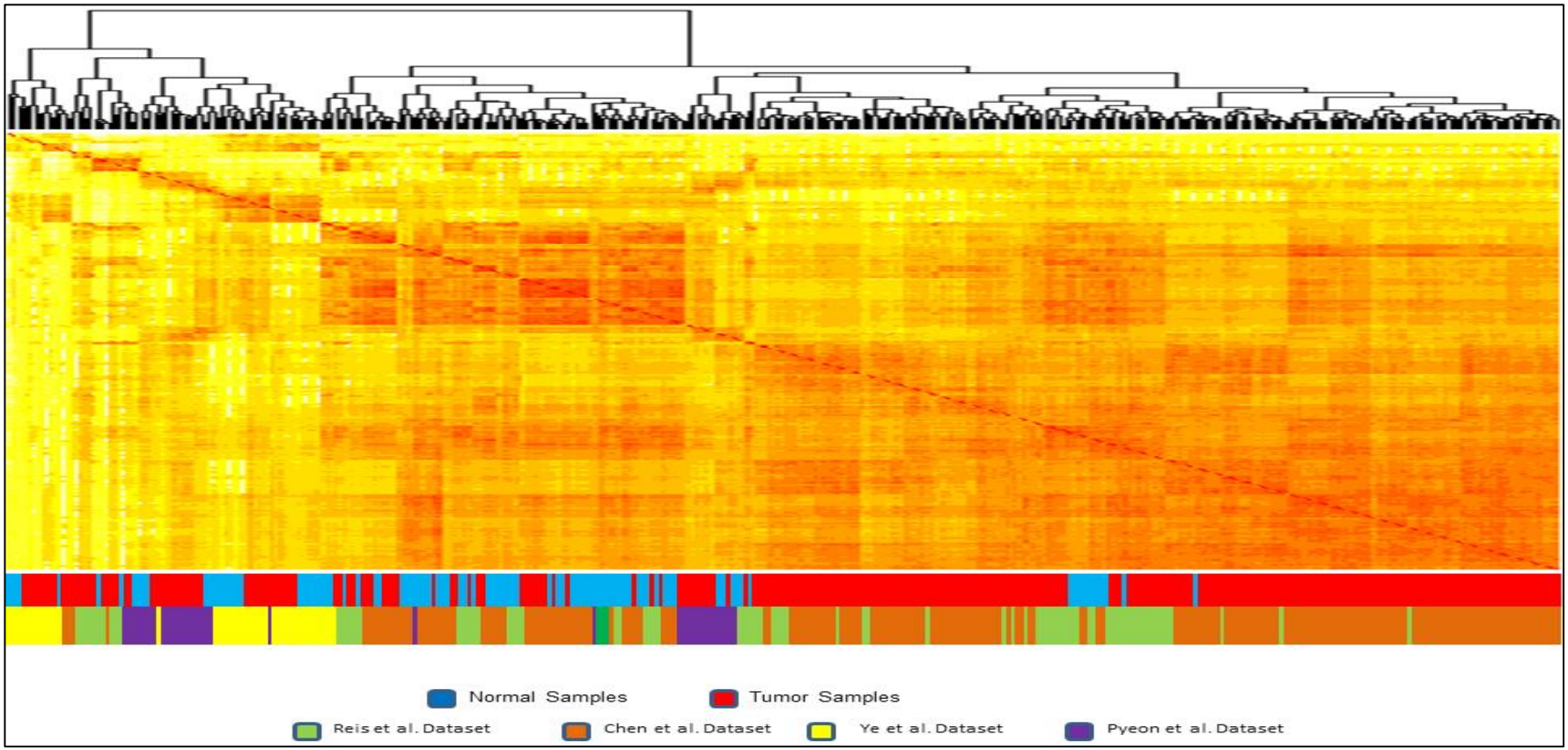


Figure 3.4: Unsupervised Pearson correlation hierarchical clustering of merged 18394 normalized genes before ComBat integration, discriminating normal oral mucosal samples (N=95) from OSCC samples (N=270).
Heatmap represents sample distribution according to the Pearson correlations (red= 1, yellow= lower than 1, white= lowest value than 1). Red and blue bars below the heatmap represent cancerous and normal samples, respectively. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

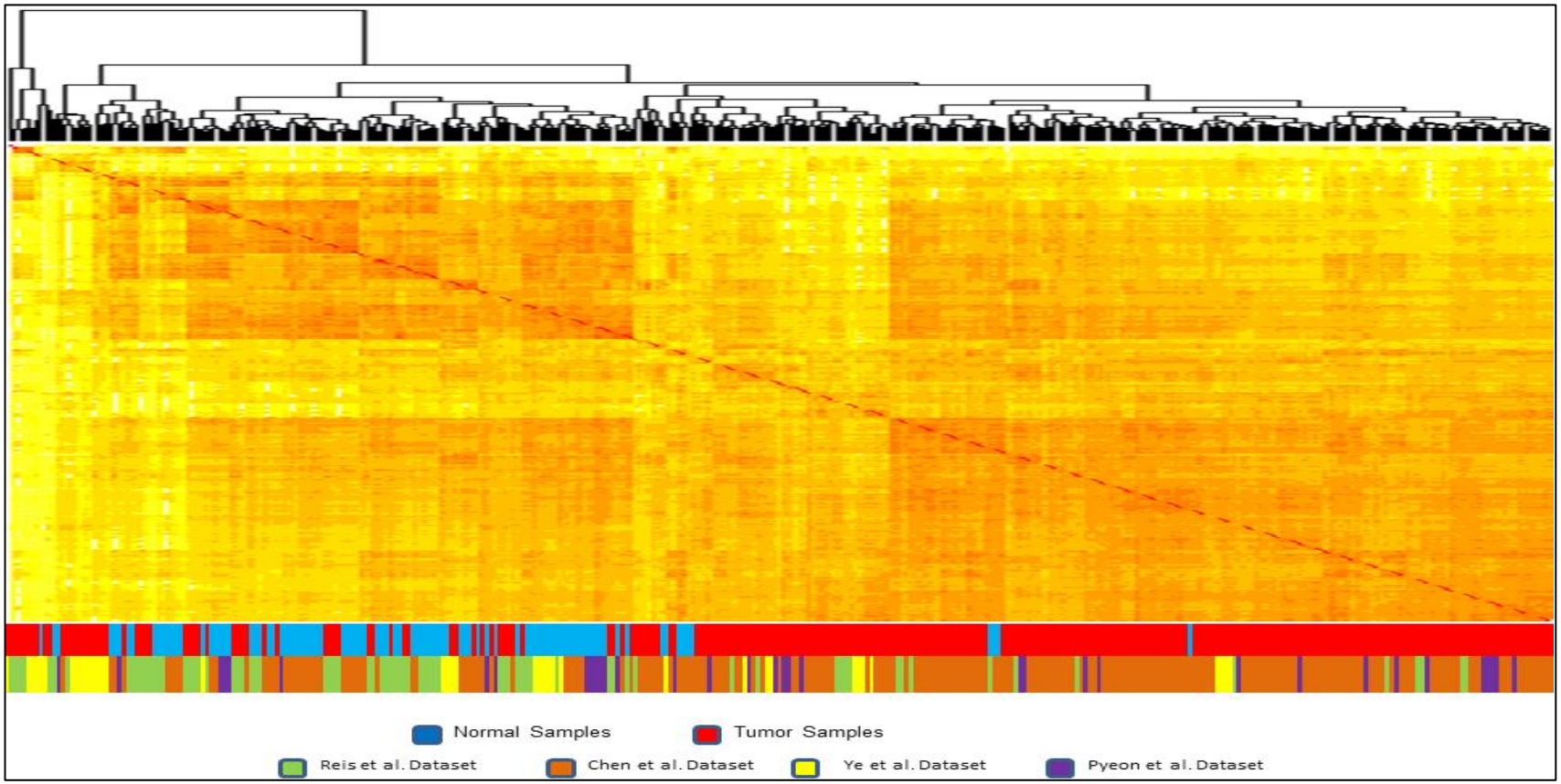


Figure 3.5: Unsupervised Pearson correlation hierarchical clustering of 12092 normalized genes after ComBat integration, discriminating normal oral mucosal samples (N=95) from OSCC samples (N=270).

Heatmap represents sample distribution according to the Pearson correlations (R-values 0.55-1; red= 1, yellow= lower than 1, white= lowest value than 1). Red and blue bars below the heatmap represent cancerous and normal samples, respectively. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

Initial supervised analysis of all tumour and normal samples was carried out using SAM method with a false discovery rate (FDR) = 0 and fold change above or below 1.5 (Figure 3.6). A group of Thirty four samples belonging to one study (Reis et al.), which comprises a part of one of the four datasets, was collected from the normal margins during removal of the tumour lesions, but as described in the original study, they had a different molecular profile from other pure normal tissue predicting tumour recurrence of individuals after surgical removal of original primary tumour. These samples were included in the study as tumour samples to see if they behave independently among the normal and tumour samples from the four datasets. However, all of them cluster with the normal rather than tumour samples, indicating that their molecular profiles are more similar to normal tissue, despite having some predicting sign of tumour recurrence according to their original study (Figure 3.6). These samples were removed from the subsequent analysis to end up with 329 samples instead of 365 totals (236 tumour and 93 normal samples).

The 329 samples were subjected to the SAM method in two groups according to sample type; OSCC and normal mucosa identifying 334 differentially expressed genes with a fold change cut-off equal to ±1.5 and FDR level = 0, where 121 genes were higher and 213 genes lower in OSCC compared to normal mucosal samples (Table 3.2, Table 3.3, and Supplementary file 5).

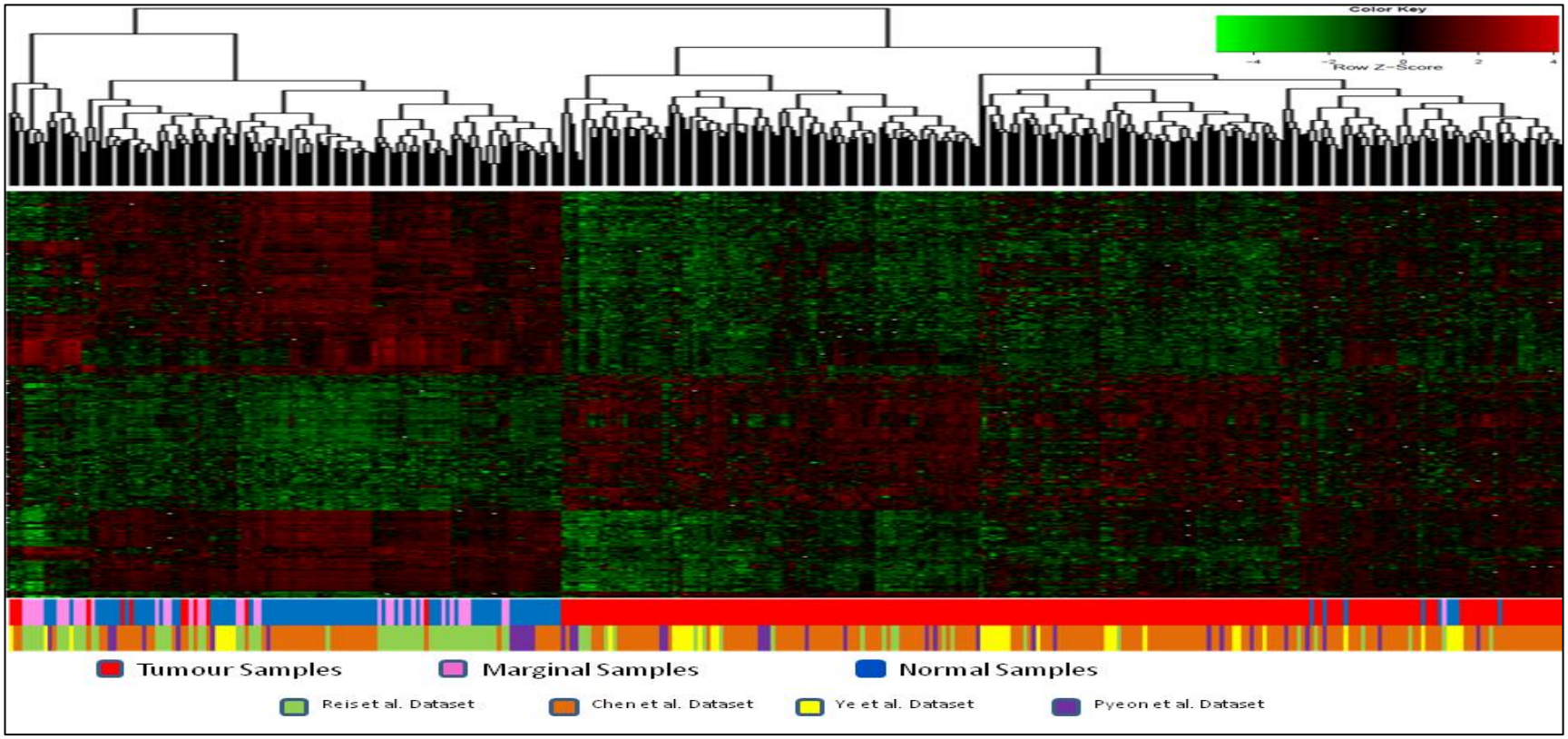


Figure 3.6: Hierarchical clustering of differentially expressed genes separating OSCC tumour (n=236) and both pure oral normal (N=95) and marginal normal oral samples (n=34) using SAM supervised analysis at FDR=0.

Heatmap represents mean-centred expression of 334 differentially expressed genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. Gene expression values were either above or below fold change magnitudes equal ± 1.5. Red, pink, and blue bars below the heatmap represent cancerous, marginal, and normal samples, respectively. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

Table 3.2: Upregulated differentially expressed meta-genes in OSCC compared to normal oral tissue samples.*

Gene*	Fold Change**	Gene*	Fold Change**	Gene*	Fold Change**
MMP1	6.16	CDH3	2.08	KRT75	1.67
MMP13	4.42	TREM2	2.07	IFI44	1.67
MMP10	4.31	CDSN	2.03	ABCA12	1.67
MMP3	4.24	RSAD2	2.02	IFI44L	1.65
MMP12	3.95	APOL1	2.02	COL4A6	1.65
COL11A1	3.43	COL5A2	2.01	LAMA1	1.64
SERPINE1	3.10	PLAU	2.01	CYP27C1	1.64
PTHLH	3.09	CXCL3	2.00	SH2D5	1.63
CXCL13	3.07	ADAM12	2.00	TGFBI	1.63
APOC1	2.92	PRR9	1.94	C7ORF10	1.63
ISG15	2.80	TDO2	1.94	EN1	1.61
IL8	2.79	CXCL9	1.94	APOBEC3A	1.59
ZIC2	2.74	ODZ2	1.93	IFIT1	1.59
CCL20	2.72	CSPG4	1.91	LEPREL1	1.58
MUCL1	2.63	CXCL6	1.89	HOXC10	1.58
CTHRC1	2.60	COL4A1	1.88	IFI27	1.57
GPR39	2.58	PMEPA1	1.87	CA2	1.57
LAMC2	2.56	SPP1	1.87	HOXB7	1.57
IFI6	2.55	AIM2	1.85	HSD17B6	1.57
POSTN	2.52	COL5A1	1.84	SLC16A1	1.56
FAP	2.49	TNC	1.83	PPP4R4	1.55
CXCL10	2.46	PPAPDC1A	1.83	WISP3	1.54
COL10A1	2.44	SERPINH1	1.83	LPCAT1	1.54
KRT17P1	2.43	HOXA10	1.82	MFAP2	1.54
CXCL11	2.39	CST1	1.79	GBP1	1.53
SCG5	2.38	ESM1	1.79	CCL11	1.52
TNFRSF12A	2.36	IDO1	1.79	COL4A2	1.52
KRT16	2.35	GBP5	1.77	IBSP	1.52
BST2	2.33	MYO1B	1.77	KANK4	1.52
GALNT6	2.31	CMPK2	1.77	CNTNAP2	1.51
CXCL1	2.27	KHDC1	1.76	C6ORF150	1.51
MAGEA4	2.27	IL1F9	1.76	ARSJ	1.51
GPR158	2.26	HOXD10	1.76	BCL2A1	1.50
SLCO1B3	2.22	SOST	1.72	TNFAIP6	1.50
SPRR2C	2.22	TREM1	1.71	TMEM45A	1.50
CXCL5	2.17	HAS3	1.71		
HOXD11	2.17	DFNA5	1.71		
MMP9	2.14	BNC1	1.70		
LOC100129940	2.11	C20ORF103	1.70		
COL1A1	2.10	PDPN	1.70		
LIPG	2.08	EPSTI1	1.68		
WDR66	2.08	CYP27B1	1.67		
GREM1	2.08	COL1A2	1.67		

* FDR value = 0. The complete lists of upregulated gene names and ENSEMBL IDs were presented in Supplementary file 5. ** Log2 folds change values.

Table 3.3: Downregulated differentially expressed meta-genes in OSCC compared to normal oral tissue samples.*

Gene*	Fold Change**	Gene*	Fold Change**	Gene*	Fold Change**
CRISP3	-5.69	ADH7	-2.56	UPK1A	-2.14
CRNN	-4.95	SFTA2	-2.55	ASPA	-2.14
TMPRSS11B	-4.95	TNNC1	-2.54	HPGD	-2.14
MAL	-4.28	GBP6	-2.53	LMOD2	-2.13
KRT4	-4.01	MAMDC2	-2.52	FUT6	-2.12
TGM3	-3.86	CHRDL1	-2.49	MUC15	-2.12
KRT78	-3.77	ANGPTL1	-2.47	DAPL1	-2.11
PRSS27	-3.68	ATP6V0A4	-2.47	CILP	-2.11
ENDOU	-3.62	SLITRK5	-2.46	PYGM	-2.11
CLCA4	-3.51	SLC6A4	-2.42	LOC643008	-2.10
FAM3B	-3.46	STATH	-2.41	ATP13A4	-2.09
SLURP1	-3.38	HLF	-2.38	ATP1A2	-2.08
SPINK7	-3.31	CLEC3B	-2.36	ECM1	-2.07
FAM3D	-3.25	SPRR3	-2.35	CA3	-2.06
KRT13	-3.23	TF	-2.34	PTN	-2.06
ADH1A	-3.20	C9ORF169	-2.34	MYOZ1	-2.06
CAPN14	-3.14	ENO3	-2.32	MYOT	-2.05
RHCG	-3.11	C2ORF40	-2.32	TOX3	-2.05
ACTA1	-3.05	RBM20	-2.32	PLP1	-2.04
SLC27A6	-3.01	SCIN	-2.30	CGNL1	-2.03
ABCA8	-3.01	CYP3A5	-2.29	CLIC3	-2.03
SCEL	-3.00	EPHX2	-2.26	UGT1A7	-2.01
CKM	-2.94	SH3BGRL2	-2.24	TCEAL2	-2.00
MYH2	-2.87	BBS5	-2.24	SLN	-2.00
PPP1R3C	-2.86	CYP4B1	-2.23	LAMB4	-1.99
PLAC9	-2.79	MFAP4	-2.22	GPD1L	-1.98
SNX31	-2.78	A2ML1	-2.21	MAOB	-1.98
MYL1	-2.76	MYRIP	-2.21	EMP1	-1.98
ALDH3A1	-2.75	CASQ1	-2.20	CNFN	-1.97
MYBPC1	-2.71	MB	-2.20	CFD	-1.96
TNNI2	-2.70	ABI3BP	-2.19	NCCRP1	-1.95
SERPINB11	-2.70	FCER1A	-2.19	PLA2G2A	-1.95
GYS2	-2.70	PADI1	-2.17	ST6GALNAC1	-1.95
TMPRSS11E	-2.69	GREM2	-2.17	TMEM100	-1.93
SPINK5	-2.68	ALOX12	-2.16	C2ORF54	-1.93
FLG	-2.64	HOPX	-2.16	OSR1	-1.92
AADAC	-2.62	C7	-2.16	CXCL17	-1.92
SCARA5	-2.58	TYRP1	-2.15	ADIPOQ	-1.92
TNNC2	-2.58	IGSF10	-2.15	IL1F6	-1.92
TSPAN8	-2.57	CEACAM5	-2.14	TMPRSS2	-1.91
GSTA1	-1.90	FOXA1	-1.67	SUSD4	-1.53
NMU	-1.90	TNNT3	-1.67	KLHDC8A	-1.52
MEOX2	-1.88	RAET1E	-1.66	DPT	-1.52
TGM1	-1.87	C10ORF116	-1.66	ETNK2	-1.51
COX7A1	-1.87	C1QTNF7	-1.65	CRISP2	-1.51
FXYD1	-1.87	SAMD5	-1.65	EXPH5	-1.51
MYOC	-1.86	PAX9	-1.65	ZBTB7C	-1.50

Continue

Gene*	Fold Change**	Gene*	Fold Change**	Gene*	Fold Change**
TPRG1	-1.85	RBP7	-1.64	KLB	-1.50
CYP4F12	-1.85	ALDH1A1	-1.64	GPR133	-1.50
CRYM	-1.84	PPL	-1.64	EYA2	-1.50
COBL	-1.83	KLK13	-1.64	IL33	-1.50
GALNT12	-1.82	PRSS3	-1.64	CWH43	-1.50
ANXA9	-1.82	CTTNBP2	-1.64	C4ORF31	-1.50
NR3C2	-1.82	EVPL	-1.64		
CYP2C18	-1.82	TM7SF2	-1.63		
CHST9	-1.81	C9ORF152	-1.63		
SPNS2	-1.81	GDPD3	-1.62		
DARC	-1.81	FMO2	-1.62		
MMRN1	-1.81	KLK12	-1.62		
ATP2A1	-1.81	FHL1	-1.62		
CLDN8	-1.80	GCNT3	-1.61		
NPY1R	-1.80	IL1RN	-1.61		
PITX1	-1.78	KRT15	-1.61		
IL12A	-1.77	GCHFR	-1.59		
NRAP	-1.77	TCP11L2	-1.59		
CEACAM7	-1.76	TGFBR3	-1.59		
DLK1	-1.76	OGN	-1.59		
CLDN17	-1.75	SASH1	-1.59		
SCNN1B	-1.75	DEPDC6	-1.57		
CKMT2	-1.72	TTC9	-1.57		
GGTA1	-1.72	BLNK	-1.57		
APOD	-1.71	VIT	-1.56		
GABRP	-1.71	PLN	-1.56		
C15ORF62	-1.70	C7ORF46	-1.56		
KLK11	-1.69	MGLL	-1.55		
SGCG	-1.69	MYL3	-1.55		
FAM149A	-1.68	BNIPL	-1.55		
LYVE1	-1.68	PAX1	-1.54		
HS3ST6	-1.68	BTC	-1.54		
CLDN10	-1.67	PPARGC1A	-1.53		

* FDR value = 0. The complete lists of downregulated gene names, ENSEMBL IDs were presented in Supplementary file 5. ** Log2 folds change values.

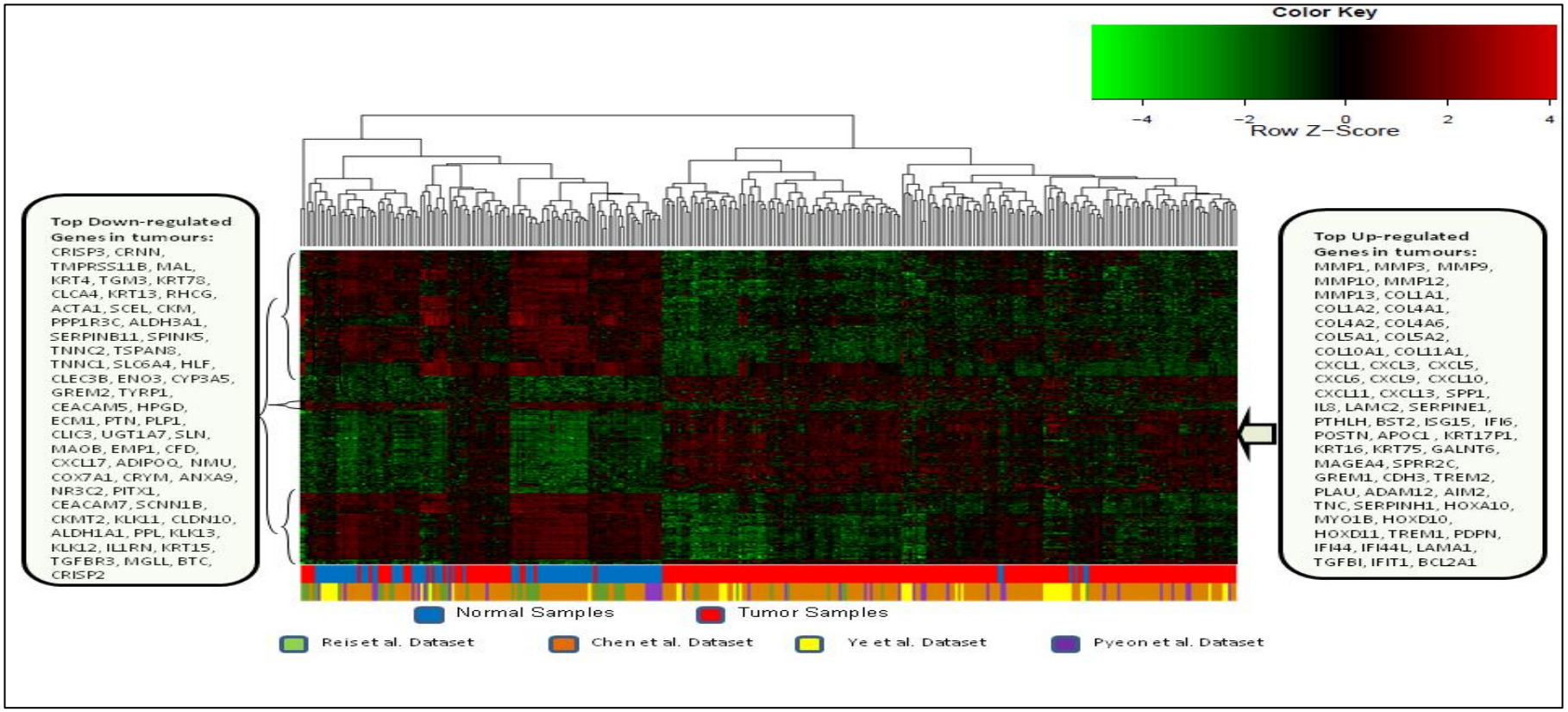


Figure 3.7: Hierarchical clustering of differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis at FDR=0.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of 334 differentially expressed genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. Gene expression values were either above or below fold change magnitudes equal ± 1.5. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

Figure 3.7 displays better discrimination between tumour and normal samples than before removal of the marginal normal samples.

At FDR level ≤ 0.05, a set of 8995 differentially expressed genes (74% of the total normalized genes) was obtained, which subsequently uploaded to IPA system for comprehensive biological interpretation of the resultant set of meta-genes differentiating tumour from normal samples of the four datasets.

3.4.5 Prediction models

Additional supervised analyses were performed in an attempt to improve the discrimination power separating tumour from normal samples with less number of genes. Delta value of SAM was further adjusted to accept only top 100, 50, 38, 8, and 3 differentially expressed genes. Despite the small number of genes, the prediction models still have the capability of discriminating OSCCs from normal oral mucosa. Figure 3.8, Figure 3.9, and Figure 3.10 display the model of top 38, 8, and 3 genes respectively. The top 8 significantly differentially expressed genes had better discriminating power between OSCC tumours and normal oral tissue than the 334 genes (Figure 3.9) with fewer normal samples clustered in the tumour area of the hierarchy positioned very close to the area of normal samples. From the set of 8 and 3 genes, LAMC2 and MMP1 reported as biomarkers overexpressed in the diagnosis of oral, head and neck cancer respectively Figure 3.16, while repression of MAL employed as diagnostic biomarker of gastric cancer. The other 5 genes (PTHLH, GPD1L, TPBG, SLC27A6, and TMPRSS11B) are suggested as candidates for diagnostic, prognostic, and disease prognosis biomarkers of OSCC. Uploading the 8 genes of the model to IPA, revealed Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization, Cellular Compromise, Cell Death and Survival, and Cellular Development functions where most of the genes involved in. Oncostatin M Signalling, Inhibition of Matrix Metalloproteases, HIF1α Signalling, Role of Tissue Factor in Cancer, and Glucocorticoid Receptor Signalling exhibited as the top, cancer related, canonical pathways where the 8 genes participated. KEGG pathway analysis (Kanehisa and Goto, 2000) of the 8 genes was performed with DAVID, online pathway annotation software based on scoring and visualization of the dysregulated

pathways collected in KEGG database (http://david.abcc.ncifcrf.gov/home.jsp). KEGG analysis highlighted additional two activated pathways; the extracellular matrix- receptor interaction (ECM Receptor Interaction) and Peroxisome proliferator-activated receptors (PPAR) signalling pathways corresponding to the model of 8 genes.

3.4.5.1 The discrimination power of the variant prediction models

The distribution of misclassified tumour and normal samples clustered in the hierarchy was used to examine the discrimination power of each prediction model. The model with 38 genes (Figure 3.8) showed 12% (11 normals) and 7.5% (18 tumours) misclassified compared to 6.6% (6 normals) and 17.6% (42 tumours) in the model of 334 genes at FDR=0 (Figure 3.7). This is considered an improvement in the discrimination power; especially as it is accompanied with a reduction in the number of discriminators (from 334 to 38 genes). Further improvement was observed with the model of 8 genes (Figure 3.9) showing only 7.6% (7 normals) and 10.9% (26 tumours) misclassified in the hierarchical cluster. However, the model consisting of 3 genes presented less improvement in the discrimination power relative to the one of 8 genes with 43.9% (40 normals) and 6.3% (15 tumours) misclassified in the hierarchy.

Table 3.4: The percentages of misclassified tumour and normal samples among the prediction models.*

Prediction Models	334 genes	38 genes	8 genes	3 genes
Misclassified tumour samples (%)	17.6	7.5	10.9	6.3
Misclassified normal samples (%)	6.6	12	7.6	43.9
Total misclassified tumour and normal Samples (%)	24.2	19.5	17.6	50.2

* The number of misclassified tumour and normal samples can be distinguished roughly from the red and blue bars presented in the hierarchical clusters of the models.

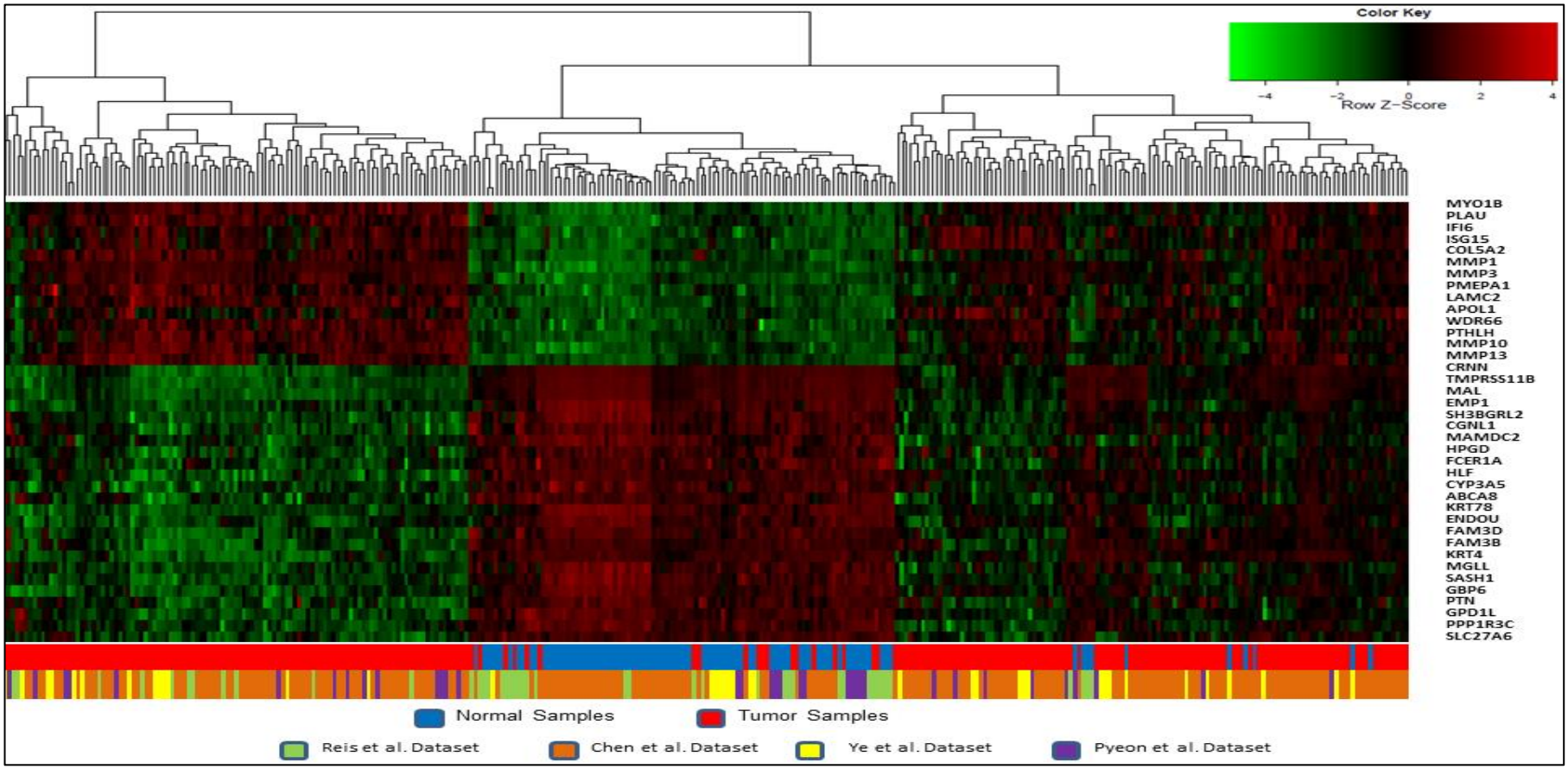


Figure 3.8: Hierarchical clustering of top 38 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

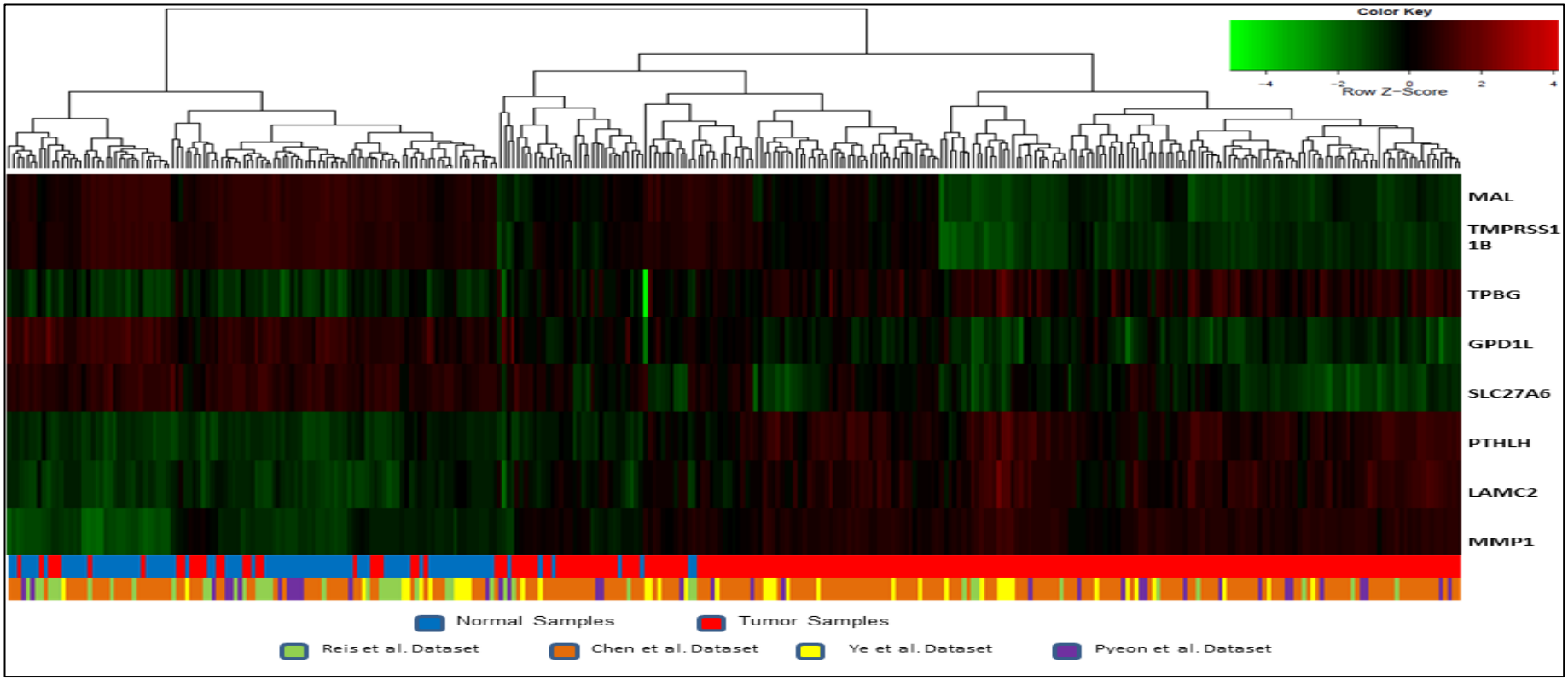


Figure 3.9: Hierarchical clustering of top 8 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

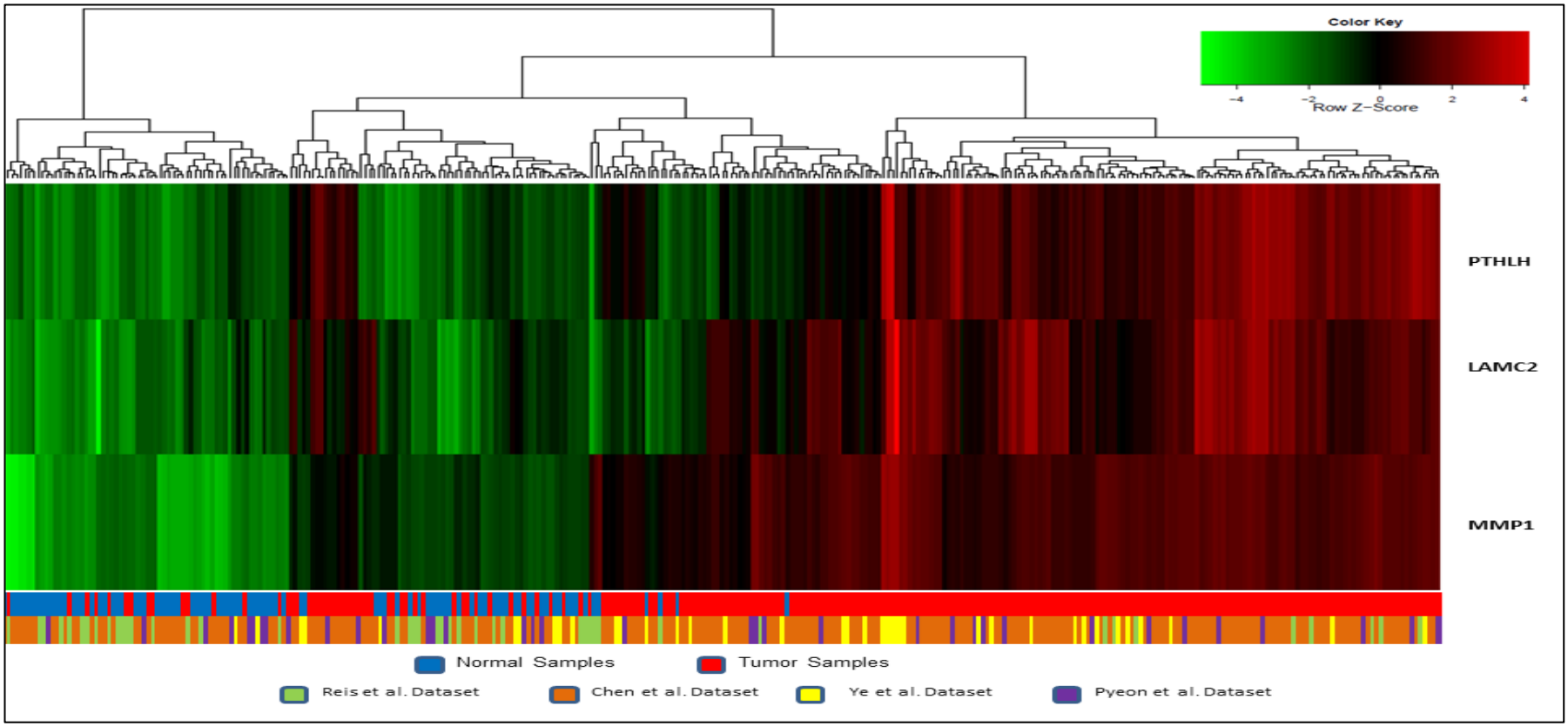


Figure 3.10: Hierarchical clustering of top 3 differentially expressed genes after removal of marginal samples separating tumour (n=238) and normal samples (n=91) using SAM supervised analysis.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. The green, brown, yellow, and purple bars on the bottom represent the samples of Reis et al., Chen et al., Ye et al., and Pyeon et al. datasets respectively.

3.4.5.2 Validation of the prediction models

The prediction models of top 8 and 3 genes were selected for validation with an external validation dataset from GEO (http://www.ncbi.nlm.nih.gov/geo/, GSE6631 containing 22 paired samples of HNSCC and normal tissue from the same donors using Affymetrix U95A platform) (Kuriakose et al., 2004). The prediction models were applied to find their altered expression in the normalised datasets. 6 out of 8 genes (MMP1, LAMC2, PTHLH, TPBG, GPD1L, and MAL) were correspondingly expressed in the dataset (Figure 3.11) while the probes of the remaining 2 genes (TMPRSS11B and SLC27A6) were absent in U95A platform. Student t-test was applied to identify the significant expression of the 6 genes discriminated tumour from normal samples of the independent dataset and whether the expression direction (up- or downregulation) of each gene in the models expressed similarly to the expression direction of the corresponding genes in the independent dataset. Boxplots showed the significant over-expression of MMP1, LAMC2, PTHLH, and TPBG6 genes along with under-expression of MAL and GPD1L genes in tumours relative to normal samples of the external independent dataset with all genes show significant differences between tumour and normal samples at p-value ≤ 0.0001 (LAMC2, TPBG, MAL, and GPD1L) and p-value ≤ 0.001 (MMP1 and PTHLH).

Both LAMC2 and PTHLH genes showed 2 probes in the external dataset where both probes confirmed the overexpression of genes in tumours relative to normal samples as seen in the boxplots. Further validation was carried out using two additional processed data sets of both UK and Sri Lanka discriminating tumour from normal samples of the oral cavity (samples of chapter 4) using HG Focus microarray platform. MMP1, PTHLH, LAMC2, TPBG, MAL, and SLC27A6 were correspondingly expressed in both UK and Sri Lankan gene sets differentiated tumour from normal samples while the probes of TMPRSS11B and GPD1L found absent in HG Focus platform.

Further information related to the 8 genes' model were gathered and summarized from different biological annotation tools including; Source (http://puma.princeton.edu/cgi-bin/source/sourceResult), Entrez Gene

(http://www.ncbi.nlm.nih.gov/gene/), and GeneCards tools (http://www.genecards.org/) in addition to IPA (Table 3.5).

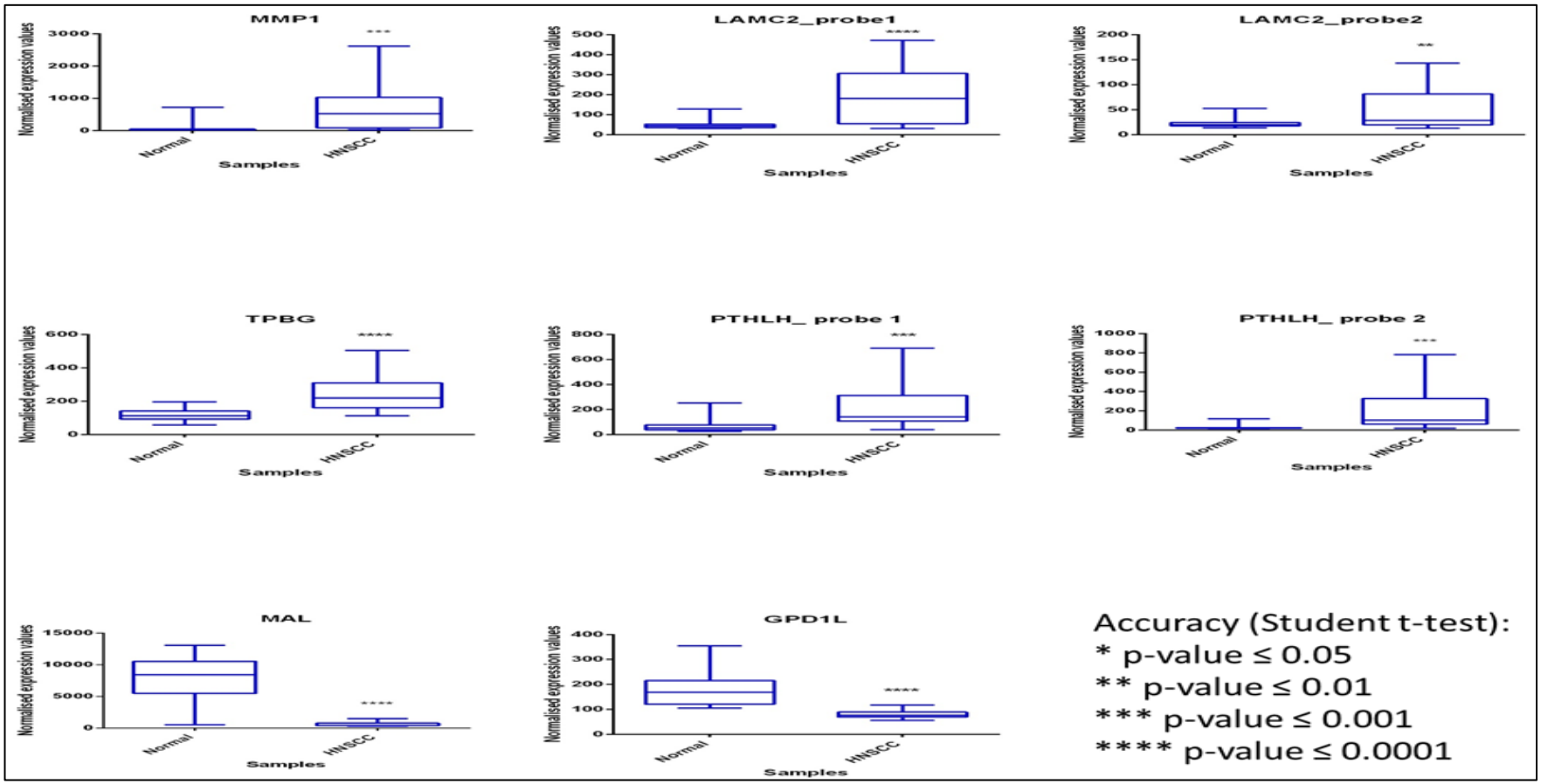


Figure 3.11: Box Plots displaying expression of 6 genes (MMP1, LAMC2, PTHLH, TPBG, GPD1L, and MAL) chosen for validation with an external dataset from GEO (GSE6631).

** P-value ≤ 0.01, *** P-value ≤ 0.001, **** P-value ≤ 0.0001 compared tumour to normal samples determined by using unpaired student t-test.

Table 3.5: Summary and biological annotations of the 8 genes' model.

Gene ID	Gene name	Biological processes	Molecular function	Chromosome/Cytoband	Entrez Gene Summary
MMP1	Matrix metallopeptidase 1 (interstitial collagenase)	Collagen catabolism	Zinc ion binding Calcium ion binding Interstitial collagenase activity	11q22.3	Proteins of the matrix metalloproteinase (MMP) family are involved in the breakdown of extracellular matrix in normal physiological processes, as well as in disease processes. This gene encodes a secreted enzyme which breaks down the interstitial collagens, types I, II, and III.
LAMC2	Laminin, gamma 2	Epidermis development Cell adhesion	Protein binding Structural molecule activity Heparin binding	1q25-q31	Laminins, a family of extracellular matrix glycoproteins, are the major noncollagenous constituent of basement membranes.
PTHLH	Parathyroid hormone-like hormone	Cell-cell signaling Positive regulation of cell proliferation Negative regulation of cell proliferation CAMP metabolism	Hormone activity	12p12.1-p11.2	This hormone regulates endochondral bone development and epithelial-mesenchymal interactions during the formation of the mammary glands and teeth. The receptor of this hormone, PTHR1, is responsible for most cases of humoral hypercalcemia of malignancy.
TPBG	Trophoblast glycoprotein	Cell motility	cell communication, playing a role in placentation or metastasis	6q14-q15	This gene encodes a leucine-rich transmembrane glycoprotein that may be involved in cell adhesion. In adults this protein is highly expressed in many tumor cells and is associated with poor clinical outcome in numerous cancers.
MAL	Mal, T-cell differentiation protein	Cell adhesion Signal transduction Cell differentiation Induction of apoptosis	Channel or pore class transporter activity Apoptotic protease activator activity	2cen-q13	The protein has been localized to the endoplasmic reticulum of T-cells and is a candidate linker protein in T-cell signal transduction. The protein plays a role in the formation, stabilization and maintenance of glycosphingolipid-enriched membrane microdomains.

Continue

Gene ID	Gene name	Biological processes	Molecular function	Chromosome/Cytoband	Entrez Gene Summary
					Continue
GPD1L	Glycerol-3-phosphate dehydrogenase 1-like	Carbohydrate metabolism Glycerol-3-phosphate catabolism	Glycerol-3-phosphate dehydrogenase (NAD+) activity, NAD binding	3p22.3	The protein encoded by this gene catalyzes the conversion of sn-glycerol 3-phosphate to glycerone phosphate. The encoded protein is found in the cytoplasm, associated with the plasma membrane, where it binds the sodium channel, voltage-gated, type V, alpha subunit (SCN5A). Defects in this gene are a cause of Brugada syndrome type 2 (BRS2) as well as sudden infant death syndrome (SIDS).
SLC27A	Solute carrier family 27 (fatty acid transporter), member 6	Metabolism Lipid metabolism Very-long-chain fatty acid metabolism	Ligase activity Long-chain-fatty-acid-CoA ligase activity	5q23.3	This gene encodes a member of the fatty acid transport protein family (FATP). FATPs are involved in the uptake of long-chain fatty acids and have unique expression patterns.
TMPRSS11B	Transmembrane protease, serine 11B	Proteolysis and peptidolysis	Chymotrypsin activity Trypsin activity	4q13.2	TMPRSS11B is active in the following subcellular-locations: extracellular region, integral to plasma membrane. It's molecular function is through catalysis of the hydrolysis of internal, alpha-peptide bonds in a polypeptide chain by a catalytic mechanism that involves a catalytic triad consisting of a serine nucleophile that is activated by a proton relay involving an acidic residue (e.g. aspartate or glutamate) and a basic residue (usually histidine).

3.4.6 Biological interpretation of the differentially expressed meta-genes

To obtain more biological understanding of why expressed genes in OSCCs are different to the normal oral tissue samples, the Ingenuity pathway analysis (IPA, http://www.ingenuity.com/) system was recruited to examine gene molecular function, biological process, canonical pathways, and transcriptional regulators analysis. IPA system depends on the knowledge database of selected functional and regulatory interactions extracted from the literature and provides integrated graphical representation of the biological relationships between genes and gene products considering both up- and downregulated genes from comparison analysis.

3.4.6.1 Functional analysis

The most relevant activated and inhibited functions at FDR<0.05 are reported in Figure 2.2. Cellular Movement, Growth, Proliferation, Development, and Cell-To-Cell Signalling and Interaction functions displayed at the top of the perturbed molecular cellular functions in relation to cancer.

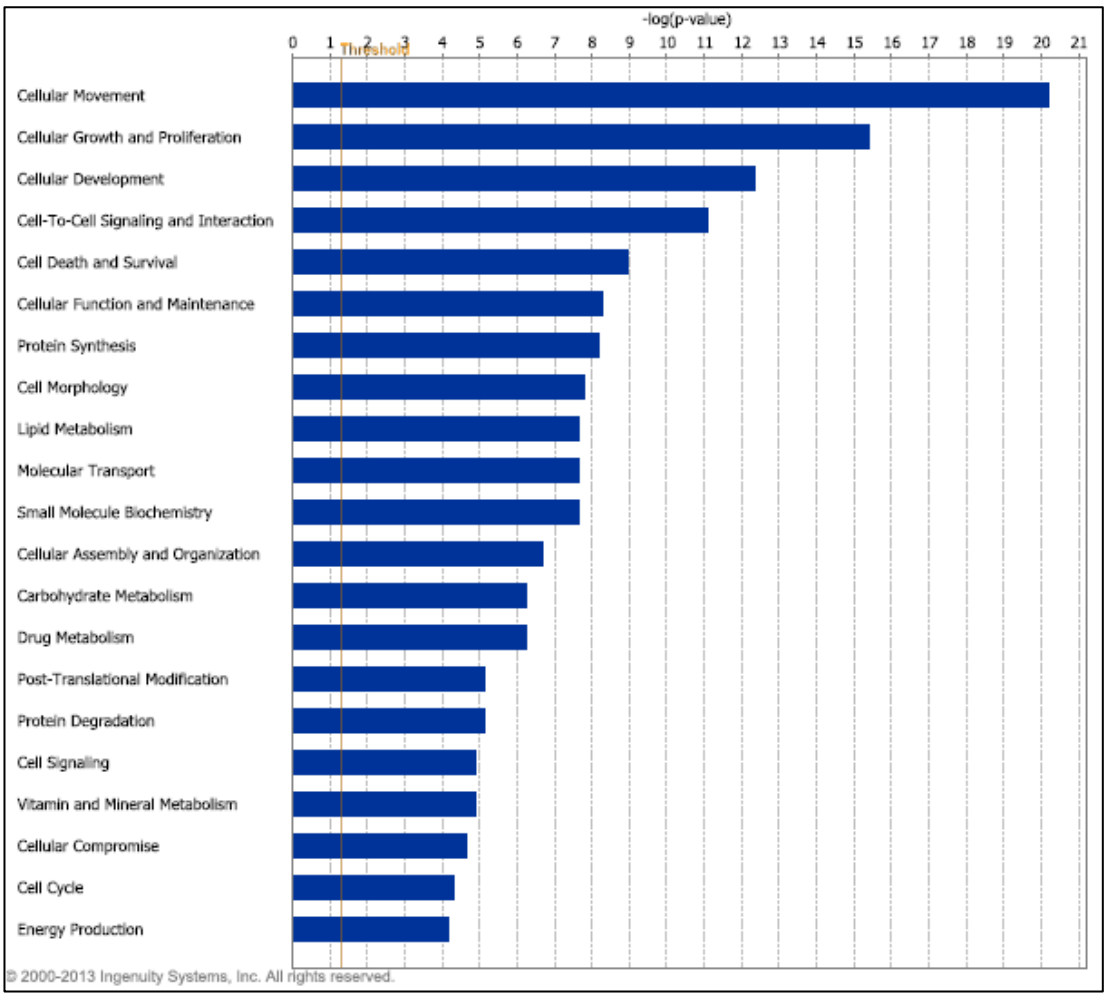


Figure 3.12: Molecular cellular functions of meta-genes related to cancer with the greatest numbers of differentially expressed genes using IPA functional annotation analysis.

Higher bars denote greater significance and the numbers of genes above the yellow line, which signifies a p value of 0.05, represent those that are significantly differentially expressed than by chance alone

3.4.6.2 Associated network analysis

Investigation of the biological relationships between genes and gene products was carried out by performing a network analysis of the meta-genes. The master network was constructed by connecting the selected genes involved in cancer using Fisher's exact test. The highly connected nodes are likely to reflect their ability to regulate large number of genes in the master network and potentially to control the gene expression pattern that identified tumour signature of the comparison. Upregulation

of IL8, IL6, IL1B, FN1, PTGS2 (COX2), SPP1, CXCL3, STAT1, SERPINE1, MMP3, MMP9, ISG15, OSM, CXCL1, CXCL2, CXCL10, PLAU, MMP1, MMP13, AHR, HOXA10, COL1A1, CDKN2A, and CCL20 along with downregulation of ADIPOQ, PLA2G2A, MAPK3, IL1RN, CXCL12, PPARGC1A, NR3C2, RORC, and MUC1 nodes were identified at the centre of the network (Figure 2.3). As molecular relationships represented on the network include not only activation or inhibition of expression, but also protein-protein interactions, DNA-protein interactions and activation, localization, inhibition of the corresponding proteins (Debily et al., 2009), it is not surprising that microarrays may fail to identify some of the hub genes as significantly differentially expressed among the tumour versus normal samples. These genes might play a major role through protein activation for instance. Alternatively, their modulations may be very subtle and below the threshold for reliable detection of differences of the microarray platform despite its high sensitivity.

Additionally, a more specific master network was constructed by connecting the selected genes involved in only OSCC and HNSCC using Fisher's exact. Upregulation of SERPINE1, MMP3, MMP9, MMP1, IL8, IL6, IL1B, PLAU, IGFBP3, FN1, PTGS2 (COX2), SPP1, STAT1, CDKN2A, and COL1A1 along with downregulation of CXCL12, CLU, PPARGC1A, and MUC1 nodes test (Figure 3.14) forming the main structure of the network. Interestingly, these nodes forming the majority of the hub genes positioned at the centre of the previous network relating to cancer disease in general.

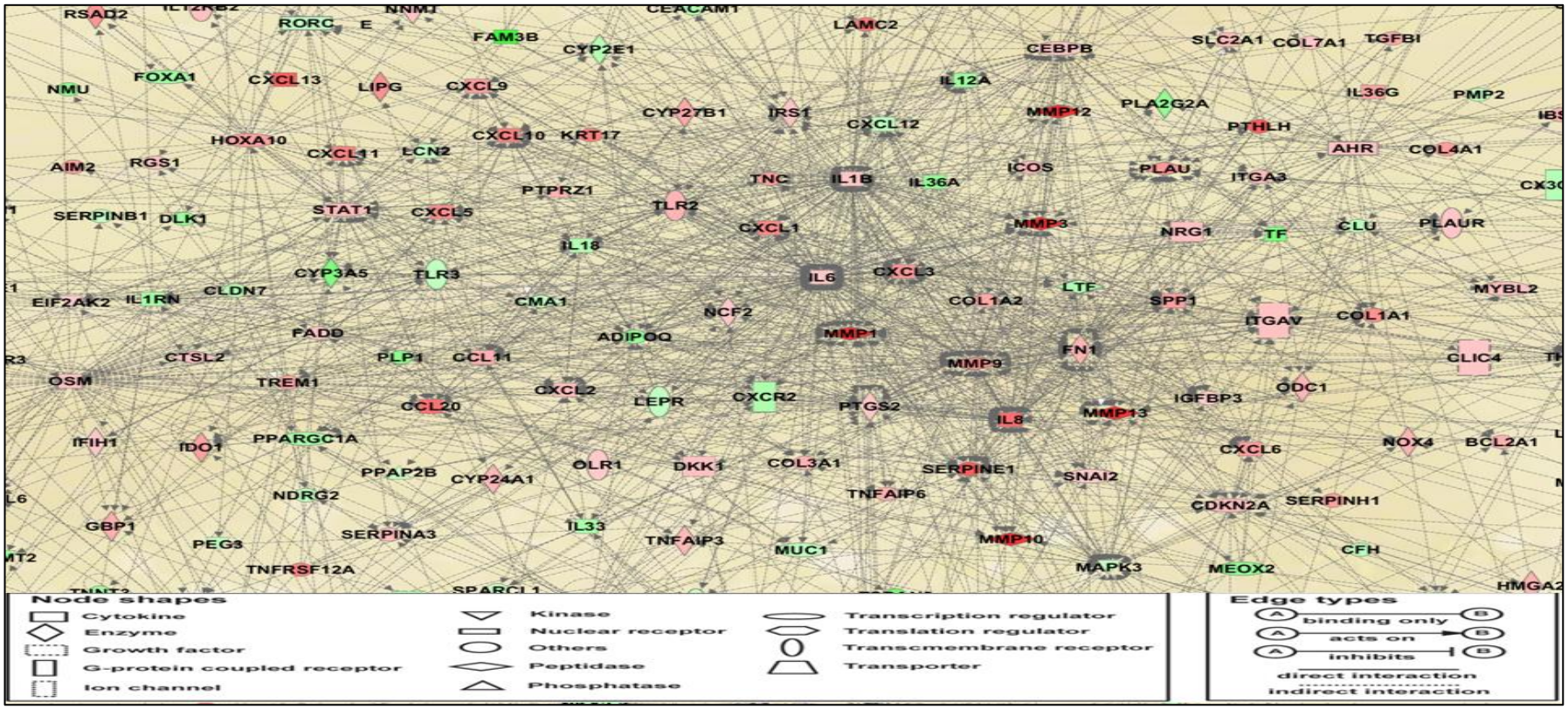


Figure 3.13: Central area of the comprehensive molecular network of meta-genes related to cancer.
Master network built by connecting significant genes involved in cancer disease, identified by IPA tool (version 9.0) from up- and downregulated genes using overlap function of core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The intensity of colours reflects the magnitude of the fold change of the expressed molecule. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The network represents significant genes at FDR level ≤ 0.05 with no fold change cut-off. For full view of the network, see supplementary file 6.

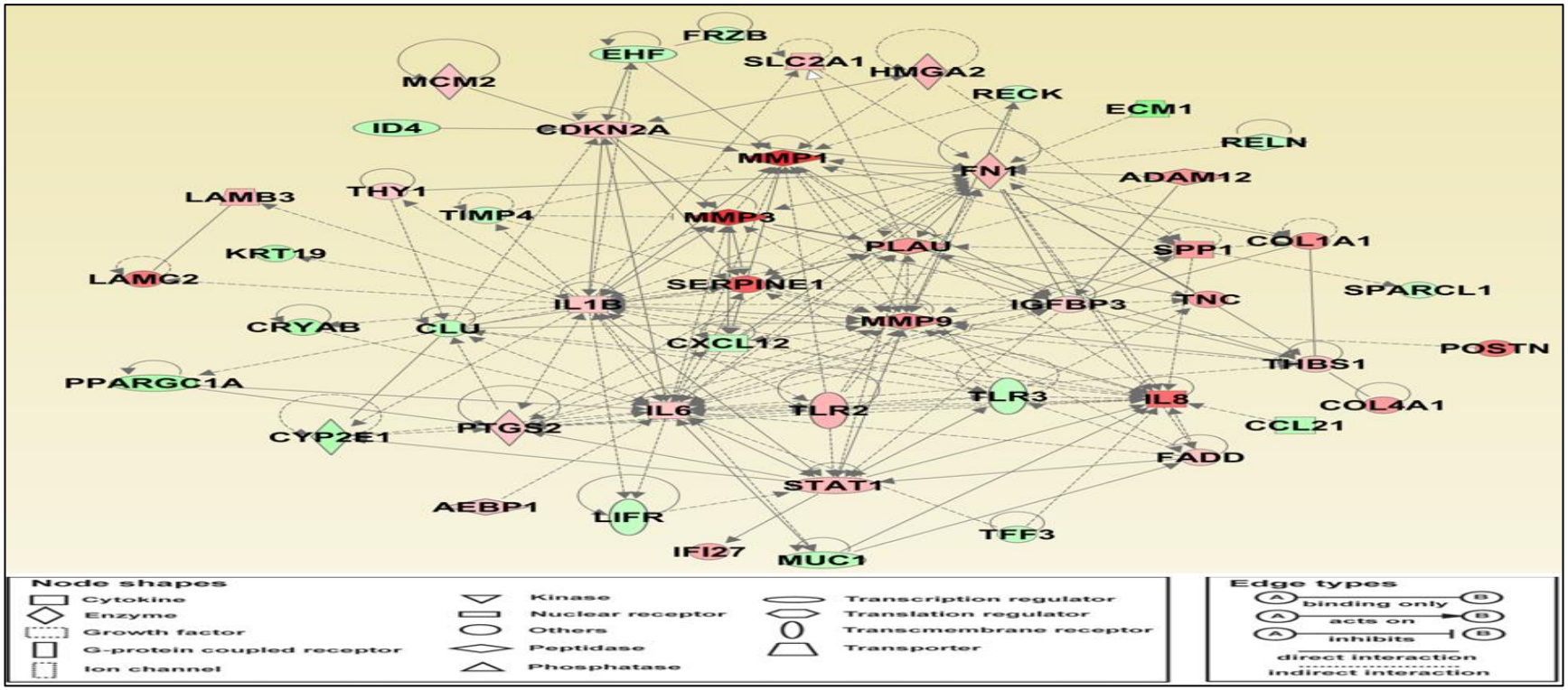


Figure 3.14: Master molecular network of meta-genes related to OSCC and head and neck cancer.
Master network built by connecting significant genes involved in OSCC, HNSCC, and oesophageal cancer, identified by IPA tool (version 9.0) from up- and down-regulated genes using overlap function of core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The intensity of colours reflects the magnitude of the fold change of the expressed molecule. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The network represents significant genes at FDR level ≤ 0.05 with no fold change cut-off.

3.4.6.3 Upstream regulators

It's well known that regulation of human genes depends on sets of transcription factors rather one individual factor. For this reason, "Transcriptional Regulator Analysis" tool available from IPA system was recruited, identifying the activation of STAT3, CTNNB1, and JUN, along with the inhibition of MEF2C and SMARCA4 potential regulators which have a major contribution in the tumorigenic process.

3.4.6.4 Canonical pathway analysis

By utilizing canonical pathway analysis tool, Role of Tissue Factor in Cancer, Inhibition of matrix metalloproteases, Calcium Signalling, Dendritic cell maturation, IL-17 Signalling, LXR/RXR activation, IL6 signalling, Oncostatin M signalling, Leukocyte extravasation signalling, Interferon signalling, xenobiotic metabolism signalling pathways, as well as FAK, P53, Wnt/β catenin transcriptional signalling pathways, were found the most common perturbed pathways involved in the process of tumour transformation. Overlying these pathways over those dysregulated from the review study in chapter 2, show considerable overlapping as well. Role of Tissue Factor in Cancer, Inhibition of matrix metalloproteases, Leukocyte extravasation signalling, and xenobiotic metabolism signalling pathways were among those top dysregulated pathways common between the two set of genes indicating the significant similarity despite the differences in the number of genes forming the two sets (only 368 common genes forming the set of genes in the review study comparing to 8995 genes of the meta-analysis signature).

3.4.6.5 Biomarker analysis

In an attempt to find reportedly potential biomarkers among the differentially expressed set of meta-genes, running biomarker filter analysis (IPA) identified 35 specific biomarkers related to OSCC, HNSCC, and oesophageal cancer (Table 3.6 and Figure 3.15). Only 14 biomarkers were differentially expressed in tumour samples at fold change cut-off log value ± 1.5 (Figure 3.16). These include ADAM12, COL1A1, COL4A1, IL8, LAMC2, MMP1, MMP3, MMP9, MMP12, PDPN, POSTN, PTGS2, SERPINE1, and STAT1 over-expression, associated with under-expression of SPINK5 and TGM3 biomarkers. Figure 3.16 displays the role of each biomarker in diagnosis, progression, prognosis, and efficacy of oral, head and

neck, and oesophageal cancer using "Path Designer" advanced tool in IPA. The vast reduction in the number of biomarkers (from 35 to 14) using fold change cut-off criterion highlights the necessity not to limit biomarker tools based on the fold change magnitude. A specific biomarker showing consistent existence in a specific type of disease, not necessarily highly expressed to be considered a reliable biomarker, as long as sufficient sensitive monitoring tool is provided.

Uploading these biomarkers to canonical pathway tool (IPA), displayed the top perturbed pathways in the following rank; Glucocorticoid receptor signalling, Inhibition of Matrix Metalloproteases, HIF1α Signalling, IL17 Signalling, Dendritic Cell Maturation, IL6, and IL8 signalling pathways. Although the number of biomarkers is not appropriate for build-up and identification of pathways, it remains displaying essential pathways reported in the tumorigenic process.

Table 3.6: Oral, oropharyngeal, and head and neck cancer biomarkers of differentially expressed meta-genes.

Symbol	Location	Family	Fold Change*	Biomarker Application(s)
MMP1	Extracellular Space	Peptidase	6.1	diagnosis, response to therapy: head and neck, oesophageal cancer
MMP3	Extracellular Space	Peptidase	4.2	diagnosis: head and neck, oesophageal cancer
MMP12	Extracellular Space	Peptidase	3.9	diagnosis: oesophageal cancer
SERPINE1	Extracellular Space	Other	3.1	diagnosis, disease progression, prognosis, response to therapy: oral, head and neck cancer
IL8	Extracellular Space	Cytokine	2.7	diagnosis, disease progression, prognosis, response to therapy: head and neck, oral cancer
LAMC2	Extracellular Space	Other	2.5	diagnosis: oral cancer
POSTN	Extracellular Space	Other	2.5	unspecified application: head and neck cancer
COL1A1	Extracellular Space	Other	2.1	diagnosis; oral cancer
MMP9	Extracellular Space	Peptidase	2.1	diagnosis, disease progression, prognosis: head and neck, oral cancer
ADAM12	Plasma Membrane	Peptidase	1.9	diagnosis; head and neck cancer
COL4A1	Extracellular Space	Other	1.8	diagnosis; oral cancer
PDPN	Plasma Membrane	Transporter	1.6	Diagnosis ,disease progression, prognosis: head and neck, oral cancer

Continue

Symbol	Location	Family	Fold Change*	Biomarker Application(s)
IFI27	Cytoplasm	Other	1.5	disease progression; head and neck cancer
TMEM45A	Plasma Membrane	Other	1.4	disease progression: head and neck cancer
RBP1	Extracellular Space	Transporter	1.3	diagnosis: oesophageal cancer
SLC2A1	Plasma Membrane	Transporter	1.3	efficacy: head and neck cancer
CDKN2A	Nucleus	transcription regulator	1.2	diagnosis, disease progression, efficacy, prognosis; oral, head and neck cancer
STAT1	Nucleus	transcription regulator	1.2	diagnosis, prognosis: head and neck cancer
DKK1	Extracellular Space	growth factor	1.1	diagnosis, prognosis; oesophageal cancer
IL1B	Extracellular Space	Cytokine	1.1	diagnosis, prognosis: oral cancer
MCM2	Nucleus	Enzyme	1.1	diagnosis, disease progression, prognosis: head and neck cancer, oropharyngeal cancer
ECT2	Nucleus	Other	1	disease progression; head and neck cancer
IL6	Extracellular Space	Cytokine	1	diagnosis, disease progression, prognosis, response to therapy: head and neck, oesophageal cancer
PTGS2	Cytoplasm	Enzyme	1	diagnosis, disease progression, prognosis: oral, head and neck, oesophageal cancer
SPARCL1	Extracellular Space	Other	-1	disease progression: head and neck cancer
EHF	Nucleus	transcription regulator	-1.1	unspecified application: head and neck cancer
TIMP4	Extracellular Space	Other	-1.1	diagnosis, prognosis: head and neck cancer
CCL21	Extracellular Space	Cytokine	-1.2	diagnosis, disease progression; oral cancer
CXCL12	Extracellular Space	Cytokine	-1.2	diagnosis, disease progression, prognosis; head and neck cancer
CRYAB	Nucleus	Other	-1.3	diagnosis, disease progression, prognosis; oral cancer, head and neck cancer
KRT19	Cytoplasm	Other	-1.4	diagnosis, disease progression, prognosis: head and neck cancer
SERPINB13	Cytoplasm	Other	-1.4	unspecified application: head and neck cancer
ZNF185	Nucleus	Other	-1.4	unspecified application: head and neck cancer
SPINK5	Extracellular Space	Other	-2.6	unspecified application: head and neck cancer
TGM3	Cytoplasm	Enzyme	-3.8	unspecified application: head and neck cancer

* Log2 folds change values.

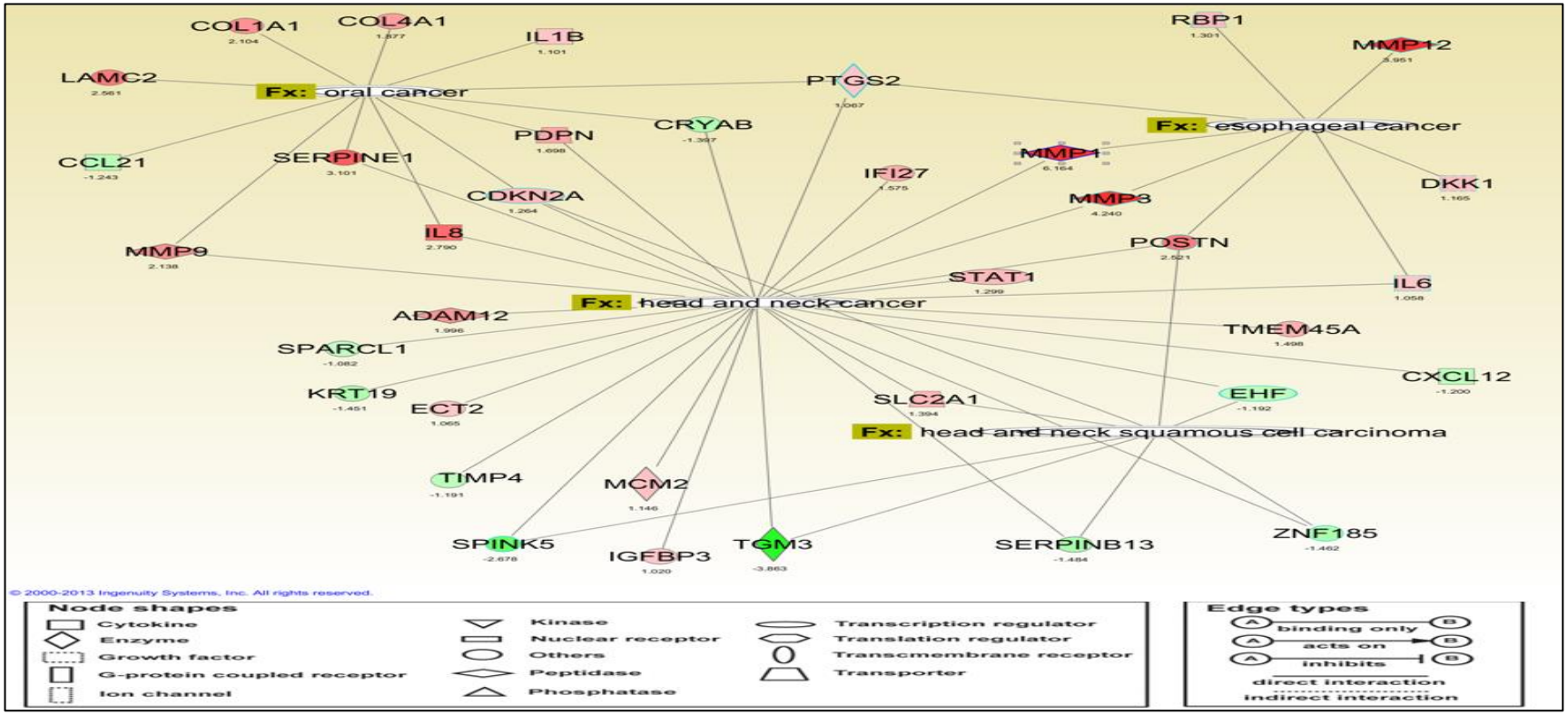


Figure 3.15: Biomarker network of meta-genes involved in OSCC and head and neck cancer

Biomarker network built by connecting significant genes involved in OSCC, HNSCC, and oesophageal cancer, identified by biomarker filter tool (IPA, version 9.0) from up- and down-regulated genes using overlap function of the biomarker analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). Each edge supported by at least one study in the literature and peer reviewed by Ingenuity team. The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. The network represents significant genes at FDR level ≤ 0.05 with no fold change cut-off.

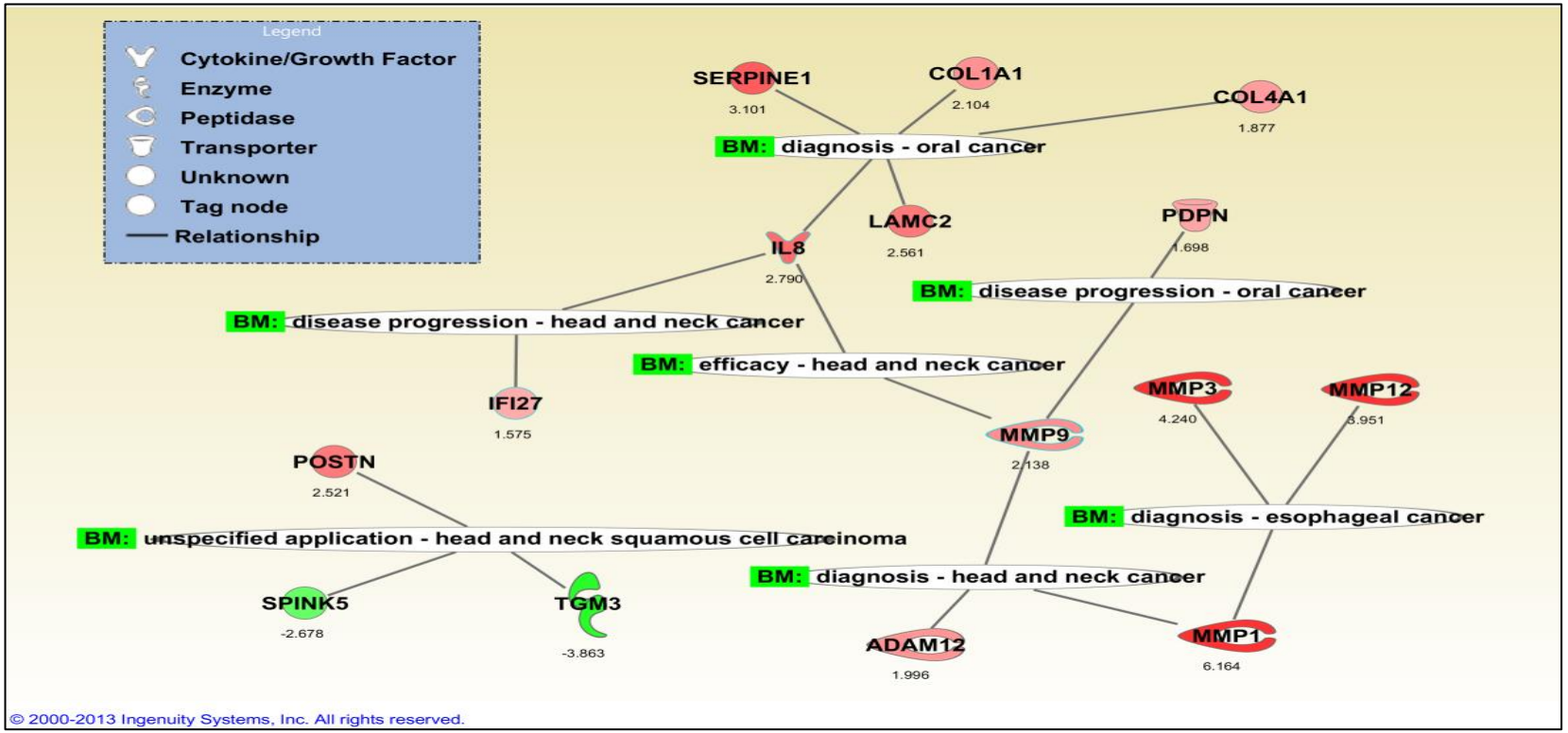


Figure 3.16: Highly expressed meta-genes biomarkers related to OSCC, HNSCC, and Oesophageal cancer.
Biomarker network built by connecting significant genes involved in OSCC, HNSCC, and oesophageal cancer, identified by biomarker filter tool (IPA, version 9.0) from up- and down-regulated genes using advanced function (Path Designer) of the biomarker analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). Each edge supported by at least one study in the literature and peer reviewed by Ingenuity team. The tumours relative to normal overexpressed genes are shaded in red and under-expressed genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. The network represents significant genes at FDR level ≤ 0.05 with 1.5 fold change cut-off.

3.4.7 Comparison between meta-genes and common genes of the review study

Comparing meta-genes set (334genes at FDR level=0 and fold change cut-off ±1.5) and the common genes procured from the review study in chapter 2 (368 genes), highlighted over a quarter (96) of common genes as differentially expressed in both datasets. Reducing FDR stringency of the differentially expressed meta-analysis to ≤ 0.05 with no fold change cut-off criteria increased the overlap to 326 genes, 89% of the total. This variation in the number of the final set of common genes highlights the impact of using different bioinformatics criteria and analysis approaches on the formulation of the final list of genes.

Among the differentially expressed genes, several potential diagnostic biomarkers were identified which have previously been reported in OSCC and HNSCC (Table 3.6). Notably, elevation of seven members of collagens (COL1A2, COL1A1, COL4A1, COL5A2, COL4A2, COL5A1, and COL11A1), six members of MMPS (MMP1, MMP3, MMP9, MMP10, MMP12, and MMP13), and five members of C-X-C motif chemokines (CXCL1, CXCL6, CXCL9, CXCL10, and CXCL11) were observed which accompanied with repression of three keratins (KRT4, KRT13, and KRT15).

Additional potential key players of tumour, including elevation of SPP1, LAMC2, IL8, IFI6, ISG15, MYO1B, SERPINE1, BS2, AIM2, and PTHLH along with repression of CRNN, MAL, TGM3, EMP1, CRISP3, PPP1R3C, ECM1, IL1RN, SCEL, SPINK5, TGFBR3, CEACAM5, CEACAM7, TGM1, and TYRP1 genes were observed in tumour relative to normal cohorts (Table 3.7).

Table 3.7: Significant genes commonly expressed between meta-genes set and common genes from the review study (chapter 2).*

ENSEMBL Gene ID	Gene Symbol	Fold Change	ENSEMBL Gene ID	Gene Symbol	Fold Change
Over-expressed genes in tumour samples			**Under-expressed genes in tumour samples**		
ENSG00000196611	MMP1	6.16	ENSG00000096006	CRISP3	-5.69
ENSG00000137745	MMP13	4.42	ENSG00000143536	CRNN	-4.95
ENSG00000166670	MMP10	4.31	ENSG00000172005	MAL	-4.28
ENSG00000149968	MMP3	4.24	ENSG00000170477	KRT4	-4.01
ENSG00000110347	MMP12	3.95	ENSG00000125780	TGM3	-3.86
ENSG00000108821	COL1A1	3.43	ENSG00000016602	CLCA4	-3.51
ENSG00000060718	COL11A1	3.43	ENSG00000126233	SLURP1	-3.38
ENSG00000106366	SERPINE1	3.10	ENSG00000171401	KRT13	-3.23
ENSG00000087494	PTHLH	3.09	ENSG00000140519	RHCG	-3.11
ENSG00000187608	ISG15	2.80	ENSG00000141338	ABCA8	-3.01
ENSG00000169429	IL8	2.79	ENSG00000136155	SCEL	-3.00
ENSG00000058085	LAMC2	2.56	ENSG00000119938	PPP1R3C	-2.86
ENSG00000126709	IFI6	2.55	ENSG00000108602	ALDH3A1	-2.75
ENSG00000133110	POSTN	2.52	ENSG00000133710	SPINK5	-2.68
ENSG00000078098	FAP	2.49	ENSG00000127324	TSPAN8	-2.57
ENSG00000169245	CXCL10	2.46	ENSG00000196344	ADH7	-2.56
ENSG00000169248	CXCL11	2.39	ENSG00000108924	HLF	-2.38
ENSG00000186832	KRT16	2.35	ENSG00000163815	CLEC3B	-2.36
ENSG00000130303	BST2	2.33	ENSG00000108515	ENO3	-2.32
ENSG00000163739	CXCL1	2.27	ENSG00000106258	CYP3A5	-2.29
ENSG00000100985	MMP9	2.14	ENSG00000108839	ALOX12	-2.16
ENSG00000062038	CDH3	2.08	ENSG00000107165	TYRP1	-2.15
ENSG00000100342	APOL1	2.02	ENSG00000105388	CEACAM5	-2.14
ENSG00000204262	COL5A2	2.01	ENSG00000164120	HPGD	-2.14
ENSG00000122861	PLAU	2.01	ENSG00000018625	ATP1A2	-2.08
ENSG00000138755	CXCL9	1.94	ENSG00000143369	ECM1	-2.07
ENSG00000124875	CXCL6	1.89	ENSG00000169583	CLIC3	-2.03
ENSG00000187498	COL4A1	1.88	ENSG00000167165	UGT1A7	-2.01
ENSG00000118785	SPP1	1.87	ENSG00000152642	GPD1L	-1. 98

Continue

ENSEMBL Gene ID	Gene Symbol	Fold Change	ENSEMBL Gene ID	Gene Symbol	Fold Change
ENSG00000163568	AIM2	1.85	ENSG00000069535	MAOB	-1.98
ENSG00000130635	COL5A1	1.84	ENSG00000134531	EMP1	-1.98
ENSG00000041982	TNC	1.83	ENSG00000172478	C2ORF54	-1.93
ENSG00000149257	SERPINH1	1.83	ENSG00000136694	IL1F6	-1.92
ENSG00000128641	MYO1B	1.77	ENSG00000109255	NMU	-1.90
ENSG00000105928	DFNA5	1.71	ENSG00000092295	TGM1	-1.87
ENSG00000162493	PDPN	1.70	ENSG00000161281	COX7A1	-1.87
ENSG00000164692	COL1A2	1.67	ENSG00000108242	CYP2C18	-1.82
ENSG00000137965	IFI44	1.67	ENSG00000007306	CEACAM7	-1.76
ENSG00000137959	IFI44L	1.65	ENSG00000168447	SCNN1B	-1.75
ENSG00000120708	TGFBI	1.63	ENSG00000167757	KLK11	-1.69
ENSG00000165949	IFI27	1.57	ENSG00000148671	C10ORF116	-1.66
ENSG00000155380	SLC16A1	1.56	ENSG00000118898	PPL	-1.64
ENSG00000117228	GBP1	1.53	ENSG00000167759	KLK13	-1.64
ENSG00000134871	COL4A2	1.52	ENSG00000010438	PRSS3	-1.64
ENSG00000123610	TNFAIP6	1.50	ENSG00000167880	EVPL	-1.64
			ENSG00000149809	TM7SF2	-1.63
			ENSG00000136689	IL1RN	-1.61
			ENSG00000171346	KRT15	-1.61
			ENSG00000069702	TGFBR3	-1.59
			ENSG00000095585	BLNK	-1.57
			ENSG00000074416	MGLL	-1.55
			ENSG00000143196	DPT	-1.52

* FDR value = 0. Fold change represents the log2 magnitude of expression in OSCC relative to normal oral samples.

3.5 Discussion

Both supervised and unsupervised analyses largely discriminated gene expression profiles between OSCC and normal mucosa samples, with the majority of tumour and normal samples clustering together. Although absolute specificity was not achieved, increasing the criteria stringency of the comparison between normal and tumour cohorts (to accept only the top 100, 50, 38, 8, and 3 genes), showed an improved discrimination power by a small set of genes as predictive discriminators. These results proved that SAM tool (*siggenes* package) is capable to elect the top discriminatory genes beyond the level of FDR = 0. As the sets of 334, 100, 50, 38, 8, and 3 genes are all significant at FDR level= 0, further adjustments of delta value (explained in materials and methods, 1.11.4) with FDR = 0, continue nominating smaller number of significant genes with better, or at least similar appropriate discrimination power of tumour from normal samples rather than selecting genes randomly. One of the four merged datasets in the meta-analysis study was used by its individual study to extract prediction models for OSCC (Chen et al., 2008a), however, their models failed in discriminating tumour from normal samples of the independent datasets (GSE6631, UK, and Sri Lankan sets comparing tumour vs. normal samples) that used for validation of the models harvested by current study (the models of 8 and 3 genes). This highlights the advantage of meta-analysis (employed large and wide range samples from different research centres) in creating more robust and comprehensive predictive signature. The model that consists of 8 genes showed the best discriminating power among the prediction models. Whilst 3 of the 8 genes are already recruited as biomarkers for OSCC, HNSCC, and gastric cancer (elevation of LAMC2 and MMP1 along with repression of MAL respectively), the other 5 genes (elevated PTHLH and TPBG along with repressed GPD1L, SLC27A6, and TMPRSS11B) might be novel biomarkers for OSCC. Cross-correlating the 8 genes of the model over the genes harvested from the review study was carried out in chapter 2, which displayed complete match in 7 out of 8 genes of the model whilst only one gene (TMPRSS11B) shown absent in the review study (chapter 2). MMP1, LAMC2, PTHLH, and TPBG genes were found regularly over-

expressed whereas, MAL, GPD1L, and SLC27A6 constantly under-expressed in tumour relative to normal samples of the review study (Table 3.8).

Table 3.8: Distribution and expression of the 8 genes model among external datasets.

Gene Symbols	Verified Biomarkers	GSE6631 Dataset*	UK Samples*	Sri Lankan samples*	Review Studies Samples*
MMP1↑	↑ HNSCC	Elevated	Elevated	Elevated	Elevated
LAMC2↑	↑ OSCC	Elevated	Elevated	Elevated	Elevated
PTHLH↑		Elevated	Elevated	Elevated	Elevated
TPBG↑		Elevated	Elevated	Elevated	Elevated
MAL↓	↓ Gastric cancer	Repressed	Repressed	Repressed	Repressed
GPD1L↓		Repressed	Absent	Absent	Repressed
SLC27A6↓		Absent	Repressed	Repressed	Repressed
TMPRSS11B↓		Absent	Absent	Absent	Absent

* Dataset refers to tumour vs. normal samples comparison. ↑ and ↓ arrows denoted to the elevated and repressed expression levels. Absent refers to the absence of the gene probe in the platform manufacturer used in the analysis of the corresponding dataset.

In fact, PTHLH exhibited as one of the potential key players over-expressed in the comprehensive molecular network of the common genes related to cancer in the review study (Chapter 2). Previous findings from both in vivo and in vitro analyses showed essential associations regarding the role of PTHLH in breast cancer, suggesting the proliferative effects of PTHLH in breast cancer cells are mediated through regulation of cell cycle and apoptosis, and that controlling PTHLH production in breast cancer may be therapeutically useful (Falzon and Du, 2000, Tabuenca et al., 1995, Sepulveda et al., 2002). Additional reports specified the possible significance of autocrine/paracrine effects of PTHLH in the progression of early-stage lung adenocarcinoma, as well as, growth and dedifferentiation of malignant cells in cervical cancer (Monego et al., 2010, Dunne et al., 1994). Some reports identified PTHLH to serve as a tumour marker in animal models of pancreatic cancer and may be a useful tumour marker for clinical pancreatic adenocarcinoma (Bouvet et al., 2002), while additional findings suggested PTHLH as a common target molecule in specific immunotherapy for patients with a wide variety of tumour types, particularly bone metastases (Arima et al., 2005). Further

results pointed to the role of PTHLH as a cancer modifier gene in skin tumourigenesis (Kaiser et al., 1994, Manenti et al., 2000). Lastly, overexpression of PTHLH reported being strongly related to tumour progression of gastric cancer (Ito et al., 1997).

TPBG gene (trophoblast glycoprotein gene) encodes 5T4 oncofetal antigen. The human oncofetal antigen 5T4 (h5T4), a transmembrane glycoprotein, is overexpressed by a wide spectrum of cancers including colorectal, ovarian, and gastric, but with a limited normal tissue expression. Such properties make 5T4 an effective putative target for cancer immunotherapy (Mulryan et al., 2002). Moreover, elevated level of TPBG reported a relation to Leiomyoma uterine cancer (Luo et al., 2005).

The reduced expression of SLC27A6 was defined in oral submucous fibrosis (OSF), a high-risk precancerous condition of the oral cavity (Li et al., 2008). Recent work identified the enzyme glycerol-3-phosphate dehydrogenase 1-like (GPD1L) as a novel regulator of HIF1-alpha stability. Whilst the ectopic expression of GPD1L attenuates the hypoxic response, preventing complete HIF1-alpha induction, decreased GPD1L protein causes increased stability of HIF-1alpha (Kelly et al., 2011). The last gene in the model, TMPRSS11B, showed, according to a recent study, differentially downregulation between both normal and dysplastic tissues, as well as, between dysplastic and tumour tissues of oral cavity and it showed consistent downregulation throughout the entire process of oral carcinogenesis (Sumino et al., 2013).

Further validations of the model of 8 genes with additional external datasets are required to confirm it’s specificity in discriminating OSCC from other tumour types of the body as well as other molecularly similar inflammatory processes (e.g., atherosclerosis and periodontitis).

Both potential biological and technical reasons are accountable for incomplete separation between normal and tumour samples. Tumours may genuinely have molecular profiles resembling normal tissue and normal samples may have gene

expression profiles similar to tumours, which could be evidence of occult disease (not identifiable by histology). In addition, non-biological explanations could result from incomplete batch correction due to various stages of microarray pre-processing including; distinct laboratory conditions, reagents, and personnel. Previous studies have demonstrated that batch effects can occur even when the same samples are processed on the same chips, in the same lab on different days (Kitchen et al., 2010). In fact, many gene expression datasets were underpowered (Ein-Dor et al., 2006) whilst other study described issues of multiple-testing as a consequence of datasets with many more features than samples (Clarke et al., 2008), though ComBat method has been shown to reduce batch effects to an absolute minimum in a number of studies (Johnson et al., 2007), (Walker et al., 2008, Chen et al., 2011a). A number of previous studies have integrated microarray datasets from cancer studies, using different integrating tools, in order to generate consensus findings independent of the cohort or platform used. According to recent published work (Turnbull et al., 2012), ComBat expressed the best performance. Furthermore, the study demonstrated that integrating comparable gene expression datasets processed on Affymetrix and Illumina platforms increases the statistical power.

Searching for genes that are both differentially expressed in the meta-analysis study and commonly identified as communal genes across published studies in chapter two, highlights the importance of the criteria used in differential gene expression analysis (fold change cut-off and FDR criteria). Altering the stringency of criteria of the differentially expressed meta-genes from FDR level =0 and fold change cut-off ±1.5 to FDR level ≤ 0.05 with no fold change cut-off criteria, increased the number from 97 to 326 cross-correlated genes forming 89% of the total common harvested genes (368) from the review study (illustrated in chapter 2), with only 22% of these expressed genes having less than 0.5 log fold change. This variation in the number of the final set of cross-correlated genes highlighted the impact of the differences in bioinformatics approaches and the criteria selected in formulating the final set of genes. The result of current study suggests dealing with fold change as cut-off criteria with caution, especially in determining the final set of genes and identifying the subsequent biological dysregulated pathways, molecular functions, and

processes, since molecular relationships in the system biology comprise not only activation or inhibition of expression, but also protein-protein interactions, DNA-protein interactions and activation, localization, and inhibition of the corresponding proteins (Debily et al., 2009), and it is not surprising that microarrays results (especially those with a high fold change cut-off) may fail to identify some of the hub genes as significantly differentially expressed among the tumour versus normal samples. These genes might play a major role through protein activation, for instance. Instead, their modulations may be very subtle and below the threshold for reliable detection of differences of the microarray platform despite its high sensitivity (Debily et al., 2009). Alternatively, study suggestion is to consider comprehensive inclusion of the whole expressed genes, accompanied with using a Biological Annotation tool (like; IPA, KEGG, STRING, etc.) in an attempt to obtain a true and complete understanding of the biological concept of the system biology from which the sample extracted at the time of collection, considering the fact that genes are highly interchangeable. However, setting a low threshold for a significant correlation between genes will result in the inclusion of many spurious links, whereas a high threshold will control the false-positive rate at the expense of omitting many genuine edges. For this reason, the suggestion of this study is the production of two separate sets; the one with less stringent criteria including large amount of genes is recruited for subsequent biological annotation purposes (like for identifying the dysregulated pathways and the affected transcriptional regulators), after which selection of a specific disease and its relative pathways will be focused on. The other set of small number of genes can be obtained by applying stringent criteria to fulfil the requirements of monitoring tool with a high specificity and can used as a biomarker identification tool (where a small number of gene set is suitable for clinical screening test). However, a set of genes with a reasonable size was obtained from the meta-analysis study as well as those cross-correlating between the review study (chapter 2) and the meta-analysis signature of this chapter with most of genes related to OSCC and HNSCC. Of these genes, potential biomarkers involved in OSCC and HNSCC were observed which confirmed the efficiency of bioinformatics approaches used in merging datasets from different studies. Importantly, these common differentially expressed genes are obtained from different collection centres, population groups,

different stages of microarray pre-processing including; distinct laboratory conditions, reagents, and personnel. Two datasets from the met-analysis study (Ye et al. and Chen et al. studies) were among the microarray datasets in the review study of chapter 2, have already been illustrated to have very low overlapping signature (22.8% of the total common genes).

Similar to the results of the review study in chapter 2, significant over-expression of chemokines in tumours was observed which has been ascribed frequently to constitutive stimulus of NFκB, the signalling pathway which plays a key role in anti-apoptosis and is widely proposed now as potential molecular tumorigenic marker in various types of tumours including colorectal, bladder, pancreatic, breast, ovarian, as well as oral cancer (He et al., 2009). Elevated level of several members of CXC family in tumours was another finding which behaves as angiogenic factors (Vandercappellen et al., 2008, Vinader and Afarinkia, 2012, Zhu et al., 2012). These include CXCL8 (IL8), CXCL1, CXCL2, CXCL3, CXCL5, CXCL6, and CXCL7. Besides their functions in the immune system, they also play a critical role in tumour initiation, promotion and progression (Vandercappellen et al., 2008). CXC chemokines 2, 3, 9, and genes encoding for extracellular matrix proteins and collagens have previously been involved in the motility and invasion of tumour (Chen et al., 2008a).

In an attempt to explore more specific signature of OSCC and understand the tumorigenic process reflected by the simultaneous perturbed genetic and metabolic system, additional analysis was conducted in chapter 5 to integrate transcriptomic signature harvested from the meta-analysis study in this chapter and the metabolomic signature obtained from the review study of metabolomic profiling (explained in chapter 5) by employing various IPA biological tools.

Chapter 4: Gene expression profiling reveals biological pathways responsible for phenotypic heterogeneity between UK and Sri Lankan oral squamous cell carcinomas

4.1 Introduction

Oral squamous cell carcinoma is responsible for a significant mortality and morbidity with poor overall survival. The pace of invasiveness, metastasis, and recurrence of OSCCs is a multifactorial event. The tumour site, type of the environmental carcinogens, immune response and resistance of the body against the disease all play important role. To date, little is known about the biological mechanisms underlying such phenotypic variation in different populations.

4.2 Aims and objectives of the study

1-To define gene expression profiles of UK and Sri Lankan OSCCs compared to their normal comparators using supervised and unsupervised strategies and to identify associated pathways.

2-To determine the impact of differences in gene expression between UK and Sri Lankan normal oral mucosa upon the above two profiles.

3-To define gene expression differences between UK and Sri Lankan OSCC and to further assess the contribution of expression differences between groups of normal tissues has on those seen between the two groups of OSCC

4-To examine the most related genetic factors, carcinogenic related pathways, and molecular processes that might be responsible for the phenotypic heterogeneity of OSCC between UK and Sri Lankan population groups which impact the vast differences in clinical behaviour of the two tumour groups.

4.3 Materials and Methods

4.3.1 Sample characteristics and biopsy specimens

All biopsy specimens of OSCC and normal oral mucosa were harvested with appropriate ethical approval and informed consent of patients. Identical protocols for tissue collection and processing were used in both countries.

Tissue samples of OSCC from UK population were obtained from sequential incident cases treated surgically by a single consultant at University Hospital Birmingham NHS Foundation Trust, Birmingham, UK Tissues specimens were harvested during the period 2001 to 2004 and stored. A total of 24 samples yielded RNA that was of suitable quality for Microarray analysis (Table 4.2). Patient's population included 13 (54%) males and 11 (46%) females. The average age of patients was 61.2 years with a range of 37 to 79 years. The detailed staging of OSCC in to group's I-IV and high / low stage is shown in Table 4.1.

Table 4.1: Factors contributing to the clinical staging of OSCC.

	Stage	T	N	M
Low stage	I	T1	N0	M0
	II	T2	N0	M0
		T1	N1	M0
	III	T2	N1	M0
		T3	N0	M0
		T3	N1	M0
High stage	IV A	T4	N0, N1	M0
	IV A	Any T	N2, N3	M0
	IV A	Any T	Any N	M1
	IV B	Any T	N3	M0
	IV C	Any T	Any N	M1
T1	Tumour less than 2cm			
T2	Tumour 2-4 cm			
T3	Tumour greater than 4 cm			
T4	Tumour invading surrounding structures			
N0	No nodal spread			
N1	Single node less than 3cm			
N2	Multiple/single nodes 3-6cm			
N3	Nodes greater than 6cm			
M0	No distant metastasis			
M1	Distant metastasis			

Table 4.2: Patient and tumour related factors in the UK cohort of oral cancers.

Study No.	Sex	Age	Site	Clinical stage	Treatment	Path staging	Classified high/low stage	Perineural invasion	Smoking	Heavy alcohol consumption	Early recurrence	Notes
OCS 001 C	M	59	NK	T2N1Mx	surgery	T4N2bMx	High	No	smoker	no	yes	
OCS 003 C	F	72	K	T4N2bMx	surgery	T4N2bMx	High	No	none	no	no	Takes oral snuff
OCS 004 C	F	67	K	T4N0Mx	surgery	T4N0Mx	High	Yes	none	no	yes	Chews betel quid
OCS 006 C	F	53	T	T1N0Mx	surgery	T1N0Mx	Low	No	none	no	no	
OCS 007 C	F	67	K	T2N0Mx	surgery	T4N0Mx	High	No	smoker	no	no	
OCS 008 C	F	65	K	T4N0Mx	radiotherapy	no surgery	High	No	none	no	yes	
OCS 011 C	F	49	T	T1N0Mx	surgery	T1N0Mx	Low	No	smoker	no	no	
OCS 012 C	F	72	T	T4N2aMx	radiotherapy	no surgery	High	No	smoker	yes	yes	
OCS 013 C	M	43	NK	T1N0Mx	surgery	T1N0Mx	Low	No	smoker	no	no	
OCS 014 C	M	46	T	T2N1Mx	surgery	T2N2bMx	High	Yes	smoker	yes	no	
OCS 015 C	M	51	T	T4N0Mx	surgery	T4N0Mx	High	Yes	none	no	yes	
OCS 016 C	M	66	T	T2N1Mx	surgery	T2N2bMx	High	Yes	smoker	yes	yes	
OCS 019 C	F	73	K	T4N0Mx	surgery	T2N0Mx	Low	No	none	no	no	Takes oral snuff
OCS 020 C	M	64	NK	T4N0Mx	surgery	T4N0Mx	High	No	smoker	no	no	
OCS 021 C	M	58	NK	T2N0Mx	surgery	T2N0Mx	Low	No	smoker	yes	no	
OCS 022 C	F	65	T	T2N0Mx	surgery	T2N0Mx	Low	No	smoker	yes	no	
OCS 023 C	M	70	K	T2N0Mx	surgery	T2N2bMx	High	No	smoker	no	no	
OCS 024 C	F	68	K	T4N2bMx	surgery	T2N1Mx	Low	Yes	smoker	no	yes	
OCS 025 C	F	79	NK	T2N0Mx	surgery	T2N0Mx	Low	No	none	no	no	
OCS 026 C	M	37	NK	T2N0Mx	surgery	T1N0Mx	Low	No	smoker	yes	no	
OCS 027 C	M	66	NK	T2N0Mx	surgery	T2N0Mx	Low	No	smoker	no	no	
OCS 029 C	M	52	NK	T4N2Mx	surgery	T4N0Mx	High	No	smoker	yes	no	
OCS 031 C	M	67	NK	T4N0Mx	radiotherapy	no surgery	High	No	smoker	yes	yes	
OCS 032 C	M	60	NK	T2N0Mx	surgery	T2N0Mx	Low	Yes	smoker	no	no	

Site of tumour was classified as tongue (T), non-keratinising mucosa (NK) or keratinising (K.).

Tissue samples of OSCC from Sri Lankan population were obtained from sequential incident cases treated by a single consultant at Kandy General Hospital, Kandy, Sri Lanka. All patient details were verified by a local translator. Biopsy specimens of tissues were harvested during the period January to May 2004. A total of 27 samples yielded RNA that was of suitable quality for Microarray analysis (Table 4.3). The study population included 6 (22%) females and 21 (78%) males with a range of 38 - 82 years and average age of 61.9 years.

Table 4.3: Patient and tumour related factors in the Sri Lankan cohort of oral cancers

Study no.	Sex	Age	Site	Clinical Stage	Treatment	Path staging	Classified high/low stage	Perineural invasion	Smoking	Alcohol consumption	Betel quid chewing & contents	Early recurrence
KC01	M	72	K	T2N0Mx	surgery	T2NxMx	low	No	yes	No	BALT	no
KC02	M	50	NK	T2N0Mx	surgery	T2NxMx	low	No	yes	No	BALT	yes
KC04	M	66	NK	T1N0Mx	surgery	T1NxMx	low	No	no	No	BALT	no
KC07	M	50	K	T2N0Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC09	F	54	T	T1N1Mx	surgery	T1NxMx	low	No	no	No	BALT	no
KC13	M	73	NK	T2N1Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC15	F	40	T	T2N0Mx	surgery	T2NxMx	low	No	no	No	No	no
KC16	F	71	NK	T2N0Mx	surgery	T2NxMx	low	No	yes	No	BALT	no
KC17	M	48	K	T2N0Mx	surgery	T2NxMx	low	No	yes	No	BALT	no
KC19	M	76	NK	T2N1Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC20	M	76	NK	T1N0Mx	surgery	T1NxMx	low	No	yes	yes	BALT	no
KC21	F	56	NK	T2N1Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC24	F	55	NK	T1N1Mx	surgery	T1NxMx	low	No	no	No	BALT	no
KC25	M	74	NK	T2N1Mx	surgery	T2NxMx	low	No	yes	No	BALT	no
KC26	F	85	K	T2N0Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC29	M	55	NK	T2N0Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC31	M	95	NK	T2N0Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC32	M	52	NK	T2N0Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC38	M	51	K	T2N0Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC39	M	64	K	T3N0Mx	surgery	T3NxMx	high	No	yes	yes	BALT	no
KC41	M	82	T	T3N1Mx	surgery	T3NxMx	high	No	yes	No	No	no
KC44	M	42	K	T3N0Mx	surgery	T3NxMx	high	No	no	No	BALT	no
KC45	M	58	T	T2N0Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC46	M	38	T	T2N1Mx	surgery	T2NxMx	low	No	yes	yes	BALT	no
KC47	M	50	NK	T2N0Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC51	M	78	NK	T2N0Mx	surgery	T2NxMx	low	No	no	No	BALT	no
KC53	M	60	NK	T3N0Mx	surgery	T3NxMx	low	No	yes	No	BALT	no

Site of tumour was classified as tongue (T), non-keratinising mucosa (NK) or keratinising (K.) Betel quid chewing contents was classified as: Betel nut (B); areca nut (A); lime (L); and Tobacco (T.)

Normal samples consist of 11 normal oral mucosa specimens (seven samples from UK & four samples from Sri Lankan population). All normal samples were harvested from patients under the age of 30 years (undergoing either a minor surgery in the mouth for a non-neoplastic condition or a dental extraction) All were non-smokers; did not chew betel quid and did not consume exceeding the national recommended weekly gender allowance of alcohol. Normal samples had no personal history of cancer and had no first degree relatives with a history of cancer. A total of 8 samples from the two countries yielded RNA that was of suitable quality for Microarray analysis (Table 4.4).

Table 4.4: Site of origin of normal oral mucosa samples from both population groups

Specimen no.	Site
M1-2	NK
M3	Nk
M4	K
M2	K
V5	T

A

Specimen no.	Site
KN2	K
KN3	NK
KN5	T

B

A- Represent UK normal group; B- Represent Sri Lankan normal group. tongue (T), non-keratinising mucosa (NK) or keratinising (K.)

After surgical extraction, samples were immediately wrapped in foil and dipped in liquid nitrogen for 20 seconds to freeze thoroughly. Then samples were transferred to labelled cryo-vials and returned to liquid nitrogen.

Using H&E-stained sections from snap-frozen tissue, all tumours were histopathologically confirmed as squamous cell carcinoma and staged according to the AJCC/UICC TNM Classification system. Tumours were pathologically designated as high/ low stage, while the patients who experienced no early

recurrence for at least 12 months post-operatively were scored as "non-recurrence cases".

To standardize stromal cell contribution, each selected tumour was confirmed histopathologically to contain ≥60% tumour tissues and <10% necrotic debris by analysis of corresponding H&E sections.

4.3.2 RNA extraction / cDNA synthesis, processing, labeling, and hybridization to microarray chip

RNA was extracted using TRIZOL RNA extraction method. The yielded RNA from the samples was quantified using Nano-spectrophotometer and stored at -80°C. The labelled cDNA was then hybridised to the *Affymetrix® Focus and U133 plus 2.0* Gene Chips (Affymetrix® Santa Clara, US). For full details of procedures for this section, see supplementary file 1b.

4.3.3 Data quality control, pre-processing and analysis

Data analysis was performed using Bioconductor packages (Gentleman et al., 2004), http://www.bioconductor.org/). The *simpleaffy* package was used to assess average background, Scale factors, Number of genes called present (% Present), 3' to 5' ratios (http://www.affymetrix.com/support/downloads/manuals/data_analysis_fundamentals_manual.pdf), 6 chips (3tumour samples and 3 normal samples) were observed of poor quality and are filtered out.

Normalization was performed using *gcRMA* algorithm from Bioconductor to extract gene expression values (two independent normalization procedures were conducted per study cohort; normal and tumour group).

Tumour samples from both UK and Sri Lanka were analysed using the Affymetrix Human Genome Focus array. The normal samples from both countries were analysed using the Affymetrix Human Genome U133 Plus2.0 Array. The probesets that identically represented on both U133 Plus2.0 array and Focus GeneChip were used for comparative analysis.

A number of pre-processing steps including background adjustment, probe filtering, normalization and quality control checks were performed using the R package, Simpleaffy (www.bioconductor.org).

It is important to mention here that the author who composed the current thesis started the work analysis of this chapter from this section, while the previous steps of methods were carried out by another author (Lopes, 2007).

4.3.4 Quality checks of microarray GeneChip results (QC)

To obtain accurate results from the subsequent analysis, Quality Checks were performed to evaluate whether to include results from each Gene chips using the *simpleaffy* package (http://bioinformatics.picr.man.ac.uk/simpleaffy/), a part of the Bioconductor project (http://www.bioconductor.org/).

The Microarray data quality reporting shown here is generated using the quality control function of the SimplyAffy software package from Affymetrix® using R programming software. OSCC from both the UK and Sri Lanka were analysed using the Affymetrix® Human Genome Focus Array that represents over 8500 verified human sequences from the NCBI RefSeq database. The normal oral mucosa specimens were analysed using the Affymetrix® Human Genome U133 Plus2.0 Array that can examine over 47,000 transcripts.

The simpleaffy function, QC, generates the most commonly used metrics which all their values are parameters computed for/from the MAS 5.0 algorithm. The standard recommendations from Affymetrix are as following:

4.3.4.1 Average background

This should be similar across all chips, if chips have significantly different average backgrounds this could be for a number of reasons. It might simply that the overall signal from the array is greater, perhaps because different amounts of cRNA were present in the hybridisation cocktails, or because the hybridisation was more efficient in one of the reactions, incorporating more label, and producing brighter chip. The accepted upper limit for background signal equal 100.

4.3.4.2 Scale factors

The default normalisation used by MAS 5.0 (and many other algorithms) makes assumption that gene expression does not change significantly for most of transcripts in an experiment. The amount of scaling applied is represented by the 'scale factor', which, therefore, provides measurement of the overall expression level for an array, and (assuming all else remains constant), a reflection of how much labelled RNA is hybridised to the chip. Large variations in scale factors signal cases where the normalisation assumptions are likely to fail due to issues with sample quality or amount of starting material. Alternatively, they might occur if there have been significant issues with RNA extraction, labelling, scanning or array manufacture. In order to successfully compare data produced using different chips, Affymetrix recommend that their scale factors should be within 3-fold of one another.

4.3.4.3 Number of genes called present (% Present)

Present/Marginal/Absent calls are generated by looking at the difference between PM and MM values for each probe pair in a probeset. Probesets are flagged "Marginal" or "Absent" when the PM values for that probeset are not considered significantly above the MM probes. As with scale factors, large differences between the numbers of genes called present on different arrays can occur when varying amounts of labelled RNA have been successfully hybridized to the chips. This can occur for similar reasons (differences in array processing pipelines, variations in the amount of starting material, etc.). The 'percent present call simply represents the percentage of probesets called Present on an array and should be between 35-55%. As with "Scale Factors", significant *variations* in % Present call across the arrays in a study should be treated with caution.

4.3.4.4 3' to 5' ratios

Most cell types ubiquitously express β-actin and GAPDH. These are relatively long genes, and most Affymetrix chips contain separate probesets targeting the 5', mid and 3' regions of their transcripts. By comparing signal number from the 3' probeset to either the mid or 5' probesets, it is possible to obtain a quality measure of the

RNA hybridised to the chip. If the ratios are high, this indicates the presence of truncated transcripts. This may occur if the *in vitro* transcription step has not performed well or if there is general degradation of the RNA. Hence, the ratio of the 3' and 5' signal gives a measure of RNA quality. GAPDH is the smaller of the two genes and the 3':5' ratio should always be at or around 1.

The β-Actin gene is a housekeeping gene which commonly used to determine array quality by measuring the 3'/5' ratio of hybridized labelled target RNA to control probes on the GeneChip array. This ratio for β-Actin is assumed representative of the ratio for other probesets on the array and a lower value ratio is generally considered desirable but no absolute limits have been determined. Affymetrix suggest that a β-actin 3':5' ratio of less than 3 is acceptable.

4.3.5 Differential gene expression analyses

Analysing gene expression of the study samples was carried out into seven specific categories of comparisons: Tumour vs. normal of the common genes exist in both UK and Sri Lankan samples; Tumour vs. normal of UK samples; Tumour vs. normal of Sri Lankan samples; Normal UK vs. normal Sri Lankan samples; UK tumour samples vs. Sri Lankan tumour samples; low vs. high/late stage tumours for both UK and Sri Lankan samples; Tongue Site vs. other Sites tumours for both UK and Sri Lankan samples; Invasive (refers to cancer spreading to the space surrounding the nerve) vs. non-invasive tumours of UK samples.

To examine differential gene expression and to extract meaningful informative results from the seven comparisons, data analysis takes two main forms. The first is 'unsupervised' algorithms where data is sorted and gene expression differences are derived based on inherent differences within the dataset. Unsupervised clustering was achieved and heatmaps generated using Pearson correlation hierarchical clustering in R. Supervised analysis was carried out using SAM (Tusher et al., 2001) in *siggenes* package available from Bioconductor. Heatmaps were generated using *gplots* package (Gentleman et al., 2004). Fold change value for "upregulated /overexpressed" and "downregulated/under expressed" genes were obtained Using Wilcoxon Rank Statistics function in siggenes R software.

The supervised methodology has the advantage of skill to define differences between known groups more easily than the unsupervised but it is possible to overlook biologically important gene changes because they do not fall in to the final significantly changed gene group.

This is a problem of statistical significance versus biological significance and attempts to define both will support all Microarray analysis.

List of genes were further narrowed by trimming out the low magnitude of the fold changes and / or reduce the FDR value of the significant genes.

4.3.6 Biological interpretation of the expressed genes

Functional annotation analysis of the differentially expressed genes was performed using Ingenuity Pathway Analysis (Ingenuity Systems, www.Ingenuity.com) tool which relies on a knowledge database of selected functional and regulatory interactions extracted from the literature and provides integrated graphical representation of the biological relationships between genes and gene products considering both up- and down-regulated genes from each comparison analysis.

IPA was used to highlight the most affected Molecular and Cellular functions, canonical pathways, Diseases and Disorders, Transcriptional Regulators, and most significant pre-generated networks associated with the uploaded group of genes including those related to OSCC development.

The master network was constructed by connecting the selected genes involved in OSCC, HNSCC, and SCC using fisher exact test. The high connectivity nodes are likely to reflect their ability to regulate an important number of genes in the master network and potentially to control the gene expression pattern / modulations identified tumour signature of that specific comparison.

It’s well known that regulation of human genes depends on sets of transcription factors rather one individual factor. For this reason, available tool from IPA system was used to identify potential transcriptional regulators. Transcriptional Regulator Analysis tool from IPA system highlighted the activation and inhibition of certain

genes acting as transcriptional regulators and have a major contribution in the tumorigenic process.

As explained in previous chapter, IPA cannot map pseudo genes, hypothetical genes, ESTs and ambiguous (mapping to multiple) genes as well as duplicates of multiple IDs snapped in one molecule or multiple molecules snapped in one ID, re-arrangement of unmapped genes was carried out by converting them using David conversion tool (Huang et al., 2009a, Huang et al., 2009b) to obtain more unmapped genes to the subsequent analysis where in some of them are substantial key players contributing to the tumorigenic process and in the master networks of the analysis.

4.3.7 Quantitative real-time PCR analysis (validation of gene expression level)

Three genes downregulated in both UK and Sri Lankan tumours relative to their normal samples including Clusterin (CLU) T-cell differentiation protein (MAL), and Protein tyrosine phosphatase, non-receptor-type, 13 (PTP313/FAP1) were selected for further investigation and qPCR validation Using a combined thermal cycler and fluorescence detector (Perkin-El mer Applied Biosystems 7700 sequence detector). Samples used in the qPCR include all Sri Lankan and UK normal oral mucosa samples. These were compared to 12 of each UK and Sri Lankan samples.

The QPCR data presented here were repeated and confirmed over four separate rounds of experimental determination consistently confirming the expression array findings that, these three genes are downregulated in OSCC compared to normal in both UK and Sri Lankan samples.

4.4 Results

4.4.1 Study populations

The OSCC cases in both UK and Sri Lankan populations were older than their controls. Compared to controls, OSCC cases were more reliable to alcoholic consumption, and current smokers. Normal control were all under the age of 30 years, non-smokers, did not consume exceeding the national recommended weekly

gender allowance of alcohol, and had no personal history of cancer and had no first degree relatives with a history of cancer.

Tumour population included 34 males (13 UK and 21 Sri Lanka) and 17 females (11 UK and 6 Sri Lanka). The UK average age was 61.2 years with a range of 37 to 79 years and Sri Lankan average age of 61.9 years with a range of 38 to 82 years. A significant male predominance of disease in both countries was observed; Table 4.2 and Table 4.3. These two groups is representative of the general population of patients with OSCCs in each country. All tumour samples of both populations were from oral cavity; including 7 (29%) UK patients presented with cancers of keratinized mucosa, 10 (42%) with cancers of non-keratinized mucosa and 7 (29%) with cancer of specialised mucosa of the tongue. Eleven patients (46%) presented with low stage disease and 13 (54%) with high stage. Eight (33%) patients had loco-regional lymph node metastasis at presentation and six (25%) samples displayed perineural invasion at histopathology. Eight (33%) patients developed early recurrence (defined as recurrence of OSCC at the same anatomical site or in the neck nodes within 12 months of completion of definitive therapy). Sri Lankan tumour samples consist of 7 (26%) patients presented with cancer of the Keratinising mucosa, 15 (55%) with cancers of the non-keratinising mucosa and 5 (19%) with cancer of the specialised mucosa of the tongue. Three patients (11%) presented with early invasive disease (perineural invasion at histopathology), 21 (78%) with low stage disease and 3 (11%) with high stage. Importantly, although eight (30%) patients had loco-regional lymph node enlargement at presentation, none of these patients had a neck dissection as the nodal enlargement is regarded as insignificant and frequently resolves after excision of the primary tumour. None of Sri Lankan samples displayed perineural invasion at histopathology and only one (4%) patient developed early recurrence (defined as recurrence of OSCC at the same anatomical site or in the neck nodes within 12 months of completion of definitive therapy).

4.4.2 Report of microarray quality control

To obtain accurate results for the subsequent analysis, Quality Checks were conducted to evaluate whether to include results from each gene chips using the simpleaffy® package (http://bioinformatics.picr.man.ac.uk/simpleaffy/) which is part of the Bio conductor Affymetrix that already mentioned in materials and methods section.

(http://www.affymetrix.com/support/downloads/manuals/data_analysis_fundamentals_manual.pdf). In order to assess the quality of data generated in this experiment, four out of the five metrics mentioned will be considered as following:

4.4.2.1 Average background

The average background for each chip shows a considerable amount of variation. This is represented in Figure 4.1and Figure 4.2 by colouring the average background values for all chips red. Specifically the average background for OCS21, OCS19, KC46, and KC31 tumour samples along with V6_1, V2A_1, and KN1 normal samples which significantly exceed the accepted upper limit for background signal equal 100. Although this may indicate a problem with these samples all other QC measures correlate well with those of its peers (as seen below).

4.4.2.2 Scale factor

To achieve overall experimental consistency the scaling factor for an array should not exceed three. The scale factors for OCS23 and OCS21 UK tumours are greater than 3-fold away from the average scale factor for all samples (Figure 4.1). This suggests that the overall intensity of these samples was lower, resulting in higher scale factors. This may result from less RNA present in these samples.

4.4.2.3 Number of genes called present

The number of genes called present (% present calls) shows a broad spread in values across the whole experiment (25.6-57%) for tumours and (17-42) for normal samples (Figure 4.1and Figure 4.2). After removal of the six bad quality chips, a good general agreement between samples was achieved in each group and between each experimental condition with new % present calls 36-60% for tumour and 30-42% for normal samples, very close to the ideal range between 35-55% and so judged within acceptable limit (Figure 4.1and Figure 4.2). The qc plot has coloured these numbers red to indicate a spread of more than 10% across the whole experiment. Considering "percent" present and scale factor together, OCS19, OCS21, and OCS23 samples that have lower % present calls also have higher scale factors. These factors certainly warrant further attention. Higher scale factors with lower "percent" present calls may indicate that less RNA has been hybridised to these chips. Accordingly, the above three samples were excluded from analysis as mentioned previously.

4.4.2.4 3' to 5' ratios for β-actin and GAPDH

At first glance the results for the 3':5' ratios for the tumour samples (Figure 4.1) appear less than satisfactory, with many of the GAPDH and β-actin ratios flagged. This may indicate that these samples contain degraded RNA. However, considering other criteria all together will give the final judgment specifically after removing the bad quality chips (Figure 4.1). All normal samples show both probes ratios within satisfactory limit before and after trimming of the bad quality chips (Figure 4.2).

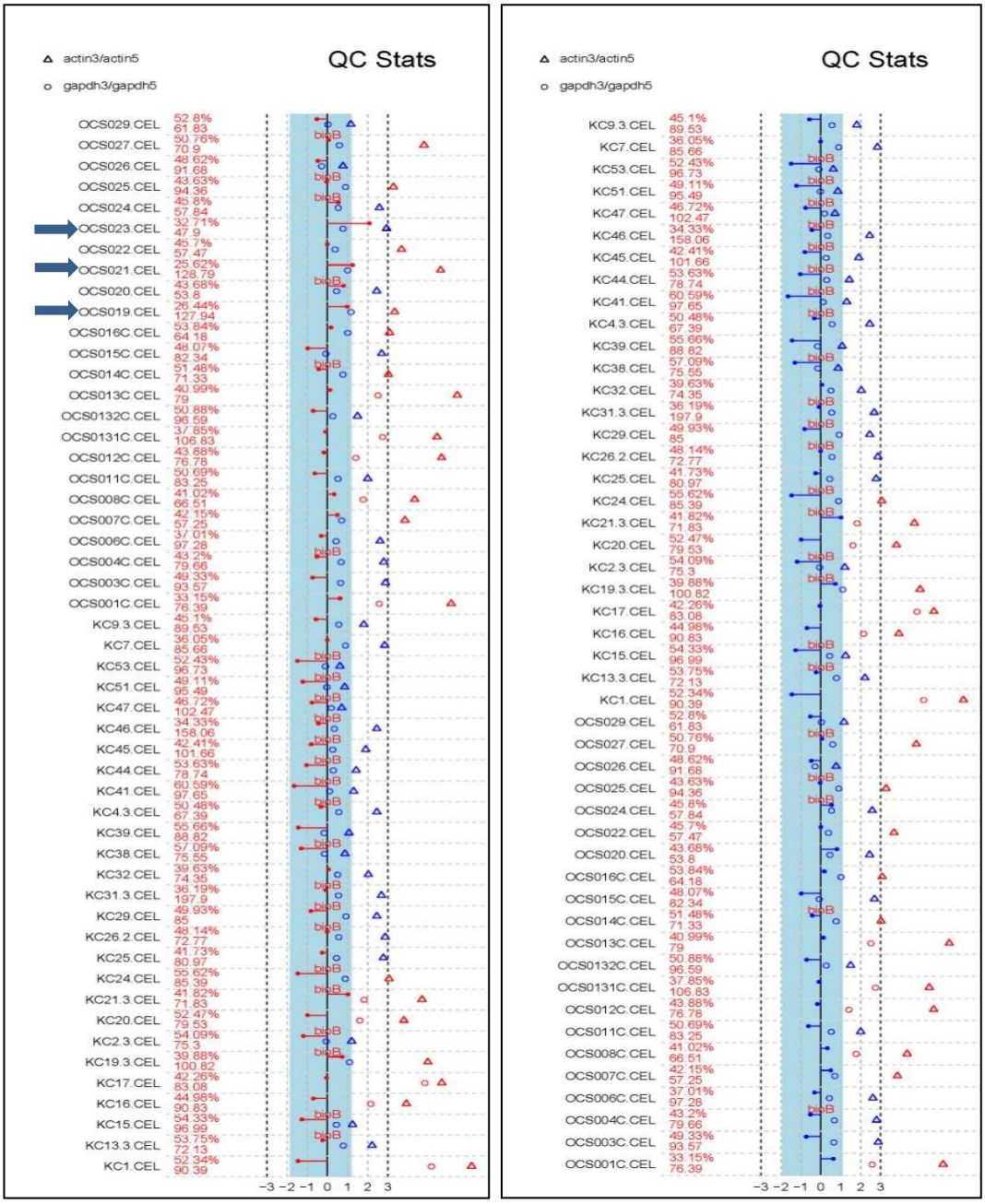


Figure 4.1:Simpleaffy plots displaying the QC of 51 and 48 tumour samples from UK and Sri Lankan population groups before (Left) and after (Right) removal of bad quality microarray chips respectively.

Each row shows the %present, average background, scale factors and GAPDH / β-actin ratios for an individual chip. GAPDH 3':5' values are plotted as circles. According to Affymetrix they should be about 1. GAPDH values that are considered potential outliers (ratio > 1.25) are coloured red, otherwise they are blue. β -actin, 3':5' ratios are plotted as triangles. 3':5' ratios recommended to be below 3; values below 3 are coloured blue, those above, red. The blue stripe in the image represents the range where scale factors are within 3-fold of the mean for all chips. Scale factors are plotted as a line from the centre line of the image. A line to the left corresponds to a down-scaling, to the right, to an up-scaling. If any scale factors fall outside this '3-fold region', they are all coloured red, otherwise they are blue. %present and average background, are listed to left of the figure.

4.4.2.5 QC conclusions

For normal and tumour group samples, 6 chips (OCS19, OCS21, OCS23, V6_1, V2A_1, and KN1; half of them related to tumour samples and the other half related to normal samples) were recognized of poor quality and are discounted from any further analysis based on these QC data. This was because OCS21, OCS23 had higher scale factors than any of other groups of samples and lower % present calls. These observations suggested a possibility of degraded RNA presents within the samples or that the amplification protocol had not worked successfully for the samples, resulting in less RNA hybridised to the chip. Furthermore, other chips (OCS19, V6_1, V2A_1, and KN1) were excluded as it had lower "percent" present call and average background which significantly exceed the accepted upper limit for background signal. The QC plots for the new data set are shown in Figure 4.1and Figure 4.2.

Although a spread in the average background can still observed, because generally, the remaining QC metrics fall within the current guidelines (specifically the scale factors for all chips are within 3-fold of one another), this was not considered significant enough to reject arrays. Similarly, tumour samples have higher β-Actin and GAPDH ratios, however, this threshold is fairly stringent, and in the context of the other QC metrics, it was decided not to fail these arrays. The final decision was to include data from 48 tumours and 8 normal arrays and therefore used in the subsequent analyses.

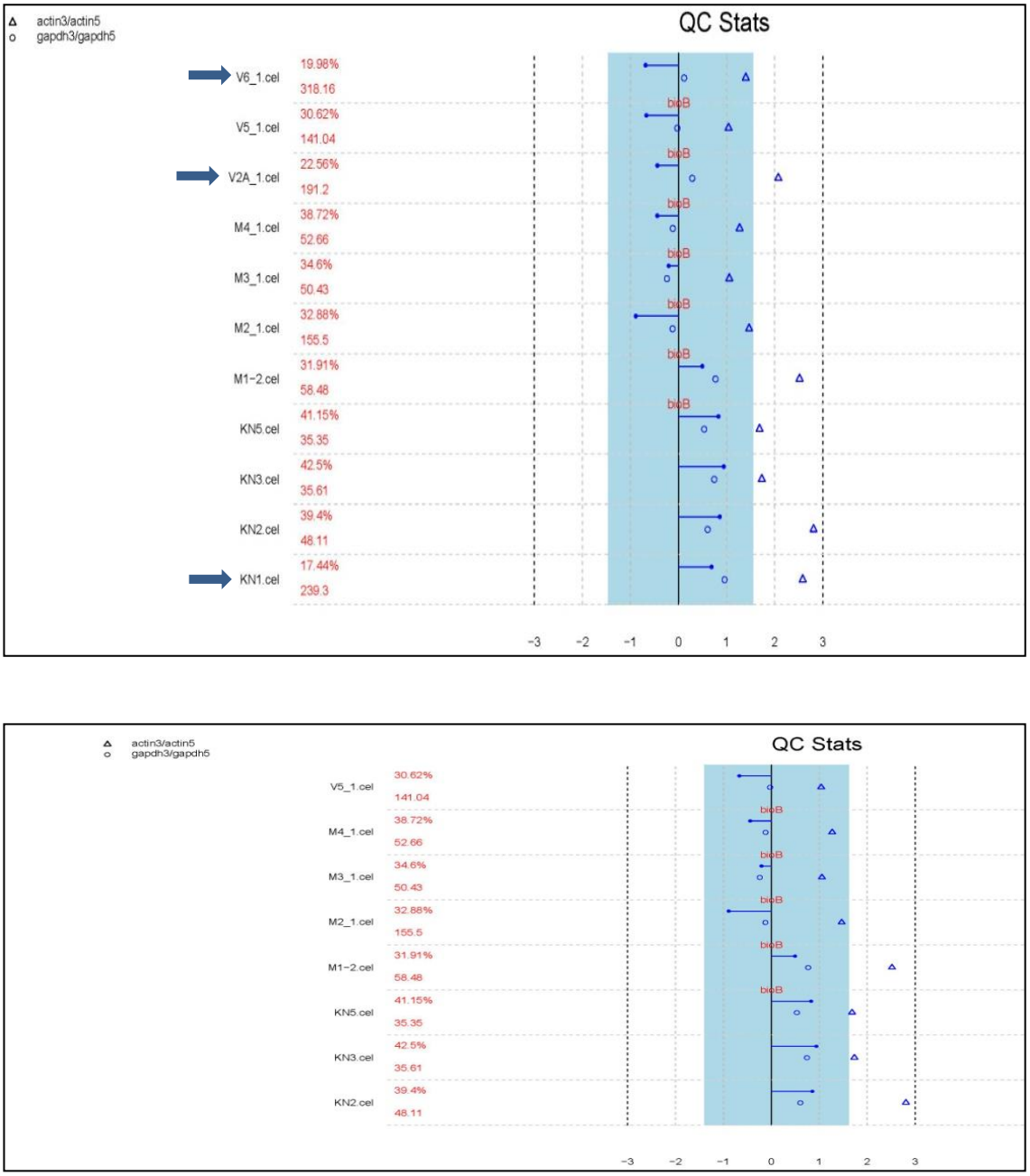


Figure 4.2: Simpleaffy plot displaying the QC of 11 and 8 normal samples from UK and Sri Lankan population groups before (top) and after (bottom) removal of bad quality microarray chips respectively.

Each row shows the %present, average background, scale factors and GAPDH / β-actin ratios for an individual chip. GAPDH 3':5' values are plotted as circles. According to Affymetrix they should be about 1. GAPDH values that are considered potential outliers (ratio > 1.25) are coloured red, otherwise they are blue. β -actin, 3':5' ratios are plotted as triangles. 3':5' ratios recommended to be below 3; values below 3 are coloured blue, those above, red. The blue stripe in the image represents the range where scale factors are within 3-fold of the mean for all chips. Scale factors are plotted as a line from the centre line of the image. A line to the left corresponds to a down-scaling, to the right, to an up-scaling. If any scale factors fall outside this '3-fold region', they are all coloured red, otherwise they are blue. %present and average background, are listed to left of the figure.

4.4.3 Filtration and pre-processing data

Following normalization, the huge number of probe sets was reduced by *Nonspecific Filtering* using the function *rowSds/ genefilter* in R statistical software package available on Bioconductor (www.bioconductor.org/packages/release/bioc/vignettes/genefilter/inst/doc/howtogenefilter.pdf). By applying the filtration method, not only a reduction in the non-useful un-diagnostic probe sets was observed, but improvement in the sample discrimination ability of the subsequent analysis.

4.4.4 Gene expression profile of OSCC versus normal oral mucosa from both population groups

Following data normalization, probe sets filtering, and removal of microarray chips having poor quality, unsupervised analysis of all tumour and normal samples from both populations altogether were performed using Pearson correlation comparison. The normal samples clearly clustered together away from the tumour samples. Moreover, the normal samples of each population were separated from each other as seen clearly from the hierarchical clustering (Figure 4.3). Examination of tumour groupings in the hierarchical clustering showed the majority but not all tumours that associated with metastasis and recurrence tend to group together. A supervised analysis of all tumour and normal samples from both populations was carried out using significance analysis of microarrays (SAM) with a false discovery rate (FDR) ≤ 0.001 and fold change above or below 1.5 (Figure 4.4).

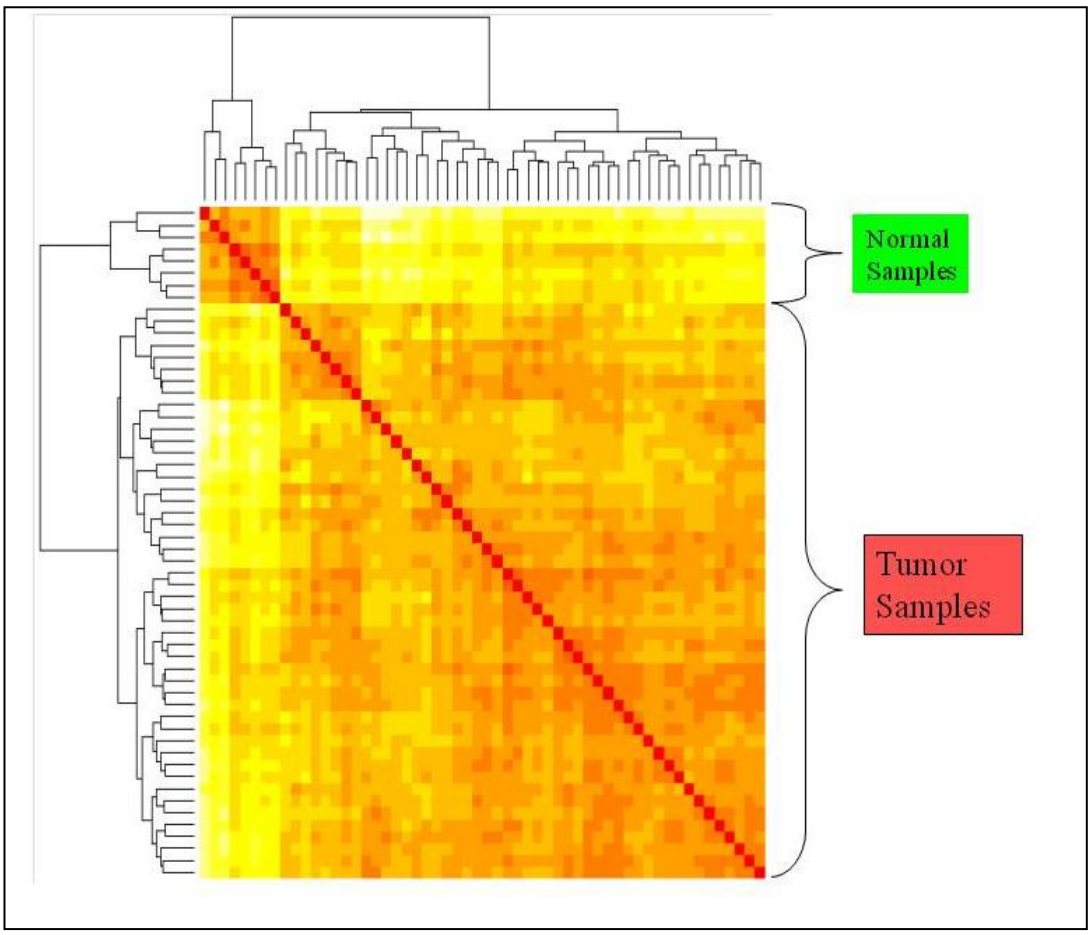


Figure 4.3: Unsupervised Pearson correlation hierarchical clustering of all normalized genes clearly discriminating normal oral mucosal samples (N=8) from OSCC samples (N=48) of both UK and Sri Lankan populations.

Heatmap represents samples distribution according to the mean-centred expression of genes. Expression (red= 1, yellow= lower than 1, white= lowest value than 1).

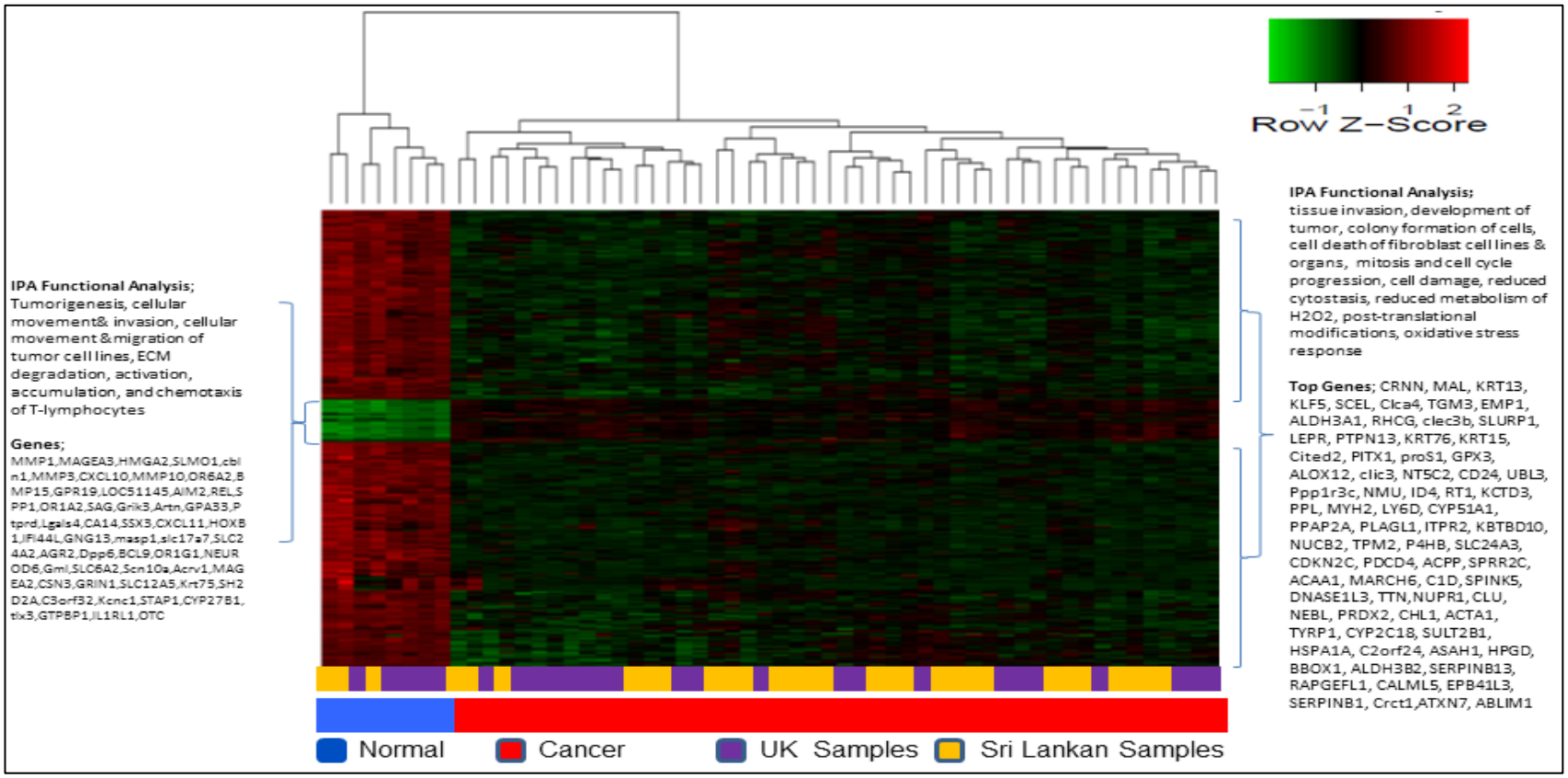


Figure 4.4: Hierarchical clustering of the 1222 most differentially expressed genes separating UK and Sri Lankan normal oral mucosa (N=8) and patient OSCC specimens (N=48) using SAM supervised analysis at FDR≤0.001.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. Gene expression values were either above or below fold change magnitudes equal ± 1.5.

4.4.5 Gene expression profile of tumour versus normal samples from UK

Using both supervised and unsupervised analysis for this comparison as mentioned in materials and methods section, both analysis separated normal (N=5) from tumour (N=21) samples clearly. The supervised analysis of the 26 samples was undertaken under different levels of false discovery rates using SAM. Supervising the analysis in two groups according to sample type; OSCC and normal mucosa identifying 925, 896, 819, and 185 differentially expressed genes with a fold change cut-off value equal to ±1.5 and FDR levels of ≤0.005, ≤0.001, ≤ 0.0001, and = 0.00 respectively. At FDR level ≤ 0.005, 60 genes were upregulated and 866 genes were downregulated in OSCC compared to normal mucosa from UK individuals (Table 4.5 and Supplementary file 7). Altering the stringency of the SAM analysis to accept only FDR level ≤ 0.0001 reduced the profile to 30 upregulated and 779 downregulated genes (Figure 4.5). At FDR level=0.0 a set of 185 genes appeared all downregulated within OSCCs in relative to their normal comparators with a fold change magnitude ≥2 (Figure 4.6).

Table 4.5: Significantly differentially expressed genes discriminating UK tumour from normal samples. *

Gene Symbol	Gene ID	Fold Change	Gene Symbol	Gene ID	Fold Change
Genes higher expressed in OSCCC			**Genes lower expressed in OSCC**		
MMP1	204475_at	4.4	CRNN	220090_at	-5.7
MMP3	205828_at	3.2	MAL	204777_s_at	-5.3
HMGA2	208025_s_at	2.9	KRT13	207935_s_at	-5.1
BMP15	221332_at	2.3	KLF5	209211_at	-4.9
REL	206035_at	2.3	SCEL	206884_s_at	-4.6
POSTN	210809_s_at	2.2	CLCA4	220026_at	-4.5
CXCL10	204533_at	2.2	TGM3	206004_at	-4.4
MMP10	205680_at	2.2	EMP1	201324_at	-4.2
SPP1	209875_s_at	2	CLEC3B	205200_at	-4.2
GRIN1	205914_s_at	1.9	RHCG	219554_at	-4.2
LAMC2	202267_at	1.9	CD24	208650_s_at	-3.7
IFI6	204415_at	1.7	SERPINB13	211362_s_at	-3
SLC6A2	215715_at	1.7	CDKN2C	204159_at	-2.9
COL11A1	37892_at	1.6	FGFR2	208228_s_at	-2.8
CXCL11	210163_at	1.5	CLU	208792_s_at	-2.8
PTGS2	204748_at	1.5	CRABP2	202575_at	-2.7
TNC	201645_at	1.5	EHF	219850_s_at	-2.6
			HSP90AB1	214359_s_at	-2.5
			YES1	202933_s_at	-2.5
			CDKN1B	209112_at	-2.5
			IGFBP2	202718_at	-2.5
			RPS6KB1	204171_at	-2.4
			SMAD4	202527_s_at	-2.3
			JUP	201015_s_at	-2.3
			VEGFA	210512_s_at	-2.2
			FGFR3	204379_s_at	-2.2
			HRAS	212983_at	-2.1
			CRYAB	209283_at	-2.1
			DUSP1	201041_s_at	-2.1
			RAD21	200608_s_at	-2
			RAF1	201244_s_at	-1.9
			HSPA1A//HSPA1B	202581_at	-1.9
			KRT19	201650_at	-1.9
			CDH1	201131_s_at	-1.8
			INS-IGF2	202409_at	-1.8
			WNT5A	205990_s_at	-1.8
			ID2	201565_s_at	-1.8
			PTPRC	212588_at	-1.8
			MAPK9	203218_at	-1.7
			MAPK13	210058_at	-1.7
			TNFSF10	202688_at	-1.7
			KRT5	201820_at	-1.7
			EZH2	203358_s_at	-1.7
			TP53	201746_at	-1.7
			RB1	203132_at	-1.5

* FDR value < 0.005.The complete lists of upregulated and downregulated genes in OSCCs were presented in Supplementary file 7.

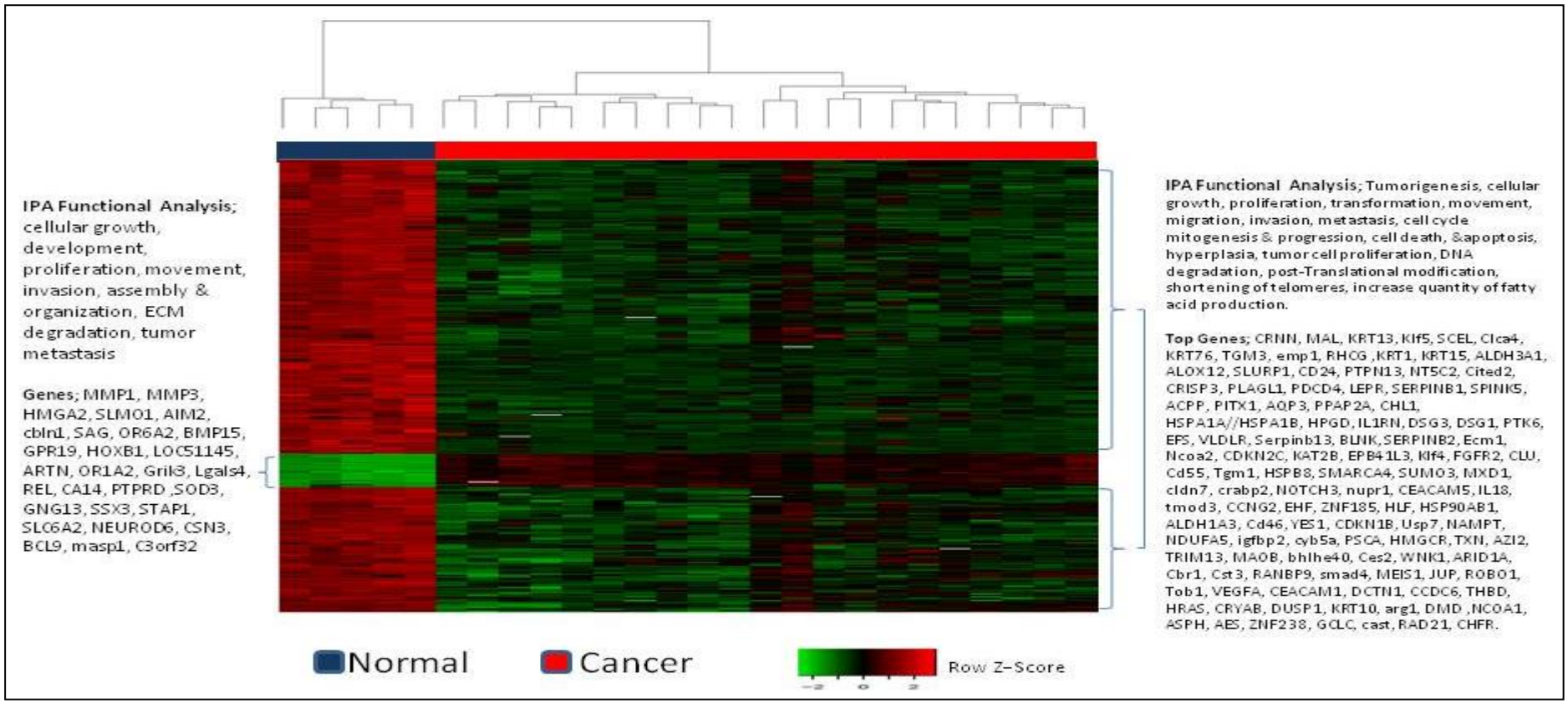


Figure 4.5: Hierarchical clustering of the 896 most differentially expressed genes separating 5 UK normal oral mucosal and 21 patient OSCC specimens using SAM supervised analysis at FDR≤0.001.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. Gene expression values were either above or below fold change magnitudes equal ± 1.5.

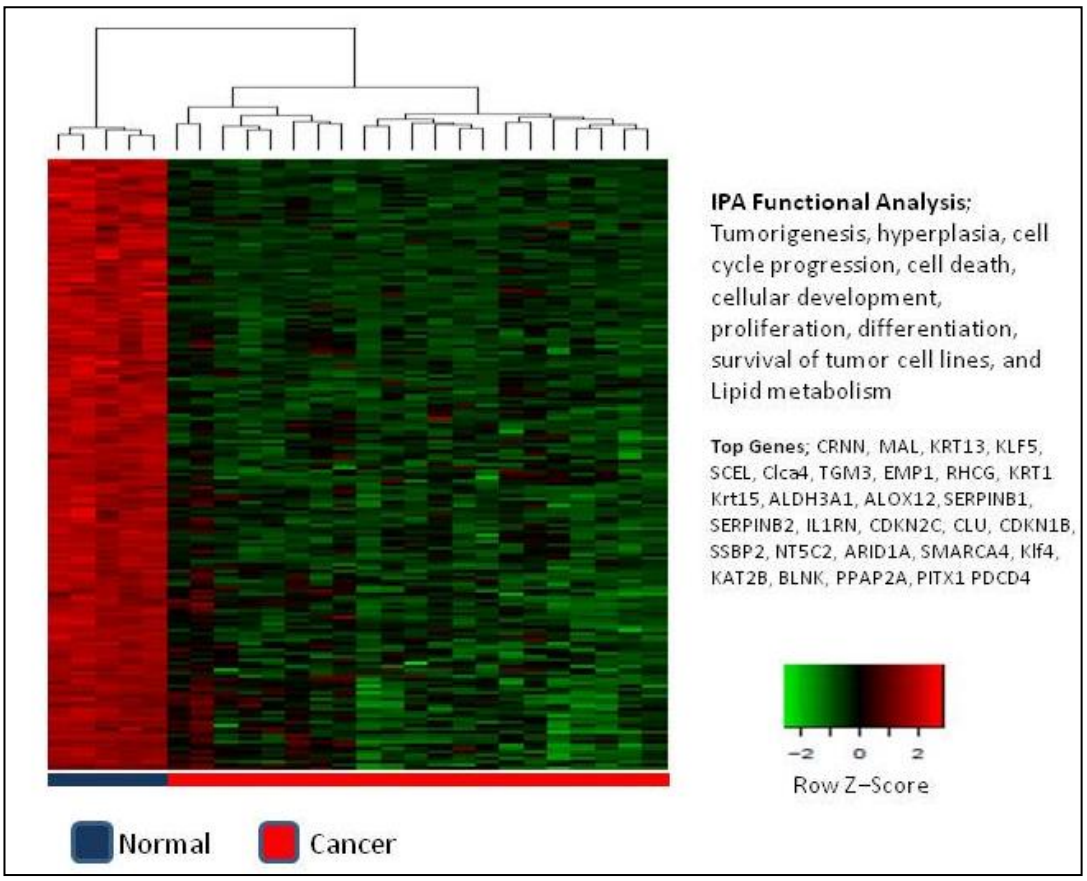


Figure 4.6: Hierarchical clustering of the 185 most differentially expressed genes separating 5 UK normal oral mucosal and 21 OSCC specimens using SAM supervised analysis at FDR=0.00

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All genes had fold change magnitudes less than 2, relative to normal samples.

Among the differentially expressed genes of UK tumour samples, several potential diagnostic biomarkers were identified which have previously been reported in OSCC and HNSCC (Figure 4.4). Of high concern, several of the above biomarkers are uniquely expressed in UK but not Sri Lankan tumour samples with their up- or downregulation reported substantially associated with tumour invasion, metastasis, and recurrence. This includes upregulation of MMP3, COX2, LAMC2, POSTN, COL11A1, and TNC along with downregulation of DUSP1 and SLPI (Wen et al., 2011, Tomasi et al., 2010, Jones and Jones, 2000, Ziober et al., 2006, Schmalbach et al., 2004, Kudo et al., 2007, Yan and Shao, 2006, Siriwardena et al., 2006, Shao et al., 2004, Bao et al., 2004, Ouyang et al., 2009, Sasaki et al., 2003, Sasaki et al.,

2001b, Sasaki et al., 2002, Sasaki et al., 2001a, Hensen et al., 2008, Mohan and Epstein, 2003, Wang, 2005, Zhi et al., 2003, Kosmehl et al., 1999, Malinoff et al., 1984, Hintermann and Quaranta, 2004, Liotta et al., 1988, Jeng et al., 2000, Grosch et al., 2001, Dannenberg et al., 2001, Brown et al., 2001, Nelson et al., 2000, Masferrer et al., 2000, Mohammed et al., 1999, Itoh et al., 2003, Gallo et al., 2002, Wiegand et al., 2005, Ye et al., 1996, Gallo et al., 2001).

Three members of Matrix Metalloproteinase family (MMP1, MMP3, and MMP10) are among the genes most significant upregulated, which may contribute to the aggressive nature of OSCC in UK samples. MMPs are well known as a large family of proteinases which remodel extracellular matrix components and play important role in tumour development, invasion and metastasis (Fingleton, 2006, Rosenthal and Matrisian, 2006, Moilanen et al., 2003). Several members of MMP family considered as important biomarkers for diagnosis and prognosis as well as potential therapeutic targets for many types of cancers, including HNSCC (Jordan et al., 2004).

Another observation in UK tumours was upregulation of genes related with tumorigenic transformation including; GREM1, downregulation of E-Cadherin, several members of Tight Junctions including CLDN1, CLDN4, CLDN5, and CLDN7. This result reflected the activation of the Hedgehog pathway which leads to an increase in Snail protein expression and a decrease in E-Cadherin and Tight Junctions (Yook et al., 2005). Hedgehog signalling appears a crucial regulator of angiogenesis and thus metastasis (Velcheti, 2007).

Central to many of the pathways (Figure 4.7) is PTGS2 (COX2); this gene has been frequently associated with OSCC and HNSCC in previous publications (Hunter et al., 2005, Chang et al., 2004, Banerjee et al., 2002, Sudbo and Reith, 2003, Nathan et al., 2001, Atula et al., 2006, Lee et al., 2002, Chan et al., 1999, Mohan and Epstein, 2003, Ye et al., 2008, Tsai et al., 2004, Wang, 2005) and suggested responsibility for the prostanoid biosynthesis involved in inflammation and mitogenesis. COX2 enzyme linked to tumour angiogenesis, prognosis, recurrence, metastasis and micro vessel density as well as postoperative disease-free survival of patients with OSCCs after surgical removal (Gallo et al., 2001, Itoh et al., 2003). Moreover, COX2 protein

expression tended higher in those tumours in advanced stage with vascularization and loss of differentiation (Gallo et al., 2002). Several theories explained the role of COX2 in a wide range of tissues including oral cancer and OSCC pre-malignancy suggesting that the effects of COX2 on tumour development and progression are most likely multifactorial and include different pathways including: Inhibition of Apoptosis, Stimulation of Angiogenesis, Immunosuppression, Enhanced Invasiveness, modulating xenobiotic metabolism, and Increased Mutagenesis (Jeng et al., 2000, Grosch et al., 2001, Dannenberg et al., 2001, Brown et al., 2001, Nelson et al., 2000, Masferrer et al., 2000). The pattern of gene expression profile in tumour samples of UK, suggesting the role of COX2 lies mainly on "Stimulation of Angiogenesis", "Inhibition of Apoptosis" by increased expression of anti-apoptotic proteins and decreased expression of pro-apoptotic proteins, and "Enhanced Invasiveness" by increased expression of various matrix metalloproteinases, a family of degradative enzymes and would play a key role in the cancer invasiveness and metastasis.

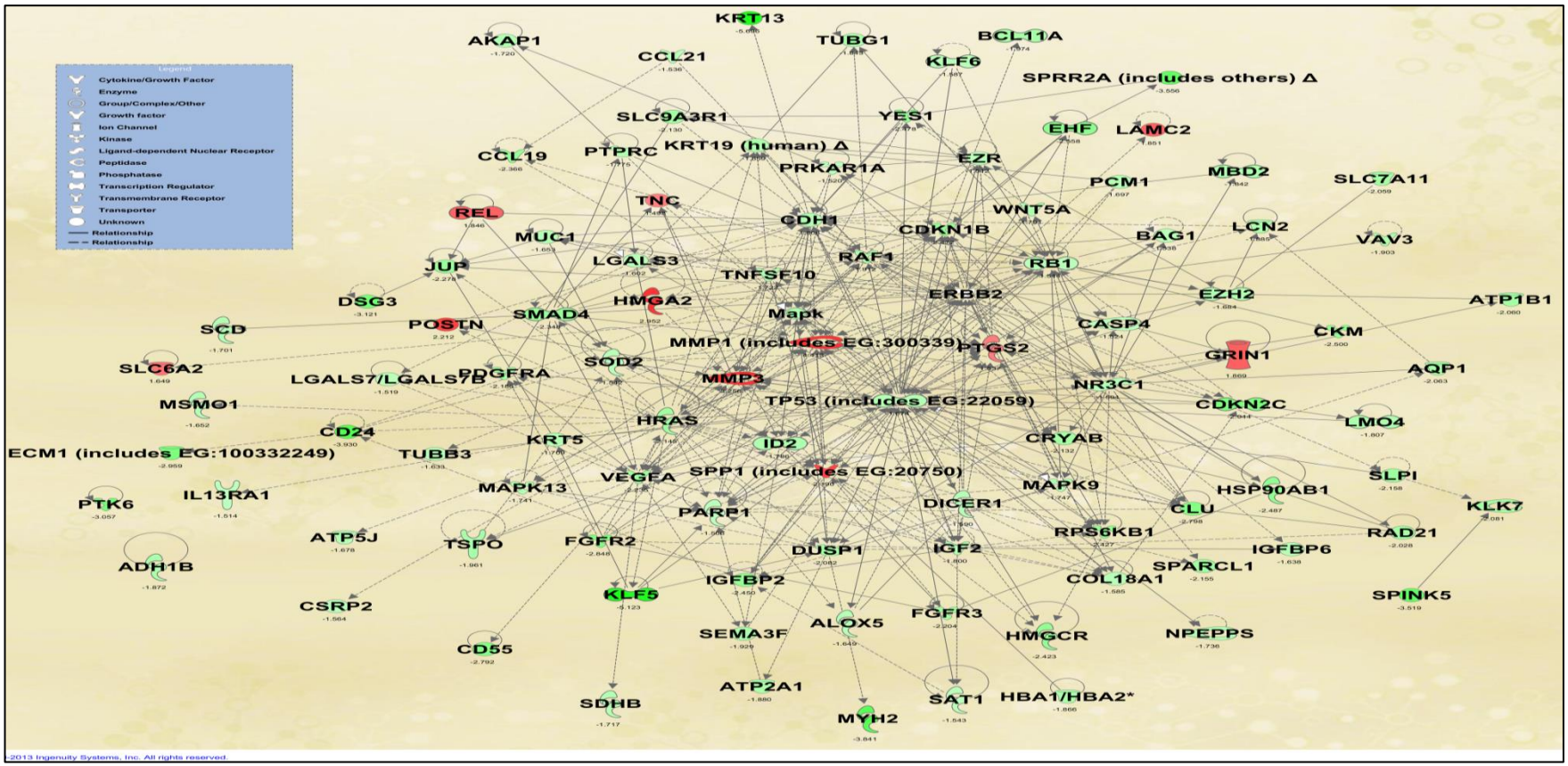


Figure 4.7: Master molecular network of significant genes of tumour samples from UK.
Master network built by connecting significant genes involved in OSCC, HNSCC, and SCC identified by IPA tool (version 8.5) from up- and down-regulated genes using overlapping core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed genes are shaded in red and downregulated genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with blue edges as potential key regulators of the master network. The network represents significant genes at FDR level ≤ 0.005 and fold change cut-off value ± 1.5.For clearer view of the network, see supplementary file 8.

Although increased expression of COX2 enzyme linked to the influence of tobacco use (a major risk factor with alcohol in UK samples) and heavy consumption of betel quid (a major risk factor in Sri Lankan samples) (Neville and Day, 2002, Jeng et al., 2000, Ayan et al., 2000, Sharan, 1996, Mashberg et al., 1993), COX2 appeared upregulated in UK but not Sri Lankan samples. This observation could be important because pathway analysis displays it at the centre of many of the associated transcriptional pathways aberrantly expressed in UK OSCC. Upregulation of PTGS2 has been reported in association with colorectal, bladder, and gastric cancer (Sudbo and Reith, 2003) as well as head and neck and OSCC (Tsai et al., 2004, Hunter et al., 2005, Chan et al., 1999). PTGS2 upregulation may considered as important contributor to the phenotypic behaviour of OSCC and this may be exclusive to that seen in western countries (where betel quid chewing is not habitual and not an aetiological agent in the pathogenesis of OSCC). In fact topical use of COX2 inhibitors on oral dysplastic lesions and early carcinoma is beginning in early clinical trials (Nelson, 2006) and represents a positive step forward in the molecular prevention and treatment of OSCC. Therefore, substantial evidence were existed for the role of PTGS2 in tumour progression and metastasis confirming the importance of its differential upregulation in UK but not Sri Lankan OSCC compared to normal mucosa.

Nuclear Factor Kappa B (NFKB) signalling pathway is also implicated in tumour transformation and CXCL10 is more highly expressed in tumours than normal tissues appear to act with NFKB1 to promote COX2 upregulation. However, NFKB1 itself is not upregulated in the SAM analysis. It has been shown that a feedback-loop requiring constitutive NFKB1 expression for the IFNγ induced expression of CXCL10 (Hiroi and Ohmori, 2003).

Upregulation of LAMC2 (a subunit gamma 2 of laminin 5) is another relevant observation. Laminin was reported frequently associated with aggressiveness, metastasis and poor prognosis of cancer in several studies (Barsky et al., 1984a, McCarthy et al., 1985, Hintermann and Quaranta, 2004, Yamamoto et al., 2001, Yamamoto et al., 2009, Kagesato et al., 2001, Aishima et al., 2004, Soini et al., 1996, Olsen et al., 2003, Malinoff et al., 1984, Barsky et al., 1984b, Haslam and

Woodward, 2003, Kaklamani and Gradishar, 2006, Gontero et al., 2004). Interestingly, this specific subunit of laminin 5 was previously reported in the cytoplasm of budding carcinoma cells at the invasive front of OSCC (Franz et al., 2010, Kosmehl et al., 1999). This gene encodes for protein that act on cell- matrix, cell-cell interaction, and cellular migration or invasion (Chen et al., 2008a). Following attachment to laminin in the basement membrane, malignant cells secrete collagenase IV that specifically breaks down type IV collagen thus facilitate cell spreading and migration (Liotta et al., 1988). Moreover, laminin fragments generated by post-translational proteolytic cleavage bind to cell surface integrins and other proteins to trigger and modulate cellular movement(Hintermann and Quaranta, 2004). Another study recruited mouse models showed that tumour cells with high levels of laminin and low level of unoccupied laminin receptor are resistant to killing by natural cytotoxic T cells, show high malignant behaviour (Malinoff et al., 1984), and by injecting laminin receptor's blockers stop lung metastasis of hematologenously introduced tumour cells (Barsky et al., 1984a). Furthermore, several types of cancers distinguished with abundance of unoccupied laminin receptors (Liotta et al., 1988).

Upregulation of Periostin (POSTN) in UK tumours relative to normal tissues is consistent with suggestions that it is a candidate for early prediction of malignant behaviour of UK cancer. Clinicopathological studies revealed that Periostin overexpression is well correlated with metastasis and poor prognosis in various cancers (Siriwardena et al., 2006, Bao et al., 2004, Sasaki et al., 2003, Sasaki et al., 2001b, Sasaki et al., 2002, Sasaki et al., 2001a). Moreover, *in vivo* studies revealed Periostin displaying a striking phenotype of greatly accelerated tumour metastatic growth by using the animal model system of metastasis (Yan and Shao, 2006, Kudo et al., 2007, Shao et al., 2004, Bao et al., 2004).

Invasion and angiogenesis promoted by Periostin may lead to metastasis of OSCC through the following steps: (I) Periostin-overexpressing OSCC cells secrete Periostin, (II) secreted Periostin binds to integrins both in OSCC cells and endothelial cells, (III) interaction between Periostin and integrins promotes invasion through inhibition of interaction between integrins and ECM and/or activation of

intracellular signal in OSCC cells, (IV) interaction between Periostin and integrins promotes angiogenesis in endothelial cells and (V) invasion and angiogenesis leads to metastasis. Overall, these findings suggest that Periostin can be a useful marker to predict metastasis in OSCC. Furthermore, it has been reported that serum levels of Periostin were elevated in patients with breast cancer, non-small-cell lung cancer and thymoma (Chen et al., 2008a, Suhr et al., 2007, Sasaki et al., 2003, Sasaki et al., 2001b, Sasaki et al., 2001a).

MMP3 also noticed highly expressed in UK tumours than normal tissue which encodes a secreted protease, matrix metalloproteinase-3, whose action is to degrade the major components of the extracellular matrix, and is thought associated with cervical lymph node metastases in HNSCC (Ye et al., 1996, Wiegand et al., 2005). MMP3 also reported upregulated in previous microarray studies which used the same criteria as this study (Nagata et al., 2003, Tsai et al., 2004, Ye et al., 2008, Suhr et al., 2007). TNC also showed overexpression which encodes an ECM protein, tenascin-C, that regulates cell adhesion, migration, and growth (Jones and Jones, 2000). Furthermore, COL11A1 demonstrated the greatest differential expression between metastatic and non- metastatic tumour (Ziober et al., 2006, Schmalbach et al., 2004).

Significantly lower expression of SLPI was observed in UK OSCCs compared to normal controls and may correlate with the invasive nature of this population group. According to a previous report, this gene showed an inverse correlation between its expression and histological parameters associated with tumour progression, including stage and pattern of invasion as well as invasive cell grade. Data suggested that SLPI may possess anti-tumorigenic activity via its ability to interfere with multiple requisite proteolytic steps underlying tumour cell invasion and may provide insight into potential stratification of oral cancer according to risk of occult metastasis, guiding treatment strategies (Abdel-fattah, Wen et al., 2011).

Additional observations that may reflect the invasive phenotypic behaviour of UK cases is reduced expression of DUSP1, a transcriptional target of tumour suppressor p53, inducing cell cycle arrest or apoptosis (Li et al., 2003) with a role in protein modification; signal transduction and oxidative stress (Markopoulos et al., 2010). It was reported that the expression of DUSP1 is decreased in primary ovarian tumours compared with corresponding normal tissues (Denkert et al., 2002). A previous microarray study showed that the expression of DUSP1 gene is decreased in OSCC and a candidate for tumour suppression, mediating PTEN signalling pathways in OSCC (Tomioka et al., 2006). In many human cancers, increased DUSP1 expression was significant in early carcinogenic stage but fell progressively with higher histologic grade and metastasis. The downregulation of DUSP1 occurs predominantly as the tumour progresses from localized to advanced disease (Tomasi et al., 2010).

To obtain biological background knowledge of the expressed genes in UK OSCCs in relative to the corresponding normal samples, IPA system was used to examine gene molecular function, biological process, canonical pathways, and transcriptional regulators analysis. 926 genes differentiating between UK tumour and normal mucosa at FDR level ≤ 0.005 and fold change cut off equal ± 1.5 were uploaded to IPA system resulted in 925 mapped genes after correction/ conversion processes of the uploaded genes. The most relevant activated and inhibited functions at FDR<0.001 are reported in Figure 4.5. Tumourigenesis, cellular growth, proliferation, transformation, movement, migration, invasion, metastasis, and ECM degradation, functions were appeared in the top of highest-level functions. Altogether these functions might influence the phenotypic behaviour of UK OSCCs related to tumour invasiveness, recurrence, and metastasis.

Utilizing the canonical pathway analysis tool, determined fatty acid metabolism, glycolysis/ gluconeogenesis, citrate cycle, xenobiotic metabolism signalling, and NRF2-mediated oxidative stress response as the main relevant significant pathways associated with UK tumours along with a significant deregulation of Arachidonic Acid Metabolism pathway.

Next, investigation of the biological relationships between genes and gene products were carried out by performing a network analysis for the genes altered in UK tumour samples. Significant genes were uploaded to IPA system (n=926). At FDR = 0.001, setting a cut off value of FC =1.5, 896 genes were mapped and considered eligible for network analysis. This network shows higher expression of essential central nodes, transcription regulators, and OSCC biomarkers including MMP1, PTGS2 (COX2), POSTN, MMP3, SPP1, HMGA2, LAMC2, TNC, GRIN1, REL, and SLC6A2 along with downregulation of TP53, RB1, HRAS, CDKN1B, DUSP1, CDH1, VEGFA,IGF2, NR3C1, PARP1, RPS6KB1, RAF1, SOD2, EZH2, SMAD4, ID2, JUP, TNFSF10, BAG1, SAT1, CASP4, CD24, KLF5, LGALS7, CDKN2C, CRYAB, MAPK9, MAPK13, FGFR2, FGFR3, HSP90, YES1, EHF, RAD21, COL18A1, PTPRC, WNT5A, KRT5, and MUC1 (Figure 4.7) highlighting MMP1, MMP3, and PTGS2 upregulation and downregulation of TP53, RB1, CDH1, and HRAS at the centre of the network. Overlapping these genes over canonical pathway tool displayed the top perturbed pathways in the following ranking; molecular mechanism of cancer, Glucocorticoid receptor signalling, axonal guidance signalling, RAR activation, IL8 signalling, PIK/ AKT signalling, PTEN signalling, and B-cell receptor signalling.

Some of the expressed genes have a special importance in the tumorigenic process acting as transcription regulators. Of the common perturbed transcription regulators, deregulation of MYC, TP53, NR3C1, and NFE2L2 were observed consistently in both populations which will be explained later on in the common gene comparison section. Activation of ESR1, SMAD3, and SMAD4 transcription regulators in UK tumour cases confirming the phenotypic behaviour of OSCCs from UK. Oestrogen receptors are over-expressed in around 70% of breast cancer cases. Two hypotheses have been proposed to explain why this causes tumourigenesis, and the available evidence suggests that both mechanisms contribute: First, binding of Oestrogen to the ER stimulates proliferation of mammary cells, with the resulting increase in cell division and DNA replication, leading to mutations. Second, Oestrogen metabolism produces genotoxic waste (Deroo and Korach, 2006). Oestrogen and the ERs have

also been implicated in breast, ovarian, colon, prostate, and endometrial cancer (Harris et al., 2003).

Activation of SMAD proteins that accumulate in the nucleus can control transcription of a large number of target genes. The ability of SMADs to target a particular gene and the decision to activate or repress gene transcription are determined by cofactors that confer discriminating properties to the SMAD complex. The upregulation of common TGF-β pathway genes is Trans activated by SMAD2/3/4. However, SMAD signalling is essential for most, but not all, TGF-β gene responses (Siegel and Massague, 2003). Induction of *CTGF* and *IL11* by TGF-β in tumour cells involves the binding of a SMAD complex to the promoter region of these genes, indicating that the SMAD pathway can involve in metastasis of breast cancer to bone by upregulation of pro-metastatic genes (Sethi et al., 2011).

Furthermore, migratory ability of epithelial cells relies on loss of cell to cell contacts and acquisition of fibroblastic characteristics, a process that is commonly referred to as the epithelial mesenchymal transition (EMT). Several signalling pathways have been implicated in TGF-β-induced EMT, including SMADs, PI3K–AKT, RHOA and p38 MAPK78. Experiments showed that SMADs are indispensable for the EMT process (Hocevar et al., 1999, Yu et al., 2002).

Inhibition of AP-1 complex (FOS) was noticed being significant in UK tumours and has been implicated in the transformation and progression of cancer. It regulates gene expression in response to a variety of stimuli, including cytokines, growth factors, and stress (Hess et al., 2004) and in turn controls a number of cellular processes including differentiation, proliferation, and apoptosis (Ameyar et al., 2003). Observations in human hepatocellular carcinoma cells indicate that c-Fos is a mediator of c-myc-induced cell death and might induce apoptosis through the p38 MAP kinase pathway (Kalra and Kumar, 2004). This result confirmed that loss of c-FOS expression is significantly and independently associated with reduced progression-free and overall survival in a cohort of optimally treated patients with epithelial ovarian cancer which suggest that c-FOS might have a function in tumour suppression in ovarian cancer. Moreover, increase evidence that AP-1 might also

have an important function in cell death (Shaulian and Karin, 2001) and the ability of AP-1 to participate in several different cellular processes requires activation of different target genes under different conditions (Mahner et al., 2008).

4.4.6 Gene expression profile of tumour versus normal samples from Sri Lanka

The analysis were supervised in two groups according to sample type; OSCC and normal mucosa identifying 2070, 1966, 480, and 28 differentially expressed genes with a fold change cut-off value equal to ±1.5 and FDR levels of ≤0.005, ≤0.001, ≤ 0.0001, and equal 0.00 respectively. At FDR level ≤ 0.005, 98 genes were upregulated and 1972 genes were downregulated in OSCC compared to normal mucosa (Table 4.6, Figure 4.8, and Supplementary file 9). Altering the stringency of SAM analysis to accept only FDR level ≤ 0.0001 reduced the profile to 480 all downregulated genes (Supplementary file 10). At FDR level=0.0 a set of 28 genes appeared all downregulated within Sri Lankan tumours relative to their normal comparators with a fold change magnitude ≥2 (Supplementary file 11). Among the differentially expressed genes represented tumour samples of Sri Lanka, Potential biomarkers were identified which previously reported a relation to OSCC and HNSCC (Table 4.6). Several of the above were significantly higher or lower in Sri Lankan but not UK tumour samples relative to normal oral tissues from the same country (Elevated expression of CD80, IL1RL1, GML, CYP27B1, SCN10A, HOXC10, and SH2D2A along with reduced expression of TWIST1, RECK, PSPH, GNG11, LSM7, and ERAP2). Interestingly, higher expression of CD80, IL1RL1 which mediate anti-tumour immunity (Tajima et al., 2003, Thomas and Wen, 2006, Vesosky and Hurwitz, 2003, Chen et al., 1993)**,** and GML upregulation which play a role in tumour suppression (Kimura et al., 1997) along with downregulation of PSPH and TWIST1 over expression is critical in SCC proliferation and metastasis (Yang et al., 2008, Yang and Wu, 2008, Yuen et al., 2007, Niu et al., 2007, Bachelor et al., 2011) and might presumably accountable in resisting tumour invasion and metastasis of Sri Lankan tumour cases.

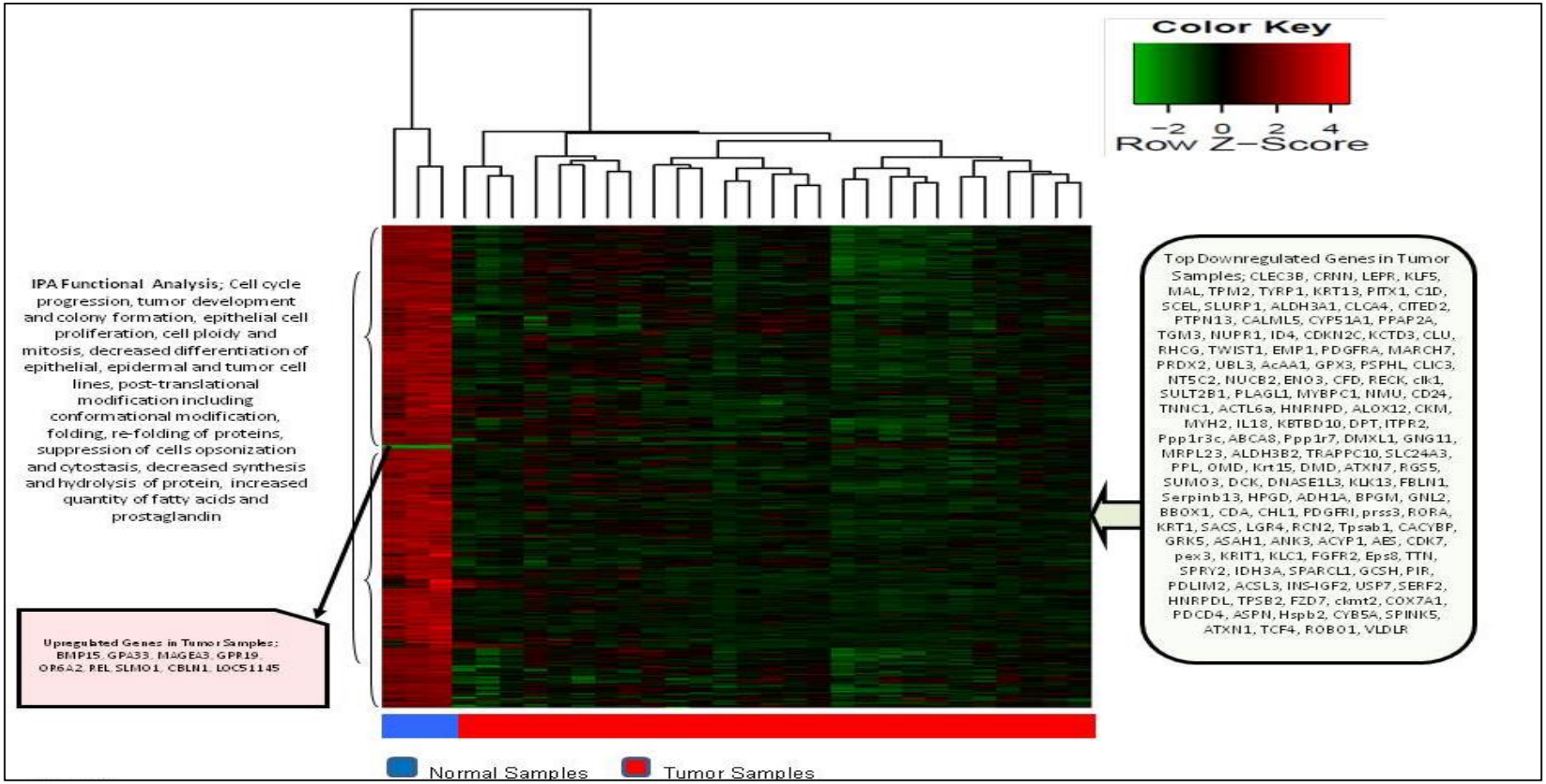


Figure 4.8: Hierarchical clustering of the1065 most differentially expressed genes separating 3 Sri Lankan normal oral mucosal and 27 patient-paired OSCC specimens using SAM supervised analysis at FDR≤0.001.

Red and blue bars represent cancerous and normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All gene expression values were either above or below fold change magnitudes equal to 2. For clearer view of the figure, see supplementary file 12.

Table 4.6: Significantly differentially expressed genes discriminating Sri Lankan tumour from normal samples.*

Gene Symbol	Gene ID	Fold Change	Gene Symbol	Gene ID	Fold Change
Genes higher expressed in OSCCC					
HMGA2	208025_s_at	2.8	SLC6A2	215715_at	1.9
REL	206035_at	2.8	SPP1	209875_s_at	1.9
CXCL10	204533_at	2.8	CXCL11	210163_at	1.8
BMP15	221332_at	2.6	RARB	205080_at	1.8
SCN10A	208578_at	2.2	OPRM1	211359_s_at	1.8
MMP1	204475_at	2.1	MMP10	205680_at	1.7
CHRM2	221330_at	1.9	CNR2	206586_at	1.5
Genes lower expressed in OSCC					
CLEC3B	205200_at	-5.3	CRYAB	209283_at	-2.4
CRNN	220090_at	-5.2	ID2	201565_s_at	-2.4
KLF5	209211_at	-4.8	TGFBR3	204731_at	-2.4
MAL	204777_s_at	-4.8	CASP4	209310_s_at	-2.3
KRT13	207935_s_at	-4.7	FGFR1	211535_s_at	-2.3
SCEL	206884_s_at	-4.3	TP53	201746_at	-2.3
CLCA4	220026_at	-4.2	BAG1	211475_s_at	-2.2
TGM3	206004_at	-4.1	EHF	219850_s_at	-2.1
CDKN2C	204159_at	-4	MAPK9	203218_at	-2.1
CLU	208792_s_at	-4	PTPRC	212588_at	-2.1
EMP1	201324_at	-4	RB1	203132_at	-2.1
PDGFRA	203131_at	-4	EZH2	203358_s_at	-2
RHCG	219554_at	-4	LGALS7	206400_at	-2
TWIST1	213943_at	-4	PTCH1	209815_at	-2
CD24	208650_s_at	-3.7	RAF1	201244_s_at	-2
SERPINB13	211362_s_at	-3.3	SOD2	216841_s_at	-2
INS-IGF2	202409_at	-3.1	LGALS3	208949_s_at	-1.9
HSPA1A//HSPA1B	200800_s_at	-3	PARP1	208644_at	-1.9
IGFBP2	202718_at	-3	BCL2	203685_at	-1.8
RPS6KB1	204171_at	-3	CDC42	208728_s_at	-1.8
KRT19	201650_at	-2.9	SMARCB1	212167_s_at	-1.8
SMAD4	202527_s_at	-2.9	WNT5A	205990_s_at	-1.8
HRAS	212983_at	-2.8	MAPK13	210058_at	-1.7
CDKN1B	209112_at	-2.7	LGALS1	201105_at	-1.7
VEGFA	210512_s_at	-2.7	PIK3CA	204369_at	-1.7
RAD21	200608_s_at	-2.6	TIMP1	201666_at	-1.7
YES1	202933_s_at	-2.6	TNFRSF10B	209295_at	-1.7
CRABP2	202575_at	-2.5	CCND1	208712_at	-1.6
HSP90AB1	214359_s_at	-2.5	JUP	201015_s_at	-1.6
NR3C1	216321_s_at	-2.5	COL1A2	202403_s_at	-1.5

* FDR value < 0.005.The complete lists of upregulated and downregulated genes in OSCCs were presented in Supplementary file 9.

The most relevant activated and inhibited functions using IPA system are reported in Figure 4.8 at FDR<0.001. Cell cycle progression, tumour development and colony formation, epithelial cell proliferation, and cell ploidy and mitosis, appeared in the top of highest-level functions. Altogether these functions show no correlation with aggressive behaviour and metastasis of cancer which might influence and confirm the phenotypic behaviour of Sri Lankan OSCCs being non-invasive with no metastasis or recurrence of the disease reported.

By utilizing canonical pathway analysis tool, mitochondrial dysfunction, oxidative phosphorylation, and ubiquinone biosynthesis are exhibited as the most common perturbed pathways involved in the process of tumour transformation in Sri Lankan group.

Mitochondrial mutations may contribute to the development of a malignant phenotype by direct effects from increased reactive oxygen species production as well as induction of aerobic glycolysis and growth promotion (Zhou et al., 2007). While normal mitochondrial function can reduce tumourigenesis, various studies showed that tumour mitochondria are structurally and functionally abnormal and unable to generate normal levels of energy (Seyfried and Shelton, 2010).

As RTG response is "off" in healthy cells with normal mitochondrial function, It turned "on" following impairment in mitochondrial energy production where transcription and signalling commences for multiple energy and anti-apoptotic related genes and proteins to include MYC, TOR, p53, RAS, CREB, NFκB, and CHOP (Wolfman et al., 2006, Kulawiec et al., 2009, Kulawiec et al., 2008, Singh et al., 2005, Miceli and Jazwinski, 2005, Butow and Avadhani, 2004). The RTG response also takes a part in the participation of multiple negative and positive regulators, which facilitate the bioenergetics' transition from respiration to substrate level Phosphorylation (Butow and Avadhani, 2004). Most cancer types are similar in expressing mitochondrial dysfunction and elevated substrate level Phosphorylation. When loss of respiration leads to glycolysis, dedifferentiation, and unbridled proliferation (Szentgyorgyi, 1977), it is expected that the presence of normal mitochondria in tumour cells would restore the cellular redox status, turn off the

RTG response, and reduce or eliminate the need for glycolysis (Warburg effect) and glutaminolysis to maintain viability. In other words, normal mitochondrial function would facilitate expression of the differentiated state thereby suppressing the tumorigenic or undifferentiated state (Seyfried and Shelton, 2010).

Mitochondria and mitochondrial functions reported an involvement in apoptotic cell death (Jacobson et al., 1994, Nguyen et al., 1993, Hockenbery et al., 1990, Hennet et al., 1993, Newmeyer et al., 1994, Schulzeosthoff et al., 1992). Moreover, recent studies suggest that mitochondrial function may play substantial role behind the commitment of cells to undergo apoptosis (Heerdt et al., 1997, Krippner et al., 1996, Liu et al., 1996, Zhang et al., 1996, Marchetti et al., 1996, Zamzami et al., 1996, Vayssiere et al., 1994, Hennet et al., 1993, Schulzeosthoff et al., 1992).

Alternative oxidations of fatty acids, ketone bodies, short-chain carboxylic acids, propionate, acetate and butyrate in tumour cells may contribute to ATP supply by oxidative phosphorylation (OXPHOS)(Ralph et al., 2010). Tumour cells maintain a significant level of OXPHOS capacity to rapidly switch from glycolysis to OXPHOS during carcinogenesis. The more realistic theory considers that OXPHOS and glycolysis cooperate to enhance energy demands along tumourigenesis (SmolkovÃ¡ et al., 2011).

According to the previous studies, glycolysis appeared a feature of several tumours associated with faster growth in high glucose environment; however, active OXPHOS is also a significant feature of other tumours occurred at a particular stage of carcinogenesis which might be more beneficial than a “glycolysis-only” metabolism in conditions of intermittent shortage in glucose supply. The metabolism of cancer cells is not constant during carcinogenesis and might depend on both the activated oncogenes and the tumour cell microenvironment (Jose et al., 2011). Moreover, several studies concluded that aerobic glycolysis is more a consequence of hypoxia rather inherent to cancer (Zu and Guppy, 2004).

Recent studies considered mitochondria as therapeutic targets for cancer chemotherapy suggested a combined strategies involving modulation of both glycolytic and mitochondrial pathways which might be required for more efficient elimination of malignant cells (Rodriguez-Enriquez et al., 2009, Galluzzi et al., 2006). As tumour cells shift among different patterns of energy metabolism (Pedersen, 1978, SmolkovÃ¡ et al., 2011, Nadege et al., 2009, Rodriguez-Enriquez et al., 2009), variant pharmacological trials could provide drug specificity via variable interference with distinctive steps of cancer energy production pathways.

Another identified observation is the relatively severe downregulation of p53 in Sri Lankan tumour samples along with the absence of COX2, raised the previous reported evidence stated that carcinogenic p53 deficiency results in a decreased level of COX2 and triggers a shift toward anaerobic metabolism. Where lactate synthesis is increased, cellular ATP levels remain stable (Zhou et al., 2003). This evidence is further confirmed by results of current study highlighting mitochondrial dysfunction, oxidative phosphorylation, citrate cycle suppression, and glycolysis/ gluconeogenesis on the top canonical pathways perturbed in Sri Lankan tumour cases.

Briefly any unspecific condition that damages a cell's oxidative phosphorylation, but is not severe enough to induce apoptosis, can potentially initiate the path to a malignant cancer. Impaired mitochondrial function can induce abnormalities in tumour suppressor genes and oncogenes. The accumulation of mitochondrial damage over time is what ultimately leads to malignant tumour formation. An increased dependency on substrate level phosphorylation for survival would follow each round of metabolic and genetic damage thus initiating uncontrolled cell growth and eventual formation of a malignant neoplasm. Basically the well-documented tumour-associated abnormalities in oncogenes, tumour suppressor genes, and chromosomal imbalances can arise as a consequence of the progressive impairment of mitochondrial function (Seyfried and Shelton, 2010).

The master network of Sri Lankan tumour samples displayed upregulation of MMP1, SPP1, HMGA2, RARB, REL, OPRM1, CNR2, SCN10A, and CHARM2 along with downregulation of TP53, BCL2, VEGFA, CDKN1B, SMAD4, LGALS1, CDKN2C, LGALS3, RB1, RAF1, HRAS, COL1A2, CDC42, PIK3CA, PTCH1, TWIST1, PDGFR, SOD2, KLF5, CASP4, BAG1, ID2, RAF1, MAPK9, TGFBR, TIMP1, CCND1, WNT5A, SMARCB1, TNFRSF10, KRT19, FGFR, JUP, YES1, and COL18A1 as major key players with MMP1 upregulation along with TP53,BCL2, RB1, CCND1, and HRAS downregulation at the centre of the network (Figure 4.9). Overlapping these genes with canonical pathway tool displayed the top affected pathways by the following rank; molecular mechanism of cancer, PTEN signalling, axonal guidance signalling, PIK/ AKT signalling, NFκB signalling, Glucocorticoid receptor signalling, RAR activation, B-cell receptor signalling, and telomerase signalling pathway as the top relevant pathways in the process of tumour transformation in Sri Lankan cases.

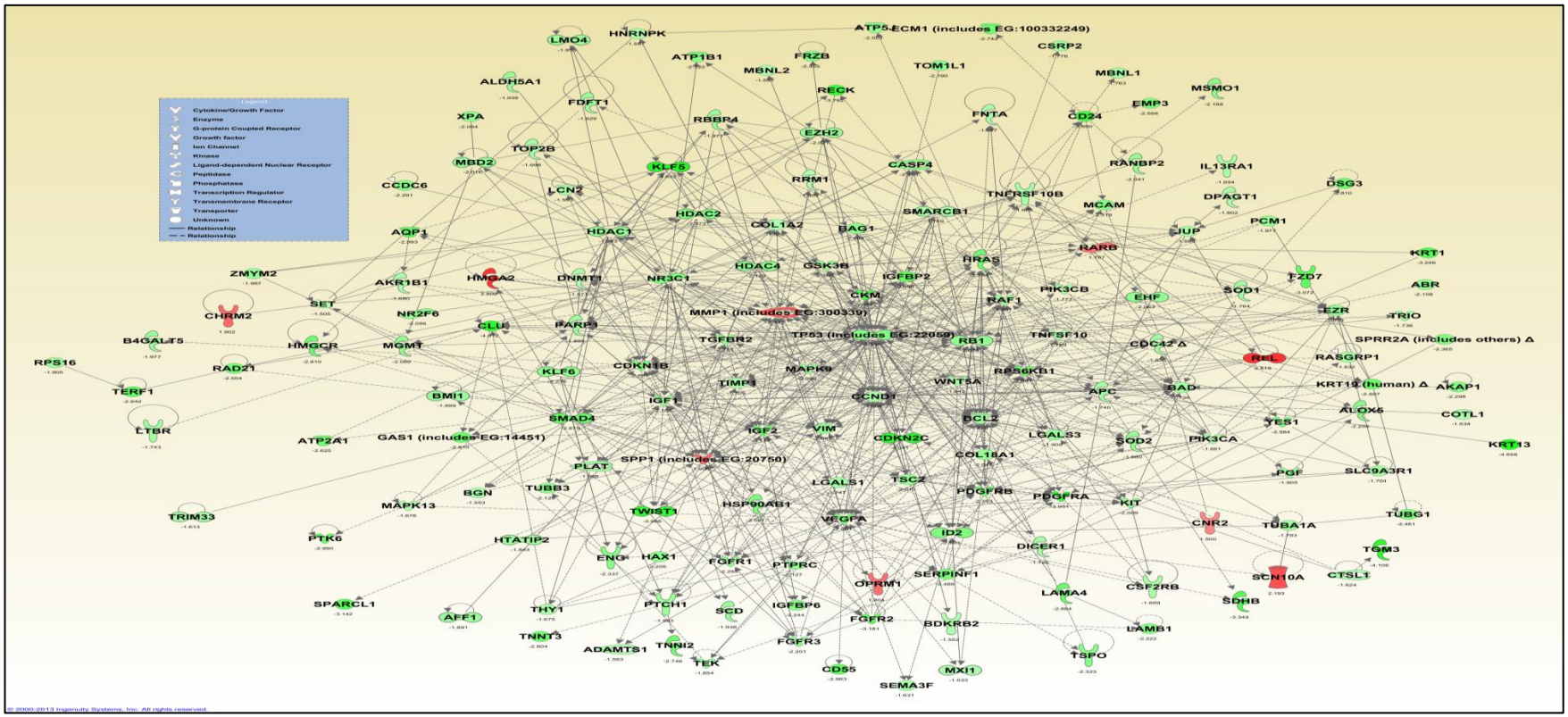


Figure 4.9: Master molecular network of significant genes of tumour samples from Sri Lanka.

Master network built by connecting significant genes involved in OSCC, HNSCC, and SCC identified by IPA tool (version 8.5) from up- and down-regulated genes using overlapping core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal upregulated genes are shaded in red and downregulated genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with blue edges as potential key regulators of the master network. The network represents significant genes at FDR level ≤ 0.005 and fold change cut-off value ± 1.5. For clearer view of the network, see supplementary file 13.

4.4.7 Common gene expression profile of tumour versus normal samples from both UK and Sri Lanka

Out of the total number of supervised differentially expressed genes at FDR ≤ 0.005 and fold change cut-off values equal to 2.0, 846 genes (807 downregulated and 39 upregulated) were observed commonly expressed in tumours versus controls for the cases from the two populations together (Table 4.7 and Supplementary file 14). The common significant biomarkers which previously reported in relation to OSCC and HNSCC are shown in Table 4.7.

MMP1 was more highly expressed in UK OSCCs compared to normal tissues and also higher, although to a lesser degree, in Sri Lankan cases. MMP1 facilitates the infiltration of tumour cells by physically destroying connective tissue and involved in the breakdown of the ECM during angiogenesis, invasion and metastasis (Seiki, 2003, Franchi et al., 2002, Thomas et al., 1999). MMP1 has been significantly associated with tumour metastasis and is widely reported in gene expression signatures of OSCC compared to normal mucosa in several studies (Alevizos et al., 2001, Mendez et al., 2002, Nagata et al., 2003, Tsai et al., 2004, Choi and Chen, 2005, Suhr et al., 2007, Ziober et al., 2006, Ye et al., 2008, Estilo et al., 2009). It is also upregulated in studies used laser capture micro dissection (LCM) to isolate pure populations of keratinocytes (Alevizos et al., 2001, Toruner et al., 2004) indicating that MMP1 is upregulated in malignant cells and therefore reinforcing its contribution to the biological behaviour of OSCC.

Table 4.7: The common significantly differentially expressed genes discriminating tumour from normal samples of both UK and Sri Lankan population groups. *

Gene Symbol	Gene ID	Fold Change	Gene Symbol	Gene ID	Fold Change
Genes lower expressed in OSCC than normal					
MMP1	204475_at	3.2	REL	206035_at	2.3
MAGEA3	209942_x_at	2.9	CXCL10	204533_at	2.2
HMGA2	208025_s_at	2.8	AIM2	206513_at	2.1
BMP15	221332_at	2.3	SPP1	209875_s_at	2
Genes lower expressed in OSCC than normal					
CRNN	220090_at	-5.7	SMAD4	202527_s_at	-2.6
MAL	204777_s_at	-5.4	CDKN1B	209112_at	-2.6
KRT13	207935_s_at	-5.2	CRABP2	202575_at	-2.5
KLF5	209211_at	-5	INS-IGF2	202409_at	-2.5
SCEL	206884_s_at	-4.7	HRAS	212983_at	-2.5
CLCA4	220026_at	-4.6	VEGFA	210512_s_at	-2.5
TGM3	206004_at	-4.5	Rgs5	209071_s_at	-2.5
EMP1	201324_at	-4.3	KRT19	201650_at	-2.4
CLEC3B	205200_at	-4.2	EHF	219850_s_at	-2.3
RHCG	219554_at	-4.2	NR3C1	216321_s_at	-2
CD24	208650_s_at	-3.8	ID2	201565_s_at	-2
CDKN2C	204159_at	-3.5	TP53	201746_at	-2
CLU	208792_s_at	-3.4	CASP4	209310_s_at	-1.9
SERPINB13	211362_s_at	-3.2	SOD2	216841_s_at	-1.8
HSPA1A//HSPA1B	200800_s_at	-3	LGALS3	208949_s_at	-1.8
IGFBP2	202718_at	-2.7	TNFSF10	202688_at	-1.7

* FDR value < 0.005. The complete lists of upregulated and downregulated genes in OSCCs were presented in Supplementary file 14.

Several investigations into tumour progression have established that AGR2 has a role in promoting cell migration and invasion (Wang et al., 2008, Maresh et al., 2010, Liu et al., 2005, Zhang et al., 2010). Recent study confirmed association of AGR2 with metastasis of HNSCC in vivo with findings indicated both CD147 and AGR2 play a role in tumourigenesis (Sweeny et al., 2012). Of specific finding is the upregulation of MAGEA3 which could be a novel biomarker in OSCC especially it's highly upregulation in both populations. Reviewing this gene uncovered its involvement in

early prediction of the relapse and prognosis in patients with hepato-cellular carcinoma (Mou et al., 2002) as well as metastasized melanoma (Roeder et al., 2005). Moreover, it considered a target in diagnosis and immunotherapy of lung cancer (Sienel et al., 2004, Vansteenkiste et al., 2007a, Vansteenkiste et al., 2007b).

The most highly downregulated gene was Cornulin (CRNN) followed by a putative tumour suppressor gene T-cell differentiation protein (MAL). CRNN functions as survival factor that participates in the clonogenicity of squamous oesophageal epithelium cell lines, attenuates deoxycholic acid (DCA)-induced apoptotic cell death. Another downregulated gene KRT13 was suggested a promising tumour marker for detecting the micro metastases in cervical lymph nodes of oral cancer (Hamakawa et al., 2000). Loss of KRT13 expression indicates tumours with a high potential for recurrence (Yanagawa et al., 2007). Several genes were similarly downregulated in both UK and Sri Lankan OSCC including the tumour suppressor gene CLU which has been associated with programmed cell death (apoptosis) (Buttyan et al., 1989). Downregulation of CLU is associated with poor outcome in breast cancer (Zhang et al., 2006) as well as OSCCs (Chin et al., 2005, Ye et al., 2008).

CRABP2 (cellular retinoic acid binding protein 2) which has been reported downregulated in prostate cancer (Okuducu et al., 2005) is also showed downregulation in both UK and Sri Lankan tumours. Retinoids have shown promising agents in the treatment of oral dysplastic lesions although toxicity has limited their clinical usefulness; retinoids may still have utility in the treatment of OSCC as they can suppress squamous differentiation in OSCC (Shalinsky et al., 1995).

RGS5 (Regulator of G-protein signalling 5) is another key player shown downregulated in Sri Lankan but not UK tumours relative to their normal controls and has been reported to play an important roles in the development of vasculature. A recent study demonstrated that levels of RGS5 inversely related to tumour metastasis in non-small cell lung cancer (Huang et al., 2012). Furthermore, RGS5 downregulation identified one of the genes capable of discriminating between lymph

node positive (N+) and lymph node negative (N0) individuals in HNSCC (Roepman et al., 2005).

846 genes differentiating between both UK and Sri Lankan OSCCs and their normal mucosa at FDR level ≤ 0.005 and fold change cut off value equal ± 1.5 were uploaded to IPA system resulted in 845 mapped genes after correction/ conversion processes of the uploaded genes with 841 genes eligible for subsequent IPA analysis.

The common key players of the comprehensive network generated by IPA system, those related to OSCC and HNSCC include upregulation of MMP1, SPP1, HMGA2, REL, and SLC6A2 along with downregulation of TP53, CLU, CDKN1B, CDKN2C, HSPA1A, CASP4, SMAD4, PARP1, PDGFRA, FGFR2, SOD2, LGALS3, KLF5, TNFSF10, NR3C1, ID2, RAF1, IGFBP2, CD24, RB1, HRAS, KRT1, WNT5A, EZR, YES1, BAG1, JUP, MAPK13, KRT19, and VEGFA (Figure 4.10).

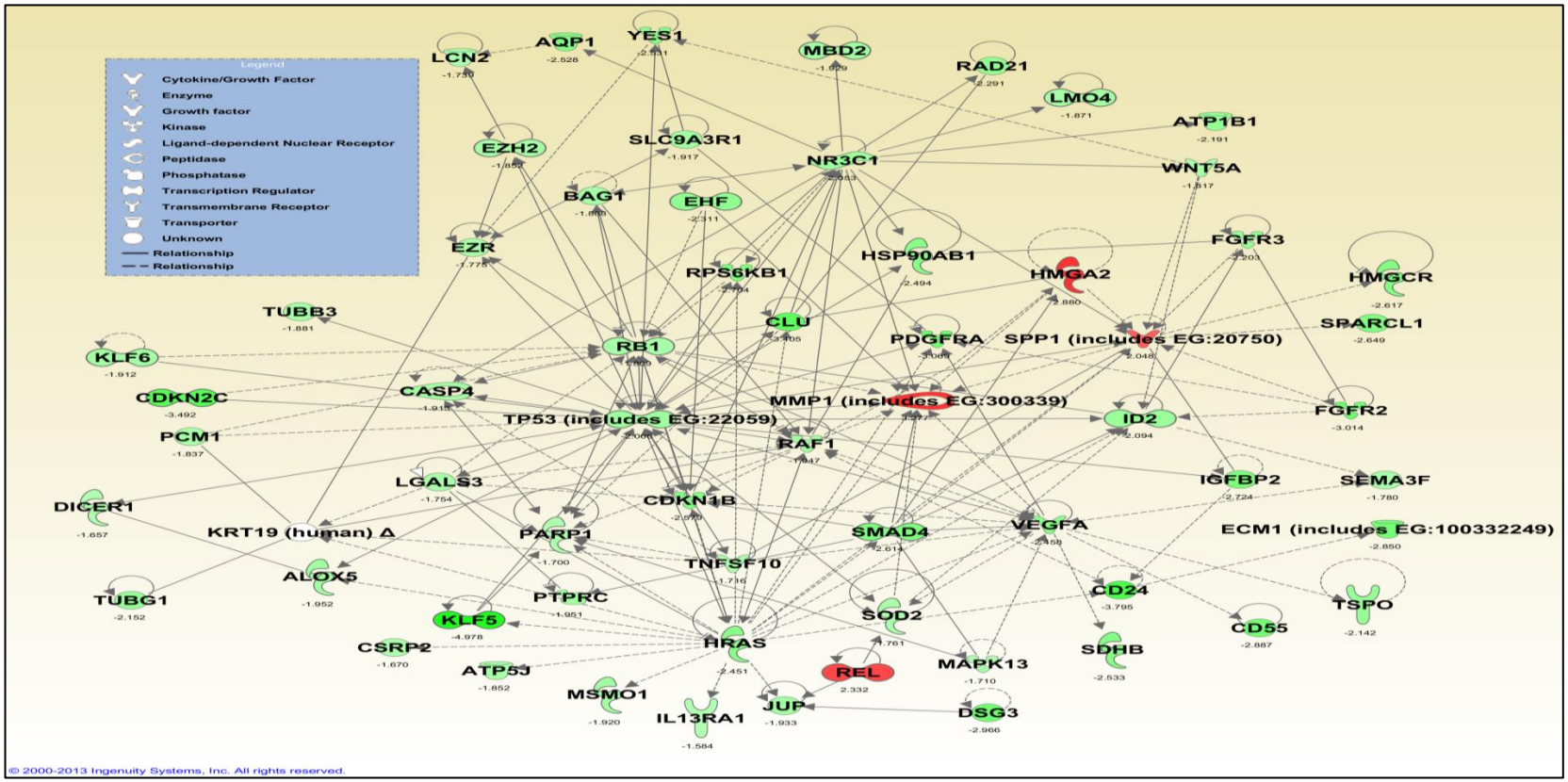


Figure 4.10: Master molecular network of common genes existed in both UK and Sri Lankan tumour samples.

Master network built by connecting significant genes involved in OSCC, HNSCC, and SCC identified by IPA tool (version 8.5) from up- and down-regulated genes using overlapping core analysis. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal upregulated genes are shaded in red and downregulated genes in green. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with blue edges as potential key regulators of the master network. The network represents significant duplicated genes existed in both populations at FDR level ≤ 0.005 and fold change cut-off value ± 1.5.

Comparison core analysis of Sri Lankan and UK tumours versus normal comparators displayed an obvious contrast in specific bio functions. This reflected by relatively higher activation of cellular compromise in Sri Lankan samples with defect and suppression of organismal function (which implements healing of lesion, wound, and oxygenation of lipid) in UK groups.

Fatty acid metabolism pathway followed by Propanoate, Butanoate metabolism, glycolysis/ gluconeogenesis, Arginine and Proline metabolism, citrate cycle, NRF2-Mediated oxidative stress response, EIF2 Signalling, Glucocorticoid receptor signalling, xenobiotic metabolism signalling are the most common perturbed pathways in the cases from the two populations. This indicates that the development of OSCCs is mediated mostly by similar biological pathways despite the differences related to racing, ethnicity, lifestyle, and/or exposure to environmental carcinogens (Gmeiner et al., 2008).

From the above mentioned pathways that not discussed in the previous sections of this study is the deregulation of NRF2-mediated oxidative response pathway. Nuclear factor (erythroid-derived 2)-like 2(NRF2)-mediated oxidative stress response pathway and genes were detected at all-time points of hypoxia suggesting activation of NRF2 are antioxidant defence system. The activation of Nrf2 results in transcriptional expression of a broad spectrum of protective enzymes including those involved in xenobiotic detoxification, antioxidative response, and proteome maintenance (Kensler et al., 2007). Low levels of Nrf2 or loss of Nrf2 activity appears to increase ROS production and DNA damage and predisposes cells to tumourigenesis (Hayes and McMahon, 2006, Klaunig et al., 2010).

Of the common perturbed transcription regulators observed consistently in both populations is MYC (v-myc myelocytomatosis viral oncogene homolog). MYC regulates the transcription of many genes including those that encode proteins involved in the control of cell adhesion, differentiation, cell cycle regulation, angiogenesis and apoptosis. The ability of MYC to modulate protein synthesis globally suggests that it has multiple roles in orchestrating cellular processes in regulating specific physiological functions (van Riggelen et al., 2010).

The MYC network can also regulate the transactivation and repression of many genes through the remodelling of chromatin structure (Knoepfler, 2007). The suppression of MYC results in global changes to chromatin structure and therefore gene expression (Knoepfler et al., 2006, Wu et al., 2007, van Riggelen et al., 2010).

A recent study demonstrated that more aggressive phenotypes of OSCC are probably the consequences of multiple deregulations of multiple copies number of amplified genes in MYC module and that MYC centred regulatory network may explain the observed differences in clinical outcomes of OSCC patients instead of direct consequences of MYC expression itself. Speculation explained that beyond MYC , multiple copy number amplified genes and abundant dysregulated genes in the MYC module may ultimately lead to dysfunction of the cell cycle, cell proliferation, and invasion in OSCC tumours (Peng et al., 2011).

NFE2L2 showed the lowest expression in tumours compared to normal controls among the transcriptional regulators distinguished in both population groups. NFE2L2 induces the expression of various genes including those that encode for several antioxidant enzymes, and it may play a physiological role in the regulation of oxidative stress.

Inhibition of NR3C1regulator was another finding. NR3C1is receptor of Glucocorticoids regulates carbohydrate, protein and fat metabolism; modulates immune responses through suppression of chemokine and cytokine production. The observation of TP53, a transcription factor, with lower expression tumours in both populations compared to normal tissue was not surprising as it is often reported to regulate the expression of genes involved in a variety of cellular functions, including cell-cycle arrest, DNA repair, and apoptosis (Kho et al., 2004, Cawley et al., 2004, Kannan et al., 2001b, Kannan et al., 2001a, Zhao et al., 2000, Yu et al., 1999).

4.4.8 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK

Using SAM supervised analysis at FDR≤0.05, only three genes separating 6 perineural invasive and 15 non-invasive OSCCs (Figure 4.11). All three genes appeared upregulated in perineural invasive tumours; two genes (PGM1, ENO2) with log2 fold change magnitudes above 1 and ILK within (0.556).

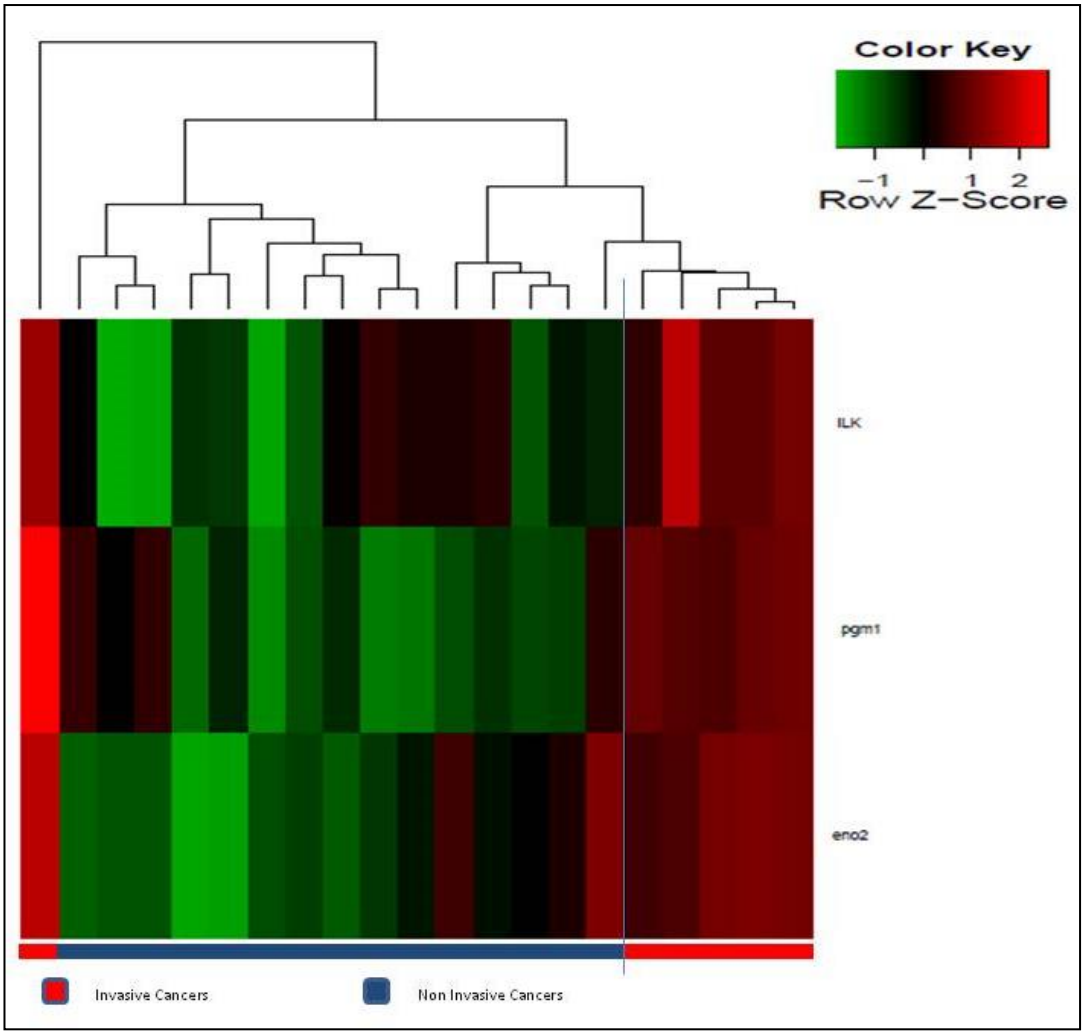


Figure 4.11: Hierarchical clustering of three differentially expressed genes separating 6 perineural invasive OSCC and 15 non-invasive OSCC samples using SAM supervised analysis at FDR≤0.05

Red and blue bars represent perineural invasive and non-invasive UK tumour samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All genes appeared over-expressed in perineural invasive tumours, two genes (PGM1, ENO2) with fold change magnitudes above 1 and ILK within 0.556 fold change magnitude.

4.4.9 Gene expression profile of high-stage versus low- stage tumour samples from both UK and Sri Lankan tumours

Using SAM supervised analysis at FDR≤0.05, 12 genes separating 15 “high stage” from 35 “low stage” OSCCs of both population groups. Most of genes appeared over-expressed in high stage tumours with fold change log2 magnitudes above or equal to 1 (Figure 4.12) and appeared associated with different cancer types and cellular processes associated with tumours.

Most of differentially upregulated genes in high stage tumours exhibited involvement in variant biological processes crucial in driven tumour lesion to late stage and metastasis with distinct roles in cellular migration, remodelling, outgrowth, invasion, differentiation, colony formation, cell cycle progression, activation, proliferation, and angiogenesis.

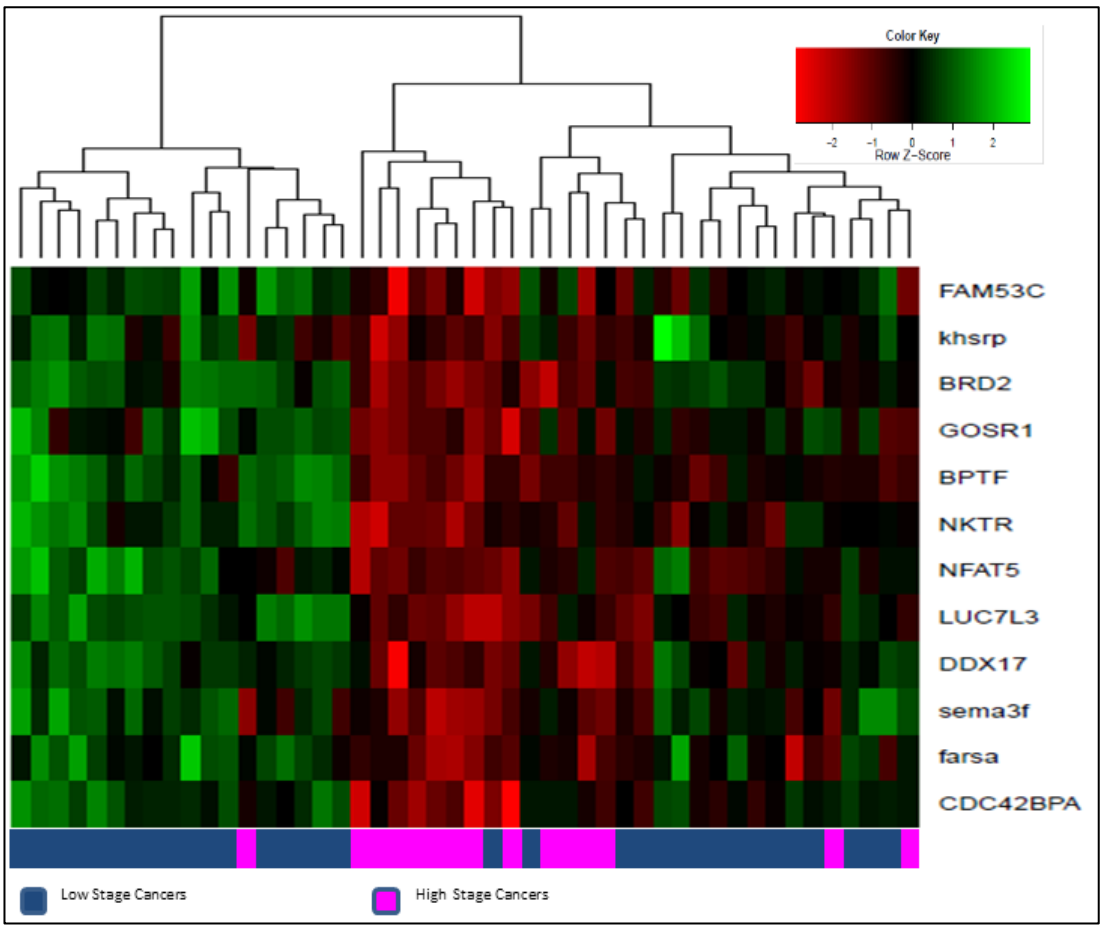


Figure 4.12: Hierarchical clustering of 12 differentially expressed genes discriminating high stage (N=15) from low stage tumours (N=33) of UK and Sri Lankan OSCCs using SAM supervised analysis at FDR≤0.05

Pink and blue bars represent high stage and low stage tumour samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. The majority of 12 genes appeared to be over-expressed in almost high stage than low stage tumour samples within expression values fold change magnitudes ≥ 0.5.

4.4.10 Gene expression profile between UK and Sri Lankan tumour samples

Using SAM supervised analysis at FDR≤0.05, 21 differentially expressed genes separating all Sri Lankan (N=27) and UK (N=21) OSCCs. Although gene expression values were only above or below log2 fold change magnitudes equal to 1, the discriminating power of these genes was weak (Figure 4.13). Altering the stringency of the SAM analysis to accept only FDR level ≤ 0.01 reduced the profile to 10 upregulated and 2 downregulated genes in Sri Lankan relative to UK cases with more powerful discrimination (Figure 4.14).

Better improvement in discrimination was observed when comparing cases from one disease stage separately. Comparing high stage cases only identified 11 upregulated genes (Figure 4.15) whereas cases of low stage separated by 31 genes, 25 up- and 6 downregulated genes in Sri Lankan versus UK tumours (Figure 4.16).

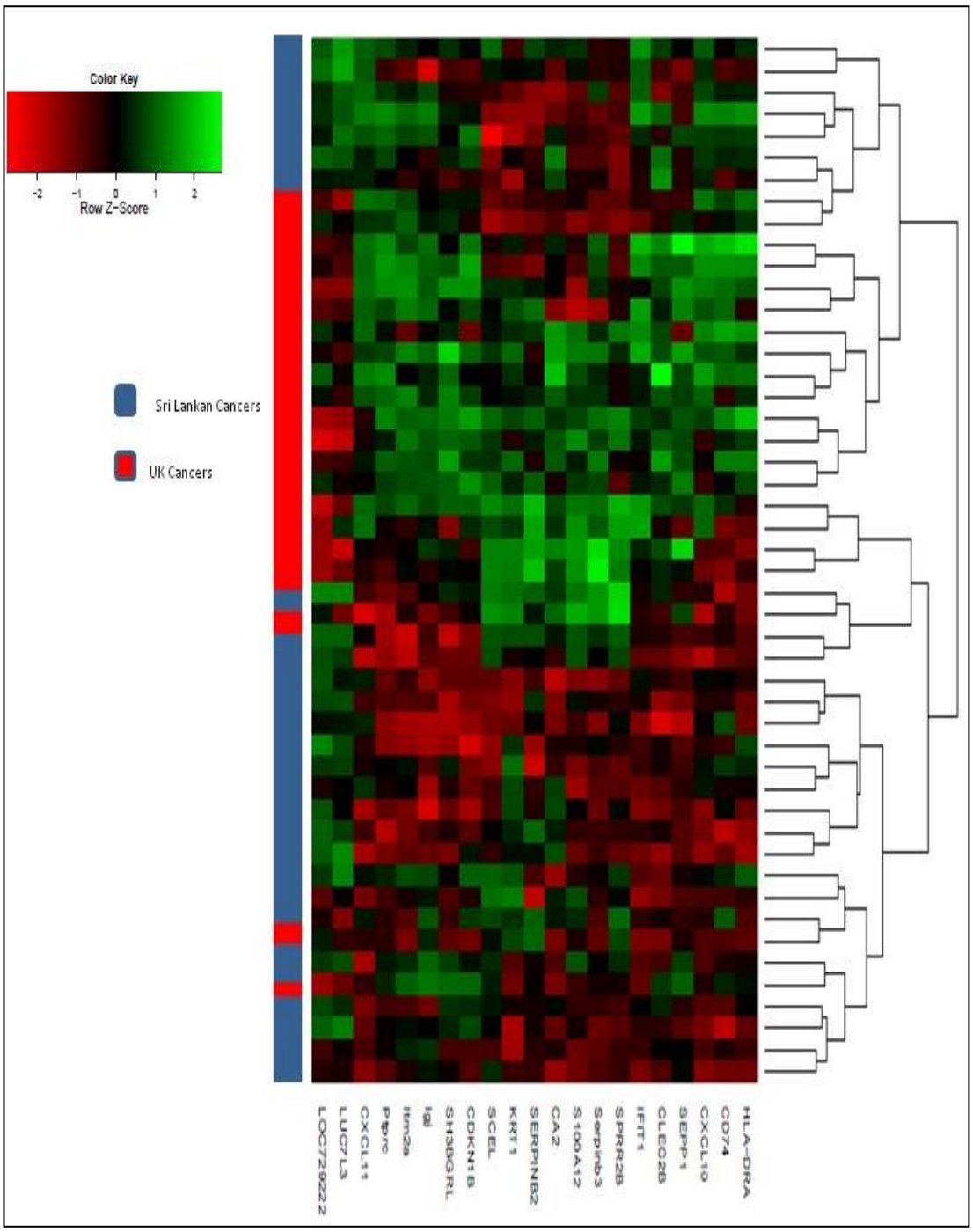


Figure 4.13: Hierarchical clustering of 21 differentially expressed genes separating Sri Lankan (N=27) and UK (N=21) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05.

Red and blue bars represent UK and Sri Lankan cancerous samples, respectively. Heatmap represents mean-centred expression of genes in columns and samples in rows. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All gene expression values were either above or below fold change magnitudes equal to 1.

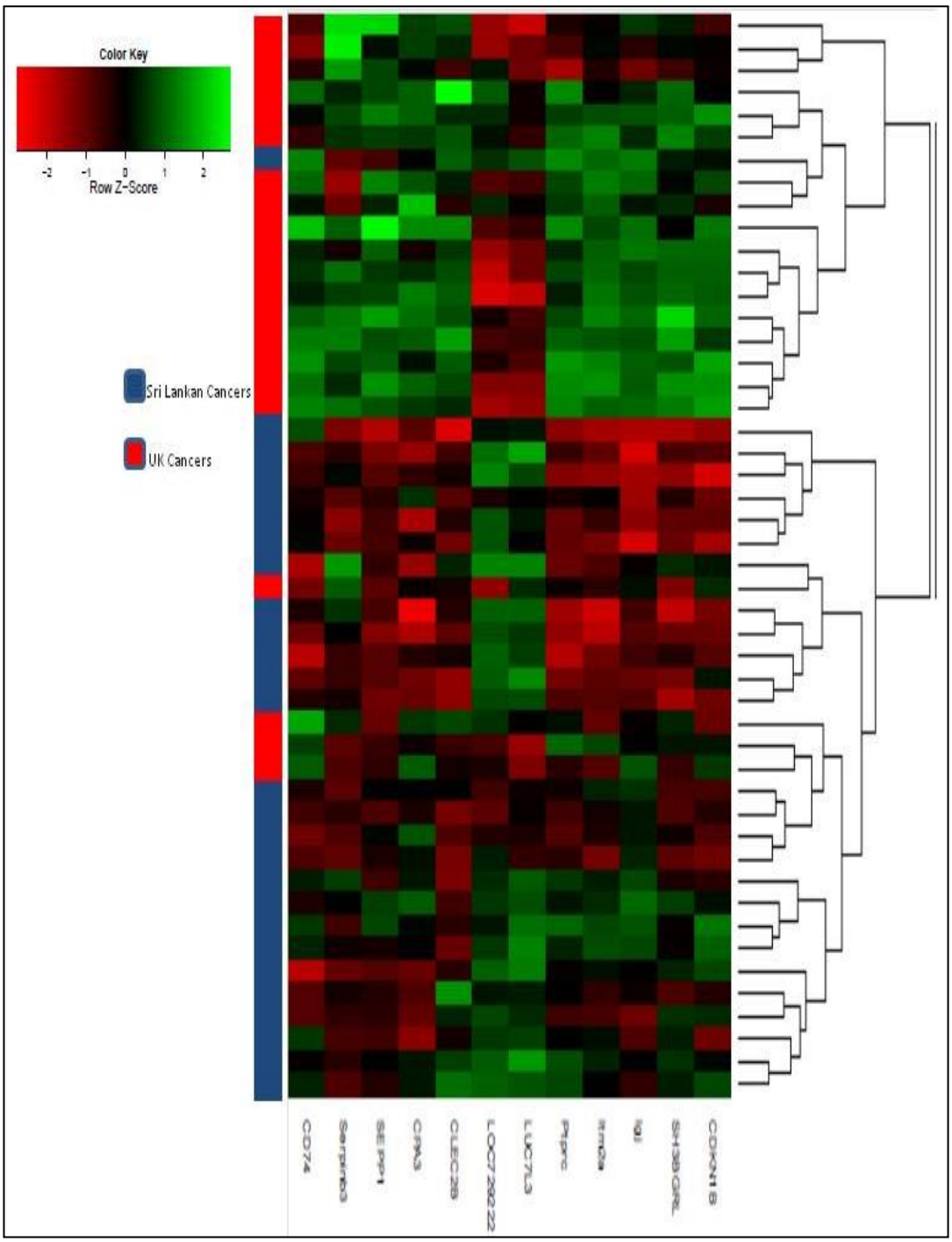


Figure 4.14: Hierarchical clustering of 12 differentially expressed genes separating Sri Lankan (N=27) and UK (N=21) OSCC samples using SAM supervised analysis at FDR≤0.01

Red and blue bars represent UK and Sri Lankan cancerous samples, respectively. Heatmap represents mean-centred expression of genes in columns and samples in rows. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All gene expression values were either above or below fold change magnitudes equal to 1.

Comparing UK and Sri Lankan OSCCs, some of the gene expression differences are attributed to cancer-associated genes, many of them are not. This comparison was carried out at a reduced fold change cut off value compared to the other supervised analyses (log2 fold change magnitudes equal to 1 at FDR level ≤ 0.01). This observation tends to support the notion that only small numbers of (tumour related) gene expression differences between UK and Sri Lankan OSCCs.

CDKN1B is only one among the genes downregulated in UK relative to Sri Lankan OSCCs which defined as being lower in high-risk tumours and is associated with poor prognosis (Chu et al., 2008). CDKN1B has already been shown as tumour suppressor gene that is downregulated in UK OSCC when compared to normal tissue, and is perhaps an important contributor to the phenotypic differences observed between UK and Sri Lankan tumours. PSIP1, SAMSN1, CCNG2, and NCOA2 play a protective role in tumorigenic process and their downregulation can favour cancer transformation (Chylack et al., 2004, Tian et al., 2002). Likewise, upregulation of CCT2 and DHCR24 in UK OSCCs contributed in rapidly proliferating tumour and melanoma metastases respectively (Yokota et al., 2001, Di Stasi et al., 2005).

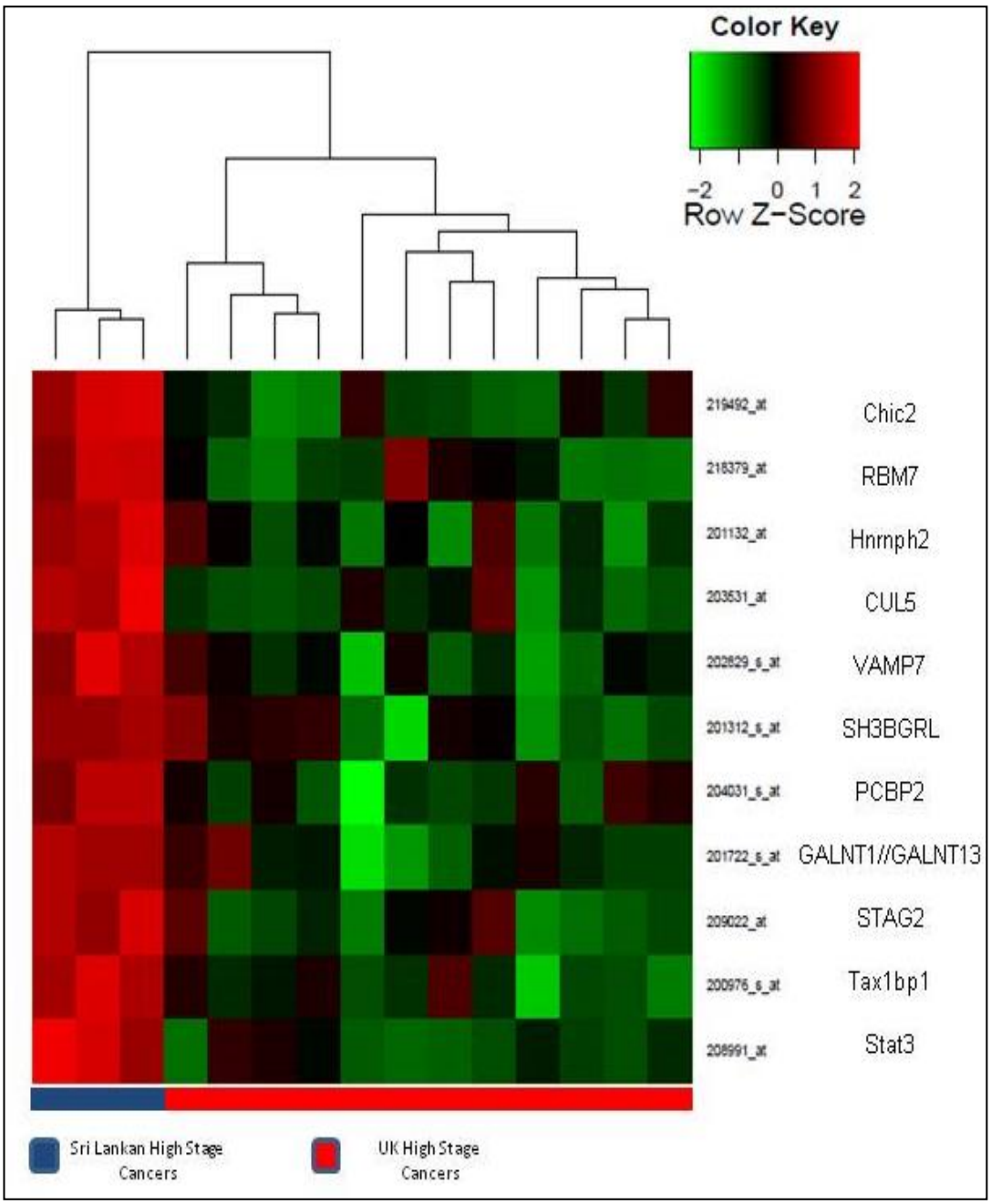


Figure 4.15: Hierarchical clustering of 11 differentially expressed genes separating "High Stage" Sri Lankan (N=3) and UK (N=12) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05

Red and blue bars represent UK and Sri Lankan high stage tumour samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All 11 gene expression values were over expressed in Sri Lankan tumours samples above fold change magnitudes equal to 1.

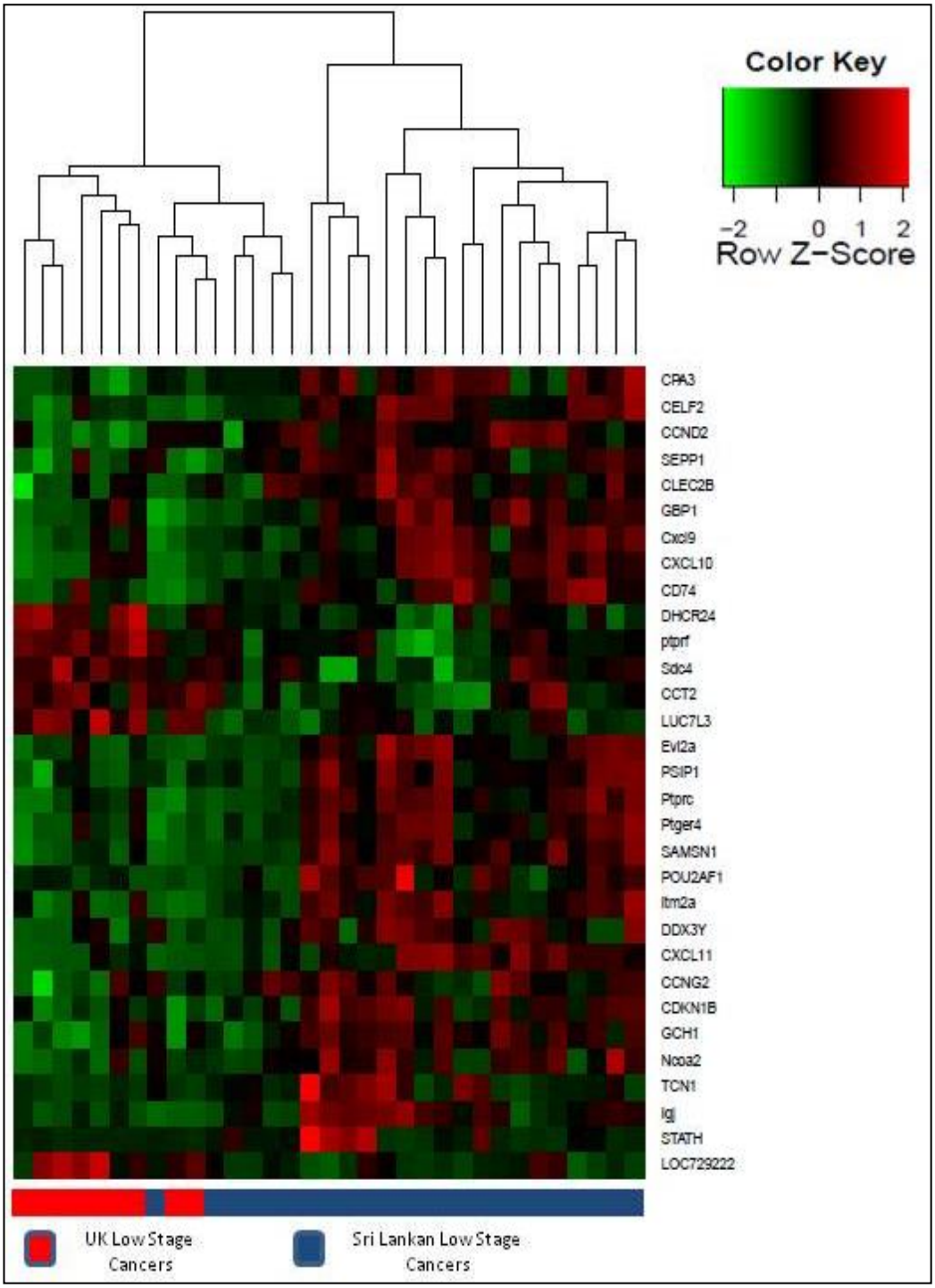


Figure 4.16: Hierarchical clustering of 11 differentially expressed genes separating "Low Stage" Sri Lankan (N=24) and UK (N=9) oral squamous cell tumour samples using SAM supervised analysis at FDR≤0.05

Red and blue bars represent UK and Sri Lankan low stage tumour samples, respectively. Heatmap represents differential mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. Gene expression values were either above or below fold change magnitudes equal to1.

Running IPA core analysis identified activated immune response functions in Sri Lankan relative to UK tumours. These include; development and mitosis of thymocytes; activation, development, differentiation, cell movement, migration, and chemotaxis of T, B-lymphocytes, leukocytes, monocytes, and other immune system cells. Curcumin consumed by Sri Lankan people for centuries as a frequent dietary component could play a role in the above observation (Pirasath et al., 2010, Jayasekera et al., 2004).

These findings might has an impact on the clinical behaviour of the two tumour groups which mainly reflected in resistant of Sri Lankan patients to the invasiveness, metastasis, and recurrence of disease as well as survival of the patient with lower morbidity and mortality relative to UK patients. The same findings are observed between Sri Lankan and UK normal groups and will be explained in details.

4.4.11 Gene expression profile between UK and Sri Lankan normal mucosa

Following the same analysis strategy used in the previous comparison categories of this study, both supervised and unsupervised analysis are used as mentioned in materials and methods section. Both analysis separated normal Sri Lankan (N=3) from normal UK (N=5) samples (Figure 4.17). Using SAM supervised method, 1439 genes were differentially expressed when comparing normal UK and Sri Lankan tissue with a fold change cut-off value equal to ±1.5 and FDR levels of ≤0.01 (Table 4.8, Figure 4.18, and Supplementary file15). Altering the stringency of the SAM analysis to accept only FDR level ≤ 0.005 reduced the profile to only 10 upregulated and 5 downregulated genes in Sri Lankan compared to UK normal mucosa (Table 4.9 and Figure 4.19).

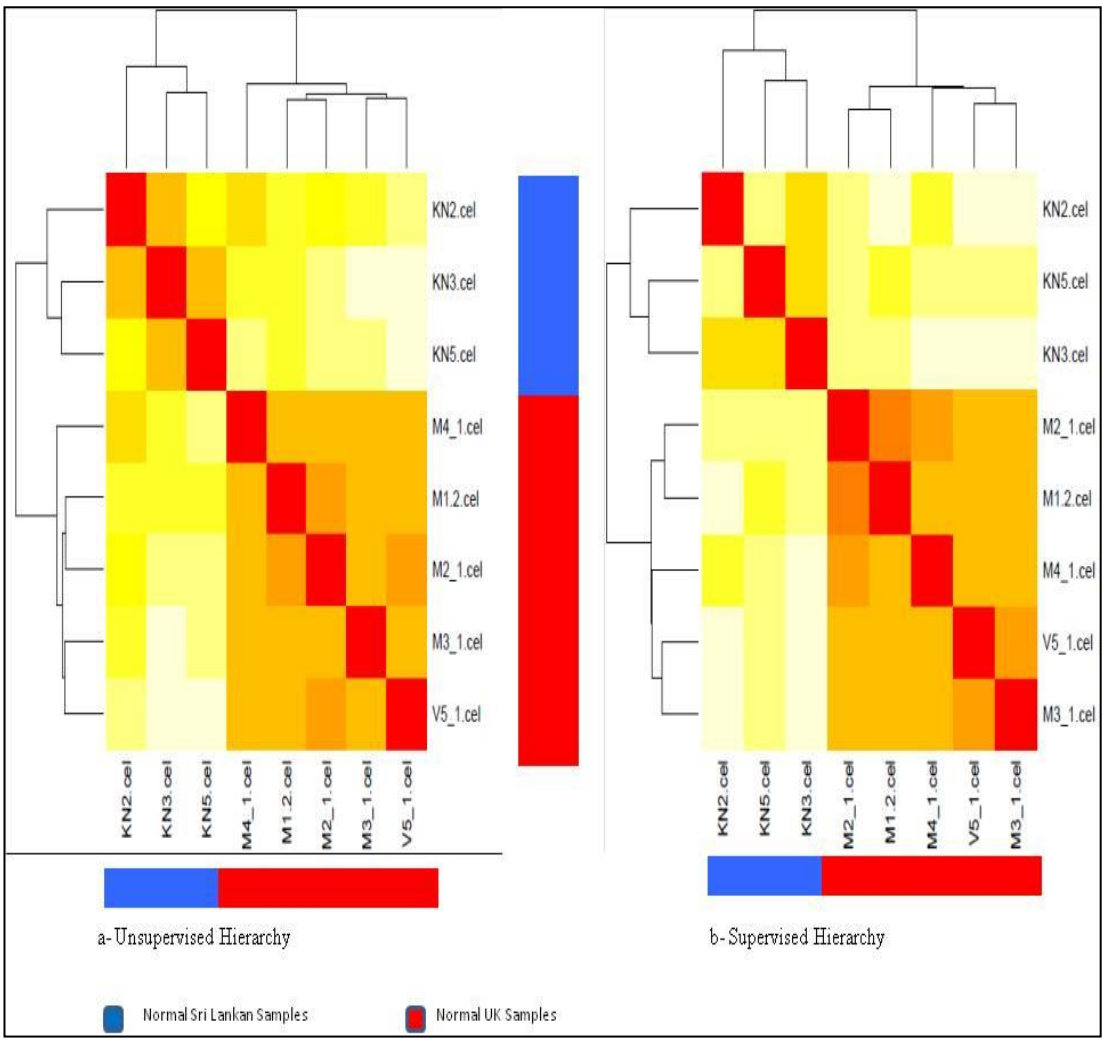


Figure 4.17: A. Unsupervised Pearson correlation hierarchical clustering of all normalized genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples. B. Supervised Pearson correlation hierarchical clustering of the 205 most differentially expressed genes using SAM supervised analysis at FDR≤0.01.

Red and blue bars represent UK and Sri Lankan normal samples, respectively. Heatmap represents samples distribution according to the mean-centred expression of genes. Expression (red= 1, yellow= lower than 1, white= lowest value than 1).

Among the differentially expressed genes discriminating normal samples of Sri Lanka from those of UK, potential biomarkers previously involved in different types of cancer were identified (Table 4.8).

Table 4.8: List of the top upregulated and downregulated genes differentially expressed between normal samples of both UK and Sri Lankan groups at FDR level ≤ 0.01.*

Gene Symbol	Gene ID	Fold Change	Gene Symbol	Gene ID	Fold Change
Genes higher expressed in Sri Lankan than UK normal samples					
PHACTR2	244774_at	4.2	LOC399959	232113_at	2.9
SCARA5	229839_at	3.7	POSTN	210809_s_at	2.9
CTHRC1	225681_at	3.5	RECK	205407_at	2.8
ITGBL1	205422_s_at	3.3	P2RY14	206637_at	2.8
IGJ	212592_at	3.3	BCAT1	226517_at	2.8
PSPH	205048_s_at	3.3	COL3A1	232458_at	2.8
FLJ40330	1569040_s_at	3.2	FERMT2	209210_s_at	2.8
GLT8D2	227070_at	3.2	TFPI	213258_at	2.8
FAM46C	226811_at	3.2	LY96	206584_at	2.8
ZBTB38	225512_at	3.1	SLC25A32	221020_s_at	2.8
PLN	204939_s_at	3.1	TYRP1	205694_at	2.8
MGP	202291_s_at	3	MGC29506	223565_at	2.8
FBN1	202766_s_at	3	LAMB1	201505_at	2.8
C12ORF23	224759_s_at	3	TMEM47	209656_s_at	2.8
C13ORF15	218723_s_at	3	LY75//CD302	203799_at	2.8
MAP1B	226084_at	2.9	POSTN	1555778_a_at	2.7
ANGPTL1	231773_at	2.9	IKBIP	227295_at	2.7
DCLK1	229800_at	2.9	EBF1	227646_at	2.7
EVI2B	211742_s_at	2.9	SELM	226051_at	2.7
ERAP2	219759_at	2.9			
MALAT1	223940_x_at	-4.1	ATP6V0D2	1553153_at	-2.2
Genes lower expressed in Sri Lankan than UK normal samples					
MALAT1	224568_x_at	-4	ZNF664	1569935_at	-2.2
EGFR	1565484_x_at	-3.6	SNORA71A	1568249_at	-2.1
EGFR	1565483_at	-3.6	WHAMML2	1557450_s_at	-2.1
PLEKHA5	1558695_at	-3	CUEDC1	1555786_s_at	-2.1
MALAT1	223578_x_at	-2.9	LOC100192378	1559966_a_at	-2.1
NEAT1	238320_at	-2.8	HNRNPD	236000_s_at	-2.1
DEFB106A//B	1552411_at	-2.8	TRIM4	224159_x_at	-2.1
FAM120A	1555944_at	-2.7	NBPF1	214693_x_at	-2
LCE1E	1559224_at	-2.6	NEAT1	214657_s_at	-2
LOC285708	1564281_at	-2.5	CLIP4	242710_at	-2
NEAT1	227062_at	-2.4	EHF	241397_at	-2
CADM3	216535_at	-2.4	MALAT1	227510_x_at	-2
HIF3A	222124_at	-2.3			

*FDR≤ 0.01.The complete lists of upregulated and downregulated genes in Sri Lankan normal samples relative to UK normal samples were presented in Supplementary file 15.

Table 4.9: List of the top upregulated and downregulated genes differentially expressed between normal samples of both UK and Sri Lankan groups at FDR level ≤ 0.005.*

Gene Symbol	Affymetrix ID	Gene Name / Description	Average Fold Change
Genes higher expressed in Sri Lankan than UK normal samples			
PHACTR2	244774_at	phosphatase and actin regulator 2	4.2
ITGBL1	205422_s_at	integrin, beta-like 1 (with EGF-like repeat domains)	3.3
GLT8D2	227070_at	glycosyltransferase 8 domain containing 2	3.2
C13ORF15	218723_s_at	chromosome 13 open reading frame 15	3.0
SELM	226051_at	selenoprotein M	2.7
NEGR1	229461_x_at	neuronal growth regulator 1	2.5
SEC62	1552790_a_at	SEC62 homolog (S. cerevisiae)	2.4
MTR	203774_at	5-methyltetrahydrofolate-homocysteine methyltransferase	2.4
POLK	223261_at	polymerase (DNA directed) kappa	2.4
RAB30	229072_at	RAB30, member RAS oncogene family	2.3
Genes lower expressed in Sri Lankan than UK normal samples			
EGFR	1565484_x_at	epidermal growth factor receptor	-3.6
EGFR	1565483_at	epidermal growth factor receptor	-3.6
PLEKHA5	1558695_at	pleckstrin homology domain containing, family A member 5	-3.0
DEFB106A//B	1552411_at	defensin, beta 106A; defensin, beta 106B	-2.8
FAM120A	1555944_at	family with sequence similarity 120A	-2.7
LCE1E	1559224_at	late cornified envelope 1E	-2.6

*FDR≤ 0.005

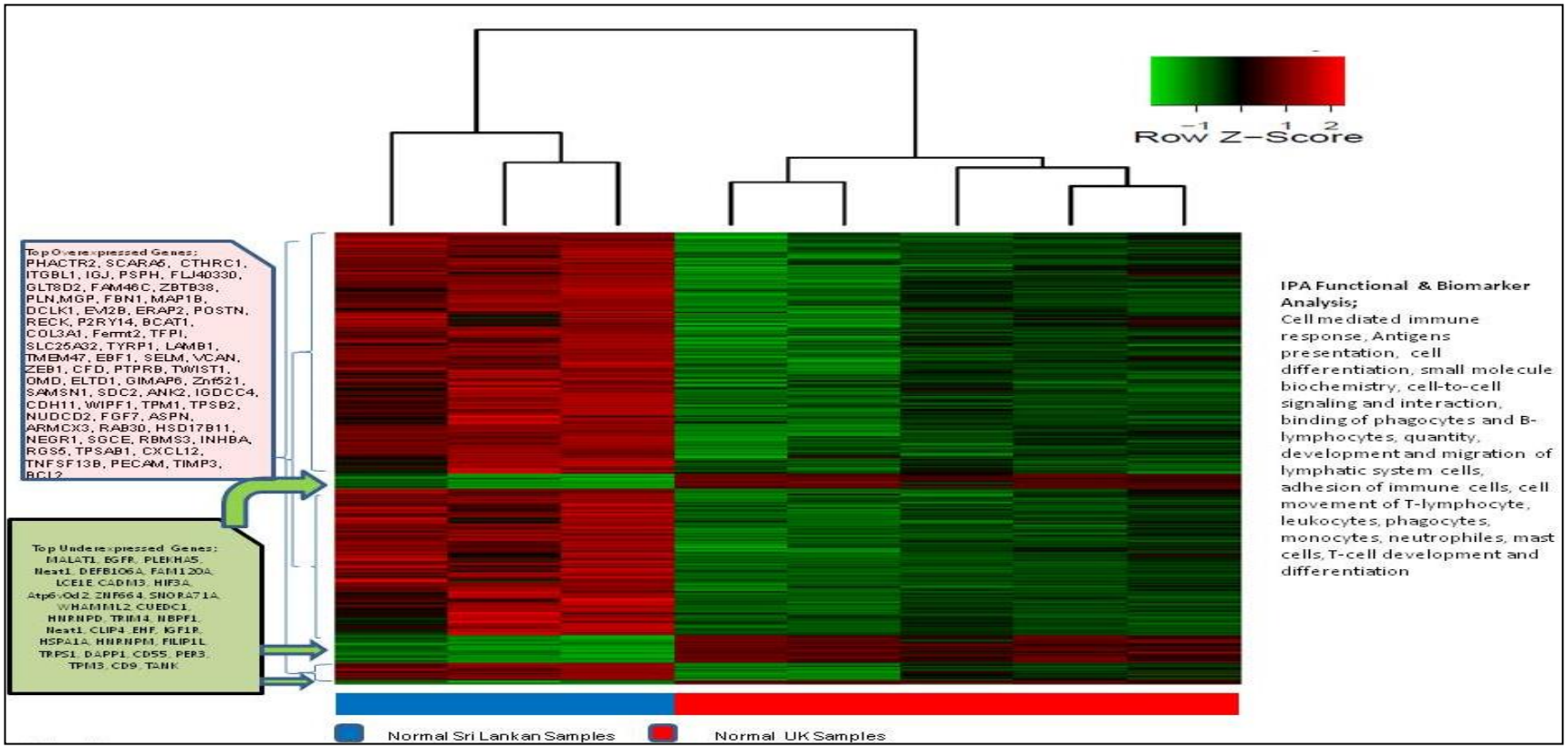


Figure 4.18: Hierarchical clustering of 1439 differentially expressed genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples using SAM supervised analysis at FDR≤0.01.

Red and blue bars represent UK and Sri Lankan normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low, black= not changed) is represented on a log2 mean- centred scale. All gene expression values were either above or below fold change magnitudes equal to ±1.5.

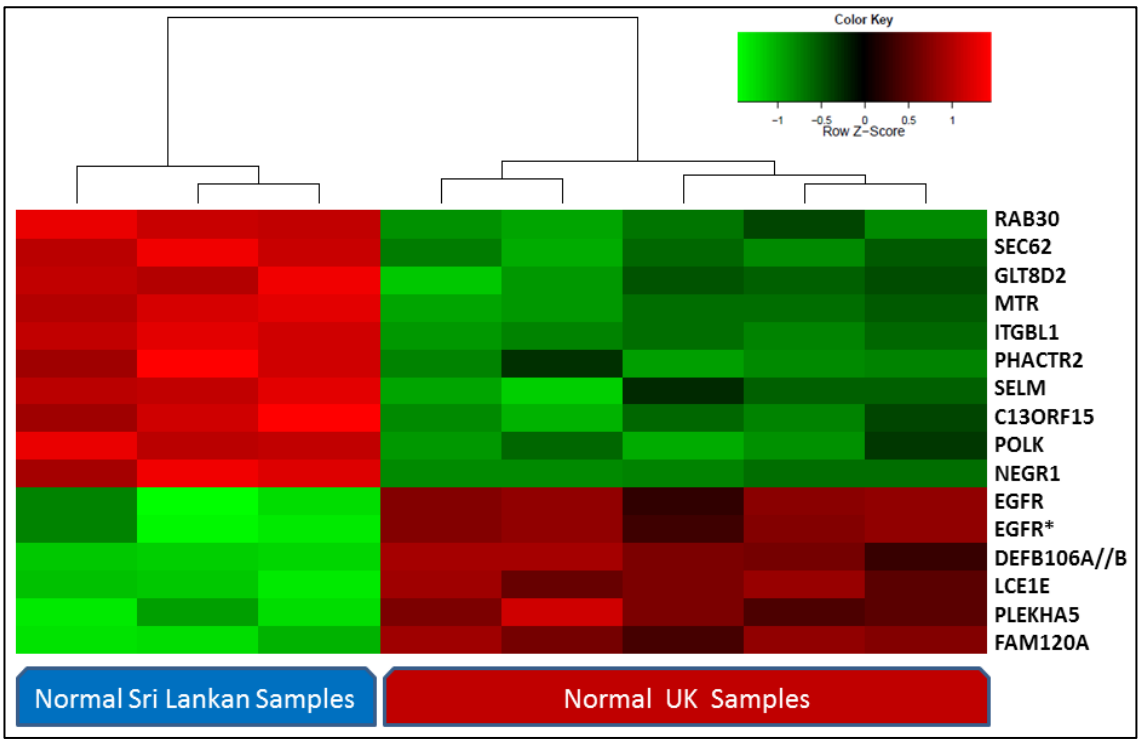


Figure 4.19: Hierarchical clustering of 16 differentially expressed genes separating 3 Sri Lankan and 5 UK normal oral mucosal samples using SAM supervised analysis at FDR≤0.005.

Red and blue bars represent UK and Sri Lankan normal samples, respectively. Heatmap represents mean-centred expression of genes in rows and samples in columns. Expression (red= high, green= low) is represented on a log2 mean- centred scale. All gene expression values were either above or below fold change magnitudes equal to 2.

Running biomarker analysis identified specific biomarkers related to OSCC and HNSCC. This include upregulation of COL1A1, COL4A1, H3F3A/B, BCL2, FGF2, ATM, IGF1, MGMT, FZD7, LGALS1, SPARKL1, CXCL12, TIMP3, MIB1, NBN, NTRK2, and TMEM38B along with downregulation of EGFR and IGF1R.

Elevated oncogenes were observed in normal samples from Sri Lanka comparing to those from UK population group. Example are; BCL2, RUNX1, AXL, ERBB2, RAS, MYC, HTERT, KIT, IGF1, MGMT, TRK, and SRC oncogenes plus other genes strongly involved in tumourigenesis like POSTN, RECK, COL1A2, COL1A3, COL5A2, WISP1, AGR2.Noticeably, suppressed oncogenes receptors, including EGFR, IGF1R, RARA, EPOR, THRA, and NMUR2 receptors, were observed which might be explained as expected biological tolerance modulation response of the body to the long term elevated expression level of oncogenes. This finding could suggest

elevated body threshold to oncogenic stimulus as protective mechanism of normal cell against tumorigenic signalling. As Sri Lankan normal individuals did not chew Betel quid, Curcumin consumed by Sri Lankan people for centuries as a frequent dietary component could play a role in the above observation (Pirasath et al., 2010, Jayasekera et al., 2004). Curry and turmeric have already been reported as anti-tumour diets which boost immune surveillance against tumour progression. Curcumin (diferuloylmethane) is reported to downregulate various growth regulatory pathways and specific genetic targets including genes for NFκB, STAT3, COX2, AKT, anti-apoptotic proteins, growth factor receptors, and multidrug-resistance proteins. The protective effects of Curcumin appear on its ability to hasten the activation of NRF2 and induce the expression of antioxidant enzymes (e.g., hemeoxygenase-1, glutathione peroxidase, modulatory subunit of gamma-glutamyl-cysteine ligase, and NAD(P)H:quinone oxidoreductase 1, increase glutathione (a product of the modulatory subunit of gamma-glutamyl-cysteine ligase), directly quench free radicals, and inhibit p300 HAT activity (Goel and Aggarwal, 2010).

A further observation is the upregulation of Selenoproteins in normal Sri Lankan samples including SELM, SELK, SEPP1, and SELT provides an additional independent explanation of the Sri Lankan population resisting tumour invasiveness. The increase in the production of reactive oxygen species (ROS) associated with low antioxidant activity has been related to several types of cancer. Selenium, an antioxidant micronutrient, may function as an anti-mutagenic agent, preventing the malignant transformation of normal cells. The protective effect of selenium is especially associated with its presence in the glutathione peroxidase and thioredoxin reductase enzymes to protect DNA and other cellular components against oxidative damage caused by ROS. The thioredoxin reductase (TrxR) selenoenzyme family is also capable of degrading hydroperoxides, as demonstrated by the reduction in hydrogen peroxide and is involved in numerous vital processes, such as DNA synthesis and regulation of apoptosis (Gromadzinska et al., 2008). Moreover, another result reported the role of selenium in inhibition of tumour growth and metastasis (Yoo et al., 2006). Several studies have shown reduced expression of these enzymes in various types of cancer, especially when associated with low intake of selenium,

which may increase the damage. Selenium supplementation appears to reduce the risk of some types of cancer by reducing oxidative stress and DNA damage (Almondes et al., 2010).

4.4.12 qPCR validation of Affymetrix GeneChip microarray gene expression data

Three genes (CLU, MAL, PTP313/FAP1) downregulated in both UK and Sri Lankan tumours were selected for further investigation and qPCR validation. The Box plot relative expression level of these 3 genes is shown in Figure 4.20.

Samples used in the qPCR were 4 Sri Lankan and 5 UK normal oral mucosa samples. These were compared to 12 samples each UK and Sri Lankan OSCC and were selected as representative of patient and tumour characteristics.

The qPCR data presented here were repeated and confirmed over four separate rounds of experimental determination consistently confirming the expression array findings that, these three genes are downregulated in OSCC compared to normal in both UK and Sri Lankan samples.

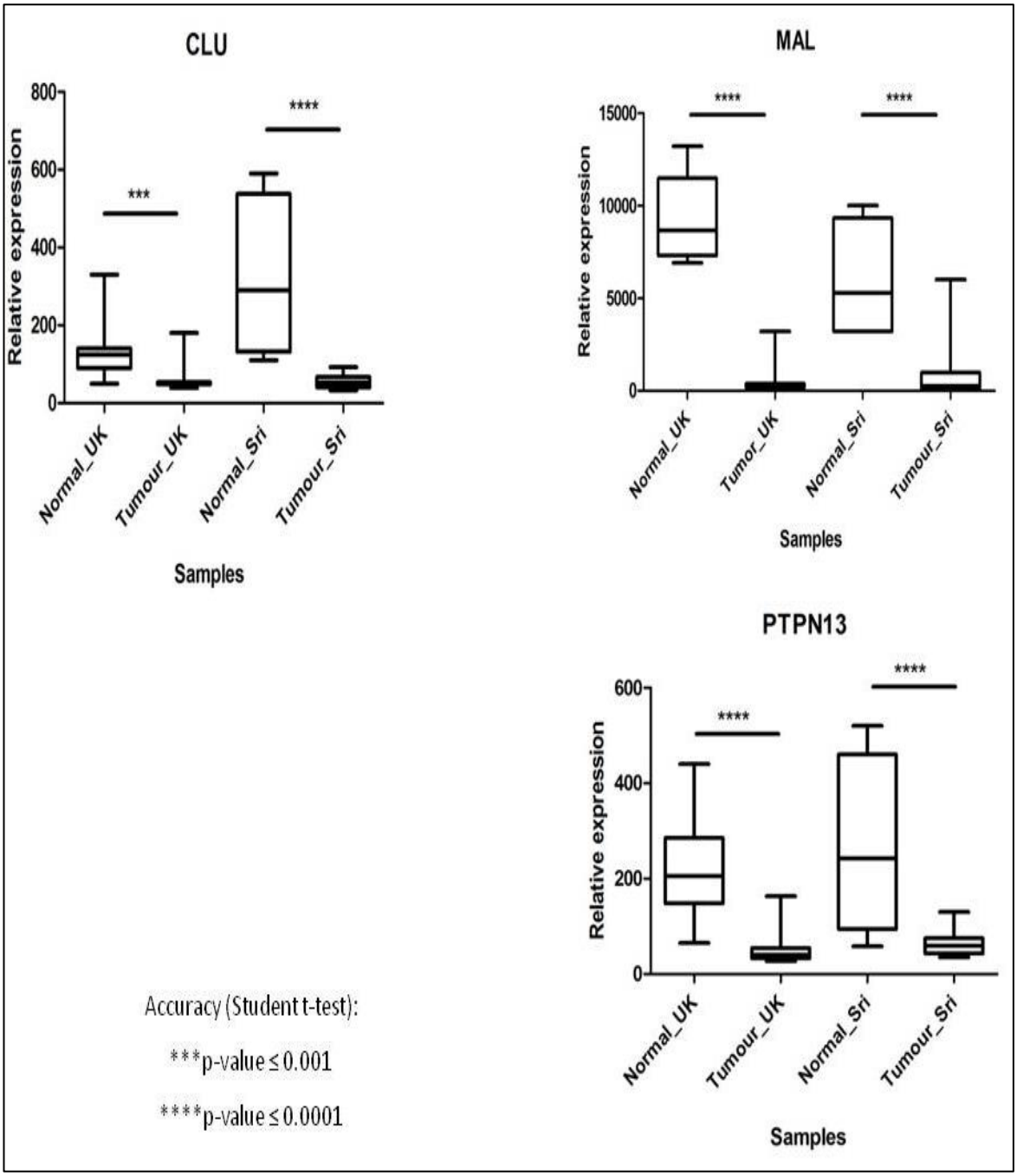


Figure 4.20: Box plots displaying results of qPCR analysis for the genes, CLU, MAL, and PTPN13 chosen for qPCR validation.

4.5 Discussion

4.5.1 Choice of normal comparator

The use of non-matched control tissue in this study was an attempt to overcome the problem of obtaining macroscopically normal tissue adjacent to a cancer that later was determined histologically abnormal; so called field cancerization. Such tissues would not display the gene expression pattern of normal mucosa and would be at increased risk of malignant transformation (Mao et al., 2004). It is suggested that there may be large gene expression differences between normal mucosa samples from different individuals that may then confound any attempt to identify expression differences between OSCC and normal mucosa.

Normal mucosa samples were clearly discriminated from OSCC samples in both supervised and unsupervised analyses. Whilst there are gene expression differences between UK and Sri Lankan normal mucosa, the normal mucosa samples all clustered together in an unsupervised comparison of all samples indicating a broad similarity of expression profiles.

Many previous studies used matched normal oral mucosa from the same patient to derive discriminatory expression profiles as it is suggested that using matched controls reduces variation in expression seen between individuals (Ziober et al., 2006). However, many patients with OSCC exhibit changes of the mucosa in many parts of the mouth with no clinical observations. Such lesions frequently display epithelial dysplasia and a proportion will transform to carcinoma (Mashberg, 1977). This observation gave rise to the concept of "field cancerization" (Slaughter et al., 1953).

Frequently second primary cancers that occur in OSCC share genetic similarities to the first primary tumour; a proportion of such similarities are also seen in normal tissue within field cancerization area adjacent to a cancer (Mao et al., 2004). This finding therefore raises a crucial question of how normal is "normal" oral mucosa in a patient with OSCC.

4.5.2 Data uploading, Functional, Networks, Canonical pathways analysis using IPA system

It is worth notable here to mention a problem experienced during uploading the gene sets to IPA system. Example for this is the uploading of 1439 genes differentiating between UK and Sri Lankan normal mucosa to IPA system resulted in 95 unmapped genes. As mentioned in previous chapters, IPA cannot map pseudogenes, hypothetical genes, ESTs and ambiguous (mapping to multiple) genes as well as duplicates of multiple IDs snapped in one molecule or multiple molecules snapped in one ID. After re-arrange the unmapped genes by converting them using David conversion tool (Huang et al., 2009a, Huang et al., 2009b) and deletion of duplicates Ids, gene names / symbols, 54 mapped out of the 95 unmapped genes were obtained, end with total of 1398 genes recognized by IPA, with 1154 eligible genes for network analysis and 982 eligible genes for functional, pathway, and list analysis .Some of these genes are substantial key players contributed in the tumorigenic process and in the master networks of the analysis. The same procedure was repeated in all gene sets for study comparisons. After communication with IPA technical support, the system seems override this problem with the new version release as the subsequent IPA analysis experienced very rare unmapped genes in the system which not recognized in other genomic tools as well. However, double check of the unmapped genes before proceed the analysis is recommended as some of these genes might be study targets.

4.5.3 Gene expression profile of tumour versus normal samples from UK

Several differentially expressed genes between tumour and normal tissues from UK are related to arachidonic acid (AA) metabolism, such as S100A10, glutathione peroxidase-3 (GPX3), phosphatidylcholine transfer protein, aldo-keto reductase family 1 (AKR1A1), and COX2. Arachidonic acid is metabolized in mammalian cells by a free radical reaction to cyclic endoperoxides via the cyclooxygenase system (Hamberg et al., 1974). The cyclooxygenase pathway converts arachidonic acid to prostaglandins and thromboxanes. These metabolites regulate diverse intracellular and extracellular biological processes in various cell types via specific

receptor molecules (Hammarstrom, 1983). Prostaglandins produced by COX2 may play an important role in the tumorigenic process of OSCCs from UK.

Previous epidemiologic studies have reported the association between the frequent usage of non-steroidal anti-inflammatory drugs (NSAIDs) and the decreased risk of oesophageal cancer (Shureiqi et al., 2001) ,the main mechanism detected is the inhibition of COX2 activity. UK tumour samples showed significant over-expression of COX2 consistent with previous studies (Banerjee et al., 2002, Nathan et al., 2001, Atula et al., 2006, Lee et al., 2002, Chan et al., 1999, Chang et al., 2004, Mohan and Epstein, 2003, Wang, 2005, Tsai et al., 2004, Ye et al., 2008, Ratnasinghe et al., 1999, Zimmermann et al., 1999).

Further study indicates a key role for arachidonic acid metabolites in metastasis. The study determined that intracellular pathways involving metabolites of arachidonic acid are required for both tumour cell invasion through basement membranes in vitro, and for their metastasis in vivo and are required for the invasive activity of several lines of malignant cells (Reich and Martin, 1996).

4.5.4 Gene expression profiling of Sri Lankan OSCCs compared to normal oral tissues

Higher expression of CD80, IL1RL1 which mediate anti-tumour immunity (Tajima et al., 2003, Thomas and Wen, 2006, Vesosky and Hurwitz, 2003, Chen et al., 1993), and GML upregulation which play a role in tumour suppression (Kimura et al., 1997) along with downregulation of PSPH and TWIST1 which their over expression is critical in SCC proliferation and metastasis (Yang et al., 2008, Yang and Wu, 2008, Yuen et al., 2007, Niu et al., 2007, Bachelor et al., 2011) and might presumably contribute in resisting tumour invasion and metastasis of Sri Lankan tumour cases. Additionally, several genes and transcription regulators which reported a role in tumour invasive behaviour, recurrence, and metastasis were upregulated in UK but not Sri Lankan tumour cases relative to their normal comparators. These include COX2, POSTN, MMP3, and LAMC2 gene biomarkers of OSCC along with activation of ESR1, SMAD3, and SMAD4 transcription regulators.

The most relevant activated and inhibited functions using IPA system are reported in Figure 4.8 at FDR<0.001. Cell cycle progression, tumour development and colony formation, epithelial cell proliferation, and cell ploidy and mitosis, appeared to be in the top of highest-level functions. Altogether these functions show no correlation with aggressive behaviour and metastasis of cancer which might influence and confirm the phenotypic behaviour of Sri Lankan OSCCs being less invasive with no metastasis or recurrence of the disease reported.

Most cancer types are similar in expressing mitochondrial dysfunction and elevated substrate level Phosphorylation, when loss of respiration leads to glycolysis, dedifferentiation, and unbridled proliferation (Szentgyorgyi, 1977). Mitochondrial mutations may contribute to the development of a malignant phenotype by direct effects from increased reactive oxygen species production as well as induction of aerobic glycolysis and growth promotion (Zhou et al., 2007). While normal mitochondrial function can reduce tumourigenesis, various studies show that tumour mitochondria are structurally and functionally abnormal and unable to generate normal levels of energy (Seyfried and Shelton, 2010).

4.5.5 Common gene expression profile of tumour versus normal samples from both UK and Sri Lanka

Among the common upregulated genes is SPP1 which is described essential in the pathway leads to type I immunity. Osteopontin is expressed and secreted by various cells, and has a role in cell adhesion, chemotaxis, prevention of apoptosis, invasion, migration and anchorage-independent growth of tumour cells. It has also been shown that SPP1 is upregulated in cervical carcinoma (Wong et al., 2006) and associated with progression from premalignancy to OSCC (Devoll et al., 1999). Although previous study has demonstrated that tumour derived SPP1 is able to inhibit macrophage function and enhance survival of SCC and metastases (Crawford et al., 1998), further study concluded that elevated expression of SPP1 was by DNA damage-induced Tp53 activity and by adenovirus-mediated transfer of the human TP53 gene and this gene has a functional Tp53-responsive element in its promoter region which suggests SPP1 a direct transcriptional target of TP53. TP53-directed

regulation of SPP1 expression suggests a novel model of TP53 participation in immunosurveillance, involving interaction with the host immune system to prevent damaged cells from undergoing malignant transformation (Morimoto et al., 2002).

MMP10 was also upregulated in both populations' tumours which already reported to promote invasion in HNSCC and that MMP10 mediates the Periostin and Wnt-5b-induced invasion (Deraz et al., 2011). Moreover, Wnt-5a which is upregulated in UK samples reported being involved in the invasion of various cancers including melanoma, breast cancer, gastric cancer, pancreatic cancer and osteosarcoma (Weeraratna et al., 2002, Pukrop et al., 2006, Kurayoshi et al., 2006, Ripka et al., 2007, Enomoto et al., 2009, Yamamoto et al., 2009, Yamamoto et al., 2010). Notably, some MMPs (upregulated in this study) are involved in Wnt-5a mediated invasion of cancer cells (Kurayoshi et al., 2006, Enomoto et al., 2009, Dissanayake et al., 2007, Masckauchan et al., 2006).

Previous studies reported MAL gene frequently extinguished in oesophageal carcinoma and other gastrointestinal cancer tissues relative to normal corresponding epithelium. The anti-tumour effect of MAL was confirmed in vitro and in vivo (Mimori et al., 2003). Recent study showed that MAL is able to suppress malignant phenotypes of OSCC, such as proliferation, apoptosis and invasion, and inhibit tumorigenicity *in vivo* by transient and stable transfection and suggested MAL might be a candidate for tumour suppression gene (Cao et al., 2010). Previous microarray studies showed frequent MAL repression in OSCC and HNSCC samples which support the existing findings (Alevizos et al., 2001, El-Naggar et al., 2002, Mendez et al., 2002, Toruner et al., 2004, Belbin et al., 2005, Tomioka et al., 2006, Ye et al., 2008, Choi and Chen, 2005).

NFE2L2 showed the lowest expression in tumours compared to normal controls among the transcriptional regulators distinguished in both population groups. NFE2L2 induces the expression of various genes including those that encode for several antioxidant enzymes, and it may play a physiological role in the regulation of oxidative stress. Genetic polymorphisms in glutathione peroxidase enzymes and their altered expressions and activities are associated with oxidative DNA damage and, as

a result, the individual's risk of cancer susceptibility (Itoh et al., 1999). Indeed, GPX3, whose main biological role is to protect the organism from oxidative damage, was massively downregulated in tumour cases of both populations.

Inhibition of NR3C1regulator was another finding, a receptor of Glucocorticoids. The major signalling pathway used by Glucocorticoids receptors is via direct DNA binding and transcriptional regulation of target genes. They can also signal by binding to other proteins, mainly with transcription factors such as NFκB, AP-1 or STAT (Lu et al., 2006, Rhen and Cidlowski, 2005, Pratt et al., 2006, Ray and Prefontaine, 1994, Lerner et al., 2003, Zhang et al., 1997, Stocklin et al., 1996). It's noteworthy that Glucocorticoids enhance glucagon action (McMahon et al., 2010) which inhibits glycolysis (a significant molecular function activated in this study) and mobilizes fats and stress response. Suppression of NR3C1 may implement in the tumorigenic process for both populations of the study.

As molecular relationships represented on the network include not only activation or inhibition of expression, but also protein-protein interactions, DNA-protein interactions and activation, localization, inhibition of the corresponding proteins (Debily et al., 2009), it is not surprising that microarrays may fail to identify some of the hub genes significantly differentially expressed among the tumour versus normal samples. These genes might play a major role through protein activation for instance. Alternatively, their modulations may be very subtle and below the threshold for reliable detection of differences of the microarray platform despite its high sensitivity.

4.5.6 Gene expression differences between UK and Sri Lankan tumour samples

It is notable that the overall number of gene expression differences between UK and Sri Lankan OSCC is smaller than that observed in either of the two other supervised analyses comparing each tumour population to its normal, suggesting that a greater gene expression differences between the cancer samples and their respective normal comparators than between the two cancer populations. IPA core analysis revealed activated immune response functions in Sri Lankan relative to UK tumours. These

include; development and mitosis of thymocytes; activation, development, differentiation, cell movement, migration, and chemotaxis of T, B-lymphocytes, leukocytes, monocytes, and other immune system cells. Curcumin consumed by Sri Lankan people for centuries as a frequent dietary component could play a role in the above observation (Pirasath et al., 2010, Jayasekera et al., 2004). More discussion will be conducted under normal samples comparison of the two population groups. These findings might have an impact on the significant differences in the clinical behaviour of the two tumour groups which mainly reflected in resistant of Sri Lankan patients to the invasiveness, metastasis, and recurrence of disease as well as survival of the patient with lower morbidity and mortality relative to UK patients.

4.5.7 Gene expression differences between UK and Sri Lankan normal mucosa and its impact on the observed phenotypic differences of the two tumour populations

First observed finding is suppressed oncogenes receptors, including EGFR, IGF1R, RARA, EPOR, THRA, and NMUR2 receptors in Sri Lankan normal controls. Further potential finding is the highly activated cell mediated immune response in Sri Lankan relative to UK normal controls. As Sri Lankan normal individuals did not chew Betel quid, as previously mentioned (4.3.1), frequent consumption of curcumin by Sri Lankan people could play a role in the above observation (Pirasath et al., 2010, Jayasekera et al., 2004). Curry and turmeric have already been reported as anti-tumour diets which boost immune surveillance against tumour progression. Curcumin (diferuloylmethane) is reported to downregulate various growth regulatory pathways and specific genetic targets including genes for NFκB, STAT3, COX2, AKT, anti-apoptotic proteins, growth factor receptors, and multidrug-resistance proteins. The protective effects of Curcumin appear mediated through its ability to induce the activation of NRF2 and induce the expression of antioxidant enzymes (e.g., hemeoxygenase-1, glutathione peroxidase, modulatory subunit of gamma-glutamyl-cysteine ligase, and NAD(P)H:quinone oxidoreductase 1, increase glutathione (a product of the modulatory subunit of gamma-glutamyl-cysteine ligase), directly quench free radicals, and inhibit p300 HAT activity (Goel and Aggarwal, 2010).

Genes detected are involved predominantly in processes such as cell mediated immune response, antigens presentation, cell differentiation, small molecule biochemistry, cell-to-cell signalling and interaction, binding of phagocytes and B-lymphocytes, elevated quantity, development and migration of lymphatic system cells, adhesion of immune cells, cell movement of T-lymphocyte, leukocytes, phagocytes, monocytes, neutrophils, mast cells, T-cell development and differentiation (Figure 4.18). The above observation might underlie differences in clinical behaviour of the two tumour groups. Indeed, the highly activated Cell-mediated Immune Response in Sri Lankan normal group presumably plays a key role in resistant of Sri Lankan patients to the invasiveness, metastasis, and recurrence of disease as well as survival of the patient with lower morbidity and mortality relative to UK patients. Previous studies on HNSCC suggested that a peri-tumoural lymphocyte infiltrate was associated with a better prognosis when compared with tumours not exhibiting immune infiltration (Wolf et al., 1986, Hiratsuka et al., 1984, Guo et al., 1987, Hirota et al., 1990, Slootweg et al., 1994). Furthermore, higher peri-tumoural lymphocyte infiltration correlated with lower stage and less invasive tumour growth (Guo et al., 1987, Slootweg et al., 1994). Other studies concluded that tumour-infiltrating lymphocytes in HNSCC acquire defects in T-cell receptor signalling and subsequent apoptotic death (Kuss et al., 1999, Gastman et al., 1999). Interestingly, circulating T lymphocytes from individuals with HNSCC exhibit defects in tumour directed cell killing suggesting that immune function modulation in the local tumour bed may extend to systemic immune system disruption in individuals with HNSCC (Reichert et al., 2002). Ginos group noticed that their tumour samples demonstrated a marked absence of an immune response signature suggesting that modulation of tumour-specific immune responses may play a role in local treatment failure of HNSCC (Ginos et al., 2004).

4.5.8 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK

Three genes appeared upregulated in perineural invasive tumours; PGM1, ENO2, and ILK. ENO2 (enolase2) reported a prognostic biomarker in metastatic, high grade, and invasiveness of various types of cancer, including breast, lung, bladder, renal, prostate, colorectal cancer, and OSCC (Abba et al., 2005, van Zandwijk et al., 1992, Di Carlo et al., 1990, Vandepol et al., 1994, Buzaid et al., 1994, Rasmuson et al., 1993, Jacobsen et al., 1990, Nishihara et al., 2009, Zufferey et al., 2001, Famulski et al., 2001). ILK has been associated with multiple cellular functions including cell migration, proliferation, and signal transduction. Recent studies reported ILK as a biomarker associated with laryngeal, lung, colon, ovarian cancer as well as metastatic behaviour of gastric carcinoma (Ito et al., 2003, Wu et al., 2006, Takanami, 2005, Bravou et al., 2006, Du et al., 2008, Loessner et al., 2009).

GO term (Gene Ontology, www.geneontology.org/) showed both PGM1 and ENO2 enzymes involved in magnesium ion binding whereas the all three genes involved in phosphorylation and interconversion activities of certain proteins and amino acids like phosphopyruvate hydratase, lyase, phosphoglucomutase, isomerase, and phosphotranferase.

4.5.9 Gene expression profile of high-stage versus low-stage tumour samples from both UK and Sri Lankan tumours

SAM supervised analysis identified 12 genes separating “high stage” from “low stage” OSCCs of both population groups. Most of genes appeared upregulated in high stage tumours and associated with different cancer types and cellular processes associated with tumours.

Most of differentially upregulated genes in high stage tumours exhibited involvement in variant biological processes crucial in driven tumour lesion to late stage and metastasis with distinct roles in cellular migration, remodelling, outgrowth, invasion, differentiation, colony formation, cell cycle progression, activation, proliferation, and angiogenesis.

Knockdown of SEMA3F (member of semaphorins) cancer cells of breast rescued the aggressive phenotypes and tumour invasion by inducing suppressive cell microenvironment (Xiong et al., 2012) as well as pancreatic (Matsushita et al., 2007). An extensive review explained how Scatter factors and semaphorins, together with their receptors, help to orchestrate metastasis programme of neoplastic cells (Trusolino and Comoglio, 2002).

Suswam group (Suswam et al., 2005) noticed that KHSRP involved in IL8 RNA stabilization, which plays an integral role in promoting the malignant phenotype in breast cancer (Chen et al., 2001, Gherzi et al., 2004).

Jauliak group identified NFAT5 isoform expressed in invasive human ductal breast carcinomas and participate in promoting invasion of human breast and colon carcinomas. The transcriptional activity of NFAT5 is induced by alpha-6 and beta-4 integrin signalling clustering in the presence of chemo-attractants, resulting in enhanced cell migration and is involved in promoting carcinoma invasion (Jauliac et al., 2002). Further study demonstrated NFAT5 contribution in progression of gastrointestinal cancer (Chen et al., 2011b).

CDC42BPA plays a role in the regulation of cytoskeleton reorganization and cell migration (Tan et al., 2008a). It has been demonstrated that invasion can be generated by Cdc42-MRCK signalling (Wilkinson et al., 2005) as well as tumour progression (Benitah et al., 2004).

4.6 Conclusion

The gene expression profiles of UK and Sri Lankan OSCC compared to normal oral mucosa are similar in many respects to other oral cancer expression profiles reported in the literature and were mainly similar to each other. Great differences in the expression profiles between OSCC and normal tissue than between OSCCs of the two populations was observed. However, molecular heterogeneity in each tumour populations (such heterogeneity is more marked in UK OSCC, which displays a wide range of clinical behaviour). Such molecular heterogeneity may act to reduce differences seen in expression profiles of the two populations. Despite that, IPA showed common perturbed canonical pathways which indicate that the development of OSCCs is mediated, to some extent by similar biological pathways despite the differences related to race, ethnicity, lifestyle, and/or exposure to environmental carcinogens.

IPA comparison core analysis of normal as well as tumour samples between the two populations displayed highly activated "Cell-mediated Immune Response" processes in Sri Lankan normal and tumours relative to UK cohorts which presumably play a key role in confrontation of Sri Lankan patients to the invasiveness, metastasis, and recurrence of the disease. Sri Lankan consumption of spicy food, including Curcumin/Curry in specific (Jayasekera et al., 2004, Pirasath et al., 2010), might influence this activation along with downregulation of growth factor receptors (EGFR, IGFR, RARA, EPOR, THRA, and NMUR2) that might as well, contributed in the phenotypic behaviour of the two tumour groups. The same findings were observed in comparison of the tumours from the two population groups which confirms the constant existence of the above events between the two populations.

Smaller number of either elevated or repressed genes in either UK or Sri Lankan OSCC but not both contributed to a number of differentially regulated cancer-associated pathways and specifically to the aggressive behaviour of UK tumours; including upregulation of MMP3, POSTN, LAMC2, and downregulation of DUSP1 and SLPI in UK. Conversely, CD80 upregulation and TWIST1 downregulation, anti-invasive associated incidents, observed in Sri Lankan tumours. It may be the case

that expression differences in key pathways contribute to the observed phenotypic differences and so expression differences described here may underpin some of the phenotypic differences between UK and Sri Lankan OSCC. Tumour heterogeneity, technical variation and data processing may all have an effect on the expression profiling data and all such variables should consider when interpreting the results.

Differences in tumour expression profiles of the two populations highlighted a few potentially genes; including lower expression of CDKN1B in UK tumours, a phenomenon associated with high-risk tumours only, along with lower expression of PSIP1, SAMSN1, CCNG2, and NCOA2 genes which played a protective role in tumorigenic process. Likewise, higher expressions of CCT2 and DHCR24 in UK tumours were contributed in rapidly proliferating tumour and melanoma metastases respectively.

Numerous results displayed in this study support the theory that considers cancer as a metabolic disorder (Warburg effect) (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956). The top perturbed pathways in both populations and each separately showed that defects in mitochondrial respiration could be the primary reason for induction of tumorigenic transformation reflecting mitochondrial dysfunction and oxidative phosphorylation pathways shifted the tumour cell to glycolysis pathway affecting Citrate cycle with remarkable dysregulation of fatty acids metabolism. As subsequence to the above defects, significant alteration in gene expression can be explained as an attempt from the tumour cells to cope with the initial disorder, repair, and overcome the fatal causative event(s). Further studies with a large size and wide range of populations are required to highlight and validate the above phenomena.

Chapter 5: Biological pathway integration of significant metabolomic and transcriptomic signatures of OSCC

5.1 Introduction

Cancer may be accompanied by the production and release of a substantial number of proteins, metabolites and/or hormones into the blood, saliva, and other body fluids that could serve as useful markers for assessing prognosis, metastasis, monitoring treatment, and detecting malignant disease at an early stage. Metabolic profile of variant biofluids, including serum, saliva, urine, and cerebrospinal fluid can be changed by different physiological processes following pathophysiological provocations, hence, comprehensive perturbation in these profiles may exhibit the presence of a specific disease. Metabolomic profiles for early detection as well as stages identification of OSCC by employing numerous methods of nuclear magnetic spectroscopy and mass spectrometry were reported to provide a coherent outlook of the wide-ranging metabolic body response to pathophysiologic effect and genetic modification.

OSCC is a complex disease with a wide heterogeneity involves multiple tissues in the oral cavity. Although molecular profiles of the disease from different tissue sites remain very similar with common biomarkers, limitation was experienced to find a unique signature with a satisfactory specificity and sensitivity of the disease. Limitation of both specificity and sensitivity is a consequence of the fact that metabolic profile is influenced by various environmental factors and physiological status of the body in addition to the biological alteration arisen from the disease.

The metabolism of OSCC reported to reflect a specific interaction arises from the biochemical status of the oral tissue and the biological condition accompanied tumourigenesis of OSCC. At present, the lack of metabolic signature required for absolute detection of different types of cancer as well as other inflammatory conditions, and physiological disorders shrinks the chance of the clinical applicability of this signature as a tool to distinguish OSCC from other type of cancer.

Linking metabolic dysfunctions to altered gene expression profiles can provide new insights into the regulatory network underlying the metabolic dysfunctions, enabled the assembly of discrete gene expression events into functional pathways. However, a major challenge in the integration of 'omics data (transcriptomics, proteomics, and metabolomics) is how best to compare and correlate these large datasets to provide the most meaningful biological context. A considerable number of studies attempted to address this challenge both at the data level, using statistical and chemometric methods, and at the biology level, using pathway analysis (Chou et al., 2003, Coen et al., 2003, Martin et al., 2003, Clish et al., 2004, Eerligh et al., 2004, Griffin et al., 2004, Go et al., 2005, Griffin, 2006, Fan et al., 2006, De Vos et al., 2007, Rezzi et al., 2007, Jalanko et al., 2006, Bylesjo et al., 2008, Gomase and Tagore, 2008, Phillips and Foster, 2008). The latter approach uses extensive databases of metabolite and gene interactions to link known metabolites and genes to relevant pathways and phenotypes, *i.e*. mapping genotype to phenotype.

5.2 Aims and objectives of the study

In this review study, an attempt was carried out to gather common differentially expressed metabolites across studies which proposed as biomarkers of OSCC. Subsequently, integration of the metabolic signature from the review with the previously harvested transcriptomic biomarker from the met-analysis study (chapter 3) was carried out in an attempt to explore more specific signature of OSCC and understand the tumorigenic process reflected by the simultaneous perturbed genetic and metabolic system using IPA biological tools. The concept behind integration approach is that different biological diseases and disorders reflecting same metabolic profiles presumably have different transcriptomic profiles. Likewise, those that give similar transcriptomic profiles expected to have different metabolic profiles. From this insight, merging both metabolic and transcriptomic profiles is carried out (data level integration). Furthermore, IPA tools were used to identify the significant biological relationships between frequently altered metabolites in OSCC from the review study and meta-genes biomarkers of OSCC harvested from chapter 3, while the shared cellular biofunctions, functional networks, canonical pathways, and upstream regulators differentiating OSCC from normal oral mucosa were identified

from the interactive network. Exploring IPA results revealed remarkable synergetic and harmonious interaction between altered gene expression during tumourigenesis and the subsequent body response coincided in its intermediate and the end product metabolites (biological level integration).

5.3 Materials and Methods

Published reports of metabolomic expression profiles of OSCC and HNSCC were identified using Medline database search. A review of the reports was performed to identify metabolites that repeatedly exhibiting altered expression in OSCC and HNSCC compared to normal oral mucosa.

5.3.1 Construction of common metabolites list

Regularly expressed metabolites in OSCC across the review studies were collected and unified in one metabolite identification number using human metabolome database (HMDB) (http://www.hmdb.ca/), a freely available electronic database containing detailed information about small molecule metabolites existed in the human body.

5.3.2 Biological annotations of the metabolic profile

Differentially expressed metabolites in OSCCs were subsequently examined using available tools from IPA to define the relative canonical pathways, biomarkers, and molecular processes that most commonly perturbed in the tumorigenic transformation of OSCC and HNSCC, as well as, to inspect the most significant pre-generated networks associated with the uploaded group of metabolites including those related to OSCC development. IPA knowledge base recruits the distinguished metabolites that regulated OSCC and HNSCC to generate and highlight the above functions. Canonical pathway analysis identified altered pathways from IPA library which are most significant to the input data set. The significance of the association between the data set and the canonical pathway was determined based on two parameters: (1) A ratio of the number of metabolites from the data set that map to the pathway divided by the total number of metabolites that map to the canonical pathway and (2) A p-value calculated using Fischer's exact test determining the

probability of the association between the metabolites in the data set and the canonical pathway is due to chance alone.

5.3.3 Integrating metabolomic and transcriptomic profiles

To assist with cross platform integration, IPA "Metabolomics Analysis" was used to interpret both metabolomic and transcriptomic data related to OSCC that harvested from the metabolic review in this chapter and the meta-analysis study in chapter 3. IPA provides biological context to the metabolomic data, links metabolic changes to disease-relevant pathways and phenotypes and assists in the selection of a subset of metabolomic markers for further study.

IPA-Metabolomics is powered by the Ingenuity Knowledge Base, a repository of molecular interactions, regulatory events, signalling and metabolic pathways, and gene to phenotype associations that provide the building blocks for pathway construction. IPA contains millions of findings from the full text of the life sciences literature that describe relationships between chemicals, proteins, genes, complexes, cells, cellular components, tissues, drugs, cellular processes, diseases and clinical phenotypes. IPA also has extensive libraries of metabolic and cell signalling pathways, an extensive contextual details, including species specificity, localization, mutations, epigenetic modifications, and experimental conditions. The structured detailed content in the knowledge base enables researchers using IPA to analyse and interpret combined 'omics data', visualize metabolite and gene interactions, and place data in a proper biological and chemical context.

5.4 Results

5.4.1 Construction of the metabolic profile set

From 1996 to 2011, 10 studies recruiting MR spectroscopy, 1HNMR, and MS spectral analyses to examine a wide metabolic expression changes associated with the development of OSCC and HNSCC were reviewed (Mukherji et al., 1996, Mukherji et al., 1997, Adalsteinsson et al., 1998, Star-Lack et al., 2000, El-Sayed et al., 2002, Bezabeh et al., 2005, Arias-Mendoza et al., 2006, Yan et al., 2008, Zhou et al., 2009, Tiziani et al., 2009, Sugimoto et al., 2010, Wei et al., 2011).

Table 5.1: Characteristics of metabolomic studies.

Studies	Sample description	Technology used
Mukherji et al., 1996	18 HNSCC and 13 normal muscle tissue specimens.	In vitro analysis of the one-dimensional proton MR spectroscopy
Mukherji et al., 1997	19 HNSCC tissue specimens, 13 normal tissue, and 3 metastatic cervical lymph nodes	In vitro 1-D and 2-D correlated proton MR spectroscopy (11 T)
	7 HNSCC and 7 healthy controls.	In vivo 1-D proton MR spectroscopy (1.5 T)
Star-Lack et al., 2000	14 HNSCC, 6 tissue normal samples	1HNMR
El-Sayed et al., 2002	85 HNSCC, 50 NORMAL tissue samples (from 40 patients)	1HNMR
Bezabeh et al., 2005	58 HNSCC and 22 normal tissue samples	1HNMR
Yan et al., 2008	Saliva samples were obtained from 20 OSCC, 20 OLP and seven OLK patients, and 11 healthy donors.	HPLC/MS analysis
Zhou et al., 2009	Plasma samples from 32 OSCC patients, 28 control group	1HNMR
Tiziani et al., 2009	Serum samples from 15 OSCC and 10 normal healthy controls	1HNMR
Sugimoto et al., 2010	Saliva fluid samples from 69 OSCC patients and 87 healthy individuals	CE-TOF-MS
Wei et al., 2011	Saliva fluid samples from 37 OSCC patients, 32 oral leukoplakia (OLK) patients, and 34 healthy individuals	UPLCQTOFMS

A set of 45 differentially expressed metabolites discriminated tumour samples from normal oral mucosa and selected for further subsequent analysis (Table 5.2). Different methodological techniques were used across studies to extract metabolomic signatures for OSCC and HNSCC disease from serum and saliva biofluids. These include ^{1}H and ^{31}P nuclear magnetic resonance spectroscopy (^{1}H and ^{31}P- NMR), Liquid chromatography–mass spectrometry (HPLC-MS), capillary electrophoresis time-of-flight mass spectrometry (CE-TOF-MS), and ultra-performance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry analytical methods. The results demonstrated a reasonable separation between the samples from healthy controls and patients with OSCC and HNSCC. Certain results went further to assign signatures discriminating advanced from early stage (Tiziani et al., 2009, Sugimoto et al., 2010, Wei et al., 2011) as well as malignant from benign

tumours (Adalsteinsson et al., 1998, Bezabeh et al., 2005, El-Sayed et al., 2002, Mukherji et al., 1997, Star-Lack et al., 2000). However, the current review focused on metabolic profiles discriminating OSCC from normal oral mucosa in an attempt to nominate frequently reported metabolic biomarkers of OSCC and HNSCC for the subsequent integration analysis with the transcriptomic biomarkers of the disease that will explain later in this chapter.

Metabolomic analysis of serum samples of OSCC patients was performed by nuclear magnetic resonance (NMR) spectroscopy with a satisfactory discrimination between OSCC and the healthy controls (Tiziani et al., 2009). Recently, a capillary electrophoresis time-of-flight mass spectrometry based salivary metabolomic analysis of oral, breast and pancreatic cancer has been performed and a number of metabolites related were identified (Sugimoto et al., 2010). Most recent work used an advanced version of LC-MS platform, ultra-performance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry (UPLCQTOFMS), to study salivary metabolomics of the three groups; OSCC, oral leukoplakia, and healthy subjects (Wei et al., 2011).

Table 5.2: 45 metabolites with altered expression in OSCC and HNSCC.*

Metabolite Name	Universal ID	Expression	Metabolite Name	Universal ID	Expression
Upregulated in OSCC RELATIVE TO NORMAL			downregulated in OSCC RELATIVE TO NORMAL		
glucose	HMDB00122	1	Valine	HMDB00883	-1
Ethanol	HMDB00108	1	Lactate	HMDB01311	-1
Methanol	HMDB01875	1	Alanine	HMDB00161	-1
Acetate	HMDB00042	1	Citrate	HMDB00094	-1
Pyruvate	HMDB00243	1	Phenylalanine	HMDB00159	-1
Acetone	HMDB01659	1	Tyrosine	HMDB00158	-1
Acetoacetate	HMDB00060	1	Formaldehyde	HMDB01426	-1
3-hydroxybutyrate	HMDB00357	1	formic acid	HMDB00142	-1
2-hydroxybutyrate	HMDB00008	1	Glycine	HMDB00123	-1
Choline	HMDB00097	1	Creatine	HMDB00064	-1
Betaine	HMDB00043	1	Carnitine	HMDB00062	-1
Dimethylglycine	HMDB00092	1	acetyle L-carnitine	HMDB00201	-1
sarcosine	HMDB00271	1	Creatinine	HMDB00562	-1
Asparagine	HMDB00168	1	Isoleucine	HMDB00172	-1
Ornithine	HMDB00214	1	Leucine	HMDB00687	-1
glutamic acid	HMDB00148	1	γ-aminobutyric acid	HMDB00112	-1
Taurine	HMDB00251	1	Threonine	HMDB00167	-1
lactic acid	HMDB01311	1	Proline	HMDB00162	-1
Piperidine	HMDB34301	1			
Putrescine	HMDB01414	1			
Cadaverine	HMDB02322	1			
Tryptophan	HMDB00929	1			
Phosphocholine	HMDB01565	1			
Glycerophosphocholine	HMDB00086	1			
Glutathione	HMDB00125	1			
Histadine	HMDB00177	1			
n-Eicosanoic acid	HMDB02212	1			

* +1 and -1 denoted to the elevated and repressed metabolites regardless the magnitude of the expressed metabolites.

5.4.2 Biological interpretation of the expressed metabolites

To obtain further biological understanding of why expressed metabolites in OSCCs are different from normal oral tissue samples, IPA tools was used to examine the dysregulated biological pathways, molecular functions, and processes. IPA system depends on the knowledge database of selected functional and regulatory interactions extracted from the literature and provides integrated graphical representation of the biological relationships between metabolites, considering both up- and downregulated metabolites from the comparison analysis.

5.4.2.1 Functional analysis of the common metabolites

"Amino Acid Metabolism, Molecular Transport, Small Molecule Biochemistry, Cell Cycle, and Protein Synthesis" functions were displayed at the top of the perturbed molecular and cellular functions involved in cancer.

5.4.2.2 Associated network analysis of the common metabolites

Investigating the biological relationships between genes and gene products, including metabolites, was carried out by performing network analysis of the commonly altered metabolites. The top molecular network having the highest score (49) among other networks was corresponded with "Amino Acid Metabolism, Molecular Transport, Small Molecule Biochemistry" functions (Figure 5.1). To visualize more comprehensive interactions between the differentially altered metabolites in OSCC, the top 4 significantly associated metabolic networks were merged (Figure 5.2). Networks 1, 2, 3, and 4 of the altered metabolites were chosen according to the highest score value (score = 49, 26, 13, and 2 respectively). The corresponding functions of the highest score networks (1, 2, and 3) were "Amino Acid Metabolism, Molecular Transport, Small Molecule Biochemistry" and "Cell Cycle, Hepatic System Development and Function, Cell-To-Cell Signalling and Interaction" respectively. These networks were constructed and merged through the overlapping genes (nodes with white colour) suggested from IPA knowledge base as hub nodes connecting the uploaded metabolites (Figure 5.2). In both top and master networks (Figure 5.1and Figure 5.2), Amino Acid Metabolism and degradation were the common detected functions. In addition, a group of metabolites altered in these studies and constantly expressed in both networks were observed reflecting common functions including "anti-oxidation, free radical scavenging, detoxification of xenobiotics and chemicals". These include metabolites of glutathione, creatine, glutamic acid, taurine, glycine, tryptophan, and pyruvic acid.

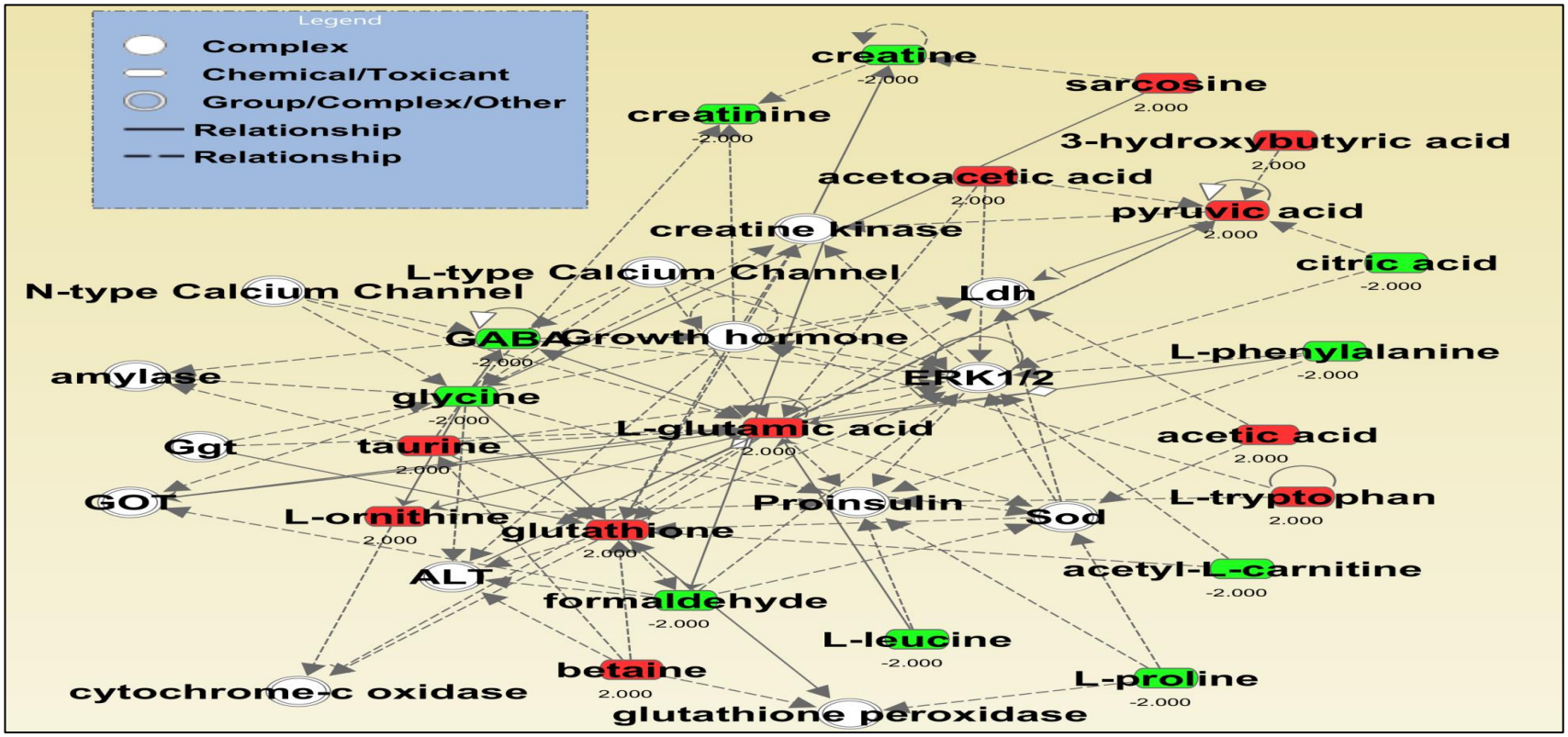


Figure 5.1: Top molecular network of the altered metabolites in OSCC.

The top molecular network identified by IPA tool (version 9.0) from up- and down-regulated metabolomic analysis using overlapping function. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed metabolites are shaded in red and underexpressed metabolites in green. White nodes represent gene/gene product suggested by IPA as putative hub nodes connecting the uploaded metabolites. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The +2 and -2 are given to the upregulated and downregulated metabolites across studies regardless the magnitude of fold change expression.

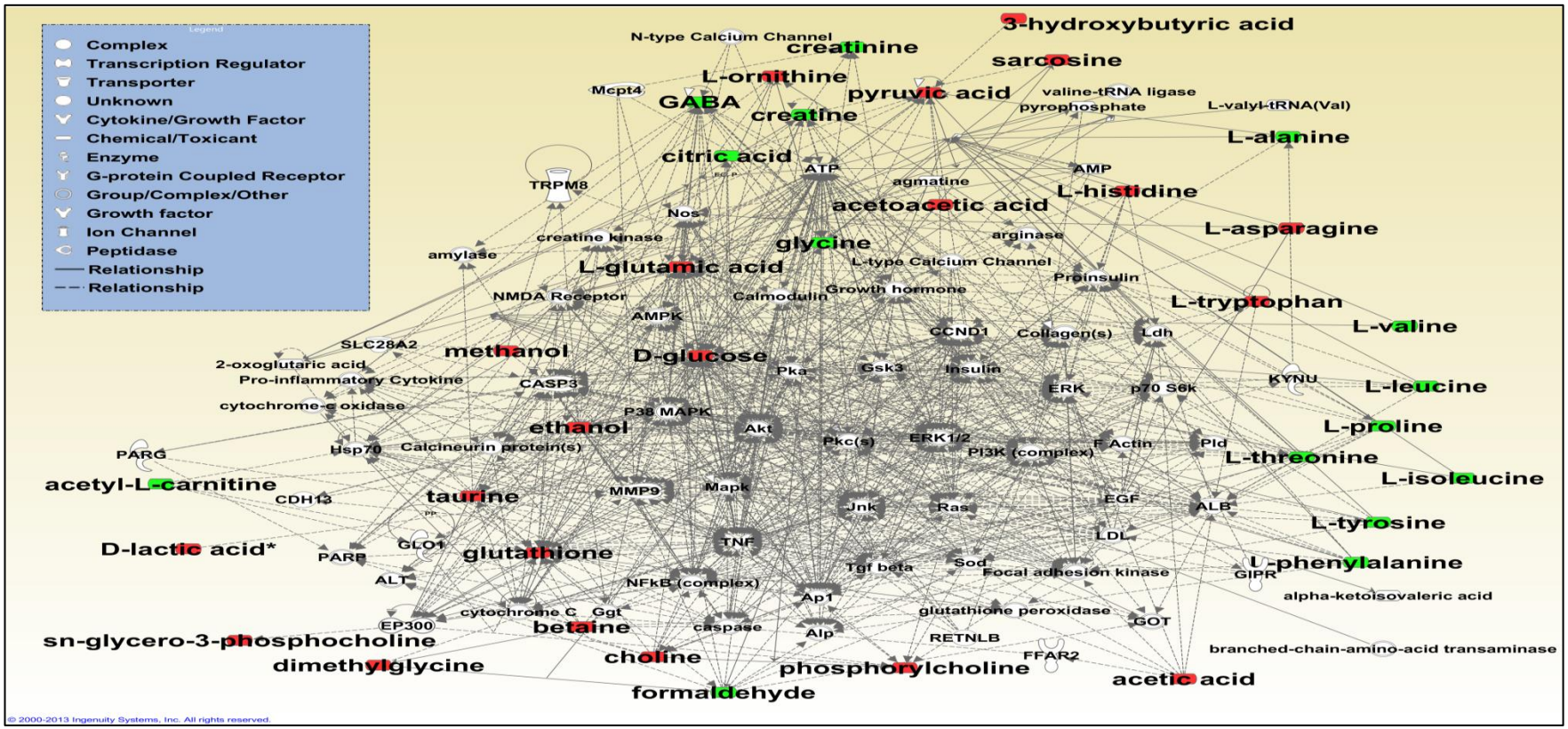


Figure 5.2: Master molecular network of the altered metabolites in OSCC.
The master molecular network assembled by merging networks 1, 2, 3, and 4 identified by IPA tool (version 9.0) from up- and down-regulated metabolomic analysis using overlapping function. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed metabolites are shaded in red and under-expressed metabolites in green. White nodes represent gene/gene product suggested by IPA as putative hub nodes connecting the uploaded metabolites. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The +2 and -2 are given to the upregulated and downregulated metabolites across studies regardless the magnitude of fold change expression. For clearer view of the network, see supplementary file 16.

5.4.2.3 Canonical pathway analysis of the common metabolites

By utilizing canonical pathway analysis tool, IPA, Glutathione-mediated detoxification and Glycine Biosynthesis pathways displayed as the most relative and common dysregulated pathway related to the uploaded metabolomic signature. The determination of the significant dysregulated pathways is accomplished by overlapping the input common metabolites over the pre-generated pathways from IPA.

5.4.2.4 Biomarker analysis of the common metabolites

In an attempt to find potential biomarkers involved in OSCC and HNSCC from the metabolic profile set, a knowledge base of IPA was unable to identify any biomarkers. However, 3 biomarkers were identified in relation to cancer disease generally (D-glucose, glutathione, and sarcosine). Figure 5.3 displays the diagnostic use of each biomarker in different tumour types using "Path Designer" advanced tool in IPA.

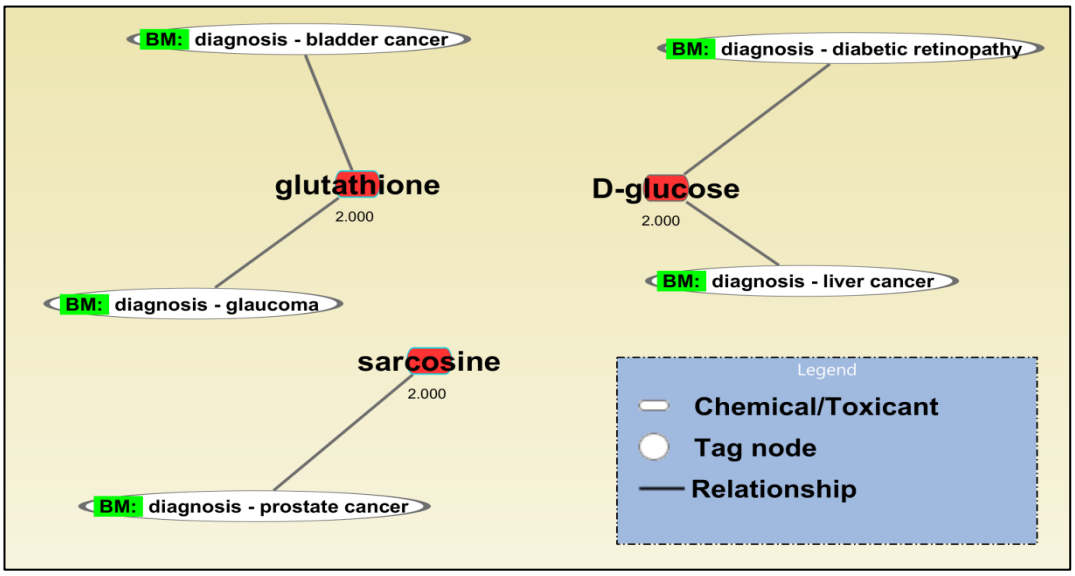


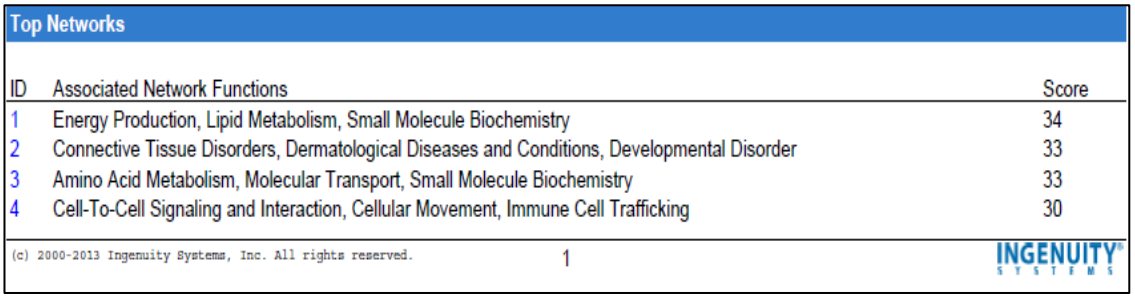
Figure 5.3: Biomarker analysis of common metabolites involved in cancer.

Biomarker network built by connecting significant metabolites involved in cancer, identified by biomarker filter tool (IPA, version 9.0) from up- and down-regulated common metabolites using overlap function of the biomarker analysis. The network is displayed graphically as nodes (metabolites) and edges (the biological relationship events between the nodes and the tumour types). The tumours relative to normal overexpressed metabolites are shaded in red and under-expressed metabolites in green. The +2 is given to the upregulated metabolites across studies regardless the magnitude of fold change expression.

5.4.3 Integrating metabolomic and transcriptomic profiles

5.4.3.1 Metabolomic analysis of both meta-genes and the 45 metabolites

IPA "Metabolomics Analysis" of the combined list of metabolites (the 45 metabolites harvested from the review study) and genes (the differentially expressed genes from the meta-analysis conducted in chapter 3) was performed. The generated networks display the interactions between the uploaded metabolites and genes (Figure 5.4).



Top Networks

ID	Associated Network Functions	Score
1	Energy Production, Lipid Metabolism, Small Molecule Biochemistry	34
2	Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder	33
3	Amino Acid Metabolism, Molecular Transport, Small Molecule Biochemistry	33
4	Cell-To-Cell Signaling and Interaction, Cellular Movement, Immune Cell Trafficking	30

1
INGENUITY SYSTEMS

Figure 5.4: Top networks of the interactive meta-genes and metabolites associated with significant molecular and cellular functions.

The top four networks were observed significantly enriched with scores ranging from 30 to 34 where the probability for a network to be selected by chance (score < 3) decreasing when its corresponding scores value increases.

Network 1, 3, and 4 exhibited the top and most relative networks to the tumorigenic process. Subsequent merging of these 3 networks generated the structure of the master network (Figure 5.5) whilst several metabolites involved, glucose symbolized a potential hub node of the network whereas HOXA10, AKT, and ERK1/2 represented the other main hubs structuring the master network. Among the selected top 3 networks, several genes and metabolites were associated with OSCC and HNSCC, including upregulation of C-X-C chemokines (CXCL1, 3, 5, 6, 9, 11, and 13), POSTN, TNC, PDPN, glucose, taurine, choline, sarcosine, betaine, and acetoacetic acid along with downregulation of PPP1R3C, GABA, creatine, glycine, formaldehyde, and acetyl-L-creatinine (Figure 5.5).

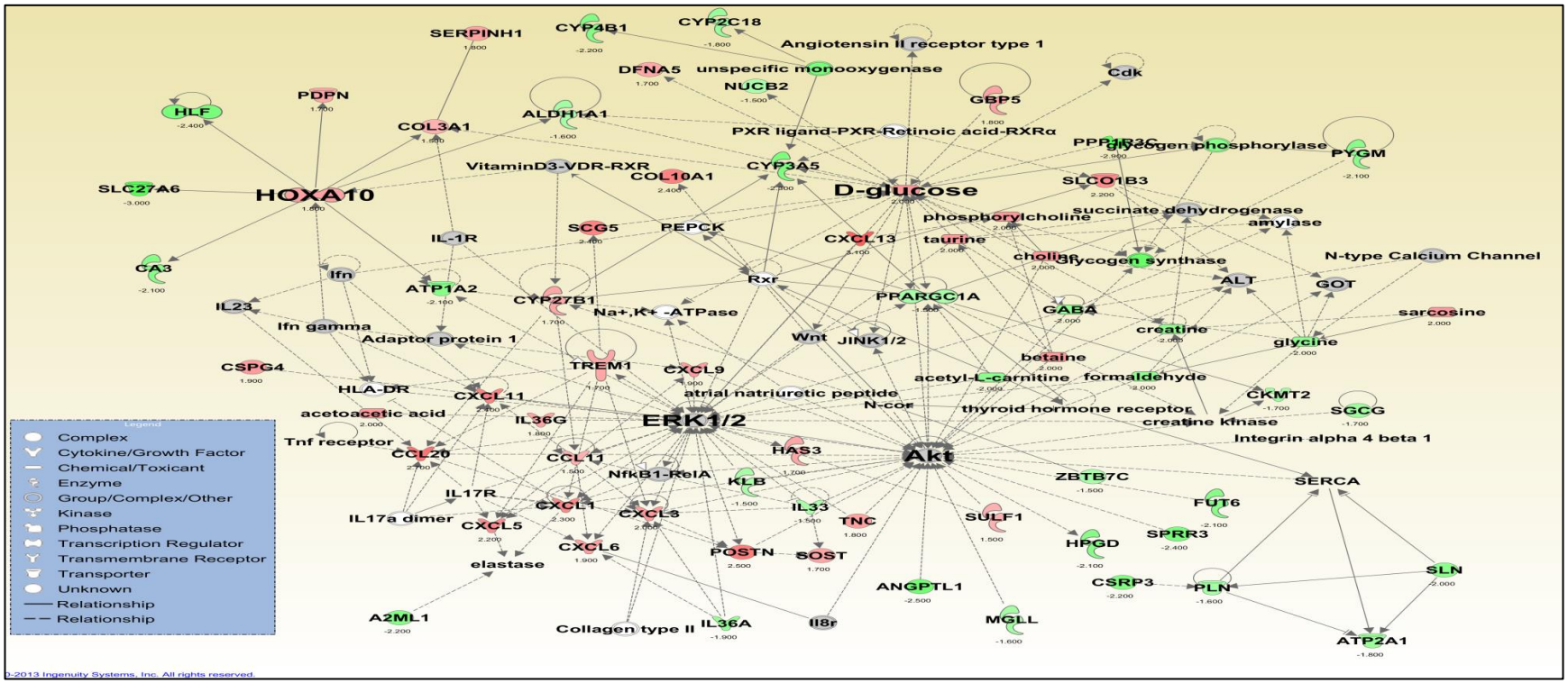


Figure 5.5: Master molecular network of meta-genes and metabolites using IPA Metabolomics analysis.
The master molecular network assembled by merging networks 1, 3, and 4 identified by IPA tool (version 9.0) from up- and down-regulated metabolomic and meta-analysis transcriptomic analysis using overlapping function. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed genes and metabolites are shaded in red and under-expressed metabolites in green. The intensity of colours reflects the magnitude of the fold change of the expressed molecule. Grey nodes represent gene/gene product exist in the dataset but don't reach the fold change cut-off criteria. White nodes represent gene/gene product suggested by IPA as putative hub nodes connecting the uploaded dataset molecules. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The +2 and -2 are given to the upregulated and downregulated metabolites across studies regardless of the magnitude of fold change expression. For clearer view of the network, see supplementary file 17.

5.4.3.2 Direct integration between the 45 metabolomic profile and the 32 OSCC and HNSCC transcriptomic biomarkers

Further interactive network was constructed by direct integration of the 45 metabolite set and the 32 transcriptomic biomarkers related to OSCC and HNSCC from the meta-analysis study. Using "connect" tool from "build" function of IPA, 14 out of 32 genes and 15 out of 45 metabolites showed at least one interactive biological relation event between transcriptomic and metabolomic profiles according to IPA knowledge base. Thorough examination of the interactive networks revealed possible interactions between metabolomic markers and several gene biomarkers involved in OSCC and HNSCC including upregulated genes with a role in ECM degradation (MMP1, MMP3, and MMP9), interleukins (IL1B, IL6, and IL8), collagens (COL1A1 and COL4A1), PTGS2, SERPINE1, SLC2A1, and STAT1 along with downregulation of CRYAB and CXCL12 (Figure 5.6). Metabolites of glucose, ethanol, and glutathione displayed the highest number of interactions with transcriptomic biomarkers. Likewise, interleukins, PTGS2, and MMP9 were the highest interactive molecules among the OSCC and HNSCC gene biomarkers.

Glucose appears at the centre of the master network with the highest number of interactions. The network displayed elevated SLC2A1, a glucose transporter, which considers a common characteristic of human malignancies.

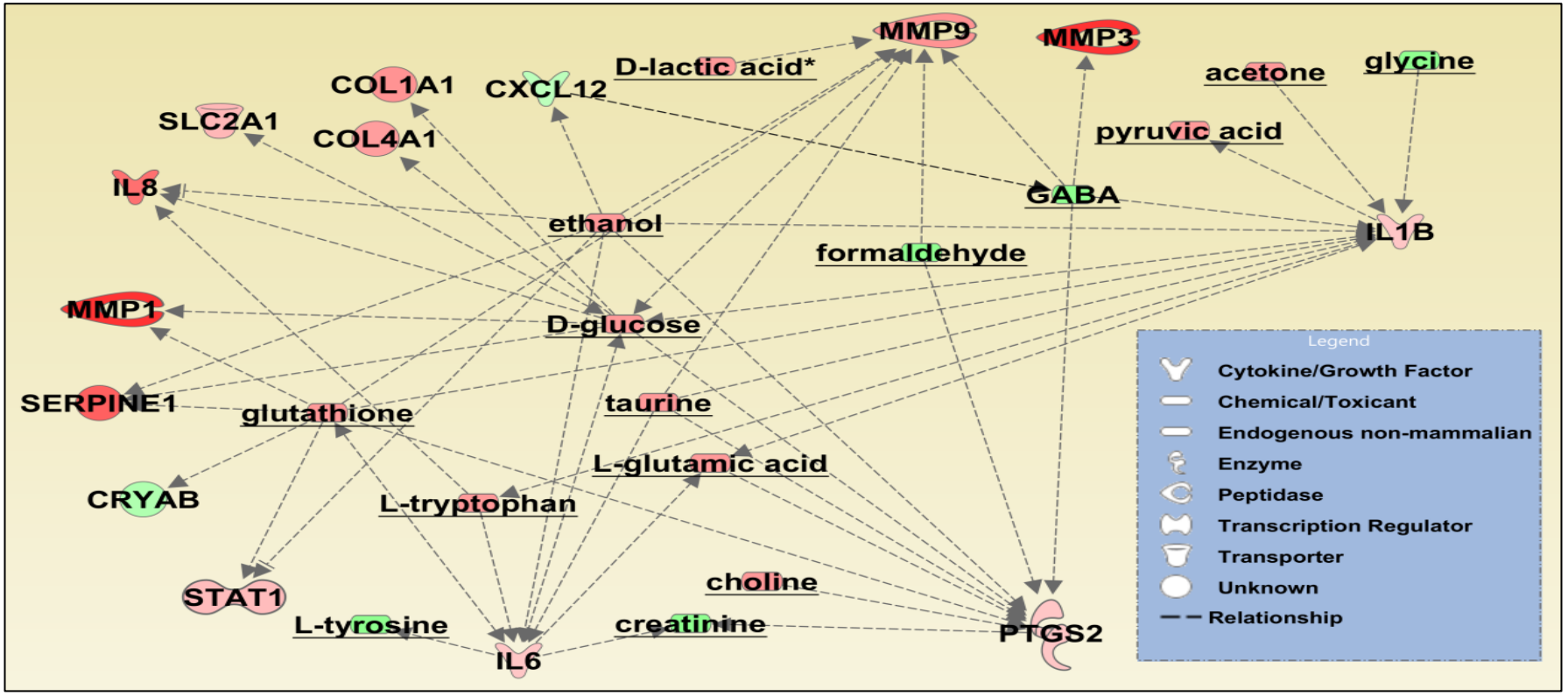


Figure 5.6: Interactive network between metabolomic profile and the meta-genes biomarkers of OSCC and HNSCC.

The interactive molecular network assembled by IPA tool (version 9.0) from up- and down-regulated metabolomic profile from the review study (chapter 5) and the 32 OSCC and HNSCC biomarkers from the meta-analysis study (chapter 3) using connect tool from build function. The network is displayed graphically as nodes (genes/gene products) and edges (the biological relationship between the nodes). The tumours relative to normal overexpressed molecules are shaded in red and under-expressed metabolites in green. The nodes are represented using various shapes that represent the functional class of the gene products. Highly interconnected nodes ('hub genes') are highlighted with black edges as potential key regulators of the master network. The underlined molecules denoted to the metabolites.

5.5 Discussion

5.5.1 Review study of metabolomics

Reviewing the literature searching for previous work using in vitro NMR and MS technologies for diagnosis and monitoring of OSCC and HNSCC, only few studies observed comparing to studies used microarray technology observed in the review study of chapter 2. Such a metabolomic profile includes aspects of host response as well as tumour behaviour. The metabolomic review study specified several metabolites and pathways displaying considerably altered expression in cancerous versus non-cancerous states across the 12 studies. Among the metabolites identified commonly altered in these studies, include several involved in anti-oxidation, free radical scavenging, detoxification of xenobiotics and chemicals (glutathione, creatine, glutamic acid, taurine, glycine, tryptophan, and pyruvic acid), fatty acid metabolism, lipid degradation, and glycolysis. These molecular pathways and cellular functions are typical signature of oral cancer and already described by previous studies. “Glutathione- mediated detoxification, methylglyoxal detoxification” pathways emphasise the crucial role of glutathione which is needed for the detoxification of methylglyoxal, a toxin produced as a by-product of metabolism.

Several of the observed metabolite markers were involved in associated network function (Figure 5.2), signalling and metabolic pathways (5.4.2.3), and biomarker analysis (Figure 5.3) representing glucose and glutathione metabolites as hub nodes associated with detoxification process, glycolysis, energy production and amino acid metabolism. Biomarker analysis identified both glucose and glutathione over-expression as verified biomarkers of other types of cancer along with sarcosine metabolite Figure 5.3.

Elevated glucose level along with significantly increased fatty acids metabolism may be linked to the unique behaviour of oral cancer. This represents a typical signature in cancer patients, and it has been previously linked to “Warburg effect” (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956), which assumes that tumours rely on glycolysis as a main source of energy, even in the presence of oxygen. Many

solid tumours showed an increase glycolytic metabolism, which has, in the case of OSCC, been associated with the over-expression of glucose transporters, especially Glut-1 (SLC2A1) (Kunkel et al., 2003), which showed an elevation in the current study as well.

Glutathione-S-transferase (GST) is considered a useful aid in early diagnosis, predicting tumour extent, and determining parameters of treatment efficacy and prognosis of oral cancer. Elevated GST was also detected in the sera of patients with cancers of oesophagus, gastrointestinal tract, liver, bile duct, pancreas, uterine cervix, colon, and lung (Hirata et al., 1992). A higher immune-activity and more affinity of GST in squamous cell carcinoma among cancers of various types of histological findings suggested GST a possible useful marker in oral cancer and cervical cancer because both have substantial involvement by squamous cell carcinoma (Eimoto et al., 1988). Further results indicated glutathione level of oral tissues may be a useful marker for oral cancer, which is in agreement with findings from lung squamous cell carcinoma, cervical squamous cell carcinoma and other squamous cell carcinomas (Wong et al., 1994).

Glutathione transferases (GSTs) are enzymes that catalyse the conjugation of glutathione to a variety of electrophilic substances. Their best known role is as cell housekeepers engaged in the detoxification of xenobiotics (redox reactions, in which portion of cell membranes protects against peroxidation) (Scholz et al., 1989). Recently, GSTs have also been shown to act as modulators of signal transduction pathways controlling cell proliferation and cell death. Their participation in tumour cell growth and differentiation, and in the development of resistance to anticancer agents, has made them attractive drug targets (Laborde, 2010). IPA canonical pathway analysis of the current study displayed Glutathione-mediated detoxification, Methylglyoxal detoxification, Glutathione-mediated formaldehyde oxidation II, Glycine Biosynthesis, and several amino acids metabolism pathways exhibited the top, carcinogenic related, dysregulated pathway that the uploaded set of metabolic profile involved in. Previous studies extensively described the role of these pathways during tumour transformation. Through this essential detoxification mechanism, glutathione binds to electrophilic chemicals, forming conjugates which exported

from the cell. These conjugation reactions have been demonstrated for a multitude of foreign chemicals, as well as endogenous reactive intermediates (Wang and Ballatori, 1998). The breakdown products of the glutathione -toxin conjugates (glutamate and glycine) are reabsorbed and can be used for glutathione synthesis (Boyland and Chasseau.Lf, 1969). The results of current study also showed glycine regularly repressed along with perturbed level of glutamic acid which confirm the detoxification abundance.

Over-expression of ethanol in tumour samples has critical relation with glutathione in tumourigenesis. Chronic alcohol consumption has been reported to decrease glutathione levels and enhances the susceptibility of the liver towards alcohol related peroxidative damage (Lieber, 1994a, Lieber, 1994b).

A pivotal relation between higher consumption of alcohol and OSCC and carcinoma of the pharynx, Larynx, and oesophagus has been distinguished since the mid1950s (1988). Alcohol may influence the proliferative cells by both intracellular (e.g., endocytosis) and intercellular (permeability) pathways. The carcinogenic exposure of the proliferating stem cells in the basal layer may be regulated through these pathways (Ogden, 2005). Ethanol can induce oxidative DNA damage in human peripheral lymphocytes in vitro, and its mechanism may be associated with the metabolism of ethanol by the ADH1B/ALDH2 pathway (Yan et al., 2011). Alcohol dehydrogenase 3 (ADH3) converts ethanol to acetaldehyde, which is a suspected oral carcinogen (Schwartz et al., 2001).

Ethanol consumption induces changes in lipid metabolism. This might be reflected locally as an alteration in the epithelial lipid barrier. The parallel changes between lipid metabolism and permeability suggests that one effect of ingested alcohol is to alter the lipid-containing permeability barrier of stratified squamous epithelium (Squier et al., 2003).

Exposure to ethanol and environmental tobacco smoke induced high increase in cytochrome P450 2E1 (CYP2E1), an enzyme involved in the metabolism of xenobiotics, and inducible nitric oxide synthase (iNOS) protein. It also increased 3-

nitrotyrosine, mtDNA damage, decreased cytochrome c oxidase protein as well as increasing HIF1alpha expression (Bailey et al., 2009).

Several evidences of elevated glutamate and interaction between inflammatory cytokines and glutamate receptors may play a role in initiating certain cancers. Additionally, recent studies demonstrated an interaction between inflammation and glutamate receptors that may enhance tumour invasion and metastasis by affecting a number of cell-signalling mechanisms. Chronic inflammation appears as a major initiating mechanism in most human cancers, involving cell-signalling pathways, which responsible for cell cycling, cancer cell migration, invasion, tumour aggressiveness, and angiogenesis (Blaylock, 2013).

The excessive growth of a tumour requires high rates of glucose uptake and glycolysis and continuous recruitment of new blood vessels. Pyruvic acid, the end product of glycolysis, exhibits strong angiogenic activity. Additionally, mRNA expression of fibroblast growth factor receptor-2 (FGFR2) and vascular endothelial growth factor (VEGF) was enhanced by pyruvic acid suggesting important role of pyruvic acid in angiogenesis of tumour growth and metastasis (Lee et al., 2001).

Further study showed pyruvate suppressed the expression of inflammatory mediators such as IL1α, IL1β, IL6 and COX2 in a culture media. In addition, pyruvate suggested inhibiting inflammatory response by inhibiting the p38 MAPK activation (Aoki-Yoshida et al., 2013).

It has been known that cancer cells produce excessive amounts of lactic acid. Researchers believe that without sufficient oxygen, cancer cells must revert to fermentation for their energy supply and this is what causes them to produce excessive lactic acid. Controversial hypothesis suggested cancer cells have dysfunctional mitochondria, which prevent their use of citric acid or Krebs cycle. Consequently, pyruvic acid, the end product of glycolysis, which normally would enter the mitochondria for its total combustion into energy, is instead converted to lactic acid. Evidence exists to support this hypothesis which, when acknowledged, could dynamically impact both cancer research and the treatment of all forms of

cancer (John, 2001). Recent results showed that lactic acid functions as an intrinsic inflammatory mediator that activates IL23-dependent and -independent pathways, resulting in the promotion of chronic inflammation in tumour microenvironments (Yabu et al., 2011). Lastly, Taurine and lysine can increase both GABA synthesis and effects, while aspartic acid and glutamic acid probably inhibit GABA effects (Nisijima and Ishiguro, 1995).

Although a promising progress of metabolomic diagnostic profiles, the sensitivity and specificity cannot meet the demands of being a clinical useful tool for monitoring disease. The multifactorial nature of oncogenesis and the heterogeneity in oncogenic pathways make detection of all cancer of a particular organ implausible to be detected when using a single biomarker with high specificity and sensitivity. Therefore, prediction models to identify combinations of biomarkers that can identify OSCC patients among all the samples can improve the task.

5.5.2 Metabolomics analysis of both meta-genes and the 45 metabolites

The corresponding functions of the highest score networks were; energy production, lipid and fatty acid metabolism and degradation processes emphasizing the significant metabolic activity of tumour relying on glycolysis as a main source of energy alternative to normal mitochondrial respiration and it has been previously linked to “Warburg effect” (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956) while lipid degradation and amino acid metabolism provide the sufficient fuel and environment for glycolysis.

5.5.3 Direct integration between the 45 metabolomic profile and the 32 transcriptomic biomarkers of OSCC and HNSCC

5.5.3.1 Glucose and gene biomarkers of OSCC and HNSCC

Elevated levels of glucose transporters, especially of Glut-1 (SLC2A1), considers a common characteristic of human malignancies, including OSCC where both glucose transport and its metabolism might play a key role in the progression of OSCC. Both glucose transport and glucose metabolism determine the glycolytic tumour

phenotype, which considers a significant negative biomarker of prognosis and overall survival in patients with OSCC (Kunkel et al., 2003). This probably due to the enhanced glycolytic metabolism of more aggressive neoplastic cells (Rojas Ayala et al., 2010). Both GLUTI and GLUT3 reported an involvement in basal glucose uptake of extracranial head and neck tumours. The increased expression of these high-affinity GLUT isoforms may be related to the growth maintenance of tumour tissue in cases of inadequate supply of substrate *e.g.*, in poorly vascularized tumour areas (Mellanen et al., 1994).

Glucose induces the upregulation of Pro-renin receptor (PRR) and its ligands, leading to increased IL1B and COX2 production via the angiotensin II-dependent pathway (Huang and Siragy, 2009). Further study demonstrated that high glucose in vitro might directly upregulate the expression of the inflammatory PTGS2 gene in pancreatic islets (Shanmugam et al., 2006). High glucose also reported significantly increasing reactive oxygen species (ROS) production, IL6 secretion, intercellular adhesion molecule-1 (ICAM-1) expression, and endothelial adhesiveness to monocytes (Chen et al., 2012). Similarly, results of ELISA showed the release of IL6 and IL8 can be inhibited by high glucose. High glucose may decrease the innate immunity through relation to toll-like receptor TLR2 and TLR4 in cornea epithelium (Ni et al., 2011). Furthermore, D-Glucose in culture medium increased the release of human IL6 protein in fibroblasts from human skin that mediated by adenosine triphosphate (Solini et al., 2000). Moreover study report concluded that IL1B-induced nitric oxide production in the presence of glucose is signalled by the p38 pathway (Sprinkel et al., 2001).

In human mesangial cells, D-Glucose indicated an increase expression of human SERPINE1 gene (Murphy et al., 1999, Jeong et al., 2011, Chen et al., 1998). Additional results specified stimulation of SERPINE1 promoter by both high glucose and glucosamine and that the hexosamine pathway may be involved in the regulation of gene expression by high glucose in glomerular mesangial cells (Goldberg et al., 2000).

The production of MMP9 and collagen IV alpha protein can be regulated by high glucose, and the extracellular signal-regulated kinase ERK1/2 transduction pathway mediates such regulation (Bai et al., 2005). Further study suggested mechanism of redox-sensitive MMP9 expression by hyperglycaemia which may provide a rationale for antioxidant therapy to modulate diabetic vascular complications (Uemura et al., 2001). In addition, a study reported that high glucose, IL6, and lipopolysaccharide act in concert to stimulate MMP1 expression. A study proposed that hyperglycaemia-induced elevation in p38 MAPK activity and Activating transcription factor-2 (ATF-2) phosphorylation which contribute to cyclic adenosine monophosphate (cAMP) response element activation and modulation of c-Jun and collagen I expression in osteoblasts (Zayzafoon et al., 2002). An attempt to elucidate underlying mechanisms, ERK and c-Jun N-terminal kinase (JNK) pathways were required for stimulation of MMP1 by IL6 and high glucose. Moreover IL6 and high glucose stimulated the expression of c-Jun, a key subunit of AP1 known essential for MMP1 transcription (Li et al., 2010). In cultured human endothelial cells, high glucose-induced overexpression of fibronectin and collagen IV were detected (Roy et al., 1990). Exposure of human renal fibroblasts to hyperglycaemic conditions resulted in the upregulation of the mRNA expression of both COL1A1and COL3A1where the increased transcription of COL1A1 mRNA by high glucose and exogenous TGF-β1 was reflected by an increase in COL1A1 protein secretion (Lam et al., 2003).

5.5.3.2 Ethanol and gene biomarkers of OSCC and HNSCC

Recently, some studies have demonstrated that ethanol and its metabolites, including acetaldehyde, can induce the production of matrix metalloproteinases (MMPs) or inhibit the tissue inhibitors of metalloproteinases (TIMPs) in the liver (Arthur, 2000, Friedman, 2004, Purohit and Brenner, 2006). Ethanol and acetaldehyde directly stimulate the production of transforming growth factor beta-1 (TGFβ-1) and several extracellular matrix (ECM) constituents and also alter TIMPs/MMPs regulation by direct activation of hepatic stellate cells (HSCs), inducing liver fibrosis (Friedman, 2004, Purohit and Brenner, 2006). In fact, acetaldehyde is a fibrogenic agent to the liver inducing the expression of type-1 and type-3 collagen genes in HSCs (Purohit and Brenner, 2006). In fibrotic liver, elevated level of TIMPs which play a

significant role in the regulation of MMPs activities, including the gelatinases 2 and 9, stimulates both proliferation and migration (Pires et al., 2008).

Additional findings on human breast cancer exhibited a higher amount of active MMP2 and MMP9 in their culture medium when exposed to ethanol giving new insights into the effects of alcohol on cell migration in breast cancer which might not due solely to an oestrogen-like activity of alcohol (Etique et al., 2006).

The ability of ethanol and arachidonic acid (AA), as inducers of oxidative stress and key factors in alcoholic liver disease, to upregulate COL1A2 gene expression was identified where the oxidant stress derived from metabolism of ethanol or AA was found directly fibrogenic. Mechanistic studies reveal a critical role for H2O2 in the upregulation of COL1A2 expression by ethanol and AA. Moreover, COX-2 mediates the AA-mediated induction of COL1A2 expression (Nieto et al., 2000). Ethanol and acetaldehyde, also induced production of type I collagen (Masamune et al., 2010).

Low doses of alcohol have been shown to alter the expression of the genes encoding tissue plasminogen activator (PLAT) and SERPINE1(Booyse et al., 1999), an OSCC biomarker, which play a critical role in oral cancer.

5.5.3.2.1 Ethanol and pro-inflammatory cytokines (IL1B, IL6, and IL8)

Increased serum levels of IL6, IL10 and, to a lesser extent IL8, declined in the few days after alcohol abstinence in patients with alcohol withdrawal syndrome. whilst serum IL8 level increased after alcohol intake in healthy subjects (Gonzalez-Quintela et al., 2000). Further study demonstrated that endotoxin-induced IL8 release is mediated by TNF alpha (Vandeventer et al., 1993). However, several studies reported acute alcohol, in vivo as well as in vitro, attenuating monocyte-derived chemokine production in response to a subsequent immune challenge. Hence, impaired IL8 induction upon an immune challenge is likely to contribute to compromised host defence after acute alcohol consumption (Szabo et al., 1999, Taieb et al., 2002, Mas et al., 2010). Additional results demonstrated that acute ethanol exposure can impair macrophage IL6 production and indicated that this

effect may result from ethanol-induced alterations in intracellular signalling through p38 and ERK1/2 (Goral et al., 2004, Norkina et al., 2008).

The abovementioned variable levels of chemokine in relation to alcohol can be explained by a study findings represented that high concentrations of ethanol inhibit IL6-activated anti-apoptotic signal, though increasing the concentrations of IL6 can overcome such inhibitory effect. These findings suggest that elevated serum IL6 levels in alcoholic liver disease may overcome the inhibitory effect of ethanol on IL6-mediated anti-apoptotic signals (Hong et al., 2002).

Moreover, a study revealed that acute alcohol treatment activated STAT1/3 signalling pathways, a significant cancer indicator, and induced suppressors of cytokine signalling levels in monocytes (Norkina et al., 2008), which support the evidence of alcohol consumption as a major risk factor in oral cancer.

5.5.3.3 Glutathione and gene biomarkers of OSCC and HNSCC

Glutathione (GSH) has multidisciplinary mechanisms in cancer. Oxidative stress may cause glutathione increased activation of pro-MMPs (pro-MMP1, 8, and 9) (Springman et al., 1990, Okamoto et al., 2001). Additional result suggested a developed synthetic Organoselenium compound, a glutathione conjugate, that inhibited colon carcinogenesis and COX1 and COX2 activities (Rao et al., 2001). Other study showed depletion of GSH increased expression of mouse IL1B mRNA, an OSCC biomarker, in mouse liver (Llacuna et al., 2009). Moreover, in fibroblasts from human lung, glutathione decreased expression of human IL6 mRNA that increased by human TGFB protein (Junn et al., 2000). Recent study demonstrated identification of IL6 as a tumour-derived molecule promoting GSH efflux in hepatocytes (Obrador et al., 2011).

Additionally, glutathione reported to decrease transcription of SERPINE1 mRNA that is increased by oxidative stress of cells (Vulin and Stanley, 2004). Moreover, the regulation of SERPINE1 expression induced by cyclosporine A might be critically related with the intracellular glutathione and the ERK-MAPK pathway (Ho et al., 2010). Further study demonstrated that STAT1 inhibition by prostaglandin

metabolite was abolished by N-acetylcysteine, glutathione, superoxide dismutase, and catalase (Chen et al., 2003). In addition, a glutathione reduction supposed to be necessary for antioxidant activity of CRYAB protein (Arrigo et al., 2007).

5.5.3.4 Taurine and gene biomarkers of OSCC and HNSCC

Anticancer effects of taurine, 2-aminoethanesulfonic acid, on cancer cells remain poorly understood. However, the protective properties of taurine have been attributed to taurine acting as a detoxifying agent against hypochloric acid by forming the stable and less toxic taurine chloramine (Quinn et al., 1996). Taurine diminished the induction of COX2 (Warskulat et al., 1997) and reduced mRNA of pro-inflammatory mediators encoding MMP9, fibronectin and the protein expression of COX2 (Song et al., 2011). Recent study reported a role of taurine in reduction of cell proliferation by apoptosis induction (Kim and Kim, 2013).

Furthermore, taurine chloramine reported inhibition of superoxide anion, IL6 and IL8 in activated human leukocytes (Park et al., 1998), and from rheumatoid arthritis patients (Park et al., 2002, Chorazy-Massalska et al., 2004) which attributed to the ability of this compound to diminish the activity of the major transcriptional regulators (NFκB and activator protein 1), which subsequently reduced the transcription of these cytokine genes (Kontny et al., 2000).

Recently, results suggested that the anti-inflammatory effect of the taurine conjugated 5-aminosalicylic acid on experimental colitis is attributed to the inhibition of the IL1B-mediated NFκB activation and the taurine effect is through taurine chloramine potentiating the ability of 5-aminosalicylic acid to inhibit IL1B dependent NFκB activation (Joo et al., 2009).

5.5.3.5 Tryptophan and gene biomarkers of OSCC and HNSCC

L-Tryptophan reduced the oxidative stress and expression of the TNF-α, IL6, IFN-γ, IL12p40, IL1B, IL17, IL8, and intracellular adhesion molecule-1, as well as increased expression of apoptosis initiators caspase-8 and BAX (Kim et al., 2010). Inversely, IL1B protein in cerebral ventricles has been increased the quantity of tryptophan in rat hippocampus (Song and Horrobin, 2004).

5.5.3.6 Glutamate and gene biomarkers of OSCC and HNSCC

Glutamate reported a significant reduction by IL1B protein (Kelly et al., 2001). IL6 has been also shown to inhibit glutamate release in the cerebral cortex (D'Arcangelo et al., 2000). Further study, demonstrated IL6, IL11, and OSM cytokines, partially inhibited glutamate-induced neuronal death that mediated by STAT3 activation (Park et al., 2012).

Neuronal COX2 expression was reported an elevation following brain insults via glutamatergic and inflammatory mechanisms (Strauss and Marini, 2002) and the elevation was reduced by 2-Arachidonoylglycerol (Zhang and Chen, 2008).

5.5.3.7 GABA and gene biomarkers of OSCC and HNSCC

GABA reported to decrease production of IL1B protein in plasma from diabetic mouse (Soltani et al., 2011). It also decreased expression of COX2 protein (Hou, 2011). However, GABA increased expression of human MMP3 and MMP9 proteins where the influence of GABA stimulation on in vitro MMP production and the invasive ability of cancer cells was investigated using human prostate cancer cell line indicating the involvement of MMP activity in GABA-induced cancer cell invasion (Azuma et al., 2003). As a final point, CXCL12 was reported to enhance the effects of GABA (Luo et al., 2008, Bhattacharyya et al., 2008).

5.5.3.8 Choline, Creatine, Creatinine, Tyrosine, and Lactic acid interactions with genetic biomarkers of OSCC and HNSCC

Choline reported to upregulate the expression of COX2 (Zhao et al., 2010) which increase the level of creatinine in blood (Norwood et al., 2000). Furthermore, mutant mouse IL6 gene (knockout) reported to decrease quantity of creatinine in mouse serum that is increased by ischemia reperfusion of kidney (Kielar et al., 2005). Additionally, human IL6 suggested to increase phosphorylation of L-tyrosine (Yin and Yang, 1994). To end with lactic acid/ MMPs interactions, extracellular matrix remodelling involving matrix metalloproteinases and wound lactate accumulation are considered as essential elements of tissue repair (Weinreich et al., 2011).

Altogether, metabolites perform a complex interaction pattern with transcriptomic biomarkers of OSCC and HNSCC and supposed to be more complex when attempting to include all altered genes related to cancer disease generally (rather than only OSCC and HNSCC biomarkers) from the differentially expressed genes discriminating tumour from normal samples of the meta-analysis. However, certain biological pathways and cellular processes appear constantly dysregulated among the vast interactions, including lipid and amino acid metabolism, molecular transport, cell-to-cell signalling interaction, cellular movement, and immune cell trafficking.

5.6 Summary

Metabolomic signature recruited NMR and MS spectral technologies to study the body end products in reflecting the behaviour of oral cancer and body response against it. The current review enriched our knowledge via better understanding of the characteristic behaviour of oral cancer and the alteration of system biology in response to the tumorigenic assault at metabolic level. Diagnostic tool biomarker from metabolomic signature of OSCC is under improving progress.

"Metabolomics Analysis" tool from IPA, provided the common molecular pathways and cellular functions related to both metabolomic candidates harvested from the review study (45 metabolites) and the differential transcriptomic profile from the meta-analysis study conducted in chapter 3 (8995 genes). The corresponding pathways and functions of the networks revealed energy production, lipid and fatty acid metabolism and degradation processes highlighting the significant metabolic activity of oral cancer relying on glycolysis as a main source of energy alternative to normal mitochondrial respiration.

Considering the transcriptional and metabolic changes that accompanied carcinogenic exposure during OSCC, direct biological integration highlighted ethanol, a major risk factor in oral cancer, created essential hub molecule in the interactive network. Acute alcohol intake suggested attenuating monocyte-derived chemokine production which responds to a subsequent immune challenge (Szabo et al., 1999, Taieb et al., 2002, Mas et al., 2010). Ethanol may influence proliferative cells, inducing oxidative DNA damage, altering the lipid-containing permeability barrier of stratified squamous epithelium, and can increase mitochondrial DNA damage (Ogden, 2005). It also induced the production of matrix metalloproteinases, stimulated both proliferation and migration or inhibited the tissue inhibitors of metalloproteinases (TIMPs) (Pires et al., 2008) as well as activation of STAT1/3 signalling pathways , a significant cancer indicator (Norkina et al., 2008), as well as elevation of COL1A2 (Nieto et al., 2000), and SERPINE1 (Booyse et al., 1999). Moreover, ethanol decreased glutathione and cytochrome c oxidase protein level, the essential detoxifying agents, (Bailey et al., 2009).

Glucose, the highest interacting metabolite with correlative transcriptomic biomarkers, denoted to glycolysis, the unique cancer behaviour including oral cancer. Elevated level of glucose with its transporters, especially SLC2A1, considers a common characteristic of human malignancies, including HNSCC where both glucose transport and its metabolism may play a key role in the progression of OSCC (Kunkel et al., 2003). Glucose also increased IL1B, COX2 (Huang and Siragy, 2009), production of reactive oxygen species (ROS) (Chen et al., 2012), IL6 secretion (Solini et al., 2000), expression of human SERPINE1 (Murphy et al., 1999, Jeong et al., 2011, Chen et al., 1998), as well as the production of MMP1 (Zayzafoon et al., 2002), MMP9 (Bai et al., 2005), fibronectin, COL1A1, and COL4A1 (Lam et al., 2003).

Elevation of glutathione, taurine, tryptophan, and L-glutamic acid levels constitute the main antioxidant and detoxification system essential to neutralise and remove the toxic by-products that produced during tumourigenesis. Best known role of glutathione enzymes is as cell housekeepers in the detoxification of xenobiotics (Scholz et al., 1989). Glutathione inhibited COX1 and COX2 activities (Rao et al., 2001), decreased transcription of SERPINE1 mRNA that increased by oxidative stress of cells (Vulin and Stanley, 2004), decreased expression of human IL6 and IL1B mRNA (Junn et al., 2000, Llacuna et al., 2009). Taurine acts as a detoxifying agent against hypochloric acid by forming the stable and less toxic taurine chloramine (Quinn et al., 1996). Taurine also diminished the induction of COX2 (Warskulat et al., 1997), reduced mRNA of pro-inflammatory mediators encoding MMP9, and fibronectin (Song et al., 2011), as well as inhibited the production of IL6, IL8, and IL1B (Park et al., 1998, Joo et al., 2009). L-Tryptophan established a reduction of oxidative stress and the expression of interleukins IL6, IL1B and IL8 (Kim et al., 2010). Similarly, glutamic acid exhibited a declination by IL1B protein (Kelly et al., 2001) and IL6 (D'Arcangelo et al., 2000).

Comprehensive reading of the aforementioned simultaneous transcriptional and metabolic interactive changes that accompanied tumourigenesis of OSCC revealed an intimate relation between genetic and metabolic system biology during carcinogenic transformation. To obtain meaningful and comprehensive

understanding, more efforts required to identify additional key players of a larger interactive network with a larger review study and validation to confirm the role of each player in this complex context. However, the current analysis already highlighted interleukins, MMPs, COX2, SERPINE1, and STAT1 as major key players represented the transcriptomic platform transecting specific players from the platform of metabolome which include; ethanol and its metabolites establishing the carcinogenic factor, glucose and its relative metabolites as representative by-products of OSCC, glutathione, creatine, glutamic acid, taurine, glycine, tryptophan, and pyruvic acid as a reparative antioxidant and detoxifying elements against tumorigenic assault of OSCC (Figure 5.6).

Such an interactive pattern may represent a clinically useful surrogate of the presence of OSCC and can be used to form the basis of diagnostic device of OSCC. Further investigation of the interactive network is required. Additionally, to get more comprehensive and complete story, further expanding of the interactive network is required by including all altered genes involved in cancer disease from the meta-analysis study (401genes rather than only 32 OSCC and HNSCC biomarkers that included in the current analysis) for integration with the 45 metabolomic OSCC biomarkers (biological level integration).

Clinical application of the current interactive signature as diagnostic tool can also carried out by simple cross-correlation of both metabolomic and transcriptomic candidate biomarkers in the interactive model across the examined samples (data level integration). A combination of part or all transcriptomic and metabolomic molecules of the interactive network in figure 5.6 can be applied on clinical independent samples to validate it’s efficacy as discriminative tool of OSCC.

Chapter 6: Summary and Discussion

6.1 Rational and objectives of the project

The pathogenic pathways that contribute to OSCC remain poorly characterised and the critical factor in the lack of prognostic improvement is that a significant proportion of cancers initially are asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. Hence, a clinically applicable gene expression signature is in high demand and improved characterization of the OSCC gene expression profile would constitute a substantial progress. For OSCC, possible themes that might be addressed using microarray data include distinguishing the disease from normal at the molecular level; determining whether specific biomarkers or profiles are predictive for tumour behaviour; and identifying biological pathways necessarily altered in tumourigenesis, potentially illuminating novel therapeutic targets. However, OSCC is a heterogeneous disease, making diagnostic biomarker development difficult. Although this phenotypic variation is striking when one compares OSCC from different geographic locales, little is known about the underpinning biological mechanisms.

The primary aim of this thesis is to investigate metabolomic and transcriptomic profiles using multiple bioinformatics approaches and biological annotation tools in an attempt to identify specific biomarkers and prediction models for OSCC from each profile as well as from the interface outcomes of integrating the two platforms. Additional aims of the thesis go further to identify the mechanisms underlying the biological changes during tumorigenic transformation of OSCC, as well as to determine the biological processes underpinning the heterogeneity of the disease among populations.

OSCCs of UK and Sri Lanka have different aetiological factors and clinical phenotypes. One aim of this project is to identify gene expression differences between the two groups of tumours as a preparatory point to understand their biological diversity. The availability and free access of the wide range bioinformatics packages from Bioconductor in R along with the existence of IPA tools as a powerful

biological annotation tools, provide a great enhancement to analyse the raw data available from both countries.

Furthermore, to investigate the interactive networks created by merging metabolomic and transcriptomic profiles to highlight the significant molecular and cellular biofunctions, pathways, and biomarkers distinguishing OSCC from normal oral mucosa. The pattern may represent a clinically useful surrogate for the presence of OSCC which might help in deciphering some of the obscure multifaceted mechanisms underlying carcinogenesis of OSCC which emerged from dysregulated genetic and metabolic system of the body.

6.2 Summary of the findings

This thesis has described findings from genome wide expression analysis, recruiting various bioinformatics approaches and biological annotation tools for the analysis of molecular behaviour of OSCC. The study reported the discovery of a number of potential candidate genes, metabolites, and a comprehensive integrative approach for further research. This final chapter presents a summary of the findings, how they relate to the molecular field of OSCC diagnosis, and a possible direction for future work based on these findings.

6.2.1 Common gene expression profiling reveals biological alterations involved in early tumourigenesis of oral squamous cell carcinomas (Chapter 2)

The review study specified several genes and pathways exhibiting substantially altered expression in cancerous versus non-cancerous states across studies. Figure 2.4 displayed 76 (over 20% of the total common genes) genes associated with OSCC and HNSCC where 27 of them are verified biomarkers (Figure 2.6). Notably, some of these changes, although identified in multiple studies, may also present in other sample studies which are not identified, basically due to the absence of probes for these genes in the platforms used for these studies. Arrays used by various studies contained probes for as few as 384 genes to as many as over 47,000 transcripts (Table 2.1). These results were encouraging considering the heterogeneity of the studies where the samples are collected.

The current review pointed out a popular significant inaccuracy among the studies using microarray technologies reflected by the underestimation of the tumour origin of HNSCC during sample collection. A study used microarray design demonstrated that HNSCCs originating from different anatomic locations can exhibit various behaviours which gave a conclusion that HNSCC originating from different anatomic sites share consistent changes in gene expression when comparing primary tumours to normal adjacent mucosa, where these common changes most likely reflect alterations required for tumour development. In contrast, once a tumour has developed, tumour-host interactions at different anatomical sites are likely responsible for the site-specific signatures associated with aggressive versus non-aggressive disease. Hence, predictions of outcomes based on gene expression profiling are therefore heavily influenced by the anatomical site of the primary tumour (Belbin et al., 2008).

The top differentially expressed gene in 11 out of 28 studies is considered a poor outcome as a unique identifier/s for OSCC. The weakness of the tumour profiling of the list of common genes originates primarily from limited sample sizes and heterogeneity in experimental design and execution as well as the differences in the size of the final set of genes published by the individual studies.

Published data from such studies should not viewed as endpoints of research accomplishments, but rather as screening tools for identifying potential genes for validation and further investigation. Nevertheless, the capability to measure the expression of thousands of genes simultaneously gives researchers a powerful method to analyse global genetic events responsible for HNSCC progression (Choi and Chen, 2005).

This review did not access the expression raw data of most of the studies, and therefore did not consider how different fold changes in gene expression might impact different clinical outcomes or biological pathways. Ideally, all published studies containing DNA microarray analyses should deposit their raw expression data in publically accessible repositories (e.g., NCBI GEO, Array express, others), to permit accomplishment of systematic and comprehensive meta-analysis. For this

reason, additional attempt was carried out (chapter 3) to perform meta-analysis profiling OSCCs relative to normal mucosa by recruiting the same bioinformatics approach on multiple raw data sets.

Conclusion of the current review suggests not only relying on the final set of genes that published by the individual studies, but to access the raw data of each study and start subsequent analysis from that stage using unified bioinformatics approaches to acquire useful and complete understanding of the systems biology and to identify the subsequent dysregulated biological pathways, molecular functions, and processes.

6.2.2 Detection of a robust gene signature for oral squamous cell carcinoma by integrating multiple microarray datasets (Chapter 3)

The successful discriminative results of bioinformatics approaches highlight the advantage of meta-analysis (employed large and wide range samples from different research centres) in creating more robust and comprehensive predictive signature. The results proved that SAM tool (*siggenes* package) is capable to elect the top discriminatory genes beyond level of FDR value = 0, with a model consisted of 8 genes showed the best discriminating power among the prediction models. Further validations of the model of 8 genes with additional external datasets are required to confirm it's specificity in discriminating OSCC from other tumour types of the body as well as other molecularly similar inflammatory processes (e.g., Atherosclerosis and periodontitis). If this model proves its ability in providing a unique signature of OSCC, a novel discovery will be reported for this analysis in the diagnosis of OSCC with emerging of new potential therapeutic targets.

The variability in the number of genes forming the final set of discriminating model following adjustment of specific criteria highlighted the impact of using different selected criteria and bioinformatics approaches in formulating the final set of gene discriminators. Hence, the result of current study suggests dealing with fold change as cut-off criteria with caution, especially in determining the final set of genes and identifying the subsequent dysregulated biological pathways, molecular functions, and processes, since molecular relationships in the system biology comprise not only

activation or inhibition of expression, but also protein-protein interactions, DNA-protein interactions and activation, localization, and inhibition of the corresponding proteins (Debily et al., 2009), and it is not surprising that microarrays results (especially those with a high fold change cut-off value criteria) may fail to identify some of the hub genes as significantly differentially expressed among the tumour versus normal samples. These genes might play a major role through protein activation, for instance. Instead, their modulations may be very subtle and below the threshold for reliable detection of differences of the microarray platform despite its high sensitivity (Debily et al., 2009). Alternatively, study suggestion is to consider comprehensive inclusion of the whole expressed genes, accompanied with using a Biological Annotation tool (like; IPA, KEGG, STRING, etc.) in an attempt to obtain a true and complete understanding of the biological concept of the system biology from which the sample extracted during collection, considering the fact that genes are highly interchangeable. However, setting a low threshold for a significant correlation between genes will result in the inclusion of many spurious links, whereas a high threshold will control the false-positive rate at the expense of omitting many genuine edges. For this reason, the suggestion of this study is the production of two separate sets; the one with less stringent criteria including large amount of genes is recruited for the purposes of subsequent biological annotation (like for identifying the dysregulated pathways and the affected transcriptional regulators), after which selection of a specific disease and its relative pathways will be focused on. Whilst the second set comprising a small number of genes can be recruited for the identification and monitoring of a specific tumour state (models of diagnostic, assessment, and prognostic biomarker).

Both common genes of the review study and the meta-genes of the meta-analysis study presented a significant common gene alteration. Among the differentially expressed genes, several potential diagnostic biomarkers were identified which have previously been reported in OSCC and HNSCC (Table 3.6). Notably, elevation of seven members of collagens (COL1A2, COL1A1, COL4A1, COL5A2, COL4A2, COL5A1, and COL11A1), six members of MMPS (MMP1, MMP3, MMP9, MMP10, MMP12, and MMP13), and five members of C-X-C motif chemokines

(CXCL1, CXCL6, CXCL9, CXCL10, and CXCL11) were elevated which accompanied with three keratins repression (KRT4, KRT13, and KRT15). Additional potential tumour key players, including elevation of SPP1, LAMC2, IL8, IFI6, ISG15, MYO1B, SERPINE1, BS2, AIM2, and PTHLH along with repression of CRNN, MAL, TGM3, EMP1, CRISP3, PPP1R3C, ECM1, IL1RN, SCEL, SPINK5, TGFBR3, CEACAM5, CEACAM7, TGM1, and TYRP1 genes were observed in tumour relative to normal cohorts.

6.2.3 Gene expression profiling reveals biological pathways responsible for phenotypic heterogeneity between UK and Sri Lankan oral squamous cell carcinomas (Chapter 4)

The gene expression profiles of UK and Sri Lankan OSCC compared to normal oral mucosa are similar in many respects to other oral cancer expression profiles reported in the literature and were mainly similar to each other. Great differences in the expression profiles between OSCC and normal tissue than between OSCCs of the two populations were observed. However, molecular heterogeneity in each tumour populations (such heterogeneity is more marked in UK OSCC, which displays a wide range of clinical behaviour). Such molecular heterogeneity may act to reduce differences seen in expression profiles of the two populations. Despite that, IPA showed common perturbed canonical pathways which indicate that the development of OSCCs is mediated, to some extent by similar biological pathways despite the differences related to race, ethnicity, lifestyle, and/or exposure to environmental carcinogens.

IPA comparison core analysis of normal as well as tumour samples between the two populations displayed highly activated “Cell-mediated Immune Response” processes in Sri Lankan normal and tumours relative to UK cohorts which presumably play a key role in resistance of Sri Lankan patients to the invasiveness, metastasis, and recurrence of the disease. Sri Lankan consumption of spicy food, including curcumin/curry in specific, might be responsible for this activation along with downregulation of growth factor receptors (EGFR, IGFR, RARA, EPOR, THRA, and NMUR2) that might as well, contributed in the phenotypic behaviour of the two

tumour groups. The same findings were observed comparing tumours from the two population groups which confirms constant existence of the events between the two populations.

Smaller number of either elevated or repressed genes in either UK or Sri Lankan OSCC but not both contributed to a number of differentially regulated cancer-associated pathways and specifically to the aggressive behaviour of UK tumours; Including upregulation of MMP3, POSTN, LAMC2, and downregulation of DUSP1 and SLPI in UK. Conversely, CD80 upregulation and TWIST1 downregulation, anti-invasive associated incidents, observed in Sri Lankan tumours. It may be the case that expression differences in key pathways contribute to the observed phenotypic differences and so expression differences described here may underpin some of the phenotypic differences between UK and Sri Lankan OSCC. Tumour heterogeneity, technical variation and data processing may all have an effect on the expression profiling data and all such variables should consider when interpreting the results.

Differences in tumour expression profiles of the two populations highlighted a few potentially genes; including lower expression of CDKN1B in UK tumours, a phenomenon associated with high-risk tumours only, along with lower expression of PSIP1, SAMSN1, CCNG2, and NCOA2 genes which played a protective role in tumorigenic process. Likewise, higher expressions of CCT2 and DHCR24 in UK tumours were contributed in rapidly proliferating tumour and melanoma metastases respectively.

6.2.4 Evidences support theory that cancer is a metabolic disease

Numerous results displayed in this study support the theory considering cancer as a metabolic disorder (Warburg effect) (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956). The top dysregulated pathways in each and both UK and Sri Lankan populations (Chapter 4) showed that defects in mitochondrial respiration could be the primary reason for induction of tumorigenic transformation reflected in dysregulation of oxidative phosphorylation and mitochondrial dysfunction pathways, shifted tumour cells to glycolysis affecting citrate cycle with remarkable

dysregulation of fatty acids metabolism. As consequent to the above defects, significant alteration in gene expression can be explained as an attempt from the tumour cells to cope with the initial disorder, repair, and overcome the fatal causative event(s). Further studies with a large size and wide range of populations are required to highlight and validate the above phenomena.

The review study of metabolic profile (Chapter 5) showed elevated glucose level along with significantly increased fatty acids metabolism which has been previously linked to "Warburg effect" (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956), and assumes that tumours are relied on glycolysis as a main source of energy, even in the presence of oxygen. Many solid tumours show an increased glycolytic metabolism, which has, for OSCC, been associated with the over-expression of glucose transporters especially Glut-1 (SLC2A1) (Kunkel et al., 2003), which showed an elevation in the current study as well.

Furthermore, IPA "Metabolomics Analysis" of combined meta-genes and the 45 metabolites (Chapter 5) displayed energy production, lipid and fatty acid metabolism and degradation processes among the corresponding functions of the highest score networks emphasizing the significant metabolic activity of tumour relying on glycolysis as a main source of energy alternative to normal mitochondrial respiration which has been previously linked to "Warburg effect" (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956) while lipid degradation and amino acid metabolism provide the sufficient fuel and environment for glycolysis.

6.2.5 Gene expression profile of perineural invasive versus non- invasive tumour samples from UK

Three genes appeared over-expressed in perineural invasive tumours of UK group; PGM1, ENO2, and ILK. ENO2 (enolase2) reported a prognostic biomarker in metastatic, high grade, and invasiveness of various types of cancer, including breast, lung, bladder, renal, prostate, colorectal cancer, and OSCC (Abba et al., 2005, van Zandwijk et al., 1992, Di Carlo et al., 1990, Vandepol et al., 1994, Buzaid et al., 1994, Rasmuson et al., 1993, Jacobsen et al., 1990, Nishihara et al., 2009, Zufferey et al., 2001, Famulski et al., 2001). ILK has been associated with multiple cellular

functions including cell migration, proliferation, and signal transduction. Recent studies reported ILK as a biomarker associated with laryngeal, lung, colon, ovarian cancer as well as metastatic behaviour of gastric carcinoma (Ito et al., 2003, Wu et al., 2006, Takanami, 2005, Bravou et al., 2006, Du et al., 2008, Loessner et al., 2009).

GO term (Gene Ontology, www.geneontology.org/) showed both PGM1 and ENO2 enzymes involved in magnesium ion binding whereas the all three genes involved in phosphorylation and interconversion activities of certain proteins and amino acids like phosphopyruvate hydratase, lyase, phosphoglucomutase, isomerase, and phosphotranferase.

6.2.6 Gene expression profile of high-stage versus early and low- stage tumour samples from both UK and Sri Lankan tumours

SAM supervised analysis identified 12 genes separating "high stage" from "low and early stage" OSCCs of both population groups. Most of genes appeared over-expressed in high stage tumours and associated with different cancer types and cellular processes associated with tumours.

Most of differentially upregulated genes in high stage tumours exhibited involvement in variant biological processes crucial in driven tumour lesion to late stage and metastasis with distinct roles in cellular migration, remodelling, outgrowth, invasion, differentiation, colony formation, cell cycle progression, activation, proliferation, and angiogenesis.

Knockdown of SEMA3F (member of semaphorins) cancer cells of breast rescued the aggressive phenotypes and tumour invasion by inducing suppressive cell microenvironment (Xiong et al., 2012) as well as pancreatic (Matsushita et al., 2007). An extensive review explained how Scatter factors and semaphorins, together with their receptors, help to orchestrate metastasis programme of neoplastic cells (Trusolino and Comoglio, 2002).

Suswam group (Suswam et al., 2005) noticed that KHSRP involved in IL8 RNA stabilization, which plays an integral role in promoting the malignant phenotype in breast cancer (Chen et al., 2001, Gherzi et al., 2004).

Jauliak group identified NFAT5 isoform expressed in invasive human ductal breast carcinomas and participate in promoting invasion of human breast and colon carcinomas. The transcriptional activity of NFAT5 is induced by alpha-6 and beta-4 integrin signalling clustering in the presence of chemo-attractants, resulting in enhanced cell migration and is involved in promoting carcinoma invasion (Jauliac et al., 2002). Further study demonstrated NFAT5 contribution in progression of gastrointestinal cancer (Chen et al., 2011b).

CDC42BPA plays a role in the regulation of cytoskeleton reorganization and cell migration (Tan et al., 2008a). It has been demonstrated that invasion can be generated by Cdc42-MRCK signalling (Wilkinson et al., 2005) as well as tumour progression (Benitah et al., 2004).

6.2.7 Review study of metabolic profile

Metabolomic signature recruited NMR and MS spectral technologies to study the body end products in reflecting the behaviour of oral cancer and body response against it. The current review enriched our knowledge via better understanding the characteristic behaviour of oral cancer and the alteration of system biology responding to tumorigenic assault at metabolic level. Diagnostic tool biomarker from metabolic signature of OSCC is continuously under improving progress.

Among the metabolites identified commonly altered in the studies of review, include several involved in anti-oxidation, free radical scavenging, detoxification of xenobiotics and chemicals, (glutathione, creatine, glutamic acid, taurine, glycine, tryptophan, and pyruvic acid), fatty acid metabolism, lipid degradation, and glycolysis. These molecular pathways and cellular functions are typical signature for oral cancer and already described by previous studies. Glutathione- mediated detoxification, methylglyoxal detoxification pathways emphasise the crucial role of

glutathione which also needed for the detoxification of methylglyoxal, a toxin produced as a by-product of metabolism.

Several detected metabolic markers were involved in associated network function (Figure 5.2), signalling and metabolic pathways (5.4.2.3), and biomarker analysis (Figure 5.3) representing glucose and glutathione metabolites as hub nodes associated with detoxification process, glycolysis, energy production, and amino acid metabolism. Biomarker analysis identified both glucose and glutathione over-expression as verified biomarkers of other types of cancer along with sarcosine metabolite (Figure 5.3).

Elevated glucose level along with significantly increased fatty acids metabolism may be linked to the unique behaviour of oral cancer. This represents a typical signature in cancer patients, and it has been previously linked to "Warburg effect" (Galluzzi et al., 2006, Seyfried and Shelton, 2010, Warburg, 1956), which assumes that tumours are relied on glycolysis as a main source of energy, even in the presence of oxygen. Many solid tumours show an increased glycolytic metabolism, which has, for OSCC, been associated with the over-expression of glucose transporters, especially Glut-1 (SLC2A1) (Kunkel et al., 2003), which showed an elevation in the current study as well.

The multifactorial nature of oncogenesis and the heterogeneity in oncogenic pathways make detection of all cancers of a particular organ unlikely detected when using a single biomarker with high specificity and sensitivity. Therefore, prediction models to identify combinations of biomarkers that can identify OSCC patients among all the samples can improve the task.

6.2.8 IPA metabolomics analysis of both meta-genes and the 45 metabolites

"Metabolomics Analysis" tool from IPA, provided the common molecular pathways and cellular functions related to both metabolomic candidates harvested from the review study (45 metabolites) and the differential transcriptomic profile from the meta-analysis study conducted in chapter 3 (8995 genes). The corresponding

pathways and functions of the networks revealed energy production, lipid and fatty acid metabolism and degradation processes highlighting the significant metabolic activity of oral cancer relying on glycolysis as a main source of energy alternative to normal mitochondrial respiration.

6.2.9 Direct integration between the 45 metabolomic profile and the 32 OSCC and HNSCC transcriptomic biomarkers

Considering the transcriptional and metabolic changes that accompanied carcinogenic exposure during OSCC, direct biological integration highlighted ethanol, a major risk factor in oral cancer, created essential hub molecule in the interactive network. Acute alcohol intake suggested attenuating monocyte-derived chemokine production responding to a subsequent immune challenge (Szabo et al., 1999, Taieb et al., 2002, Mas et al., 2010). Ethanol may influence proliferative cells, inducing oxidative DNA damage, altering the lipid-containing permeability barrier of stratified squamous epithelium, and can increase mitochondrial DNA damage (Ogden, 2005). It also induced the production of matrix metalloproteinases, stimulated both proliferation and migration or inhibited the tissue inhibitors of metalloproteinases (TIMPs) (Pires et al., 2008) as well as activation of STAT1/3 signalling pathways , a significant cancer indicator (Norkina et al., 2008), as well as elevation of COL1A2 (Nieto et al., 2000), and SERPINE1 (Booyse et al., 1999). Moreover, ethanol decreased glutathione and cytochrome c oxidase protein level, the essential detoxifying agents, (Bailey et al., 2009).

Glucose, the highest interacting metabolite with correlative transcriptomic biomarkers, denoted to glycolysis, the unique cancer behaviour including oral cancer. Elevated level of glucose with its transporters, especially SLC2A1, considers a common characteristic of human malignancies, including HNSCC where both glucose transport and its metabolism may play a key role in the progression of OSCC (Kunkel et al., 2003). Glucose also increased IL1B, COX2 (Huang and Siragy, 2009), production of reactive oxygen species (ROS) (Chen et al., 2012), IL6 secretion (Solini et al., 2000), expression of human SERPINE1 (Murphy et al., 1999, Jeong et al., 2011, Chen et al., 1998), as well as the production of MMP1 (Zayzafoon

et al., 2002), MMP9 (Bai et al., 2005), fibronectin, COL1A1, and COL4A1 (Lam et al., 2003).

Elevation of glutathione, taurine, tryptophan, and L-glutamic acid levels constitute the main antioxidant and detoxification system essential to neutralise and remove the toxic by-products that produced during tumourigenesis. Best known role of glutathione enzymes is as cell housekeepers in the detoxification of xenobiotics (Scholz et al., 1989). Glutathione inhibited COX1 and COX2 activities (Rao et al., 2001), decreased transcription of SERPINE1 mRNA that increased by oxidative stress of cells (Vulin and Stanley, 2004), decreased expression of human IL6 and IL1B mRNA (Junn et al., 2000, Llacuna et al., 2009). Taurine acts as a detoxifying agent against hypochloric acid by forming the stable and less toxic taurine chloramine (Quinn et al., 1996). Taurine also diminished the induction of COX2 (Warskulat et al., 1997), reduced mRNA of pro-inflammatory mediators encoding MMP9, and fibronectin (Song et al., 2011), as well as inhibited the production of IL6, IL8, and IL1B (Park et al., 1998, Joo et al., 2009). L-Tryptophan established a reduction of oxidative stress and the expression of interleukins IL6, IL1B and IL8 (Kim et al., 2010). Similarly, glutamic acid exhibited a declination by IL1B protein (Kelly et al., 2001) and IL6 (D'Arcangelo et al., 2000).

Profound reading of the simultaneous transcriptional and metabolic interactive changes that accompanied tumourigenesis of OSCC revealed an intimate relation between genetic and metabolic system biology during carcinogenic transformation. To obtain meaningful and comprehensive understanding, more efforts required to identify additional key players from a larger interactive network alongside larger review study and validation to confirm the role of each player in this complex context. However, the current analysis already highlighted interleukins, MMPs, COX2, SERPINE1, and STAT1 as major key players from transcriptomic profile platform transecting specific players from the platform of metabolome which include; ethanol and its metabolites establishing the carcinogenic factor, glucose and its relative metabolites as representative by-products of OSCC, glutathione, creatine, glutamic acid, taurine, glycine, tryptophan, and pyruvic acid as a reparative

antioxidant and detoxifying elements against tumorigenic assault of OSCC (Figure 5.6).

Such an interactive pattern may represent a clinically useful surrogate of the presence of OSCC and can be employed to form the basis of diagnostic device of OSCC. Further investigation of the interactive network is required. Additionally, to get more comprehensive and complete story, further expanding of the interactive network is required by including all altered genes involved in cancer disease from the meta-analysis study (401genes rather than only 32 OSCC and HNSCC biomarkers that included in the current analysis) for integration with the 45 metabolomic OSCC biomarkers (biological level integration). Clinical application of the current interactive signature as diagnostic tool can also carried out by simple cross-correlation of both metabolomic and transcriptomic candidate biomarkers in the interactive model across the examined samples (data level integration).

6.3 Suggestion for the future work

The various findings of this investigation have also raised new questions which should be addressed in the future studies to fully understand the mechanisms involved in the tumorigenic transformation of OSCC, as well as to improve the predictive models for diagnosis, assessment, and prognosis of the disease. Some of the suggestions are listed below:

1- Development of a new meta-analysis study through implementation of a larger number of samples as well as establishment and investigation of the signatures responsible for the invasion, recurrence, and metastasis of OSCC.

2- Further validation of the predictive models (especially the models consisting of the 8 and 38 genes) using different methods to prepare for their employment as diagnostic tool for OSCC.

3- Additional assessment and validation of the harvested gene expression models responsible for the discrimination of perineural invasion from non- or lesser-invasive OSCC.

4- Additional comprehensive investigation of the interactive networks correlating metabolic and transcriptomic profiles to improve and further understand the interactive signature of what can represent a clinically useful surrogate of the presence of OSCC and can be employed to form the basis of the diagnostic device of the disease.

5- Collection of additional wide range samples from Indo-Asian and Western countries to validate the results defining the heterogenic phenotypic behaviours of OSCC across distinct populations.

6- Establishment of a new meta-analysis study differentiating gene expression profile of LCM samples and those harvested from tumour bulk samples of OSCC in an attempt to understand the tumour-stromal interaction of the disease.

7- As various results in the current work suggested a defect in mitochondrial function which might be responsible, or even, the primary cause of tumour, further studies required to measure gene expression profiles of tumour patients who receive a therapeutic dietary supplement accompanied with a periodic fasting.

Supplementary Materials

For supplementary files and manuscripts refer to Supplementary Files CD

References

1988. Alcohol drinking. IARC Working Group, Lyon, 13-20 October 1987. *IARC Monogr Eval Carcinog Risks Hum,* 44**,** 1-378.

Abba, M. C., Hu, Y., Sun, H., Drake, J. A., Gaddis, S., Baggerly, K., Sahin, A. & Aldaz, C. M. 2005. Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC Genomics,* 6**,** 37.

Abdel-fattah, M. Reply from Authors re: Ricarda M. Bauer. Female Slings: Where Do We Stand? Eur Urol. In press. http://dx.doi.org/10.1016/ j.eururo.2012.05.036: Synthetic Midurethral Slings Stand Firmly on Solid Ground as a Surgical Treatment for Stress Urinary Incontinence in Women. *European Urology*.

Ackerstaff, E., Glunde, K. & Bhujwalla, Z. M. 2003. Choline phospholipid metabolism: a target in cancer cells? *J Cell Biochem,* 90**,** 525-33.

Adalsteinsson, E., Spielman, D. M., Pauly, J. M., Terris, D. J., Sommer, G. & Macovski, A. 1998. Feasibility study of lactate imaging of head and neck tumors. *NMR Biomed,* 11**,** 360-9.

Aishima, S., Matsuura, S., Terashi, T., Taguchi, K., Shimada, M., Maehara, Y. & Tsuneyoshi, M. 2004. Aberrant expression of laminin gamma 2 chain and its prognostic significance in intrahepatic cholangiocarcinoma according to growth morphology. *Mod Pathol,* 17**,** 938-45.

Al Moustafa, A. E., Alaoui-Jamali, M. A., Batist, G., Hernandez-Perez, M., Serruya, C., Alpert, L., Black, M. J., Sladek, R. & Foulkes, W. D. 2002. Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene,* 21**,** 2634-2640.

Alevizos, I., Mahadevappa, M., Zhang, X., Ohyama, H., Kohno, Y., Posner, M., Gallagher, G. T., Varvares, M., Cohen, D., Kim, D., Kent, R., Donoff, R. B., Todd, R., Yung, C. M., Warrington, J. A. & Wong, D. T. 2001. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene,* 20**,** 6196-204.

Almondes, K. G. d. S., Leal, G. V. d. S., Cozzolino, S. M. F., Philippi, S. T. & Rondó, P. H. d. C. 2010. O papel das selenoproteínas no câncer. *Revista da Associação Médica Brasileira,* 56**,** 484-488.

Ameyar, M., Wisniewska, M. & Weitzman, J. B. 2003. A role for AP-1 in apoptosis: the case for and against. *Biochimie,* 85**,** 747-52.

Ang, K. K., Berkey, B. A., Tu, X. Y., Zhang, H. Z., Katz, R., Hammond, E. H., Fu, K. K. & Milas, L. 2002. Impact of epidermal growth factor receptor expression on survival and pattern of relapse in patients with advanced head and neck carcinoma. *Cancer Research,* 62**,** 7350-7356.

Anneroth, G., Batsakis, J. & Luna, M. 1987. Review of the literature and a recommended system of malignancy grading in oral squamous cell carcinomas. *Scand J Dent Res,* 95**,** 229-49.

Aoki-Kinoshita, K. & Kanehisa, M. 2007. Gene Annotation and Pathway Mapping in KEGG. *In:* BERGMAN, N. (ed.) *Comparative Genomics.* Humana Press.

Aoki-Yoshida, A., Aoki, R. & Takayama, Y. 2013. Protective effect of pyruvate against UVB-induced damage in HaCaT human keratinocytes. *J Biosci Bioeng,* 115**,** 442-8.

Arias-Mendoza, F., Payne, G. S., Zakian, K. L., Schwarz, A. J., Stubbs, M., Stoyanova, R., Ballon, D., Howe, F. A., Koutcher, J. A., Leach, M. O., Griffiths, J. R., Heerschap, A., Glickson, J. D., Nelson, S. J., Evelhoch, J. L., Charles, H. C. & Brown, T. R. 2006. In vivo 31P MR spectral patterns and reproducibility in cancer patients studied in a multi-institutional trial. *NMR Biomed,* 19**,** 504-12.

Arima, Y., Matsueda, S., Yano, H., Harada, M. & Itoh, K. 2005. Parathyroid hormone-related protein as a common target molecule in specific immunotherapy for a wide variety of tumor types. *International Journal of Oncology,* 27**,** 981-988.

Arrigo, A. P., Simon, S., Gibert, B., Kretz-Remy, C., Nivon, M., Czekalla, A., Guillet, D., Moulin, M., Diaz-Latoud, C. & Vicart, P. 2007. Hsp27 (HspB1) and alphaB-crystallin (HspB5) as therapeutic targets. *FEBS Lett,* 581**,** 3665-74.

Arthur, M. J. P. 2000. Fibrogenesis - II. Metalloproteinases and their inhibitors in liver fibrosis. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 279**,** G245-G249.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. & Gene Ontology, C. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics,* 25**,** 25-29.

Assaily, W. & Benchimol, S. 2006. Differential utilization of two ATP-generating pathways is regulated by p53. *Cancer Cell,* 10**,** 4-6.

Atula, T., Hedstrom, J., Ristimaki, A., Finne, P., Leivo, I., Markkanen-Leppanen, M. & Haglund, C. 2006. Cyclooxygenase-2 expression in squamous cell carcinoma of the oral cavity and pharynx: association to p53 and clinical outcome. *Oncol Rep,* 16**,** 485-90.

Ayan, N., Ayan, I., Alatli, C., Guler, S. D., Dalkilic, C., Cinar, F. & Dogan, O. 2000. P53 overexpression in normal oral mucosa of heavy smokers. *J Exp Clin Cancer Res,* 19**,** 525-9.

Azuma, H., Inamoto, T., Sakamoto, T., Kiyama, S., Ubai, T., Shinohara, Y., Maemura, K., Tsuji, M., Segawa, N., Masuda, H., Takahara, K., Katsuoka, Y. & Watanabe, M. 2003. gamma-Aminobutyric acid as a promoting factor of cancer metastasis; Induction of matrix metalloproteinase production is potentially its underlying mechanism. *Cancer Research,* 63**,** 8090-8096.

Bachelor, M. A., Lu, Y. & Owens, D. M. 2011. L-3-Phosphoserine phosphatase (PSPH) regulates cutaneous squamous cell carcinoma proliferation independent of L-serine biosynthesis. *J Dermatol Sci,* 63**,** 164-72.

Bagchi, A. & Mills, A. A. 2008. The quest for the 1p36 tumor suppressor. *Cancer Res,* 68**,** 2551-6.

Bai, Y. L., Huang, H. C., Li, J. Z., Zhao, Y. Y. & Wang, H. Y. 2005. [High glucose regulates the production of MMP-9 in podocyte through ERK1/2 signal pathway]. *Zhonghua Yi Xue Za Zhi,* 85**,** 1451-5.

Bailey, S. M., Mantena, S. K., Millender-Swain, T., Cakir, Y., Jhala, N. C., Chhieng, D., Pinkerton, K. E. & Ballinger, S. W. 2009. Ethanol and tobacco smoke increase hepatic steatosis and hypoxia in the hypercholesterolemic apoE(-/-) mouse: Implications for a "multihit" hypothesis of fatty liver disease. *Free Radical Biology and Medicine,* 46**,** 928-938.

Bales, J. R., Higham, D. P., Howe, I., Nicholson, J. K. & Sadler, P. J. 1984. Use of High-Resolution Proton Nuclear Magnetic-Resonance Spectroscopy for Rapid Multi-Component Analysis of Urine. *Clinical Chemistry,* 30**,** 426-432.

Banerjee, A. G., Bhattacharyya, I., Lydiatt, W. M. & Vishwanatha, J. K. 2003. Aberrant expression and localization of decorin in human oral dysplasia and squamous cell carcinoma. *Cancer Res,* 63**,** 7769-76.

Banerjee, A. G., Gopalakrishnan, V. K., Bhattacharya, I. & Vishwanatha, J. K. 2002. Deregulated cyclooxygenase-2 expression in oral premalignant tissues. *Mol Cancer Ther,* 1**,** 1265-71.

Banerji, U., de Bono, J., Judson, I., Kaye, S. & Workman, P. 2008. Biomarkers in early clinical trials: The committed and the skeptics. *Clinical Cancer Research,* 14**,** 2512-2513.

Bao, S., Ouyang, G., Bai, X., Huang, Z., Ma, C., Liu, M., Shao, R., Anderson, R. M., Rich, J. N. & Wang, X. F. 2004. Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. *Cancer Cell,* 5**,** 329-39.

Barsky, S. H., Rao, C. N., Hyams, D. & Liotta, L. A. 1984a. Characterization of a laminin receptor from human breast carcinoma tissue. *Breast Cancer Res Treat,* 4**,** 181-8.

Barsky, S. H., Rao, C. N., Williams, J. E. & Liotta, L. A. 1984b. Laminin Molecular Domains Which Alter Metastasis in a Murine Model. *Journal of Clinical Investigation,* 74**,** 843-848.

Bathen, T. F., Engan, T., Krane, J. & Axelson, D. 2000. Analysis and classification of proton NMR spectra of lipoprotein fractions from healthy volunteers and patients with cancer or CHD. *Anticancer Res,* 20**,** 2393-408.

Bathen, T. F., Jensen, L. R., Sitter, B., Fjosne, H. E., Halgunset, J., Axelson, D. E., Gribbestad, I. S. & Lundgren, S. 2007. MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. *Breast Cancer Res Treat,* 104**,** 181-9.

Beckonert, O., Monnerjahn, J., Bonk, U. & Leibfritz, D. 2003. Visualizing metabolic changes in breast-cancer tissue using 1H-NMR spectroscopy and self-organizing maps. *NMR Biomed,* 16**,** 1-11.

Beckwith-Hall, B. M., Thompson, N. A., Nicholson, J. K., Lindon, J. C. & Holmes, E. 2003. A metabonomic investigation of hepatotoxicity using diffusion-edited H-1 NMR spectroscopy of blood serum. *Analyst,* 128**,** 814-818.

Belbin, T. J., Schlecht, N. F., Smith, R. V., Adrien, L. R., Kawachi, N., Brandwein-Gensler, M., Bergman, A., Chen, Q., Childs, G. & Prystowsky, M. B. 2008. Site-specific molecular signatures predict aggressive disease in HNSCC. *Head Neck Pathol,* 2**,** 243-56.

Belbin, T. J., Singh, B., Barber, I., Socci, N., Wenig, B., Smith, R., Prystowsky, M. B. & Childs, G. 2002. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Research,* 62**,** 1184-1190.

Belbin, T. J., Singh, B., Smith, R. V., Socci, N. D., Wreesmann, V. B., Sanchez-Carbayo, M., Masterson, J., Patel, S., Cordon-Cardo, C., Prystowsky, M. B. & Childs, G. 2005. Molecular profiling of tumor progression in head and neck cancer. *Arch Otolaryngol Head Neck Surg,* 131**,** 10-8.

Benitah, S. A., Valeron, P. F., van Aelst, L., Marshall, C. J. & Lacal, J. C. 2004. Rho GTPases in human cancer: an unresolved link to upstream and downstream transcriptional regulation. *Biochim Biophys Acta,* 1705**,** 121-32.

Bergmann, C., Strauss, L., Wang, Y., Szczepanski, M. J., Lang, S., Johnson, J. T. & Whiteside, T. L. 2008. T regulatory type 1 cells in squamous cell carcinoma of the head and neck: mechanisms of suppression and expansion in advanced disease. *Clin Cancer Res,* 14**,** 3706-15.

Bernard, P. S. & Wittwer, C. T. 2002. Real-time PCR technology for cancer diagnostics. *Clin Chem,* 48**,** 1178-85.

Bezabeh, T., Odlum, O., Nason, R., Kerr, P., Sutherland, D., Patel, R. & Smith, I. C. 2005. Prediction of treatment response in head and neck cancer by magnetic resonance spectroscopy. *AJNR Am J Neuroradiol,* 26**,** 2108-13.

Bhattacharyya, B. J., Banisadr, G., Jung, H., Ren, D., Cronshaw, D. G., Zou, Y. & Miller, R. J. 2008. The chemokine stromal cell-derived factor-1 regulates GABAergic inputs to neural progenitors in the postnatal dentate gyrus. *J Neurosci,* 28**,** 6720-30.

Bhawal, U. K., Ozaki, Y., Nishimura, M., Sugiyama, M., Sasahira, T., Nomura, Y., Sato, F., Fujimoto, K., Sasaki, N., Ikeda, M. A., Tsuji, K., Kuniyasu, H. & Kato, Y. 2005. Association of expression of receptor for advanced glycation end products and invasive activity of oral squamous cell carcinoma. *Oncology,* 69**,** 246-55.

Blaylock, R. L. 2013. Immunoexcitatory mechanisms in glioma proliferation, invasion and occasional metastasis. *Surg Neurol Int,* 4**,** 15.

Blot, W. J., McLaughlin, J. K., Winn, D. M., Austin, D. F., Greenberg, R. S., Preston-Martin, S., Bernstein, L., Schoenberg, J. B., Stemhagen, A. & Fraumeni, J. F., Jr. 1988. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res,* 48**,** 3282-7.

Bolan, P. J., Meisamy, S., Baker, E. H., Lin, J., Emory, T., Nelson, M., Everson, L. I., Yee, D. & Garwood, M. 2003. In vivo quantification of choline compounds in the breast with 1H MR spectroscopy. *Magn Reson Med,* 50**,** 1134-43.

Bonifazi, M., Malvezzi, M., Bertuccio, P., Edefonti, V., Garavello, W., Levi, F., La Vecchia, C. & Negri, E. 2011. Age-period-cohort analysis of oral cancer mortality in Europe: The end of an epidemic? *Oral Oncology,* 47**,** 400-407.

Booyse, F. M., Aikens, M. L. & Grenett, H. E. 1999. Endothelial cell fibrinolysis: Transcriptional regulation of fibrinolytic protein gene expression (t-PA, n-PA, and PAI-1) by low alcohol. *Alcoholism-Clinical and Experimental Research,* 23**,** 1119-1124.

Boros, L. G., Lerner, M. R., Morgan, D. L., Taylor, S. L., Smith, B. J., Postier, R. G. & Brackett, D. J. 2005. [1,2-C-13(2)]-D-glucose profiles of the serum, liver, pancreas, and DMBA-induced pancreatic tumors of rats. *Pancreas,* 31**,** 337-343.

Bouvet, M., Nardin, S. R., Burton, D. W., Lee, N. C., Yang, M., Wang, X. O., Baranov, E., Behling, C., Moossa, A. R., Hoffman, R. M. & Deftos, L. J. 2002. Parathyroid hormone-related protein as a novel tumor marker in pancreatic adenocarcinoma. *Pancreas,* 24**,** 284-290.

Boyland, E. & Chasseau.Lf 1969. Role of Glutathione and Gluthathione S-Transferases in Mercapturic Acid Biosynthesis. *Advances in Enzymology and Related Areas of Molecular Biology,* 32**,** 173-&.

Bravou, V., Klironomos, G., Papadaki, E., Taraviras, S. & Varakis, J. 2006. ILK over-expression in human colon cancer progression correlates with activation of beta-catenin, down-regulation of E-cadherin and activation of the Akt-FKHR pathway. *J Pathol,* 208**,** 91-9.

Brindle, J. T., Nicholson, J. K., Schofield, P. M., Grainger, D. J. & Holmes, E. 2003. Application of chemometrics to H-1 NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *Analyst,* 128**,** 32-36.

Broders, A. C. 1926. Carcinoma - Grading and practical application. *Archives of Pathology & Laboratory Medicine,* 2**,** 376-381.

Brown, K., Strathdee, D., Bryson, S., Lambie, W. & Balmain, A. 1998. The malignant capacity of skin tumours induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr Biol,* 8**,** 516-24.

Brown, W. A., Skinner, S. A., Malcontenti-Wilson, C., Vogiagis, D. & O'Brien, P. E. 2001. Non-steroidal anti-inflammatory drugs with activity against either cyclooxygenase 1 or cyclooxygenase 2 inhibit colorectal cancer in a DMH rodent model by inducing apoptosis and inhibiting cell proliferation. *Gut,* 48**,** 660-666.

Bryne, M., Koppang, H. S., Lilleng, R., Stene, T., Bang, G. & Dabelsteen, E. 1989. New Malignancy Grading Is a Better Prognostic Indicator Than Broders Grading in Oral Squamous-Cell Carcinomas. *Journal of Oral Pathology & Medicine,* 18**,** 432-437.

Buchholz, A., Hurlebaus, J., Wandrey, C. & Takors, R. 2002. Metabolomics: quantification of intracellular metabolite dynamics. *Biomol Eng,* 19**,** 5-15.

Bunn, P. J., Jr. 2003. Early detection of lung cancer using serum RNA or DNA markers: ready for "prime time" or for validation? *J Clin Oncol,* 21**,** 3891-3.

Butow, R. A. & Avadhani, N. G. 2004. Mitochondrial signaling: the retrograde response. *Mol Cell,* 14**,** 1-15.

Buttyan, R., Olsson, C. A., Pintar, J., Chang, C., Bandyk, M., Ng, P. Y. & Sawczuk, I. S. 1989. Induction of the TRPM-2 gene in cells undergoing programmed death. *Mol Cell Biol,* 9**,** 3473-81.

Buzaid, A. C., Sandler, A. B., Hayden, C. L., Scinto, J., Poo, W. J., Clark, M. B. & Hotchkiss, S. 1994. Neuron-Specific Enolase as a Tumor-Marker in Metastatic Melanoma. *American Journal of Clinical Oncology-Cancer Clinical Trials,* 17**,** 430-431.

Bylesjo, M., Rantalainen, M., Nicholson, J. K., Holmes, E. & Trygg, J. 2008. K-OPLS package: kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space. *BMC Bioinformatics,* 9**,** 106.

Califano, J., van der Riet, P., Westra, W., Nawroz, H., Clayman, G., Piantadosi, S., Corio, R., Lee, D., Greenberg, B., Koch, W. & Sidransky, D. 1996. Genetic

progression model for head and neck cancer: implications for field cancerization. *Cancer Res,* 56**,** 2488-92.

Cao, W., Zhang, Z. Y., Xu, Q., Sun, Q., Yan, M., Zhang, J., Zhang, P., Han, Z. G. & Chen, W. T. 2010. Epigenetic silencing of MAL, a putative tumor suppressor gene, can contribute to human epithelium cell carcinoma. *Mol Cancer,* 9**,** 296.

Carinci, F., Lo Muzio, L., Piattelli, A., Rubini, C., Chiesa, F., Ionna, F., Palmieri, A., Maiorano, E., Pastore, A., Laino, G., Dolci, M. & Pezzetti, F. 2005. Potential markers of tongue tumor progression selected by cDNA microarray. *Int J Immunopathol Pharmacol,* 18**,** 513-24.

Carraro, S., Rezzi, S., Reniero, F., Heberger, K., Giordano, G., Zanconato, S., Guillou, C. & Baraldi, E. 2007. Metabolomics applied to exhaled breath condensate in childhood asthma. *Am J Respir Crit Care Med,* 175**,** 986-90.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K. & Gingeras, T. R. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell,* 116**,** 499-509.

Chan, G., Boyle, J. O., Yang, E. K., Zhang, F., Sacks, P. G., Shah, J. P., Edelstein, D., Soslow, R. A., Koki, A. T., Woerner, B. M., Masferrer, J. L. & Dannenberg, A. J. 1999. Cyclooxygenase-2 expression is up-regulated in squamous cell carcinoma of the head and neck. *Cancer Research,* 59**,** 991-994.

Chang, B. W., Kim, D. H., Kowalski, D. P., Burleson, J. A., Son, Y. H., Wilson, L. D. & Haffty, B. G. 2004. Prognostic significance of cyclooxygenase-2 in oropharyngeal squamous cell carcinoma. *Clin Cancer Res,* 10**,** 1678-84.

Chen, C., Grennan, K., Badner, J., Zhang, D., Gershon, E., Jin, L. & Liu, C. 2011a. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One,* 6**,** e17238.

Chen, C., Mendez, E., Houck, J., Fan, W., Lohavanichbutr, P., Doody, D., Yueh, B., Futran, N. D., Upton, M., Farwell, D. G., Schwartz, S. M. & Zhao, L. P. 2008a. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev,* 17**,** 2152-62.

Chen, C. W., Chang, Y. H., Tsi, C. J. & Lin, W. W. 2003. Inhibition of IFN-gamma-mediated inducible nitric oxide synthase induction by the peroxisome proliferator-activated receptor gamma agonist, 15-deoxy-delta 12,14-prostaglandin J2, involves inhibition of the upstream Janus kinase/STAT1 signaling pathway. *J Immunol,* 171**,** 979-88.

Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., Stoecklin, G., Moroni, C., Mann, M. & Karin, M. 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell,* 107**,** 451-64.

Chen, J. S., Chen, Y. H., Huang, P. H., Tsai, H. Y., Chen, Y. L., Lin, S. J. & Chen, J. W. 2012. Ginkgo biloba extract reduces high-glucose-induced endothelial adhesion by inhibiting the redox-dependent interleukin-6 pathways. *Cardiovasc Diabetol,* 11**,** 49.

Chen, L., Linsley, P. S. & Hellstrom, K. E. 1993. Costimulation of T cells for tumor immunity. *Immunol Today,* 14**,** 483-6.

Chen, M., Sastry, S. K. & O'Connor, K. L. 2011b. Src kinase pathway is involved in NFAT5-mediated S100A4 induction by hyperosmotic stress in colon cancer cells. *Am J Physiol Cell Physiol,* 300**,** C1155-63.

Chen, Y. Q., Su, M., Walia, R. R., Hao, Q., Covington, J. W. & Vaughan, D. E. 1998. Sp1 sites mediate activation of the plasminogen activator inhibitor-1 promoter by glucose in vascular smooth muscle cells. *J Biol Chem,* 273**,** 8225-31.

Chen, Y. W., Yu, E. H., Wu, T. H., Lo, W. L., Li, W. Y. & Kao, S. Y. 2008b. Histopathological factors affecting nodal metastasis in tongue cancer: analysis of 94 patients in Taiwan. *Int J Oral Maxillofac Surg,* 37**,** 912-6.

Chiang, Y. Y., Tsai, M. H., Lin, T. Y. & Chiang, I. P. 2008. Expression profile of metastasis-related genes in invasive oral cancers. *Histol Histopathol,* 23**,** 1213-22.

Chin, D., Boyle, G. M., Williams, R. M., Ferguson, K., Pandeya, N., Pedley, J., Campbell, C. M., Theile, D. R., Parsons, P. G. & Coman, W. B. 2005. Novel markers for poor prognosis in head and neck cancer. *Int J Cancer,* 113**,** 789-97.

Choi, P. & Chen, C. 2005. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. *Cancer,* 104**,** 1113-28.

Choi, P., Jordan, C. D., Mendez, E., Houck, J., Yueh, B., Farwell, D. G., Futran, N. & Chen, C. 2008. Examination of oral cancer biomarkers by tissue microarray analysis. *Arch Otolaryngol Head Neck Surg,* 134**,** 539-46.

Chorazy-Massalska, M., Kontny, E., Janicka, I., Marcinkiewicz, J. & Maslinski, W. 2004. Effects of taurine bromamine and taurine chloramine on cytokine production by fibroblast like synoviocytes (FLS) isolated from RA patients. *Annals of the Rheumatic Diseases,* 63**,** 135-136.

Chou, W. H., Yan, F. X., Robbins-Weilert, D. K., Ryder, T. B., Liu, W. W., Perbost, C., Fairchild, M., de Leon, J., Koch, W. H. & Wedlund, P. J. 2003. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. *Clin Chem,* 49**,** 542-51.

Chu, I. M., Hengst, L. & Slingerland, J. M. 2008. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer,* 8**,** 253-67.

Chung, L. W., Baseman, A., Assikis, V. & Zhau, H. E. 2005. Molecular insights into prostate cancer progression: the missing link of tumor microenvironment. *J Urol,* 173**,** 10-20.

Chylack, L. T., Jr., Fu, L., Mancini, R., Martin-Rehrmann, M. D., Saunders, A. J., Konopka, G., Tian, D., Hedley-Whyte, E. T., Folkerth, R. D. & Goldstein, L. E. 2004. Lens epithelium-derived growth factor (LEDGF/p75) expression in fetal and adult human brain. *Exp Eye Res,* 79**,** 941-8.

Clarke, R., Ressom, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A. & Wang, Y. 2008. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer,* 8**,** 37-49.

Clish, C. B., Davidov, E., Oresic, M., Plasterer, T. N., Lavine, G., Londo, T., Meys, M., Snell, P., Stochaj, W., Adourian, A., Zhang, X., Morel, N., Neumann, E.,

Verheij, E., Vogels, J. T., Havekes, L. M., Afeyan, N., Regnier, F., van der Greef, J. & Naylor, S. 2004. Integrative biological analysis of the APOE*3-leiden transgenic mouse. *OMICS,* 8**,** 3-13.

Coen, M., Lenz, E. M., Nicholson, J. K., Wilson, I. D., Pognan, F. & Lindon, J. C. 2003. An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. *Chemical Research in Toxicology,* 16**,** 295-303.

Crawford, H. C., Matrisian, L. M. & Liaw, L. 1998. Distinct roles of osteopontin in host defense activity and tumor survival during squamous cell carcinoma progression in vivo. *Cancer Res,* 58**,** 5206-15.

Cruz, I., Van den Brule, A. J. C., Brink, A., Snijders, P. J. F., Walboomers, J. M. M., Van der Waal, I. & Meijer, C. 2000. No direct role for Epstein-Barr virus in oral carcinogenesis: A study at the DNA, RNA and protein levels. *International Journal of Cancer,* 86**,** 356-361.

Curwen, V., Eyras, E., Andrews, T. D., Clarke, L., Mongin, E., Searle, S. M. J. & Clamp, M. 2004. The Ensembl automatic gene annotation system. *Genome Research,* 14**,** 942-950.

D'Arcangelo, G., Tancredi, V., Onofri, F., D'Antuono, M., Giovedi, S. & Benfenati, F. 2000. Interleukin-6 inhibits neurotransmitter release and the spread of excitation in the rat cerebral cortex. *Eur J Neurosci,* 12**,** 1241-52.

Dai, M. H., Wang, P. L., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., Bunney, W. E., Myers, R. M., Speed, T. P., Akil, H., Watson, S. J. & Meng, F. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research,* 33.

Dannenberg, A. J., Altorki, N. K., Boyle, J. O., Dang, C., Howe, L. R., Weksler, B. B. & Subbaramaiah, K. 2001. Cyclo-oxygenase 2: a pharmacological target for the prevention of cancer. *Lancet Oncol,* 2**,** 544-51.

De Vos, R. C., Moco, S., Lommen, A., Keurentjes, J. J., Bino, R. J. & Hall, R. D. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc,* 2**,** 778-91.

Debily, M. A., Marhomy, S. E., Boulanger, V., Eveno, E., Mariage-Samson, R., Camarca, A., Auffray, C., Piatier-Tonneau, D. & Imbeaud, S. 2009. A functional and regulatory network associated with PIP expression in human breast cancer. *PLoS One,* 4**,** e4696.

Deng, G., Lu, Y., Zlotnikov, G., Thor, A. D. & Smith, H. S. 1996. Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science,* 274**,** 2057-9.

Denkert, C., Budczies, J., Kind, T., Weichert, W., Tablack, P., Sehouli, J., Niesporek, S., Konsgen, D., Dietel, M. & Fiehn, O. 2006. Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res,* 66**,** 10795-804.

Denkert, C., Schmitt, W. D., Berger, S., Reles, A., Pest, S., Siegert, A., Lichtenegger, W., Dietel, M. & Hauptmann, S. 2002. Expression of mitogen-activated protein kinase phosphatase-1 (MKP-1) in primary human ovarian carcinoma. *Int J Cancer,* 102**,** 507-13.

Deraz, E. M., Kudo, Y., Yoshida, M., Obayashi, M., Tsunematsu, T., Tani, H., Siriwardena, S. B., Keikhaee, M. R., Qi, G., Iizuka, S., Ogawa, I., Campisi,

G., Lo Muzio, L., Abiko, Y., Kikuchi, A. & Takata, T. 2011. MMP-10/stromelysin-2 promotes invasion of head and neck cancer. *PLoS One,* 6**,** e25438.

Deroo, B. J. & Korach, K. S. 2006. Estrogen receptors and human disease. *J Clin Invest,* 116**,** 561-70.

Desiere, F. 2004. Towards a systems biology understanding of human health: interplay between genotype, environment and nutrition. *Biotechnol Annu Rev,* 10**,** 51-84.

Dettmer, K., Aronov, P. A. & Hammock, B. D. 2007. Mass spectrometry-based metabolomics. *Mass Spectrom Rev,* 26**,** 51-78.

Devoll, R. E., Li, W., Woods, K. V., Pinero, G. J., Butler, W. T., Farach-Carson, M. C. & Happonen, R. P. 1999. Osteopontin (OPN) distribution in premalignant and malignant lesions of oral epithelium and expression in cell lines derived from squamous cell carcinoma of the oral cavity. *J Oral Pathol Med,* 28**,** 97-101.

Di Carlo, A., Mariano, A., Coppa, A., Macchia, V., Maiuri, F., Iaconetta, G. & D'Andrea, F. 1990. Epidermal growth factor receptor and tissue tumor markers in human intracranial neoplasms. *J Neurosurg Sci,* 34**,** 181-5.

Di Stasi, D., Vallacchi, V., Campi, V., Ranzani, T., Daniotti, M., Chiodini, E., Fiorentini, S., Greeve, I., Prinetti, A., Rivoltini, L., Pierotti, M. A. & Rodolfo, M. 2005. DHCR24 gene expression is upregulated in melanoma metastases and associated to resistance to oxidative stress-induced apoptosis. *International Journal of Cancer,* 115**,** 224-230.

Diamond, M. E., Sun, L., Ottaviano, A. J., Joseph, M. J. & Munshi, H. G. 2008. Differential growth factor regulation of N-cadherin expression and motility in normal and malignant oral epithelium. *J Cell Sci,* 121**,** 2197-207.

Dissanayake, S. K., Wade, M., Johnson, C. E., O'Connell, M. P., Leotlela, P. D., French, A. D., Shah, K. V., Hewitt, K. J., Rosenthal, D. T., Indig, F. E., Jiang, Y., Nickoloff, B. J., Taub, D. D., Trent, J. M., Moon, R. T., Bittner, M. & Weeraratna, A. T. 2007. The Wnt5A/protein kinase C pathway mediates motility in melanoma cells via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal transition. *J Biol Chem,* 282**,** 17259-71.

Du, Y., Yang, C., Zhang, X. & Zhang, L. 2008. [Expression and significance of integrin beta1 and integrin-linked kinase in laryngeal carcinoma]. *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi,* 22**,** 500-3.

Duarte, I. F., Goodfellow, B. J., Barros, A., Jones, J. G., Barosa, C., Diogo, L., Garcia, P. & Gil, A. M. 2007. Metabolic characterisation of plasma in juveniles with glycogen storage disease type 1a (GSD1a) by high-resolution (1)H NMR spectroscopy. *NMR Biomed,* 20**,** 401-12.

Dunn, W. B., Bailey, N. J. & Johnson, H. E. 2005. Measuring the metabolome: current analytical technologies. *Analyst,* 130**,** 606-25.

Dunne, F. P., Rollason, T., Ratcliffe, W. A., Marshall, T. & Heath, D. A. 1994. Parathyroid Hormone-Related Protein Gene-Expression in Invasive Cervical Turners. *Cancer,* 74**,** 83-89.

Dysvik, B., Vasstrand, E. N., Lovlie, R., Elgindi, O. A., Kross, K. W., Aarstad, H. J., Johannessen, A. C., Jonassen, I. & Ibrahim, S. O. 2006. Gene expression

profiles of head and neck carcinomas from Sudanese and Norwegian patients reveal common biological pathways regardless of race and lifestyle. *Clin Cancer Res,* 12**,** 1109-20.

Eerligh, P., Koeleman, B. P. C., Dudbridge, F., Bruining, G. J., Roep, B. O. & Giphart, M. J. 2004. Functional genetic polymorphisms in cytokines and metabolic genes as additional genetic markers for susceptibility to develop type 1 diabetes. *Genes and Immunity,* 5**,** 36-40.

Eimoto, H., Tsutsumi, M., Nakajima, A., Yamamoto, K., Takashima, Y., Maruyama, H. & Konishi, Y. 1988. Expression of the glutathione S-transferase placental form in human lung carcinomas. *Carcinogenesis,* 9**,** 2325-7.

Ein-Dor, L., Kela, I., Getz, G., Givol, D. & Domany, E. 2005. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics,* 21**,** 171-178.

Ein-Dor, L., Zuk, O. & Domany, E. 2006. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A,* 103**,** 5923-8.

El-Naggar, A. K., Kim, H. W., Clayman, G. L., Coombes, M. M., Le, B., Lai, S., Zhan, F., Luna, M. A., Hong, W. K. & Lee, J. J. 2002. Differential expression profiling of head and neck squamous carcinoma: significance in their phenotypic and biological classification. *Oncogene,* 21**,** 8206-19.

El-Sayed, S., Bezabeh, T., Odlum, O., Patel, R., Ahing, S., MacDonald, K., Somorjai, R. L. & Smith, I. C. 2002. An ex vivo study exploring the diagnostic potential of 1H magnetic resonance spectroscopy in squamous cell carcinoma of the head and neck region. *Head Neck,* 24**,** 766-72.

ElDeredy, W., Ashmore, S. M., Branston, N. M., Darling, J. L., Williams, S. R. & Thomas, D. G. T. 1997. Pretreatment prediction of the chemotherapeutic response of human glioma cell cultures using nuclear magnetic resonance spectroscopy and artificial neural networks. *Cancer Research,* 57**,** 4196-4199.

Enomoto, M., Hayakawa, S., Itsukushima, S., Ren, D. Y., Matsuo, M., Tamada, K., Oneyama, C., Okada, M., Takumi, T., Nishita, M. & Minami, Y. 2009. Autonomous regulation of osteosarcoma cell invasiveness by Wnt5a/Ror2 signaling. *Oncogene,* 28**,** 3197-208.

Eriksson, L., Andersson, P. L., Johansson, E. & Tysklind, M. 2006. Megavariate analysis of environmental QSAR data. Part II--investigating very complex problem formulations using hierarchical, non-linear and batch-wise extensions of PCA and PLS. *Mol Divers,* 10**,** 187-205.

Estilo, C. L., P, O. c., Talbot, S., Socci, N. D., Carlson, D. L., Ghossein, R., Williams, T., Yonekawa, Y., Ramanathan, Y., Boyle, J. O., Kraus, D. H., Patel, S., Shaha, A. R., Wong, R. J., Huryn, J. M., Shah, J. P. & Singh, B. 2009. Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *BMC Cancer,* 9**,** 11.

Etique, N., Grillier-Vuissoz, I. & Flament, S. 2006. Ethanol stimulates the secretion of matrix metalloproteinases 2 and 9 in MCF-7 human breast cancer cells. *Oncol Rep,* 15**,** 603-8.

Falzon, M. & Du, P. 2000. Enhanced growth of MCF-7 breast cancer cells overexpressing parathyroid hormone-related peptide. *Endocrinology,* 141**,** 1882-92.

Famulski, W., Sulkowska, M., Miller-Famulska, D., Kisielewski, W. & Sulkowski, S. 2001. Correlation between chromogranin A, neuron-specific enolase and synaptophysin expression, and some clinico-pathological features of colorectal cancer. *Folia Histochemica Et Cytobiologica,* 39**,** 155-156.

Fan, T. W., Higashi, R. M. & Lane, A. N. 2006. Integrating metabolomics and transcriptomics for probing SE anticancer mechanisms. *Drug Metab Rev,* 38**,** 707-32.

Fedele, S. 2009. Diagnostic aids in the screening of oral cancer. *Head Neck Oncol,* 1**,** 5.

Fiehn, O. 2002. Metabolomics - the link between genotypes and phenotypes. *Plant Molecular Biology,* 48**,** 155-171.

Fingleton, B. 2006. Matrix metalloproteinases: roles in cancer and metastasis. *Front Biosci,* 11**,** 479-91.

Foulkes, W. D., Brunet, J. S., Sieh, W., Black, M. J., Shenouda, G. & Narod, S. A. 1996. Familial risks of squamous cell carcinoma of the head and neck: Retrospective case-control study. *British Medical Journal,* 313**,** 716-721.

Foxall, P. J., Price, R. G., Jones, J. K., Neild, G. H., Thompson, F. D. & Nicholson, J. K. 1992. High resolution proton magnetic resonance spectroscopy of cyst fluids from patients with polycystic kidney disease. *Biochim Biophys Acta,* 1138**,** 305-14.

Franchi, A., Santucci, M., Masini, E., Sardi, I., Paglierani, M. & Gallo, O. 2002. Expression of matrix metalloproteinase 1, matrix metalloproteinase 2, and matrix metalloproteinase 9 in carcinoma of the head and neck. *Cancer,* 95**,** 1902-10.

Franz, M., Wolheim, A., Richter, P., Umbreit, C., Dahse, R., Driemel, O., Hyckel, P., Virtanen, I., Kosmehl, H. & Berndt, A. 2010. Stromal laminin chain distribution in normal, hyperplastic and malignant oral mucosa: relation to myofibroblast occurrence and vessel formation. *J Oral Pathol Med,* 39**,** 290-8.

Friedman, S. L. 2004. Mechanisms of disease: Mechanisms of hepatic fibrosis and therapeutic implications. *Nat Clin Pract Gastroenterol Hepatol,* 1**,** 98-105.

Gallo, O., Franchi, A., Magnelli, L., Sardi, I., Vannacci, A., Boddi, V., Chiarugi, V. & Masini, E. 2001. Cyclooxygenase-2 pathway correlates with VEGF expression in head and neck cancer. Implications for tumor angiogenesis and metastasis. *Neoplasia,* 3**,** 53-61.

Gallo, O., Masini, E., Bianchi, B., Bruschini, L., Paglierani, M. & Franchi, A. 2002. Prognostic significance of cyclooxygenase-2 pathway and angiogenesis in head and neck squamous cell carcinoma. *Hum Pathol,* 33**,** 708-14.

Galluzzi, L., Larochette, N., Zamzami, N. & Kroemer, G. 2006. Mitochondria as therapeutic targets for cancer chemotherapy. *Oncogene,* 25**,** 4812-30.

Gartland, K. P. R., Bonner, F. W. & Nicholson, J. K. 1989. Investigations into the Biochemical Effects of Region-Specific Nephrotoxins. *Molecular Pharmacology,* 35**,** 242-250.

Gastman, B. R., Atarashi, Y., Reichert, T. E., Saito, T., Balkir, L., Rabinowich, H. & Whiteside, T. L. 1999. Fas ligand is expressed on human squamous cell carcinomas of the head and neck, and it promotes apoptosis of T lymphocytes. *Cancer Research,* 59**,** 5356-5364.

Gauci, V. J., Wright, E. P. & Coorssen, J. R. 2011. Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. *Journal of chemical biology,* 4**,** 3-29.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. & Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol,* 5**,** R80.

Gherzi, R., Lee, K. Y., Briata, P., Wegmuller, D., Moroni, C., Karin, M. & Chen, C. Y. 2004. A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. *Mol Cell,* 14**,** 571-83.

Ginos, M. A., Page, G. P., Michalowicz, B. S., Patel, K. J., Volker, S. E., Pambuccian, S. E., Ondrey, F. G., Adams, G. L. & Gaffney, P. M. 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.

Glunde, K. & Serkova, N. J. 2006. Therapeutic targets and biomarkers identified in cancer choline phospholipid metabolism. *Pharmacogenomics,* 7**,** 1109-23.

Gmeiner, W. H., Hellmann, G. M. & Shen, P. 2008. Tissue-dependent and -independent gene expression changes in metastatic colon cancer. *Oncol Rep,* 19**,** 245-51.

Go, V. L., Nguyen, C. T., Harris, D. M. & Lee, W. N. 2005. Nutrient-gene interaction: metabolic genotype-phenotype relationship. *J Nutr,* 135**,** 3016S-3020S.

Goel, A. & Aggarwal, B. B. 2010. Curcumin, the Golden Spice From Indian Saffron, Is a Chemosensitizer and Radiosensitizer for Tumors and Chemoprotector and Radioprotector for Normal Organs. *Nutrition and Cancer-an International Journal,* 62**,** 919-930.

Goldberg, H. J., Scholey, J. & Fantus, I. G. 2000. Glucosamine activates the plasminogen activator inhibitor 1 gene promoter through Sp1 DNA binding sites in glomerular mesangial cells. *Diabetes,* 49**,** 863-871.

Gomase, V. S. & Tagore, S. 2008. Toxicogenomics. *Curr Drug Metab,* 9**,** 250-4.

Gong, G., DeVries, S., Chew, K. L., Cha, I., Ljung, B. M. & Waldman, F. M. 2001. Genetic changes in paired atypical and usual ductal hyperplasia of the breast by comparative genomic hybridization. *Clin Cancer Res,* 7**,** 2410-4.

Gontero, P., Banisadr, S., Frea, B. & Brausi, M. 2004. Metastasis markers in bladder cancer: a review of the literature and clinical considerations. *Eur Urol,* 46**,** 296-311.

Gonzalez-Quintela, A., Dominguez-Santalla, M. J., Perez, L. F., Vidal, C., Lojo, S. & Barrio, E. 2000. Influence of acute alcohol intake and alcohol withdrawal on circulating levels of IL-6, IL-8, IL-10 and IL-12. *Cytokine,* 12**,** 1437-40.

Gonzalez, H. E., Gujrati, M., Frederick, M., Henderson, Y., Arumugam, J., Spring, P. W., Mitsudo, K., Kim, H. W. & Clayman, G. L. 2003. Identification of 9 genes differentially expressed in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg,* 129**,** 754-9.

Goral, J., Choudhry, M. A. & Kovacs, E. J. 2004. Acute ethanol exposure inhibits macrophage IL-6 production: role of p38 and ERK1/2 MAPK. *J Leukoc Biol,* 75**,** 553-9.

Grandis, J. R., Melhem, M. F., Gooding, W. E., Day, R., Holst, V. A., Wagener, M. M., Drenning, S. D. & Tweardy, D. J. 1998. Levels of TGF-alpha and EGFR protein in head and neck squamous cell carcinoma and patient survival. *Journal of the National Cancer Institute,* 90**,** 824-832.

Griffin, J. L. 2003. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol,* 7**,** 648-54.

Griffin, J. L. 2006. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philosophical Transactions of the Royal Society B-Biological Sciences,* 361**,** 147-161.

Griffin, J. L., Bonney, S. A., Mann, C., Hebbachi, A. M., Gibbons, G. F., Nicholson, J. K., Shoulders, C. C. & Scott, J. 2004. An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiol Genomics,* 17**,** 140-9.

Griffin, J. L. & Shockcor, J. P. 2004. Metabolic profiles of cancer cells. *Nat Rev Cancer,* 4**,** 551-61.

Gromadzinska, J., Reszka, E., Bruzelius, K., Wasowicz, W. & Akesson, B. 2008. Selenium and cancer: biomarkers of selenium status and molecular action of selenium supplements. *Eur J Nutr,* 47 Suppl 2**,** 29-50.

Grosch, S., Tegeder, I., Niederberger, E., Brautigam, L. & Geisslinger, G. 2001. COX-2 independent induction of cell cycle arrest and apoptosis in colon cancer cells by the selective COX-2 inhibitor celecoxib. *FASEB J,* 15**,** 2742-4.

Gu, X., Coates, P. J., Boldrup, L. & Nylander, K. 2008. p63 contributes to cell invasion and migration in squamous cell carcinoma of the head and neck. *Cancer Lett,* 263**,** 26-34.

Gunther, U. L., Ludwig, C. & Ruterjans, H. 2002. WAVEWAT-improved solvent suppression in NMR spectra employing wavelet transforms. *J Magn Reson,* 156**,** 19-25.

Guo, M., Rabin, B. S., Johnson, J. T. & Paradis, I. L. 1987. Lymphocyte Phenotypes at Tumor Margins in Patients with Head and Neck-Cancer. *Head & Neck Surgery,* 9**,** 265-271.

Ha, P. K., Benoit, N. E., Yochem, R., Sciubba, J., Zahurak, M., Sidransky, D., Pevsner, J., Westra, W. H. & Califano, J. 2003. A transcriptional progression model for head and neck cancer. *Clin Cancer Res,* 9**,** 3058-64.

Ha, P. K. & Califano, J. A. 2004. The role of human papillomavirus in oral carcinogenesis. *Critical Reviews in Oral Biology & Medicine,* 15**,** 188-196.

Hamakawa, H., Fukuzumi, M., Bao, Y., Sumida, T., Kayahara, H., Onishi, A. & Sogawa, K. 2000. Keratin mRNA for detecting micrometastasis in cervical lymph nodes of oral cancer. *Cancer Lett,* 160**,** 115-23.

Hamberg, M., Svensson, J., Wakabaya.T & Samuelss.B 1974. Isolation and Structure of 2 Prostaglandin Endoperoxides That Cause Platelet-Aggregation. *Proceedings of the National Academy of Sciences of the United States of America,* 71**,** 345-349.

Hammarstrom, S. 1983. Leukotrienes. *Annual Review of Biochemistry,* 52**,** 355-377.
Hanahan, D. & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell,* 100**,** 57-70.
Harris, H. A., Albert, L. M., Leathurby, Y., Malamas, M. S., Mewshaw, R. E., Miller, C. P., Kharode, Y. P., Marzolf, J., Komm, B. S., Winneker, R. C., Frail, D. E., Henderson, R. A., Zhu, Y. & Keith, J. C., Jr. 2003. Evaluation of an estrogen receptor-beta agonist in animal models of human disease. *Endocrinology,* 144**,** 4241-9.
Harrison, L. B., Sessions, R. B. & Hong, W. K. 1999. Head and neck cancer: A multidisciplinary approach. *Head and neck cancer: A multidisciplinary approach***,** xx+1096p.
Haslam, S. Z. & Woodward, T. L. 2003. Host microenvironment in breast cancer development: epithelial-cell-stromal-cell interactions and steroid hormone action in normal and cancerous mammary gland. *Breast Cancer Res,* 5**,** 208-15.
Hayes, J. D. & McMahon, M. 2006. The double-edged sword of Nrf2: subversion of redox homeostasis during the evolution of cancer. *Mol Cell,* 21**,** 732-4.
He, D., Xu, Q., Yan, M., Zhang, P., Zhou, X., Zhang, Z., Duan, W., Zhong, L., Ye, D. & Chen, W. 2009. The NF-kappa B inhibitor, celastrol, could enhance the anti-cancer effect of gambogic acid on oral squamous cell carcinoma. *BMC Cancer,* 9**,** 343.
Heerdt, B. G., Houston, M. A. & Augenlicht, L. H. 1997. Short-chain fatty acid-initiated cell cycle arrest and apoptosis of colonic epithelial cells is linked to mitochondrial function. *Cell Growth & Differentiation,* 8**,** 523-532.
Hennet, T., Bertoni, G., Richter, C. & Peterhans, E. 1993. Expression of BCL-2 protein enhances the survival of mouse fibrosarcoid cells in tumor necrosis factor-mediated cytotoxicity. *Cancer Res,* 53**,** 1456-60.
Hensen, E. F., De Herdt, M. J., Goeman, J. J., Oosting, J., Smit, V. T., Cornelisse, C. J. & Baatenburg de Jong, R. J. 2008. Gene-expression of metastasized versus non-metastasized primary head and neck squamous cell carcinomas: a pathway-based analysis. *BMC Cancer,* 8**,** 168.
Hess, J., Angel, P. & Schorpp-Kistner, M. 2004. AP-1 subunits: quarrel and harmony among siblings. *Journal of Cell Science,* 117**,** 5965-5973.
Hill, B. T. & Price, L. A. 1994. Lack of Survival Advantage in Patients with Advanced Squamous-Cell Carcinomas of the Oral Cavity Receiving Neoadjuvant Chemotherapy Prior to Local Therapy, Despite Achieving an Initial High Clinical Complete Remission Rate. *American Journal of Clinical Oncology-Cancer Clinical Trials,* 17**,** 1-5.
Hintermann, E. & Quaranta, V. 2004. Epithelial cell motility on laminin-5: regulation by matrix assembly, proteolysis, integrins and erbB receptors. *Matrix Biol,* 23**,** 75-85.
Hirata, S., Odajima, T., Kohama, G. I., Ishigaki, S. & Niitsu, Y. 1992. Significance of Glutathione-S-Transferase-Pi as a Tumor-Marker in Patients with Oral-Cancer. *Cancer,* 70**,** 2381-2387.
Hiratsuka, H., Imamura, M., Kasai, K., Kamiya, H., Ishii, Y., Kohama, G. & Kikuchi, K. 1984. Lymphocyte subpopulations and T-cell subsets in human oral cancer tissues: immunohistologic analysis by monoclonal antibodies. *Am J Clin Pathol,* 81**,** 464-70.

Hiroi, M. & Ohmori, Y. 2003. Constitutive nuclear factor kappa B activity is required to e,licit interferon-gamma-induced expression of chemokine CXC ligand 9 (CXCL9) and CXCL10 in human tumour cell lines. *Biochemical Journal,* 376**,** 393-402.

Hirota, J., Ueta, E., Osaki, T. & Ogawa, Y. 1990. Immunohistologic study of mononuclear cell infiltrates in oral squamous cell carcinomas. *Head Neck,* 12**,** 118-25.

Ho, C. M., Lam, K. H., Wei, W. I., Lau, S. K. & Lam, L. K. 1992. Occult lymph node metastasis in small oral tongue cancers. *Head Neck,* 14**,** 359-63.

Ho, Y. C., Lin, H. J., Tsai, C. H. & Chang, Y. C. 2010. Regulation of type I plasminogen activator inhibitor in human gingival fibroblasts with cyclosporine A. *Oral Dis,* 16**,** 396-401.

Hocevar, B. A., Brown, T. L. & Howe, P. H. 1999. TGF-beta induces fibronectin synthesis through a c-Jun N-terminal kinase-dependent, Smad4-independent pathway. *Embo Journal,* 18**,** 1345-1356.

Hockenbery, D., Nunez, G., Milliman, C., Schreiber, R. D. & Korsmeyer, S. J. 1990. Bcl-2 Is an Inner Mitochondrial-Membrane Protein That Blocks Programmed Cell-Death. *Nature,* 348**,** 334-336.

Holmes, E. & Antti, H. 2002. Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analyst,* 127**,** 1549-57.

Hong, F., Kim, W. H., Tian, Z., Jaruga, B., Ishac, E., Shen, X. & Gao, B. 2002. Elevated interleukin-6 during ethanol consumption acts as a potential endogenous protective cytokine against ethanol-induced apoptosis in the liver: involvement of induction of Bcl-2 and Bcl-x(L) proteins. *Oncogene,* 21**,** 32-43.

Hou, C. W. 2011. Pu-Erh tea and GABA attenuates oxidative stress in kainic acid-induced status epilepticus. *J Biomed Sci,* 18**,** 75.

Howe, F. A., Barton, S. J., Cudlip, S. A., Stubbs, M., Saunders, D. E., Murphy, M., Wilkins, P., Opstad, K. S., Doyle, V. L., McLean, M. A., Bell, B. A. & Griffiths, J. R. 2003. Metabolic profiles of human brain tumors using quantitative in vivo H-1 magnetic resonance spectroscopy. *Magnetic Resonance in Medicine,* 49**,** 223-232.

Huang, D. W., Sherman, B. T. & Lempicki, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research,* 37**,** 1-13.

Huang, D. W., Sherman, B. T. & Lempicki, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4**,** 44-57.

Huang, G. C., Song, H. Z., Wang, R., Han, X. D. & Chen, L. B. 2012. The relationship between RGS5 expression and cancer differentiation and metastasis in non-small cell lung cancer. *Journal of Surgical Oncology,* 105**,** 420-424.

Huang, J. & Siragy, H. M. 2009. Glucose promotes the production of interleukine-1beta and cyclooxygenase-2 in mesangial cells via enhanced (Pro)renin receptor expression. *Endocrinology,* 150**,** 5557-65.

Huang, X., Gollin, S. M., Raja, S. & Godfrey, T. E. 2002. High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 11369-11374.

Hunter, K. D., Parkinson, E. K. & Harrison, P. R. 2005. Opinion - Profiling early head and neck cancer. *Nature Reviews Cancer,* 5**,** 127-135.

Hwang, D., Alevizos, I., Schmitt, W. A., Misra, J., Ohyama, H., Todd, R., Mahadevappa, M., Warrington, J. A., Stephanopoulos, G., Wong, D. T. & Stephanopoulos, G. 2003. Genomic dissection for characterization of cancerous oral epithelium tissues using transcription profiling. *Oral Oncology,* 39**,** 259-268.

Idle, J. R. & Gonzalez, F. J. 2007. Metabolomics. *Cell Metabolism,* 6**,** 348-351.

Irie, T., Aida, T. & Tachikawa, T. 2004. Gene expression profiling of oral squamous cell carcinoma using laser microdissection and cDNA microarray. *Medical Electron Microscopy,* 37**,** 89-96.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. & Speed, T. P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-264.

Ito, M., Nakashima, M., Alipov, G. K., Matsuzaki, S., Ohtsuru, A., Yano, H., Yamashita, S. & Sekine, I. 1997. Gastric cancer associated with overexpression of parathyroid hormone-related peptide (PTHrP) and PTH/PTHrP receptor in relation to tumor progression. *J Gastroenterol,* 32**,** 396-400.

Ito, R., Oue, N., Zhu, X. D., Yoshida, K., Nakayama, H., Yokozaki, H. & Yasui, W. 2003. Expression of integrin-linked kinase is closely correlated with invasion and metastasis of gastric carcinoma. *Virchows Archiv,* 442**,** 118-123.

Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., Igarashi, K., Engel, J. D. & Yamamoto, M. 1999. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes & Development,* 13**,** 76-86.

Itoh, S., Matsui, K., Furuta, I. & Takano, Y. 2003. Immunohistochemical study on overexpression of cyclooxygenase-2 in squamous cell carcinoma of the oral cavity: its importance as a prognostic predictor. *Oral Oncology,* 39**,** 829-835.

Jacobs, M. A., Barker, P. B., Bottomley, P. A., Bhujwalla, Z. & Bluemke, D. A. 2004. Proton magnetic resonance spectroscopic imaging of human breast cancer: a preliminary study. *J Magn Reson Imaging,* 19**,** 68-75.

Jacobsen, A. B., Nesland, J. M., Fossa, S. D. & Pettersen, E. O. 1990. Human chorionic gonadotropin, neuron specific enolase and deoxyribonucleic acid flow cytometry in patients with high grade bladder carcinoma. *J Urol,* 143**,** 706-9.

Jacobson, M. D., Burne, J. F. & Raff, M. C. 1994. Programmed cell death and Bcl-2 protection in the absence of a nucleus. *EMBO J,* 13**,** 1899-910.

Jalanko, A., Tyynela, J. & Peltonen, L. 2006. From genes to systems: new global strategies for the characterization of NCL biology. *Biochim Biophys Acta,* 1762**,** 934-44.

Jaramillo, M. L., Leon, Z., Grothe, S., Paul-Roc, B., Abulrob, A. & McCourt, M. O. C. 2006. Effect of the anti-receptor ligand-blocking 225 monoclonal antibody on EGF receptor endocytosis and sorting. *Experimental Cell Research,* 312**,** 2778-2790.

Jauliac, S., Lopez-Rodriguez, C., Shaw, L. M., Brown, L. F., Rao, A. & Toker, A. 2002. The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nat Cell Biol,* 4**,** 540-4.

Jayasekera, R., Freitas, M. C. & Araujo, M. F. 2004. Bulk and trace element analysis of spices: the applicability of k0-standardization and energy dispersive X-ray fluorescence. *J Trace Elem Med Biol,* 17**,** 221-8.

Jeffery, I. B., Higgins, D. G. & Culhane, A. C. 2006. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics,* 7**,** 359.

Jeng, J. H., Ho, Y. S., Chan, C. P., Wang, Y. J., Hahn, L. J., Lei, D., Hsu, C. C. & Chang, M. C. 2000. Areca nut extract up-regulates prostaglandin production, cyclooxygenase-2 mRNA and protein expression of human oral keratinocytes. *Carcinogenesis,* 21**,** 1365-70.

Jeon, G. A., Lee, J. S., Patel, V., Gutkind, J. S., Thorgeirsson, S. S., Kim, E. C., Chu, I. S., Amornphimoltham, P. & Park, M. H. 2004. Global gene expression profiles of human head and neck squamous carcinoma cell lines. *Int J Cancer,* 112**,** 249-58.

Jeong, I. K., Oh da, H., Park, S. J., Kang, J. H., Kim, S., Lee, M. S., Kim, M. J., Hwang, Y. C., Ahn, K. J., Chung, H. Y., Chae, M. K. & Yoo, H. J. 2011. Inhibition of NF-kappaB prevents high glucose-induced proliferation and plasminogen activator inhibitor-1 expression in vascular smooth muscle cells. *Exp Mol Med,* 43**,** 684-92.

John, A. P. 2001. Dysfunctional mitochondria, not oxygen insufficiency, cause cancer cells to produce inordinate amounts of lactic acid: the impact of this on the treatment of cancer. *Med Hypotheses,* 57**,** 429-31.

Johnson, W. E., Li, C. & Rabinovic, A. 2007a. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8**,** 118-127.

Johnson, W. E., Li, C. & Rabinovic, A. 2007b. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8**,** 118-27.

Jonason, A. S., Kunala, S., Price, G. J., Restifo, R. J., Spinelli, H. M., Persing, J. A., Leffell, D. J., Tarone, R. E. & Brash, D. E. 1996. Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc Natl Acad Sci U S A,* 93**,** 14025-9.

Jones, F. S. & Jones, P. L. 2000. The tenascin family of ECM glycoproteins: structure, function, and regulation during embryonic development and tissue remodeling. *Dev Dyn,* 218**,** 235-59.

Joo, K., Lee, Y., Choi, D., Han, J., Hong, S., Kim, Y. M. & Jung, Y. 2009. An anti-inflammatory mechanism of taurine conjugated 5-aminosalicylic acid against experimental colitis: taurine chloramine potentiates inhibitory effect of 5-aminosalicylic acid on IL-1beta-mediated NFkappaB activation. *Eur J Pharmacol,* 618**,** 91-7.

Jordan, R. C. K., Macabeo-Ong, M., Shiboski, C. H., Dekker, N., Ginzinger, D. G., Wong, D. T. W. & Schmidt, B. L. 2004. Overexpression of matrix

metalloproteinase-1 and-9 mRNA is associated with progression of oral dysplasia to cancer. *Clinical Cancer Research,* 10**,** 6460-6465.

Jose, C., Bellance, N. & Rossignol, R. 2011. Choosing between glycolysis and oxidative phosphorylation: A tumor's dilemma? *Biochimica Et Biophysica Acta-Bioenergetics,* 1807**,** 552-561.

Junn, E., Lee, K. N., Ju, H. R., Han, S. H., Im, J. Y., Kang, H. S., Lee, T. H., Bae, Y. S., Ha, K. S., Lee, Z. W., Rhee, S. G. & Choi, I. 2000. Requirement of hydrogen peroxide generation in TGF-beta 1 signal transduction in human lung fibroblast cells: involvement of hydrogen peroxide and Ca2+ in TGF-beta 1-induced IL-6 expression. *J Immunol,* 165**,** 2190-7.

Jurs, P. C. 1986. Pattern-Recognition Used to Investigate Multivariate Data in Analytical-Chemistry. *Science,* 232**,** 1219-1224.

Kagesato, Y., Mizushima, H., Koshikawa, N., Kitamura, H., Hayashi, H., Ogawa, N., Tsukuda, M. & Miyazaki, K. 2001. Sole expression of laminin gamma 2 chain in invading tumor cells and its association with stromal fibrosis in lung adenocarcinomas. *Jpn J Cancer Res,* 92**,** 184-92.

Kaiser, S. M., Sebag, M., Rhim, J. S., Kremer, R. & Goltzman, D. 1994. Antisense-mediated inhibition of parathyroid hormone-related peptide production in a keratinocyte cell line impedes differentiation. *Mol Endocrinol,* 8**,** 139-47.

Kaklamani, V. G. & Gradishar, W. J. 2006. Gene expression in breast cancer. *Curr Treat Options Oncol,* 7**,** 123-8.

Kalra, N. & Kumar, V. 2004. c-Fos is a mediator of the c-Myc-induced apoptotic signaling in serum-deprived hepatoma cells via the p38 mitogen-activated protein kinase pathway. *Journal of Biological Chemistry,* 279**,** 25313-25319.

Kalyankrishna, S. & Grandis, J. R. 2006. Epidermal growth factor receptor biology in head and neck cancer. *Journal of Clinical Oncology,* 24**,** 2666-2672.

Kanehisa, M. & Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

Kannan, K., Amariglio, N., Rechavi, G., Jakob-Hirsch, J., Kela, I., Kaminski, N., Getz, G., Domany, E. & Givol, D. 2001a. DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene,* 20**,** 2225-34.

Kannan, K., Kaminski, N., Rechavi, G., Jakob-Hirsch, J., Amariglio, N. & Givol, D. 2001b. DNA microarray analysis of genes involved in p53 mediated apoptosis: activation of Apaf-1. *Oncogene,* 20**,** 3449-3455.

Kao, S. Y., Chu, Y. W., Chen, Y. W., Chang, K. W. & Liu, T. Y. 2009. Detection and Screening of Oral Cancer and Pre-cancerous Lesions. *Journal of the Chinese Medical Association,* 72**,** 227-233.

Katakura, A., Kamiyama, I., Takano, N., Shibahara, T., Muramatsu, T., Ishihara, K., Takagi, R. & Shouno, T. 2007. Comparison of salivary cytokine levels in oral cancer patients and healthy subjects. *Bull Tokyo Dent Coll,* 48**,** 199-203.

Kato, Y., Uzawa, K., Saito, K., Nakashima, D., Kato, M., Nimura, Y., Seki, N. & Tanzawa, H. 2006. Gene expression pattern in oral cancer cervical lymph node metastasis. *Oncology Reports,* 16**,** 1009-1014.

Kell, D. B. & Mendes, P. 2000. Snapshots of systems - Metabolic control analysis and biotechnology in the post-genomic era. *Technological and Medical Implications of Metabolic Control Analysis,* 74**,** 3-25.

Kelloff, G. J., Sigman, C. C. & Contag, C. H. 2009. Early Detection of Oral Neoplasia: Watching with New Eyes. *Cancer Prevention Research,* 2**,** 405-408.

Kelly, A., Lynch, A., Vereker, E., Nolan, Y., Queeman, P., Whittaker, E., O'Neill, L. A. J. & Lynch, M. A. 2001. The anti-inflammatory cytokine, interleukin (IL)-10, blocks the inhibitory effect of IL-1 beta on long term potentiation - A role for JNK. *Journal of Biological Chemistry,* 276**,** 45564-45572.

Kelly, T. J., Souza, A. L., Clish, C. B. & Puigserver, P. 2011. A Hypoxia-Induced Positive Feedback Loop Promotes Hypoxia-Inducible Factor 1 alpha Stability through miR-210 Suppression of Glycerol-3-Phosphate Dehydrogenase 1-Like. *Molecular and Cellular Biology,* 31**,** 2696-2706.

Kensler, T. W., Wakabayash, N. & Biswal, S. 2007. Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. *Annual Review of Pharmacology and Toxicology.*

Kho, P. S., Wang, Z., Zhuang, L., Li, Y., Chew, J.-L., Ng, H.-H., Liu, E. T. & Yu, Q. 2004. p53-regulated Transcriptional Program Associated with Genotoxic Stress-induced Apoptosis. *Journal of Biological Chemistry,* 279**,** 21183-21192.

Kielar, M. L., John, R., Bennett, M., Richardson, J. A., Shelton, J. M., Chen, L., Jeyarajah, D. R., Zhou, X. J., Zhou, H., Chiquett, B., Nagami, G. T. & Lu, C. Y. 2005. Maladaptive Role of IL-6 in Ischemic Acute Renal Failure. *Journal of the American Society of Nephrology,* 16**,** 3315-3325.

Kim, C. J., Kovacs-Nolan, J. A., Yang, C., Archbold, T., Fan, M. Z. & Mine, Y. 2010. L-Tryptophan exhibits therapeutic function in a porcine model of dextran sodium sulfate (DSS)-induced colitis. *Journal of Nutritional Biochemistry,* 21**,** 468-475.

Kim, S. H., Cho, N. H., Kim, K., Lee, J. S., Koo, B. S., Kim, J. H., Chang, J. H. & Choi, E. C. 2006. Correlations of oral tongue cancer invasion with matrix metalloproteinases (MMPs) and vascular endothelial growth factor (VEGF) expression. *Journal of Surgical Oncology,* 93**,** 330-337.

Kim, T. & Kim, A. K. 2013. Taurine enhances anticancer activity of Cisplatin in human cervical cancer cells. *Advances in experimental medicine and biology,* 776**,** 189-98.

Kimura, Y., Furuhata, T., Urano, T., Hirata, K., Nakamura, Y. & Tokino, T. 1997. Genomic structure and chromosomal localization of GML (GPI-anchored molecule-like protein), a gene induced by p53. *Genomics,* 41**,** 477-480.

Kitajima, S., Kudo, Y., Ogawa, I., Bashir, T., Kitagawa, M., Miyauchi, M., Pagano, M. & Takata, T. 2004. Role of Cks1 overexpression in oral squamous cell carcinomas - Cooperation with Skp2 in promoting p27 degradation. *American Journal of Pathology,* 165**,** 2147-2155.

Kitchen, R. R., Sabine, V. S., Sims, A. H., Macaskill, E. J., Renshaw, L., Thomas, J. S., van Hemert, J. I., Dixon, J. M. & Bartlett, J. M. S. 2010. Correcting for intra-experiment variation in Illumina BeadChip data is necessary to generate robust gene-expression profiles. *Bmc Genomics,* 11.

Klaunig, J. E., Kamendulis, L. M. & Hocevar, B. A. 2010. Oxidative Stress and Oxidative Damage in Carcinogenesis. *Toxicologic Pathology,* 38**,** 96-109.

Kline, E. E., Treat, E. G., Averna, T. A., Davis, M. S., Smith, A. Y. & Sillerud, L. O. 2006. Citrate concentrations in human seminal fluid and expressed prostatic fluid determined via H-1 nuclear magnetic resonance spectroscopy outperform prostate specific antigen in prostate cancer detection. *Journal of Urology,* 176**,** 2274-2279.

Knoepfler, P. S. 2007. Myc goes global: New tricks for an old oncogene. *Cancer Research,* 67**,** 5061-5063.

Knoepfler, P. S., Zhang, X.-y., Cheng, P. F., Gafken, P. R., McMahon, S. B. & Eisenman, R. N. 2006. Myc influences global chromatin structure. *Embo Journal,* 25**,** 2723-2734.

Kobayashi, H., Sagara, J., Kurita, H., Morifuji, M., Ohishi, M., Kurashina, K. & Taniguchi, S. 2004. Clinical significance of cellular distribution of moesin in patients with oral squamous cell carcinoma. *Clinical Cancer Research,* 10**,** 572-580.

Kontny, E., Szczepanska, K., Kowalczewski, J., Kurowska, M., Janicka, I., Marcinkiewicz, J. & Maslinski, W. 2000. The mechanism of taurine chloramine inhibition of cytokine (interleukin-6, interleukin-8) production by rheumatoid arthritis fibroblast-like synoviocytes. *Arthritis and Rheumatism,* 43**,** 2169-2177.

Kopreski, M. S., Benko, F. A. & Gocke, C. D. 2001. Circulating RNA as a tumor marker - Detection of 5T4 mRNA in breast and lung cancer patient serum. *Circulating Nucleic Acids in Plasma or Serum Ii,* 945**,** 172-178.

Kosmehl, H., Berndt, A., Strassburger, S., Borsi, L., Rousselle, P., Mandel, U., Hyckel, P., Zardi, L. & Katenkamp, D. 1999. Distribution of laminin and fibronectin isoforms in oral mucosa and oral squamous cell carcinoma. *British Journal of Cancer,* 81**,** 1071-1079.

Krippner, A., MatsunoYagi, A., Gottlieb, R. A. & Babior, B. M. 1996. Loss of function of cytochrome c in Jurkat cells undergoing Fas-mediated apoptosis. *Journal of Biological Chemistry,* 271**,** 21629-21636.

Kudo, Y., Siriwardena, B. S. M. S., Hatano, H., Ogawa, I. & Takata, T. 2007. Periostin: Novel diagnostic and therapeutic target for cancer. *Histology and Histopathology,* 22**,** 1167-1174.

Kulawiec, M., Ayyasamy, V. & Singh, K. K. 2009. p53 regulates mtDNA copy number and mitocheckpoint pathway. *Journal of carcinogenesis,* 8**,** 8-8.

Kulawiec, M., Safina, A., Desouki, M. M., Still, I., Matsui, S.-I., Bakin, A. & Singh, K. K. 2008. Tumorigenic transformation of human breast epithelial cells induced by mitochondrial DNA depletion. *Cancer Biology & Therapy,* 7**,** 1732-1743.

Kunkel, M., Reichert, T. E., Benz, P., Lehr, H. A., Jeong, J. H., Wieand, S., Bartenstein, P., Wagner, W. & Whiteside, T. L. 2003. Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral spuamous cell carcinoma. *Cancer,* 97**,** 1015-1024.

Kuo, W. P., Jenssen, T. K., Park, P. J., Lingen, M. W., Hasina, R. & Ohno-Machado, L. 2002. Gene expression levels in different stages of progression in oral squamous cell carcinoma. *Amia 2002 Symposium, Proceedings***,** 415-419.

Kuo, W. P., Mendez, E., Chen, C., Whipple, M. E., Farell, G., Agoff, N. & Park, P. J. 2003. Functional relationships between gene pairs in oral squamous cell carcinoma. *AMIA Annu Symp Proc*, 371-5.

Kurayoshi, M., Oue, N., Yamamoto, H., Kishida, M., Inoue, A., Asahara, T., Yasui, W. & Kikuchi, A. 2006. Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. *Cancer Research,* 66**,** 10439-10448.

Kuriakose, M. A., Chen, W. T., He, Z. M., Sikora, A. G., Zhang, P., Zhang, Z. Y., Qiu, W. L., Hsu, D. F., McMunn-Coffran, C., Brown, S. M., Elango, E. M., Delacure, M. D. & Chen, F. A. 2004. Selection and validation of differentially expressed genes in head and neck cancer. *Cellular and Molecular Life Sciences,* 61**,** 1372-1383.

Kuss, I., Saito, T., Johnson, J. T. & Whiteside, T. L. 1999. Clinical significance of decreased zeta chain expression in peripheral blood lymphocytes of patients with head and neck cancer. *Clinical Cancer Research,* 5**,** 329-334.

La Vecchia, C., Lucchini, F., Negri, E. & Levi, F. 2004. Trends in oral cancer mortality in Europe. *Oral Oncology,* 40**,** 433-439.

Laborde, E. 2010. Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death. *Cell Death and Differentiation,* 17**,** 1373-1380.

Lam, S., van der Geest, R. N., Verhagen, N. A. M., van Nieuwenhoven, F. A., Blom, I. E., Aten, J., Goldschmeding, R., Daha, M. R. & van Kooten, C. 2003. Connective tissue growth factor and IGF-I are produced by human renal fibroblasts and cooperate in the induction of collagen production by high glucose. *Diabetes,* 52**,** 2975-2983.

Lambert, R., Sauvaget, C., Cancela, M. d. C. & Sankaranarayanan, R. 2011. Epidemiology of cancer from the oral cavity and oropharynx. *European Journal of Gastroenterology & Hepatology,* 23**,** 633-641.

Landry, J. R. & Mager, D. L. 2002. Widely spaced alternative promoters, conserved between human and rodent, control expression of the Opitz syndrome gene MID1. *Genomics,* 80**,** 499-508.

LaVecchia, C., Tavani, A., Franceschi, S., Levi, F., Corrao, G. & Negri, E. 1997. Epidemiology and prevention of oral cancer. *Oral Oncology,* 33**,** 302-312.

Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., Weiss-Solís, D. Y., Duque, R., Bersini, H. & Nowé, A. 2012. Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings in Bioinformatics*.

Le Cao, K. A., Martin, P. G. P., Robert-Granie, C. & Besse, P. 2009. Sparse canonical methods for biological data integration: application to a cross-platform study. *Bmc Bioinformatics,* 10.

Lee, D. W., Sung, M. W., Park, S. W., Seong, W. J., Roh, J. L., Park, B., Heo, D. S. & Kim, K. H. 2002. Increased cyclooxygenase-2 expression in human squamous cell carcinomas of the head and neck and inhibition of proliferation by nonsteroidal anti-inflammatory drugs. *Anticancer Research,* 22**,** 2089-2096.

Lee, J. K., Williams, P. D. & Cheon, S. 2008. Data mining in genomics. *Clinics in Laboratory Medicine,* 28**,** 145-+.

Lee, K.-D., Lee, H.-S. & Jeon, C.-H. 2011. Body Fluid Biomarkers for Early Detection of Head and Neck Squamous Cell Carcinomas. *Anticancer Research,* 31**,** 1161-1167.

Lee, M.-S., Moon, E.-J., Lee, S.-W., Kim, M. S., Kim, K.-W. & Kim, Y.-J. 2001. Angiogenic Activity of Pyruvic Acid in in Vivo and in Vitro Angiogenesis Models. *Cancer Research,* 61**,** 3290-3293.

Lee, W. N. P. & Go, V. L. W. 2005. Nutrient-gene interaction: Tracer-based metabolomics. *Journal of Nutrition,* 135**,** 3027S-3032S.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K. & Irizarry, R. A. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics,* 11**,** 733-739.

Leemans, C. R., Tiwari, R., Nauta, J. J. P., Vanderwaal, I. & Snow, G. B. 1994. Recurrence at the Primary Site in Head and Neck-Cancer and the Significance of Neck Lymph-Node Metastases as a Prognostic Factor. *Cancer,* 73**,** 187-190.

Leethanakul, C., Knezevic, V., Patel, V., Amornphimoltham, P., Gillespie, J., Shillitoe, E. J., Emko, P., Park, M. H., Emmert-Buck, M. R., Strausberg, R. L., Krizman, D. B. & Gutkind, J. S. 2003. Gene discovery in oral squamous cell carcinoma through the Head and Neck Cancer Genome Anatomy Project: confirmation by microarray analysis. *Oral Oncology,* 39**,** 248-258.

Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J. F., Koontongkaew, S., Liotta, L. A., Emmert-Buck, M. & Gutkind, J. S. 2000a. Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene,* 19**,** 3220-3224.

Leethanakul, C., Patel, V., Gillespie, J., Shillitoe, E., Kellman, R. M., Ensley, J. F., Limwongse, V., Emmert-Buck, M. R., Krizman, D. B. & Gutkind, J. S. 2000b. Gene expression profiles in squamous cell carcinomas of the oral cavity: use of laser capture microdissection for the construction and analysis of stage-specific cDNA libraries. *Oral Oncology,* 36**,** 474-483.

Lerner, L., Henriksen, M. A., Zhang, X. K. & Darnell, J. E. 2003. STAT3-dependent enhanceosome assembly and disassembly: synergy with GR for full transcriptional increase of the alpha 2-macroglobulin gene. *Genes & Development,* 17**,** 2564-2577.

Li, M. X., Zhou, J. Y., Ge, Y. B., Matherly, L. H. & Wu, G. S. 2003. The phosphatase MKP1 is a transcriptional target of p53 involved in cell cycle regulation. *Journal of Biological Chemistry,* 278**,** 41059-41068.

Li, N., Jian, X., Hu, Y., Xu, C., Yao, Z. & Zhong, X. 2008. Discovery of novel biomarkers in oral submucous fibrosis by microarray analysis. *Cancer Epidemiology Biomarkers & Prevention,* 17**,** 2249-2259.

Li, Y., Elashoff, D., Oh, M., Sinha, U., St John, M. A. R., Zhou, X. F., Abemayor, E. & Wong, D. T. 2006. Serum circulating human mRNA profiling and its utility for oral cancer detection. *Journal of Clinical Oncology,* 24**,** 1754-1760.

Li, Y., Samuvel, D. J., Sundararaj, K. P., Lopes-Virella, M. F. & Huang, Y. 2010. IL-6 and High Glucose Synergistically Upregulate MMP-1 Expression by

U937 Mononuclear Phagocytes via ERK1/2 and JNK Pathways and c-Jun. *Journal of Cellular Biochemistry,* 110**,** 248-259.

Li, Y., St John, M. A. R., Zhou, X. F., Kim, Y., Sinha, U., Jordan, R. C. K., Eisele, D., Abemayor, E., Elashoff, D., Park, N. H. & Wong, D. T. 2004a. Salivary transcriptome diagnostics for oral cancer detection. *Clinical Cancer Research,* 10**,** 8442-8450.

Li, Y., Zhou, X., John, M. & Wong, D. T. W. 2004b. RNA profiling of cell-free saliva using microarray technology. *Journal of Dental Research,* 83**,** 199-203.

Lieber, C. S. 1994a. Alcohol and the liver: 1994 update. *Gastroenterology,* 106**,** 1085-105.

Lieber, C. S. 1994b. Metabolic Consequences of Ethanol. *Endocrinologist,* 4**,** 127-139.

Lilenbaum, R. C. 2006. The evolving role of cetuximab in non-small cell lung cancer. *Clinical Cancer Research,* 12**,** 4432S-4435S.

Lindon, J. C., Holmes, E. & Nicholson, J. K. 2003. So whats the deal with metabonomics? Metabonomics measures the fingerprint of biochemical perturbations caused by disease, drugs, and toxins. *Analytical Chemistry,* 75**,** 384A-391A.

Liotta, L. A., Wewer, U., Rao, N. C., Schiffmann, E., Stracke, M., Guirguis, R., Thorgeirsson, U., Muschel, R. & Sobel, M. 1988. Biochemical mechanisms of tumor invasion and metastases. *Progress in clinical and biological research,* 256**,** 3-16.

Liu, D., Rudland, P. S., Sibson, D. R., Platt-Higgins, A. & Barraclough, R. 2005. Human Homologue of Cement Gland Protein, a Novel Metastasis Inducer Associated with Breast Carcinomas. *Cancer Research,* 65**,** 3796-3805.

Liu, X. S., Kim, C. N., Yang, J., Jemmerson, R. & Wang, X. D. 1996. Induction of apoptotic program in cell-free extracts: Requirement for dATP and cytochrome c. *Cell,* 86**,** 147-157.

Liu, Z., Niu, Y., Li, C., Yang, Y. & Gao, C. 2012. Integrating multiple microarray datasets on oral squamous cell carcinoma to reveal dysregulated networks. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 34**,** 1789-1797.

Llacuna, L., Mari, M., Lluis, J. M., Garcia-Ruiz, C., Fernandez-Checa, J. C. & Morales, A. 2009. Reactive Oxygen Species Mediate Liver Injury Through Parenchymal Nuclear Factor-kappa B Inactivation in Prolonged Ischemia/Reperfusion. *American Journal of Pathology,* 174**,** 1776-1785.

Loessner, D., Abou-Ajram, C., Benge, A., Aumercier, M., Schmitt, M. & Reuning, U. 2009. Integrin alpha v beta 3 Upregulates Integrin-Linked Kinase Expression in Human Ovarian Cancer Cells Via Enhancement of ILK Gene Transcription 2. *Journal of Cellular Physiology,* 220**,** 367-375.

Lopes, V. 2007. *An investigation of molecular biological factors underlying the heterogeneity of oral cancer.* PhD, University of Birmingham.

Lu, N. Z., Wardell, S. E., Burnstein, K. L., Defranco, D., Fuller, P. J., Giguere, V., Hochberg, R. B., McKay, L., Renoir, J.-M., Weigel, N. L., Wilson, E. M., McDonnell, D. P. & Cidlowski, J. A. 2006. International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear

receptor superfamily: Glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. *Pharmacological Reviews,* 58**,** 782-797.

Luo, X. P., Ding, L., Xu, J. X., Williams, R. S. & Chegini, N. 2005. Leiomyoma and myometrial gene expression profiles and their responses to gonadotropin-releasing hormone analog therapy. *Endocrinology,* 146**,** 1074-1096.

Luo, Y., Lathia, J., Mughal, M. & Mattson, M. P. 2008. SDF1 alpha/CXCR4 Signaling, via ERKs and the Transcription Factor Egr1, Induces Expression of a 67-kDa Form of Glutamic Acid Decarboxylase in Embryonic Hippocampal Neurons. *Journal of Biological Chemistry,* 283**,** 24789-24800.

Lutz, N. W., Viola, A., Malikova, I., Confort-Gouny, S., Audoin, B., Ranjeva, J.-P., Pelletier, J. & Cozzone, P. J. 2007. Inflammatory multiple-sclerosis plaques generate characteristic metabolic profiles in cerebrospinal fluid. *PLoS One,* 2**,** e595.

Madan, R., Brandwein-Gensler, M., Schlecht, N. F., Elias, K., Gorbovitsky, E., Belbin, T. J., Mahmood, R., Breining, D., Qian, H., Childs, G., Locker, J., Smith, R., Haigentz, M., Gunn-Moore, F. & Prystowsky, M. B. 2006. Differential tissue and subcellular expression of ERM proteins in normal and malignant tissues: Cytoplasmic ezrin expression has prognostic signficance for head and neck squamous cell carcinoma. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 28**,** 1018-1027.

Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. 2007. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research,* 35**,** D26-D31.

Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. 2011. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research,* 39**,** D52-D57.

Maheshwari, S. R., Fatterpekar, G. M., Castillo, M. & Mukherji, S. K. 2000a. Proton MR spectroscopy of the brain. *Seminars in Ultrasound Ct and Mri,* 21**,** 434-451.

Maheshwari, S. R., Mukherji, S. K., Neelon, B., Schiro, S., Fatterpekar, G. M., Stone, J. A. & Castillo, M. 2000b. The choline/creatine ratio in five benign neoplasms: Comparison with squamous cell carcinoma by use of in vitro MR spectroscopy. *American Journal of Neuroradiology,* 21**,** 1930-1935.

Mahner, S., Baasch, C., Schwarz, J., Hein, S., Woelber, L., Jaenicke, F. & Milde-Langosch, K. 2008. C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. *British Journal of Cancer,* 99**,** 1269-1275.

Makinen, V. P., Soininen, P., Forsblom, C., Parkkonen, M., Ingman, P., Kaski, K., Groop, P. H., Ala-Korpela, M. & FinnDiane Study, G. 2006. Diagnosing diabetic nephropathy by H-1 NMR metabonomics of serum. *Magnetic Resonance Materials in Physics Biology and Medicine,* 19**,** 281-296.

Malinoff, H. L., McCoy, J. P., Varani, J. & Wicha, M. S. 1984. Metastatic Potential of Murine Fibro-Sarcoma Cells Is Influenced by Cell-Surface Laminin. *International Journal of Cancer,* 33**,** 651-655.

Mallia, R. J., Thomas, S. S., Mathews, A., Kumar, R., Sebastian, P., Madhavan, J. & Subhash, N. 2008. Laser-induced autofluorescence spectral ratio reference standard for early discrimination of oral cancer. *Cancer,* 112**,** 1503-12.

Manenti, G., Peissel, B., Gariboldi, M., Falvella, F. S., Zaffaroni, D., Allaria, B., Pazzaglia, S., Rebessi, S., Covelli, V., Saran, A. & Dragani, T. A. 2000. A

cancer modifier role for parathyroid hormone-related protein. *Oncogene,* 19**,** 5324-8.

Manyam. 2009. *Global Patterns Of Changes In The Gene Expression Associated With Genesis Of Cancer.*

Mao, L., Hong, W. K. & Papadimitrakopoulou, V. A. 2004. Focus on head and neck cancer. *Cancer Cell,* 5**,** 311-316.

Marchetti, P., Susin, S. A., Decaudin, D., Gamen, S., Castedo, M., Hirsch, T., Zamzami, N., Naval, J., Senik, A. & Kroemer, G. 1996. Apoptosis-associated derangement of mitochondrial function in cells lacking mitochondrial DNA. *Cancer Research,* 56**,** 2033-2038.

Marcus, B., Arenberg, D., Lee, J., Kleer, C., Chepeha, D. B., Schmalbach, C. E., Islam, M., Paul, S., Pan, Q., Hanash, S., Kuick, R., Merajver, S. D. & Teknos, T. N. 2004. Prognostic factors in oral cavity and oropharyngeal squamous cell carcinoma - The impact of tumor-associated macrophages. *Cancer,* 101**,** 2779-2787.

Maresh, E., Mah, V., Alavi, M., Horvath, S., Bagryanova, L., Liebeskind, E., Knutzen, L., Zhou, Y., Chia, D., Liu, A. & Goodglick, L. 2010. Differential expression of anterior gradient gene AGR2 in prostate cancer. *Bmc Cancer,* 10**,** 680.

Markopoulos, A. K., Michailidou, E. Z. & Tzimagiorgis, G. 2010. Salivary markers for oral cancer detection. *The open dentistry journal,* 4**,** 172-8.

Martin, L. J., North, K. E., Dyer, T., Blangero, J., Comuzzie, A. G. & Williams, J. 2003. Phenotypic, genetic, and genome-wide structure in the metabolic syndrome. *Bmc Genetics,* 4.

Mas, V. R., Fassnacht, R., Archer, K. J. & Maluf, D. 2010. Molecular Mechanisms Involved in the Interaction Effects of Alcohol and Hepatitis C Virus in Liver Cirrhosis. *Molecular Medicine,* 16**,** 287-297.

Masamune, A., Satoh, A., Watanabe, T., Kikuta, K., Satoh, M., Suzuki, N., Satoh, K. & Shimosegawa, T. 2010. Effects of ethanol and its metabolites on human pancreatic stellate cells. *Dig Dis Sci,* 55**,** 204-11.

Masckauchan, T. N. H., Agalliu, D., Vorontchikhina, M., Ahn, A., Parmalee, N. L., Li, C.-M., Khoo, A., Tycko, B., Brown, A. M. C. & Kitajewski, J. 2006. Wnt5a signaling induces proliferation and survival of endothelial cells in vitro and expression of MMP-1 and Tie-2. *Molecular Biology of the Cell,* 17**,** 5163-5172.

Masferrer, J. L., Leahy, K. M., Koki, A. T., Zweifel, B. S., Settle, S. L., Woerner, B. M., Edwards, D. A., Flickinger, A. G., Moore, R. J. & Seibert, K. 2000. Antiangiogenic and Antitumor Activities of Cyclooxygenase-2 Inhibitors. *Cancer Research,* 60**,** 1306-1311.

Mashberg, A. 1977. Erythroplasia Vs Leukoplakia in Diagnosis of Early Asymptomatic Oral Squamous Carcinoma. *New England Journal of Medicine,* 297**,** 109-110.

Mashberg, A., Boffetta, P., Winkelman, R. & Garfinkel, L. 1993. Tobacco Smoking, Alcohol-Drinking, and Cancer of the Oral Cavity and Oropharynx among United-States Veterans. *Cancer,* 72**,** 1369-1375.

Matlin, A. J., Clark, F. & Smith, C. W. J. 2005. Understanding alternative splicing: Towards a cellular code. *Nature Reviews Molecular Cell Biology,* 6**,** 386-398.

Matrisian, L. M., Cunha, G. R. & Mohla, S. 2001. Epithelial-stromal interactions and tumor progression: Meeting summary and future directions. *Cancer Research,* 61**,** 3844-3846.

Matsushita, A., Gotze, T. & Korc, M. 2007. Hepatocyte growth factor-mediated cell invasion in pancreatic cancer cells is dependent on neuropilin-1. *Cancer Research,* 67**,** 10309-10316.

Mazurek, S. & Eigenbrodt, E. 2003. The tumor metabolome. *Anticancer Research,* 23**,** 1149-1154.

McCarthy, J. B., Basara, M. L., Palm, S. L., Sas, D. F. & Furcht, L. T. 1985. The Role of Cell-Adhesion Proteins - Laminin and Fibronectin - in the Movement of Malignant and Metastatic Cells. *Cancer and Metastasis Reviews,* 4**,** 125-152.

McCoy, G. D. & Wynder, E. L. 1979. Etiological and preventive implications in alcohol carcinogenesis. *Cancer Res,* 39**,** 2844-50.

McMahon, S. K., Pretorius, C. J., Ungerer, J. P. J., Salmon, N. J., Conwell, L. S., Pearen, M. A. & Batch, J. A. 2010. Neonatal Complete Generalized Glucocorticoid Resistance and Growth Hormone Deficiency Caused by a Novel Homozygous Mutation in Helix 12 of the Ligand Binding Domain of the Glucocorticoid Receptor Gene (NR3C1). *Journal of Clinical Endocrinology & Metabolism,* 95**,** 297-302.

Mellanen, P., Minn, H., Grenman, R. & Harkonen, P. 1994. Expression of glucose transporters in head-and-neck tumors. *Int J Cancer,* 56**,** 622-9.

Mendelsohn, J. 1997. Epidermal growth factor receptor inhibition by a monoclonal antibody as anticancer therapy. *Clinical Cancer Research,* 3**,** 2703-2707.

Mendes, P., Kell, D. B. & Westerhoff, H. V. 1992. Channeling Can Decrease Pool Size. *European Journal of Biochemistry,* 204**,** 257-266.

Mendes, P., Kell, D. B. & Westerhoff, H. V. 1996. Why and when channelling can decrease pool size at constant net flux in a simple dynamic channel. *Biochimica Et Biophysica Acta-General Subjects,* 1289**,** 175-186.

Mendez, E., Cheng, C., Farwell, D. G., Ricks, S., Agoff, S. N., Futran, N. D., Weymuller, E. A., Maronian, N. C., Zhao, L. P. & Chen, C. 2002. Transcriptional expression profiles of oral squamous cell carcinomas. *Cancer,* 95**,** 1482-1494.

Miceli, M. V. & Jazwinski, S. M. 2005. Common and cell type-specific responses of human cells to mitochondrial dysfunction. *Experimental Cell Research,* 302**,** 270-280.

Michiels, S., Koscielny, S. & Hill, C. 2005. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet,* 365**,** 488-492.

Mimori, K., Shiraishi, T., Mashino, K., Sonoda, H., Yamashita, K., Yoshinaga, K., Masuda, T., Utsunomiya, T., Alonso, M. A., Inoue, H. & Mori, M. 2003. MAL gene expression in esophageal cancer suppresses motility, invasion and tumorigenicity and enhances apoptosis through the Fas pathway. *Oncogene,* 22**,** 3463-3471.

Modrek, B. & Lee, C. 2002. A genomic view of alternative splicing. *Nature Genetics,* 30**,** 13-19.

Mohammed, S. I., Knapp, D. W., Bostwick, D. G., Foster, R. S., Khan, K. N. M., Masferrer, J. L., Woerner, B. M., Snyder, P. W. & Koki, A. T. 1999.

Expression of cyclooxygenase-2 (COX-2) in human invasive transitional cell carcinoma (TCC) of the urinary bladder. *Cancer Research,* 59**,** 5647-5650.

Mohan, S. & Epstein, J. B. 2003. Carcinogenesis and cyclooxygenase: the potential role of COX-2 inhibition in upper aerodigestive tract cancer. *Oral oncology,* 39**,** 537-546.

Moilanen, M., Sorsa, T., Stenman, M., Nyberg, P., Lindy, O., Vesterinen, J., Paju, A., Konttinen, Y. T., Stenman, U. H. & Salo, T. 2003. Tumor-associated trypsinogen-2 (trypsinogen-2) activates procollagenases (MMP-1,-8,-13) and stromelysin-1 (MMP-3) and degrades type I collagen. *Biochemistry,* 42**,** 5414-5420.

Moinfar, F., Man, Y. G., Arnould, L., Bratthauer, G. L., Ratschek, M. & Tavassoli, F. A. 2000. Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: Implications for tumorigenesis. *Cancer Research,* 60**,** 2562-2566.

Moka, D., Vorreuther, R., Schicha, H., Spraul, M., Humpfer, E., Lipinski, M., Foxall, P. J. D., Nicholson, J. K. & Lindon, J. C. 1998. Biochemical classification of kidney carcinoma biopsy samples using magic-angle-spinning H-1 nuclear magnetic resonance spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis,* 17**,** 125-132.

Monego, G., Lauriola, L., Ramella, S., D'Angelillo, R. M., Lanza, P., Granone, P. & Ranelletti, F. O. 2010. Parathyroid Hormone-Related Peptide and Parathyroid Hormone-Related Peptide Receptor Type 1 Expression in Human Lung Adenocarcinoma. *Chest,* 137**,** 898-908.

Moore, M. A., Yoo, K.-Y., Tuncer, M. & Sobue, T. 2010. Overview of Players and Information in the Cancer Epidemiology and Control World in Asia INTRODUCTION. *Asian Pacific Journal of Cancer Prevention,* 11**,** 1-10.

Moore, S. R., Johnson, N. W., Pierce, A. M. & Wilson, D. F. 2000. The epidemiology of mouth cancer: a review of global incidence. *Oral Diseases,* 6**,** 65-74.

Morimoto, I., Sasaki, Y., Ishida, S., Imai, K. & Tokino, T. 2002. Identification of the osteopontin gene as a direct target of TP53. *Genes Chromosomes & Cancer,* 33**,** 270-278.

Moriya, T., Seki, N., Shimada, K., Kato, M., Yakushiji, T., Nimura, Y., Uzawa, K., Takiguchi, M. & Tanzawa, H. 2003. In-house cDNA microarray analysis of gene expression profiles involved in SCC cell lines. *International Journal of Molecular Medicine,* 12**,** 429-435.

Morrison, N., Bearden, D., Bundy, J. G., Collette, T., Currie, F., Davey, M. P., Haigh, N. S., Hancock, D., Jones, O. A. H., Rochfort, S., Sansone, S.-A., Stys, D., Teng, Q., Field, D. & Viant, M. R. 2007. Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics,* 3**,** 203-210.

Morse, D. E., Psoter, W. J., Cleveland, D., Cohen, D., Mohit-Tabatabai, M., Kosis, D. L. & Eisenberg, E. 2007. Smoking and drinking in relation to oral cancer and oral epithelial dysplasia. *Cancer Causes & Control,* 18**,** 919-929.

Morvan, D. & Demidem, A. 2007. Metabolomics by proton nuclear magnetic resonance spectroscopy of the response to chloroethylnitrosourea reveals

drug efficacy and tumor adaptive metabolic pathways. *Cancer Research,* 67**,** 2150-2159.

Mou, D. C., Cai, S. L., Peng, J. R., Wang, Y., Chen, H. S., Pang, X. W., Leng, X. S. & Chen, W. F. 2002. Evaluation of MAGE-1 and MAGE-3 as tumour-specific markers to detect blood dissemination of hepatocellular carcinoma cells. *British Journal of Cancer,* 86**,** 110-116.

Mueller, M. M. & Fusenig, N. E. 2002. Tumor-stroma interactions directing phenotype and progression of epithelial skin tumor cells. *Differentiation,* 70**,** 486-497.

Muir, C. S. & Kirk, R. 1960. Betel, Tobacco, and Cancer of the Mouth. *British Journal of Cancer,* 14**,** 597-&.

Mukherji, S. K., Schiro, S., Castillo, M., Kwock, L., Muller, K. E. & Blackstock, W. 1997. Proton MR spectroscopy of squamous cell carcinoma of the extracranial head and neck: In vitro and in vivo studies. *American Journal of Neuroradiology,* 18**,** 1057-1072.

Mukherji, S. K., Schiro, S., Castillo, M., Kwock, L., Soper, R., Blackstock, W., Shockley, W. & Weissler, M. 1996. Proton MR spectroscopy of squamous cell carcinoma of the upper aerodigestive tract: In vitro characteristics. *American Journal of Neuroradiology,* 17**,** 1485-1490.

Mulryan, K., Ryan, M. G., Myers, K. A., Shaw, D., Wang, W., Kingsman, S. M., Stern, P. L. & Carroll, M. W. 2002. Attenuated recombinant vaccinia virus expressing oncofetal antigen (tumor-associated antigen) 5T4 induces active therapy of established tumors. *Molecular Cancer Therapeutics,* 1**,** 1129-1137.

Murphy, M., Godson, C., Cannon, S., Kato, S., Mackenzie, H. S., Martin, F. & Brady, H. R. 1999. Suppression subtractive hybridization identifies high glucose levels as a stimulus for expression of connective tissue growth factor and other genes in human mesangial cells. *Journal of Biological Chemistry,* 274**,** 5830-5834.

Nadege, B., Patrick, L. & Rodrigue, R. 2009. Mitochondria: from bioenergetics to the metabolic regulation of carcinogenesis. *Frontiers in Bioscience,* 14**,** 4015-4034.

Nagata, M., Fujita, H., Ida, H., Hoshina, H., Inoue, T., Seki, Y., Ohnishi, M., Ohyama, T., Shingaki, S., Kaji, M., Saku, T. & Takagi, R. 2003. Identification of potential biomarkers of lymph node metastasis in oral squamous cell carcinoma by cDNA microarray analysis. *International Journal of Cancer,* 106**,** 683-689.

Nathan, C. A. O., Leskov, I. L., Lin, M. H., Abreo, F. W., Shi, R. H., Hartman, G. H. & Glass, J. 2001. COX-2 expression in dysplasia of the head and neck - Correlation with elF4E. *Cancer,* 92**,** 1888-1895.

Nelson, A. R., Fingleton, B., Rothenberg, M. L. & Matrisian, L. M. 2000. Matrix Metalloproteinases: Biologic Activity and Clinical Implications. *Journal of Clinical Oncology,* 18**,** 1135.

Nelson, N. J. 2006. Years of research come to fruition with launch of oral cancer prevention trial. *Journal of the National Cancer Institute,* 98**,** 88-89.

Neville, B. W. & Day, T. A. 2002. Oral cancer and precancerous lesions. *Ca-a Cancer Journal for Clinicians,* 52**,** 195-215.

Newmeyer, D. D., Farschon, D. M. & Reed, J. C. 1994. Cell-free apoptosis in Xenopus egg extracts: Inhibition by Bcl-2 and requirement for an organelle fraction enriched in mitochondria. *Cell,* 79**,** 353-364.

Nguyen, M., Millar, D. G., Yong, V. W., Korsmeyer, S. J. & Shore, G. C. 1993. Targeting of Bcl-2 to the mitochondrial outer membrane by a COOH-terminal signal anchor sequence. *Journal of Biological Chemistry,* 268**,** 25265-8.

Ni, H. L., Yan, X. R., Lin, Z. Y. & Jin, X. M. 2011. High glucose may decrease the innate immune through TLRs in cornea epithelium. *Molecular Vision,* 17**,** 3384-3391.

Nicholson, J. K., Buckingham, M. J. & Sadler, P. J. 1983. High-Resolution H-1-Nmr Studies of Vertebrate Blood and Plasma. *Biochemical Journal,* 211**,** 605-615.

Nicholson, J. K., Lindon, J. C. & Holmes, E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica,* 29**,** 1181-1189.

Nicholson, J. K., Timbrell, J. A. & Sadler, P. J. 1985. Proton Nmr-Spectra of Urine as Indicators of Renal Damage - Mercury-Induced Nephrotoxicity in Rats. *Molecular Pharmacology,* 27**,** 644-651.

Nielsen, J. 2003. It is all about metabolic fluxes. *Journal of Bacteriology,* 185**,** 7031-7035.

Nieto, N., Greenwel, P., Friedman, S. L., Zhang, F., Dannenberg, A. J. & Cederbaum, A. I. 2000. Ethanol and arachidonic acid increase alpha 2(I) collagen expression in rat hepatic stellate cells overexpressing cytochrome P450 2E1 - Role of H2O2 and cyclooxygenase-2. *Journal of Biological Chemistry,* 275**,** 20136-20145.

Nishihara, K., Nozoe, E., Hirayama, Y., Miyawalki, A., Semba, I. & Nakamura, N. 2009. A case of small cell carcinoma in the buccal region. *International Journal of Oral and Maxillofacial Surgery,* 38**,** 1000-1003.

Nisijima, K. & Ishiguro, T. 1995. Cerebrospinal fluid levels of monoamine metabolites and gamma-aminobutyric acid in neuroleptic malignant syndrome. *Journal of Psychiatric Research,* 29**,** 233-244.

Niu, R. F., Zhang, L., Xi, G. M., Wei, X. Y., Yang, Y., Shi, Y. R. & Hao, X. S. 2007. Up-regulation of twist induces angiogenesis and correlates with metastasis in hepatocellular carcinoma. *Journal of Experimental & Clinical Cancer Research,* 26**,** 385-394.

Nordstrom, A. & Lewensohn, R. 2010. Metabolomics: moving to the clinic. *J Neuroimmune Pharmacol,* 5**,** 4-17.

Norkina, O., Dolganiuc, A., Catalano, D., Kodys, K., Mandrekar, P., Syed, A., Efros, M. & Szabo, G. 2008. Acute alcohol intake induces SOCS1 and SOCS3 and inhibits cytokine-induced STAT1 and STAT3 signaling in human monocytes. *Alcoholism-Clinical and Experimental Research,* 32**,** 1565-1573.

Norwood, V. F., Morham, S. G. & Smithies, O. 2000. Postnatal development and progression of renal dysplasia in cyclooxygenase-2 null mice. *Kidney Int,* 58**,** 2291-2300.

Obrador, E., Benlloch, M., Pellicer, J. A., Asensi, M. & Estrela, J. M. 2011. Intertissue flow of glutathione (GSH) as a tumor growth-promoting

mechanism: interleukin 6 induces GSH release from hepatocytes in metastatic B16 melanoma-bearing mice. *J Biol Chem,* 286**,** 15716-27.

Odani, T., Ito, D., Li, M. H., Kawamata, A. I., Isobe, T., Iwase, M. & Nagumo, M. 2006. Gene expression profiles of oral leukoplakia and carcinoma: Genome-wide comparison analysis using oligonucleotide microarray technology. *International Journal of Oncology,* 28**,** 619-624.

Odunsi, K., Wollman, R. M., Ambrosone, C. B., Hutson, A., McCann, S. E., Tammela, J., Geisler, J. P., Miller, G., Sellers, T., Cliby, W., Qian, F., Keitz, B., Intengan, M., Lele, S. & Alderfer, J. L. 2005. Detection of epithelial ovarian cancer using H-1-NMR-based metabonomics. *International Journal of Cancer,* 113**,** 782-788.

Ogden, G. R. 2005. Alcohol and oral cancer. *Alcohol,* 35**,** 169-173.

Oh, Y. & Mao, L. 1997. Biomarkers in head and neck carcinoma. *Current Opinion in Oncology,* 9**,** 247-256.

Okamoto, T., Akaike, T., Sawa, T., Miyamoto, Y., van der Vliet, A. & Maeda, H. 2001. Activation of matrix metalloproteinases by peroxynitrite-induced protein S-glutathiolation via disulfide S-oxide formation. *J Biol Chem,* 276**,** 29596-602.

Okuducu, A. F., Janzen, V., Ko, Y., Hahne, J. C., Lu, H., Ma, Z. L., Albers, P., Sahin, A., Wellmann, A., Scheinert, P. & Wernert, N. 2005. Cellular retinoic acid-binding protein 2 is down-regulated in prostate cancer. *International Journal of Oncology,* 27**,** 1273-1282.

Olsen, J., Kirkeby, L. T., Brorsson, M. M., Dabelsteen, S., Troelsen, J. T., Bordoy, R., Fenger, K., Larsson, L. I. & Simon-Assmann, P. 2003. Converging signals synergistically activate the LAMC2 promoter and lead to accumulation of the laminin gamma 2 chain in human colon carcinoma cells. *Biochemical Journal,* 371**,** 211-221.

Ouyang, G., Liu, M., Ruan, K., Song, G., Mao, Y. & Bao, S. 2009. Upregulated expression of periostin by hypoxia in non-small-cell lung cancer cells promotes cell survival via the Akt/PKB pathway. *Cancer Letters,* 281**,** 213-219.

Park, E., Alberti, J., Quinn, M. R. & Schuller-Levis, G. 1998. Taurine chloramine inhibits the production of superoxide anion, IL-6 and IL-8 in activated human polymorphonuclear leukocytes. *In:* SCHAFFER, S., LOMBARDINI, J. B. & HUXTABLE, R. J. (eds.) *Taurine 3: Cellular and Regulatory Mechanisms.*

Park, E., Jia, J., Quinn, M. R. & Schuller-Levis, G. 2002. Taurine chloramine inhibits lymphocyte proliferation and decreases cytokine production in activated human leukocytes. *Clinical Immunology,* 102**,** 179-184.

Park, K. W., Nozell, S. E. & Benveniste, E. N. 2012. Protective Role of STAT3 in NMDA and Glutamate-Induced Neuronal Death: Negative Regulatory Effect of SOCS3. *Plos One,* 7.

Patti, M. E., Butte, A. J., Crunkhorn, S., Cusi, K., Berria, R., Kashyap, S., Miyazaki, Y., Kohane, I., Costello, M., Saccone, R., Landaker, E. J., Goldfine, A. B., Mun, E., DeFronzo, R., Finlayson, J., Kahn, C. R. & Mandarino, L. J. 2003. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1.

Proceedings of the National Academy of Sciences of the United States of America, 100**,** 8466-8471.

Pavlova, I., Williams, M., El-Naggar, A., Richards-Kortum, R. & Gillenwater, A. 2008. Understanding the biological basis of autofluorescence imaging for oral cancer detection: High-resolution fluorescence microscopy in viable tissue. *Clinical Cancer Research,* 14**,** 2396-2404.

Pedersen, P. L. 1978. Tumor mitochondria and the bioenergetics of cancer cells. *Progress in experimental tumor research,* 22**,** 190-274.

Pendyala, G., Want, E. J., Webb, W., Siuzdak, G. & Fox, H. S. 2007. Biomarkers for neuroAIDS: The widening scope of metabolomics. *Journal of Neuroimmune Pharmacology,* 2**,** 72-80.

Peng, C.-H., Liao, C.-T., Peng, S.-C., Chen, Y.-J., Cheng, A.-J., Juang, J.-L., Tsai, C.-Y., Chen, T.-C., Chuang, Y.-J., Tang, C.-Y., Hsieh, W.-P. & Yen, T.-C. 2011. A Novel Molecular Signature Identified by Systems Genetics Approach Predicts Prognosis in Oral Squamous Cell Carcinoma. *PLoS One,* 6.

Pettus, J. R., Johnson, J. J., Shi, Z. G., Davis, J. W., Koblinski, J., Ghosh, S., Liu, Y. Y., Ravosa, M. J., Frazier, S. & Stack, M. S. 2009. Multiple kallikrein (KLK 5, 7, 8, and 10) expression in squamous cell carcinoma of the oral cavity. *Histology and Histopathology,* 24**,** 197-207.

Phillips, K. P. & Foster, W. G. 2008. Key developments in endocrine disrupter research and human health. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews,* 11**,** 322-344.

Piazza, C., Dessouky, O., Peretti, G., Cocco, D., De Benedetto, L. & Nicolai, P. 2008. Narrow-band imaging: a new tool for evaluation of head and neck squamous cell carcinomas. Review of the literature. *Acta Otorhinolaryngologica Italica,* 28**,** 49-54.

Pirasath, S., Thayaananthan, K., Balakumar, S. & Arasaratnam, V. 2010. Effect of dietary curries on the glycaemic index. *Ceylon Medical Journal,* 55**,** 118-122.

Pires, P. W., Furtado, K. S., Justullin, L. A., Jr., Marchesan Rodrigues, M. A., Felisbino, S. L. & Barbisan, L. F. 2008. Chronic ethanol intake promotes double gluthatione S-transferase transforming growth factor-alpha-positive hepatocellular lesions in male Wistar rats. *Cancer Science,* 99**,** 221-228.

Pratt, W. B., Morishima, Y., Murphy, M. & Harrell, M. 2006. Chaperoning of glucocorticoid receptors. *Handbook of Experimental Pharmacology.*

Prewett, M., Rockwell, P., Rose, C. & Goldstein, N. I. 1996. Anti-tumor and cell cycle responses in KB cells treated with a chimeric anti-EGFR monoclonal antibody in combination with cisplatin. *International Journal of Oncology,* 9**,** 217-224.

Psyrri, A., Yu, Z. W., Weinberger, P. M., Sasaki, C., Haffty, B., Camp, R., Rimm, D. & Burtness, B. A. 2005. Quantitative determination of nuclear and cytoplasmic epidermal growth factor receptor expression in oropharyngeal squamous cell cancer by using automated quantitative analysis. *Clinical Cancer Research,* 11**,** 5856-5862.

Pukrop, T., Klemm, F., Hagemann, T., Gradl, D., Schulz, M., Siemes, S., Trumper, L. & Binder, C. 2006. Wnt 5a signaling is critical for macrophage-induced

invasion of breast cancer cell lines. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 5454-5459.

Purohit, V. & Brenner, D. A. 2006. Mechanisms of alcohol-induced hepatic fibrosis: A summary of the Ron Thurman Symposium. *Hepatology,* 43**,** 872-878.

Pyeon, D., Newton, N. A., Lambert, P. F., den Boon, J. A., Sengupta, S., Marsit, C. J., Woodworth, C. D., Connor, J. P., Haugen, T. H., Smith, E. M., Kelsey, K. T., Turek, L. P. & Ahlquist, P. 2007. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Research,* 67**,** 4605-4619.

Quinn, M. R., Park, E. & SchullerLevis, G. 1996. Taurine chloramine inhibits prostaglandin E(2) production in activated RAW 264.7 cells by post-transcriptional effects on inducible cyclooxygenase expression. *Immunology Letters,* 50**,** 185-188.

Ralph, S. J., Rodriguez-Enriquez, S., Neuzil, J. & Moreno-Sanchez, R. 2010. Bioenergetic pathways in tumor mitochondria as targets for cancer therapy and the importance of the ROS-induced apoptotic trigger. *Molecular Aspects of Medicine,* 31**,** 29-59.

Ranasinghe, A. W., Warnakulasuriya, K. A. & Johnson, N. W. 1993. Low prevalence of expression of p53 oncoprotein in oral carcinomas from Sri Lanka associated with betel and tobacco chewing. *European journal of cancer. Part B, Oral oncology,* 29B**,** 147-50.

Ransohoff, D. F. 2005. Lessons from controversy: Ovarian cancer screening and serum proteomics. *Journal of the National Cancer Institute,* 97**,** 315-319.

Rao, C. V., Wang, C. Q., Simi, B., Rodriguez, J. G., Cooma, I., El-Bayoumy, K. & Reddy, B. S. 2001. Chemoprevention of colon cancer by a glutathione conjugate of 1,4-phenylenebis(methylene)selenocyanate, a novel organoselenium compound with low toxicity. *Cancer Research,* 61**,** 3647-3652.

Rasmuson, T., Grankvist, K. & Ljungberg, B. 1993. Serum Gamma-Enolase and Prognosis of Patients with Renal-Cell Carcinoma. *Cancer,* 72**,** 1324-1328.

Ratnasinghe, D., Tangrea, J., Roth, M. J., Dawsey, S., Hu, N., Anver, M., Wang, Q. H. & Taylor, P. R. 1999. Expression of cyclooxygenase-2 in human squamous cell carcinoma of the esophagus; an immunohistochemical survey. *Anticancer Research,* 19**,** 171-174.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. 2002. Hierarchical organization of modularity in metabolic networks. *Science,* 297**,** 1551-1555.

Ray, A. & Prefontaine, K. E. 1994. Physical Association and Functional Antagonism between the P65 Subunit of Transcription Factor Nf-Kappa-B and the Glucocorticoid Receptor. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 752-756.

Rehman, I., Quinn, A. G., Healy, E. & Rees, J. L. 1994. High-Frequency of Loss of Heterozygosity in Actinic Keratoses, a Usually Benign Disease. *Lancet,* 344**,** 788-789.

Reich, R. & Martin, G. R. 1996. Identification of arachidonic acid pathways required for the invasive and metastatic activity of malignant tumor cells. *Prostaglandins & Other Lipid Mediators,* 51**,** 1-17.

Reichard, K. W., Joseph, K. T., Cohen, M. & Greager, J. A. 1993. Squamous-Cell Carcinoma of the Tongue - Experience with 86 Consecutive Cases. *Journal of Surgical Oncology,* 54**,** 239-242.

Reichert, T. E., Strauss, L., Wagner, E. M., Gooding, W. & Whiteside, T. L. 2002. Signaling abnormalities, apoptosis, and reduced proliferation of circulating and tumor-infiltrating lymphocytes in patients with oral carcinoma. *Clinical Cancer Research,* 8**,** 3137-3145.

Reis, E. M., Ojopi, E. P. B., Alberto, F. L., Rahal, P., Tsukumo, F., Mancini, U. M., Guimaraes, G. S., Thompson, G. M. A., Camacho, C., Miracca, E., Carvalho, A. L., Machado, A. A., Paquola, A. C. M., Cerutti, J. M., da Silva, A. M., Pereira, G. G., Valentini, S. R., Nagai, M. A., Kowalski, L. P., Verjovski-Almeida, S., Tajara, E. H., Dias-Neto, E., Bengtson, M. H., Canevari, R. A., Carazzolle, M. F., Colin, C., Costa, F. F., Costa, M. C., Estecio, M. R., Esteves, L. I., Federico, M. H., Guimaraes, P. E., Hackel, C., Kimura, E. T., Leoni, S. G., Maciel, R. M., Maistro, S., Mangone, F. R., Massirer, K. B., Matsuo, S. E., Nobrega, F. G., Nobrega, M. P., Nunes, D. N., Nunes, F., Pandolfi, J. R., Pardini, M. I., Pasini, F. S., Peres, T., Rainho, C. A., Dos Reis, P. P., Rodrigus-Lisoni, F. C., Rogatto, S. R., dos Santos, A., dos Santos, P. C., Sogayar, M. C., Zanelli, C. F. & Head and Neck Annotation, C. 2005. Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. *Cancer Research,* 65**,** 1693-1699.

Reis, P. P., Waldron, L., Perez-Ordonez, B., Pintilie, M., Galloni, N. N., Xuan, Y., Cervigne, N. K., Warner, G. C., Makitie, A. A., Simpson, C., Goldstein, D., Brown, D., Gilbert, R., Gullane, P., Irish, J., Jurisica, I. & Kamel-Reid, S. 2011. A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *Bmc Cancer,* 11.

Rezzi, S., Ramadan, Z., Fay, L. B. & Kochhar, S. 2007. Nutritional metabonomics: Applications and perspectives. *Journal of Proteome Research,* 6**,** 513-525.

Rhen, T. & Cidlowski, J. A. 2005. Antiinflammatory action of glucocorticoids - New mechanisms for old drugs. *New England Journal of Medicine,* 353**,** 1711-1723.

Ripka, S., Koenig, A., Buchholz, M., Wagner, M., Sipos, B., Kloeppel, G., Downward, J., Gress, T. M. & Michl, P. 2007. WNT5A - target of CUTL1 and potent modulator of tumor cell migration and invasion in pancreatic cancer. *Carcinogenesis,* 28**,** 1178-1187.

Roblyer, D., Richards-Kortum, R., Sokolov, K., El-Naggar, A. K., Williams, M. D., Kurachi, C. & Gillenwater, A. M. 2008. Multispectral optical imaging device for in vivo detection of oral neoplasia. *Journal of Biomedical Optics,* 13.

Rodriguez-Enriquez, S., Marin-Hernandez, A., Carlos Gallardo-Perez, J., Carreno-Fuentes, L. & Moreno-Sanchez, R. 2009. Targeting of cancer energy metabolism. *Molecular Nutrition & Food Research,* 53**,** 29-48.

Roeder, C., Schuler-Thurner, B., Berchtold, S., Vieth, G., den Driesch, P. V., Schuler, G. & Luftl, M. 2005. MAGE-A3 is a frequent tumor antigen of metastasized melanoma. *Archives of Dermatological Research,* 296**,** 314-319.

Roepman, P. & Holstege, F. C. P. 2005. Tumor profiling turmoil. *Cell Cycle,* 4**,** 659-660.

Roepman, P., Kemmeren, P., Wessels, L. F. A., Slootweg, P. J. & Holstege, F. C. P. 2006. Multiple robust signatures for detecting lymph node metastasis in head and neck cancer. *Cancer Research,* 66**,** 2361-2366.

Roepman, P., Wessels, L. F. A., Kettelarij, N., Kemmeren, P., Miles, A. J., Lijnzaad, P., Tilanus, M. G. J., Koole, R., Hordijk, G. J., van der Vliet, P. C., Reinders, M. J. T., Slootweg, P. J. & Holstege, F. C. P. 2005. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nature Genetics,* 37**,** 182-186.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. & Fernie, A. R. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell,* 13**,** 11-29.

Rojas Ayala, F. R., Rocha, R. M., Carvalho, K. C., Carvalho, A. L., da Cunha, I. W., Lourenco, S. V. & Soares, F. A. 2010. Glut1 and Glut3 as Potential Prognostic Markers for Oral Squamous Cell Carcinoma. *Molecules,* 15**,** 2374-2387.

Romero, P., Wagg, J., Green, M. L., Kaiser, D., Krummenacker, M. & Karp, P. D. 2005. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology,* 6.

Rose, B. R., Thompson, C. H., Tattersall, M. H., O'Brien, C. J. & Cossart, Y. E. 2000. Squamous carcinoma of the head and neck: Molecular mechanisms and potential biomarkers. *Australian and New Zealand Journal of Surgery,* 70**,** 601-606.

Rosenthal, E. L. & Matrisian, L. M. 2006. Matrix metalloproteases in head and neck cancer. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 28**,** 639-648.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C. E., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D. & Brown, P. O. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics,* 24**,** 227-235.

Roy, S., Sala, R., Cagliero, E. & Lorenzi, M. 1990. Overexpression of fibronectin induced by diabetes or high glucose: phenomenon with a memory. *Proc Natl Acad Sci U S A,* 87**,** 404-8.

Ryan, D. & Robards, K. 2006. Metabolomics: The greatest omics of them all? *Analytical Chemistry,* 78**,** 7954-7958.

Safran, M., Solomon, I., Shmueli, O., Lapidot, M., Shen-Orr, S., Adato, A., Ben-Dor, U., Esterman, N., Rosen, N., Peter, I., Olender, T., Chalifa-Caspi, V. & Lancet, D. 2002. GeneCards (TM) 2002: towards a complete, object-oriented, human gene compendium. *Bioinformatics,* 18**,** 1542-1543.

Sartini, D., Santarelli, A., Rossi, V., Goteri, G., Rubini, C., Ciavarella, D., Lo Muzio, L. & Emanuelli, M. 2007. Nicotinamide N-methyltransferase upregulation inversely correlates with lymph node metastasis in oral squamous cell carcinoma. *Molecular Medicine,* 13**,** 415-421.

Sasahira, T., Kirita, T., Bhawal, U. K., Ikeda, M., Nagasawa, A., Yamamoto, K. & Kuniyasu, H. 2007. The expression of receptor for advanced glycation end products is associated with angiogenesis in human oral squamous cell carcinoma. *Virchows Archiv,* 450**,** 287-295.

Sasaki, H., Auclair, D., Fukai, I., Kiriyama, M., Yamakawa, Y., Fujii, Y. & Chen, L. B. 2001a. Serum level of the periostin, a homologue of an insect cell adhesion molecule, as a prognostic marker in nonsmall cell lung carcinomas. *Cancer,* 92**,** 843-848.

Sasaki, H., Lo, K. M., Chen, L. B., Auclair, D., Nakashima, Y., Moriyama, S., Fukai, I., Tam, C., Loda, M. & Fujii, Y. 2001b. Expression of Periostin, homologous with an insect cell adhesion molecule, as a prognostic marker in non-small cell lung cancers. *Japanese Journal of Cancer Research,* 92**,** 869-873.

Sasaki, H., Sato, Y., Kondo, S., Fukai, I., Kiriyama, M., Yamakawa, Y. & Fuji, Y. 2002. Expression of the periostin mRNA level in neuroblastoma. *Journal of Pediatric Surgery,* 37**,** 1293-1297.

Sasaki, H., Yu, C.-Y., Dai, M., Tam, C., Loda, M., Auclair, D., Chen, L. B. & Elias, A. 2003. Elevated Serum Periostin Levels in Patients with Bone Metastases from Breast but not Lung Cancer. *Breast Cancer Research and Treatment,* 77**,** 245-252.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. 1995. Quantitative Monitoring of Gene-Expression Patterns with a Complementary-DNA Microarray. *Science,* 270**,** 467-470.

Schmalbach, C. E., Chepeha, D. B., Giordano, T. J., Rubin, M. A., Teknos, T. N., Bradford, C. R., Wolf, G. T., Kuick, R., Misek, D. E., Trask, D. K. & Hanash, S. 2004. Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. *Archives of Otolaryngology-Head & Neck Surgery,* 130**,** 295-302.

Schneider-Stock, R., Mawrin, C., Motsch, C., Boltze, C., Peters, B., Hartig, R., Buhtz, P., Giers, A., Rohrbeck, A., Freigang, B. & Roessner, A. 2004. Retention of the arginine allele in codon 72 of the p53 gene correlates with poor apoptosis in head and neck cancer. *American Journal of Pathology,* 164**,** 1233-1241.

Scholz, R. W., Graham, K. S., Gumpricht, E. & Reddy, C. C. 1989. Mechanism of Interaction of Vitamin E and Glutathione in the Protection against Membrane Lipid Peroxidation. *Annals of the New York Academy of Sciences,* 570**,** 514-517.

Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. & Herzel, H. 2000. Normalization strategies for cDNA microarrays. *Nucleic Acids Research,* 28**,** e47.

Schulzeosthoff, K., Bakker, A. C., Vanhaesebroeck, B., Beyaert, R., Jacob, W. A. & Fiers, W. 1992. Cytotoxic Activity of Tumor-Necrosis-Factor Is Mediated by Early Damage of Mitochondrial Functions - Evidence for the Involvement of Mitochondrial Radical Generation. *Journal of Biological Chemistry,* 267**,** 5317-5323.

Schwartz, S. M., Doody, D. R., Fitzgibbons, E. D., Ricks, S., Porter, P. L. & Chen, C. 2001. Oral squamous cell cancer risk in relation to alcohol consumption and alcohol dehydrogenase-3 genotypes. *Cancer Epidemiology Biomarkers & Prevention,* 10**,** 1137-1144.

Schwarz, R. A., Gao, W., Daye, D., Williams, M. D., Richards-Kortum, R. & Gillenwater, A. M. 2008. Autofluorescence and diffuse reflectance

spectroscopy of oral epithelial tissue using a depth-sensitive fiber-optic probe. *Applied Optics,* 47**,** 825-834.

Seiki, M. 2003. Membrane-type 1 matrix metalloproteinase: a key enzyme for tumor invasion. *Cancer Letters,* 194**,** 1-11.

Sepulveda, V. A. T., Shen, X. L. & Falzon, M. 2002. Intracrine PTHrp protects against serum starvation-induced apoptosis and regulates the cell cycle in MCF-7 breast cancer cells. *Endocrinology,* 143**,** 596-606.

Serkova, N. J. & Niemann, C. U. 2006. Pattern recognition and biomarker validation using quantitative H-1-NMR-based metabolomics. *Expert Review of Molecular Diagnostics,* 6**,** 717-731.

Serkova, N. J., Spratlin, J. L. & Eckhardt, S. G. 2007. NMR-based metabolomics: Translational application and treatment of cancer. *Current Opinion in Molecular Therapeutics,* 9**,** 572-585.

Sethi, N., Dai, X., Winter, C. G. & Kang, Y. 2011. Tumor-Derived Jagged1 Promotes Osteolytic Bone Metastasis of Breast Cancer by Engaging Notch Signaling in Bone Cells. *Cancer Cell,* 19**,** 192-205.

Severino, P., Alvares, A. M., Michaluart, P., Jr., Okamoto, O. K., Nunes, F. D., Moreira-Filho, C. A., Tajara, E. H., Head & Neck Genome Project, G. 2008. Global gene expression profiling of oral cavity cancers suggests molecular heterogeneity within anatomic subsites. *BMC research notes,* 1**,** 113-113.

Seyfried, T. & Shelton, L. 2010. Cancer as a metabolic disease. *Nutrition & Metabolism,* 7**,** 7.

Shaaban, A. M., Sloane, J. P., West, C. R., Moore, F. R., Jarvis, C., Williams, E. M. I. & Foster, C. S. 2002. Histopathologic types of benign breast lesions and the risk of breast cancer - Case-control study. *American Journal of Surgical Pathology,* 26**,** 421-430.

Shalinsky, D. R., Bischoff, E. D., Gregory, M. L., Gottardis, M. M., Hayes, J. S., Lamph, W. W., Heyman, R. A., Shirley, M. A., Cooke, T. A. & Davies, P. J. A. 1995. Retinoid-Induced Suppression of Squamous-Cell Differentiation in Human Oral Squamous-Cell Carcinoma Xenografts (Line-1483) in Athymic Nude-Mice. *Cancer Research,* 55**,** 3183-3191.

Shanmugam, N., Todorov, I. T., Nair, I., Omori, K., Reddy, M. A. & Natarajan, R. 2006. Increased expression of cyclooxygenase-2 in human pancreatic islets treated with high glucose or ligands of the advanced glycation endproduct-specific receptor (AGER), and in islets from diabetic mice. *Diabetologia,* 49**,** 100-107.

Shao, R., Bao, S. D., Bai, X. F., Blanchette, C., Anderson, R. M., Dang, T. Y., Gishizky, M. L., Marks, J. R. & Wang, X. F. 2004. Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. *Molecular and Cellular Biology,* 24**,** 3992-4003.

Sharan, R. N. 1996. Association of betel nut with carcinogenesis. *Cancer Journal,* 9**,** 13-19.

Sharwani, A., Jerjes, W., Salih, V., Swinson, B., Bigio, I. J., El-Maaytah, M. & Hopper, C. 2006. Assessment of oral premalignancy using elastic scattering spectroscopy. *Oral Oncology,* 42**,** 343-349.

Shaulian, E. & Karin, M. 2001. AP-1 in cell proliferation and survival. *Oncogene,* 20**,** 2390-2400.

Shibata, M., Kodani, I., Osalki, M., Araki, K., Adachi, H., Ryoke, K. & Ito, H. 2005. Cyclo-oxygenase-1 and-2 expression in human oral mucosa, dysplasias and squamous cell carcinomas and their pathological significance. *Oral Oncology,* 41**,** 304-312.

Shigeishi, H., Fujimoto, S., Hiraoka, M., Ono, S., Taki, M., Ohta, K., Higashikawa, K. & Kamata, N. 2009. Overexpression of the receptor for hyaluronan-mediated motility, correlates with expression of microtubule-associated protein in human oral squamous cell carcinomas. *International Journal of Oncology,* 34**,** 1565-1571.

Shimada, K., Uzawa, K., Kato, M., Endo, Y., Shiiba, M., Bukawa, H., Yokoe, H., Seki, N. & Tanzawa, H. 2005. Aberrant expression of RAB1A in human tongue cancer. *British Journal of Cancer,* 92**,** 1915-1921.

Shpitzer, T., Bahar, G., Feinmesser, R. & Nagler, R. M. 2007. A comprehensive salivary analysis for oral cancer diagnosis. *Journal of Cancer Research and Clinical Oncology,* 133**,** 613-617.

Shureiqi, I., Xu, X. C., Chen, D. N., Lotan, R., Morris, J. S., Fischer, S. M. & Lippman, S. M. 2001. Nonsteroidal anti-inflammatory drugs induce apoptosis in esophageal cancer cells by restoring 15-lipoxygenase-1 expression. *Cancer Research,* 61**,** 4879-4884.

Siegel, P. M. & Massague, J. 2003. Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. *Nature Reviews Cancer,* 3**,** 807-820.

Sienel, W., Varwerk, C., Linder, A., Kaiser, D., Teschner, A., Delire, A., Stamatis, G. & Passlick, B. 2004. Melanoma associated antigen (MAGE)-A3 expression in Stages I and II non-small cell lung cancer: results of a multi-center study. *European Journal of Cardio-Thoracic Surgery,* 25**,** 131-134.

Sims, A. H., Smethurst, G. J., Hey, Y., Okoniewski, M. J., Pepper, S. D., Howell, A., Miller, C. J. & Clarke, R. B. 2008. The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets - improving meta-analysis and prediction of prognosis. *Bmc Medical Genomics,* 1.

Singh, K. K., Kulawiec, M., Still, I., Desouki, M. M., Geradts, J. & Matsui, S.-I. 2005. Inter-genomic cross talk between mitochondria and the nucleus plays an important role in tumorigenesis. *Gene,* 354**,** 140-146.

Singh, R. & Valcarcel, J. 2005. Building specificity with nonspecific RNA-binding proteins. *Nature Structural & Molecular Biology,* 12**,** 645-653.

Sinpitaksakul, S. N., Pimkhaokham, A., Sanchavanakit, N. & Pavasant, P. 2008. TGF-beta 1 induced NIMP-9 expression in HNSCC cell lines via Smad/MLCK pathway. *Biochemical and Biophysical Research Communications,* 371**,** 713-718.

Siriwardena, B. S. M. S., Kudo, Y., Ogawa, I., Kitagawa, M., Kitajima, S., Hatano, H., Tilakaratne, W. M., Miyauchi, M. & Takata, T. 2006. Periostin is frequently overexpressed and enhances invasion and angiogenesis in oral cancer. *British Journal of Cancer,* 95**,** 1396-1403.

Slaughter, D. P., Southwick, H. W. & Smejkal, W. 1953. Field Cancerization in Oral Stratified Squamous Epithelium - Clinical Implications of Multicentric Origin. *Cancer,* 6**,** 963-968.

Slootweg, P. J., Depagter, M., Deweger, R. A. & Dewilde, P. C. M. 1994. Lymphocytes at Tumor Margins in Patients with Head and Neck-Cancer - Relationship with Tumor Size, Human Lymphocyte Antigen Molecules, and Metastasis. *International Journal of Oral and Maxillofacial Surgery,* 23**,** 286-289.

SmolkovÃ¡, K. n., PlecitÃ¡-HlavatÃ¡, L., Bellance, N. g., Benard, G., Rossignol, R. & JeÅ¾ek, P. 2011. Waves of gene regulation suppress and then restore oxidative phosphorylation in cancer cells. *The International Journal of Biochemistry & Cell Biology,* 43**,** 950-968.

Soini, Y., Maatta, M., Salo, S., Tryggvason, K. & AutioHarmainen, H. 1996. Expression of the laminin gamma 2 chain in pancreatic adenocarcinoma. *Journal of Pathology,* 180**,** 290-294.

Sok, J. C., Kuriakose, M. A., Mahajan, V. B., Pearlman, A. N., DeLacure, M. D. & Chen, F. A. 2003. Tissue-specific gene expression of head and neck squamous cell carcinoma in vivo by complementary DNA microarray analysis. *Archives of Otolaryngology-Head & Neck Surgery,* 129**,** 760-770.

Solini, A., Chiozzi, P., Falzoni, S., Morelli, A., Fellin, R. & Di Virgilio, F. 2000. High glucose modulates P2X(7) receptor-mediated function in human primary fibroblasts. *Diabetologia,* 43**,** 1248-1256.

Soltani, N., Qiu, H., Aleksic, M., Glinka, Y., Zhao, F., Liu, R., Li, Y., Zhang, N., Chakrabarti, R., Ng, T., Jin, T., Zhang, H., Lu, W.-Y., Feng, Z.-P., Prud'homme, G. J. & Wang, Q. 2011. GABA exerts protective and regenerative effects on islet beta cells and reverses diabetes. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 11692-11697.

Song, C. & Horrobin, D. 2004. Omega-3 fatty acid ethyl -eicosapentaenoate, but not soybean oil, attenuates memory impairment induced by central IL-1 beta administration. *Journal of Lipid Research,* 45**,** 1112-1121.

Song, M. K., Salam, N. K., Roufogalis, B. D. & Huang, T. H. W. 2011. Lycium barbarum (Goji Berry) extracts and its taurine component inhibit PPAR-γ-dependent gene transcription in human retinal pigment epithelial cells: Possible implications for diabetic retinopathy treatment. *Biochemical Pharmacology,* 82**,** 1209-1218.

Spandidos, D. A., Lamothe, A. & Field, J. K. 1985. Multiple Transcriptional Activation of Cellular Oncogenes in Human Head and Neck Solid Tumors. *Anticancer Research,* 5**,** 221-224.

Spratlin, J. L., Serkova, N. J. & Eckhardt, S. G. 2009. Clinical Applications of Metabolomics in Oncology: A Review. *Clinical Cancer Research,* 15**,** 431-440.

Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. & Van Wart, H. E. 1990. Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. *Proc Natl Acad Sci U S A,* 87**,** 364-8.

Sprinkel, A. M. E., Andersen, N. A. & Mandrup-Poulsen, T. 2001. Glucose potentiates interleukin-1 beta (IL-1 beta)-induced p38 mitogen-activated protein kinase activity in rat pancreatic islets of Langerhans. *European Cytokine Network,* 12**,** 331-339.

Squier, C. A., Kremer, M. J. & Wertz, P. W. 2003. Effect of ethanol on lipid metabolism and epithelial permeability barrier of skin and oral mucosa in the rat. *Journal of Oral Pathology & Medicine,* 32**,** 595-599.

St John, M. A. R., Li, Y., Zhou, X. F., Denny, P., Ho, C. M., Montemagno, C., Shi, W. Y., Qi, F. X., Wu, B., Sinha, U., Jordan, R., Wolinsky, L., Park, N. H., Liu, H. H., Abemayor, E. & Wong, D. T. W. 2004. Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. *Archives of Otolaryngology-Head & Neck Surgery,* 130**,** 929-935.

Stanwell, P., Gluch, L., Clark, D., Tomanek, B., Baker, L., Giuffre, B., Lean, C., Malycha, P. & Mountford, C. 2005. Specificity of choline metabolites for in vivo diagnosis of breast cancer using H-1 MRS at 1.5 T. *European Radiology,* 15**,** 1037-1043.

Star-Lack, J. M., Adalsteinsson, E., Adam, M. F., Terris, D. J., Pinto, H. A., Brown, J. M. & Spielman, D. M. 2000. In vivo H-1 MR spectroscopy of human head and neck lymph node metastasis and comparison with oxygen tension measurements. *American Journal of Neuroradiology,* 21**,** 183-193.

Stocklin, E., Wissler, M., Gouilleux, F. & Groner, B. 1996. Functional interactions between Stat5 and the glucocorticoid receptor. *Nature,* 383**,** 726-728.

Strauss, K. I. & Marini, A. M. 2002. Cyclooxygenase-2 inhibition protects cultured cerebellar granule neurons from glutamate-mediated cell death. *Journal of Neurotrauma,* 19**,** 627-638.

Stubbs, M., Robinson, S. P., Hui, C., Price, N. M., Rodrigues, L. M., Howe, F. A. & Griffiths, J. R. 2002. The importance of tumor metabolism in cancer prognosis and therapy; pre-clinical studies on rodent tumors with agents that improve tumor oxygenation. *Advances in Enzyme Regulation, Vol 42, Proceedings,* 42**,** 131-141.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 15545-15550.

Sudbo, J. & Reith, A. 2003. Which putatively pre-malignant oral lesions become oral cancers? Clinical relevance of early targeting of high-risk individuals. *Journal of Oral Pathology & Medicine,* 32**,** 63-70.

Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T. & Tomita, M. 2010. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics,* 6**,** 78-95.

Suhr, M. L., Dysvik, B., Bruland, O., Warnakulasuriya, S., Amaratunga, A. N., Jonassen, I., Vasstrand, E. N. & Ibrahim, S. O. 2007. Gene expression profile of oral squamous cell carcinomas from Sri Lankan betel quid users. *Oncology Reports,* 18**,** 1061-1075.

Sumino, J., Uzawa, N., Okada, N., Miyaguchi, K., Mogushi, K., Takahashi, K.-I., Sato, H., Michikawa, C., Nakata, Y., Tanaka, H. & Amagasa, T. 2013. Gene expression changes in initiation and progression of oral squamous cell carcinomas revealed by laser microdissection and oligonucleotide microarray analysis. *International Journal of Cancer,* 132**,** 540-548.

Sunada, H., Magun, B. E., Mendelsohn, J. & Macleod, C. L. 1986. Monoclonal-Antibody against Epidermal Growth-Factor Receptor Is Internalized without Stimulating Receptor Phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 3825-3829.

Suswam, E. A., Nabors, L. B., Huang, Y. Y., Yang, X. H. & King, P. H. 2005. IL-1 beta induces stabilization of IL-8 mRNA in malignant breast cancer cells via the 3 ' untranslated region: Involvement of divergent RNA-binding factors HuR, KSRP and TIAR. *International Journal of Cancer,* 113**,** 911-919.

Sweeny, L., Liu, Z., Bush, B. D., Hartman, Y., Zhou, T. & Rosenthal, E. L. 2012. CD147 and AGR2 expression promote cellular proliferation and metastasis of head and neck squamous cell carcinoma. *Experimental Cell Research,* 318**,** 1788-1798.

Syrjanen, S. 2005. Human papillomavirus (HPV) in head and neck cancer. *Journal of Clinical Virology,* 32**,** S59-S66.

Szabo, G., Chavan, S., Mandrekar, P. & Catalano, D. 1999. Acute alcohol consumption attenuates interleukin-8 (IL-8) and monocyte chemoattractant peptide-1 (MCP-1) induction in response to ex vivo stimulation: A potential role for nuclear regulatory factor-kappa B inhibition G. *Gastroenterology,* 116**,** A828-A829.

Szentgyorgyi, A. 1977. Living State and Cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 2844-2847.

Tabuenca, A., Mohan, S., Garberoglio, C. A., Borgen, P. I., Rosol, T. & Linkhart, T. A. 1995. Parathyroid hormone-related protein: primary osteolytic factor produced by breast tumor cells in vitro? *World J Surg,* 19**,** 292-7; discussion 297-8.

Taieb, J., Delarche, C., Ethuin, F., Selloum, S., Poynard, T., Gougerot-Pocidalo, M. A. & Chollet-Martin, S. 2002. Ethanol-induced inhibition of cytokine release and protein degranulation in human neutrophils. *Journal of Leukocyte Biology,* 72**,** 1142-1147.

Tajima, S., Oshikawa, K., Tominaga, S. & Sugiyama, Y. 2003. The increase in serum soluble ST2 protein upon acute exacerbation of idiopathic pulmonary fibrosis. *Chest,* 124**,** 1206-1214.

Takanami, I. 2005. Increased expression of integrin-linked kinase is associated with shorter survival in non-small cell lung cancer. *Bmc Cancer,* 5.

Tan, I., Yong, J., Dong, J. M., Lim, L. & Leung, T. 2008a. A tripartite complex containing MRCK modulates lamellar actomyosin retrograde flow. *Cell,* 135**,** 123-136.

Tan, W., Sabet, L., Li, Y., Yu, T., Klokkevold, P. R., Wong, D. T. & Ho, C. M. 2008b. Optical protein sensor for detecting cancer markers in saliva. *Biosensors & Bioelectronics,* 24**,** 266-271.

Thomas, G. R. & Wen, J. 2006. Endogenous expression of CD80 co-stimulatory molecule facilitates in vivo tumor regression of oral squamous carcinoma. *Anticancer Research,* 26**,** 4093-4101.

Thomas, G. T., Lewis, M. P. & Speight, P. M. 1999. Matrix metalloproteinases and oral cancer. *Oral oncology,* 35**,** 227-233.

Tian, Y.-l., Liu, F.-r., Liu, J., Jiang, L., Luo, Y. & Zhang, X. 2002. [Ectopic expression of cyclin G2 inhibits cell proliferation in HeLa cancer cell line]. *Ai zheng = Aizheng = Chinese journal of cancer,* 21**,** 577-81.

Tiziani, S., Lopes, V. & Gunther, U. L. 2009. Early Stage Diagnosis of Oral Cancer Using H-1 NMR-Based Metabolomics. *Neoplasia,* 11**,** 269-U69.

Tomasi, M. L., Ramani, K., Lopitz-Otsoa, F., Rodríguez, M. S., Li, T. W. H., Ko, K., Yang, H., Bardag-Gorce, F., Iglesias-Ara, A., Feo, F., Pascale, M. R., Mato, J. M. & Lu, S. C. 2010. S-adenosylmethionine regulates dual-specificity mitogen-activated protein kinase phosphatase expression in mouse and human hepatocytes. *Hepatology,* 51**,** 2152-2161.

Tomioka, H., Morita, K., Hasegawa, S. & Omura, K. 2006. Gene expression analysis by cDNA microarray in oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine,* 35**,** 206-211.

Toruner, G. A., Ulger, C., Alkan, M., Galante, A. T., Rinaggio, J., Wilk, R., Tian, B., Soteropoulos, P., Hameed, M. R., Schwalb, M. N. & Dermody, J. J. 2004. Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer Genetics and Cytogenetics,* 154**,** 27-35.

Trusolino, L. & Comoglio, P. M. 2002. Scatter-factor and semaphorin receptors: Cell signalling for invasive growth. *Nature Reviews Cancer,* 2**,** 289-300.

Tsai, W. C., Tsai, S. T., Ko, J. Y., Jin, Y. T., Li, C., Huang, W. Y., Young, K. C., Lai, M. D., Liu, H. S. & Wu, L. W. 2004. The mRNA profile of genes in betel quid chewing oral cancer patients. *Oral Oncology,* 40**,** 418-426.

Tsang, T. M., Huang, J. T. J., Holmes, E. & Bahn, S. 2006. Metabolic profiling of plasma from discordant schizophrenia twins: Correlation between lipid signals and global functioning in female schizophrenia patients. *Journal of Proteome Research,* 5**,** 756-760.

Tseng, G. C., Ghosh, D. & Feingold, E. 2012. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Research,* 40**,** 3785-3799.

Turnbull, A. K., Kitchen, R. R., Larionov, A. A., Renshaw, L., Dixon, J. M. & Sims, A. H. 2012. Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. *Bmc Medical Genomics,* 5.

Tusher, V. G., Tibshirani, R. & Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response (vol 98, pg 5116, 2001). *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 10515-10515.

Uemura, S., Matsushita, H., Li, W., Glassford, A. J., Asagami, T., Lee, K. H., Harrison, D. G. & Tsao, P. S. 2001. Diabetes mellitus enhances vascular matrix metalloproteinase activity - Role of oxidative stress. *Circulation Research,* 88**,** 1291-1298.

Ujpal, M., Barabas, J., Kovalszky, I., Szabo, G., Nemeth, Z., Gabris, K. & Suba, Z. S. 2007. A preliminary comparative study of the prognostic implications of type 2 diabetes mellitus for patients with primary gingival carcinoma treated with surgery and radiation therapy. *Journal of Oral and Maxillofacial Surgery,* 65**,** 452-456.

Umar, A. 2004. Applications of Bioinformatics in Cancer Detection: A Lexicon of Bioinformatics Terms. *Annals of the New York Academy of Sciences,* 1020**,** 263-276.

Vairaktaris, E., Yapijakis, C., Serefoglou, Z., Avgoustidis, D., Critselis, E., Spyridonidou, S., Vylliotis, A., Derka, S., Vassiliou, S., Nkenke, E. & Patsouris, E. 2008. Gene expression polymorphisms of interleukins-1 beta, -4, -6, -8, -10, and tumor necrosis factors-alpha, -beta: regression analysis of their effect upon oral squamous cell carcinoma. *Journal of Cancer Research and Clinical Oncology,* 134**,** 821-832.

van Riggelen, J., Yetil, A. & Felsher, D. W. 2010. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature Reviews Cancer,* 10**,** 301-309.

van Zandwijk, N., Jassem, E., Bonfrer, J. M., Mooi, W. J. & van Tinteren, H. 1992. Serum neuron-specific enolase and lactate dehydrogenase as predictors of response to chemotherapy and survival in non-small cell lung cancer. *Seminars in oncology,* 19**,** 37-43.

Vandepol, M., Twijnstra, A., Tenvelde, G. P. M. & Menheere, P. 1994. Neuron-Specific Enolase as a Marker of Brain Metastasis in Patients with Small-Cell Lung-Carcinoma. *Journal of Neuro-Oncology,* 19**,** 149-154.

Vandercappellen, J., Van Damme, J. & Struyf, S. 2008. The role of CXC chemokines and their receptors in cancer. *Cancer Letters,* 267**,** 226-244.

Vandeventer, S. J. H., Hart, M., Vanderpoll, T., Hack, C. E. & Aarden, L. A. 1993. Endotoxin and Tumor Necrosis Factor-Alpha-Induced Interleukin-8 Release in Humans. *Journal of Infectious Diseases,* 167**,** 461-464.

Vansteenkiste, J., Zielinksi, M., Linder, A., Dahabre, J., Esteban, E., Malinowski, W., Jassem, J., Passlick, B., Lehmann, F. & Brichard, V. G. 2007a. Activity of MAGE-A3 cancer immunotherapeutic as adjuvant therapy in stage IB/II non-small cell lung cancer (NSCLC): final results of a multi-center, double-blind, randomized, placebo-controlled phase II study. *Journal of Thoracic Oncology,* 2**,** S334-S335.

Vansteenkiste, J., Zielinski, M., Linder, A., Dahabre, J., Esteban, E., Malinowski, W., Jassem, J., Passlick, B., Lopez-Brea, M. & Vanakesa, T. 2007b. Adjuvant therapy in stage IB/II non-small cell lung cancer (NSCLC): final results of a multi-center, double-blind, randomized, placebocontrolled Phase II study evaluating the MAGE-A3 cancer immunotherapeutic. *Ejc Supplements,* 5**,** 361-361.

Vayssiere, J. L., Petit, P. X., Risler, Y. & Mignotte, B. 1994. Commitment to Apoptosis Is Associated with Changes in Mitochondrial Biogenesis and Activity in Cell-Lines Conditionally Immortalized with Simian-Virus-40. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 11752-11756.

Velcheti, V. 2007. Hedgehog signaling is a potent regulator of angiogenesis in small cell lung cancer. *Medical Hypotheses,* 69**,** 948-949.

Vesosky, B. & Hurwitz, A. A. 2003. Modulation of costimulation to enhance tumor immunity. *Cancer Immunology Immunotherapy,* 52**,** 663-669.

Vigneswaran, N., Wu, J., Sacks, P., Gilcrease, M. & Zacharias, W. 2005. Microarray gene expression profiling of cell lines from primary and metastatic tongue

squamous cell carcinoma: possible insights from emerging technology. *Journal of Oral Pathology & Medicine,* 34**,** 77-86.

Villaret, D. B., Wang, T. T., Dillon, D., Xu, J. C., Sivam, D., Cheever, M. A. & Reed, S. G. 2000. Identification of genes overexpressed in head and neck squamous cell carcinoma using a combination of complementary DNA subtraction and microarray analysis. *Laryngoscope,* 110**,** 374-381.

Vinader, V. & Afarinkia, K. 2012. The emerging role of CXC chemokines and their receptors in cancer. *Future Medicinal Chemistry,* 4**,** 853-867.

Vora, H. H., Shah, N. G., Patel, D. D., Trivedi, T. I. & Chikhlikar, P. R. 2003. Prognostic significance of biomarkers in squamous cell carcinoma of the tongue: Multivariate analysis. *Journal of Surgical Oncology,* 82**,** 34-50.

Vulin, A. I. & Stanley, F. M. 2004. Oxidative stress activates the plasminogen activator inhibitor type 1 (PAI-1) promoter through an AP-1 response element and cooperates with insulin for additive effects on PAI-1 transcription. *Journal of Biological Chemistry,* 279**,** 25172-25178.

Walker, W. L., Liao, I. H., Gilbert, D. L., Wong, B., Pollard, K. S., McCulloch, C. E., Lit, L. & Sharp, F. R. 2008. Empirical Bayes accomodation of batch-effects in microarray data using identical replicate reference samples: application to RNA expression profiling of blood from Duchenne muscular dystrophy patients. *Bmc Genomics,* 9.

Walsh, M. C., Brennan, L., Malthouse, J. P. G., Roche, H. M. & Gibney, M. J. 2006. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *American Journal of Clinical Nutrition,* 84**,** 531-539.

Wang, W. & Ballatori, N. 1998. Endogenous glutathione conjugates: Occurrence and biological functions. *Pharmacological Reviews,* 50**,** 335-355.

Wang, Z. 2005. The role of COX-2 in oral cancer development, and chemoprevention/treatment of oral cancer by selective COX-2 inhibitors. *Current Pharmaceutical Design,* 11**,** 1771-1777.

Wang, Z., Hao, Y. & Lowe, A. W. 2008. The Adenocarcinoma-Associated Antigen, AGR2, Promotes Tumor Growth, Cell Migration, and Cellular Transformation. *Cancer Research,* 68**,** 492-497.

Warburg, O. 1956. Origin of Cancer Cells. *Science,* 123**,** 309-314.

Wardbooth, R. P., Gabriel, A. & Tanner, N. S. B. 1982. Unexpectedly Low-Frequency of Lymph-Node Spread in Oral-Cancer in Sri-Lanka. *Lancet,* 1**,** 165-165.

Warner, G. C., Reis, P. P., Jurisica, I., Sultan, M., Arora, S., MacMillan, C., Makitie, A. A., Grenman, R., Reid, N., Sukhai, M., Freeman, J., Gullane, P., Irish, J. & Kamel-Reid, S. 2004. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. *International Journal of Cancer,* 110**,** 857-868.

Warskulat, U., Zhang, F. & Haussinger, D. 1997. Taurine is an osmolyte in rat liver macrophages (Kupffer cells). *Journal of Hepatology,* 26**,** 1340-1347.

Weeraratna, A. T., Jiang, Y. A., Hostetter, G., Rosenblatt, K., Duray, P., Bittner, M. & Trent, J. M. 2002. Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma. *Cancer Cell,* 1**,** 279-288.

Wei, J., Xie, G., Zhou, Z., Shi, P., Qiu, Y., Zheng, X., Chen, T., Su, M., Zhao, A. & Jia, W. 2011. Salivary metabolite signatures of oral cancer and leukoplakia. *International Journal of Cancer,* 129**,** 2207-2217.

Weinreich, J., Löb, S., Löffler, M., Königsrainer, I., Zieker, D., Königsrainer, A., Coerper, S. & Beckert, S. 2011. Rapamycin-Induced Impaired Wound Healing Is Associated with Compromised Tissue Lactate Accumulation and Extracellular Matrix Remodeling. *European Surgical Research,* 47**,** 39-44.

Wen, J., Nikitakis, N. G., Chaisuparat, R., Greenwell-Wild, T., Gliozzi, M., Jin, W., Adli, A., Moutsopoulos, N., Wu, T., Warburton, G. & Wahl, S. M. 2011. Secretory Leukocyte Protease Inhibitor (SLPI) Expression and Tumor Invasion in Oral Squamous Cell Carcinoma. *American Journal of Pathology,* 178**,** 2866-2878.

Whelehan, O. P., Earll, M. E., Johansson, E., Toft, M. & Eriksson, L. 2006. Detection of ovarian cancer using chemometric analysis of proteomic profiles. *Chemometrics and Intelligent Laboratory Systems,* 84**,** 82-87.

Whipple, M. E., Mendez, E., Farwell, D. G., Agoff, S. N. & Chen, C. 2004. A log likelihood predictor for genomic classification of oral cancer using principle component analysis for feature selection. *Medinfo 2004: Proceedings of the 11th World Congress on Medical Informatics, Pt 1 and 2,* 107**,** 823-826.

Whitehead, T. L. & Kieber-Emmons, T. 2005. Applying in vitro NMR spectroscopy and H-1 NMR metabonomics to breast cancer characterization and detection. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 47**,** 165-174.

Who 1995. *ICD-DA: Application of the international classification of diseases to dentistry and stomatology, Third edition,* WHO {a}, Geneva, Switzerland.

Wiegand, S., Dunne, A. A., Muller, H. H., Mandic, R., Barth, P., Davis, R. K. & Werner, J. A. 2005. Metaanalysis of the significance of matrix metalloproteinases for lymph node disease in patients with head and neck squamous cell carcinoma. *Cancer,* 104**,** 94-100.

Wilkinson, S., Paterson, H. F. & Marshall, C. J. 2005. Cdc42-MRCK and Rho-ROCK signalling cooperate in myosin phosphorylation and cell invasion. *Nature Cell Biology,* 7**,** 255-U45.

Williams, H. K. 2000. Molecular pathogenesis of oral squamous carcinoma. *Journal of Clinical Pathology-Molecular Pathology,* 53**,** 165-172.

Williams, M. D. 2010. Integration of biomarkers including molecular targeted therapies in head and neck cancer. *Head and neck pathology,* 4**,** 62-9.

Wilson, C. L. & Miller, C. J. 2005. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics,* 21**,** 3683-3685.

Wogan, G. N., Hecht, S. S., Felton, J. S., Conney, A. H. & Loeb, L. A. 2004. Environmental and chemical carcinogenesis. *Seminars in Cancer Biology,* 14**,** 473-486.

Wolf, G. T., Hudson, J. L., Peterson, K. A., Miller, H. L. & McClatchey, K. D. 1986. Lymphocyte subpopulations infiltrating squamous carcinomas of the head and neck: correlations with extent of tumor and prognosis. *Otolaryngol Head Neck Surg,* 95**,** 142-52.

Wolfman, J. C., Planchon, S. M., Liao, J. & Wolfman, A. 2006. Structural and functional consequences of c-N-Ras constitutively associated with intact

mitochondria. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1763**,** 1108-1124.

Wong, D. Y., Hsiao, Y. L., Poon, C. K., Kwan, P. C., Chao, S. Y., Chou, S. T. & Yang, C. S. 1994. Glutathione concentration in oral cancer tissues. *Cancer Lett,* 81**,** 111-6.

Wong, Y. F., Cheung, T. H., Tsao, G. S. W., Lo, K. W. K., Yim, S. F., Wang, V. W., Heung, M. M. S., Chan, S. C. S., Chan, L. K. Y., Ho, T. W. F., Wong, K. Y. W., Li, C., Guo, Y., Chung, T. K. H. & Smith, D. I. 2006. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. *International Journal of Cancer,* 118**,** 2461-2469.

Woolgar, J. A. 2006. Histopathological prognosticators in oral and oropharyngeal squamous cell carcinoma. *Oral Oncology,* 42**,** 229-239.

Wu, C.-H., van Riggelen, J., Yetil, A., Fan, A. C., Bachireddy, P. & Felsher, D. W. 2007. Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 13028-13033.

Wu, C., Zhai, N., Guan, C. & Ji, W. 2006. [Expression of integrin-linked kinase is closely correlated with laryngeal squamous cell carcinomas]. *Lin chuang er bi yan hou ke za zhi = Journal of clinical otorhinolaryngology,* 20**,** 393-5.

Wu, T. D. 2001. Bioinformatics in the post-genomic era. *Trends in Biotechnology,* 19**,** 479-480.

Xiong, G., Wang, C., Evers, B. M., Zhou, B. P. & Xu, R. 2012. RORÎ± Suppresses Breast Tumor Invasion by Inducing SEMA3F Expression. *Cancer Research,* 72**,** 1728-1739.

Yabu, M., Shime, H., Hara, H., Saito, T., Matsumoto, M., Seya, T., Akazawa, T. & Inoue, N. 2011. IL-23-dependent and -independent enhancement pathways of IL-17A production by lactic acid. *International Immunology,* 23**,** 29-41.

Yamamoto, H., Itoh, F., Iku, S., Hosokawa, M. & Imai, K. 2001. Expression of the gamma(2) chain of laminin-5 at the invasive front is associated T-with recurrence and poor prognosis in human esophageal squamous cell carcinoma. *Clinical Cancer Research,* 7**,** 896-900.

Yamamoto, H., Kitadai, Y., Oue, N., Ohdan, H., Yasui, W. & Kikuchi, A. 2009. Laminin gamma 2 Mediates Wnt5a-Induced Invasion of Gastric Cancer Cells. *Gastroenterology,* 137**,** 242-252.

Yamamoto, H., Oue, N., Sato, A., Hasegawa, Y., Matsubara, A., Yasui, W. & Kikuchi, A. 2010. Wnt5a signaling is involved in the aggressiveness of prostate cancer and expression of metalloproteinase. *Oncogene,* 29**,** 2036-2046.

Yan, S.-K., Wei, B.-J., Lin, Z.-Y., Yang, Y., Zhou, Z.-T. & Zhang, W.-D. 2008. A metabonomic approach to the diagnosis of oral squamous cell carcinoma, oral lichen planus and oral leukoplakia. *Oral oncology,* 44**,** 477-483.

Yan, W. & Shao, R. 2006. Transduction of a mesenchyme-specific gene periostin into 293T cells induces cell invasive activity through epithelial-mesenchymal transformation. *Journal of Biological Chemistry,* 281**,** 19700-19708.

Yan, Y., Yang, J.-Y., Mou, Y.-H., Wang, L.-H., Zhang, H. & Wu, C.-F. 2011. Possible Metabolic Pathways of Ethanol Responsible for Oxidative DNA

Damage in Human Peripheral Lymphocytes. *Alcoholism-Clinical and Experimental Research,* 35**,** 1-9.

Yanagawa, T., Yoshida, H., Yamagata, K., Onizawa, K., Tabuchi, K., Koyama, Y., Iwasa, S., Shimoyamada, H., Harada, H. & Omura, K. 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,* 26**,** 215-220.

Yang, M.-H. & Wu, K.-J. 2008. TWIST activation by hypoxia inducible factor-1 (HIF-1). *Cell Cycle,* 7**,** 2090-2096.

Yang, M.-H., Wu, M.-Z., Chiou, S.-H., Chen, P.-M., Chang, S.-Y., Liu, C.-J., Teng, S.-C. & Wu, K.-J. 2008. Direct regulation of TWIST by HIF-1 alpha promotes metastasis. *Nature Cell Biology,* 10**,** 295-305.

Yavrouian, E. J. & Sinha, U. K. 2012. Recent advances in biomarkers and potential targeted therapies in head and neck squamous cell carcinoma. *ISRN surgery,* 2012**,** 715743-715743.

Ye, H., Yu, T., Temam, S., Ziober, B. L., Wang, J., Schwartz, J. L., Mao, L., Wong, D. T. & Zhou, X. 2008. Transcriptomic dissection of tongue squamous cell carcinoma. *BMC Genomics,* 9**,** Article No.: 69.

Ye, S., Eriksson, P., Hamsten, A., Kurkinen, M., Humphries, S. E. & Henney, A. M. 1996. Progression of coronary atherosclerosis is associated with a common genetic variant of the human stromelysin-1 promoter which results in reduced gene expression. *Journal of Biological Chemistry,* 271**,** 13055-13060.

Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. 2004. Variation in alternative splicing across human tissues. *Genome Biology,* 5.

Yeung, D. K. W., Cheung, H. S. & Tse, G. M. K. 2001. Human breast lesions: Characterization with contrast-enhanced in vivo proton MR spectroscopy - Initial results. *Radiology,* 220**,** 40-46.

Yi, M., Horton, J. D., Cohen, J. C., Hobbs, H. H. & Stephens, R. M. 2006. WholePathwayScope: a comprehensive pathway-based analysis tool for high-throughput data. *Bmc Bioinformatics,* 7.

Yin, T. & Yang, Y. C. 1994. Mitogen-activated protein kinases and ribosomal S6 protein kinases are involved in signaling pathways shared by interleukin-11, interleukin-6, leukemia inhibitory factor, and oncostatin M in mouse 3T3-L1 cells. *J Biol Chem,* 269**,** 3731-8.

Yokota, H., Guo, J. F., Matoba, M., Higashi, K., Tonami, H. & Nagao, Y. 2007. Lactate, choline, and creatine levels measured by vitro H-1-MRS as prognostic parameters in patients with non-small-cell lung cancer. *Journal of Magnetic Resonance Imaging,* 25**,** 992-999.

Yokota, S., Yamamoto, Y., Shimizu, K., Momoi, H., Kamikawa, T., Yamaoka, Y., Yanagi, H., Yura, T. & Kubota, H. 2001. Increased expression of cytosolic chaperonin CCT in human hepatocellular and colonic carcinoma. *Cell Stress & Chaperones,* 6**,** 345-350.

Yoo, M. H., Xu, X. M., Carlson, B. A., Gladyshev, V. N. & Hatfield, D. L. 2006. Thioredoxin reductase 1 deficiency reverses tumor phenotype and tumorigenicity of lung carcinoma cells. *Journal of Biological Chemistry,* 281**,** 13005-13008.

Yook, J. I., Li, X. Y., Ota, I., Fearon, E. R. & Weiss, S. J. 2005. Wnt-dependent regulation of the E-cadherin repressor snail. *Journal of Biological Chemistry,* 280**,** 11740-11748.

Yu, J., Zhang, L., Hwang, P. M., Rago, C., Kinzler, K. W. & Vogelstein, B. 1999. Identification and classification of p53-regulated genes. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 14517-14522.

Yu, L., Hebert, M. C. & Zhang, Y. E. 2002. TGF-beta receptor-activated p38 MAP kinase mediates smad-independent TGF-beta responses. *Embo Journal,* 21**,** 3749-3759.

Yuan, P., Temam, S., El-Naggar, A., Zhou, X., Liu, D. D., Lee, J. J. & Mao, L. 2006. Overexpression of podoplanin in oral cancer and its association with poor clinical outcome. *Cancer,* 107**,** 563-569.

Yuen, H.-F., Chan, Y.-P., Wong, M. L.-Y., Kwok, W.-K., Chan, K.-K., Lee, P.-Y., Srivastava, G., Law, S. Y.-K., Wong, Y.-C., Wang, X. & Chan, K.-W. 2007. Upregulation of Twist in oesophageal squamous cell carcinoma is associated with neoplastic transformation and distant metastasis. *Journal of Clinical Pathology,* 60**,** 510-514.

Zamzami, N., Susin, S. A., Marchetti, P., Hirsch, T., GomezMonterrey, I., Castedo, M. & Kroemer, G. 1996. Mitochondrial control of nuclear apoptosis. *Journal of Experimental Medicine,* 183**,** 1533-1544.

Zayzafoon, M., Botolin, S. & McCabe, L. R. 2002. p38 and activating transcription factor-2 involvement in osteoblast osmotic response to elevated extracellular glucose. *Journal of Biological Chemistry,* 277**,** 37212-37218.

Zhang, C. H., Ao, Z. H., Seth, A. & Schlossman, S. F. 1996. Mitochondrial membrane protein defined by a novel monoclonal antibody is preferentially detected in apoptotic cells. *Journal of Immunology,* 157**,** 3980-3987.

Zhang, J. & Chen, C. 2008. Endocannabinoid 2-Arachidonoylglycerol Protects Neurons by Limiting COX-2 Elevation. *Journal of Biological Chemistry,* 283**,** 22601-22611.

Zhang, L. Q. & Li, W. H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution,* 21**,** 236-239.

Zhang, S., Zhang, D., Zhu, Y., Guo, H., Zhao, X. & Sun, B. 2006. Clusterin expression and univariate analysis of overall survival in human breast cancer. *Technology in Cancer Research & Treatment,* 5**,** 573-578.

Zhang, X., Liu, Y. N., Gilcrease, M. Z., Yuan, X. H., Clayman, G. L., Adler-Storthz, K. & Chen, Z. 2002. A lymph node metastatic mouse model reveals alterations of metastasis-related gene expression in metastatic human oral carcinoma sublines selected from a poorly metastatic parental cell line. *Cancer,* 95**,** 1663-1672.

Zhang, Y., Ali, T. Z., Zhou, H., D'Souza, D. R., Lu, Y., Jaffe, J., Liu, Z., Passaniti, A. & Hamburger, A. W. 2010. ErbB3 Binding Protein 1 Represses Metastasis-Promoting Gene Anterior Gradient Protein 2 in Prostate Cancer. *Cancer Research,* 70**,** 240-248.

Zhang, Z. X., Jones, S., Hagood, J. S., Fuentes, N. L. & Fuller, G. M. 1997. STAT3 acts as a co-activator of glucocorticoid receptor signaling. *Journal of Biological Chemistry,* 272**,** 30607-30610.

Zhao, J., Su, Y., Zhang, Y., Pan, Z., Yang, L., Chen, X., Liu, Y., Lu, Y., Du, Z. & Yang, B. 2010. Activation of cardiac muscarinic M3 receptors induces delayed cardioprotection by preserving phosphorylated connexin43 and up-regulating cyclooxygenase-2 expression. *British Journal of Pharmacology,* 159**,** 1217-1225.

Zhao, M., Begum, S., Ha, P. K., Westra, W. & Califano, J. 2008. Downregulation of RAD17 in head and neck cancer. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 30**,** 35-42.

Zhao, R. B., Gish, K., Murphy, M., Yin, Y. X., Notterman, D., Hoffman, W. H., Tom, E., Mack, D. H. & Levine, A. J. 2000. Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes & Development,* 14**,** 981-993.

Zhi, H. Y., Zhang, J., Hu, G. X., Lu, J. Y., Wang, X. Q., Zhou, C. N., Wu, M. & Liu, Z. H. 2003. The deregulation of arachidonic acid metabolism-related genes in human esophageal squamous cell carcinoma. *International Journal of Cancer,* 106**,** 327-333.

Zhong, L. P., Yang, X., Zhang, L., Wei, K. J., Pan, H. Y., Zhou, X. J., Li, J., Chen, W. T. & Zhang, Z. Y. 2008. Overexpression of insulin-like growth factor binding protein 3 in oral squamous cell carcinoma. *Oncology Reports,* 20**,** 1441-1447.

Zhou, J. L., Xu, B., Huang, J., Jia, X. M., Xue, J., Shi, X. C., Xiao, L. Y. & Li, W. 2009. H-1 NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma. *Clinica Chimica Acta,* 401**,** 8-13.

Zhou, S., Kachhap, S. & Singh, K. K. 2003. Mitochondrial impairment in p53-deficient human cancer cells. *Mutagenesis,* 18**,** 287-292.

Zhou, S., Kachhap, S., Sun, W., Wu, G., Chuang, A., Poeta, L., Grumbine, L., Mithani, S. K., Chatterjee, A., Koch, W., Westra, W. H., Maitra, A., Glazer, C., Carducci, M., Sidransky, D., McFate, T., Verma, A. & Califano, J. A. 2007. Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7540-7545.

Zhu, Q., Han, X., Peng, J., Qin, H. & Wang, Y. 2012. The role of CXC chemokines and their receptors in the progression and treatment of tumors. *Journal of Molecular Histology,* 43**,** 699-713.

Zimmermann, K. C., Sarbia, M., Weber, A. A., Borchard, F., Gabbert, H. E. & Schror, K. 1999. Cyclooxygenase-2 expression in human esophageal carcinoma. *Cancer Research,* 59**,** 198-204.

Ziober, A. F., Patel, K. R., Alawi, F., Gimotty, P., Weber, R. S., Feldman, M. M., Chalian, A. A., Weinstein, G. S., Hunt, J. & Ziober, B. L. 2006. Identification of a gene signature for rapid screening of oral spuamous cell carcinoma. *Clinical Cancer Research,* 12**,** 5960-5971.

Zu, X. L. & Guppy, M. 2004. Cancer metabolism: facts, fantasy, and fiction. *Biochemical and Biophysical Research Communications,* 313**,** 459-465.

Zufferey, L., Roques, B., Flechon, A. & Droz, J. P. 2001. Study of the hormone-refractory prostate cancer clinical practice in an anti-cancer center. *Bulletin Du Cancer,* 88**,** 1213-1221.