

https://dspace.org/

Molecular profiling of oesophageal squamous cell carcinomas in the South African population

Brown, Jacqueline

2012-03-08
http://hdl.handle.net/10539/11405

Downloaded from a DSpace repository. https://dspace.org/

MOLECULAR PROFILING OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS IN THE SOUTH AFRICAN POPULATION

Jacqueline Brown

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, November 2011

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before any degree or examination in any other university.

_____

_____Day of_____20_____

DEDICATION

In loving memory of my loved ones who suffered from cancer, Eileen Longmore, Ian Brown and Myrna Carey.

I also dedicate this work to all those who have suffered from oesophageal cancer.

PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS STUDY

Publication

1. **J Brown**, H Bothma, R Veale, P Willem. 2011. Genomic imbalances in esophageal carcinoma cell lines involve Wnt pathway genes. *World journal of Gastroentrology*, 17(24), pp. 2909-2923.

Conference and poster presentations

1. **Brown J**, Veale R, Willem P. Molecular Characterisation of Five South African Oesophageal Squamous Cancer Cell Lines. 6th International Cancer Conference of the African Organisation for Research and Training in Cancer. 24-28 October 2007, Cape Town.
2. **J Brown,** R Veale, P Willem. Detection of copy number changes in five South African oesophageal squamous cancer cell lines. CANSA in action congress, Emperors Palace, Johannesburg, August 2008.
3. **J Brown**, R Veale, A Stepien, A Kafuko, P Willem. Oesophageal Cancer in South Africa: Gene Copy Number Analysis uncovers some candidate genes. The 50th Annual conference of the federation of South African societies of pathology (Pathvine). 2-5 September 2010, Somerset West.

ABSTRACT

Oesophageal squamous cell carcinoma (OSCC) has a high prevalence in the Asiatic belt and areas of Africa. In South Africa (SA), the incidence of this cancer in the Eastern Cape is one of the highest in the world. The molecular carcinogenesis of this disease remains unresolved. Single nucleotide polymorphism (SNP) array technology provides a high resolution technique to determine DNA copy number imbalances across the whole genome. DNA copy number changes can affect oncogenes and tumour suppressor genes, contributing to carcinogenesis. The aim of this study was to map common chromosomal break points previously identified in five SA OSCC cell lines by multi colour fluorescence *in situ* hybridisation (FISH) and to characterise copy number changes in these cell lines and OSCC patient's specimens using SNP array technology. Genome wide copy number analysis was performed on the cell lines and 51 OSCC retrospective samples from the Eastern Cape region using Affymetrix® 500K SNP arrays. A number of genes were significantly affected by copy number changes across specimens. The copy number status of some of these candidate genes identified by arrays, were verified by (FISH) in a subset of the samples. Expression of the *EPHA3, FGF3, FGF4, FGF19* and *C-MYC* candidate genes was assessed in the cell lines and four fresh samples. The common translocation break point previously detected in 5 cell lines involving chromosome 3p11.2 correlated with deletions affecting the *EPHA3* gene in 4 of the 5 cell lines and was deleted in 74% of the OSCC cohort. *EPHA3* is an ephrin A3 receptor tyrosine kinase that has been shown to have both oncogenic and tumour suppressor functionality. In addition, significant regions of amplification and deletion identified genes (*CCND1, C-MYC, FHIT, SFRP1, SFRP2, FGF3, FGF4, FGF19, SMAD4, SMAD6* and *FBXW7*) involved in the Wnt, TGF-β and FGF. Deletion of the genes, *WRN*, *ATM*, *RAD18*

and *XRCC4* involved in DNA repair pathways, may contribute to genetic instability that is characteristic of OSCC. This study has highlighted some molecular pathways that may contribute to better understanding carcinogenesis of OSCC in South Africa.

ACKNOWLEDGMENTS

I would like to say a huge thank you to Dr Pascale Willem who has given me encouraging support all through the years for not only this project but others and always helped me to see the light at the end of the tunnel. In addition to her inspiring passion for our subject and the intellectual support, I greatly appreciate her friendship. Thanks must go to my honours student for 2010, Kealeboga Mokolobate, for assisting in production of some of the FISH probes in this research. Thank you to my co-supervisor, Professor Rob Veale, for his support and providing the cell lines. I would also like to thank Elsabe Scott for preparing the cell lines for me. Thanks very much to Professor Andrej Stepien from NHLS at Nelson Mandela Academic Hospital in Umtata for providing the retrospective samples and taking the time to ring the tumour areas for me.

Thanks to the funding bodies, the Cancer Association of South Africa, the Medical Research Council of South Africa and the National Health Laboratory Services. Thanks also go to the University of the Witwatersrand for financial assistance for tuition fees and the project.

Then last but not least, thank you to my wonderful mom, dad, brother, Imogen and Greig who always reminded me of my capabilities when I felt I couldn't do it and their love and support have kept me going.

ETHICS

Ethics approval was obtained for this project from the University of the Witwatersrand Human Research Ethics committee. The certificate is attached at the back of this thesis.

Clearance certificate: **M090658**

Title: MOLECULAR PROFILING OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS IN THE SOUTH AFRICAN POPULATION

CONTENTS

DECLARATION ii

DEDICATION Iii

PUBLICATIONS AND PRESENTATIONS iv

ABSTRACT v

ACKNOWLEDGEMENTS vii

ETHICS viii

TABLE OF CONTENTS ix

LIST OF FIGURES xv

LIST OF TABLES xviii

LIST OF ABBREVIATIONS xix

CHAPTER 1-INTRODUCTION 1

1.1 Oesophageal Squamous Carcinoma: Epidemiology and risk factors 2

Overview 2

Maize consumption and OSCC 4

Smoking, alcohol consumption and OSCC 5

Human Papillomavirus and OSCC 7

Chronic inflammation and OSCC 8

1.2 Molecular biology 9

Adhesion molecules 10

Repair genes 15

Chemokines 16

Genetic polymorphisms associated with risk of OSCC development 17

Oncogenes and tumour suppressor genes implicated in OSCC 21

1.3 Chromosomal aberrations and OSCC 34

Mechanisms of chromosomal aberrations 34

Chromosomal aberrations detected in OSCC 38

1.4 Project rationale 42

Hypotheses 42

Aim 42

Specific objectives 43

CHAPTER 2-Characterisation of common chromosomal aberrations in 5 OSCC cell lines 44

2.1 Introduction 45

2.2 Materials and methods 46

Cell lines 46

Isolation of cell lines for metaphase preparation 47

Isolation of cell lines for DNA 47

Conventional cytogenetics 48

Preparation metaphase slides for Fluorescence *in situ* Hybridisation (FISH) 48

FISH procedures 48

BAC probe development and FISH analysis for *EPHA3* gene deletions **49**

DNA isolation **54**

Assessment of DNA integrity **55**

DNA quantification **55**

Sample preparation **55**

Affymetrix 250K genotyping assay for the assessment of DNA copy number changes in the five cell lines **55**

250K data analysis **56**

2.3 Results **57**

Cytogenetics and M-FISH **57**

Copy number changes **70**

Significant gains **71**

Significant losses **73**

2.4 Discussion **80**

CHAPTER 3- Assessment of expression levels of selected candidate genes in the 5 OSCC cell lines and fresh tissue samples **88**

3.1 Introduction **89**

3.2 Materials and methods **90**

RNA extraction from cell lines **90**

Fresh OSCC tumour samples **91**

RNA extraction from tissue samples **91**

Target genes and primer design for RQ-PCR **92**

DNase treatment of RNA **92**

cDNA synthesis **93**

RQ-PCR experiments **93**

Data analysis **94**

3.3 Results **95**

Expression levels of the candidate genes, *EPHA3, CCND1, FGF19, FGF3, FGF4* and *C-MYC* **95**

EPHA3 gene expression **98**

CCND1 expression **99**

FGF3, FGF4 and *FGF19* expression **100**

C-MYC expression **103**

3.4 Discussion **104**

Expression of the *EPHA3* gene in the five cell lines **104**

Expression of the *CCND1, FGF3, FGF4, FGF19* and *C-MYC* candidate genes **105**

CHAPTER 4- Characterisation of copy number profiles in an OSCC cohort from the Eastern Cape **109**

4.1 Introduction **110**

Copy number studies in OSCC **110**

Challenges of OSCC pathology sampling **112**

4.2 Materials and methods **114**

Oesophageal cancer cohort **114**

Control samples **114**

DNA isolation **115**

Assessment of DNA integrity **116**

DNA quantification **116**

Sample preparation **116**

Multiplex PCR to assess PCR performance of DNA from FFPE samples **117**

Affymetrix 500K genotyping assay for copy number analysis of the OSCC cohort **118**

500K data analysis **119**

FISH for confirmation of array results on the OSCC samples **122**

4.3 Results **123**

Demographics of the OSCC patient cohort from Eastern Cape **123**

DNA assessment and copy number analysis of the OSCC FFPE cohort from the Eastern Cape **123**

Copy number gains **124**

Copy number losses **142**

Functional associations of candidate genes in amplified and deleted regions **155**

4.4 Discussion **161**

4.4.1 OSCC demographics in SA **161**

4.4.2 SNP arrays as a method for copy number analysis for FFPE samples **161**

4.4.3 Similarities between the five cell lines and the Eastern Cape cohort **163**

4.4.4 Comparison of the present study findings to published literature on copy number changes in OSCC **164**

4.4.5 *EPHA3* as a candidate gene in OSCC **169**

Structure and function of EphA receptors **169**

Role of Epha/ephrin signalling in cancer **171**

4.4.6 Common copy number aberrations and potential affected pathways **173**

Chromosome 11q amplification **174**

Chromosome 19 amplification **175**

Amplifications and deletions affecting the TGF-β, Hedgehog and Wnt signalling pathways **176**

Genes with aberrant copy number involved in genomic stability **181**

Potential markers for prognosis, diagnosis and treatment **182**

4.4.7 Summary **183**

CHAPTER 5- Final conclusions and future directions **185**

5.1 Conclusions **186**

5.2 Future directions **188**

REFERENCES **190**

APPENDIX A **220**

APPENDIX B **228**

APPENDIX C **249**

LIST OF FIGURES

Figure 1.1. Diagrammatic representation of the BFB cycle adapted from Lo et al (2002). **37**

Figure 2.1 Schematic representations of the *EPHA3* gene locus and the BAC clones selected for FISH experiments. **53**

Figure 2.2 Normal metaphase and interphase FISH for RP11-598N21 and RP11-784B9. **53**

Figure 2.3 Normal metaphase and interphase FISH for RP11535I05 and RP11-547K2. **54**

Figure 2.4 G-banded and M-FISH representative karyotypes of cell line WHCO1 **63**

Figure 2.5 M-FISH representative karyotypes showing complex rearrangements in four cell lines. **64**

Figure 2.6 Partial G-banded karyotypes of the five cell lines showing shared chromosomal aberrations across cell lines. **65**

Figure 2.7 DAPI stained metaphasic chromosomes showing the der(3)t(1;3)(p11-12;q11) in cell lines SNO and WHCO5. **66**

Figure 2.8 CNAT plot of SNP hybridization results showing the map of the region of deletion on chromosome 3p11.2-12.1 in cell lines WHCO1, WHCO3, WHCO5, WHCO6 and SNO. **67**

Figure 2.9 Representative images of FISH results for RP11-589N21 (green) and RP11-784B9 (red). **68**

Figure 2.10 Representative interphase images for each cell line for RP11-535I05 (red) and RP11-547K2 (green). **69**

Figure 2.11 DAPI stained metaphase from cell line SNO showing FISH results with the locus specific *EGFR* probe (red) and Cep7 alpha (green). **70**

Figure 2.12 Plots of recurrent genomic amplifications (A) and deletions (B) detected in the OSCC cell lines from GISTIC analysis of SNP array data. **79**

Figure 2.13 DAPI stained interphase nuclei hybridised with locus specific probes for C-*MYC* Spectrum Orange (red) (A-D) and *CCND1* (red)/IGH (green) (E-H). **80**

Figure 3.1 Relative expression levels of the *EPHA3* gene as measured by RQ-PCR. **99**

Figure 3.2 Relative expression levels of the *CCND1* gene as measured by RQ-PCR. **100**

Figure 3.3 Relative expression levels of the *FGF3* gene as measured y RQ-PCR. **101**

Figure 3.4 Relative expression levels of the *FGF3* gene as measured by RQ-PCR for sample WITSCA9. **102**

Figure 3.5 Relative expression levels of the *FGF4* gene as measured by RQ-PCR. **102**

Figure 3.6 Relative expression levels of the *FGF19* gene as measured by RQ-PCR. **103**

Figure 3.7 Relative expression levels of the C-*MYC* gene as measured by RQ-PCR. **104**

Figure 4.1 Example of the results for a multiplex PCR on 8 FFPE samples. **118**

Figure 4.2 Chromosomal mapping of regions of copy number alteration across all 51 samples. **133**

Figure 4.3 Graph representing the regions of high amplification (> 10 copies) with the number of samples affected and the average copy number. **136**

Figure 4.4 Graph representing the regions of moderate amplification (>5<10 copies) with the number of samples affected and average copy numbers for that region. **136**

Figure 4.5 Chromosome 11 mapping of copy number changes detected in one representative FFPE sample by 500K SNP arrays **139**

Figure 4.6 Selected images of DAPI stained nuclei for each of the 10 tumours chosen for FISH analysis with the Vysis LSI t(11;14) dual colour probe **140**

Figure 4.7 Selected images of DAPI stained nuclei for each of the 10 tumours chosen for FISH analysis with the Vysis LSI MYC probe. **141**

Figure 4.8 Graph showing the focal regions of deletion and the copy number for that region. **148**

Figure 4.9 Graph showing the regions of hemizygous deletion and the number of samples with that deletion. **149**

Figure 4.10 Representative images from 6 cases selected for FISH analysis of *EPHA3* gene deletions. **153**

Figure 4.11 Pie chart showing the percentage of the 20 candidate genes involved in

each biological process listed. **156**

Figure 4.12 Pie chart showing the percentage of deleted genes associated with the listed biological pathways. **157**

Figure 4.13 Predicted functional interactions for the selected candidate genes from amplified and deleted regions using STRING 8.3 (http://string.embl.de) accessed July 2010. **158**

Figure 4.14 Predicted functional interactions for the amplified genes as determined by STRING 8.3. **159**

Figure 4.15 Predicted functional interactions for the deleted genes as determined by STRING 8.3. **160**

Figure 4.16 Schematic representation of Eph/ephrin signalling **171**

Figure 4.17 Schematic representation of TGF-β signalling. **178**

Figure 4.18 Schematic representations of how the Wnt, FGF and Hedgehog pathways can interact. **180**

LIST OF TABLES

Table 1.1 Summary of studies done on molecular markers in OSCC. **29**

Table 2.1 Cell lines and their patho-clinical information. **46**

Table 2.2 Composite karyotypes. **61**

Table 2.3 FISH results for the BAC probes RP11-784B9 (red) and RP11-598N21 green). **62**

Table 2.4 FISH results for the BAC probes RP11-535I05 (red) and RP11-547K2 (green). **62**

Table 2.5 Amplification peaks as detected by the GISTIC algorithm. **76**

Table 2.6 Deletion peaks as detected by the GISTIC algorithm. **77**

Table 2.7 FISH for detection of *CCND1* and *C-MYC* amplification. **78**

Table 3.1 Reference genes and target genes for RQ-PCR analysis **92**

Table 3.2 Table of samples used in expression analysis by RQ-PCR. **97**

Table 4.1 Summary of studies on copy number changes in OSCC. **111**

Table 4.2 Multiplex primers and final concentrations used for PCR. **117**

Table 4.3 Regions affected by high amplification in a minimum of 10 samples and an average copy number larger than 10 copies. **134**

Table 4.4 Regions of moderate amplification detected in minimum of 10 samples (copy number >5<10). **135**

Table 4.5 FISH results for detection of *CCND1/FGF4* gene amplification in 10 FFPE samples. **137**

Table 4.6 FISH results for detection of c-*MYC* gene amplification in 10 FFPE samples. **138**

Table 4.7 Regions of hemizygous deletion (<1.3 copies) detected in a minimum of 10 samples. **150**

Table 4.8 FISH results for analysis of *EPHA3* gene deletions using the BAC clones RP11-598N21 (red) and RP11-547K2 (green). **152**

Table 4.9 Biological pathways and the amplified genes associated with these. **156**

Table 4.10 Biological pathways and the deleted genes associated with each pathway. **157**

LIST OF ABBREVIATIONS

5FU	Fluorouracil
5-methyl THF	5-methyltetrahydrofolate
A1BG	alpha-IB-glycoprotein precursor gene
A1BG	alpha-IB-glycoprotein precursor
ABTB1	ankyrin repeat and BTB (POZ) domain containing 1
ADAMTS1	thrombospondin type 1 motif1
ADAP1	ArfGAP with dual PH domains 1
ADH	Alcohol dehydrogenases
ADH2	Alcohol dehydrogenase 2
ALDH	aldehyde dehydrogenase
ALKBH8	alkB, alkylation repair homolog 8 gene
ANNOVA	analysis of variance
APC	Adenomatosis Polyposis Coli
ATM	ataxia telangiectasia mutated gene
BACs	Bacterial Artificial Chromosomes
BAD	bcl-xl/bcl2-antagonist causing cell death
BCL6	B-cell CLL/lymphoma 6 gene).
BCL8	B-cell CLL/lymphoma 8 gene
BFB	Breakage-fusion-bridge
BIND	biomolecular interaction network database
BMPR1B	Bone morphogenetic protein receptor 1B
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like
CAMs	cell adhesion molecules
CASP3	Caspase 3 gene
CASP6	Caspase 6
CCND1	Cyclin D1
CDKN2A	Cyclin dependent kinase 2A (also p16)
CDKN2B	cyclin dependent kinase 2B
cDNA	complimentary DNA
CGGBP1	CGG triplet repeat binding protein
CGH	comparative genomic hybridisation
CLPTM1L	cleft lip and palate transmembrane protein 1-like protein gene
C-*MYC*	v-myc myelocytomatosis viral oncogene homolog
CNAT	copy number analysis tool
CNRQ	calibrated normalised relative quantities
COX-2	Cyclooxygenase-2
CREB5	cAMP responsive element binding protein 5
CTH	cystathionine γ-lyase
CTNND2	Delta catenin

CTTN	cortactin
CYP2E1	cytochrome p450 2E21
DAPI	4', 6-diamino-2-phenylindole
DCC	deleted in colorectal cancer
DCK	deoxycytidine kinase
DIP	the database of interacting proteins
DIRAS3	DIRAS family, GTP-binding RAS-like 3
DNA	deoxyribonucleic acid
DSB	double-stranded breaks
dTMP	deoxythymidine monophosphate
dUMP	deoxyuridine monophosphate
EGFR	epidermal growth factor receptor
EGR-1	early growth response-1
EPs	G-protein coupled receptors
ERK	extracellular signal regulated kinase
ESRRA	estrogen related receptor alpha gene
EYA2	eyes absent homologue 2
FAM84B	family with sequence similarity 84, member B
FBXW7	F-box and WD repeat domain containing 7 gene
FFPE	formalin fixed paraffin embedded
FGF	fibroblast growth factor
FHIT	fragile histidine triad
FISH	fluorescence in situ hybridisation
GADD45alpha	growth arrest and DNA damage-45 alpha
GISTIC	Genomic Identification of Significant Targets in Cancer
GLAD	Gain and Loss analysis of DNA
GPER	G protein-coupled estrogen receptor 1 gene
GROα	Growth-related oncogene alpha
GTW	G-bands by trypsin and wright's stain
HHAT	Hedgehog acetyltransferase gene
HIPK2	homeodomain-interacting protein kinase-2
HMGB2	high-mobility group box 2 gene
hMLH1	mutL homolog 1
HPRD	Human protein reference database
HPV	human papillomavirus
HR	homologous recombination
HUNK	hormonally upregulated neu-associated kinase
KCl	potassium chloride
KCNMA1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1
KLF2	Kruppel-like factor 2 gene
LITAF	lipopolysaccharide-induced TNF factor gene

LOH	Loss of heterozygosity
LRP16	leukaemia related protein 16
LRP1B	low density lipoprotein 1B
LZTS1	leucine zipper, putative tumour suppressor 1
MACROD1	macro domain containing 1
MCPH1	microcephalin 1 gene
M-FISH	Multicolour fluorescence *in situ* hybridisation
MINA	MYC induced nuclear antigen
MINT	molecular interaction database
MLPA	multi ligation probe analysis
MMP	matrix metalloproteinases
mRNA	messenger RNA
MSI	microsatellite instability
MSR1	macrophage scavenger receptor 1 gene
MTHFR	Methylenetrahydrofolate reductase
NAT2	N-acetyltransferase 2 [arylamine N-acetyltransferase]
NHEJ	non homologous end joining
OSCC	Oesophageal squamous cell carcinoma
PACRG	Park2 co-regulated gene
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDGFRL	platelet-derived growth factor receptor-like gene
PGE2	precursor of prostaglandin E2
PGS	Partek® Genomics Suite
PPFIA1	protein tyrosine phosphatase, receptor type, polypeptide, interacting protein alpha 1
PSCA	prostate-specific membrane antigen gene
RAD18	E3 ubiquitin-protein ligase RAD18
RASA1	RAS p21 protein activator [GTPase activating protein] 1
RNA	ribonucleic acid
RQ	relative quantities
RQ-PCR	relative quantification polymerase chain reaction
RT-PCR	reverse transcriptase polymerase chain reaction
SA	South Africa
SFRP2	secreted frizzled-related protein 2
SHANK2	SH3 and multiple ankyrin repeat domains 2 gene
SKY	spectral karyotyping
SMAD1	SMAD protein
SMAD4	Smad protein family 4
SMAD6	Smad family protein 6 gene
SNP	single nucleotide polymorphism
SP1	splicing factor 1 gene
SPTA1	spectrin, alpha, erythrocytic 1

TERT	telomerase reverse transcriptase
TGF-β	transforming growth factor-β
TIMP4	TIMP metallopeptidase inhibitor 4
TNFRS	tumor necrosis factor related superfamily
TNK2	tyrosine kinase non-receptor 2
TP53	Tumour suppressor 53 (p53)
TS	thymidylate synthase
TSP1	thrombospondin 1
TTC4	tetratricopeptide repeat motif containing protein 4 gene
TYMS	thymidylate synthase
VEGF	vascular endothelial growth factor
WRN	Werner syndrome helicase/exonuclease protein
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4
YES1	Yamaguchi sarcoma virus v-*yes*
ZNF248	zinc finger protein 248

CHAPTER ONE

Introduction

1.1 Oesophageal Squamous Carcinoma: Epidemiology and risk factors

Overview

Oesophageal squamous cell carcinoma (OSCC) is a major cause of cancer death worldwide and has been linked to geographically endemic regions including Northern Iran, China, Japan, Central Asia, South America, African Americans in the US and Northern France (Islami et al., 2009). In South Africa, it has one of the highest incidences in the world, in particular it is endemic to the Eastern Cape and to a lesser extent, Johannesburg and the Western Cape regions (Denver Hendricks & M Iqbal Parker 2002.; van Rensburg 1987). According to the most recent cancer registry operating in the Eastern Cape alone, OSCC is the leading cancer affecting men and second most common cancer in woman residing in the region, with a prevalence of 31.3 and 18 per 100 000 individuals respectively (Somdyala et al., 2010).

Tumours are diagnosed at very advanced stages of the cancer and thus these patients have a poor survival rate (Khushalani, 2008). Surgery has low cure rates due to this factor, as the disease has in most cases metastasized. The aetiology of this cancer is unresolved and while the most common risk factors associated with OSCC include smoking and alcohol consumption, these factors are surprisingly lacking in areas of high incidence such as in Iran and northern parts of China (Islami et al., 2009). Additional risk factors have been proposed to play a role in some regions. In particular, exposure to fumonisin, a *Fusarium* fungi toxin that grows on maize, is reported in South Africa and China (W. F. Marasas et al., 1979; G. Sun et al., 2007) as well as human papillomavirus (HPV) infection (Sitas et al. 2007; Yao et al. 2006). Poor nutrition is associated with OSCC in most parts of the world (Pacella-Norman et al., 2002; Islami et al., 2009) and chronic inflammation was described in endemic parts of SA (Matsha et al., 2006). The respective part played by environmental risk factors and a potential genetic susceptibility remains unclear and it is possible that different combinations

of factors may be at play in different areas of the world. In the USA, the incidence of OSCC is 48% higher in Hispanic males than non-Hispanic white males and the incidence in African – American males is 70% higher than that of Hispanic males (Wu et al., 2007). Similarly, white South Africans smoke more than black South Africans, including those residing in the Eastern Cape region, however smoking is strongly associated with OSCC but the incidence of OSCC in white individuals is much lower (Sammon, 2007). These facts would suggest that there might be an ethnic predisposition to this form of cancer. The risk of OSCC is certainly linked to the socioeconomic status, with poorer populations being the most affected (Islami *et al*., 2009; K. M. De Groot, 2006).

Although numerous studies on OSCC have been done, which have highlighted the genetic complexity of this disease, very little has been translated into clinical, prognostic and treatment practices. Resection was the main mode of OSCC treatment in the past. However, the failure rates were high with survival not exceeding 18 months (Khushalani, 2008). Now the common practice internationally is neoadjuvant chemoradiation therapy, i.e. administration of radiation-sensitising chemotherapeutic drugs and radiation prior to resection. The 5 year survival rate, which is rarely above 25%, is poor no matter what mode of treatment is currently used (Khushalani, 2008). This highlights the need for a better understanding of the underlying mechanisms involved in the development and progression of this disease. Some of the main risk factors and molecular markers associated to OSCC are discussed below. Despite the numerous studies, no clear genetic aberrations that are linked to common molecular pathways have been identified in OSCC.

Maize consumption and OSCC

In the past sorghum was the staple diet in South African regions endemic for OSCC; however this changed to maize since the onset of the 20th century, similarly to the onset of the OSCC epidemic (Isaacson 2005). Maize is now the dietary staple of populations residing in rural and subsistence farming regions of Africa and in particular the Eastern Cape region of South Africa is known as a high maize consuming area with high levels of fumonisin contamination (G S Shephard et al., 2007). The fumonisin mycotoxins are produced in maize by the *Fusarium verticilliodes* and *Fusarium moniliforme* fungi. These toxins have been associated with the risk of acquiring OSCC (W. F. Marasas *et al*., 1979). Even though maize used for consumption is visibly separated out from mouldy maize in households, the estimated fumonisin exposures were detected to be higher than the provisional maximum tolerable daily intake for fumonisin B1 and B2 of 2ugkg^{-1}body weight day^{-1} (G S Shephard *et al*., 2007). Fumonisins can be present in maize beer and could be a significant additional source of toxin. This beer, along with non-alcoholic lactic acid fermented beverage made from local maize, could contribute to fumonisin intake (Gordon S Shephard et al., 2005; G S Shephard et al., 2007). Despite these mycotoxins being shown to be hepatocarcinogenic and nephrocarcinogenic in rats (Howard *et al*., 2001; W C Gelderblom et al., 1991), these studies could not detect changes to the oesophageal epithelium in these rats. There is still a need to determine a direct relationship between Fumonisin toxin and OSCC carcinogenesis.

Another mechanism proposed for OSCC risk associated with maize consumption is that maize meal is high in linoleic acid content, a precursor of prostaglandin E2 (PGE2) (A M Sammon & Iputo, 2006). The consumption of maize and a low fat diet are associated with elevated PGE2 production (Sammon & A. Morgan 2002). Elevated PGE2 production causes lower gastric acid production and causes relaxation of the lower oesophageal sphincter.

Lowered gastric acid production and reflux due to a relaxed sphincter has been associated with a risk to develop OSCC (A M Sammon & Iputo, 2006). However, the authors show no experimental evidence to support their hypothesis that elevated PGE2 levels are directly linked to reflux of gastrointestinal juices into the oesophagus.

Perhaps a more solid implication for elevated PGE2 is that PGE2 itself has been implicated in tumour cell growth. PGE2 stimulates G-protein coupled receptors (EPs), which are proposed to mediate over expression of C-*MYC* through protein kinase C and extracellular signal regulated kinase (ERK) thereby stimulating tumour growth (L. Yu et al., 2009). Over expression of EPs is also detected in 43.4% of OSCC cases (n=226), which was detected by immunohistochemical analysis and verified with western blotting (Kuo et al., 2009). Additional studies to corroborate these results would strengthen PGE2 as a candidate involved in OSCC.

An additional factor at play in prostaglandin synthesis is Cyclooxygenase-2 (*COX*-2), which is a rate limiting enzyme in prostaglandin synthesis. *COX*-2 is often over expressed in many cancers (M. Sahin et al., 2009). *COX-2* over expression is detected in 34-40% of OSCC tumours (n=68 and 108 respectively for the two studies) (N. Hashimoto et al., 2007; Boone et al., 2009). COX-2 inhibition with aspirin has been proposed as therapy for OSCC due to the observed *COX*-2 over expression (Umar & Fleischer, 2008).

Smoking, alcohol consumption and OSCC

Smoking and alcohol consumption are established risk factors associated with the development of OSCC and have a combined effect as well as acting independently with a dose-response relationship (Ganesh, Talole, & Dikshit, 2009; Pacella-Norman et al., 2002; A.H. Wang et al., 2004). Smoking is thought to contribute to the early and late stages of

OSCC carcinogenesis while alcohol consumption may only be associated with a later stage of carcinogenesis (Vos, Adams, Victor, & van Helden, 2003).

Generally in the Eastern Cape region the tobacco is home grown and is said to be more carcinogenic than commercial cigarettes, possibly due to the higher content of N-Nitrosamine compounds (Vos et al*.,* 2003). Cigarette carcinogens can cause a number of DNA lesions such as apurinic/apyramidinic (AP) sites, modified bases, adducts, cross links and DNA strand breaks (Hang et al. 2010). N-nitrosamine compounds are one of more than 60 known carcinogenic compounds in cigarette smoke and react with DNA to form DNA adducts. DNA adducts are regions of the DNA that are covalently bonded to a chemical, these affect the helical structure of DNA and lead to defective DNA pairing and synthesis (Hang et al*.,* 2010). DNA adducts are known to be elevated in smokers and increase the risk for development of cancer (Hang *et al*., 2010). In addition to being a risk factor to develop OSCC, smoking is found to affect the outcome of patients on treatment. Patients with a history of heavy smoking had significantly lower 3 and 5 year survival rates compared to non-smokers or light smokers (Shitara et al., 2010). The exact mechanism is not known but likely, over expression of DNA repair enzymes may affect the efficacy of chemoradiation therapy (Shitara et al., 2010).

Alcohol can act as a carcinogen through various mechanisms, including the DNA damaging effects of the alcohol metabolite, acetaldehyde, which causes oxidative stress (Millonig et al., 2011). Several enzymes are involved in metabolising alcohol, which have been implicated in the production of reactive oxygen species associated with alcohol induced carcinogenesis. A few of these enzymes are discussed later in this chapter. Alcohol consumption also aggravates mineral deficiencies, for example folate and folate deficiency

is another factor associated with the incidence of OSCC (van Rensburg, 1987). Folate (a water soluble vitamin B) is required for DNA synthesis, repair and methylation by donation of one carbon units, which are involved in cellular metabolism (Bailey 2003). Alcohol contributes to folate deficiency by inhibiting its absorption and increasing its excretion (Bailey, 2003). It can also impair folate metabolism by inhibiting folate metabolising enzymes, e.g. Methylenetrahydrofolatereductase (MTHFR) (van Rensburg, 1987; Bailey, 2003). MTHFR is responsible for the conversion of folate to 5-methyltetrahydrofolate (5-methyl THF), which is required for methionine production. The 5-methyl THF compound is metabolised by thymidylate synthase (TS), which uses the methyl group to convert deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). Depletion in folate causes a lack of this process and accumulation of dUMP, which is mis-incorporated into DNA in place of thymidine (Duthie, 2010). DNA repair enzymes excise the dUMP, which results in single stranded breaks and if folate remains at low levels, this cycle continues leading to chromosomal aberrations (Duthie, 2010). In addition to high levels of alcohol consumption, rural endemic regions do show dietary nutrient deficiencies, including folate (van Rensburg, 1987). Alcohol may therefore aggravate this deficiency and in this way, amongst other mechanisms, add to the risk of developing OSCC.

Human Papillomavirus infection and OSCC

Human Papillomavirus (HPV) infection has been associated with a number of cancers, especially cervical cancer (Matsha et al., 2002). The high risk HPV subtypes (higher risk association for development of cancer) are HPV16 and 18 while HPV6, 11 and 13 are lower risk types (McCabe & Dlamini, 2005). HPV has been linked to the aetiology of OSCC (P. F. Yao et al., 2006) but the reported prevalence of HPV in OSCC varies considerably. The frequency

ranges from 3.6% (n= 222) to, 30% (n=82) to, 56.1% (n= 435) in Western and Chinese populations (Antonsson et al., 2010; P.F. Yao et al., 2006; Xueqian Wang et al., 2010) and one Chinese study did not detect HPV in 272 OSCC cases analysed (Koshiol et al., 2010). The discrepancy in reports is thought to be attributed to the difference and accuracy of detection methods, different sample types and the regions where samples are collected (Xueqian Wang et al., 2010). In SA, HPV subtype 11 is found to be the predominant subtype associated with OSCC, with 46% (n=50) and 44.7% (n=114) of cases being positive by PCR detection in two studies (Matsha et al., 2007; Matsha et al., 2002). Only 4.3% of cases were positive for the high risk subtype 16 (Matsha et al., 2002). These findings would suggest that HPV infection may be associated with OSCC incidence, although this is unlikely as the low risk subtype 11 is predominant in SA.

The mechanisms proposed for HPV in carcinogenesis is through the E6 and E7 proteins. Expression of the HPV16 oncoproteins, E6 and E7, was observed to induce chromosomal aberrations (H. Zhang et al., 2006) and the expression of proteins such as telomerase, c-myc, ras, bcl-2 and p53 are altered in the presence of HPV18 oncoproteins E6 and E7 (Z. Y. Shen et al., 2001). The de-regulated expression of these proteins results in loss of apoptosis and uncontrolled cell growth (Z. Y. Shen et al., 2001). Oesophageal cells immortalised by HPV16 also showed gain of chromosome 20q, which suggested that this aberration was an early event in the process of oesophageal squamous cell immortalisation (Zhang et al., 2006). There is still no evidence of HPV directly causing OSCC.

Chronic inflammation and OSCC

Chronic inflammation can be induced by biological, chemical and physical factors and induces oxidative stress and lipid peroxidation (Bartsch & Nair, 2006). These processes can

lead to the production of reactive oxygen species and DNA reactive aldehydes, which cause DNA damage (Helmut Bartsch & J. Nair, 2006). Chronic inflammation is proposed to be associated to, or to act as a precursor to the development of OSCC (Matsha et al., 2006). The endemic area of South Africa, the Eastern Cape, is populated by a majority of Xhosa speaking people who follow a cultural tradition of self–induced vomiting. This ritual is considered as either a biological cleansing or spiritual cleansing and the most common mode of induction is through the consumption of sea water or salt water (Matsha et al., 2006). Only one study showed that self induced vomiting is strongly associated with chronic inflammation of the oesophagus, however the methodology used in the study by Matsha et al (2006) did not differentiate between other possible causes of inflammation such as infection or exposure to irritants such as smoking and alcohol, which could have affected their results (Matsha et al., 2006). Additional studies on chronic inflammation in OSCC are still required to establish if self-induced vomiting has a direct impact on OSCC development.

1.2 OSCC Molecular biology

OSCC is difficult to stage on pathological features alone and the heterogeneous nature of the tumours presents a problem using standard methods (Nair et al., 2005). Patients with lymph node metastases and organ infiltration do poorly in comparison to those patients with minimal tumour invasion (Nair et al., 2005). It would be of great benefit to identify molecular markers related to tumour invasion in the interest of using them as prognostic and diagnostic indicators. Many studies have investigated various molecular factors involved in tumour development and progression such as, factors of cell adhesion, repair genes, chemokines, gene polymorphisms, oncogenes and tumour suppressor genes. Most

studies have been done abroad and a few studies were done in SA. Some of the most important molecular factors that emerged in these studies are discussed below and the details of publications describing these factors from multiple studies are summarised in Table 1.1.

Adhesion molecules

The first step for cancer cells to metastasize is for them to detach from the tumour, which can occur when there is de-regulation of cell-cell adhesion and cell-matrix interactions (K. S. Nair et al. 2005). The cell adhesion molecules (CAMS) are critical in this process. There are several families of CAMS involved in adhesion, which include integrins, cadherins, immunoglobulins and selectins.

The integrins are transmembrane glycoproteins that are important in the control of attachment, migration, cell division and apoptosis. They exist as alpha or beta heterodimers at focal adhesion sites and have extracellular domains (acting as receptors in the extracellular matrix) and intracellular domains, which bind to the actin cytoskeleton (K S Nair et al., 2005). Under normal conditions, integrins are stimulated by the extracellular matrix to mediate adhesion and inactivation of integrins prevents adhesion at inappropriate times. It only was shown by one study in SA that α2 and β1 integrin subunits were under expressed, suggesting that there is decreased adhesion in OSCC (S. E. Miller & Veale, 2001). In another study, alpha4-integrin was hypermethylated in 21% of OSCC tumours (n=251) (E. J. Lee et al., 2008). This study only assessed the methylation status of the gene promoter and did not show whether gene expression was affected by the methylation status although methylation was associated with disease recurrence and a poor survival (E. J. Lee at al., 2008). The β1 integrin subunit was also shown to have low expression in 71 OSCC lymph

node metastases in Japan (Takayama et al., 2003). These three independent studies have not been repeated and provide little evidence to support a role for decreased expression of integrins in OSCC.

Increased expression of some integrins has been associated with the ability of cancerous cells to adhere to the vascular endothelium, which is required in the process of migration. One study showed strong expression of the α5 integrin in tumour samples and not in the normal epithelium, this is likely to allow for the adhesion to endothelial vessels required for migration and as a result this may increase the metastatic potential of OSCC cells (S. E. Miller & Veale, 2001). In another study, α2β1 and α3β1 integrins were similarly over expressed in OSCC cells cultured with endothelial cells and the authors observed in culture that this enhances migration of cells to extravascular tissues (M Sato et al., 1996).

The cadherins are a family of membrane glycoproteins that are required for cell-cell adhesion in normal cells. E-cadherin is one of the most important cadherins in epithelial cells and has been the most extensively studied in OSCC. Reduced expression of E-cadherin results in OSCC cells having a higher metastatic potential (K S Nair et al., 2005; Kriebashne S Nair et al., 2006). E-cadherin was found to be absent in 57% of SA OSCC tumours (n=100) (Kriebashne S Nair et al., 2006), however expression was assessed by immunohistochemical analysis and could not be directly related to prognosis. A mechanistic role for reduced E-cadherin in OSCC carcinogenesis was shown by one study, where exposure of cells to endothelial proteins (e.g. SHEAR) resulted in internalisation of E-cadherin in an OSCC cell line, which increased invasiveness of the cells (Lawler et al., 2009). Methylation of the *CDH1* gene (encoding E-cadherin) was detected in 43% (n=251) and 42.3% (n=71) of OSCC tumours (E. J. Lee et al., 2008, Chung et al., 2007). Two studies showed that the methylation status of

the *CDH1* gene in stage I OSCC was strongly associated with a high risk of recurrence (E. J. Lee et al., 2008) and lymph node metastasis (Takayama et al., 2003.). These studies all applied immunohistochemistry to establish the expression levels of E-cadherin and only one study (Lawler et al., 2009) looked at the functional impact of E-cadherin expression in a cell line.

Catenins are cytoplasmic proteins closely related to the cadherins that bind cytoplasmic domains of cadherins to the cytoskeleton. One study showed that reduced expression of E-cadherin correlated with lower levels of β-catenin in the cytoplasm (Takayama et al., 2003). It was found that in 46 primary OSCC tumours, 56% of cases had similar expression of E-cadherin and α-catenin while 46% had expression of E-cadherin in some of the tumour cells and no α-catenin expression (Kadowaki et al., 1994). Kadowki et al (1994) showed that α-catenin was only expressed when normal E-cadherin was present, suggesting that E-cadherin is required for α-catenin stability. Regardless, there are no functional studies to corroborate this hypothesis and it is only based on one method, immunohistochemistry. Forty eight percent of tumours showed no expression of α-catenin while E-cadherin was still expressed; suggesting that α-catenin is downregulated earlier in metastasis than E-cadherin. Reduced expression of α-catenin was correlated with invasion and lymph node metastasis (Kadowaki et al., 1994).

CD44 is a transmembrane glycoprotein which binds ligands in the extracellular matrix and is involved in cell-matrix adhesion. CD44 has a variable affinity to bind hyaluronic acid, which allows for binding to other ligands. The down regulation of CD44v2 is associated with a poor prognosis in OSCC (K S Nair et al., 2005). Two studies detected lowered expression of CD44v6 in OSCC (Roye et al., 1996; Takayama et al., 2003). One study showed that low

expression of CD44v6 had a significant correlation with lymph node metastasis (n=71) (Takayama et al., 2003). CD44v3 and CD44v6 expression patterns differed from normal, dysplastic, moderately differentiated carcinoma and poorly differentiated tumours with the highest expression of these variants observed in the normal epithelium (Roye et al., 1996), suggesting that the lowered expression of these variants is associated with disease progression. In contrast, three studies showed that CD44v6 was over expressed in 64.6% of OSCC (n=82) cases, 71.43% of cases (n=21) and 46.9% of cases (n=38), which had a correlation with lymph node metastasis (Wen-Kang Liu et al., 2009; DM. Li et al., 2005; Nozoe et al., 2004). All these studies were performed using the subjective method, immunohistochemistry, which could account for the contradictory results and highlight the need to clarify the exact role of CD44 and its isoforms in OSCC.

A newer candidate gene being investigated in OSCC is cortactin (CTTN), a cytoplasmic protein also involved in cell adhesion and cytoskeletal structure. This gene is on chromosome 11q13.3, a region frequently amplified at the DNA level in OSCC (Man-Li Luo et al., 2006). *CTTN* is highly over expressed in 43.5% of OSCC cases (n=46) and 47.8% had high expression but at a lower level (K. F. Hsu et al., 2009). Over expression of *CTTN* has been detected in pre-malignant lesions, suggesting its involvement in carcinogenesis (N. Y. Hsu et al., 2008). Another study studied the expression of *CTTN* by immunohistochemical analysis, reverse transcriptase PCR (RT-PCR) and gene amplification by PCR and fluorescence *in situ* hybridisation (FISH) analysis. Fifty percent of primary tumours (n=16) had *CTTN* gene over expression by PCR analysis, 68.8% (n=125) had a strong immunohistochemical staining, which correlated with tumour stage and lymph node metastasis while amplification (as detected by FISH and PCR) was detected in 70% of samples (n=20) (Man-Li Luo et al., 2006).

This study showed that gene amplification correlated with over expression in 70% of tumours and over expression of *CTTN* was associated with lymph node metastasis, a strong indication that *CTTN* has a role in OSCC. Luo *et al* (2006) also performed RNA interference (RNAi) experiments in EC9706 cells (OSCC cell line), which were inoculated into nude mice. It was found that knock down of *CTTN* resulted in these cells growing slower in nude mice than in mice inoculated with EC9706 cells without RNAi against *CTTN* (Man-Li Luo et al., 2006). They concluded that *CTTN* is an independent marker for lymph node metastasis and promotes tumour invasiveness in OSCC (Man-Li Luo et al., 2006). *CTTN* expression was correlated with activation of the PI3K/Akt pathway, which is proposed to facilitate resistance to anoikis (programmed cell death initiated by cell detachment), which may allow survival of cells in circulation (X. L. Du et al., 2009; Man-Li Luo et al., 2006). These studies show good evidence to highlight *CTTN* as a promising marker for further investigation in OSCC as it appears to be functionally relevant in OSCC carcinogenesis.

It is evident from the literature that E-cadherin is the adhesion molecule with the most corroborative data to show that it is involved in the metastasis of OSCC. However the studies are not supported with functional analyses. On ther other hand, *CTTN* has good evidence to show a functional role in OSCC. It is unclear from the work on the other molecules in the adhesion network whether they are of real importance. More information and a comprehensive understanding of the interactions and impact of all these adhesion molecules are required to see if they can be used as markers of diagnosis and prognosis. Studies involving more samples, alternative methods and a collective effort on the adhesion molecule network could help in this regard. Specifically, whole genome studies could show

aberrations in multiple molecules involved in adhesion, which could facilitate identifying those that may work in combination.

Repair genes

The mismatch repair genes are important for the detection and correction of errors in DNA sequence that arise during cell division. Mutation of both alleles of mismatch repair genes results in DNA mismatch repair being defective and consequently, the accumulation of mutations occurs in the DNA (Vasavi et al. 2010). The mismatch repair gene most well studied in oesophageal cancer is *hMLH1*. Most studies have been performed in China and India assessing the methylation status and microsatellite instability (MSI) of this gene. A series of studies show methylation of the *hMLH1* gene in 27-85.7% of OSCC cases (case numbers ranged from 30 to 125) with this frequency increasing with disease progression. Methylation of *hMLH1* is therefore associated with a poor prognosis (Mohan Vasavi et al., 2006; Uehara et al., 2005; M Vasavi et al., 2010). Fifty percent of pre-cancerous lesions (n=13) in one study had methylation of *hMLH1* with MSI (M Vasavi et al., 2010), the authors suggest that *hMLH1* inactivation may occur early in oncogenesis, but due to the small number studied, this would need to be analysed in a larger sample size to prove that *hMLH1* inactivation is an early event. None of these studies were verified by an alternative technique to establish whether expression of *hMLH1* is affected and whether this does indeed result in defective DNA repair in patients with OSCC. This gene has been studied in SA in one study. Microsatellite instability (MSI) analysis of 6 markers located at mismatch repair gene loci (*hMLH*, *hPMS1*, *hPMS2*, *hMSH2*, *HMSH3* and *hMSH6*) showed a low frequency of aberrations in 100 OSCC patients in South Africa. Loss of heterozygosity (LOH) was found more frequently than MSI in the microsatellite loci analysed and specifically in chromosome 3 (D3S659 and D3S1255, *hMLH1*), which was detected in 40% of the cases (R

Naidoo et al., 2005). Naidoo et al (2005) concluded that no predisposition could be associated with any of these loci in OSCC and that the molecular pathway for OSCC is something involving genes other than the mismatch repair genes (R Naidoo et al., 2005). However, the study only investigated 100 tumours and normal epithelium, it is unclear if the normal epithelium was obtained from the same patients or not and only a clear case control study could determine whether MSI at these loci contributes to a predisposition for OSCC or not.

Chemokines

Chemokines and their receptors stimulate the recruitment of leukocytes through chemo attraction and CXC chemokines (chemokines with a C-X-C motif) play a role in inflammatory responses (B. Wang et al., 2006). They have also been linked to cancer and tend to show over expression in various cancer types for example, the CXC chemokine family member, Growth-related oncogene alpha (GROα) and its receptor chemokine receptor 2 (CXCR2) act as growth factors and are over expressed in melanoma and prostate cancer cells (B. Wang et al., 2006). In South Africa, one study examined the chemokine GROα and the receptor CXCR2 in OSCC. Only 4 tumours and matching normal epithelium were studied and a two to five fold increase in expression of GRO α and β was found in these tumours (B. Wang et al., 2006). RNA interference against GROα and GROβ in the South African OSCC cell line WHCO1, resulted in 20% and 50% decrease in cell proliferation respectively. The authors of this study suggested that GROα-CXCR2and GROβ-CXCR2 signalling pathways contribute to OSCC proliferation and may be involved in OSCC development (Wang et al., 2006). These studies are too small to show real significance of these proteins in OSCC and their results need to be corroborated by larger studies.

Wang et al (2009) investigated expression of the *EGR-1* (early growth response-1) gene, which they reported to be over expressed in 80% of OSCC tumours (n=65) and four OSCC cell lines (WHCO1, WHCO5, WHCO6 and KYSE 450) (B. Wang et al., 2009). EGR-1 protein knockdown in the WHCO1 cell line resulted in a 50% reduction in cell proliferation, 50% reduction in Cdkn-4 levels and mediated GRO/CXCR2 proliferative signalling (Wang et al., 2009). Wang et al (2009) suggested that this protein could be a therapeutic target in OSCC. This study also needs to be corroborated with additional research by other groups.

Five studies from China and Japan investigated the expression of CXCL12 and its receptor CXCR4. All three studies found that both CXCL12 and CXCR4 were strongly expressed in OSCC primary tumours (numbers were 86 to 214 samples) and two of the studies found that CXCL12 expression correlated with lymph node metastasis while CXCR4 did not (Lu et al. 2010; D.F. Wang et al. 2009; K. Sasaki et al. 2008; K. Sasaki et al. 2009; Gockel et al. 2007).

Again, these studies are lacking in additional methods to support their immunohistochemistry results. There is still not enough evidence to say that chemokines are clinically relevant to OSCC.

Genetic polymorphisms associated with risk of OSCC development.

Metabolic enzymes responsible for metabolism of carcinogenic compounds, such as those in cigarette smoke and alcohol, have been investigated due to risk association to develop OSCC with smoking and alcohol consumption. Genetic polymorphisms in these genes are thought to contribute to this risk of developing OSCC.

Sulfate conjugation is involved in pathways responsible for the metabolism of carcinogenic compounds such as catecholamines, aromatic amines and certain drugs. This sulfation is

carried out by sulfotransferases, encoded by the *SULT* gene family (Glatt et al. 2001). The enzyme is important in the metabolism of heterocyclic amines and polycyclic aromatic hydrocarbons, components of cigarette smoke (Glatt et al. 2001). *SULT1A1* polymorphisms are associated with a higher risk to develop breast cancer and colorectal cancer (M.T. Wu et al., 2003; Cleary et al., 2010) and specifically the G--→A transition at codon 213 (*SULT1A1**2 allele) confers a reduced enzyme activity and stability (Glatt et al., 2001; Weinshilboum et al., 1997). Only two studies have looked at these genes in OSCC, one in Taiwan and one in South Africa. The *SULT1A1**2 allele was found in 27.8% of OSCC cases (n=187) in Taiwan, which was significantly higher than controls (n=308). This allele (*SULTA1**2) was associated with a 3.53 fold higher risk of developing OSCC in Taiwanese men (M.T. Wu et al., 2003). The *SULT1A1**2 allele was investigated in the black and mixed ancestry populations in SA. The polymorphism in a homozygous state (*SULT1A1**2*/*2*) was detected with a frequency of 32% in blacks and 31% in mixed ancestry individuals with OSCC compared to 20% (n=245) and 14% (n=288) in controls respectively. The presence of this homozygous genotype was associated with an increased risk for OSCC development in smokers (Dandara et al., 2006). Additional studies to corroborate this data are required.

Alcohol dehydrogenases (ADH) metabolise alcohol into acetyldehyde, which is further metabolised into acetic acid by aldehyde dehydrogenase (ALDH). Acetyldehyde is a known carcinogenic compound in this pathway (Jelski et al., 2009). Highly active ADH enzymes are associated with alcohol intolerance and low active ADH enzymes are associated with an increased risk for cancer development (Brooks et al., 2009). Three meta-analyses of *ADH1B* and *ALDH2* gene polymorphisms in China and Japan from several studies, collectively looking at over 1000 patients and controls, showed that *ADH1B* (Arg/Arg) and *ALDH2*

(Lysine) genotypes have an increased risk to develop OSCC and interact significantly with alcohol consumption (G. H. Zhang et al., 2010; F. Tanaka et al., 2010; S. J. Yang et al., 2010). These particular genes have not been investigated in SA apart from *ALDH2* polymorphisms.

The Alcohol dehydrogenase 2 (*ADH2*) gene has two polymorphisms (*ADH2*2* and *ADH2*3*) associated with increased activity of the enzyme. The *ADH2*2* allele was associated with an increased risk to develop OSCC in a Chinese study of 221 cases and 191 controls (J. hua Ding et al., 2010). A case control study in 238 patients and 268 controls from South African black (n=125) and mixed ancestry (n=268) populations investigated frequencies of *ADH2*, *ADH3* and *ALDH2* genotypes. The *ADH2*1*, *ADH3*2, ADH3*2/*2* and *ALDH2*2* alleles were significantly associated with an increased risk to develop OSCC in the black individuals, while the *ADH2*2* and *ADH2*3* alleles were only associated with increased production of acetaldehyde and a higher intolerance to alcohol (D. P. Li et al., 2008). Li et al (2008) suggest that the mechanism by which the *ADH2*1* and *ADH3*2* alleles may increase the risk for OSCC development could be due to a higher tolerance to alcohol. The higher tolerance of individuals for alcohol may result in prolonged consumption and high volumes of alcohol could act as a solvent for other carcinogenic compounds such as those from cigarettes and lead to induction of other carcinogen metabolising enzymes such as cytochrome p450 (D. P. Li et al., 2008). Interactions between polymorphisms in the ADH and ALDH enzymes with those polymorphisms in the cytochrome p450's may add to the risk of OSCC development.

The cytochrome p450 2E21 (CYP2E1) enzyme in particular, metabolises ethanol as well as nitrosamines, and may bio activate these potential carcinogens (D. P. Li et al., 2008). The gene contains several polymorphisms with contradicting reports on the effects on the expression of the gene and its protein function. *CYP2E1* was recently shown to be induced in

a dose dependant manner in the oesophagus by ethanol (Millonig et al., 2011) and various studies in Chinese and Japanese populations show that the *CYP2E1**5 polymorphism (1293G→C) is associated with an elevated risk to develop OSCC (X. M. Lu et al., 2005; W. Tan et al., 2000; J. M. Qin et al., 2008; Y. M. Guo et al., 2008; R. Liu et al., 2007). *CYP2E1* has not been extensively studied in the SA population. One study investigated three polymorphisms and have conflicting reports on the effects on expression and function of the gene, 1053C→T, 1293G→C and 7632T→A. The study was performed on 189 patients with OSCC and 198 controls from black and mixed ancestry populations (D. Li et al., 2005). The polymorphism with the highest frequency was the 7632T→A polymorphism, detected in 18% of patients (n=189) as opposed to 7% (n=198) in control individuals. The *CYP2E1**6 polymorphism was associated with an increased risk of developing OSCC in SA as opposed to the more frequent *CYP2E1**5 polymorphism detected in Chinese and Japanese populations (D. Li et al., 2005). A second study performed in SA, on 331 black individuals (70 cancers and 261 controls), reported a higher frequency (95%) of the *CYP2E1*1* polymorphism (1053C→T) also in contrast to the Chinese and Japanese studies (Chelule et al., 2006). However both these studies in SA have relatively small numbers and the first being on both black and mixed ancestry individuals does not clarify the frequency of these alleles in the black population where OSCC has the highest incidence.

Overall, studies of genetic polymorphisms related to disease susceptibility require large numbers of subjects with and without disease to determine the significance of these alleles in specific populations. Although some of these genes involved in chemical metabolism maybe convincing in the Chinese and Japanese populations, further work is required to establish their role in the South African population affected by OSCC.

Oncogenes and tumour suppressor genes implicated in OSCC

Oncogenes can be categorised into several classes including growth factors, growth factor receptors, nuclear factors and signal transducers all of which have been implicated in carcinogenesis.

The growth factors and growth factor receptors stimulate cells to enter into the cell cycle. Growth factors have not been extensively studied in OSCC, however vascular endothelial growth factor (*VEGF*) has been implicated in OSCC in China and Japan by immunohistochemical anlaysis (summarised in Table 1.1). The growth factor, VEGF, is important in angiogenesis, an important process for tumour survival. The *VEGF* gene is over expressed in a high number of tumours (43-100%) and is associated with tumour progression (Y Shimada et al., 1999; Okazawa et al., 2008.; P. Liu et al., 2009; Gholamin et al., 2009; Bedoya et al., 2009). These studies could only show a prognostic value for *VEGF*, no functional studies are available to show the direct impact of *VEGF* expression on OSCC carcinogenesis.

The epidermal growth factor receptor (*EGFR* gene) is a tyrosine kinase transmembrane receptor and is often over expressed in lung and colorectal cancers (Ding et al. 2008; Dragovich & Campen, 2009). This receptor has sparked interest due to the development of targeted therapies against EGFR protein in the form of small molecule tyrosine kinase inhibitors and monoclonal antibodies (Dragovich & Campen, 2009). It is a promising gene in OSCC however, most studies only looked at over expression with immunohistochemistry and so should be evaluated using an alternative method as there is variation in the reported data. Over expression of *EGFR* is detected in OSCC as evidenced by a few studies from Japan, China, the Netherlands and France (summarised in table 1.1). Over expression was

detected at frequencies ranging from 34-75.8% (Gotoh et al., 2007; Hoshino et al., 2007; Gongyuan Zhang et al., 2010; Boone et al., 2009; Gibault et al., 2005). Hoshino et al (2007) specified that 71.2% of OSCC specimens (n=52) show EGFR protein localisation at the cell surface and only 36.5% showed phosphorylated EGFR in the nucleus and the latter was correlated with a poor prognosis (Hoshino et al., 2007). Zhang et al (2010) detected EGFR mRNA by *in situ* hybridisation in 75.8% of cases (n=62), which correlated with tumour grade, invasion and lymph node metastasis (Gongyuan Zhang et al., 2010). Over expression of *EGFR* in primary OSCC was also correlated with the response of patients to chemoradiation therapy (Gotoh et al., 2007). The activation of EGFR protein is thought to contribute to transformation of oesophageal epithelium to squamous carcinoma under inflammatory conditions (Parthasarathy et al., 2010). *EGFR* gene amplification has also been reported, which is associated with over expression in 85-88.7% of the OSCC cases that had gene amplification (Hanawa et al., 2006; Carneiro, Isinger, Karlsson, J. Johansson, Jönsson, et al. 2008). The sensitivity of patients to treatment using *EGFR* targeted therapies is still unclear and further work is required on the mechanisms by which *EGFR* acts in carcinogenesis in order to establish reliable molecular predictors of patient responses to this kind of therapy (Dragovich & Campen, 2009).

The nuclear factor Cyclin D1 (encoded by *CCND1*) forms complexes with the cyclin dependent kinases, CDK4 and CDK6, which control the transition of cells from G1 to the S phase of the cell cycle. Over expression of the *CCND1* gene results in deregulated cell division. A study looking at the effects of cigarette smoke extracts on the expression of *CCND1* in an OSCC cell line (EC109), showed an increase in proliferation and *CCND1* protein

levels in a dose responsive manner (H. Hu et al., 2009). The cyclin D1 protein levels reduced upon treatment with aspirin and resulted in decreased cell proliferation (H. Hu et al., 2009).

CCND1 over expression and gene amplification with over expression, have been well documented in OSCC in more than 10 studies mostly conducted in USA, Japan and China (summarised in Table 1.1). *CCND1* gene amplification and over expression is detected in 22-73% of OSCC tumours and cell lines and several of these studies show a correlation with lymph node metastasis and a poor survival (S. Zhu et al., 1998; S. Fujii et al., 1998; H. Suzuki et al., 1998; Takeuchi et al., 1997; Roncalli et al., 1998; Shamma et al., 1998; Inomata et al., 1998; R Chetty & S. Chetty, 1997; Sheyn et al., 1997; J. Zheng et al., 1996; Shinozaki et al., 1996; Kitahara et al., 1996; H Nakagawa et al., 1995; W Jiang et al., 1993; W Jiang et al., 1992; De-Chen Lin et al., 2010). One study from South Africa reported a minority of cases (29%, n=80) to have over expression by immunohistochemical analysis and over expression could not be correlated with histological grade (R Chetty & S. Chetty, 1997). This gene requires further investigation in the SA setting with only one study to date however, the *CCND1* gene is a significant gene involved in OSCC pathogenesis and appears to be a useful prognostic marker.

Over expression of the *CCND1* gene can also occur through deregulation of the Wnt pathway as *CCND1* is a transcriptional target of the TCF/β-catenin complex. The Adenomatosis Polyposis Coli (*APC*) gene is a tumour suppressor gene that is involved in antagonising the Wnt pathway. Loss of the APC protein contributes to accumulation of non-phosphorylated β-catenin in the nucleus, which leads to deregulated cell division through transcription of target genes by the TCF/β-catenin complex. *APC* can be inactivated by hypermethylation of its promoter or by LOH/deletion (studies are summarised in Table 1.1).

Hypermethylation was found in 44.4% (n=45) and 46% (n=50) of OSCC tumours in two studies, however it is unclear whether this correlates with metastasis or a poor survival as each of these studies reports a different finding in this regard (Zare et al., 2009; Y. T. Kim et al., 2009). Expression of *APC* was lost in 30.4% (n=199) of late stage OSCC according to a study in China. Indeed *APC* expression decreases with undifferentiated tumours; therefore nuclear β-catenin levels increase (H. Peng et al., 2009). *APC* has not been well studied in SA, only one study evaluated MSI in the *APC* gene in 100 cases of OSCC. Nair et al (2006) showed an overall LOH frequency of 62.48% and 41.27% MSI for at least one of the 4 markers analysed (Kriebashne S Nair et al., 2006). *APC* gene expression and methylation could still be investigated in the South African setting using more than one method to verify the international findings.

In addition to *APC*, the fragile histidine triad (*FHIT*) gene was also shown to inhibit the oncogenic TCF/β- catenin complex (Weiske et al., 2007) and several studies (summarised in table 1.1) show that loss of *FHIT* by deletion or methylation occurs in OSCC at a frequency of 14-89% and correlates with loss of expression of the gene (X. Q. Guo et al., 2007; M. Morita et al., 2006; Kuroki et al., 2003; Noguchi et al., 2003; Kitamura et al., 2001; Y Shimada et al., 2000; M Mori et al., 2000; Menin et al., 2000; H. Tanaka et al., 1998). These studies showed varying degrees of *FHIT* involvement (from 14% to 89%). This variation in frequency is likely due to the differing methodologies used to detect *FHIT* deletions, never the less, *FHIT* seems to be an important tumour suppressor in OSCC.

The well known oncogene, *C-MYC* encodes a transcription factor involved in transcription of numerous genes involved in cell division, apoptosis and transformation. The *C-MYC* gene is deregulated through amplification or translocation in many cancers. C-*MYC* amplification

was detected in 39% (n=41) and 45% (n=20) of OSCC samples (J. He et al., 1995; H. Arai et al., 2003) while two other studies on 20 and 46 OSCC samples did not show a significant amplification or over expression of C-*MYC* and report that C-*MYC* may not be important for OSCC carcinogenesis (S. Miyazaki et al., 1992; X.P. Huang et al., 2006). Only one study was done on C-*MYC* expression in four South African OSCC cell lines using one technique, western blot analysis, which showed over expression of C-*MYC* in two of the cell lines (G. J. Jones et al., 1993). The limited study of *C-MYC* in OSCC would not suggest a role for C-*MYC* in the disease pathogenesis, however these studies had relatively small samples sizes and further study of this oncogene may clarify its role.

Additional tumour suppressor genes inactivated, either by mutation, LOH or deletion that are also important in the deregulation of the cell cycle and apoptosis in cancer cells are the *TP53* and *CDKN2A* genes, both previously investigated in OSCC (Metzger et al. 2004). Although p53 is usually active in oncogenesis by deletion, mutations of the *TP53* gene have also been correlated with abnormal over expression of this protein (Wagata et al., 1993), which accumulates in the nucleus (von Brevern et al., 1996) and possibly inhibits apoptosis (Hamada et al., 1996), although there has not been experimental data to support this hypothesis. Expression analysis of p53 has mostly been conducted by immunohistochemistry, a limitation to these studies. These studies show over expression of p53 in 39 to 88% of tumours (K Y Lam et al., 1995; Casson et al., 1998; Wagata et al., 1993; von Brevern et al., 1996; Muro et al., 1996; K Y Lam, S Law, et al., 1997). Mutated p53 is suggested to have a longer half life and accumulates in less differentiated cells (A. K. Lam, 2000; K Y Lam et al., 1995). Patients with abnormally expressed p53 were suggested to have a shorter survival (K Y Lam, S W Tsao, et al., 1997; Casson et al., 1998).

TP53 gene missense mutations (mostly correlating with p53 over expression) have been described in OSCC in more than 20 studies, 2 of which were from South Africa (summarised in table 1.1) (K Y Lam et al., 1995; Casson et al., 1998; Wagata et al., 1993; Montesano et al., 1996; W. P. Bennett et al., 1997; Hamada et al., 1996; Muro et al., 1996; Seta et al., 1998; Matsha et al., 2007; Vos et al., 2003; R Chetty & Simelane, 1999; Gamieldien et al., 1998; Jaskiewicz & K. M. De Groot, 1994). Overall, from these studies, at least 50% of tumours have *TP53* gene aberrations internationally, with most mutations occurring within exons 5-8 (A. K. Lam, 2000).

Two SA studies investigated *TP53* mutation status; the first analysed exons 5-8 and the second the flanking regions of exons 5-8. Gamieldien et al (1998) detected a relatively low mutation frequency of 17% (n=76) (Gamieldien et al., 1998). The second study by Vos et al (2003), on 73 tumours, only detected 2 mutations in the exon 5-8 flanking regions (Vos et al., 2003). These studies do not suggest a role for *TP53* mutation in OSCC. However, more comprehensive analysis of the *TP53* gene is required in SA. Additional *TP53* mutations in the intronic sequences or *TP53* expression may be affected by other mechanisms, which were not addressed in these studies. Three SA studies showed that expression of p53 was absent in 30-40% of OSCC tumours (n=50 to 155) (Matsha et al., 2007; R Chetty & Simelane, 1999; Jaskiewicz & K. M. De Groot, 1994). These studies did not report any association with tumour grade or depth of tumour invasion, which may be due to the limitation of only using one technique, immunohistochemistry. Overall, p53 overexpression appears to be the major mode of action in carcinogenesis however, the molecular basis for p53 involvement in OSCC carcinogenesis needs clarification with functional studies.

The p16 protein (encoded by *CDKN2A*) is an inhibitor of the cyclin dependant kinases and inactivation of this protein facilitates the transition of cells into S-phase. The expression of p16 is often lost in OSCC and deletions and/or mutations are not frequently detected suggesting that other mechanisms of p16 inactivation occur in OSCC (A. K. Lam, 2000). One of the mechanisms reported is hypermethylation of the *CDKN2A* gene, which correlates with loss of expression in 60%-80% of OSCC tumours (studies are summarised in table 1.1) (Taghavi et al., 2010; Salam et al., 2009; X. Q. Guo et al., 2007; Jianming Wang et al., 2008; Fukuoka et al., 2006; M. Guo et al., 2006). Only one study was performed on *CDKN2A* in SA, the mutation status of exons 1 and 2 was investigated. Seventy six OSCC samples were analysed and mutations were found in 28% of the samples (Gamieldien et al., 1998). These mutations resulted in missense mutations, deletions, stop codons and insertions, which would result in loss of functional protein (Gamieldien et al., 1998).

Overall, from the published data, genes that play a role in oncogenesis have been identified. Most of these studies focused on association with prognosis. However, the results from these studies are difficult to apply in a clinical setting and the understanding of the molecular pathways underlying carcinogenesis in OSCC is still not clarified. As mentioned before, the tumours are difficult to stage, which may contribute to the variation in results obtained.

In South Africa specifically, the focus of research on OSCC, as described above, has been on gene polymorphisms linked to the risk of developing disease, mutations in the *TP53*, *CDKN2A* and *APC* genes, expression of some of the adhesion molecules and the association of HPV with the disease. These studies have not yet elucidated clear pathways for the

molecular carcinogenesis of OSCC in South Africa. Whole genome studies could help to identify genes linked to key pathways, which collectively contribute to OSCC carcinogenesis.

Table 1.1 Summary of studies done on molecular markers in OSCC

Author and country	Number of samples	Aberration	Method	Findings/comments
ADHESION				
β1 integrin				
(S. E. Miller & Veale 2001), SA	?	Under expressed (?)	?	Could only obtain an abstract. Increased metastasis
(Takayama et al.2003), Japan	71	Under expressed (?)	immunohistochemistry	Could only obtain an abstract. Correlated with lymph node metastasis
CD44				
(Takayama et al. 2003), Japan	71	Reduced expression of CD44 and CD44v6 (?)	immunohistochemistry	metastasis
(Roye et al. 1996), USA	50	Over expressed CD44 (100%)	immunohistochemistry	OSCC development
(W-K Liu et al. 2005), China	40	CD44v6 over expressed (65%)	immunohistochemistry	Increased expression linked to HPV16 infection
(D.-M. Li et al. 2005), China	42	CD44v6 over expressed (71.43%)	immunohistochemistry	
(Nozoe et al. 2004), Japan	81	Over expressed (46.9%)	immunohistochemistry	
E-Cadherin				
(Kriebashne S Nair et al. 2006), SA	100	Low or no expression (81%)	immunohistochemistry	
(Lawler et al. 2009), Ireland	Oc-1 cell line	Showed shear induced internalisation of E-cadherin	Cell culture experiments	Endocytosis of E-cad associated with invasiveness
(E. J. Lee et al. 2008), China	251	Methylated (43%)	Methylation specific PCR	
(Chung et al. 2007), China	71	Under expressed (42.3%)	immunohistochemistry	
(Takayama et al. 2003.), Japan	71	Decreased expression	immunohistochemistry	
(Kadowaki et al. 1994), Japan	46	Reduced expression (46%)	immunohistochemistry	

Table 1.1 Summary of studies done on molecular markers in OSCC

ONCOGENES				
VEGF				
(Y Shimada et al. 1999), Japan	116	Over expressed (68.9%)	immunohistochemistry	Tumour invasion
(Gholamin et al. 2009), Iran	49	Over expressed (44.9%)	mRNA detection	
(Bedoya et al. 2009), Spain	29	Over expressed (100%)	immunohistochemistry	
(P. Liu et al. 2009), China	73	Over expressed (53.4%)	immunohistochemistry	
(Okazawa et al. 2008.), Japan	100	Over expressed (43%)	immunohistochemistry	
EGFR				
(Gotoh et al. 2007), Japan	62	Over expressed (34%)	immunohistochemistry	Predicts response to chemoradiation therapy
(Hoshino et al. 2007), Japan	52	36% over expression of p-EGFR in nucleus	immunohistochemistry	p-EGFR associated with a poor prognosis
(Gongyuan Zhang et al. 2010), China	62	Over expressed (75.8%)	in situ hybridisation	
(Boone et al. 2009), Netherlands	108	Over expressed (40%)	immunohistochemistry	
(Gibault et al. 2005), France	107	Over expressed (68.2%)	immunohistochemistry	
(Hanawa et al. 2006), Japan	106	Over expressed (50%)	immunohistochemistry	
(Carneiro, Isinger, Karlsson, J. Johansson, Jönsson, et al. 2008), Denmark	30	Amplified (46.6%)	CGH array (32K BAC array)	
CCND1				
(S. Zhu et al. 1998), China	104	Over expressed (62.5%)	immunohistochemistry	
(S. Fujii et al. 1998), Japan	26 cell lines	Over expressed (65.4%)	Flow cytometry	
(H. Suzuki et al. 1998), Japan	4	Amplified (100%)	Fluorescence in situ hybridisation (FISH), differential PCR	

Table 1.1 Summary of studies done on molecular markers in OSCC

(Takeuchi et al. 1997), Japan	111	Over expressed (25%)	immunohistochemistry	
(Roncalli et al. 1998), Italy	74	Amplified and over expressed (31%)	FISH, immunohistochemistry	
(Shamma et al. 1998), Japan	66	Over expressed (42%)	immunohistochemistry	
(Inomata et al. 1998.), Japan	45	Over expressed (31.2%)	Southern, northern, western blotting & immunohistochemistry	
(R Chetty & S. Chetty 1997,), SA	80	Over expressed (29%)	immunohistochemistry	
(Sheyn et al. 1997), USA	20	Amplified (65%), over expressed (75%)	FISH, immunohistochemistry	Correlation between amplification and over expression
(J. Zheng et al. 1996), China	21	Amplified and Over expressed (57%)	Northern blot, immunohistochemistry	
(Shinozaki et al. 1996), Japan	122	Amplified (23%)	Slot blot	
(Kitahara et al. 1996), Japan	8 cell lines	Over expressed (67.5%)	?	Could only obtain abstract
(H Nakagawa et al. 1995), USA	?	Over expressed (50%)	?	Could only obtain abstract
(W Jiang et al. 1992), USA	20	Amplified (25%)	?	Could only obtain abstract
(W Jiang et al. 1993), USA	50	Amplified and over expressed (32%)	immunohistochemistry	
(De-Chen Lin et al. 2010), China	148	Over expressed and amplified (45.5%)	Immunohistochemistry, array CGH	
C-MYC				
(J. He et al. 1995), China	41	Amplified (39%)	PCR	
(H. Arai et al. 2003), Japan	20	Amplified (45%)	Array CGH	
(S. Miyazaki et al. 1992.), Japan	20	Expression not significant	Northern blot	
(X.-P. Huang et al. 2006), China	46	Amplified (100%) Over expression (8.6%)	FISH, immunohistochemistry, RT-PCR	C-*MYC* not target of amplification
TUMOUR SUPPRESSORS				
TP53				

Table 1.1 Summary of studies done on molecular markers in OSCC

(Wagata et al. 1993), Japan	32	Mutations (47%)	PCR, sequencing	
(von Brevern et al. 1996), Germany	62	Mutation and nuclear over expression (88%)	PCR, immunohistochemistry	
(W. P. Bennett et al. 1997), USA	29	Mutations (69%)	PCR, sequencing	
(Hamada et al. 1996), Japan	25			
(K Y Lam, S W Tsao, et al. 1997), China	42	Over expressed (67%)	immunohistochemistry	
(K Y Lam et al. 1995), China	99	Over expressed (50%)	immunohistochemistry	
(Casson et al. 1998), UK	61	Over expressed (39%)		
(Montesano et al. 1996),				
(Muro et al. 1996), Japan	18	Over expressed (56%)	immunohistochemistry	
(Seta et al. 1998), Japan	15	Over expressed (33%)	immunohistochemistry	
(Matsha et al. 2007), SA	114	Over expressed (70%)	PCR	
(Vos et al. 2003), SA	74	2 mutations (flanking regions to exon 5-8)	PCR, sequencing	
(R Chetty & Simelane 1999), SA	50	Over expressed (42%)	immunohistochemistry	
(Gamieldien et al. 1998), SA	76	Mutations (17%)	PCR, HEX SSCP, sequencing	
(Jaskiewicz & De Groot 1994), SA	155		immunohistochemistry	
CDKN2A				
(Taghavi et al. 2010), Iran	50	Methylation (62%)	Methylation specific PCR	
(Salam et al. 2009), India	69	Methylation (67%)	Methylation specific PCR, Immunoblotting	Correlated with loss of expression
(Jianming Wang et al. 2008), China	125	Methylation (88%)	Methylation specific PCR	
(Fukuoka et al. 2006.), Japan	35	Methylation (80%)	Methylation specific PCR	

Table 1.1 Summary of studies done on molecular markers in OSCC

(Gamieldien et al. 1998), SA	76	Mutations (32%)	Sequencing, HEX SSCP	
APC				
(Zare et al. 2009), Iran	45	Methylated (44.4%)	Methylation specific PCR	Poor survival
(Y. T. Kim et al. 2009), Korea	50	Methylated (46%)	Methylation specific PCR	Good prognosis
(H. Peng et al. 2009), China	199	Low expression (30.4%)	Immunohistochemistry	
(Kriebashne S Nair et al. 2006), SA	100	LOH (62.48%), MSI (41.27%)	MSI and LOH analysis	
FHIT				
(X. Q. Guo et al. 2007), China	95	No expression (81.62%)	Methylation specific PCR, immunohistochemistry	
(Morita et al. 2006.), Japan	55	Loss of protein	Immunohistochemistry	
(Kuroki et al. 2003), USA	47	Methylated (45%)	Methylation specific PCR, LOH	
(Noguchi et al. 2003), Japan	36	Deleted (69.41%)	Methylation specific PCR	
(H. Tanaka et al. 1998), Japan	35	Methylated (14%)	Methylation specific PCR	
(M Mori et al. 2000), Japan	46	No expression (70%)	Immunohistochemistry	
(Menin et al. 2000), Italy	21	LOH (80%), aberrant proteins (77%)	LOH and RT-PCR	

1.3 Chromosomal aberrations and OSCC

OSCC is reported to be a genetically complex cancer (Carneiro, et al., 2008). The high number of genetic aberrations detected is likely due to genetic instability, which could be due to one or more of several mechanisms that lead to complex chromosomal aberrations.

It is widely accepted that recurrent chromosomal aberrations seen in malignancies pinpoint the genes involved in the initiation and progression of cancer (Felix Mitelman et al., 2007). Gene fusions in solid tumours have not been studied in depth previously due to the difficulties in obtaining metaphase chromosomes from solid tumours, but increased access to better analysis tools has shown that translocations do occur more frequently in solid tumours than previously thought and contribute to cancer morbidity at an approximate rate of about 17% (Mitelman et al., 2007). In addition to chromosomal rearrangements, amplifications and deletions are frequent events in cancer. These events often lead to over expression of oncogenes and inactivation of tumour suppressor genes.

Mechanisms of chromosomal aberrations

Genetic instability is defined as an excess of structural or numerical chromosomal abnormalities. Intra-tumour heterogeneity is common amongst cancers, probably as a result of genetic instability due to the dysfunction of genes responsible for maintaining the integrity of the genome (D Gisselsson et al., 2000).

Chromosomal rearrangements can be generated by the formation of double-stranded breaks (DSB) in the DNA, which are normally repaired by two main DNA repair mechanisms, non homologous end joining (NHEJ) and homologous recombination (HR) (Ferguson & Alt 2001). Double stranded breaks are usually initiated by various DNA damaging agents such

as, ionising radiation and oxidative reactive species. Defects in the repair pathways can lead to rearrangements, LOH and deletions (Ferguson & Alt, 2001).

NEJH involves the maintenance of DNA ends in close proximity, removal of 5' and 3' overhangs, fill in synthesis and ligation of the DNA strands (Lieber et al., 2003). NHEJ may create a perfect repair of a DSB or there can be loss of material surrounding the lesion, which often occurs at the sites of translocation break points (Lieber et al. 2003). HR on the other hand, can fill in gaps of DNA by copying missing information from a sister chromatid or homologous chromosomes, avoiding the formation of deletion (Lieber et al. 2003). This process increases the chance of generating LOH by gene conversion (base mismatch repair during recombination) or gene rearrangements by filling in DNA sequences that are similar but are not from homologous chromosomes (Ferguson & Alt, 2001). There is interplay between these pathways in maintaining the genome integrity and preventing tumourigenesis, with each of these pathways predominating in different phases of the cell cycle (Ferguson & Alt, 2001).

Animal models have shown that when there is a deficiency of one or more of the proteins required for NHEJ, radiation sensitivity occurred mainly in the G1 and early S-phase while *RAD54* inactivation, leading to disruption of the HR pathway, affects the survival of cells irradiated in late S phase and G2 (Takata et al. 1998; Sonoda et al. 1998). In addition, deletion of *RAD51* lead to cells in the G2/M phase with numerous single chromatid breaks and lack of HR repair prior to DNA replication (Takata et al., 1998; Sonoda et al., 1998). The loss of these and other enzymes in the NHEJ and HR pathways can lead to higher levels of chromosome aberrations (Takata et al., 1998).

Another mechanism involved in genetic instability and the major mechanism for the formation of amplifications is the breakage-fusion bridge cycle. This mechanism was first described by Barbara McClintock in 1938 and 1940 and represents one of the most described mechanisms of variability in chromosome structure (McClintock, 1941). Breakage-fusion-bridge (BFB) cycles can be described as a break forming within a single chromatid of a chromosome and a subsequent fusion forming between the two sister chromatids during the anaphase stage of cell division (McClintock 1941). This results in the inability for these two chromatids to separate during the following mitotic anaphase and a bridge is formed as the two centromeres migrate to opposite poles of the nuclei (Figure 1.1). Breakage will occur at the point of tension (McClintock, 1941). This results again in broken ends resulting in fusions and bridge formation in subsequent anaphases. The position of the breakage during successive cycles can vary, thus giving rise to the heterogeneity of aberrations seen in cancer cells (Selvarajah et al., 2006; D Gisselsson et al., 2000).

Gisselsson et al (2000) showed that BFB's were frequently detected in tumours with highly heterogeneous abnormalities i.e. unspecific patterns of chromosome rearrangements while BFB's were less frequent in tumours with more specific and recurrent patterns of rearrangement. This suggests that there are possibly several mechanisms of development of chromosomal aberrations in solid tumours (Gisselsson et al., 2000). In healthy cells, those carrying chromosomal damage are usually forced into apoptosis. In tumours, the mechanisms responsible for DNA damage checkpoint, repair and induction of apoptosis are often defective (Bayani et al., 2007).

BFBs are also associated with telomere length, where excessive shortening of telomeres and loss of telomere capping function lead to BFB cycles (Bayani et al., 2007; David Gisselsson,

2005). BFB cycles definitely contribute to genomic instability and promote the formation of complex rearrangements and homogenously staining regions (HSRs) (A. W. I. Lo et al., 2002; Bayani et al., 2007). Lo et al (2002) demonstrated that BFB formation was associated with amplification of small regions and can continue into at least 20 cycles of cell division (Lo et al., 2002).

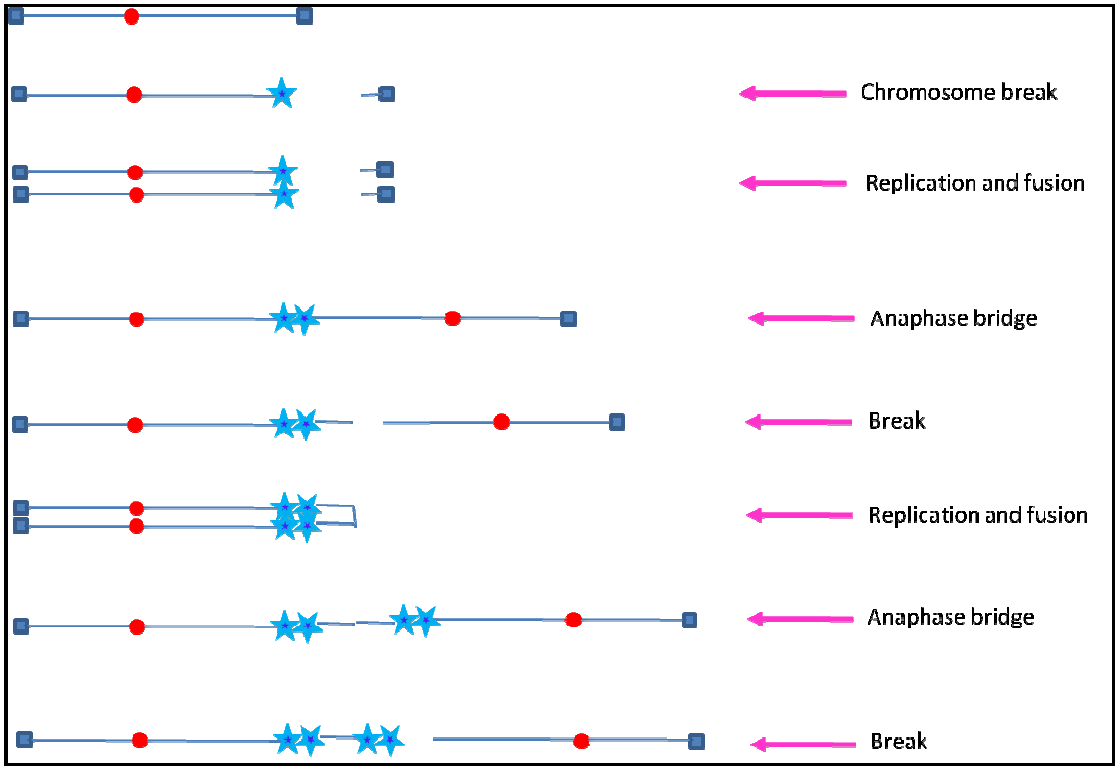

Figure 1.1. Diagrammatic representation of the BFB cycle adapted from Lo et al (2002). The blue squares indicate the telomeres and red dots, the centromeres. The stars indicate the region of initial break, which is copied and amplified in subsequent cell divisions.

Another mechanism associated with genomic instability is the genomic architecture itself, where GC rich regions have been linked to translocations and AT rich regions with deletion events (Abeysinghe et al., 2003). Frequent translocation events at pericentric regions would also suggest that repetitive sequences facilitate translocation events with the sequence

homology of these regions across chromosomes acting as sites to receive and donate in translocation events and jumping translocations (Bayani et al., 2007).

Fragile sites are chromosomal regions expressed as gaps or breaks appearing under specific culture conditions such as folate deficiency or aphidicolin treatment (Yunis & Soreng, 1984). They are prone to breakage under replication stress and exposure to clastogenic compounds. They are also prone to translocation events, deletions and viral integration and have been shown to coincide with amplicon boundaries (Myllykangas et al., 2007). Common fragile sites include FRA3B, FRA6E, FRA8C and FRA16D, which harbour the tumour suppressor genes *FHIT*, *WWOX, PACRG* and the oncogene C-*MYC* respectively, all known to be potentially involved in cancers (Toma et al., 2008). Amplification patterns seem to correspond to tissue-specific lineage, however tissues exposed to similar environmental conditions have similar patterns of amplification suggesting their link to environmental carcinogen exposure (Myllykangas et al., 2007).

Chromosomal aberrations detected in OSCC

A whole genome view would allow patterns of copy number change to be detected in tumours and since the detection of amplifications and deletions could help identify the key genes involved in the carcinogenic process, a newer approach to study cancers has evolved. The methodology of comparative genomic hybridisation (CGH) was developed for this purpose. In this method, hybridising labelled tumour DNA comparatively with normal labelled DNA to normal metaphases allows for detection of copy number changes at a resolution of 10Mb (Albertson & Pinkel, 2003). Large regions of deletions or amplifications can be detected but smaller regions (< 10Mb) will be missed by conventional CGH. Conceptually based on the same principle as CGH, array CGH has been developed. In array

CGH, tumour and control DNA are co-hybridised to slides printed with either BACs (Bacterial Artificial Chromosomes) or oligonucleotide sequences, either covering the whole genome or spanning it at regular intervals (Pinkel & Albertson, 2005). Depending upon the type of 'array CGH', the resolution can range from 1 Megabase to a few hundred bases. Small regions of gene imbalances can be detected and the delineation of critical regions of deletion and/or amplification can be refined (Pinkel & Albertson, 2005). SNP (single nucleotide polymorphism) array technology has provided a platform, which can be used to analyse the whole genome at high resolution and allow DNA copy number analysis by evaluating the signal intensities over a given genomic window for cancer samples by comparing them to the signal intensities of a normal reference set of samples (G. R. Bignell et al., 2004). The Affymetrix® 500K assay consists of two microarray chips covering approximately 500 000 SNPs. The features are 5µm in size and each SNP has 24 to 40 probes (matched and mismatched probes for the sense and anti-sense strands) at a mean spacing of 5.8kb. This technique is now widely used to identify common DNA copy number aberrations across samples and identify genes that may be targeted by these amplicons or deletions.

Regarding whole genome analysis of oesophageal cancer, to date, 15 studies have been performed analysing copy number changes, one of which was performed in the South African population using conventional CGH. Eight of these studies used conventional CGH and the other 7 have utilised various types of higher resolution array CGH from 10K SNP arrays, 32K BAC arrays to the highest resolution of 250K SNP arrays. Two studies scrutinised specific gene loci i.e. 766 known cancer genes and another specifically investigated 386 markers on chromosome 3p. There is much variation in the results from study to study,

possibly due to the varying resolutions of the techniques or it may be that different geographic, ethnic or etiologic factors that contribute to this variation. One theory, suggested by Chattopadhyay et al (2010), was that the disease arises from mutations in genes involved in maintaining genetic stability. As a result, genomic instability increases as the tumours progress and different clones expand by selection of those cells with abnormalities that provide a growth advantage (Chattopadhyay et al., 2010).

The most common aberrations across these studies were amplicons affecting chromosomes 1p36.32-p36.13, 2q31-32, 3q28, 5p15.2, 6p25.3, 7p14, 8q24, 11q12.3, 20p11.1, 20q13.1 and deletions affecting 1p, 4p14-p15.3, 3p14.2, 3p26.3, 3p24.2, 5q33, 8p23.2, 9p21.3 and 18q21.2-22 (C. C. Yen et al., 2001; C. C. Yen et al., 2005; Xu Zhang et al., 2008; J. Chen et al., 2008; Y. R. Qin et al., 2008; Du Plessis et al., 1999; N. Hu et al., 2006; Carneiro, et al., 2008; Chattopadhyay et al., 2010). Aberrations that are consistently found include 3q amplification (8 studies) and 3p deletion (8 studies), which is not exclusive to OSCC but also common to other epithelial cancers. Abnormalities of chromosome 3 may therefore be pivotal in the carcinogenesis of epithelial tissues (Chattopadhyay et al., 2010).

Chromosome 3q26 amplification is the most commonly detected 3q aberration in OSCC. The *SOX2* gene, on chromosome 3q26, encodes a transcription factor and is amplified in primary OSCC's at a rate of 15% (Bass et al., 2009; Gen et al., 2010) but over expression is detected in 70% of tumours and is found to promote proliferation (Gen et al., 2010), suggesting that *SOX2* expression is not only activated by gene amplification and other genes may be the primary targets of 3q26 amplification.

Chromosome 8q24 amplification and 8p deletions are also consistently detected in 8 different studies, highlighting these regions as important for OSCC carcinogenesis.

Chromosome 8q24 amplification may target the C-*MYC* oncogene. However other genes on 8q24 are proposed to be the targets for amplification due to a lack of over expression of C-*MYC* in 59 cases analysed (X. P. Huang et al., 2006). Huang et al (2006) suggested that the *FAM84B* gene was the probable target since 66% of the cases had over expression of this gene, which correlated with gene amplification (X. P. Huang et al., 2006).

One study analysed both copy number aberrations and the expression of genes affected by copy number aberrations but the analysis was restricted to amplicons (T. Sugimoto et al., 2007). The study was performed in 10 OSCC cell lines using oligo-array CGH and Acegene human oligo Chip 30K for the expression analysis. Correlations were detected for 97 genes, (i.e. the DNA copy number correlated with expression of genes within the aberrant region), in the top ranked region, namely 11q12-14. Twenty four of these genes were selected for verification of amplification and expression with quantitative PCR. Twenty of the genes in the 11q12-14 region are significantly over expressed (*SSRP1*, *MS4A7*, *LOC51035*, *HRASLS2*, *LOC144097*, *MEN1, EHD1, CFL1, BANF1, SF3B2, SUV420H1, FGF19, PPFIA1, PDE2A, STARD10, RAB6A, WDR71, NEU3, SERPINH1* and *PCF11*), correlating with gene amplification (T. Sugimoto et al., 2007).

The variation from these studies makes it difficult to identify genes that may be targets of the aberrations and apart from one (Chattopadhyay et al., 2010), these studies did not attempt to link the genes in aberrant regions to common biological pathways that may be implicated in OSCC. Recurrent chromosomal amplifications and deletions need clarification in OSCC with the varying results that have been highlighted in the literature and specifically, a high resolution analysis of such aberrations in SA OSCC could be valuable in assisting in the

unravelling of OSCC pathogenesis through identification of affected genes and their specific pathways.

1.4 Project Rationale

Hypotheses

Previous work on five oesophageal squamous cancer cell lines developed in SA (Bey et al., 1976; R Veale, 1989) were analysed for the presence of deletions at fragile sites (Willem et al., 2006) and for the presence of common translocations using Multicolour fluorescence *in situ* hybridisation (M-FISH) (J Brown, MSc thesis, WITS, 2005). Chromosomal aberrations were detected across cell lines, specifically a t(1;3), t(1;15), t(1;14), t(3;22), 7q deletion, and 20q deletion, which could target common breakpoints on chromosome 1p, 3p, 7q and 20q. These common chromosomal rearrangements may target novel genes involved in carcinogenesis and further investigation of these breakpoints is required. As previously mentioned, only one study analysing copy number patterns in OSCC has been performed in SA and this study used low resolution conventional CGH. There is a need to look at common copy number aberrations at high resolution in OSCC patients in SA.

It is hypothesised that breakpoints detected in the cell lines and common regions of amplifications and deletions in cell lines and patient samples will pin point genes and pathways involved in OSCC carcinogenesis.

Aim

The project aim is to map common rearrangements detected in five SA OSCC cell lines and establish the gene copy number profile of these 5 cell lines and 50 OSCC clinical specimens in order to identify the most common loci undergoing copy number changes.

Specific Objectives

1. To map the common 1p11 and 3p11 breakpoints in the OSCC cell lines, which may reveal a common oncogene/s non-randomly involved in OSCC.
2. To establish the DNA copy number profiles of the five OSCC cell lines and combine the results with the M-FISH data to characterise their cytogenetics and molecular profile.
3. To assess the expression levels of candidate genes in the 5 OSCC cell lines to verify if copy number amplification/deletion correlates with expression levels of genes at these loci.
4. To compare and identify regions of amplification and deletion in the genome of 50 clinical specimens. Compare the patterns of abnormality to that of the 5 OSCC cell lines and to the findings of other studies.
5. To validate the copy number microarray results using an alternative method. Here, being fluorescence in situ hybridisation (FISH).

The project is broken into 3 major chapters; chapter 2 describes the work done on the 5 cell lines for objectives 1 and 2 combining molecular cytogenetic techniques. Chapter 3 deals with objective 3, the expression of some candidate genes identified in the cell line study and the fourth chapter describes objectives 4 and 5, the in depth copy number analysis in a patient cohort. These chapters are linked together in the final concluding chapter.

CHAPTER 2

Characterisation of common chromosomal aberrations in 5 OSCC cell lines

2.1 Introduction

It is widely accepted that recurrent chromosomal breakpoints in malignancies often pinpoint genes involved in the initiation and progression of cancer (Felix Mitelman et al., 2007). A major limitation to assess the chromosome complement in OSCC specimens is the difficulty to obtain metaphases from fresh tumours and established cell lines provide a unique resource for such investigations. The number of OSCC cell lines that have been reported to date remains limited and they were all established in China (Y. C. Hu et al., 2002; Y. P. Wu et al., 2006; S. Xiao et al., 1991) and Japan (K. Tada et al., 2000). These cell lines have been investigated with one or several low resolution molecular cytogenetic techniques including cytogenetics, fluorescence *in situ* hybridization (FISH), multicolor FISH or spectral karyotyping (SKY) and conventional comparative genomic hybridization (CGH). Various clonal aberrations have been identified and the most common abnormalities across studies involved over representation of chromosomes 1q, 3q, 11q and 8q as well as breakpoints in the centromeric or near centromeric regions of chromosomes 1, 3 and 8 (Y. C. Hu et al., 2002; Y. P. Wu et al., 2006; S. Xiao et al., 1991; H. Zhang et al., 2006).

Five cell lines have previously been established from South African OSCC patients (Bey et al., 1976; R Veale, 1989) but apart from cell line SNO, which was karyotyped, these cell lines were never characterized for their genetic constitution. Conventional cytogenetics, fluorescence *in situ* hybridization (FISH), and multicolor fluorescence *in situ* hybridization (M-FISH) were used to identify common chromosome structural abnormalities in these five cell lines. Affymetrix 250K SNP arrays were performed to investigate DNA copy number changes and interrogate the common aberrations previously detected by M-FISH (results from MSc work) and conventional cytogenetics. Here clonal aberrations shared by these cell

lines and highlighted preferential targets for chromosomal rearrangements and copy number changes are described. These are the first OSCC cell lines from Africa genetically characterized.

2.2 Materials and Methods

Cell lines

Five oesophageal carcinoma cell lines were isolated from moderately differentiated OSCCs as previously described (Bey et al., 1976; R Veale, 1989). These cell lines are designated WHCO1, WHCO3, WHCO5, WHCO6 and SNO (Table 2.1). The five cell lines were cultured at 37°C, 5% CO_2 in Dulbecco's Modified Eagles medium (DME): HAMS F12 (3:1) (GIBCO®) containing 10% fetal calf serum (FCS) (GIBCO®), and 100ug/ml streptomycin (ICN) and 100 IU/ml of penicillin (ICN).

Table 2.1. Cell lines and their patho-clinical information

CELL LINE	SEX	RACE	AGE	GRADE	SOURCE	YEAR	Passage
WHCO1	M	Black	47	Moderately differentiated SSC	Professor R Veale (WITS University)	1986	141
WHCO3	M	Black	71	Moderately differentiated SSC	Professor R Veale	1987	51
WHCO5	M	Black	47	Moderately differentiated SSC	Professor R Veale	1988	66
WHCO6	M	Black	39	Moderately differentiated SSC	Professor R Veale	1989	77
SNO	M	Black	62	Well differentiated SSC	Bey et al. (Virus Cancer Research Unit, Johannesburg)	1976	Unknown

Isolation of cell lines for metaphase preparation

The cell lines were cultured as above. When the cultures were half confluent, the cells were incubated with a final concentration of 0.44µg/ml of Karyomax® Colcemid® (Invitrogen Corporation) for 4 hours. Cells were harvested using Nunc radiation sterilised scrapers (Thermo Fisher Scientific Incorporated) and poured into tubes for centrifugation at 900rpm for 10 minutes (Eppendorff centrifuge 5810R, A-4-62 rotor). The supernatant was removed and 8ml of 0.075M KCl and 1ml of Foetal Calf Serum (GIBCO®) was added. The cells were incubated for 30 minutes at 37°C. Five drops of fixative (3:1 methanol: acetic acid) was added prior to centrifugation at 900rpm for 10 minutes (Eppendorff centrifuge 5810R, A-4-62 rotor). The supernatant was discarded and the cells re-suspended in 5ml of fixative and placed at -20°C for 1 hour. Two more fixative changes were performed and the cells were transferred to 1.5ml centrifuge tubes. Subsequent fixative changes were performed and centrifugation was performed at 6000rpm for 2 minutes (Eppendorff centrifuge 5810R, CL021 rotor). This was done until the pellets were clean in appearance. The cells were stored at -20°C until slides were prepared either for FISH or conventional cytogenetics.

Isolation of cell lines for DNA

When the cultures were confluent, the cells were isolated for DNA by scraping the cells from the plate and centrifuging at 900rpm (Eppendorff centrifuge 5810R, A-4-62 rotor) for 10 minutes to remove the supernatant. The cells were washed in 1X PBS (Appendix A) and centrifuged again at 1500rpm for 10 minutes (Eppendorff centrifuge 5810R, A-4-62 rotor). New PBS was added to cover the cells and the cells were then frozen at -20°C until DNA isolation.

Conventional Cytogenetics

Slides were prepared for conventional cytogenetics by dropping 30µl of the cell suspension onto clean glass slides flooded with 3:1 Methanol: Acetic acid. The slides were steamed for 30 seconds to spread the metaphase chromosomes sufficiently. The slides were dehydrated in an alcohol series (70, 90 and 100%) for 5 minutes each. The slides were aged overnight in an oven at 60°C in preparation for GTW banding. Slides were placed in trypsin (Appendix A) for 1 minute and stained with Wright's stain (Appendix A), rinsed with water and air dried. CytoVision 3.0 software (Applied Imaging Corporation San Jose, California, USA) was used for image acquisition and karyotypic analysis.

Preparation of metaphase slides for Fluorescence in situ hybridisation (FISH)

Slides were prepared, for metaphase analysis by FISH, by standard procedure. Slides were prepared a day prior to FISH experiments by dropping three drops of the cell suspension onto a glass slide and steaming for an appropriate amount of time to spread the metaphase chromosomes appropriately. The slides were dehydrated in an alcohol series 70, 90 and 100% ethanol for 5 minutes each and then aged overnight at room temperature.

FISH procedures

FISH was performed on metaphase chromosomes from all cell lines in order to confirm or refine translocation derivatives' breakpoints and composition. Probes specific for the short and long arms of chromosome 3, the short arm of chromosome 1 and the long arm of chromosome 22 (Qbiogene) as well as the Vysis ® Cep 3 Alpha and Cep 1 Alpha SpectrumOrange probes (Abbott Molecular Inc) were hybridized to further map translocation breakpoints in all cell lines. The Vysis® LSI IGH and Vysis® LSI RARA, both dual colour break apart rearrangement probes (Abbott Molecular Inc) were used to confirm the

involvement of chromosomes 14 and 15 in translocation derivatives seen in cell lines WHCO1 and WHCO3 respectively. FISH was also performed on interphase nuclei from all the cell lines using the Vysis® LSI C-MYC SpectrumOrange probe (Abbott Molecular Inc) and the Vysis® LSI t(11;14) dual colour probe (Abbott Molecular Inc). These probes target the *MYC* gene (Spectrum Orange) on 8q24 and the *CCND1* and *FGF4* genes (Spectrum Orange) on chromosome 11 respectively and were used to confirm SNP array copy number results. A hundred interphase nuclei were analyzed in all the cell lines. All FISH experiments were performed according to the manufacturer's instructions and analyzed on an Olympus BX41 fluorescent microscope equipped with appropriate fluorescence filters. The Genus™ CytoVision 3.0 software (Applied Imaging Corporation) was used for image acquisition and analysis.

The slides were denatured at 76˚C for 5 minutes in denaturing buffer (Appendix A) and then dehydrated in ice-cold 70, 90 and 100% ethanol for 5 minutes each. The slides were air-dried and the 10µl of probe was applied to the slides. Hybridisation was overnight or at least 15 hours at 37˚C. Washing was done at 42˚C, three washes in 50% formamide (Appendix A) for 10 minutes each, one wash in 2X SSC (Appendix A) for 10 minutes and one wash in 2X SSC with Tween® 20 (Appendix A) for 5 minutes. The slides were then stained in DAPI (Appendix A) for 15 minutes and washed in DAPI wash solution (Appendix A) for 2 minutes. The slides were mounted with Vectashield (Vecta Laboratories) and a coverslip. Analysis was performed on a fluorescent microscope and all images were captured using the Genus™ CytoVision 3.0 program from Applied Imaging.

BAC probe development and FISH analysis for the EPHA3 gene deletions

This section of the work was done by an honours student supervised by myself.

The bacterial artificial chromosome is based on the fertility plasmid of bacteria. This cloning vector is an artificial, self-replicating chromosome, which can accept an insert of about 300kb. The vector carries an antibiotic resistance marker allowing for the selection of those clones carrying the insert. *E. Coli* serves as the host cells for the generation of BAC clones. The BAC clones chosen for assessment of *EPHA3* deletions are shown in figure 2.1 (Obtained from BACPAC resource centre, Children's Hospital Oakland Research Institute, CA, USA). The BAC clone, RP11-598N21, maps to exons 1, 2 and 3. The BAC clone, RP11-535I05 maps telomeric to the *EPHA3* gene and the RP11-547K2 clone maps from exons 11-17. BAC clone RP11-784B9 was mapped to *EPHA3,* however, in these experiments, the clone we received mapped to a distant locus on chromosome 3p and was used as a 3p internal control for hybridisation of RP11-598N21 (Figure 2.2).

Two hundred millilitres of BAC growth medium, LB top agar (Appendix A), was aliquoted into autoclaved flasks. 12.5µg/ml of chloramphenicol (Sigma®) was added to the growth medium. Five millilitres of the 200ml volume was aliquoted into 50ml Nunc® tubes for inoculation with BACs from frozen glycerol stock cultures (Appendix A). The inoculated medium was incubated at 37˚C, orbiting at 200rpm for 2-6 hours. When the medium was turbid, the cultures were transferred to the flasks containing 195ml of medium. If they were not turbid they were left overnight. The 200ml cultures were incubated at 37˚C, orbiting at 180rpm overnight.

The cultures were harvested by aliquoting into 50ml Nunc® tubes and centrifuging at 5000 rpm for 15 minutes (Eppendorff centrifuge 5810R, F34-6-38 rotor). DNA extraction was done using the Qiagen® Plasmid Purification Kit. Utilisation of the QIAfilter™ Plasmid Midi procedure resulted in very pure, protein and RNA free insert DNA with no chromosomal

DNA. The protocol for the kit was strictly followed. The lysate was cleared through the Qiagen®-tip 100. Plasmid DNA is bound to the anion- exchange resin, while RNA, proteins and other impurities are removed by low salt washing. The high salt buffers then elute the DNA. The DNA is precipitated to remove the salt by adding isopropanol (Merck) and centrifuging at 12000xg for 30 minutes at 4˚C (Eppendorff centrifuge 5810R, CL021 rotor). Washing with 70% ethanol followed and then centrifugation at 12000xg for 10 minutes at 4˚ C (Eppendorff centrifuge 5810R, CL021 rotor). The pellets were air dried and re-suspended in an appropriate volume of 1X TE buffer (Promega) for the size of the DNA pellet.

The concentration and purity of DNA extracted needs to be clarified for the efficient labelling of DNA in probe production. This was achieved by 2% agarose (Bioline) gel electrophoresis. The spectrophotometer was found to be inaccurate for BAC DNA estimation, the readings given, were 2 fold that given by estimation on gel electrophoresis and so the latter was used instead.

To estimate the concentration of DNA, the sample DNA was compared to a control DNA with known concentration. Lambda DNA (Sigma®) was diluted to a 1/10 dilution, which was approximately 17.5ng/µl of DNA. 1µl of BAC DNA was loaded into a 2% agarose gel along with 17.5ng, 35ng and 70ng of Lambda DNA. The gel was run for 10 minutes at 100V. The electrophoresis tank contained 1X TAE buffer (Appendix A) in which the agarose had also been dissolved. DNA integrity could be determined by the presence of a high molecular weight band of DNA or a smear, in which case the DNA would be degraded.

The BAC DNA was directly labelled using the nick translation method. BAC clones, RP11-535IO5 and RP11-784B9, were labelled with SpectrumOrange-dUTP (Abbott Laboratories). The BAC clones RP11-598N21 and RP11-547K2 were labelled with SpectrumGreen-dUTP

(Abbott Laboratories). The reaction was made up to a final volume of 100µl, which consisted of final concentrations of 1x nick translation buffer (Appendix A), 0.01M β-mercaptoethanol (BDH), 0.8X nucleotide stock, which contained the fluorescent nucleotides (the fluorescent nucleotides were kept in a 1:1 ratio to thymine nucleotides to minimise steric hindrance (Appendix A), 210u/ml of DNA Polymerase (Promega), 1µl of a 3µl/1000µl dilution of DNase I solution (Appendix A) and 1µg/100µl of DNA. The reaction was incubated at 15˚C for 2 hours in the Eppendorf Mastercycler. The optimum probe size is 200-500bp for efficient penetration of the nucleus and hybridisation. The probe size and fluorescence incorporation was evaluated by 2% agarose gel electrophoresis with a 100bp ladder (Fermentas Ruler™) (Appendix A).

The probe was purified to remove unincorporated nucleotides by co-precipitation with 20µg of Cot1-Human DNA (Roche). To precipitate the labelled DNA, 1/10 volumes of 3M-sodium acetate and 2, 5 volumes of ice-cold 100% ethanol were added and then incubated at -70°C for 30 minutes and centrifuging at 13000xg for 30 minutes (Eppendorff centrifuge 5810R, CL021 rotor). The precipitated DNA was then washed in 70% ethanol to remove salts by centrifuging at 13000xg for 10 minutes (Eppendorff centrifuge 5810R, CL021 rotor). The probes were air dried and re-suspended in hybridisation buffer (Appendix A), usually 160ng of DNA per 10µl of hybridisation buffer. Ten microlitres of probe was hybridised to slides, with denaturing hybridisation and washing all done as above for FISH procedures. One hundred cells per cell line were analysed. Each probe was hybridised to normal metaphase and interphase cells to confirm the correct chromosomal location and interphase pattern (Figure 2.2 and 2.3).

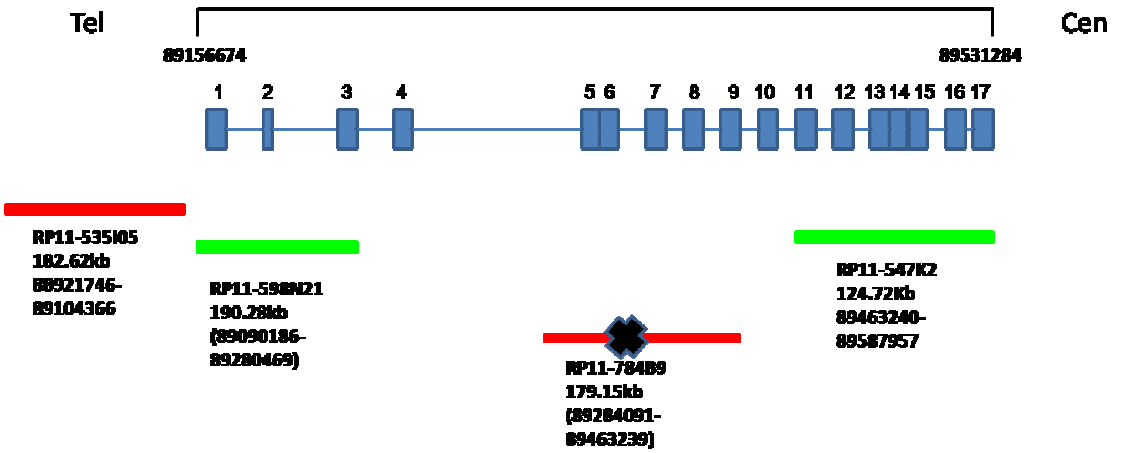

Figure 2.1 Schematic representations of the *EPHA3* gene locus and the BAC clones selected for FISH experiments. The red and green delineate the colour of the fluorescent label for each clone. The blue blocks represent each of the 17 exons that make up the *EPHA3* gene. The black cross indicates that this clone did not map to this location.

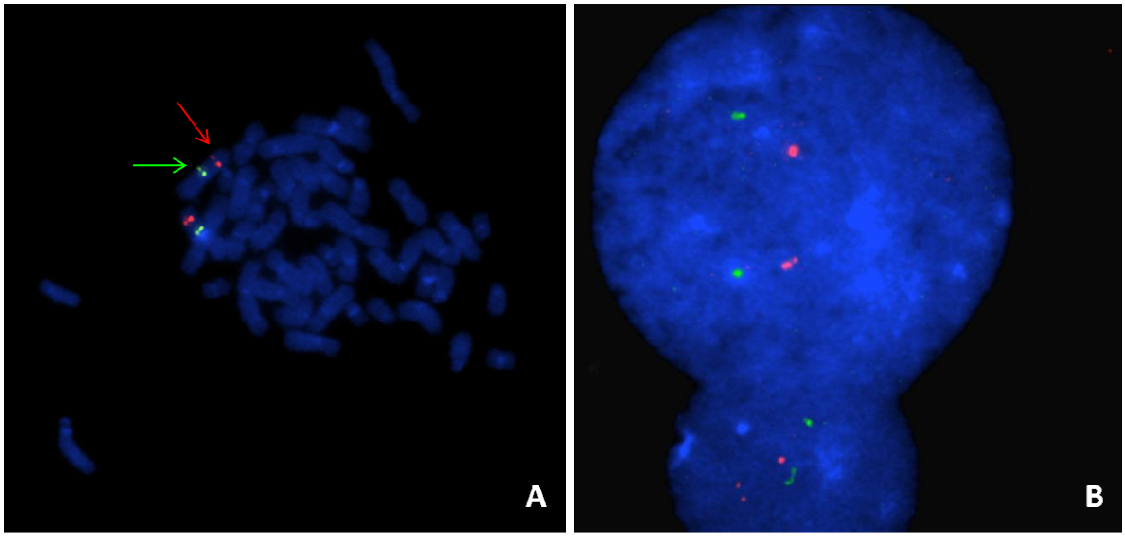

Figure 2.2. Normal metaphase and interphase FISH for RP11-598N21 and RP11-784B9. DAPI stained metaphase image showing the confirmed position of RP11-598N21 at 3p11.2 (green arrow) (A). The red arrow indicates RP11-784B9, which is not positioned in 3p11.2 but on a distant band on chromosome 3p**.** DAPI stained interphase nuclei (B) showing the pattern of hybridisation for RP11-598N21 (green) and RP11-784B9 (red) in normal nuclei.

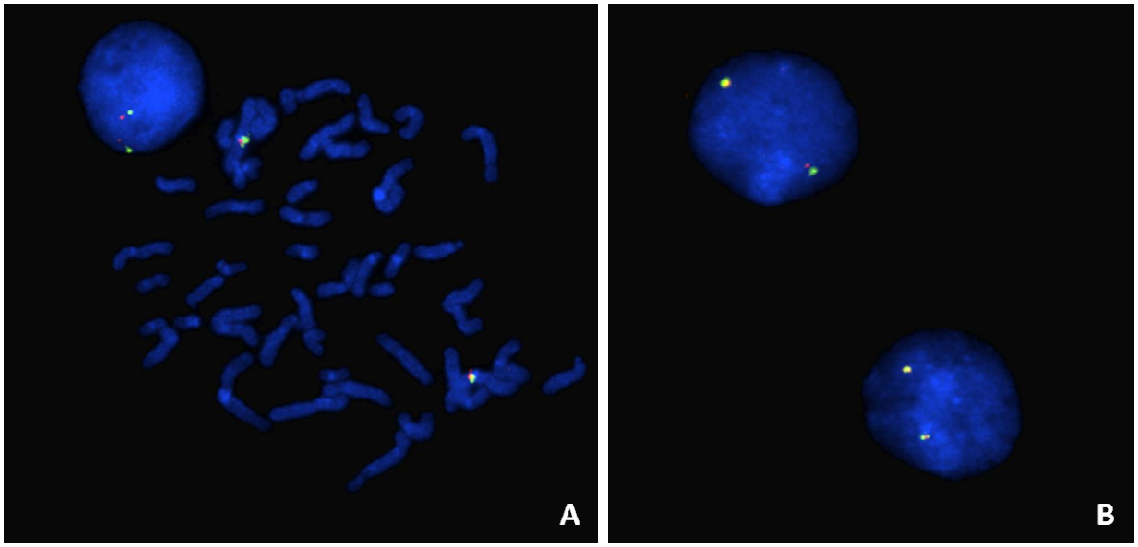

Figure 2.3. Normal metaphase and interphase FISH for RP11535I05 and RP11-547K2. DAPI stained metaphase image, showing the location of RP11-535I05 (red) and RP11-547K2 (green) clones at 3p11.2 (A). DAPI stained interphase nuclei (B) showing the FISH pattern of two fusion (yellow) signals in normal nuclei.

DNA isolation

The cells were thawed and centrifuged at 1500rpm for 10 minutes (Eppendorff centrifuge, A-4-62 rotor), the supernatant was removed and 4ml of lysis buffer (Appendix A) was added to the pellets. One hundred micro litres of Proteinase K (10mg/ml) (Roche Diagnostics) was added to the samples and incubated at 55°C until homogenous. Four millilitres of phenol (Sigma®) and chloroform (Fluka® Analytical) were added and the samples were vortexed and spun at 6000rpm for 10 minutes (Eppendorff centrifuge 5810R, F34-6-38 rotor). The aqueous phase was transferred to a new tube and 2ml of phenol and chloroform were added to the samples. They were vortexed before centrifugation at 6000rpm for 10 minutes (Eppendorff centrifuge 5810R, F34-6-38 rotor). The aqueous phase was again transferred to a new tube and 4ml of chloroform was added. The samples were vortexed and centrifuged at 6000rpm for 10 minutes. The aqueous phase was transferred to a new tube and 2.5 volumes of ice-cold, 100% ethanol was added to each sample. The samples were inverted to mix and then centrifuged at 6000rpm for 15 minutes at 4°C (Eppendorff centrifuge 5810R,

F34-6-38 rotor). The ethanol was decanted off and the pellets were washed in 200µl of 70%, ice-cold ethanol. The samples were centrifuged at 6000rpm for 10 minutes (Eppendorff centrifuge 5810R, F34-6-38 rotor). The DNA pellets were allowed to air-dry and then re-suspended in 1XTE buffer (Promega).

Assessment of DNA integrity

All the DNA samples were assessed by gel electrophoresis and run on a 2% agarose gel at 100mV for 45 minutes in 1X TAE buffer (Appendix A) to assess the DNA integrity.

DNA Quantification

The samples were quantified by spectrophotometry using the ND-1000 spectrophotometer (Nanodrop® Technologies). The A260/280 and A260/230 ratios were recorded.

Sample preparation

The DNA samples were diluted in a separate aliquot to 50ng/µl using molecular grade water (Accugene®, Lonza). They were assessed on the Nanodrop to check the concentration.

Affymetrix 250K Genotyping assay for the assessment of DNA copy number changes in the five cell lines

The 250K Nsp chips were used to assess copy number changes in the five cell lines. DNA extracted from blood of 6 normal male black individuals was used as a reference group for estimation of the copy numbers in the cell lines.

Two hundred and fifty nanograms of DNA was used and the Affymetrix® GeneChip® 500K assay manual (P/N 701930 Rev. 3) was followed strictly for the Nsp part of the assay. The assay involves a restriction enzyme digest with Nsp I (New England Biolabs® Inc), adaptor ligation, PCR amplification using a universal primer, fragmentation and labelling. The

restriction enzyme digest, PCR amplification and fragmentation require quality assessment by gel electrophoresis before proceeding to the next step (full details on the 500K assay can be seen in Appendix C). Once the samples were labelled they were mixed with a hybridisation buffer and hybridised to the chips. Hybridisation took place overnight in an Affymetrix hybridisation chamber for 18 hours. Post hybridisation washing was performed using the Affymetrix fluidics station 450. Scanning was performed by the Affymetrix® Genechip scanner 7G. Quality control for the assay is done by generating the call rates for the genotyping calls in Genotyping Console™ (Affymetrix®).

250K data analysis

The Affymetrix®, Genotyping Console™ 2.0 was used for data analysis. The algorithm used for copy number estimation was the copy number analysis tool (CNAT 4.0) program (Affymetrix®). Subsequently the data was analyzed with third party software, Genepattern (Reich et al., 2006). The signal intensities from the CEL files were normalized by the PM-MM (perfect match minus mismatch) probe intensity and invariant set normalization against the median intensity of the controls. The raw copy number was then estimated from the signal intensities of the normals and smoothed by GLAD (Gain and Loss analysis of DNA) (Hupé et al., 2004).

To determine the significant common regions of amplification and deletion (i.e. driver aberrations as opposed to random passenger aberrations) amongst the five cell lines, the GISTIC (Genomic Identification of Significant Targets in Cancer) algorithm (Beroukhim et al., 2007) was applied to the smoothed data. The algorithm first scores the regions of copy number change according to their frequency and amplitude, which indicates the likelihood of it being a driver aberration (G-score). The statistical significance of each G-score is

calculated by comparison of these scores against a null model of random aberrations. This significance is represented as a q-value (False discovery rate), which represents the likelihood that the data was generated by chance. The aberrations that occur frequently and are highly aberrant are therefore considered to most likely contain driver aberrations. The most likely locations for oncogenes or tumour suppressors are identified by calculating the minimal common regions of aberration, which are most significantly altered i.e. high amplitude change. It also distinguishes broad aberrations (can be the length of a chromosome arm) from focal aberrations. The aberrations with high G-scores and minimal q-values (less than 0.25) are more likely to contain target genes. It also identifies possible overlapping aberrations that are statistically significant but independent. Each sample is then classified, according to its copy number status at each of these peak regions, as aberrant or not. The samples are classified as high or low level amplification with the thresholds: where low level amplification is greater than 0.1 but smaller than 0.9. High level amplification is >0.9 (0.9 corresponds to at least 3.7 copies per diploid cell). Similarly low level deletions (hemizygous) were -0.1 and high level deletions were <-1.3 (-1.3 corresponds to less than 0.9 copies per diploid cell). For further information refer to the publication and supplementary data from Beroukhim et al, 2007.

2.3 RESULTS

Cytogenetics and M-FISH

Cytogenetic analysis revealed complex numerical and structural chromosome aberrations in all cell lines with a high variability observed between cells from the same cell line (Figures 2.4 and 2.5). The G-banded karyotype of cell line WHCO1 and a corresponding M-FISH karyotype are shown in figure 2.4 that illustrates the degree of variation from one

metaphase to another (Figure 2.4). The M-FISH data confirmed the complexity of the karyotypes and revealed the genetic composition of recurrent chromosome markers that could not be identified on G-banded metaphases (Figure 2.5); some of these markers are further discussed below. The detailed composite karyotypes obtained from twenty metaphases in each cell line are summarized in Table 2.2. All cell lines were near diploid except for cell line WHCO5 which was near tetraploid (Figure 2.5). Across the five cell lines a total of 97 translocations, 19 trisomies and 11 monosomies were detected. The breakpoints amounted to 203, with 78 of these clustering around the centromeric regions of chromosomes. The chromosomes involving the highest number of abnormalities were chromosomes 1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 15, 18, 19, 20 and 22. G-banded metaphases also revealed the presence of isochromosomes involving the D group acrocentric chromosomes in all the cell lines. In particular isochromosome 13q, i(13)(p10) was common to cell lines WHCO3, WHCO5 and WHCO6, isochromosome 14q, i(14)(p10) was common to cell lines WHCO1, WHCO5 and WHCO6, and isochromosome 15q, i(15)(p10) was seen in cell lines WHCO3 and WHCO5.

The combined results of cytogenetics, FISH and M-FISH revealed a common translocation derivative, der(3)t(1;3)(p11;q11) that combined chromosome 3 and chromosome 1 short arms in cell lines WHCO5 and SNO (Figures 2.5, 2.6 and 2.7). The SNP array copy number data did not bring further information on these breakpoints, possibly due to a lack of SNP probes in this region. Interestingly, the 1p11 breakpoint was also involved in cell lines WHCO1 and WHCO6 in differing unbalanced translocations (Figure 2.5). The M-FISH and FISH results with probes for the respective partner chromosomes and centromeric Cep1, confirmed the interpretation of these derivatives as der(1)t(1;14)(p11;q11) in cell line WHCO1 and der(1)t(1;8)(p11;p11) in cell line WHCO6 (results not shown). SNP array copy

number data did not provide further information on the specific location of the breakpoints, which could have helped pin point possible target genes involved in the rearrangement.

In contrast, a translocation derivative also involving the chromosome 1 pericentromeric region in cell line WHCO3 (Figure 2.7) was shown not to involve 1p11 but 1q11, and was interpreted as der(15)t(1;15)(q11-12;p11) (result not shown).

Translocation derivatives involving chromosomes 3 and 22 were seen in cell lines WHCO3 and WHCO5 and were interpreted as der(16)t(3;16;22)(p11;?;q?11) and t(3;22)(p11;q11) respectively (Table 2.3). The array data showed corresponding hemizygous deletions at 3p11.2, in cell line WHCO3. Interestingly deletions at 3p11.2 were also observed in cell lines WHCO1, SNO and WHCO6. The minimal region of overlap was 343kb (88184220-88527215bp) in size and involved *c3orf38* and CGG triplet repeat binding protein (*CGGBP1)* genes. Deletions affecting *EPHA3* gene were detected by array copy number analysis in WHCO1, WHCO3, WHCO6 and SNO (Figure 2.8). The *EPHA3* gene encodes a tyrosine kinase, which is mutated in lung and breast cancers (Davies et al., 2005;Wood et al. 2006). *EPHA3* locus deletions were confirmed in the cell lines WHCO1, WHCO3 and SNO (Figure 2.9 and 2.10) by the locus specific FISH probes designed and produced in-house. The signal patterns observed can be seen in table 2.3 and 2.4. The WHCO1 cell line showed loss of exons 1-3 (loss of RP11-598N21). The WHCO3 cell line showed loss of exons 1-3 (loss of RP11-598N21) and exons 11-17 (RP11-547K2). The SNO cell line had loss of RP11-535I05, which maps 5' of *EPHA3*. WHCO5 and WHCO6 didn't show loss of any of the probes utilized here (Figure 2.9 and 2.10).

Cell lines WHCO1, WHCO3, WHCO5 and WHCO6 all showed a deletion of chromosome 7 long arm with varying breakpoints, q21 to q31, on G-banded metaphases (Figure 2.6). However GISTIC analysis of the SNP array data identified a significant common focal deletion (q-value of 0.09) of approximately 5.16Mb at 7q33-q34 (133721542-138880555bp) in only three of these cell lines. This region contained 26 genes including the potential target gene homeodomain-interacting protein kinase-2 (*HIPK2)* the product of which activates p53 expression and is pro apoptotic (Puca et al., 2009).

Although deletions at 20q11.2 were detected in cell lines WHCO5, WHCO6 and at 20q12 in cell line SNO on G-banded metaphases (Figure 2.6), copy number analysis (CNAT) revealed that there was in fact amplification of 20q sequences in all cell lines implying that complex rearrangement occurred for these sequences to be relocated elsewhere in the genome. Two minimal regions of amplification were identified, the 20q11.21-q11.22, of approximately 1.12Mb in size (31799867-32906584bp), which contained 14 genes and a smaller region at 20q13.12 of 149.48kb which contained 5 genes. Both these regions were amplified in all cell lines, yet these amplifications were not found to be significant on GISTIC analysis.

Cell line SNO showed a large marker chromosome 7 on metaphases analyzed by M-FISH. This marker was interpreted as a possible inverted duplication of chromosome 7p sequences (Figure 2.5D). FISH with an epidermal growth factor receptor (*EGFR)* probe, revealed a high amplification of the *EGFR* gene in 14% of the cells (Figure 2.11). A high level amplification at 7p13-7p11.2 (genomic location of *EGFR*) was confirmed by GISTIC (q-value 0.14) on array analysis in this cell line, while low level amplification was observed in the remaining four cell lines in agreement with the presence of 4 to 7 copies on FISH analysis.

Table 2.2. Composite karyotypes

Cell line	Composite combined cytogenetics and M-FISH karyotype
WHCO1	42~73, XY, +X, +X [4], +1 [10], del 1(?) [4], + der(1)t(1;14)(p11;q11) x2 [18], t(1;14;22)(?;?;q?) [4], t(1;22)(?;q?) [6], + 2 [8], t(2;8)(?;?) [4], +3 [8], +del 3(p21p13) x2 [4], der(3;22)(q10;q10) [4], -4 [8], t(5;21)(?;?) [8], + der(5)t(5;?)(q23;?) [4], + del(6)(q?21) x2 [4], t(6;12)(?q;?p) [16], t(6;13)(?q25;?q14) x2 [18], +7 [4], + del(7)(q21) [4], + der(7;14)(p10;q10) [20], der(7)t(7;18)(p21;q23) [16], der(7)t(7;19)(p22;p13) [4], der(8)t(8;?19)(q?24;q?11) [4], t(8;22)(?;?) [18], +9 [10], + der(10)t(9;10)(q13;q?11.2) [4], +11 [8], t(11;17;20)(?;?;?) [4], +12 [10], t(13;22)(?;?) [4], add(14)(q?) [4], i(14)(q10) [6], der(15;19)(q10;q10) [20], der(16)t(5;16)(?;?) [8], +18 [4], der(19)t(5;19)(p?12;q11) [12], der(19) t(5,19)(q33;q13.4) [18], + der(19)t(9;19)(q?13;q?13) x2 [4], t(19;21)(?;?) [14], +20, +20 [20], + der (20)t(1;11;20)(?;?;?) [4], +1~7mar [20] [cp20].
WHCO3	46~50 , X, -Y, +2 [12], der(5)t(5;8,18)(q?;?;?) [20], + del 7(q22) [4], der(7)t(7;9)(?;?) [6], t(7;9;16;18)(?;?;?;?) [14], t(7;15)(?;?) [4], +12 [6], der(12)t(6;12)(?;?) [16], i(13)(q10) [6], t(13;14)(?;?) [6], t(13;14;20)(?;?;?) [6], +14 [12], + der(15)t(1;15)(q11;p11) [6], + der(15)t(1;15;11)(?;?;?) [10], der(15;22)(q10;q10) [20], i(15)(q10) [4], der(16)t(3;16;22)(p?11.2;?;q?10) [20], +17 [10], der(?20)t(9;13;20)(?;?;?) [10], der(21)t(13;21)(?;?) [12] [cp20].
WHCO5*	99~108, XY, t(X;4;10.22)(?;?;?;q?) [14], t(1;19)(?;?) [8], t(1;18)(?;?) [6], del 2 [2], der(2)t(2;9)(q12;q13) [10], t(2;9)(?q31;?q34) [8], +der(3)t(1;3)(p11-12;q11) [12], t(3;11;13;22)(?;?;?;?) [8], t(3;11;22)(?;?;?) [16], t(3;22)(p11;q11) [12], der(5;20)(p10;p10) [4], t(6;13)(?;?) [4], del(7)(q31) [4]; t(8;14;18)(?;?;?) [16], t(8;18)(?;?) [10], der(9)t(9;14)(?;?) [16], t(9;15)(q?;q?) [10], t(9;19)(?;?) [6], t(12;19)(?;?) [4], i(13)(q10) [6], i(14)(q10) [8], der(15)t(Y;15)(?;?)[12], der(15)t(7;15)(?;?) x2 [14], i(15)(q10) [4], der(19)del(19)(q13.?2)t(5;19)(p?12;p11) [12], del(20)(q11.2) [12] [cp20].
WHCO6	44~54, Y, -X, der(1)t(1;8)(p11;q11) [14], t(3;10)(?;?) [4], t(3;10)(?q13.3;p10) [6], t(4;10)(?;?) [4], t(5;10)(?;?) [10], t(5;22)(?;q?) [4], +6 [2], del(6)(q?21) [4]; t(6;11)(p12;q13) [6], der(22)t(6;22)(?;?) [6], +7[7], del(7)(q31) [8]; +8 [8], t(9;15)(?;?) [10], t(10;14)(?;?) [6], -11 [6], t(11;22)(p?;q?) [6], +12 [10], i(13)(q10) [6]; i(14)(q10) [6], +16 [8], t(17;19)(?;?) [10], +18 [4], del(20)(q?11.2) [6]; -21 [12], t(21;22)(?;?) [6], der(22)t(6;22)(?;?) [6], [cp14].
SNO	29~43, XY, +X, del X(?) [14], der(Y;15)(q10;q10) [12], t(1;16)(?;?) [14], der(2)t(X;2)(?;?) [16], der(2)t(1;2)(?;?) [18], der(3)t(1;3)(p11-12;q11) [20], t(3;9)(?;?) [4], t(3;10)(?;?) [14], t(3;12)(?;?) [20], t(3;20)(?;?) [8], t(4;9;11)(?;?;?) [20], t(4;11)(?;?) [12], der(5)t(1;5)(?;?) [20], -6 [18], del(6)(q?23) [20], der(7)t(3;7)(?q25;p22) [18], t(7;20;11;8;2)(?;?;?;?;?) [18], der(8)t(2;8)(?;?) [14], t(8;18)(?;?) [8], -9 [10], t(9;18)(?;?) [4], t(10;22)(q?;q?) [12], t(11;13)(?;?)[8], t(12;15)(?;?)[8], t(12;21)(?;?)[20], t(13;11;20)(?;?;?) [14], der(14;22)(p10;q10) x2 [20], t(14;19)(?;?) [20], -16 [14], der(16)t(9;16)(q?22;q?13) [20], der(17)t(6;17)(?;?) [12], -18 [16], -19 [16], der(19)del(19)(q13.?2)t(5;19)(q?12;q10) [16], +20 [10], del(20)(q?12) [16]; der(20;21)(q10;p10) [6], -21 [18], -22 [20] [cp20].

* Only structural rearrangements are listed due to the complexity of this cell line.

Table 2.3. FISH results for the BAC probes RP11-784B9 (red) and RP11-598N21 (green).

Cell line	Signal pattern	No. Cells (n=100)	Conclusion
WHCO1			
	4 red 3 green	93	**93% loss of RP11-598N21**
	2 red 2 green	7	
WHCO3			
	2 red 1 green	100	**100% loss of RP11-598N21**
WHCO5			
	4reds 4green	100	**100% no loss**
WHCO6			
	2reds 2greens	100	**100% no loss**
SNO			
	3 reds 3 greens	100	**100% no loss**

Table 2.4. FISH results for the BAC probes RP11-535I05 (red) and RP11-547K2 (green).

Cell line	Signal pattern	No. Cells (n=100)	Conclusion
WHCO1			
	2yellow2green	93	
	2yellow	7	**93% loss of RP11-535I05**
WHCO3			
	1Yellow	74	
	1yellow1red	15	
	1 red	4	
	2yellow	7	**74% loss of both RP11-535 and RP11-547K2**
WHCO5			
	4yellow	73	
	3 yellow	27	**100% no loss**
WHCO6			
	2yellow	100	**100% no loss**
SNO			
	2 yellow 2 green	45	
	4 green	28	
	3yellow	15	
	5green	12	**88% loss of RP11-535I05**

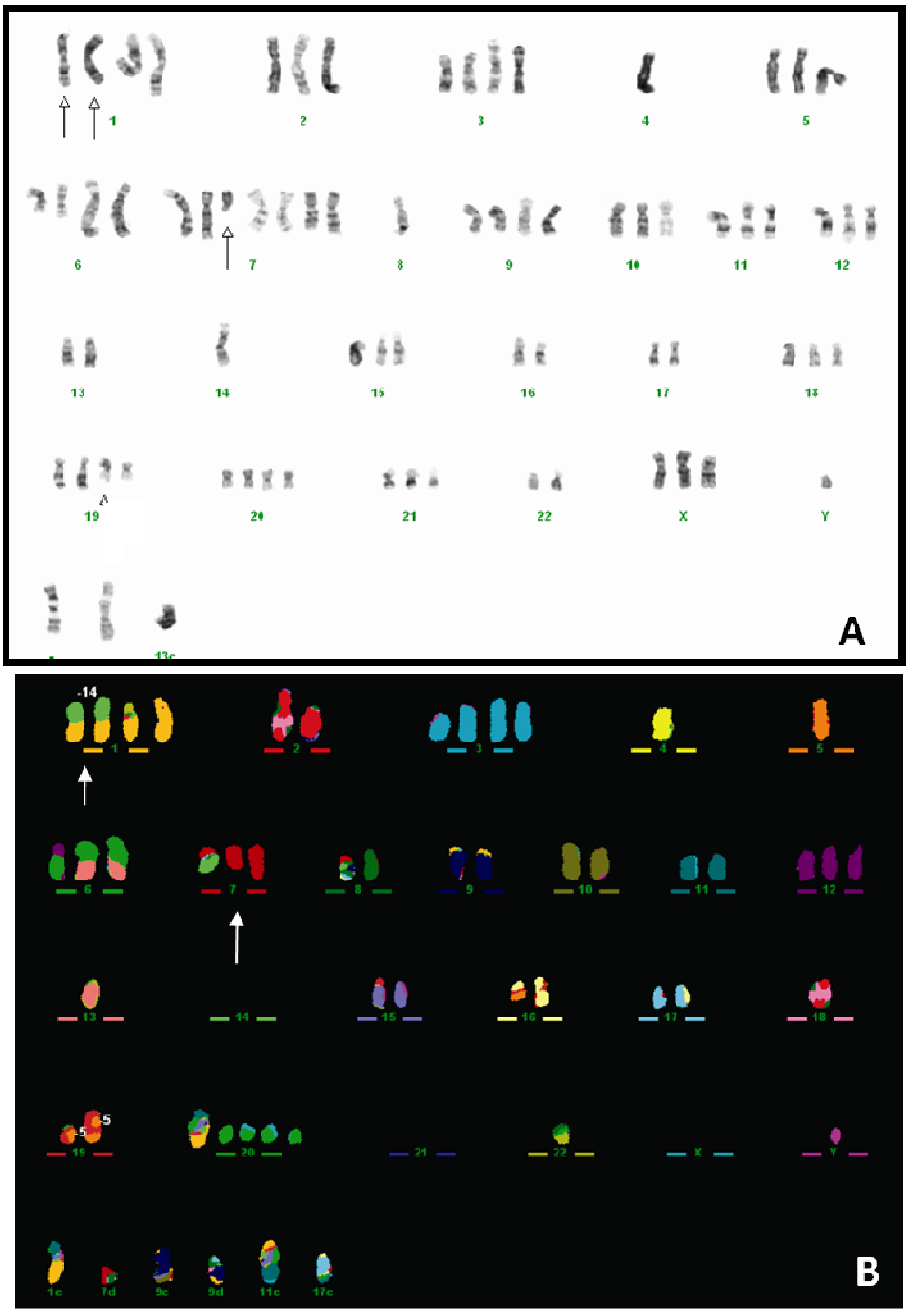

Figure 2.4. G-banded and M-FISH representative karyotypes of cell line WHCO1. (A) G-banded karyotype and (B) M-FISH karyotype, the arrows indicate the marker chromosomes, der(1) t(1;14)(p11;q11) and del(7)(q21).

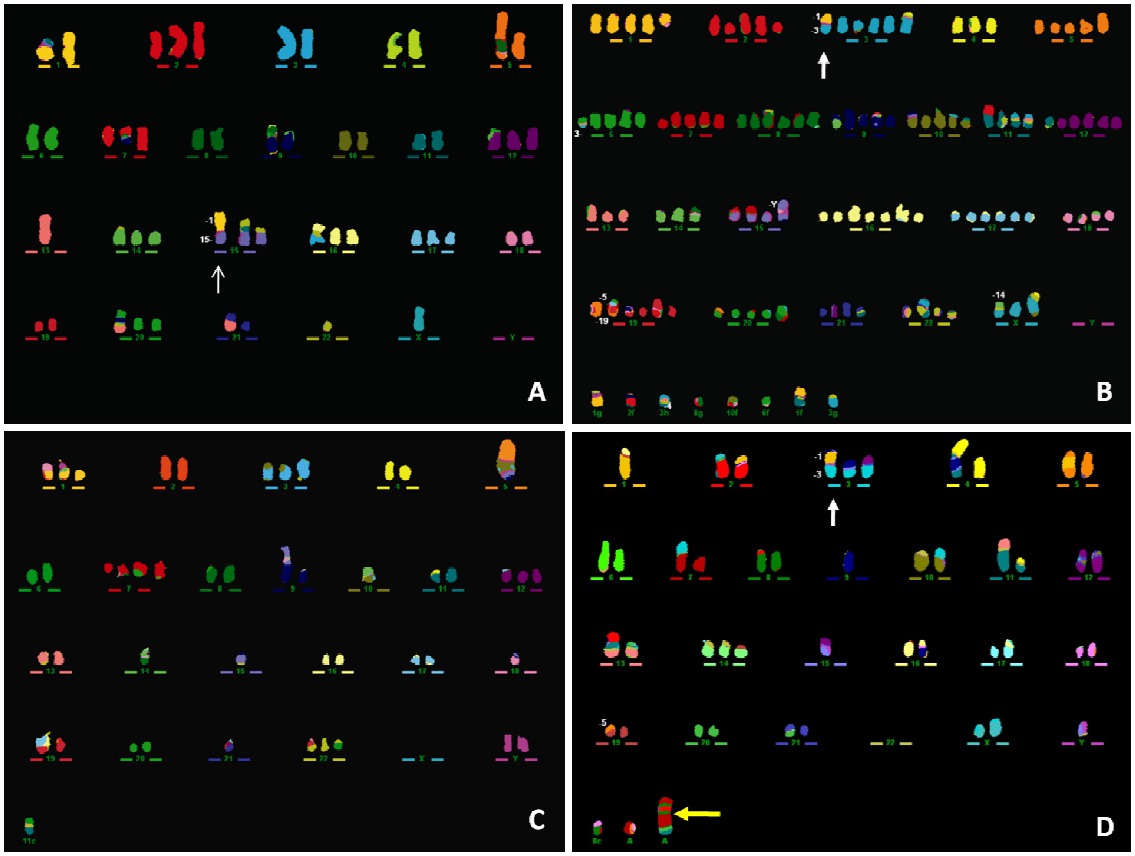

Figure 2.5. **M-FISH representative karyotypes showing complex rearrangements in four cell lines.** (A) Cell line WHCO3**,** the white arrow indicates the der(15)t(1;15)(q11;p11). (B) Cell line WHCO5**,** the white arrow points to the der(3)t(1;3)(p11-12;q11). (C) Cell line WHCO6 and (D) Cell line SNO, the white arrow points to the der(3)t(1;3)(p11-12;p11) and the yellow arrow indicates the marker chromosome 7, mar(7), which involves the *EGFR* locus (see figure 2.11).

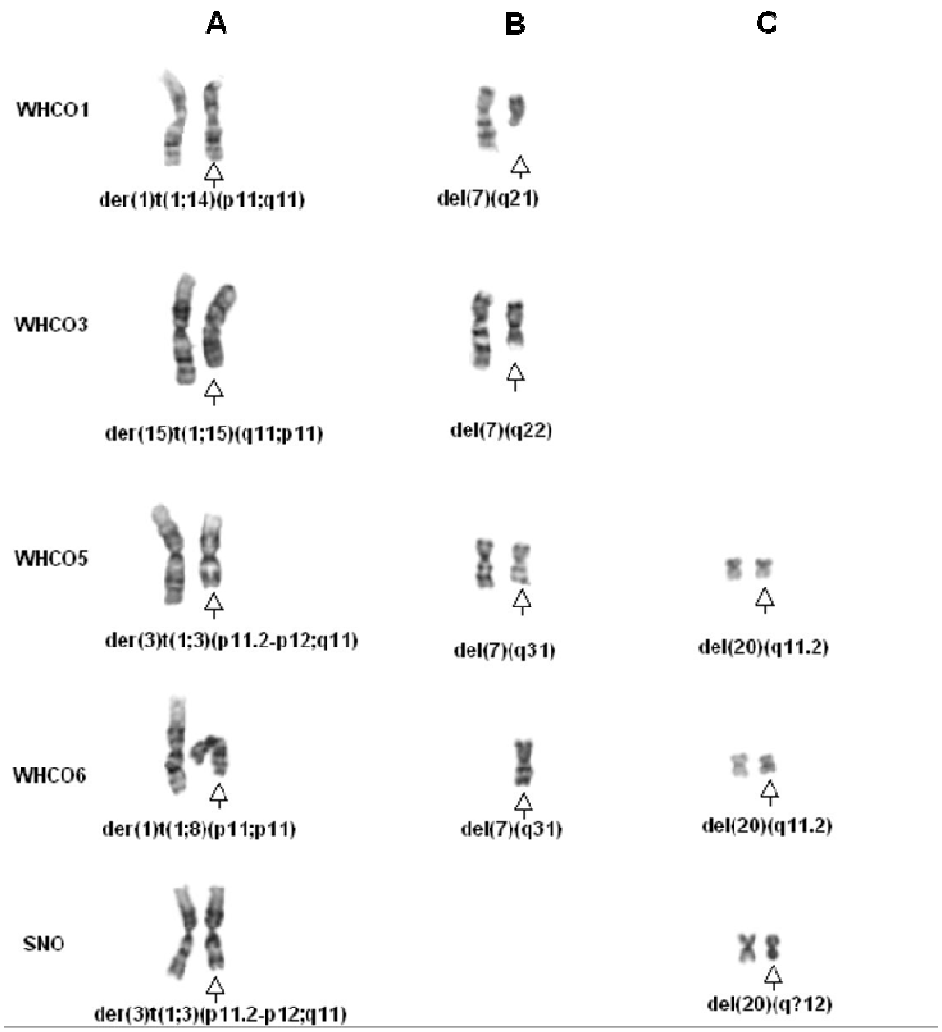

Figure 2.6. Partial G-banded karyotypes of the five cell lines showing shared chromosomal aberrations across cell lines. Chromosome derivatives from unbalanced translocations involving (A) chromosome 1p11-12 breakpoints in 4 cell lines, 1q11 in cell line WHCO3, (B) deletions of chromosome 7q in four cell lines, and (C) deletions of chromosome 20q in 3 cell lines.

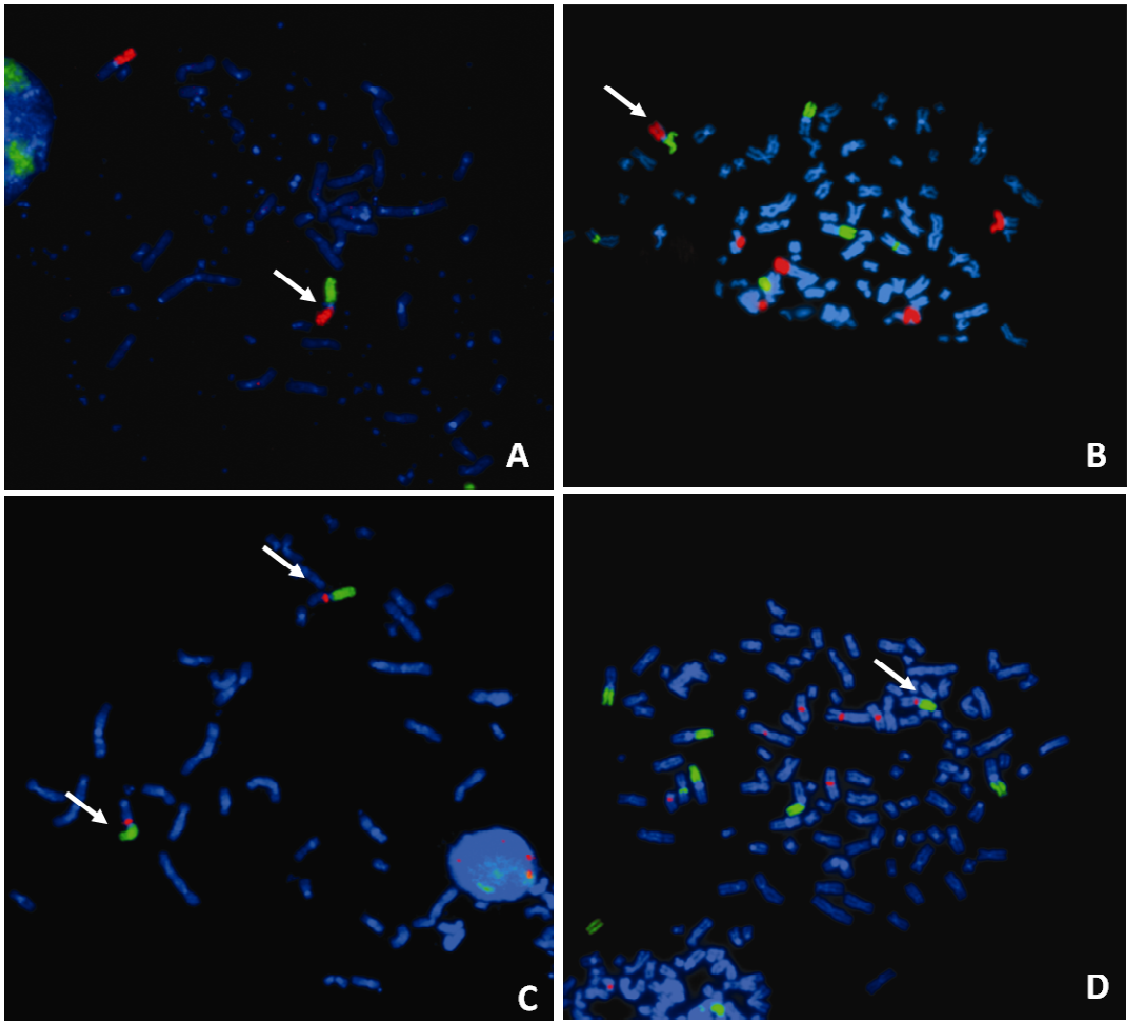

Figure 2.7. DAPI stained metaphasic chromosomes showing the der(3)t(1;3)(p11-12;q11) in cell lines SNO and WHCO5. (A) and (B) Arm-specific paint for chromosomes 1p (green) and 3p (red) in cell lines SNO and WHCO5 respectively showing the derivatives der(3)t(1;3)(p11-12;q11) arrowed. (C) and (D), arm specific paint for chromosome 1p (green) and Cep 3 alpha probe (red) in cell lines SNO and WHCO5 respectively showing that the centromere of chromosome 3 is retained on the derivatives der(3)t(1;3)(p11-12;q11).



Figure 2.8. CNAT plot of SNP hybridization results showing the map of the region of deletion on chromosome 3p11.2-12.1 in cell lines WHCO1, WHCO3, WHCO5, WHCO6 and SNO. The red bars indicate the segment of deletion as determined by CNAT 4.0. The dotted lines indicate the copy number status for each SNP probe for each cell line. The blue block indicates the minimal common region of deletion and the red arrow indicates the two genes in this region, *c3orf38* and *CGGBP1*.

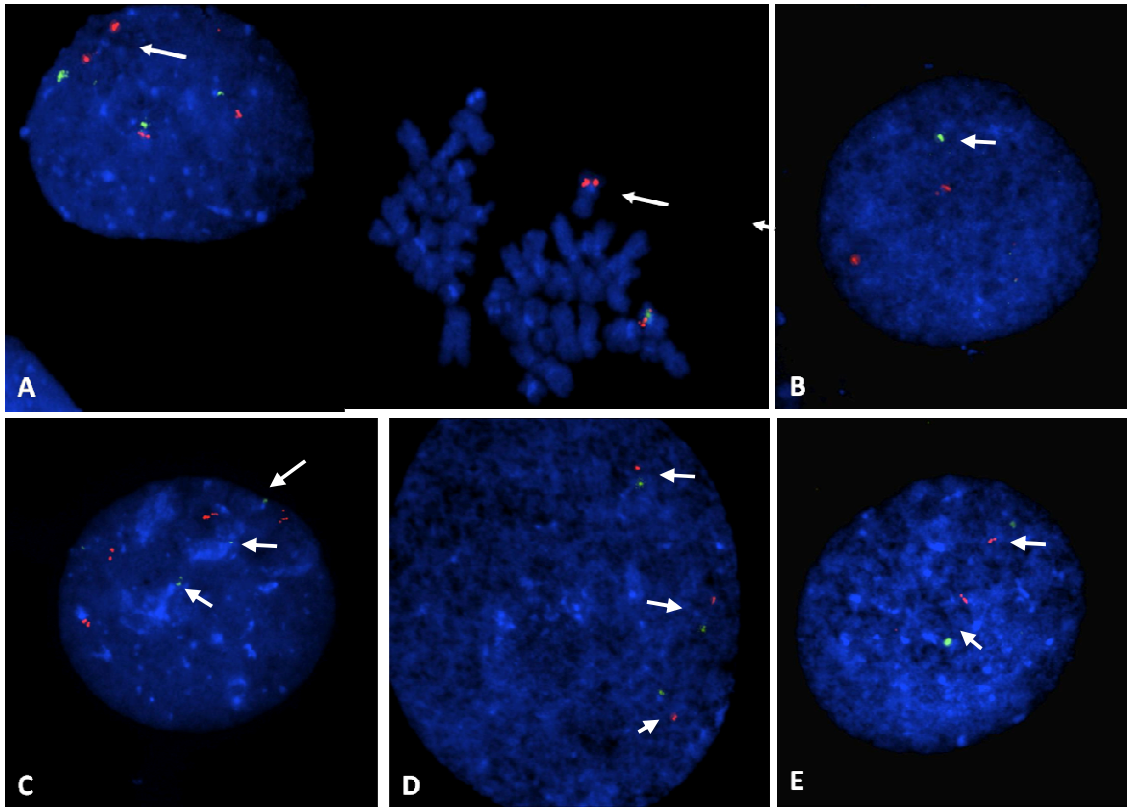

Figure 2.9. Representative images of DAPI stained interphase nuclei showing FISH results for all five cell lines with BAC clones, RP11-589N21 (green) and RP11-784B9 (red). The WHCO1 cell line (A) showing an interphase cell with 4 copies of chromosome 3p and loss of RP11-598N21 BAC (green) signal, which indicates loss of exon 1-3 sequences in one chromosome 3 homologs. The metaphase shows chromosome 3p with loss of RP11-598N21 as indicated by the white arrow. The WHCO3 cell line (B) showing two copies of chromosome 3p and loss of one RP11-598N21 BAC signal (loss of exon 1-3 sequences, indicated by the white arrow). The WHCO5 cell line (C) showing 4 copies of chromosome 3p and no loss of RP11-598N21 BAC signal (white arrows). The SNO cell line (D) showing 3 copies of both BAC probes (white arrows), indicating that there is no loss of RP11-598N21. The WHCO6 cell line (E) with both copies of both probes, indicating that there is no loss of either. Cell lines WHCO5 and SNO had 3 to 4 copies of chromosome 3 per cell as determined by M-FISH.

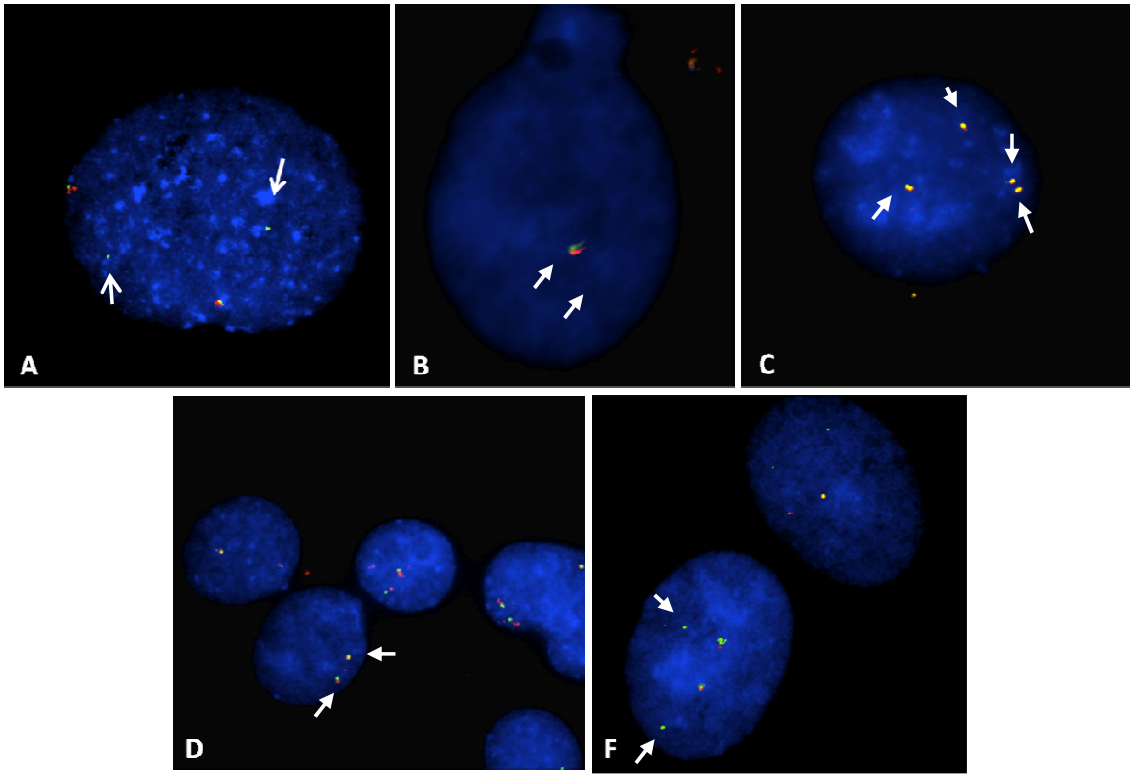

Figure 2.10. Representative images for DAPI stained interphase cells from each cell line showing results for FISH analysis with BAC clones, RP11-535I05 (red) and RP11-547K2 (green). In the WHCO1 cell line (A), the pattern was 2 fusion and 2 green signals indicating loss of the RP11-535I05 BAC, indicating loss of sequences upstream of the *EPHA3* gene. In the WHCO3 cell line (B) the pattern was 1 fusion showing loss of both RP-11535I05 and RP-11547K2 BACs on chromosome 3 homolog (white arrow), indicating a loss of exon 11-17 and upstream sequences of *EPHA3*. In the WHCO5 cell line (C) the pattern was 4 fusion signals showing no loss (white arrows). In the WHCO6 cell line (D) 2 fusions were seen, showing no loss (white arrows). In the SNO cell line (F), 2 to 3 fusions and two greens (white arrows) were observed showing loss of the RP11535I05 clone.

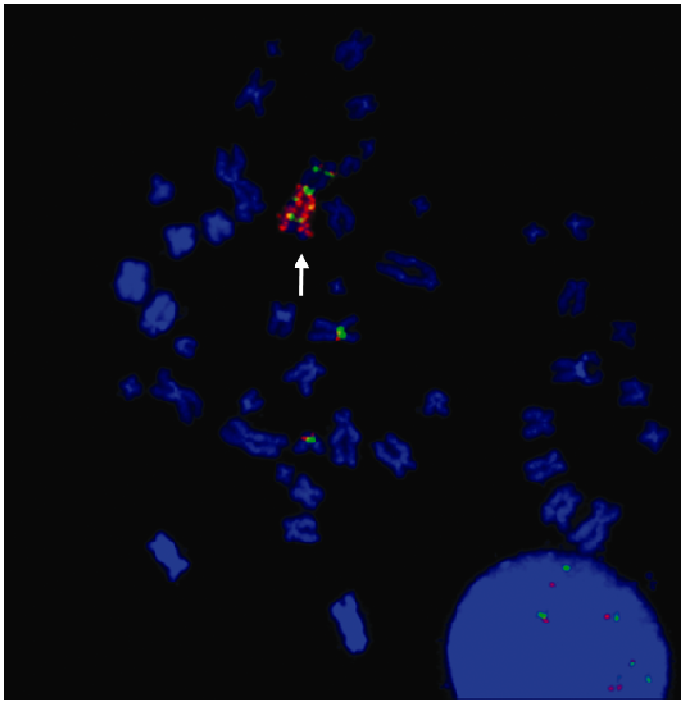

Figure 2.11. DAPI stained metaphase from cell line SNO showing FISH results with the locus specific EGFR probe (red) and Cep7 alpha (green). The arrow indicates the chromosome 7 marker with a homogeneously stained region (HSR) that contains the *EGFR* locus.

Copy number changes

In view of the high cell to cell heterogeneity in each cell line, and in order to separate the potential driver aberrations out of the background of aberrations that may have occurred by chance, we used the software GISTIC specifically designed for this purpose. Target genes, defined as genes whose alteration confers a cell growth advantage, are likely to reside in the regions amplified or deleted to the highest degree in a majority of cells and within a common region of overlap amongst all cell lines. Fourteen common regions of amplification and 20 regions of deletion were identified (Figure 2.12 and Tables 2.5 and 2.6). The most significant chromosomal regions of amplification were, in descending order of significance: 11q13.3 and 8q24.21, in 4 and 5 cell lines respectively, 11q22.1-q22.3 and 5p15.2 both

detected in 2 cell lines. Chromosomal regions of less significant amplification were 10p12.33-q21.3, 18p11.32, 20p11.1-p11.22, 22q11.21-q11.22, 22q12.3, 9q31.1, 22q11.21, 3q11.2-q12.2 and 1p34.2 (Figure 2.12). Similarly, chromosomal regions of deletion were seen in the following order of significance: 1p31.1-p31.2, 2q22.1, 3p12.1-p14.2, 4q22.1-q32.1, 8q11.1-p23.2, 14q21.2, 18q21.1-q21.2 and 21p13-q21.3 detected in all or 4 cell lines. Other regions of less significant deletion, and only seen in 1 or 2 cell lines, included: 10p12.31-p14, 14p13-q32.33, 3p26.3-q29, 9p24.2-p24.3, 6q24.3-q27, 12p13.33-q24.33, 7q33-q34, 8p23.3-q24.3, 10q11.23-q22.1, 11p11.12-q12.2 and 13q21.33-q34 (Figure 2.12). Together the regions of amplification and deletion encompassed a total of 4595 genes.

Significant Gains

The five regions of amplification that were the most significant on GISTIC analysis (q-value <0.25) are selectively described below (Figure 2.12).

Chromosomes 11q13.3 (68753086-69985447bp) and 8q24.21 (127445828-129661846) were the two most amplified and most significant regions with a q-value of $5.76E^{-05}$ and 0.0007 respectively. The 11q13.3 region was 1.23Mb in size and was highly amplified in cell lines WHCO3, WHCO5, WHCO6 and SNO while amplified to a lesser degree in cell line WHCO1. This region harbours seven candidate genes including the Cyclin D1 (*CCND1*), the Cortactin (*CTTN)*, the protein tyrosine phosphatase, receptor type, polypeptide, interacting protein alpha 1 (*PPFIA1*), the fibroblast growth factor *FGF3*, *FGF4* and *FGF19* and the myeloma overexpressed (*MYEOV*) genes, which could all play a role in OSCC oncogenesis. FISH validated these findings and confirmed the amplification of *CCND1* and *FGF4* (Table 2.6,

Figure 2.13). The common amplicon of 2.22Mb at 8q24.21 was highly amplified in cell lines WHCO1 and WHCO3 and moderately amplified in cell lines WHCO5, WHCO6 and SNO. This amplicon involved the oncogene v-myc myelocytomatosis viral oncogene homolog (*C-MYC*) and the family with sequence similarity 84, member B (*FAM84B*) gene. Locus specific FISH confirmed the amplification of *C-MYC* in cell lines WHCO1, WHCO3, WHCO5, WHCO6 and SNO (Table 2.7, Figure 2.13).

A second focal region of high amplification on chromosome 11 was observed at 11q22.1-q22.3 (2.23Mb in size) in cell lines WHCO5 and SNO. This region included the regulators of apoptosis *BIRC2* (*cIAP1*) and *BIRC3* (*cIAP2*), the matrix metalloproteinases (*MMP*) and the Yes associated protein *(YAP-1)* genes all potential target genes. The *BIRC2* gene was previously described as a target of amplification /increased expression in cervical cancers (Imoto et al., 2002) and the *YAP*-1 gene product is a cellular adaptor protein, which can induce *BIRC2* expression. *YAP-1* was reported to be over expressed in hepatic and mammary cancers (Da et al., 2009). In turn, the MMP genes, which include *MMP1*, *MMP7*, and *MMP13*, have been shown to be co-expressed in early stage OSCC correlating with a poorer prognosis (Z. D. Gu et al., 2005).

A 1.75Mb region on chromosome 5p15.2 (10051329-11800765bp) was highly amplified in cell lines WHCO6 and WHCO5 and moderately amplified in cell lines WHCO1 and SNO. This region hosts the potential target gene, Delta catenin (*CTNND2*), over expressed in prostate cancer (M. J. Burger et al. 2002).

Four cell lines, (WHCO6, WHCO3, WHCO5 and SNO) had focal gain on chromosome 3q. The minimal common region of amplification mapped at 3q11.2-12.2, and was 6.02Mb in size (95917505-101945216bp) (q-value=0.17). This region is commonly amplified in a variety of

cancers (Haverty et al., 2009; Or et al., 2005) including oesophageal squamous carcinoma (J Chen et al., 2008) and it involved the potential oncogene, MYC induced nuclear antigen (*MINA*).

The 18p11.32 sub-band was highly amplified in cell lines WHCO3 and WHCO6 and moderately amplified in the remaining 3 cell lines. This region of 1.12Mb in size (1-1118244bp) has previously been described in OSCC (Nakakuki et al., 2002) and involves the potential oncogenes *TYMS* and *YES-1.* Both genes have been implicated in gastro intestinal cancer. The *TYMS* gene codes for a thymidylate synthase involved in DNA synthesis and targeted by the chemotherapy agent Fluorouracil (5FU). *TYMS* overexpression leads to 5FU treatment's resistance (Nakakuki et al., 2002) and affects colorectal cancer treatment (S. A. Jensen et al., 2008). *YES-1* is a homologue of the Yamaguchi sarcoma virus v-*yes* amplified and overexpressed in gastric cancers and OSCC (Nakakuki et al., 2002).

Significant Losses

The most significantly deleted chromosomal regions were 1p31.1-p31.2, 2q22.1, 3p12.1-p14.2, 4q22.1-q32.1, 8p23.2-q11.1, 14q21.2 and 18q21.1-q21.2 (in descending order of significance) in all or four of the cell lines (Figure 2.12). Less significant regions detected in less than 4 cell lines are depicted in table 2.6.

Three small regions of deletions involved chromosomes 1p, 2q and 18q.

The 4.5 Mb deletion on chromosome 1 short arm, 1p31.2-p31.1 (66691991-71187083bp) was seen in four cell lines (homozygous deletion in cell line SNO and hemizygous in cell lines WHCO1, WHCO5, WHCO6) with a high significant q-value of 0.02. Three genes with a

reported tumour suppressor activity were involved in this deletion, the cystathionine γ-lyase (*CTH*), the growth arrest and DNA damage-45 alpha (*GADD45alpha*) and the DIRAS family, GTP-binding RAS-like 3 (*DIRAS3*) genes. The *DIRAS3* gene was shown to be down regulated in hepatocellular carcinoma and breast cancers (Jian Huang et al., 2009; Z. Shi et al., 2002) and is postulated to have a tumour suppressive activity. Both the *CTH* and *GADD45* genes were shown to negatively control cell growth (Rosemary Siafakas & Richardson, 2009; G. Yang et al., 2004).

Chromosome 2, sub-band q22.1 (141590067-141951947bp) was lost in all cell lines (homozygous deletion in cell lines WHCO3 and WHCO5 and hemizygous in the other 3 cell lines) (q-value= 0.019). The low density lipoprotein 1B (*LRP1B*) tumour suppressor gene maps in this region and is deleted in lung cancer (C. X. Liu et al., 2007; Nagayama et al., 2007).

A chromosome 18q sub-band, q21.1-q21.2 (46081464-51919972bp) (5.8Mb) was hemizygously deleted in all five cell lines, involving both the *SMAD4* and deleted in colorectal carcinoma (*DCC*) genes.

Three large regions of deletion involved chromosomes 3p, 4q and 8p.

First, the 24.7Mb region of deletion at 3p12.1-p14.2 (60424050-85108679) was significant (q-value of 0.02) in all cell lines (homozygous deletion in cell line SNO). This region houses the FRA3B associated gene, Fragile Histidine triad (*FHIT*), whose deletions were previously detected by MLPA analysis in these cell lines (Willem et al., 2006). The potential tumour suppressor *ADAMTS9* gene, a metalloproteinase family member involved in inhibition of angiogenesis (P. H. Y. Lo et al., 2010) was also involved in this deletion.

Second, the 4q22.1-q32.1 (91972774-162358674bp) (75Mb) region was homozygously deleted in cell lines WHCO3 and WHCO5 and hemizygously deleted in the 3 other cell lines with a q-value of 0.02. This region encompasses many genes but of interest are the Bone morphogenetic protein receptor 1B (*BMPR1B*), the Caspase 6 (*CASP6*), the secreted frizzled-related protein 2 (*SFRP2*) and the SMAD protein 1 (*SMAD1*) genes, all potential tumour suppressor genes.

Lastly, a chromosome 8p23.2-q11.1 (4078057-47043375bp) (43Mb) deletion was seen in all cell lines (homozygous in cell line WHCO3 and hemizygous in the other 4 cell lines) (q value of 0.02). This region is very large and could target numerous genes. Genes that have been implicated as tumour suppressors include, secreted frizzled-related protein 1 (*SFRP1*), the BCL2/adenovirus E1B 19kDa interacting protein 3-like (*BNIP3L*), the leucine zipper tumor suppressor 1 (*LZTS1*) and the three tumor necrosis factor related superfamily genes *(TNFRS), TNFRSF10A, TNFRSF10B* and *TNFRSF10C*.

Table 2.5. Amplification peaks as detected by the GISTIC algorithm.

Cytoband	Location(kb)	Approx Size (Mb)	Frequency(n/5)	Mean log2ratio[a]	q-value	Genes[b]
1p34.2	Chr1:39027237-40780163	1.75	3	1.1	0.18	*MYCL1*
3q11.2-q12.2	chr3:95917505-101945216	6.02	4	1.4	0.17	*MINA*
5p15.2	chr5:10051329-11800765	1.75	2	1.36	0.09	*CTNND2*
7p11.2-p13	Chr7:45294289-57299457	12.01	1	1.24	0.13	*EGFR*
8q24.21	chr8:127445828-129661846	2.22	5	2.1	0.0007	*MYC*
9q31.1	chr9:100905143-102152148	1.25	1	2.57	0.15	
10q12.33-q21.3	chr10:17811791-65388337	47.58	1	3.05	0.13	
11q13.3	chr11:68753086-69985447	1.23	4	1.81	$5.76E^{-05}$	*CCND1, CTTN, FGF3, FGF4, FGF19, MYEOV*
11q22.1-q22.3	Chr11:100815801-103042620	2.23	2	2.35	0.03	*BIRC2, BIRC3, YAP1*
18p11.32	chr18:1-1118244	1.12	2	1.04	0.13	*TYMS, YES1*
20p11.1-p11.22	chr20:22140447-26145930	4.01	2	1.24	0.13	*PYGB*
22q11.21	chr22:16558724-17937900	1.38	2	1.2	0.21	*BID, CLDN5*
22q11.21-q11.22	Chr22:18577713-20667607	2.1	2	1.2	0.13	*CRKL, MAPK1*
22q12.3	Chr22:31889314-32003182	0.11	2	1.15	0.13	*LARGE*

[a]The mean log2ratio for the samples with a log2ratio>0.9 (equivalent to 3.7 copies per diploid cell).
[b]The selected genes from the peak region

Table 2.6. Deletion peaks as detected by the GISTIC algorithm.

Cytoband	Location(kb)	Size (Mb)	Frequency(n/5)		Mean log2ratio*	q-value	Genes[b]
			Hemi	Homo			
1p31.1-p31.2	chr1:66691991-71187083	4.5	3	1	-3.03	0.02	*CTH*
2q22.1	chr2:141590067-141951947	0.36	3	2	-1.66	0.19	*LRP1B*
3p26.3-q29	Chr3:1-199344050	199.34	2	1	-2.9	0.022	
3p12.1-p14.2	chr7:60424050-85108679	24.7	4	1	-2.99	0.02	*FHIT, ADAMTS9*
4q22.1-q32.1	Chr4:91972774-162358674	70.4	3	2	-2.19	0.02	*CASP6, SMAD1*
6q24.3-q27	Chr6:147967444-170914576	22.95	2	1	-2.22	0.07	
7q33-q34	Chr7:133721542-138880555	5.16	2	1	-1.72	0.09	*HIPK2*
8p23.2-q11.1	Chr8:4078057-47043375	43	4	1	-2.47	0.02	*SFRP1, BNIP3L, INDO*
8p23.3-q21.13	Chr8:1-146308819	146.31	1	1	-2.27	0.09	
9p24.2-p24.3	Chr9:1151516-2459741	1.31	2	1	-2.53	0.03	
10p12.31-p14	Chr10:10309026-19155158	408.85	2	1	-2.74	0.02	
10q11.23-q22.1	Chr10:50393324-70694787	20.3	1	1	-2.35	0.12	
11p11.12-q12.2	Chr11:50256798-61426521	11.2	1	1	-1.9	0.12	
12p13.33-q24.33	1-132078379	132.1		1	-2.8	0.08	
13q21.33-q34	68772537-113042980	44.3	1	1	-1.7	0.24	
14q21.2	42828345-44176016	1.35	3	1	-4.5	0.02	
14p13-q32.33	1-105311216	105.3		1	-5.1	0.02	
18q21.1-q21.2	46081464-51919972	5.8	5		-1.03	0.02	DCC, SMAD4
21p13-q21.3	1-29932926	29.9	3			0.12	BAGE

[a]The mean of the log2ratio of those samples with log2ratios <-1.3 (<0.9 copies per diploid cell). [b]The selected genes within the deletion peaks.

Table 2.7. FISH for detection of *CCND1, FGF4* and *C-MYC* amplification.

Cell line	FISH signals	Amplitude#	Log2Ratio
CCND1			
WHCO1	5-15	2	1.56
WHCO3	>20	2	2.57
WHCO5	10-20	1	0.78
WHCO6	4-8	1	0.55
SNO	15-20	1	0.79
C-MYC			
WHCO1	4-6	1	0.12
WHCO3	15->20	2	1.83
WHCO5	10->20	2	1.83
WHCO6	10->20	2	2.06
SNO	4->20	2	1.71

Amplitude threshold where log2ratio<0.1 =0, log2ratio>0.1<0.9= 1 and log2ratio>0.9=2 as determined by SNP array copy number analysis.

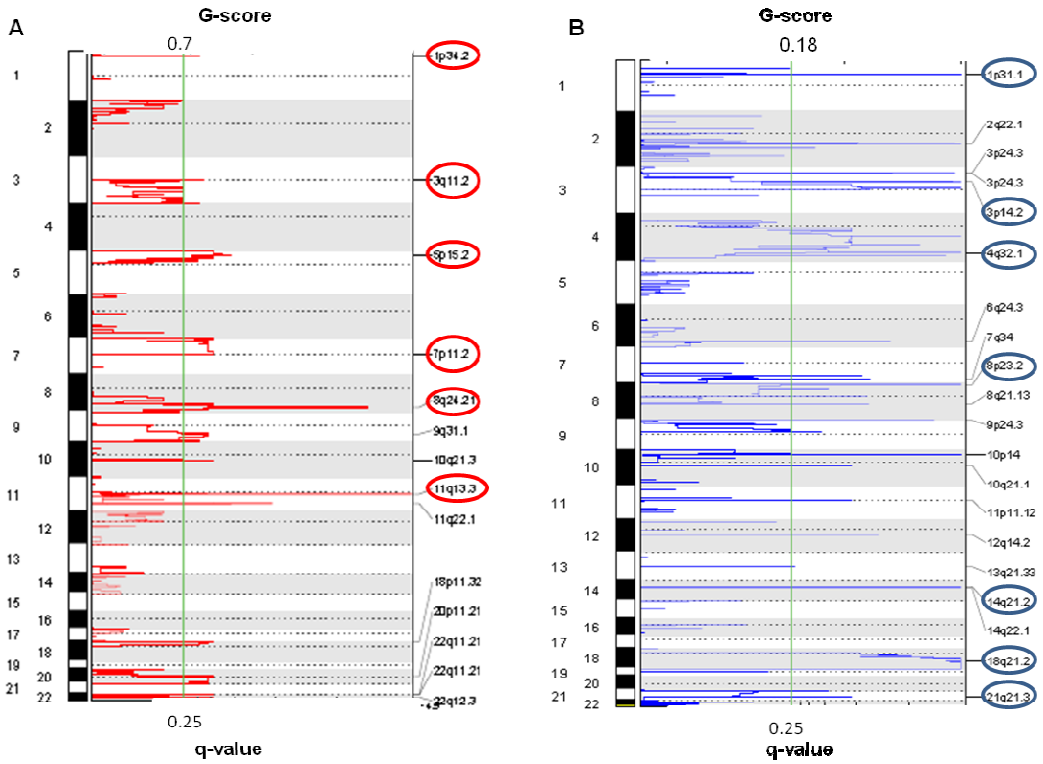

Figure 2.12. Plots of recurrent genomic amplifications (A) and deletions (B) detected in the OSCC cell lines from GISTIC analysis of SNP array data. The x-axis shows the G-score (top) and false discovery rate (q value; bottom). The green line indicates the false discovery rate cut off of 0.25. The circles indicate the peaks of the most significantly aberrant chromosomal regions.

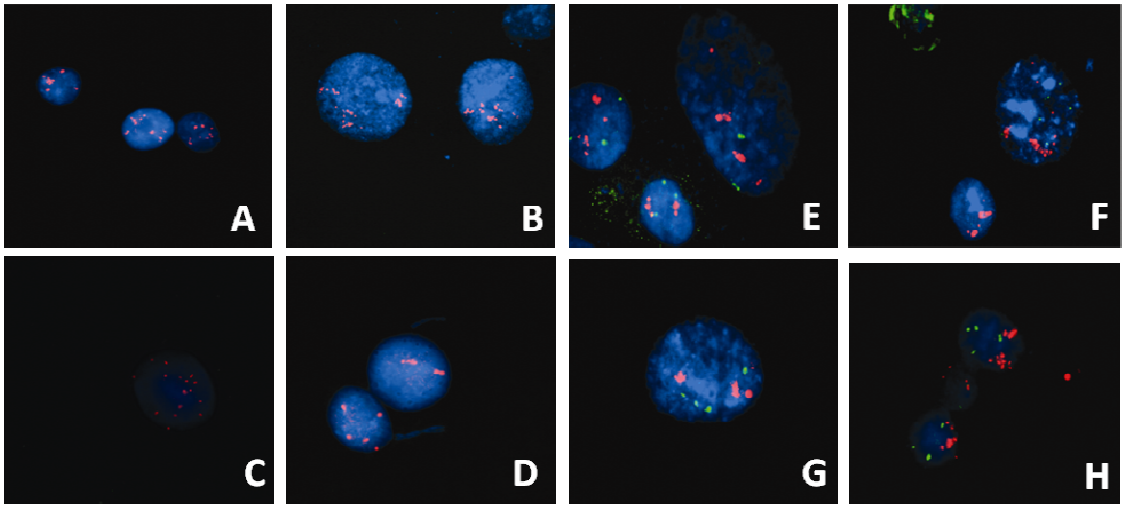

Figure 2.13. DAPI stained interphase nuclei hybridised with locus specific probes for C-*MYC* Spectrum Orange (red) (A-D) and *CCND1* (red)/*IGH* (green) (E-H). C-*MYC* amplification can be seen in cell lines WHCO1 (A), WHCO3 (B), WHCO5 (C) and SNO (D). *CCND1* amplification was detected in cell lines WHCO3 (E), WHCO5 (F), WHCO6 (G) and SNO (H).

2.4 DISCUSSION

The karyotype and genomic constitution of five OSCC cell lines established in SA using a combination of traditional cytogenetics, M-FISH, FISH and SNP arrays has been characterised. The number of OSCC cell lines genetically described worldwide is limited. Only eight OSCC cell lines have been investigated previously with traditional cytogenetics to our knowledge (Y. C. Hu et al., 2002; Y. P. Wu et al., 2006; S. Xiao et al., 1991; L. C. M. Cheung et al., 2007) and 10 with multi color FISH (Y. P. Wu et al., 2006; Yang et al., 2008; C. C. Yen et al., 2001), these are the only two techniques that can detect recurrent translocation breakpoints. In this study the 5 cell lines had complex karyotypes and were hyperploid with WHCO5 being near tetraploid. There was a high level of intra cell-line heterogeneity. The chromosomes most frequently involved in translocations were

chromosomes 1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 15, 18, 19, 20 and 22. These features were comparable to OSCC cell lines previously characterized (Y. P. Wu et al., 2006; S. Xiao et al., 1991; L. C. M. Cheung et al., 2007) and across all studies that involved karyotyping, including the karyotyping of fresh OSCC tumour samples (Yuesheng Jin et al., 2004), chromosomes 1, 3 and 8 were the most commonly affected by translocation breakpoints (Y. C. Hu et al., 2002; Y. P. Wu et al., 2006; S. Xiao et al., 1991; Yang et al., 2008).

Forty percent of translocation breakpoints occurred in near centromeric regions (Table 2.2). These were represented by unbalanced whole arm chromosome translocations, frequently involving chromosomes 1 and 3, and by isochromosomes for the D group acrocentric chromosomes 13, 14 and 15. Indeed, frequent centromeric breakpoints have been described in squamous carcinoma including OSCC (Y. C. Hu et al., 2002; S. Xiao et al., 1991) with up to 60 % of all breakpoints being in centromeric regions (Yang et al., 2008) supporting the idea that centromeric disruption is a frequent event in epithelial cancers. It has been suggested that environmental factors may preferentially interact with centromeric sequences (Y Jin et al., 2000), and clastogenic compounds, such as mitomycin C, induce breaks in centromeric of chromosomes 1, 9 and 16 (B Johansson & F Mertens, 1988). Smoking is a major risk factor associated with OSCC and nicotine is known to induce single strand DNA breaks (Kleinsasser et al., 2006). Although active smokers exhibit an increased number of breaks at fragile sites (C. K. Stein et al., 2002), it is not known if centromeric regions are also targeted.

Two chromosomal breakpoints were shared across cell lines. First, chromosome 1p11 was involved in a translocation in four cell lines, translocation t(1;3)(p11.2-12; q11) in cell lines

SNO and WHCO5 and translocations with differing partners in cell lines WHCO1 and WHCO6. Second, chromosome 3p11-12 was involved in translocations in two cell lines and deletions in, or near the *EPHA3* locus was seen in three cell lines. The *EPHA3* gene codes for a receptor tyrosine kinase, with a tumour suppressor activity (Clifford et al., 2008). It was found to be mutated in lung and breast cancers (Davies et al., 2005; P.; L. D. Wood et al., 2006) and interestingly, found to be deleted in 18.2% of OSCC patients in a previous study (Chen et al., 2008).

Breakpoints, in or near the centromeric regions of chromosomes 1 and 3, have previously been reported in several OSCC cell lines (Y. C. Hu et al., 2002; Y.P. Wu et al., 2006; S. Xiao et al., 1991), as well as in fresh OSCC tumour samples that were karyotyped (Yuesheng Jin et al., 2004) and in OSCC cell lines obtained by in vitro transformation with the human papillomavirus (HPV) (H. Zhang et al., 2006). This strongly points to 1p11 and 3p11 translocation hotspots in OSCC that may affect genes and/or regulatory sequences not identified by SNP array. Deletions affecting or near the *EPHA3* gene may point to a role for this gene which was previously found to be deleted in 18.2% of OSCC patients (Chen et al., 2008).

It is known from previous studies that these OSCC cell lines, all over express the *EGFR* gene (R Veale, 1989). *EGFR* DNA amplification observed in cell line SNO is likely to contribute to *EGFR* over expression whilst amplification in other cell lines other factors are likely to be involved in the other 4 cell lines where low levels of *EGFR* amplification were observed.

In view of the high clonal heterogeneity observed in each cell line, we used the GISTIC software in addition to CNAT to analyze SNP array data and evidence the significant targets of amplification and deletions. The most stable genetic rearrangements are thought to reflect a proliferative cell growth advantage. In this context, the GISTIC algorithm allowed us to prioritize amplicons and regions of deletions in term of their likelihood to host driver genes. The five most interesting significant regions of amplification included chromosomal regions: 11q13.3, 8q24.21, 11q22.1-q22.3, 5p15.2 and 3q11.2-q12.2 in decreasing order of significance.

The 8q24.21, 11q13.3 and 3q11.2-q12.2 regions have all previously been reported in a variety of carcinomas (Burkhardt et al., 2010; Freier et al., 2003; Bass et al., 2009) and they often co-exist with one another. Genomic amplification at 8q24 occurs in a large variety of cancers (Burkhardt et al., 2010; Bass et al., 2009; Camps et al., 2009; X. P. Huang et al., 2006; van Duin et al., 2007), and most amplicons described in the literature involve both the *C-MYC* and *FAM84B* genes, as was observed in this study in four OSCC cell lines. In previous reports the target of amplification has been attributed to either both genes (Camps et al., 2009), or to one or the other (X. P. Huang et al., 2006; van Duin et al., 2007) based on their respective increased transcription.

The 11q13.3 amplicon covered a large region containing a number of potential target genes. The *CCND1* and *MYEOV* genes (11q13.3) were co-amplified in four cell lines. Co-amplification of these genes has been reported in multiple myeloma, breast cancer and OSCC (Yuesheng Jin et al., 2004). *CCND1* is a downstream effector in the Wnt2/beta-catenin pathway and the most frequent target of amplification in several OSCC studies (Yuesheng Jin

et al., 2004; Bass et al., 2009; J. W. G. Janssen, Cuny, et al., 2002). While the *MYEOV* gene has been associated with cell proliferation in colon cancer (Moss et al., 2006), its amplification in OSCC is not always matched by increased transcription due to its silencing by epigenetic mechanisms (J. W. G. Janssen, Imoto, et al., 2002). The cortactin gene, *CTTN,* involved in cell motility (P. Hofman et al., 2008), was previously shown to be over expressed in OSCC pre-cancerous lesions, as well as in carcinogen induced murine OSCC supporting a role for this gene in OSCC carcinogenesis (N. Y. Hsu et al., 2008). The three fibroblast growth factor (FGF) genes, *FGF3*, *FGF4* and *FGF19* were part of the 11q13.3 amplicon. FGFs and Wnt signalling pathways cross talk in a number of carcinogenesis scenarios (Masuko Katoh & Masaru Katoh, 2006). Activated FGF receptors activate the FRS-GR2-GAB1-PI3K-AKT signaling cascade, and down regulate GSK-3β protein activity, thus hampering β-catenin phosphorylation and degradation (Masuko Katoh & Masaru Katoh, 2006). In particular, FGF19 ligand down regulates GSK-3β activity, which results in the release and nuclear accumulation of β-catenin. Nuclear β-catenin activates the transcription of downstream genes including *C-MYC* and *CCND1* (Novak & Dedhar, 1999).

In addition to the amplification of Wnt pathway activators, the *SFRP1* and *SFRP2* tumour suppressor genes' loci were deleted at chromosomes 8p23.2-q11.1 and 4q22.1-q32.1 respectively (Figure 2.7). The *SFRP1* and *SFRP2* genes encode frizzled-related proteins and are part of the SFRP family of Wnt inhibitors (Rattner et al. 1997; Kawano & Kypta 2003). Inactivation of *SFRP1* by methylation and lower expression is observed in gastric cancer (Kinoshita et al. 2011). Loss of *SFRP2* is detected in medulloblastoma and is suggested to contribute to carcinogenesis through loss of inhibition of the Wnt pathway (Kongkham et al., 2010).

The copy number data therefore suggests that the Wnt signalling pathway may be at work in these OSCC cell lines through one or the combined effects of genes activating the β-catenin transcriptional activity and/or the FGF signalling pathways as well as deletions of genes, at 4q22.1-q32.1 and 8p23.2-q11.1, inhibiting this pathway. Amplicons at both 8q24 and 11q13.3-13.4 have been described in a variety of squamous cell carcinoma (Camps et al., 2009; W. Liu et al., 2008; J. W. Janssen et al., 2000; Fantozzi et al., 2008) suggesting that the activation of pathways through the combined effects of genes at 8q24 and 11q13.3-13.4 contributes to the development and aggressiveness of SCC.

In addition to the *SFRP2* gene, the three tumour suppressor genes *BMR1B, SMAD1* and *CASP6* were also targets of deletion at 4q22.1-q32.1. Both *BMR1B* and *SMAD1* genes have previously been reported to have decreased expression in gliomas correlating with poor survival (S. Liu et al., 2009), and *BMPR1B* decreased expression in breast cancer is associated with increased cell proliferation and poor prognosis (Bokobza et al., 2009). The *CASP6* gene encodes the pro apoptotic caspase-6 protein (J. Y. Chan et al., 2008).

Although large 3q amplicons are commonly observed in squamous carcinoma (Haverty et al., 2009; Kanao et al., 2005) in this study the 6Mb, 3q11.2-12.2, amplicon was focal and involved the *MINA* gene. This gene has previously been reported to be over expressed in 83% of OSCC in one study and its inhibition was shown to suppress OSCC cell proliferation (Tsuneoka et al., 2004).

A 43Mb region of deletion at 8p23.2-q11.1 (4078057-47043375bb) was observed in the five cell lines and involved 5 potential target genes in addition to the *SFRP1* gene. These included the *BNIP3L* gene deleted or down regulated in prostate cancer and malignant melanomas respectively (W. Liu et al., 2008; D. M. Su et al., 2009), the *LZTS1* gene, deleted in oral

squamous cell carcinomas and down regulated in breast carcinomas (L. Chen et al., 2009; Kanae Ono et al., 2003); and the three tumour necrosis factor receptors *TNFRSF10A*, *TNFRSF10B* and *TNFRSF10C* whose epigenetic inactivation was reported in gastric cancers (K. H. Lee et al., 2009). An 8p loss was previously detected by conventional CGH in a study performed on 29 South African black and coloured OSCC patients (Du Plessis et al., 1999). Chromosome 8p22 loss has also been reported in prostate, breast, lymphoma, hepatocellular and colorectal cancers (Bova et al., 1996; Di Benedetto et al., 2006; J. M. Flanagan et al., 2004; Thomassen et al., 2009; Ying et al., 2007).

Active smoking is linked to increased fragile site expression (Sozzi et al., 1997) and is also one of the primary risk factors associated with OSCC in South Africa (Pacella-Norman et al., 2002). We previously hypothesized that deletions affecting anti-oncogenes located at fragile sites may contribute to the etiology of OSCC in South Africa and reported *FHIT* intragenic deletions in these cell lines and a small cohort of patients (Willem et al., 2006). *FHIT* gene deletions were confirmed here supporting its role in OSCC carcinogenesis.

Deletion at 18q23 involved the *SMAD4* and *DCC* genes in all cell lines. Both genes have previously been reported to be down regulated in OSCC either by deletion, mutation or methylation (Metzger et al., 2004). Decreased expression of the *SMAD4* gene, a tumour suppressor of the TGF-β family signalling pathway, has been associated with OSCC tumour invasion (Fukuchi et al., 2002). The *DCC* gene was shown to be frequently methylated in OSCC tumour specimens (H. L. Park et al., 2008).

In summary of this chapter, breakpoints at 1p11 and 3p11 were recurrent in the 5 OSCC cell lines and may point to genes such as *EPHA3* that may be involved in OSCC carcinogenesis. Copy number alterations involved both amplicons previously reported in squamous cell carcinoma (8q24, 11q13 and 3q11) as well as novel regions of significant amplification (11q22.1-q22.3, 5p15.2 and 18p11.32). The finding that a significant number of genes that were amplified (*FGF3, FGF4, FGF19, CCND* and *C-MYC)* or deleted (*SFRP1* and *SFRP2* genes) are involved in the Wnt and FGF signaling pathways suggest that these pathways may be activated in these OSCC cell lines. These results warranted investigation in patient tumour samples and expression studies of these genes in both cell lines and patients' specimens.

CHAPTER 3

Assessment of expression levels of selected candidate genes in the 5 OSCC cell lines and fresh tissue samples

3.1 Introduction

Many potential candidate genes were identified from the copy number data obtained from the five OSCC cell lines WHCO1, WHCO3, WHCO5, WHCO6 and SNO in chapter 2. The *EPHA3* candidate tumour suppressor gene had deletions by array copy number analysis in four of the five cell lines. It is expected that these deletions may negatively affect the expression of the *EPHA3* gene in the respective cell lines WHCO1, WHCO3, WHCO6 and SNO. The 11q13.3 chromosomal region was amplified in the five cell lines and spanned the known oncogenes, *CCND1*, *FGF3*, *FGF4* and *FGF19*. *CCND1* and *FGF4* gene amplification was confirmed by FISH analysis in chapter 2. Similarly, chromosome 8q24 amplification was detected and amplification of the *c-MYC* oncogene in this region was also confirmed by FISH analysis. The *EPHA3*, *FGF3*, *FGF4*, *FGF19*, *CCND1* and c-*MYC* genes were selected to establish whether gene copy number correlated with expression levels.

Real time relative quantification by reverse transcriptase polymerase chain reaction (RT-PCR) is the most frequently used methodology to assess relative gene expression levels. Relative quantification allows for the assessment of the levels of expression of particular mRNAs in one sample relative to another i.e. there is more or less RNA present in the cancer sample compared to that of the normal sample. The levels of expression of a target gene are measured as a ratio to the reference genes. The purpose of this study was to assess whether the expression of selected genes was lower or higher in cancer cell lines compared to normal oesophageal tissue and relative quantification was sufficient to answer this question.

This chapter describes the real time RT-PCR experiments performed to assess the expression levels of the 6 mentioned genes in the five cell lines, WHCO1, WHCO3, WHCO5, WHCO6 and

SNO. The gene expression levels in the 5 cell lines as well as in 4 fresh tumour samples were compared to the expression levels in 5 samples from normal oesophageal tissue.

3.2 Materials and Methods

RNA extraction from cell lines

The cells from the five cell lines were isolated as described in chapter 2, section 2.2 and lysed in 2ml of RLT buffer (Qiagen RNeasy® midi kit). Frozen lysates were thawed at 37°C and the RNA was extracted using the RNeasy® midi kit as per the manufacturer's protocol. RNA was eluted into 100µl of RNase-free water; the elution was then repeated using 50µl of RNase-free water. RNA quantity and quality was then assessed using the Nanodrop-1000 and with the RNA 6000 Nano kit (Agilent Technologies Inc) on the Bioanalyser 2100 (Agilent technologies Inc).

Fresh OSCC tumour samples

Eleven biopsy samples were collected from patients presenting with dysphagia at the Department of Medical Gastroenterology, Charlotte Maxexe Johannesburg Academic Hospital. Upon patient informed consent and during standard medical procedure, biopsies from the tumour region as well as the normal appearing epithelium were collected, immediately snap frozen in liquid nitrogen and stored at -70°C. Four of these patients were confirmed to have OSCC by histopathology and one patient had atypia with features of basal cell hyperplasia. These four samples were used in the expression study. The remaining six samples had a normal epithelium and one of these was used as a control (WITSCA10N).

Fifty five samples were collected from patients presenting at the Chris Hani Baragwanath Hospital, Soweto upon informed consent. These samples were collected in collaboration with the Department of Gastroenterology. Two tumour biopsies were collected and two biopsies from normal epithelium were collected from each patient. One tumour and one normal specimen were placed into RNAlater® stabilisation reagent and frozen at -20°C. The other two samples were stored at -20°C for DNA extraction.

RNA extraction from tissue samples

The RNA stabilised samples were removed from RNAlater® (Qiagen®) and weighed to obtain the appropriate weight of no more than 30mg of tissue. A tissue weight of 30mg or less was placed into a new tube with 450µl of RLT buffer (from RNA extraction kit) with β-mercaptoethanol (10µl per 1ml of buffer RLT). Snap frozen samples were thawed on ice and 450µl of RLT with β-mercaptoethanol was immediately added. One 5mm stainless steel bead (Qiagen®) was inserted into each tube. The tissue was homogenised using the TissueLyser (Qiagen®) at 25Hertz for 2 minutes, the adaptor set was rotated and then homogenisation repeated for another 2 minutes. The tubes were spun down at 14000rpm for 3 minutes (Eppendorff centrifuge 5810R, CL021 rotor) and the supernatant was transferred to a new tube. The RNA was then extracted according to the RNeasy® mini protocol (Qiagen®). RNA quantity and quality was then assessed using the nanodrop-1000 and the RNA 6000 Nano kit (Agilent Technologies Inc) on the Bioanalyser 2100 (Agilent technologies Inc).

Target genes and primer design for RQ-PCR

The primers were designed for candidate and reference genes using the RTPrimerDB public database (Pattyn et al., 2003) (Table 3.1), taking into account the presence of SNPs in the region of primer binding, secondary structures and specificity to the target by BLAST analysis.

Table 3.1. Reference genes and target genes for RQ-PCR analysis.

Gene	Expected expression in normal oesophagus	Forward primer (5'-3')	Reverse Primer (5'-3')	Expected size
ACTB	High	AGCCTCGCCTTTGCCGA	GCGCGGCGATATCATCATC	87bp
YWHAZ	Median	ACTTTTGGTACATTGTGGCTTCAA	CCGCCAGGACAAACCAGTAT	94bp
PGK1	Low	GAGATTGGCACTTCTCTGTT	CAGGCAAGGTAATCTTCACA	95bp
MYC	Low	AAGACAGCGGCAGCCCGAAC	TGGGCGAGCTGCTGTCGTTG	147bp
CCND1	Low/Absent	CGTGCCCGTGTGCATGTCCT	GCTGCAGGGCCGTTGGGTAG	294bp
EPHA3	Low	CATGGTAACTTCTCCAGCAA	GAGAGCTGACAATCCATGTT	112bp
FGF3	Absent	TTGGCGAACCGTCCTTTCT	TCCACCCAGGAGCAATGAA	64bp
FGF4	Absent	CCAACAACTACAACGCCTACGA	CCCTTCTTGGTCTTCCCATTCT	83bp
FGF19	Absent	ATGCAGGGGCTGCTTCAGTA	AGCCATCTGGGCGGATCT	67bp

DNase treatment of RNA

RNA was first treated with DNase I post RNA extraction in order to eliminate contaminating DNA from the samples and to prevent amplification of pseudogenes. One microgram of RNA was treated at a time in a final volume of 10µl with 1U of DNase I (Invitrogen), 1x buffer (Invitrogen) and AccuGENE® molecular grade water (Lonza). The samples were incubated at room temperature for 15 minutes. DNAse I was inactivated by addition of EDTA (Invitrogen)

at a final concentration of 5mM and heated at 65°C for 10 minutes. cDNA synthesis was then immediately performed.

cDNA synthesis

cDNA was synthesized from DNAse I treated RNA immediately prior to RQ-PCR experiments. Sufficient cDNA was synthesized for each sample to perform all RQ-PCR runs for all the genes tested. For 1µg of RNA, 100 units of MMLV reverse transcriptase, 1x RT buffer, 1mM dNTPs, 24µM random primers, 10mM DTT and 20 units of RNase out was added in a final volume of 25µl. All the reagents are from Invitrogen. One microgram of RNA was set up in a 10µl reaction volume with RNAse free water (Invitrogen) and incubated for 10 minutes at 70°C, then cooled to 4°C. The cDNA master mix was then added to the RNA. The reaction was incubated at 25°C for 10 minutes, 42°C for 45 minutes, 99°C for 3 minutes and cooled to 4°C. The cDNA was kept at 4°C until RQ-PCR set up, for long term storage it was placed at -20°C.

RQ-PCR experiments

RQ-PCR was optimised for all the genes using the MCF7 breast cancer cell line as a template as this cell line was previously described to express all the targets of interest (Wang et al., 2003; Fox & Kandpal, 2004; Lukyanova et al., 2009). Melt-curve analysis and gel electrophoresis were used to determine that the correct product was being amplified based on expected size and presence of one melting peak (Appendix B). Reference gene stability was assessed (Appendix B), however all three reference genes were included in each RQ-PCR run and the most stable reference gene from the run was used in normalisation. The RQ-PCR experiments were set up on a 96 well plate, however, some samples had to be

carried over to additional plates. To correct for variation between runs, 3 samples (inter-run calibrators) were repeated on each plate. Each sample was set up in triplicate. Nine plates were run, 2 genes were tested on each plate and non template controls were included for each gene (Appendix B). The same cDNA batch for each sample was used for the analysis of all genes and 120ng of cDNA was added to the reaction. The five cell lines, 4 fresh OSCC samples (delineated OC8BT, OC7BT, WITSCA9 and WITSCA2) and 5 samples from normal oesophageal tissue (delineated WITS35N, WITS36N, WITS39N, WITS40N and WITSCA10N) were analysed. The final reaction volume (15µl) consisted of 7.5µl of Power SYBR®Green PCR master mix (Applied Biosystems) and forward and reverse primers (Operon) at the following concentrations, 0.227mM for c-*MYC*, *CCND1* and *EPHA3* and 0.454mM for the *FGF3*, *FGF4* and *FGF19* primers. The reaction was performed on an Applied Biosystems 7500 real time PCR system in standard run mode, 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds and 62°C for 1 minute (fluorescence data was collected at this step). A melt curve analysis was performed at the end by heating from 60°C to 95°C at 0.1^{o}C increments. The passive reference dye was ROX. Data was captured using the Applied Biosystems Sequence Detection Software version 1.3.1.

Data analysis

Run data was imported from the CSV files taken from the Applied Biosystems sequence analysis software into Biogazelle Qbase PLUS. The cycle threshold values (Cq values) were used for subsequent data normalisation and calibration to obtain calibrated normalised relative quantities (CNRQ) values. Qbase PLUS has built in algorithms to normalise the data based on the replicates, reference genes and the inter-run calibrators (Hellemans et al., 2007). First, the quality control of the runs was assessed by analysing the raw data to

evaluate the non-template controls and the difference in Cq values between replicates (experimental/pipetting error). The non-template controls were undetected and the cut-off value for the difference in Cq values between replicates was 0.5, a difference of 0.5 of a cycle between replicates was considered acceptable. The mean of the Cq values from at least 2 replicates were used to calculate relative quantities (RQ values). The RQ values for the samples were normalised by Qbase PLUS by dividing the RQ value for each sample by the normalisation factor calculated from the reference genes (Hellemans et al., 2007). The most stable reference gene (*YWHAZ*), as determined by GNORM (Appendix B), was used for normalisation. Inter-run calibration was then performed, using the Qbase PLUS algorithms, to adjust for the variation between runs using the three inter-run calibrators (WHCO1, WHCO3 and WHCO5) in order to obtain the calibrated normalised relative quantities (CNRQ) (Hellemans et al., 2007). A two tailed t-test was performed to test whether the relative expression values (CNRQ values) of the samples were different to the median relative expression of the controls using GraphPad InStat® version 3.10.

3.3 Results

Expression levels of the candidate genes, EPHA3, CCND1, FGF19, FGF3, FGF4 and C-MYC.

RNA from 4 fresh OSCC samples were of good enough quality and quantity to use in RQ-PCR analysis and although it was attempted to obtain additional specimens from normal epithelium, several biopsies were contaminated with cancerous cells on touch preparations analysed by a pathologist. Thus, the expression levels in the five cell lines and 4 OSCC samples were compared to the average expression of 5 normal oesophageal samples (Table

3.2). The normal OSCC controls, delineated as WITS35N, WITS36N, WITS39N and WITS40N were biopsy samples taken from neighbouring normal epithelium in patients with suspected OSCC. The WITSCA10N sample was a normal tissue sample subsequently found to be from a patient with normal oesophagus. The normalised, calibrated expression values (CNRQ) were log transformed and statistically analysed by a two tailed t-test to establish whether the expression levels of each of the candidate genes were statistically different from the median expression level observed in the normal samples.

Table 3.2. Table of samples used in expression analysis by RQ-PCR.

Sample ID	Sample type	Histology	RNA concentration	RIN
WHCO1	OSCC cell line	MD SCC	575ng/µl	9.5
WHCO3	OSCC cell line	MD SCC	329.7ng/µl	9.51
WHCO5	OSCC cell line	MD SCC	304.7ng/µl	9.5
WHCO6	OSCC cell line	MD SCC	463.2ng/µl	9.9
SNO	OSCC cell line	MD SCC	562ng/µl	9.4
MCF7	Breast Adenocracinoma cell line	Adenocarcinoma	1ug/µl	N/A
OC7BT	OSCC tumour tissue	MD SCC	419ng/µl	5.2
OC8BT	OSCC tumour tissue	Invasive SCC	289ng/µl	5.7
WITSCA2	OSCC tumour tissue		594.4ng/µl	
WITSCA9	OSCC tumour tissue		197.1ng/µl	8.4
WITSCA10N	Normal oesophageal tissue		122.6ng/µl	9.4
WITS35N	Normal oesophageal tissue		280.58ng/µl	
WITS36N	Normal oesophageal tissue		157.09ng/µl	
WITS39N	Normal oesophageal tissue		401.51ng/µl	
WITS40N	Normal oesophageal tissue		223.86ng/µl	

RIN numbers indicate the level of degradation of the RNA and high numbers close to 10 have low degradation and are considered good quality RNA.

MD= moderately differentiated

SCC= squamous cell carcinoma

***EPHA3* gene expression**

The *EPHA3* gene is usually highly expressed in the developing embryo. The expression levels were expected to be low in normal oesophageal tissue and late Cq values were obtained for the normal samples averaging at cycle 32 (Appendix B, Table B2). *EPHA3* gene expression was only detectable in the cell lines WHCO1 and WHCO3 with cell line WHCO1 showing higher expression relative to the controls ($p=0.0001$) and cell line WHCO3 showed significantly lower expression than the controls ($p=0.0076$) (Figure 3.1). The remaining samples had Cq values above the threshold limit in Qbase Plus and were considered as not expressed ($Cq>35$). The expression results correlated with deletions detected in the WHCO3, WHCO6 and SNO cell lines in chapter 2 and would suggest that these deletions may result in the lower or lack of expression observed in these cell lines. The finding that cell line WHCO5 does not show detectable levels of *EPHA3* may suggest that transcription is inactivated by other mechanisms, for example, inactivating mutations or silencing through promoter hypermethylation. The over expression observed in WHCO1, despite the observed deletion of exons 1-3 (chapter2), may be explained by the additional copies of the gene due to the precence of additional copies of chromosome 3 as determined by cytogenetics and FISH (chapter 2). It is important to note that the limitation of this experiment was that the primers were designed to bind in exon 1, which may not represent the expression of transcripts that have exon 1 deletions.

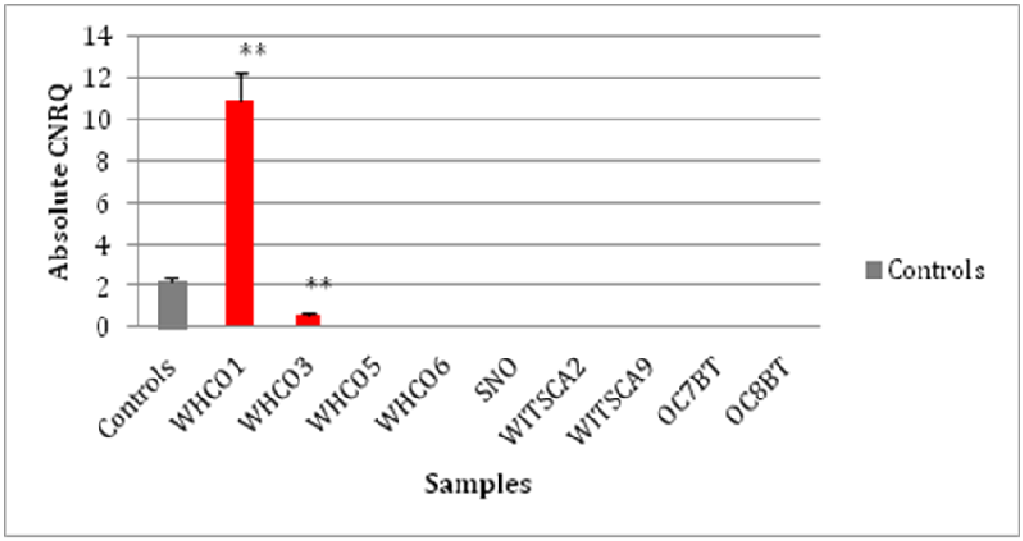

Figure 3.1 Relative expression levels of the *EPHA3* gene as measured by RQ-PCR. The corrected normalised relative quantities (CNRQ) are plotted standard error of the mean (±SEM). The expression levels are significantly higher than the controls for cell line WHCO1 *(p=0.0001) and significantly lower for cell line WHCO3 (p=0.0076). The other cell lines and OSCC samples had undetectable *EPHA3*.

CCND1 expression

The relative expression of *CCND1* was significantly increased compared to the controls for all the cell lines and three of the fresh samples (WITSCA2, OC7BT and OC8BT) (Figure 3.2). The most significantly different expression was detected in the cell lines WHCO6 and SNO ($p<0.0001$).

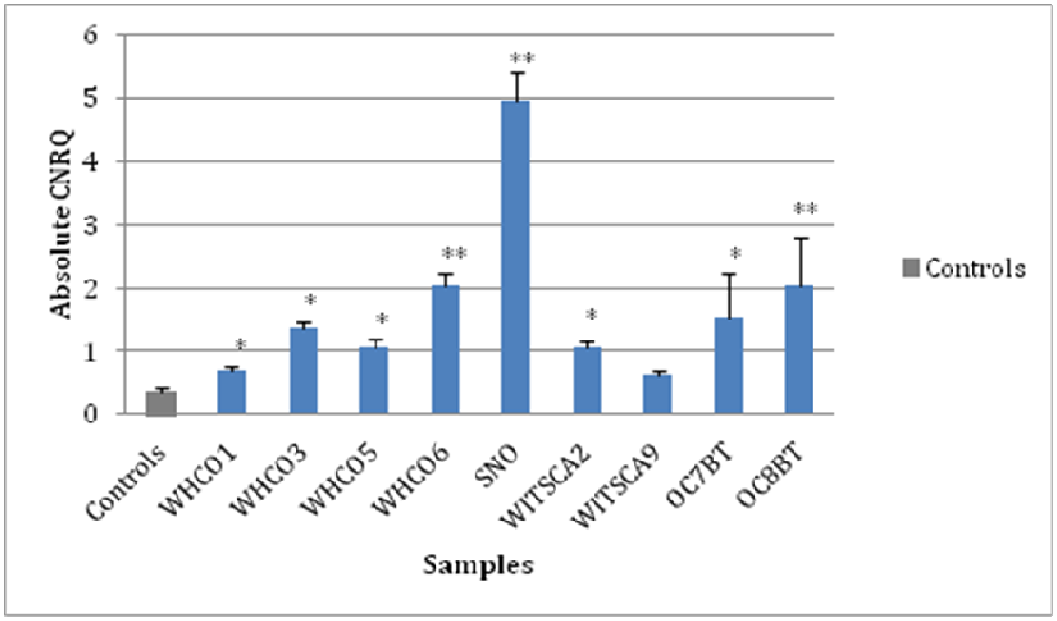

Figure 3.2 Relative expression levels of the *CCND1* gene as measured by RQ-PCR. The corrected normalised relative quantities (CNRQ) are plotted (±SEM). The expression levels are significantly higher *($p<0.05$) for cell lines WHCO1, WHCO3, WHCO5 and samples WITSCA2 and OC7BT and highly significant ** ($p<0.0001$) for the cell lines WHCO6, SNO and sample OC8BT compared to the controls.

FGF3, FGF4 and *FGF19* expression

The fibroblast growth factors are only expected to be expressed during embryological development and were not detected in the normal samples. *FGF3* gene expression was only detectable for the cell lines WHCO6 and SNO, which were significantly different from the controls, $p=0.0046$ and $p= 0.0013$ respectively, as well as in fresh samples OC8BT ($p= 0.0028$) and WITSCA9 ($p= 0.0007$) (Figure 3.3 and figure 3.4). The fresh sample WITSCA9 had the highest expression with more than 500 fold the expression compared to normal oesophagus, which had very high cq values above the detectable limit (averaged at around 36 cycles). *FGF4* gene expression was only significantly increased relative to the normal samples in cell line, WHCO1, and samples OC8BT and WITSCA9 (Figure 3.5).

The expression of the *FGF19* gene was only analysed in the 5 cell lines and two normal controls due to lack of sufficient RNA. *FGF19* gene expression was almost undetectable in the controls and all the cell lines had significantly higher levels of expression compared to the two normal samples run (WITSCA10N and WITS35N) (Figure 3.6). The cell lines WHCO5 and SNO showed the highest expression, almost 50 and 30 fold respectively.

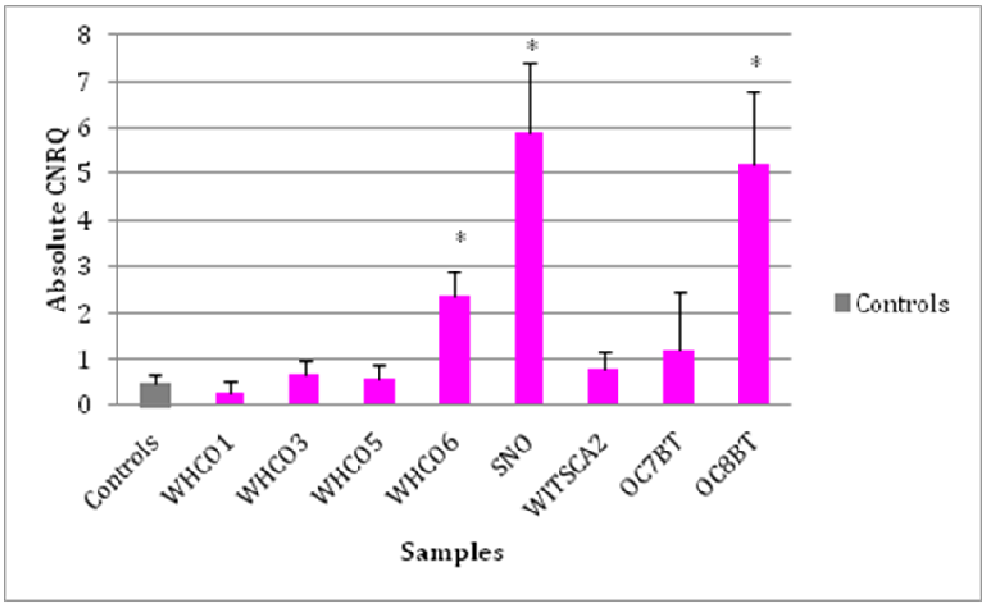

Figure 3.3 Relative expression levels of the *FGF3* gene as measured by RQ-PCR. The corrected normalised relative quantities (CNRQ) are plotted (±SEM). The expression levels are highly significant * ($p<0.005$) for cell lines WHCO6, SNO and sample OC8BT compared to the controls.

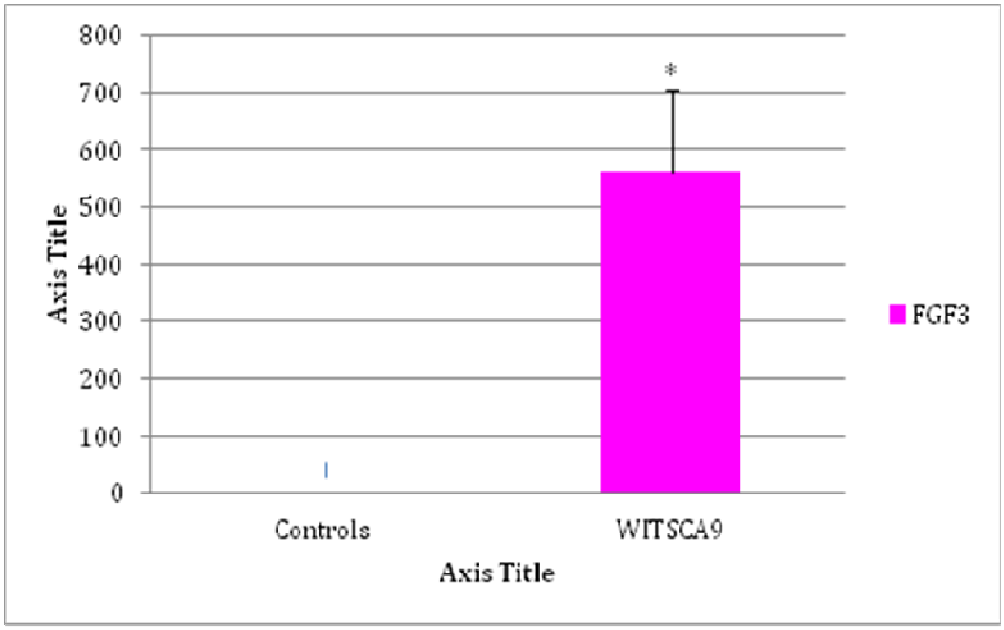

Figure 3.4 Relative expression levels of the *FGF3* gene as measured by RQ-PCR for the sample WITSCA9. The corrected normalised relative quantities (CNRQ) are plotted (±SEM) relative to the controls (indicated by the blue line due to the scale). The expression levels are highly significant * (p=0.0007) compared to the controls.

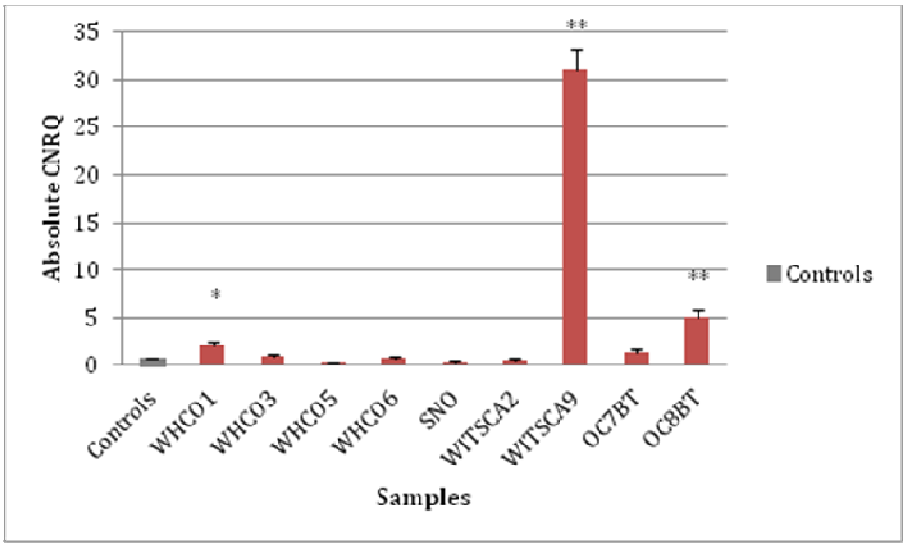

Figure 3.5 Relative expression levels of the *FGF4* gene as measured by RQ-PCR for the. The corrected normalised relative quantities (CNRQ) are plotted (±SEM). The expression levels are significant * for cell line WHCO1 (p=0.0019) and highly significant for samples WITSCA9 and OC8BT (p<0.0001) compared to the controls.

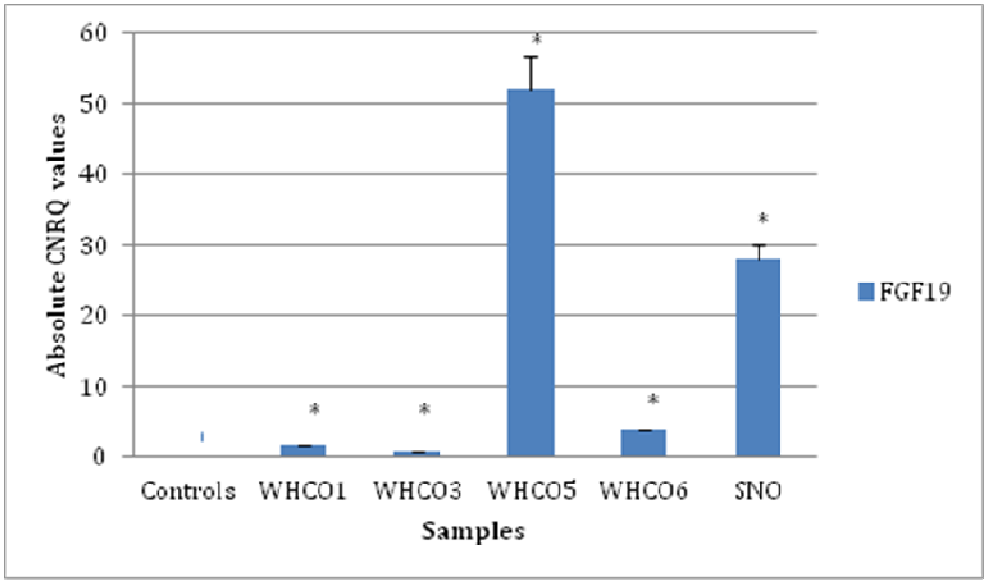

Figure 3.6 Relative expression levels of the *FGF19* gene as measured by RQ-PCR. The corrected normalised relative quantities (CNRQ) are plotted (±SEM), the controls are indicated by the blue line due to the scale. The expression levels are highly significant * ($p<0.01$) for all five cell lines compared to the controls.

C-MYC expression

The expression of the C-*MYC* gene was statistically significantly higher than the controls in the cell lines WHCO1 ($p<0.0001$), WHCO3 ($p=0.0002$), WHCO5 ($p<0.0001$) and SNO ($p=0.0004$) and sample WITSCA2 ($p=0.0004$) (Figure 3.7).

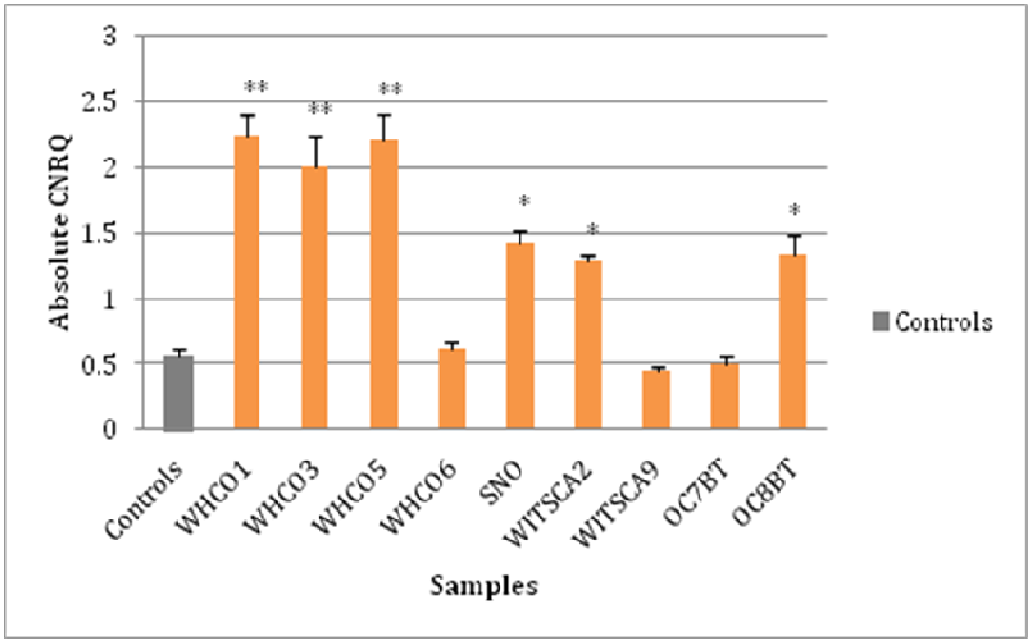

Figure 3.7 Relative expression levels of the *C-MYC* gene as measured by RQ-PCR. The corrected normalised relative quantities (CNRQ) are plotted (±SEM). The expression levels are highly significant **($p<0.0001$) for cell lines WHCO1, WHCO3 and WHCO5 and also highly significant *($p<0.001$) for cell line SNO and the samples WITSCA2 and OC8BT compared to the controls.

3.3 Discussion

Expression of the EPHA3 gene in the 5 cell lines

EPHA3 gene expression was expressed at low levels in normal oesophageal tissue (late cycle thresholds of between 31 and 33 cycles for the 5 normal samples). The *EPHA3* gene had deletions in cell lines WHCO1, WHCO3, WHCO6 and SNO as determined by array copy number array analysis in chapter 2. Apart from the cell line WHCO1, which had higher expression levels than the controls (possibly explained by the prescence of additional copies of chromosome 3), the cell lines WHCO3, WHCO5, WHCO6 and SNO had low or no expression of the *EPHA3* gene. The copy number results therefore correlated well with the

expression data for cell lines WHCO3, WHCO6 and SNO and would suggest that the observed deletions do affect the expression of the *EPHA3* gene in these cell lines. Cell line WHCO5, where no deletion was observed in the *EPHA3* gene, may have alternative mechanisms for the lack of *EPHA3* gene expression, such as epigenetic changes or inctvating mutations. The four fresh tissue samples from OSCC patients also had no detectable levels of *EPHA3* expression. These results would have to be confirmed in a large number of samples before any conclusion can be made about *EPHA3* expression in OSSC samples. However, these results taken together with the array results highlight the importance of testing *EPHA3* levels of expression , as well as its known isoform (consists of exons 1-7) and the other genes in the EPH/ephrin network in both cell lines and fresh tumour samples. It would also be beneficial to do RT-PCR to assess the expression of all possible variants in these cell lines.

Expression of the CCND1, FGF3, FGF4, FGF19 and c-MYC candidate genes

The 11q13.3 amplicon detected in the cell lines is frequently reported in OSCC . There are a number of genes within this amplicon and it is important to establish which of them are specifically targeted by amplification. It is hypothesised that genes likely to drive oncogenesis are specifically targeted for amplification and would be expected to be over expressed.

The chromosomal region 11q13.3 was highly amplified in cell lines WHCO1, WHCO3, WHCO5 and SNO, and included the *CCND1*, *FGF3*, *FGF4* and *FGF19* genes. The increased relative expression of the *CCND1* gene was found and correlated with amplification of this gene (as detected by array and FISH analysis), where 15 to more than 20 copies of the gene were detected in cell lines WHCO1, WHCO3, WHCO5 and SNO. However, WHCO1 and

WHCO3 cell lines had high *CCND1* gene copy numbers (15-20 copies as determined by FISH in chapter 2), but did not show the same high level of expression as the other cell lines with high copy numbers. Mechanisms for the lack of correlation between gene copy number and gene expression have been suggested through study of copy number variants and their effects on gene expression (Henrichsen et al. 2009). One mechanism, which may be applicable to *CCND1* in WHCO1 and WHCO3 where copy number did not show complete correlation with expression, could be that extra copies of the gene cause steric hinderence, impairing access to the transcriptional machinery (Sexton et al. 2007). Cell line WHCO6 had 4-8 copies of *CCND1*, reflecting low to mild amplification; however this cell line still showed high expression of the *CCND1* gene. In contrast *FGF4* gene expression levels were not significantly higher for 4 of the cell lines and therefore the expression levels for this gene did not correlate with the *FGF4* amplification status in the cell lines, which could possibly explained by a loss of the regulatory elements of the gene during the amplification event, a mechanism reviewed by Henrichsen et al (2009).

Similarly, the *FGF3* gene did not show an increased expression as would be expected from the array copy number data, while the *FGF19* gene showed over expression relative to the normal oesophageal tissue in all 5 cell lines. This would suggest that the *CCND1* and *FGF19* genes are important targets of amplification in the 11q13.3 amplicon in the 5 cell lines. Fresh oesophageal tumour samples showed higher expression of *CCND1* in three of four samples and of the *FGF3* and *FGF4* genes in two of four samples. *FGF19* gene expression was not assessed in the fresh samples due to lack of RNA. The expression of all these genes therefore needs to be assessed in a larger set of fresh samples in order to establish their importance in OSCC.

The FISH and array data showed the presence of 4 to more than 20 copies of the C-*MYC* oncogene in cell lines WHCO3, WHCO5, WHCO6 and SNO while cell line WHCO1 had 4-6 copies. All cell lines, except cell line WHCO6, had significantly higher expression of c-*MYC* compared to normal oesophageal tissue showing a correlation with the gene copy numbers. Cell line WHCO6, which had 10 to >20 copies, did not show statistically significant increased c-*MYC* expression. Higher expression was detected for cell line WHCO1, which had 4-6 copies. There was therefore no consistent correlation between C-*MYC* increased copy number and expression. The C-*MYC* oncogene is therefore likely to be targeted for amplification in the cell lines WHCO3, WHCO5 and SNO but amplification may be compounded by other mechanisms for C-*MYC* increased transcription as was observed for cell line WHCO1. In addition gene amplification may not always result in over expression, as observed for cell line WHCO6. The C-*MYC* gene was highly expressed in 2 out of 4 OSCC samples and more samples will need to be analysed to establish the frequency of C-*MYC* over expression in SA OSCC.

In summary, deletions in the *EPHA3* gene correlated with a low expression in the cell lines. In addition, four fresh samples showed no expression of *EPHA3*. The lowered expression of *EPHA3* compared to normal oesophageal tissue further supports a role for this gene in OSCC and warrants further investigation in a larger cohort of fresh samples. The 11q amplicon and 8q amplicon correlated with higher expression of the *CCND1*, *FGF19* and C-*MYC* genes respectively, pin pointing these genes as likely targets for amplification in the cell lines. The high expression of the *FGF19* gene observed in the cell lines warrants assessment of *FGF19* gene expression in fresh samples. Similarly, *CCND1* and C-*MYC* genes were over expressed in 3/4 and 2/4 of the fresh samples analysed respectively, together with the cell lines

suggesting their involvement in OSCC. Again, expression levels need verification in a larger sample size.

This was a pilot study to assess the expression levels of the *EPHA3* gene and the genes in amplified regions in the cell lines. The results do indicate that C-*MYC*, *CCND1* and *FGF19* genes are likely to be over expressed due to gene amplification in both cell lines and tumour samples. The *CTTN* gene, which was part of the 11q13.3 amplicon, is also an important candidate gene over expressed in pre-cancerous lesions of OSCC (Man-Li Luo et al. 2006). This gene should also be investigated for over expression in these cell lines and fresh tumours in a future study.

The *EPHA3* gene deletions resulted in low or no expression. This part of the study was however limited by the number of fresh samples available and conclusions regarding the involvement of these genes in OSCC in SA cannot be made at present. Additional limitations were that only 5 normal samples were available for comparison and only one reference gene could be used for data normalisation. In future studies, a larger selection of reference genes will be tested to obtain at least three stably expressed genes for more accurate normalisation of RQ-PCR data and a larger collection of fresh OSCC tumour samples and normal tissue will be assessed.

CHAPTER 4

Characterisation of copy number profiles in an OSCC cohort from the Eastern Cape

4.1 Introduction

Copy number studies in OSCC

The development of comparative genomic hybridisation and high resolution DNA copy number microarrays has allowed for the assessment of patterns of common copy number changes in a wide variety of cancers. Detection of these patterns of DNA copy number aberrations in tumours can highlight key genes and molecular pathways involved in carcinogenesis. DNA copy number aberrations have been assessed in OSCC in approximately 15 studies, which utilised a variety of platforms, from low resolution conventional CGH (8 studies) to 250K SNP array technology (1 study) (summarised in table 4.1). There is a high degree of variation in the results from each group, which makes it difficult to interpret regions of importance in OSCC carcinogenesis. However, the most commonly affected regions are amplification of 3q, 5p, 7p, 8q, 11q, 20q and deletion of 1p, 3p, 5q, 8p, 9p and 18q (Z. W. Chang et al., 2009; Y. R. Qin et al., 2008; Y. R. Qin et al., 2005; C. C. Yen et al., 2001; C. C. Yen et al., 2005; Xu Zhang et al., 2008; Chen, et al., 2008; Du Plessis et al., 1999; N. Hu et al., 2006; Carneiro, et al., 2008; Chattopadhyay et al., 2010; Y. C. Chang et al., 2010; Sakai et al., 2010). It is possible that variations in copy number results across these studies reflect geographical differences in OSCC carcinogenesis. Only one study was performed in SA using low resolution conventional CGH to establish DNA copy number patterns from 29 patients (Du Plessis et al., 1999). This study was performed on a mixed population including mixed ancestry, Xhosa speaking black patients and Caucasian patients.

Table 4.1 Summary of studies on copy number changes in OSCC

Author and country	Method	No. samples	Reported gains	Reported losses
(Du Plessis et al. 1999), South Africa	Conventional CGH	29, Xhosa, mixed ancestry and caucasian	3q, 2q, 5p, 7p, 7q, 8q, X, 1p in Xhosas	1q, 1p, 4p, 18q, 22q, 1p36-pter, 8p in males, Xp in black males, 5q21-22, 5q31-qter
(C. C. Yen et al. 2001), China	Conventional CGH	46, Chinese	1q32-42, 2q31-q32, 3q26-qter, 5p13-p14, 7p14-pter, 8q24-qter, 11q12.3-q13, 12p12, 14q23-qter, 17q24-qter, 20q13.1-qter, xq27-q28	1p34.1-pter, 3p23-pter, 4p14-p15.3, 5q32-qter, 8p22-pter, 9p21, 9q21, 11q23-qter, 13q12-q14, 16p13.1-pter, 18q22, 19p, 19q
(Y.-R. Qin et al. 2005), China	Conventional CGH	52, Chinese	3q, 8q, 5p, 1q, 6q, 18p, 20q	3p, 1p, 9q, 19p, 4p, 8p
(Y.-R. Qin et al. 2008), China	Conventional CGH	37, Chinese	1p31, 1p11-q22, 3q22-qter, 4p13, 5p, 6p12-14, 8q, 9p, 11q13-q23, 16p11, 18p, 20p	3p, 8p, 9q, 11q, 13q, 5q, 1p, 4p, 18q, 16q, 17p, 19p
(Y.-R. Qin et al. 2008)	Conventional CGH	37 ,Chinese	8q, 3q, 5p	3p, 8p, 9q
(Z.-W. Chang et al. 2009), China	Conventional CGH	13, Chinese	3p, 5p, 7p, 8q, 10q, 15q	3p, 9q, 19q
(Y. C. Chang et al. 2010), Taiwan	Conventional CGH	40, Taiwanese	2q, 3q, 8q, 13q	1q, 4q, 3p, 5q, 18q
(Sakai et al. 2010), Japan	Conventional CGH	51, Japanese	17q12, 17q21, 3q29, 3q28, 8q24.2, 22q12, 3q27, 8q24.3, 1q22, 5p15.3, 22q13, 3q26.3, 8q23, 8q24.1, 9q34, 11q13, 17p12, 17q25, 20q12, 20q13.1, 1q32, 1q42, 20q13.2	3p26, 13q21, 4q32, 13q22
(C.-C. Yen et al. 2005), China	Copy number changes of target genes, Q-PCR, IHC	60, Chinese	3q25.3-qter, *TP63* gain and over expressed, *ECT2*, *PIK3CA*	
(Carneiro, Isinger, Karlsson, J. Johansson, Jönsson, et al. 2008), Sweden	32K BAC array	30, Swedish	5p15.33, 7p22.3, 7p11.2, 8q24, 10q, 11q13.3-q13.4, 12p13.33, 14q32, 16p13.3, 17p, 19p, 19q, 20q13.33	3p26.3-p24.2, 3p14.2-p14.1, 3p14.1-p13, 5q12.13, 8p23.2, 9p24.3-p24.2, 9p21, 11q25
(Jing Chen, L. Guo, Peiffer, L. Zhou, O. T. M. Chan, Bibikova, Wickham-Garcia, S.-H. Lu, Q. Zhan, Wang-Rodriguez, Wei Jiang & J.-B. Fan 2008b), China	Illumina SNP bead array, 766 genes	33 FFPE ,Chinese	*DVL3, ABCC5, TNK2, SKIL, TERT, CCND1, FGF12*	*FANCD2, EPHA3, CTNNB1, FGF2, APC, CDKN2A, CDKN2B, DCC*
(Y. R. Qin et al. 2008)	SNP, LOH analysis of 386 markers on 3p	100, Chinese		3p26.3, 3p22, 3p21.3, 3p14.2
(N. Hu et al. 2006), China	10K SNP array	26 Chinese	1p36, 2p14, 2q11.2, 3q, 5p15.2, 6p25.3, 7p14, 7p12, 7q11.21, 8q24.2, 9q31, 14q21, 17p13, 18p11, 20p11.1-q11.1, 22q12.1	1p13.3, 3p24, 4q28, 5q33, 9p21.3, 10p12.1, 11p15.5, 11q21, 13q12.2, 18q
(Chattopadhyay et al. 2010), India	10K SNP arrays	20, Indian	1p36.13, 1p21, 1q21,1, 2p25.3, 2q14.1, 3q28, 3q29, 4p15.2, 4q21.23, 5p15.2, 6p25.3, 7p21, 10q21.2, 11p13, 11q12.2, 12p13, 17q21.2, 18q11.2, 20p13-p11.12	1p36.32, 3p26-p14.3, 5q32, 6p21, 6q13, 6q23, 8p23, 8p21.1-p12, 10p15.3-p11.21, 12q15, 13q12.11, 17q21.31, 18q21.1, 22q11.21
(Xu Zhang et al. 2008), China	250K Nsp SNP array	5, China	1p, 1q, 2p, 2q, 3q, 4q, 5p, 6p, 6q, 7q, 8q, 11p, 11q, 12p, 12q, 14q, 20q, 18p, 22q	1p, 2q, 3p, 3q, 4p, 4q, 8p, 9p, 9q, 10q, 11p, 13q, 16p, 18q, 19p, 19q, 22q

The variation in chromosomal regions affected in studies across the world and the lack of high resolution data, warrant new studies, especially in South Africa where the incidence of OSCC is one of the highest in the world . The last objective of this project was therefore to identify common copy number changes in 50 tumour samples obtained from patients with OSCC using high resolution Affymetrix® 500K SNP array technology. Patients' samples were all collected from the Eastern Cape where the highest OSCC incidence is reported.

Challenges of OSCC pathology sampling

OSCC is a malignancy of the epithelium of the oesophagus with squamous cell differentiation (Domanowski G, www.eMedicine.medscape.com, Accessed 21.04.2009). Below are the basic stages and histological characteristics of these tumours. They may have extracellular or intracellular keratinisation. Keratin is deposited at the site of the trauma on the surface of the epithelium; these traumas may be physical, thermal or chemical. Dysplasia describes a lesion where the cells show atypia i.e. enlargement and darkening of the nuclei and distortion of the nucleus shape. In dysplasia the architecture is abnormal; there are an increased number of cell layers described as hyperplasia. The thickness of the atypical cell layers determines the severity of dysplasia and if the bottom of the lesion is not distinguishable, looking at the epithelium alone, this is usually considered as severe dysplasia (Domanowski G, www.eMedicine.medscape.com, Accessed 21.04.2009). Carcinoma *in situ* is represented by full thickness, atypical squamous epithelium without invasion of cancerous cells beyond the basement membrane. Invasive carcinoma is characterised by atypical squamous cells having penetrated the basement membrane. One of the problems with invasive SCC is that not all of these tumours have gone through a progression from dysplasia, to carcinoma *in situ*, to invasive carcinoma, often resulting in a

normal appearing surface epithelium hiding a de novo invasive SCC (Domanowski G, eMedicine.medscape.com, Accessed 21.04.2009). The degree of differentiation is a subjective feature of the tumour which can be difficult to determine yet it helps in evaluating the potential of a tumour to metastasize which affects treatment outcome. Generally, if it is difficult to determine that the cells are squamous in origin the tumour is considered poorly differentiated, if there is obvious squamous differentiation then the lesion is classified as well differentiated and anything in between is moderately differentiated (Domanowski G, eMedicine.medscape.com, Accessed 21.04.2009). There is often a lot of variability in one tumour, with varying degrees of differentiation occurring in different areas. Some of these particularities make it difficult for researchers to obtain fresh samples from tumours with a particular stage of disease or differentiation type as well as obtaining normal epithelium from the same patient. Initially, the objective was to collect fresh tumour samples and neighbouring normal epithelium. However, the size and quality of the samples we obtained was unsuitable for copy number analysis due to the presence of admixture of normal cells in tumour samples and the presence of tumour cells in what was apparently normal epithelium.

In view of the above, formalin fixed paraffin embedded (FFPE) tissue specimens were considered for this study. The tumour area can be ringed on these specimens and a clear histological examination is made to appropriately select the tumour material for analysis. In addition, a set of normal FFPE samples from a matched population has to be used as a reference set for copy number estimation. Using a matched population should account for population specific copy number variations that naturally exist and therefore highlight only somatic cell derived copy number aberrations.

Common copy number aberrations in an OSCC cohort of 51 (FFPE) tumour samples collected in the Eastern Cape region, the similarities/differences of these copy number changes to those of the cell lines and the comparison to published data are described in this chapter.

4.2 Materials and Methods

Oesophageal cancer cohort

One hundred and seventy seven FFPE tissue specimens were collected from the Nelson Mandela Academic Hospital in Umtata, Eastern Cape. These samples were biopsies taken from patients presenting at the hospital with dysphagia and were confirmed as OSCC by the histo-pathologists from the Department of Anatomical Pathology, Nelson Mandela Academic Hospital. The samples were selected from archive, based on histo-pathology reports. Each sample had a Haematoxylin and Eosin stained slide from routine diagnosis. These slides were used by the collaborating pathologist to ring the tumour area containing more than 80% tumour cells. The ringed slide was used to match the tumour region in the block. The tumour region was then macro-dissected out using a scalpel.

Control samples

A collection of normal samples from a matched population was retrieved from archived material and used as a reference set for copy number analysis. Fifty seven FFPE tissue specimens from a range of tissue sites including fallopian tube, ovary, prostate, appendix, epididymis, oesophagus, liver, oral mucosa, uterus, small bowel and skin were selected. All these blocks were confirmed to be normal by histopathology examination at the Nelson Mandela Academic Hospital, Department of Anatomical Pathology. The normal tissue was macro-dissected for DNA extraction in the same manner as for the tumour sample sin order to keep the methodology consistent for all FFPE samples.

DNA isolation

The macro-dissected tissue was placed in a 1,5ml centrifuge tube. An area of similar size to the patient specimens in the FFPE block was chosen for each control specimen and macro-dissected in the same way. De-paraffinisation was performed by adding 1ml of xylene (Merck laboratory supplies) and incubating at room temperature for 5 minutes. The samples were centrifuged for 3 minutes at 13000rpm (Eppendorff centrifuge 5810R, CL021 rotor) and the ethanol removed. This step with xylene was repeated twice more. To wash the pellets, 1ml of 100% ethanol was added and the samples vortexed. They were then centrifuged for 3 minutes at 13000rpm (Eppendorff centrifuge 5810R, CL021 rotor). The ethanol was poured off and the pellets allowed to air dry. One millilitre of 1M sodium thiocyanate (SMM) was added and the samples were incubated in an Eppendorf Thermomixer® Compact (shaking at 1000rpm) overnight at 39°C. The samples were then centrifuged for 3 minutes at 13000rpm (Eppendorff centrifuge 5810R, CL021 rotor), the sodium thiocyanate was removed completely and allowed to air dry. Five hundred microlitres of lysis buffer (Appendix A) was added as well as 40µl of proteinase K (Roche) (10mg/ml). The samples were incubated overnight at 56°C in the shaking heating block. The samples were checked for homogenisation, if longer digestion was required additional proteinase K was added and the samples monitored until complete homogenisation was achieved. Phenol-chloroform extraction was then performed. Addition of equal volumes of phenol and chloroform (dependant on the total volume of the lysate) was followed by vortexing and then centrifugation was performed at 13000rpm for 3 minutes (Eppendorff centrifuge 5810R, CL021 rotor). The aqueous phase was transferred to a new tube and half volumes of phenol and chloroform was added. The samples were vortexed and then centrifuged at 13000rpm for 3 minutes (Eppendorff centrifuge 5810R, CL021 rotor). The

aqueous phase was transferred to a new tube and 500µl of chloroform was added. The samples were vortexed and centrifuged at 13000rpm for 3 minutes (Eppendorff centrifuge 5810R, CL021 rotor). The aqueous phase was transferred to a new tube and the DNA was precipitated by adding 1/10 volume of 3M Sodium Acetate (Merck) and 2.5 volumes of ice-cold, 100% ethanol. The samples were mixed by inversion and then placed at -70°C overnight. The samples were then centrifuged at 13000rpm for 30 minutes at 4°C (Eppendorff centrifuge 5810R, CL021 rotor). The supernatant was removed and the DNA pellets were washed in 200µl of ice-cold 70% ethanol. Centrifugation was performed at 13000rpm for 10 minutes to pellet the DNA (Eppendorff centrifuge 5810R, CL021 rotor). The ethanol was removed and the DNA was allowed to air dry before re-suspending the DNA in an appropriate volume of 1XTE at pH7 (Promega), depending on the pellet size, 30-100µl.

Assessment of DNA integrity

FFPE samples were assessed for the proportion of high molecular weight DNA as they were generally degraded to some extent. They were assessed by gel electrophoresis and run on a 2% agarose (Bioline) gel at 100mV for 45 minutes in 1X TAE buffer (Appendix A).

DNA Quantification

The samples were quantified by spectrophotometry using the ND-1000 spectrophotometer (Nanodrop® Technologies). The A260/280 and A260/230 ratios were recorded.

Sample preparation

The DNA samples were diluted in a separate aliquot to 50ng/µl using AccuGENE® molecular grade water (Lonza). They were assessed on the Nanodrop-1000 to check the concentration.

Multiplex PCR to assess PCR performance of DNA from FFPE samples

An additional step to assess the performance of PCR amplification of large fragments in DNA from FFPE was done. The performance in PCR is linked to degradation and cross-linking in FFPE tissue. A multiplex PCR designed to amplify different regions of the *GAPDH* gene on chromosome 12 was used to assess this (van Beers et al., 2006). Seven fragments, at increments of 100bp were amplified from 100 to 700bp in size. Each primer was added at a specific concentration (Table 4.2) as advised by the published method (van Beers et al., 2006). The reaction was performed in a total reaction volume of 30µl, and was made of 1x PCR buffer (Bioline), 1mM dNTPs (Promega), 1.4mM $MgCl_2$ (Bioline), 1 unit of Biotaq taq DNA polymerase (Bioline), primers (Operon) and 50ng of DNA.

Table 4.2 Multiplex primers and final concentrations of primers used for PCR

Primers	Forward	Reverse	Final Concentration
100bp	5'-GTTCCAATATGATTCCACCC-3'	5'-CTCCTGGAAGATGGTGATGG-3'	266nM
200bp	5'-AGGTGGAGCGAGGCTAGC-3'	5'-TTTTGCGGTGGAAATGTCCT-3'	133nM
300bp	5'-AGGTGAGACATTCTTGCTGG-3'	5'-TCCACTAACCAGTCAGCGTC-3'	133nM
400bp	5'-ACAGTCCATGCCATCACTGC-3'	5'-GCTTGACAAAGTGGTCGTTG-3'	67.5nM
500bp	5'-AGCCCCTAAGGTCTTCAAGC-3'	5'-CATGCCTGTAGCTGGGACTA-3'	266nM
600bp	5'-GGCTCCCTTGGGTATATGGT-3'	5'-GGAGCCAGTCTTGGATGAGA-3'	266nM
700bp	5'-CCCCACACACATGCACTTAC-3'	5'-AATGAAGGGGTCATTGATGG-3'	266nM

The reaction was performed with the following cycles, initial denaturing for 14 minutes at 95°C, 35 cycles of 94°C for 1 minute, 57°C for 1 minute and 72°C for 3 minutes. Final elongation was done at 72°C for 10 minutes and held at 15°C. PCR products were resolved

on a 2% agarose gel, run at 100mV for 1 hour. An example of the results for this PCR is shown in Figure 4.1.

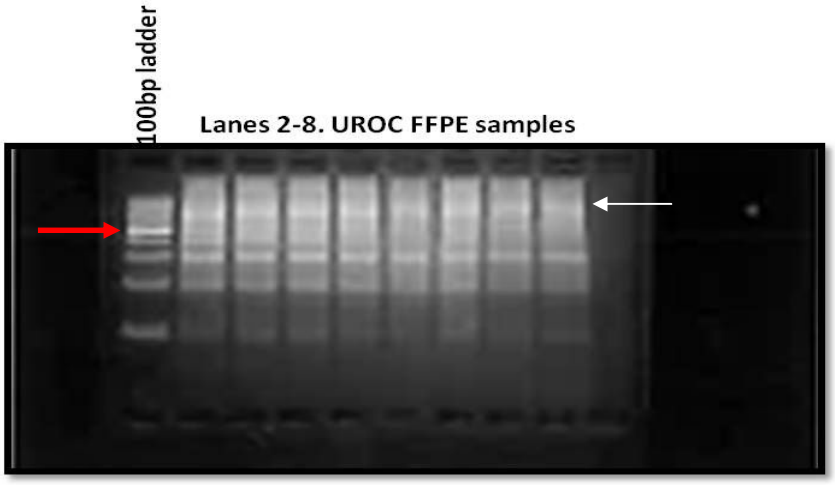

Figure 4.1 Example of the results for a multiplex PCR on 8 FFPE OSCC samples that amplified all 7 fragments of the *GAPDH* gene. The red arrow indicates the 500bp mark. Fifty five of the extracted samples amplified up to 700bp (white arrow).

Affymetrix 500K Genotyping assay for copy number analysis of the OSCC cohort

The 500K assay, consisting of the Affymetrix® 250K Nsp I and 250K Sty I GeneChips, was used for interrogation of copy number changes in the cohort of OSCC samples. The FFPE control samples were used as the reference data set for copy number estimation. The Affymetrix Genechip® mapping 500K protocol (P/N 701930 Rev. 3) was followed strictly apart from a modification in the number of PCR reactions per sample. Five hundred nanograms of DNA was used in total. The protocol involved restriction digestion with either Nsp I or Sty I , adaptor ligation, PCR amplification, fragmentation, labelling , hybridisation to the GeneChips® and washing the following day (Appendix C). Quality control gels were run after both the PCR reaction and after fragmentation. The protocol stipulates three PCR reactions per sample to yield an optimal 90µg of PCR product. FFPE samples do not amplify

as efficiently as other DNA sources and it was found that 6 separate PCR reactions per sample were optimal to yield 90µg of PCR product. Six PCR reactions were set as the limit due to cost and samples which failed to yield a sufficient quantity of PCR product were excluded from further experiments. Hybridisation was performed in the Affymetrix® hybridisation chamber for 18 hours. Post hybridisation washing was performed using the Affymetrix® Fluidics 450 station. Scanning was done by the Affymetrix® GeneChip® Scanner 7G. The scanned chips' images were converted to CEL files (signal intensity files); these files were used to assess the assay performance based on the SNP call rates which were generated in Genotyping Console™ (Affymetrix®).

500K data analysis

The Affymetrix®, Genotyping Console™ 2.0 (Affymetrix®) was used to generate call rates for quality control. Subsequently the data was analyzed with third party software, Partek® Genomics Suite (PGS). The performance of the assay on FFPE samples determines the quality of the array data. Call rates of 60% or more for FFPE were accepted. The Nsp call rates were above 70% for 51% of the samples and the Sty call rate was above 70% for 76.5% of the samples (Tables C1 and C2 in appendix C). Affymetrix array based PCR amplification should yield products ranging from 100bp to 1100bp, however, due to degradation and cross-linking of DNA extracted from FFPE specimens, the largest fragment size amplified was always less than 1100bp and varied from sample to sample. All samples amplified up to at least 800bp. In order to be stringent, SNP probes that reside on fragment sizes larger than 700bp were excluded from the analysis in PGS. This lowered the array resolution by 30% (since approximately 30% of the SNP data were lost), there was however sufficient data for analysis as per reported validation (Jacobs et al., 2007). The CEL files generated by the

Affymetrix command console were imported into PGS. During importation, quantile normalisation was performed to adjust for background. At this stage exclusion of large PCR fragments was performed. There may be variation in signal intensity from one batch of arrays to another since batches are run at different times. This variation was corrected in PGS by analysis of variance (ANNOVA), using the scan date for Sty and Nsp arrays. The raw copy number was then estimated from the model file generated from the intensities of the normal controls. Copy number data were then smoothed using the segmentation algorithm (Picard et al., 2011), a segment was defined as a minimum of 10 markers with the same copy number and the p-value cut off for a neighbouring segment having a different mean copy number was set to the default (p= 0.001). Segments with a minimum of 10 data points were reported when a small p-value (<0.001) was obtained in comparing signal intensity of the adjacent regions (breakpoints/ boundaries) (i.e. significant difference in copy number from the next set of data points, segments could then cover multiple data points (>10)). Regions of significant gain or loss on individual samples were reported using a default range of <1.7 or >2.3 with a p-value threshold of 0.01 (i.e. the copy number was significantly different from an expected value of 2 copies). Common regions of gain or loss across multiple samples were reported when a minimum of 10 samples showed the same aberration and the genes included within these common regions were reported using the Refseq database, genome build hg18. For the multiple sample analysis, PGS reports the average copy number for a region across all samples and specified the number of samples with that aberration. In a homogeneous sample with no contaminating normal cells, the presence of 1 copy would be equivalent to a heterozygous loss; however in cancer samples heterogeneity for any particular abnormality can be expected, which may affect the average copy number. For this reason, in manual analysis the cut off value for homozygous deletion

was chosen as <0.7 and >0.7<1.3 for hemizygous deletion. Copy numbers considered highly amplified range in the literature from above 3.7 copies to above 7 copies (Beroukhim et al., 2007; Xiaojun Zhao et al., 2005). In the manual analysis, the cut off values for amplification were chosen as >10 copies for highly amplified and >5 but <10 copies for moderately amplified.

The publically available software, EASE version 2 was used to determine the biological pathways in which genes identified in regions of copy number change, may be involved. EASE is free software from the National Institute of Health in the USA (Hosack et al., 2003). The EASE software looks for biological themes (as specified by the user, for example apoptosis), in the specific list of genes entered by the user. The genes significantly linked to a particular biological pathway are then listed (significance was assessed using Bonferoni multiple comparison test). The software also generates an output list with all the gene annotations (describing gene features) for the genes entered by the user. The database accessed by the EASE software for gene ontology is DAVID (database for annotation, visualisation and integrated discovery) (Dennis et al., 2003). To identify and visualise functional associations of candidate genes, the STRING software, a functional protein association networks database and web tool, was used with a confidence score set at 0.4 (confidence limit at which the associations were reported was 0.4, a confidence score of 0 would indicate 100% confidence) (http://string-db.org) (L. J. Jensen et al., 2009). The STRING software identifies interactions between two proteins that jointly contribute to the same process using databases of biological pathways. These include the molecular interaction database (MINT), the database of interacting proteins (DIP), Human protein reference database (HPRD), biomolecular interaction network database (BIND), BioGRID and KEGG.

The STRING software only assesses one protein per gene, the longest isoform for a particular gene. The functional relationships are determined by a set of prediction algorithms and text mining (L. J. Jensen et al., 2009).

FISH for confirmation of the array results on OSCC samples

Ten patient's blocks were selected per candidate gene on the basis of their array data analysis. The *CCND1*, *MYC* and *EPHA3* genes copy number were evaluated with FISH. Three micron slides were cut from the blocks and placed on positively charged slides. The slides were baked in the oven at 60°C overnight. The slides were de-waxed by immersing them in Xylene (Merck) for 15 minutes at room temperature and this was repeated twice more. Slides were then dehydrated twice in 100% ethanol for 5 minutes at room temperature and air dried for 2-5 minutes at 42°C. Thereafter, the slides were treated in 0.2N HCl (Appendix A) for 20 minutes, rinsed in distilled water for 3 minutes and then 2X SSC (Appendix A) for 3 minutes at room temperature. The slides were then immersed in 1M sodium thiocyanate at 80°C for 30 minutes, rinsed in distilled water at room temperature for 1 minute and then twice in 2xSSC for 5 minutes at room temperature. The slides were air dried before treating them in Pepsin (Roche) (0.5mg/ml) dissolved in 0.1N HCl (Appendix A) for 20 minutes to 1h30 minutes depending on the tissue. This buffer was pre-warmed to 37°C. The slides were then rinsed twice in 2xSSC for 5 minutes each at room temperature. Slides were dried at 42°C for 2-5 minutes and then fixed in 1% formaldehyde (Appendix A) at room temperature. Slides were rinsed once more, twice in 2XSSC for 5 minutes and briefly dipped in distilled water. The slides were placed on a heating block at 42°C and then denatured as described in chapter 2, section 2.1. The Vysis® LSI C-MYC SpectrumOrange probe (Abbott Molecular Inc) and the Vysis® LSI t(11;14) dual colour probe (Abbott Molecular Inc) were denatured at

76°C for 5 minutes and 10µl was applied to the denatured slides.After a minimum of 16hrs hybridisation, the slides were washed in 0.3% Tween20 in 2XSSC for 4 minutes at 74°C, counterstained in DAPI solution (Appendix A) for 15 minutes and rinsed in 2xSSC with 0.3% Tween20® for 2 minutes at room temperature. The slides were mounted in Vectashield (Vecta Laboratories) and analysed on the fluorescent microscope. Images were captured using the Cytovision 4.0 software (Applied Imaging).

The *EPHA3* gene deletion status was assessed using the "in-house" probe combination (prepared according to the methods in chapter 2, section 2.1 on BAC probe development for analysis of *EPHA3* gene deletions) of RP11-598N21 (spectrum orange) and RP11-547K2 (spectrum green).This probe combination covered exons 1-3 and 11-17 of the gene. The slides were treated, hybridised and washed as above.

4.3 Results

Demographics of the OSCC Patient cohort from the Eastern Cape

The 177 samples were all from black individuals from the Eastern Cape region, collected in the years 2004-2006. The mean age was 59.74 years with a male to female ratio of 1:0.97. All samples collected were from invasive, keratinising squamous cell carcinoma.

DNA assessment and copy number analysis of the OSCC FFPE cohort from the Eastern Cape.

DNA was extracted from all samples and 84 samples had sufficient DNA concentration, purity and integrity to continue with the array assay. These samples were further assessed for PCR performance using the multiplex PCR assay for amplification of seven fragments

(100bp to 700bp) of the *GAPDH* gene. There was inter-sample variability in amplifying the various fragment sizes thought to be due to varying length of fixation in formalin, differing methods of fixation and embedding, which could affect the level of DNA degradation. A minimum of six separate PCR reactions were required for each sample in order to obtain the 90µg necessary for hybridisation. Fifty five samples had sufficient DNA amplification and were hybridised to microarrays. Four of these samples had a call rate below 60% and were removed from further analysis (Appendix C). The male to female gender ratio for the 51 samples analysed was 1:1.16. The controls had a male to female gender ratio of 1:0.94.

Copy number gains

The genetic heterogeneity of OSCC was made evident by the number and complexity of copy number changes present in this cohort (Figure 4.2). To simplify data interpretation and to pull out important aberrations more likely to contribute to OSCC carcinogenesis as opposed to being random carrier aberrations, data were analysed as follows. As mentioned in materials and methods, copy number gains were characterised as high and moderate and were reported on the basis that a minimum of 10 samples (±20%) had the same region affected. Most studies do not stipulate a minimum number of samples affected by a particular aberration, in this study 20% was selected based on the average of a few studies (Y. C. Chang et al., 2010; Z. W. Chang et al., 2009; Carneiro et al., 2008; Xiaojun Zhao et al., 2005).

Fifteen regions with high amplification (>10 copies) and 21 regions with moderate amplification (>5<10 copies) were detected (Table 4.3, 4.4 and figures 4.3 and 4.4). The aberrations most likely to target key genes involved in OSCC carcinogenesis are those with high copy number changes that occur most frequently across samples. The most important

regions of high amplification, in descending order of frequency were 11q13.3 (66.7%) (70061246-70310057bp), 11q13.3-q13.4 (57%) (70310057-71374778bp), 2p21 (47%), 11q13.3 (37.2%) (68884395-70061246bp), 19p13.11-p13.12 (33.3%) (15972003-16438337bp), 19q13.43 (33.3%), 10q24.1 (33.3%), 1q32.2 (31.4%), 8p23.1 (31.4%), 1p32.3 (29.4%), 16p13.3 (25.5%), 15q22.31 (25.5%), 4q35.1 (21.6%), 10q22.1 (19.6%) and 11q13.3 (19.6%) (68405316-68471560bp) (Figure 4.3).

The moderate amplifications sorted in order of frequency were (Figure 4.4), 8q24.3 (70.6%) (140629509-144457635bp), 3p25.1-p25.2 (13232589-13450155bp) (68.6%), 3p25.2 (12775362-13086971bp) (68.6%), 2q37.3 (56.9%), 5p15.33 (43.1%), 7p22.2-p22.3 (41.2%), 1p34.2 (39.2%), 17q25.1-q25.2 (29.4%), 12q24.31 (29.4%), 3q21.3 (29.4%) (127551516-129106253bp), 12q24.31 (29.4%), 4p16.1 (27.5%), 20q13.3 (25.5%), 3q29 (25.5%), 19p13.2 (25.5%), 11q13.4 (25.5%) (71428559-72219067bp), 12p13.32 (21.6%), 2q14.2 (21.6%), 5q31.1 (21.6%), 10q22.3 (19.6%), 3q27.3 (19.6%) and 11q13.1 (19.6%). The amplicons with potential oncogenic genes in regions occurring at the highest frequency and with the highest amplitudes are described in more detail below.

Chromosome 11q gains

Four sub bands on chromosome 11q were amplified (an example of 11q amplification on copy number data can be seen in figure 3.5).

The region with the highest amplitude involved in the highest number of samples was a 248.81kb region at 11q13.3 (70061246-70310057bp). This region includes the known oncogene, *CTTN* gene, amplified in 66.7% of cases. Other amplicons affecting 11q included the 11q13.3 region from 68884395-70061246bp in 19 samples (37.2%), which included the

known oncogenes, Fibroblast growth factor 3 (*FGF3*) and fibroblast growth factor 4 (*FGF4*), fibroblast growth factor 19 (*FGF19*), and the cyclin D1 (*CCND1*) genes. Amplification of *CCND1* and *FGF4* genes was verified in 10 samples with a FISH probe covering the region that included both the *CCND1* and *FGF4* loci. On FISH, the average copy number for *CCND1* and *FGF4* was 15.65 copies per cell in 100 cells analysed per sample. The array results correlated closely (average of 13.3 copies) with the FISH results (Table 4.5 and Figure 4.6). Chromosome 11q13.3-q13.4 (70310057-71374778bp) included the SH3 and multiple ankyrin repeat domains 2 gene (*SHANK2*), which was amplified in 56.9% of cases. This gene encodes a protein involved in the regulation of trans-epithelial absorption in normal epithelial cells (W. Han et al., 2006).

The fourth 11q region, 11q13.1 (63678703-64497435bp), was moderately amplified in 19.6% of cases with an average copy number of 8.3. The region has 2 genes potentially targeted (*BAD,* and *SP1*). The potential target gene, the bcl-xl/bcl2-antagonist causing cell death (*BAD*) gene, is a regulator of the pro-apoptotic molecule Bcl-2. In a phosphorylated state, the Bad protein no longer promotes apoptosis. Phosphorylated Bad protein expression is elevated in colon cancers, which may deregulate apoptosis in colorectal tumours (M. R. Kim et al., 2007). An increase in phosphorylated BAD expression was also detected in prostate cancer (A. J. Smith et al., 2009). The splicing factor 1 gene (*SP1*) is the last of the potential oncogenes in this region.

Chromosome 8 gains

Three regions on chromosome 8 were affected by amplifications.

The 8q24.3 region (140629509-144457635bp) affected the most number of samples (70.6% of cases) and was moderately amplified with an average copy number of 5.7. A potential oncogene reported in cancer in this region is the prostate-specific membrane antigen gene (*PSCA*).

The 8p23.1 (6482787-6495845kb) focal region of amplification (13.06kb in size) was amplified in 31.4% of cases. This region harbours the microcephalin 1 gene (*MCPH1*) which is involved in chromatin remodelling in response to DNA damage (G. Peng et al., 2009). *MCPH1* has been shown to mediate BRCA2-Rad51 dependent homologous recombination in the repair of DNA damage and it is suggested to function as a tumour suppressor gene and not as an oncogene which makes the relevance of this particular gene in OSCC carcinogenesis questionable (Xianglin Wu et al., 2009).

The region harbouring the c-*MYC* oncogene was increased in copy number in the cohort below the cut off value of 5 copies for moderate amplification at an average of only 3.9 copies. This region on 8q24.13-q21.33 (123539799-137116755bp) was 13.58Mb in size and was detected in 21.57% of cases (11 cases). Ten of these cases were selected for FISH analysis using a locus specific probe for the c-*MYC* oncogene. It was found that across these samples an average copy number of 8.5 copies in 100 cells analysed was present (Table 4.6 and Figure 4.7). The samples had heterogeneity for c-*MYC* copy numbers within the tumour regions. For example, UROC16, had a copy number ranging from 5 copies to more than 20 copies in 100 cells analysed (Figure 4.5). This heterogeneity could explain the discrepancy with the array results, the number of cells affected by high copy number could have been diluted out by cells with low to normal copy number. There may have been pockets of cells with high amplification within the tumour population.

Chromosome 3 gains

Two regions on chromosome 3p and three regions on chromosome 3q were affected by amplification.

The 3p25.1-p25.2 (13232589-13450155bp) and 3p25.2 (12775362-13086971bp) regions were amplified in 68.6% of cases with an average amplitude of 8.4 and 8.5 copies respectively. No genes described in cancer map in these regions currently. Chromosome 3q29 (195427099-195612723bp) was amplified in 25.5% of the cases, the average copy number was 6.1 and the tyrosine kinase non-receptor 2 gene (*TNK2*) is located in this region.

Chromosome 3q21.3 (127551516-129106253bp) region had the highest average copy number of the moderately amplified regions with 9.3 copies in 29.4% of cases. The region was 1554.74kb in size, harbouring 22 genes, none of which have previously been described in cancer.

The chromosome 3q27.3 (187252322-187509741bp) sub-band had an average of 8.2 copies in 19.6% of cases. One potential key gene is in this region is the B-cell CLL/lymphoma 6 gene (*BCL6*).

Chromosome 2 gains

Chromosome 2p21 (45016287-45119536bp) had an average copy number of 14.2 copies in 47.1% of cases. Chromosome 2q37.3 (238430400-239560898bp) was amplified in 56.9% of cases with an average amplitude of 5.4 copies. Currently, there are no known oncogenes reported in either of these regions.

Chromosome 7p gain

Chromosome 7p22.2-p22.3 (141311-2851256bp) was a large region that showed moderate amplification (average of 5.2 copies) in 41.2% of cases. Two potential oncogenes locate in this region, the *ADAP1* and *CYP2W1* genes. The ArfGAP with dual PH domains 1 gene (*ADAP1*) encodes a phosphotidylinositol 3,4,5- triphosphate binding protein and its expression results in Erk (elk-related tyrosine kinase) activation (Hideko Hayashi et al., 2006). The *CYP2W1* gene encodes a cytochrome p450 enzyme expressed in colon and adrenal cancers and is a potential drug target (Karlgren et al., 2006).

Chromosome 5p gain

An average copy number of 5.6 for the chromosomal region 5p15.33 (1036116-1678508bp) was detected in 41.2% of cases. This region harbours the oncogene *TERT* (telomerase reverse transcriptase) and the cleft lip and palate transmembrane protein 1-like protein gene (*CLPTM1L*) which has been described as pro-apoptotic in lung cells (McKay et al., 2008).

Chromosome 19p gains

Chromosome 19 was affected by the highest level of amplification. Two regions on chromosome 19 were amplified.

Chromosome 19p13.11-p13.12 (15972003-16438337bp) affected 33.3% of samples and had the highest level of amplification (an average of 110 copies). The Kruppel-like factor 2 gene (*KLF2*) overlaps with this 273.93kb region along with 16 other genes and several micro RNAs. The third region on chromosome 19 amplified in 33.3% of cases was 270.81kb in size,

at position 58858174-59128983bp on 19q13.43, and covered one potential target described in cancer, the alpha-IB-glycoprotein precursor gene (*A1BG*).

Chromosome 1 gains

Chromosome 1 was affected by two amplicons, one on 1p and one on 1q.

Chromosome 1q32.2 (207893164-207891765bp) was amplified in 31.37% of the cases with a mean copy number of 13.61. This region was only 1.4Kb in size with one target gene, the complement component (3b/4b) receptor 1-like gene (*CR1L*). This gene's function is poorly understood and has no current involvement in cancer.

Amplification of 1p32.3 (54534339-54745164bp) was detected in 29.41% of cases with an average copy number of 18.2 copies. The tetratricopeptide repeat motif containing protein 4 gene (*TTC4*) interacts with heat shock proteins and the Cyclin dependant kinase 6. The gene is over expressed in cancer cell lines, whereas it is undetectable in normal non-proliferating tissues (Crevel et al., 2008).

Chromosome 20q gain

Chromosome 20q13.3 (45146901-46502912bp) was moderately amplified in 25.5% of cases with an average copy number of 5.7. The region contains the developmental gene eyes absent homologue 2 (*EYA2*) and over expression is reported in epithelial ovarian cancer (L. Zhang et al., 2005).

Chromosome 16p gain

In 25.5% of samples, amplification of a 71.5kb region on chromosome 16p13.3 (11574742-11646325bp) was detected, with an average copy number of 13.63 copies. This region has

one gene, the lipopolysaccharide-induced TNF factor gene (*LITAF*). This gene is induced by p53 and is involved in apoptosis (Moriwaki et al., 2001). Deletion of the *LITAF* gene was detected in B-cell lymphomas and it has tumour suppressor activity in diffuse large B-cell lymphomas and Burkitt's lymphomas (Mestre-Escorihuela et al., 2007). The role of the above three genes as oncogenes has not been reported.

Chromosome 15q gain

A focal region of 69.08kb on chromosome 15q22.31 (64782565-66994674bp) was also amplified in this cohort, with one gene reported in cancer, the Smad family protein 6 gene (*SMAD6*). The region was highly amplified in 23.53% of cases with an average of 24.7 copies. The Smad proteins are involved in the transforming growth factor-β pathway. Over expression of *SMAD6* was previously described in pancreatic cancers and acts in an oncogenic way by blocking the cell growth inhibiting activity of the *TGF-β* gene (Kleeff et al., 1999).

Chromosome 4q gain

Chromosome 4q35.1 (186169423-186303602bp) was amplified in 21.6% of cases with an average amplitude of 15.2 copies. There are no potential oncogenes previously described in this region.

Chromosome 10q gains

Three regions on chromosome 10q were amplified.

Chromosome 10q24.1 (98872389-98967971bp) was gained in 33.3% of cases with average amplitudes of 34.1 copies. No apparent potential oncogenes are described here.

Chromosome 10q22.1 (70839876-70921831bp) was gained in 19.6% of cases with average amplitude of 19.7 copies. Chromosome 10q22.3 (78583974-78724496bp) was moderately amplified in 19.61% of cases with average amplitudes of 19.7 copies. The potassium large conductance calcium-activated channel, subfamily M, alpha member 1 gene (*KCNMA1*) is located in this region. The *KCNMA1* gene is overexpressed in breast cancer and is involved in metastasis (Khaitan et al., 2009).

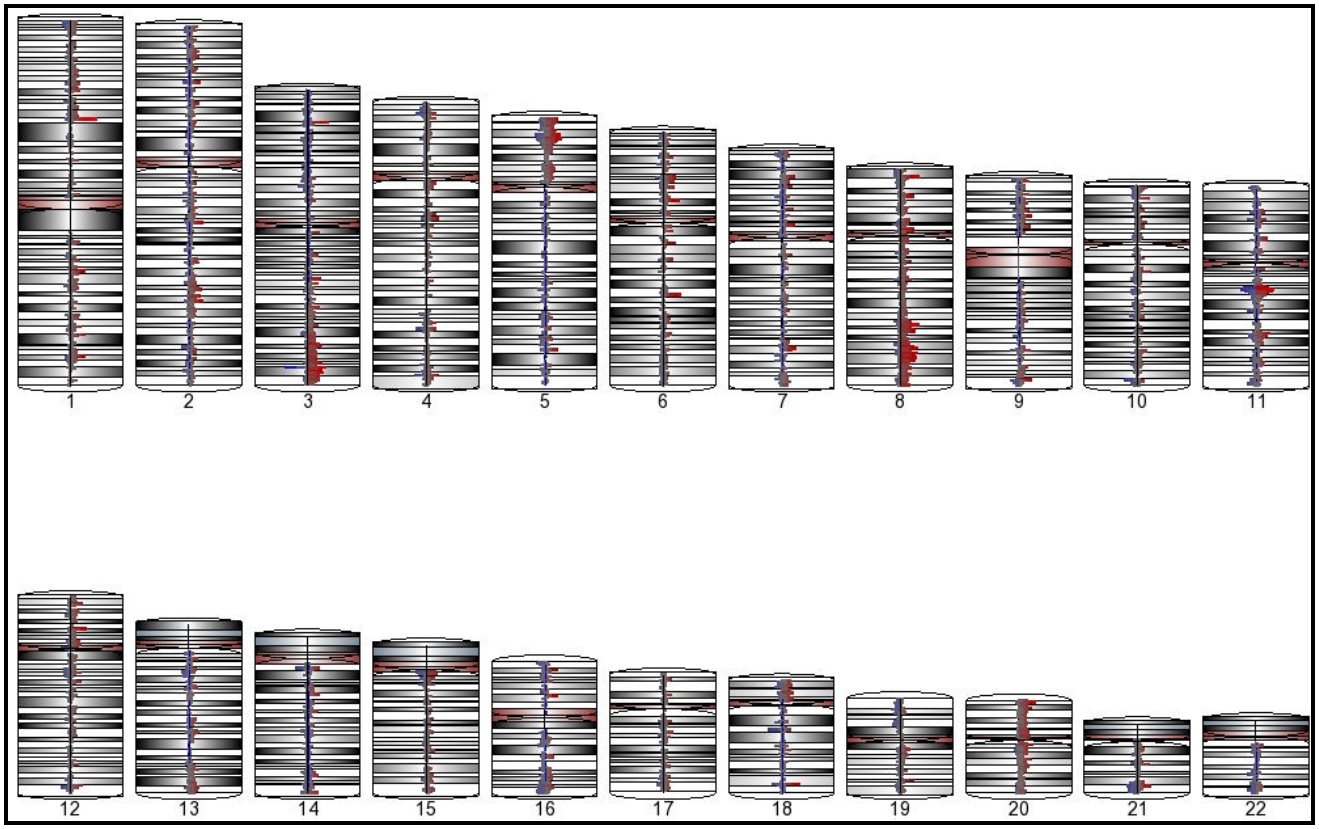

Figure 4.2 Chromosomal mapping of regions of copy number alteration across all 51 samples (generated in PGS). Increased copy numbers are indicated in red and blue refers to decreased copy number. The higher the peak for the red/blue bar the more samples affected in that region.

Table 4.3 Regions affected by high amplification in a minimum of 10 samples and with an average copy number larger than 10 copies. The regions are arranged in order of chromosome.

Cytoband	Location (bp)	Approximate size (Kb)	Frequency (n=51)	Mean copy no.	No of genes	Genes
1p32.3	54,534,339-54,745,164	210.82	15	18.2	7	*SSBP3*
1q32.2	207,891,765-207,893,164	1.4	16	13.61	1	*CR1L*
2p21	45,016,287-45,119,536	103.25	24	14.2	0	
4q35.1	186,169,423-186,303,602	134.18	11	15.2	2	
8p23.1	6,482,787-6,495,845	13.06	16	25.35	1	*MCPH1*
10q22.1	70,839,876-70,921,831	81.95	10	19.68	2	
10q24.1	98,872,389-98,967,971	95.58	17	34.12	3	
11q13.3	68,405,316-68,471,560	66.25	10	13.53	1	
11q13.3	68,884,395-70,061,246	1176.85	19	14.66	13	*CCND1,MYEOV, FGF19,FGF4,FGF3*
11q13.3	70,061,246-70,310,057	248.81	34	57.24	3	*CTTN*
11q13.3-q13.4	70,310,057-71,374,778	1064.72	29	14.62	13	*SHANK2*
15q22.31	64,782,565-66,994,674	69.1	13	26.9	2	*SMAD6*
16p13.13	11,389,965-11,646,325	256.36	13	14.18	3	*LITAF*
19p13.11-p13.12	15,972,003-16,438,337	466.33	17	110.8	16	*KLF2*
19q13.43	58,858,174-59,128,983	270.81	17	12	21	*A1BG*

Table 4.4 Regions of moderate amplification detected in a minimum of 10 samples (>5<10 average copies). The regions are ordered by chromosome.

Cytoband	Location (bp)	Approximate size (Kb)	Frequency (n=51)	Mean copy no.	No of genes	Genes
1p34.2	41680889-42044382	363.5	20	5.67	4	
2q14.2	121307843-121587908	280.06	11	8.7	1	*GLI2*
2q37.3	238430400-239560898	1130.5	29	5.4	52	*HES6, PER2*
3p25.1-p25.2	13232589-13450155	217.57	29	8.4	1	
3p25.2	12775362-13086971	311.61	35	8.5	4	
3q21.3	127551516-129106253	1554.74	15	8.65	22	
3q27.3	187252322-187509741	257.42	10	5.7	3	*BCL6*
3q29	195427099-195612723	185.63	13	6.1	6	*TNK2*
4p16.1	7528891-7648574	119.68	14	5.5	1	
5p15.33	1321181-1678508	642.4	22	5.6	14	*TERT, CLPTM1L,*
5q31.1	134,552,169-135,480,185	928.02	11	6	17	*H2AFY, TGFBI, SMAD5*
7p22.2-p22.3	141322-2851256	2709.94	21	5.2	36	*ADAP1, CYP2W1, GPER,*
8q24.3	140,629,509-144457635	3828.13	36	7.1	28	*PSCA*
10q22.3	78583974-78724496	140.52	10	7	1	*KCNMA1*
11q13.1	63678703-64497435	818.73	10	8.3	36	*BAD, ESRRA, MACROD1, MIR192, SF1*
11q13.4	71428559-72219067	790.51	13	8.8	5	
12p13.32	3115831-3332452	216.62	11	7.5	2	
12q23.31	123310702-123885293	574.6	15	6.55	15	
17q25.1-q25.2	73571719-75133860	1562.14	15	5.3	54	
19p13.2	7745413-8317060	571.65	13	7.1	17	*ELAVL1, MAP2K7*
20q13.13	45146901-46502912	1356.01	13	5.7	9	*EYA2*

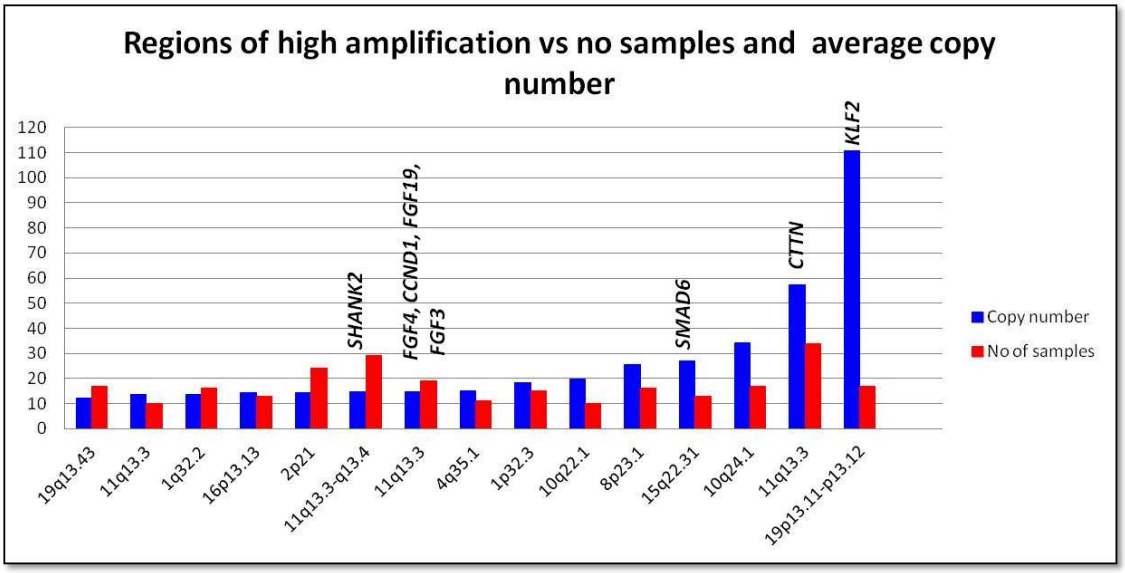

Figure 4.3 Graph representing the regions of high amplification (> 10 copies) with the number of samples affected and the average copy number per sample. The regions were sorted from the smallest to largest copy number.

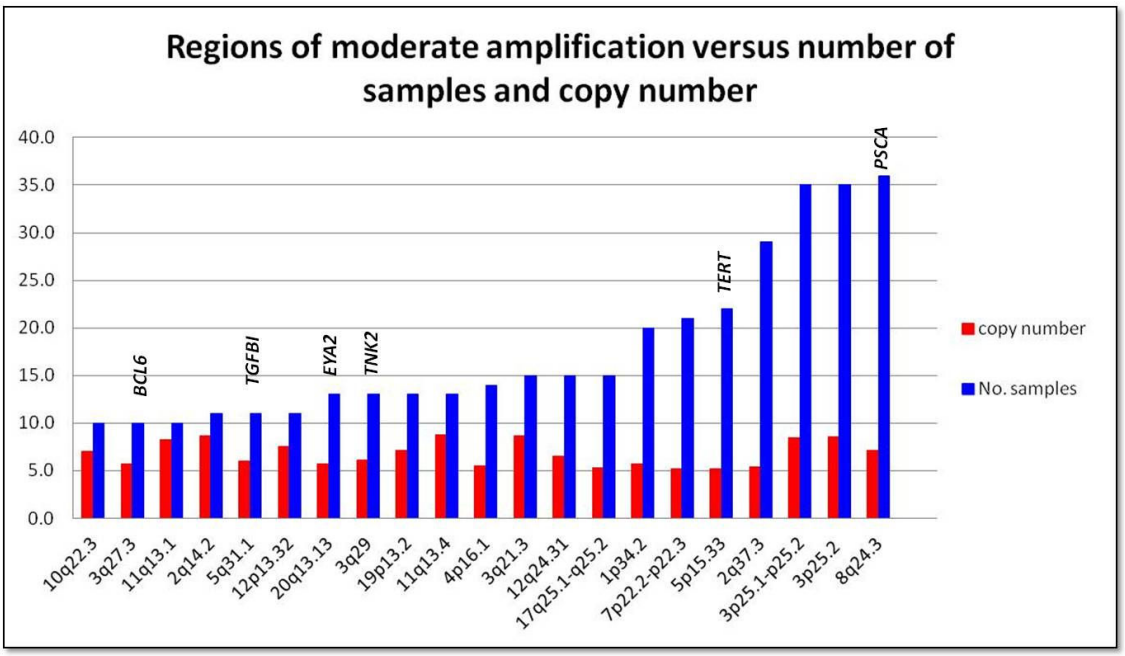

Figure 4.4 Graph representing the regions of moderate amplification (>5<10 copies) with the number of samples affected and average copy numbers per sample for that region. The regions were sorted from smallest to largest number of samples affected.

Table 4.5 FISH results for detection of *CCND1/FGF4* gene amplification in 10 FFPE samples.

Sample	Average copy no. (100 cells analysed)	Overall result for FISH analysis*
UROC26	17.5	High Amplification
UROC34	15	High Amplification
UROC37	20	High Amplification
UROC46	6.5	Moderately amplified
UROC62	10	Moderately Amplified
UROC71	17.5	High Amplification
UROC87	12.5	High Amplification
UROC124	20	High Amplification
UROC156	20	High amplification
UROC171	17.5	High amplification
Average copy number for 10 samples= 15.65		

*moderate amplification was defined as 5 to 10 copies and high amplification was defined as more than 10 copies in keeping with the array interpretation.

Table 4.6 FISH results for detection of c-*MYC* gene amplification in 10 FFPE samples.

Sample	Average copy no. (100 cells analysed)	Overall result for FISH analyis*
UROC16	12.5	High Amplification
UROC26	15	High Amplification
UROC30	6.5	Moderate Amplification
UROC33	4.5	Polyploidy
UROC47	8	Moderate Amplification
UROC55	8.5	Moderate Amplification
UROC84	7.5	Moderate Amplification
UROC98	6.5	Moderate Amplification
UROC125	10.5	High amplification
UROC130	6	Moderate amplification
Average copy number for 10 samples= 8.55		

*moderate amplification was defined as 5 to 10 copies and high amplification was defined as more than 10 copies

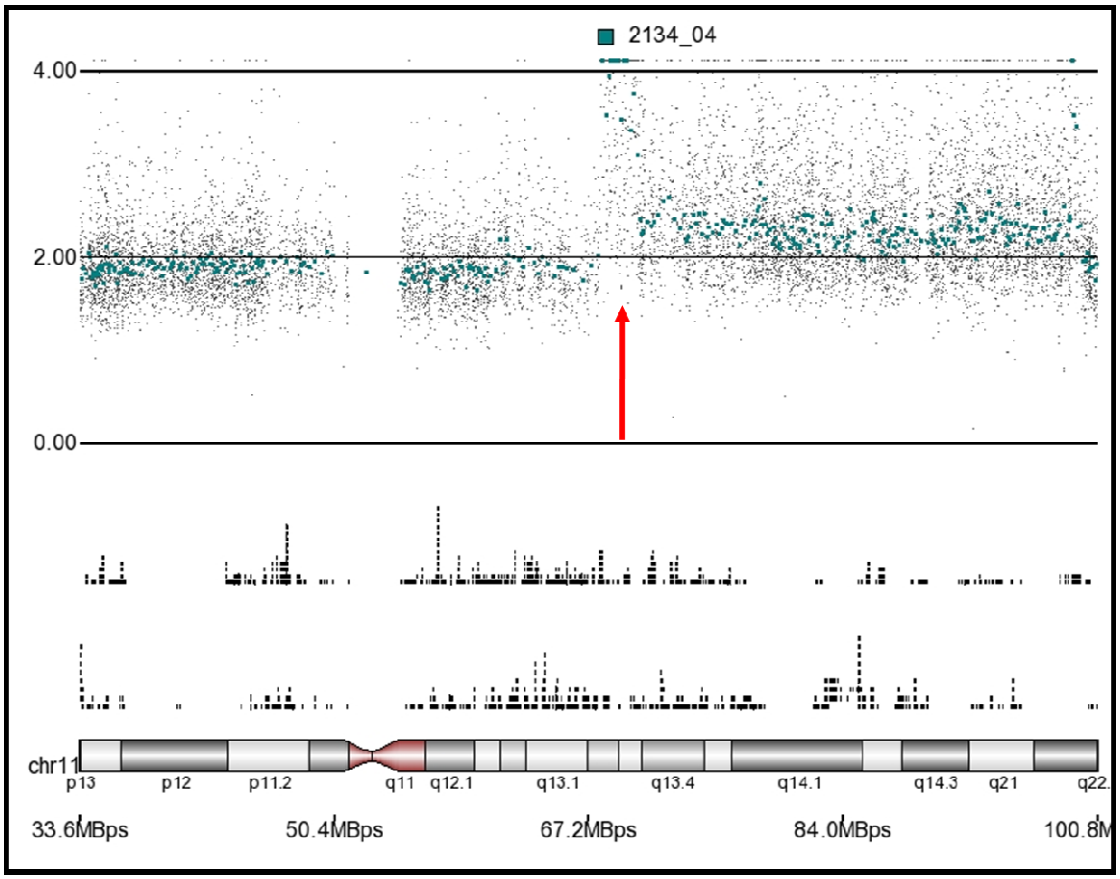

Figure 4.5 Chromosome 11 mapping of copy number changes detected in one representative FFPE sample by 500K SNP arrays (generated in PGS). The red arrow indicates the significant region of high amplification detected at 11q13.3 (67226192-71208231bp). The blue dots represent the smoothed data points while the grey dots reflect the unsmoothed data points.

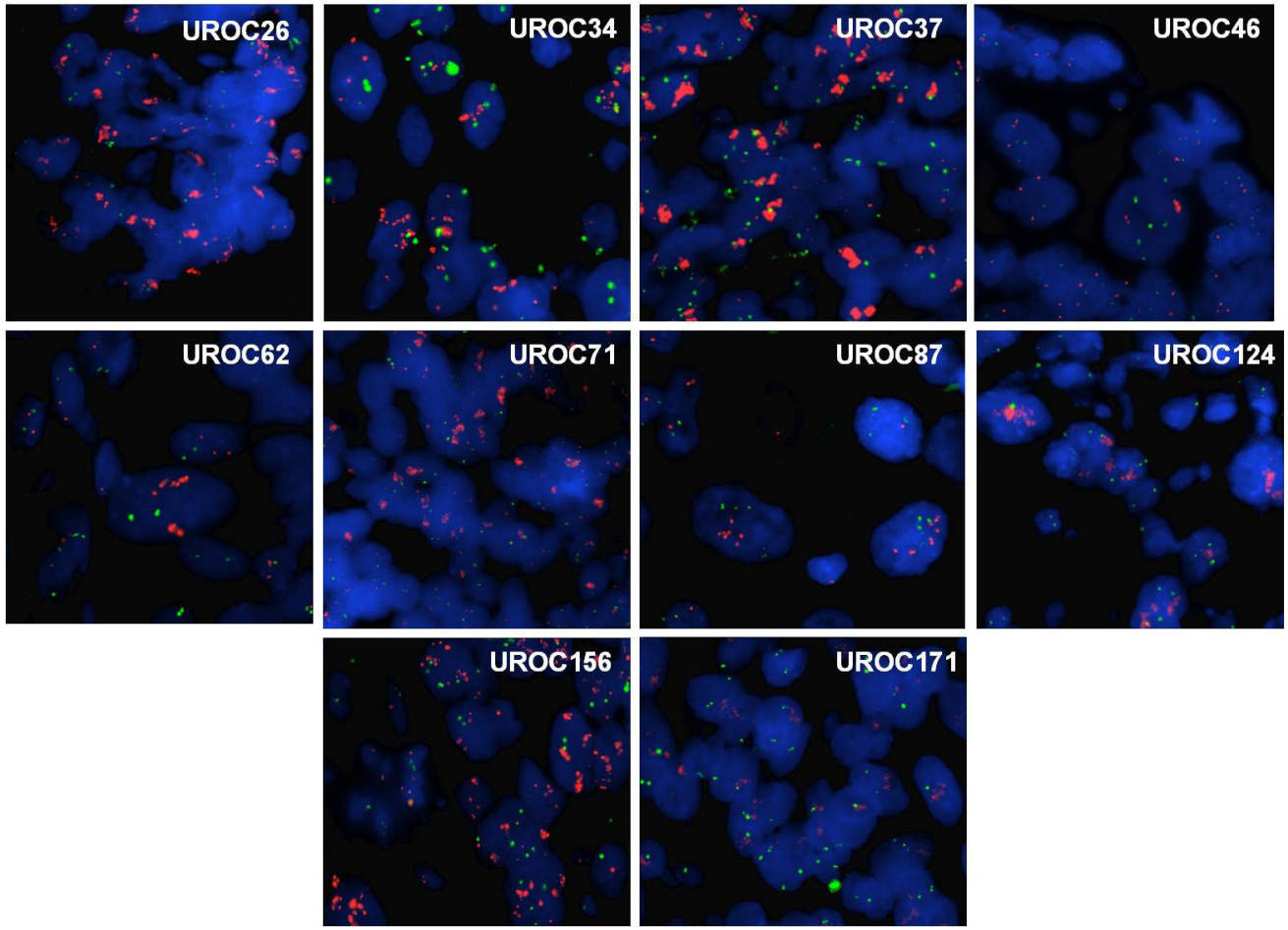

Figure 4.6 Selected images of DAPI stained nuclei for each of the 10 tumours chosen for FISH analysis with the Vysis LSI t(11;14) dual colour probe showing *CCND1* and *FGF4* (red) gene amplification. The green signal is the locus specific probe for the *IGH* gene on chromosome 14 acting as an internal control. Each tumour shows varying copy number with an average copy number of 15.65 across the 10 samples.

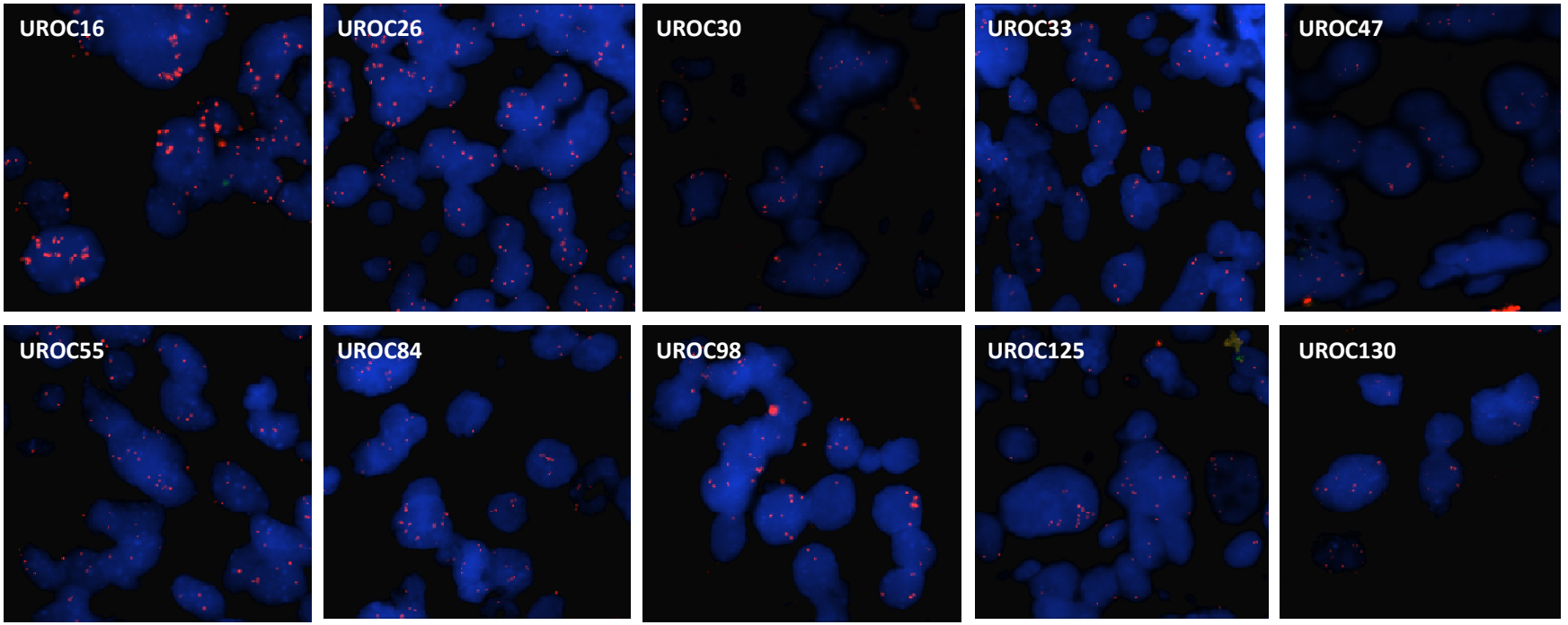

Figure 4.7 Selected images of DAPI stained nuclei for each of the 10 tumours chosen for FISH analysis using the Vysis LSI MYC probe. Moderate to low copy numbers of the c-*MYC* gene (Red) were seen. The average copy number across the 10 samples was 8.5.

Copy number losses

Two homozygous deletions (copy number <0.7) were detected. The first was a 1.41Mb region on chromosome 15q11.2 (18427103-19837012bp) in 12 samples (23.53%), which involved 8 genes. One of these genes, the B-cell CLL/lymphoma 8 gene (*BCL8*), has previously been reported in cancer but not in OSCC. The second homozygous deletion on chromosome 19p13.2 (11628007-11922344) was detected in 10 samples (19.6%), and included 9 genes, five of which encoding zinc finger proteins and the other four genes being the *ACP5, ELOF1, ECSIT* and *CNN1* genes. Apart from *CNN1*, these genes do not have well understood functions and no reported involvement in cancer. The Calponin 1 (*CNN1*) gene encodes an actin binding protein, which is down-regulated in pre-cancerous lesions of the rat esophagus under conditions of zinc deficiency (C.G. Liu et al. 2005). No hemizygous deletions were observed for either of these regions.

Thirty one regions of hemizygous deletion (copy number >0.7<1.3) were detected (Table 4.7 and Figure 4.8 and 4.9).

Deletions frequently involved large portions of chromosomes. It is thought that more focal regions affecting the smallest minimal region of overlap across many samples would be regions more likely to target genes involved in carcinogenesis (Beroukhim et al., 2007; Weir et al., 2007). For this reason the deletions in focal regions (less than a whole chromosome arm) that harbour potential tumour suppressor genes and larger regions that encompass well described tumour suppressor genes are reported.

Chromosome 3p deletion

Chromosome 3p deletion is commonly reported in cancer and four major regions of deletion on chromosome 3p were detected.

The sub band 3p11.1-p12.1 (85803132-90338670bp) was deleted in 74.5% of cases. This region affects the same gene reported in the OSCC cell lines (see chapter 2), *EPHA3,* as well as 6 other genes. The *CADM2* gene is the only other gene reported to have a tumour suppressor activity in prostate cancer (G. Chang et al. 2010). *EPHA3* gene deletions were confirmed in a selected 8 cases by FISH analysis using BAC clones, RP11-598N21 and RP11-547K2 (Figure 4.10 and Table 4.8). The hemizygous deletions were detected in 50-100% of 70-100 cells analysed per case and the average copy number across the 8 samples was 1.25. There were hemizygous deletions in the *EPHA3* gene in the 8 samples analysed but there were areas of the tumour that still had both copies of the gene. This supports the selection for a copy number cut off of <1.3 used in this study, where the cells with normal copy numbers may dilute out those cells where there is a deletion.

Another region on 3p25.2-p26.3 (966545-12654748bp) was deleted in 68.6% of cases, and affects the E3 ubiquitin-protein ligase RAD18 (*RAD18*) and TIMP metallopeptidase inhibitor 4 (*TIMP4*) genes. The *TIMP4* gene was shown to inhibit metalloproteinase activity and inhibit growth of breast cancer cells (M. Wang et al., 1997).

The larger region of deletion affecting 3p12.2-p21.2 (50688518-81859001bp) was detected in 21.6% of the cases. This region included the known tumour suppressor gene, fragile histidine triad (*FHIT*).

Chromosome 6q deletion

The smallest region of deletion detected was on chromosome 6q26 (163464809-163471799bp), which was 6.99kb in size. This region was deleted in 66.67% of the samples and affects the Park2 co-regulated gene (*PACRG*). This region was reported to be deleted in

renal cell carcinomas and correlated with a lowered expression of *PACRG* (Toma et al., 2008). The common fragile site FRA6E is located in 6q26 and encompasses the *PACRG* locus.

Chromosome 4q deletion

Chromosome 4 had a deletion which included the whole q arm, 4p14-q35.2. The region was deleted in the largest proportion of samples (84.3%). This large region contains 5 potential gene targets, the *CASP3*, *FBXW7, SFRP2*, *HMGB2* and *DCK* genes. The Caspase 3 gene (*CASP3*) is involved in apoptosis and lower expression is detected in colorectal carcinoma (J. ting Guan et al., 2009). The F-box and WD repeat domain containing 7 gene (*FBXW7*) is a tumour suppressor, which targets proliferative factors for degradation through ubiquitination (N. Liu et al., 2010). The secreted frizzled-related protein gene (*SFRP2)* was also deleted in the OSCC cell lines, (chapter 2), and is involved in Wnt signalling control. The high-mobility group box 2 gene (*HMGB2*) product is shown to stabilise p53 by preventing HPV E6 protein mediated ubiquitination and degradation of p53 (D. Lee et al., 2010). The *HMGB2* gene was only activated in HPV infected cells (D. Lee et al., 2010). The deoxycytidine kinasegene (*DCK*) is in this region, this enzyme is involved in rate-limiting catabolism of chemotherapies like gemcitabine, which is used in treatment of pancreatic adenocarcinoma. *DCK* gene expression has been associated with prolonged survival post resection in patients with pancreatic adenocarcinoma and patients with adenocarcinoma of the oesophagus (Maréchal et al., 2010; C. J. Peters et al., 2010).

Chromosome 18q deletion

Chromosome 18q21.1-q21.2 (46295514-53583643bp) was deleted in 70.6% of cases affecting the well known tumour suppressors deleted in colorectal cancer (*DCC*) and Smad protein family 4 (*SMAD4*).

Chromosome 5q deletion

Three regions of deletion on chromosome 5q were detected.

Chromosome 5q14.2 (82203607-82649577bp) was deleted in 54.9% of the samples. The X-ray repair complementing defective repair in Chinese hamster cells 4 gene (*XRCC4*) is located in this region. This is an important gene involved in DNA repair by non-homologous end joining.

The second region on chromosome 5, 5q14.3-q21.2 (83674394-103025813bp) was deleted in 65% of cases. This region encompasses the RAS p21 protein activator [GTPase activating protein] 1 (*RASA1*) gene, which is a suppressor of the oncogenic *RAS*. Loss of heterozygosity of this gene has been reported in *HER2* positive breast cancers (Xiaolan Hu et al., 2009).

The third region on chromosome 5 was 5q21.2-q23.3, deleted in 47% of samples. The 26Mb base region is large but includes the important tumour suppressor genes, *APC* and *MCC*.

Chromosome 1q deletion

A deletion of chromosome 1q23.1 (156789994-156870545bp) of 80.55kb in size was detected in 52.9% of samples. This deletion may target the spectrin, alpha, erythrocytic 1 gene (*SPTA1*) encoding a molecular scaffolding protein. Chromosome 1q23.1 deletions have been reported previously in breast cancer (L. C. Chen et al., 1989) and a translocation

mapping to the *SPTA1* locus, t(1;x), was detected in renal carcinoma (Weterman et al., 1996).

Chromosome 11q deletion

The ataxia telangiectasia mutated gene (*ATM*) (11q22.3) was examined in this data set as it is an important tumour suppressor reported to be deleted in OSCC. Forty nine percent of the samples had a copy number of 1.34 for the 11q22.3 (107663328-107745036bp) region, which is 2.96Mb in size and harbours the *ATM* gene. This region was just above the cut off value of 1.3 for hemizygous deletion. Confirmation of this deletion by an alternative method would be needed.

Chromosome 10p deletion

A small (161.6kb) region on 10p11.11 (38013117-38174719bp) was deleted in 31.4% of cases. The zinc finger protein 248 gene (*ZNF248*) is located in this region. This gene has not been described previously in cancer.

Chromosome 8p deletion

Two discrete regions on chromosome 8p were deleted.

Chromosome 8p22 deletions have been reported in the literature in a variety of cancer types and specifically in oesophageal cancers (Du Plessis et al., 1999; Di Benedetto et al., 2006; Bova et al., 1996). Chromosome 8p22 (16014369-16057542bp) was deleted in 21.6% of samples. The region is 43.17kb in size and involves the macrophage scavenger receptor 1 gene (*MSR1*). Perhaps more relevant is another region on chromosome 8, 8p12-p21.3 (22587295-32018598bp), deleted in 21.57% of the samples. This region harbours the

Werner syndrome helicase/exonuclease protein (*WRN*) gene. Loss of this gene is associated with Werner syndrome, a premature ageing and cancer predisposing syndrome (Franchitto et al., 2008). Chromosome 8p12-p21.3 also harbours the platelet-derived growth factor receptor-like gene (*PDGFRL*), which is down-regulated in colorectal cancers and was shown to inhibit proliferation of colon cancer cells (F. J. Guo et al., 2010). This 8p12-p21.3 region harbours two other potential tumour suppressors, the leucine zipper, putative tumour suppressor 1 (*LZTS1*) and the N-acetyltransferase 2 [arylamine N-acetyltransferase] (*NAT2*) genes. The *NAT2* gene is involved in the metabolism of carcinogens (X. Zhong et al., 2010).

Chromosome 7p deletion

A deletion of chromosome 7p15.1 (28142894-28372542bp), of 229.65kb in size, was detected in 19.6% of the cases. This region harbours one gene, the cAMP responsive element binding protein 5 (*CREB5*). Currently it has no reported involvement in cancer.

Chromosome 9p deletion

Chromosome 9p21.1-p21.3 (21792634-27938527bp), a 6.15Mb region, was deleted in 15.7% (8 samples) of the samples, which was below the cut off of 10 samples. This region includes the well known tumour suppressors cyclin dependent kinase 2A and B (*CDKN2A* and *CDKN2B*) genes commonly deleted in OSCC.

Figure 4.8 Graph showing the focal regions of deletion and the copy number for that region. Hemizygous deletions were defined as those regions with a copy number < 1.3 (red line). Homozygous deletions were defined as regions with a copy number <0.7 (black line).

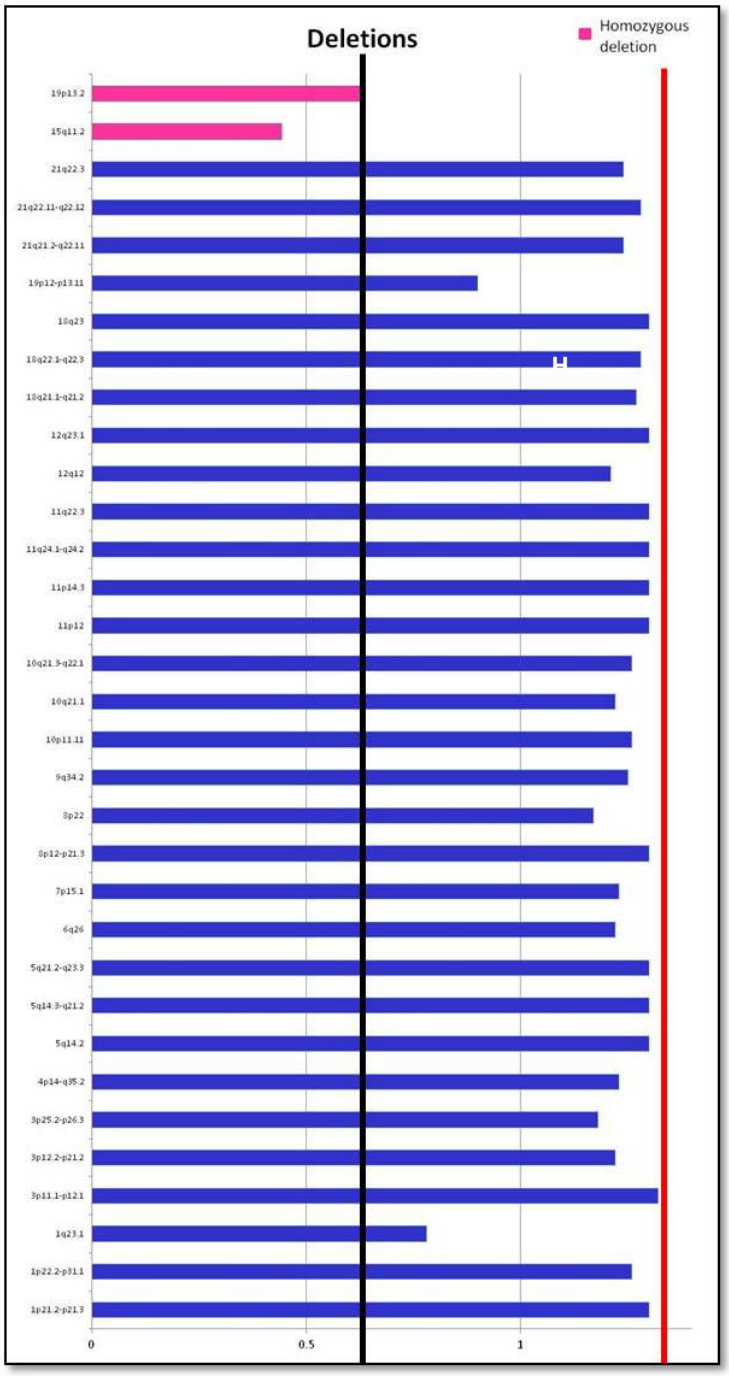

Figure 4.9 Graph showing the regions of hemizygous deletion and the number of samples with that deletion. The minimum number of samples affected was 10. The data was sorted from smallest number of samples to largest number of samples affected.

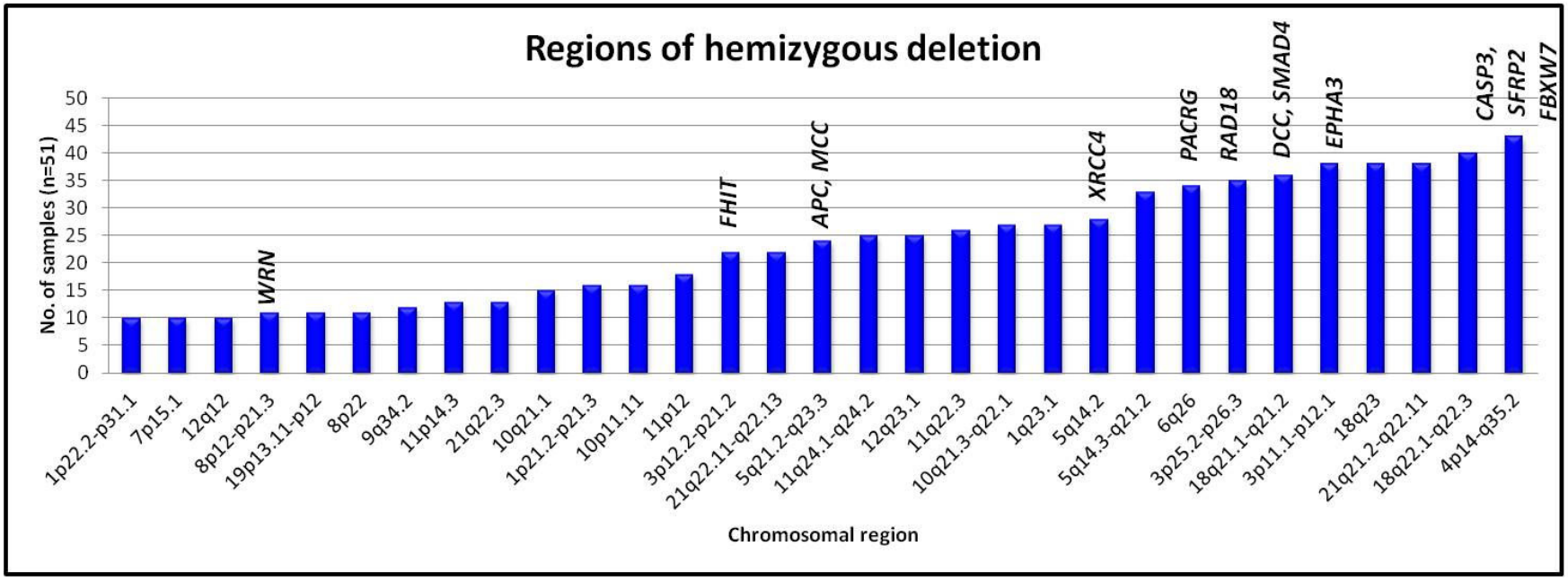

Table 4.7 Regions of hemizygous (<1.3 copies) deletion detected in a minimum of 10 samples. Regions are arranged by chromosome.

Cytoband	Location (bp)	Approx size (Mb)	Frequency (n=51)	Mean copy no.	No of genes	Genes	Cytoband	Location (bp)	Approx size (Mb)	Frequency (n=51)	Mean copy no.	No of genes	Genes
1p21.2-p21.3	95128120-102094945	6.97	16	1.3	35	*DPYD*	**10q21.1**	53471227-53487091	0.02	15	1.22	1	
1p22.2-p31.1	70258477-89926222	19.7	10	1.26	84	*IFI44,MSH4*	**10q21.3-q22.1**	69697598-70167452	0.47	27	1.26	7	
1q23.1	156789994-156870545	0.81	27	0.78	3	*NTRK1*	**11p12**	37029613-42832684	5.8	18	1.3	3	
3p11.1-p12.1	85803132-90338670	4.53	38	1.32	7	*EPHA3, CADM2*	**11p14.3**	22795608-23901634	1.1	13	1.3	3	*SVIP*
3p12.2-p21.2	50688518-81859001	31.2	22	1.22	166	*FHIT*	**11q24.1-q24.2**	123583522-123975617	0.39	25	1.3	12	
3p25.2-p26.3	966545-12654748	11.7	35	1.18	61	*RAD18, HRH1,TIMP4*	**11q22.3**	106893561-106962829	0.07	26	1.3	0	
4p14-q35.2	37962765-191152645	153.2	43	1.23	483	*RFC1, KIT, DCK, SMAD1, CASP3, FBXW7, SFRP2*	**12q12**	40815395-41047532	0.23	10	1.21	3	
5q14.2	82203607-82649577	0.44	28	1.3	3	*XRCC4*	**12q23.1**	98,292,180-98450546	0.16	25	1.3	0	
5q14.3-q21.2	83674394-103025813	19.35	33	1.3	54	*RASA1, PCSK1*	**18q21.1-q21.2**	46295514-53583643	7.3	36	1.27	3	*DCC, SMAD4*
5q21.2-q23.3	104063492-13006733	26	24	1.3	49	*APC, MCC*	**18q22.1-q22.3**	64854005-68986680	4.13	40	1.28	10	

	7											
6q26	163464809-163471799	0.007	34	1.22	1	*PACRG*	**18q23**	75902093-76115139	0.21	38	1.3	0
7p15.1	28142894-28372542	0.23	10	1.23	3	*CREB5*	**19p12-p13.11**	19,898,571-22510844	2.6	11	0.9	21
8p12-p21.3	22587295-32018598	9.43	11	1.3	51	*WRN*	**21q21.2-q22.11**	25435155-33624585	8.2	38	1.24	81
8p22	16014369-16057542	0.043	11	1.17	1	*MSR1*	**21q22.11-q22.12**	35177445-38271413	3.1	22	1.28	20
9q34.2	136572266-136884847	0.312	12	1.25	2		**21q22.3**	42986775-44046571	1.06	13	1.24	21
10p11.11	38013117-38174719	0.16	16	1.26	2	*ZNF248*						

Table 4.8 FISH results for analysis of *EPHA3* gene deletions using the BAC clones RP11-598N21 (red) and RP11-547K2 (green).

Sample	Signal pattern	# Cells	Average copy no.	Overall result
UROC156	2 yellow	50	1.5	50% hemizygous deletion
	1 yellow	50		
UROC128	2 yellow	10	1.1	90% hemizygous deletion
	1 yellow	90		
UROC125	2 yellow	12	1.05	82% hemizygous deletion
	1 yellow	85		
	0 yellow	7		
UROC16	1 yellow	30	1.57	75% hemizygous deletion (n=70)
	2 yellow	40		
UROC171	1 Yellow	30	1.0	42.8% hemizygous deletion and 28.6% loss of exons 11-17 (n=70)
	2 yellow	20		
	2 red	20		
UROC130	1 yellow	40	1.6	40% hemizygous deletion (n=100)
	2 yellow	60		
UROC124	1 red	60	1.0	100% Complete loss of exons 11-17 and hemizygous loss of exons 1-3. (n=100)
	2 red	20		
	0 signal	20		
UROC158	1 yellow	59	1.14	60.2% Hemizygous deletion (n=98)
	2 yellow	21		
	0 yellow	5		
	2 green	4		
	1 yellow, 1 red	8		
	1 green	1		

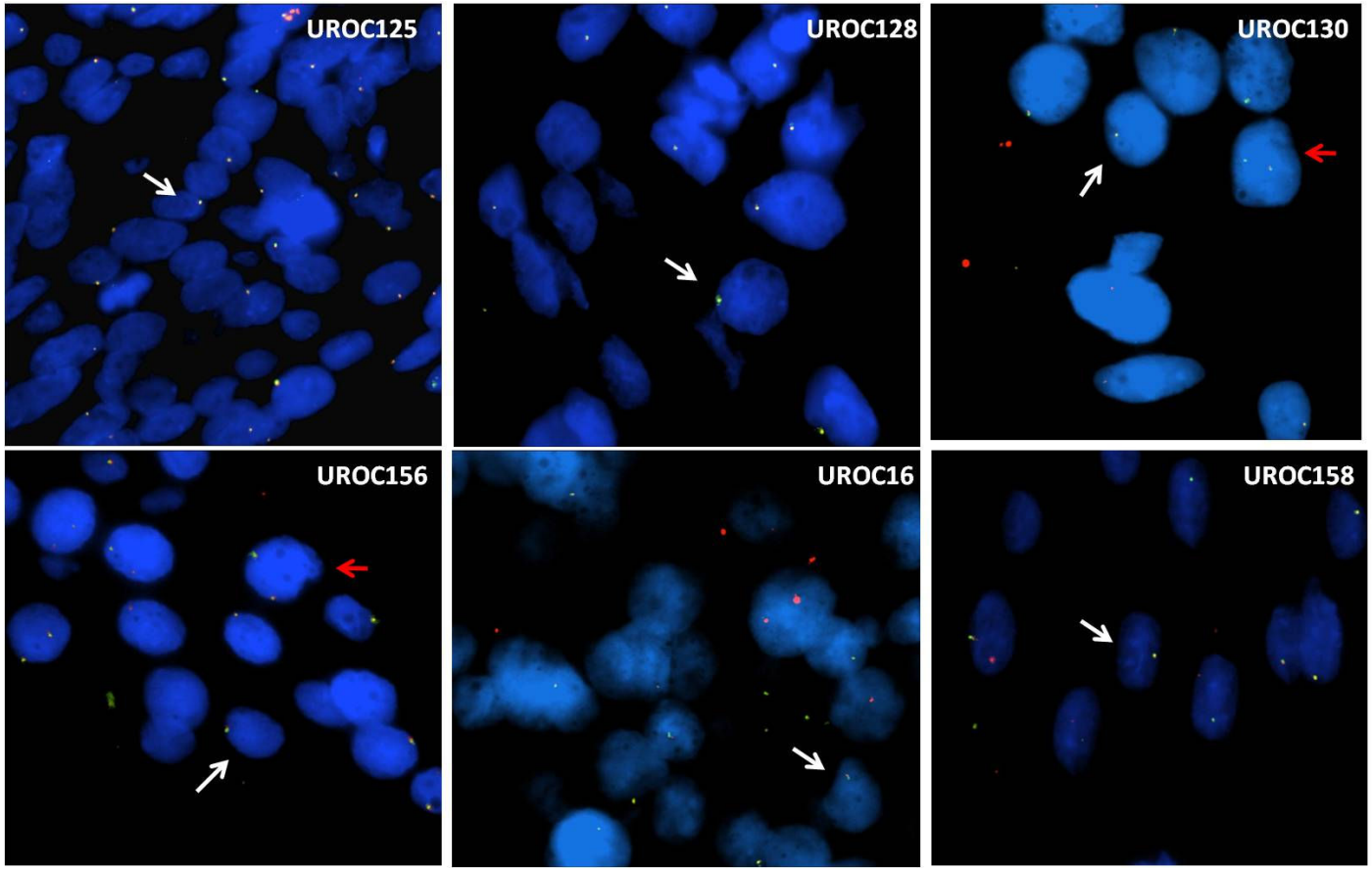

Figure 4.10 Representative images from 6 of the 10 cases selected for FISH analysis of *EPHA3* gene deletions. DAPI stained interphase nuclei for UROC125, UROC 128, UROC130, UROC156, UROC16 AND UROC158, showing hemizygous deletion within *EPHA3*. The white arrows show cells with one yellow signal representing the presence of exon 1-3 and exon 11-17 sequences for one allele. The red arrows indicate nuclei where both alleles are intact.

Functional associations of candidate genes in amplified and deleted regions

The list of genes, from segmented copy number data that was generated by Partek Genomics Suite, was screened for potential tumour suppressors and oncogenes by reviewing the published literature. This refined list of genes was then analysed in the data mining software programs, EASE and STRING, to identify biological pathways in which these genes are involved as well as identify functional associations between them.

Using the EASE software to elucidate the common biological pathways amongst the selected candidate genes from amplified regions, twenty of the 36 candidate genes could be linked to common pathways. The EASE software establishes links to pathways selected by the user through text mining and by accessing the DAVID database for gene annotations. The most common pathways highlighted by EASE were DNA metabolism, RNA transcription, cell proliferation, cell growth and cell maintenance (Table 4.9 and figure 4.11). Other pathways involved included cell signalling, chromatin modelling, RNA processing, cell motility and cell development. In terms of the regions of deletion, 21 of 27 candidate tumour suppressor genes identified could be linked to pathways involved in cell growth and maintenance, nucleic acid metabolism and cell communication, RNA transcription, DNA damage responses and apoptosis (Table 4.10 and Figure 4.12).

The STRING 8.0 software was then used to identify functional relationships between candidate genes and graphically represent these associations. The associations between genes are reported by strength of experimental evidence or by literature based evidence. First, analysis was done between the 36 amplified genes and the 27 deleted genes, results can be seen in figure 4.13. The *CCND1, NCOR2, FGF19, FGF3, FGF4, CTTN, TERT* and *SHANK2* genes were the most strongly linked among amplified genes with *CLPTM1L, PSCA, H2AFY,*

BAD and *TGFBI* linking with a weaker confidence (Figure 4.14). Of interest, the majority of strongly associated genes, *FGF19*, *FGF3*, *FGF4*, *CTTN*, *CCND1* and *SHANK2* were all closely located on chromosome 11q13.3. Functional relationships between *CTTN* and *SHANK2* are reported in the literature, where CTTN is reported to bind SHANK2 and together are involved in cytoskeletal reorganisation in neurons (Y. Du et al., 1998). The CTTN protein may also indirectly interact with Cyclin D1 as CTTN is found to regulate transcription of the cyclin-dependant kinase inhibitors p21, p27 and p57 (which bind CyclinD1), allowing for cell cycle progression in a head and neck carcinoma cell line (Croucher et al., 2010). The FGF growth pathway can interact with and activate the Wnt pathway, which results in transcription of the *CCND1* gene (Masaru Katoh, 2007), which would explain the functional relationship observed between the *FGF* genes and *CCND1*.

The deleted genes were weakly associated. Stronger associations were shown for *KIT* and *RFC1* as well as *DCC* and *CASP3* genes. The *FHIT*, *PDGFRL*, *NDEA*, *PACRG*, *LZTS1*, *SMAD4*, *RFC1*, *RAD18, WRN, XRCC4, RASA1* and *EPHA3* genes linked in weakly into this network of proteins (Figure 4.15). The analysis of functional relationships between both amplified and deleted genes showed strong associations between the deleted genes *FHIT*, *SMAD4, SMAD6* and amplified *TGFBI* and *CCND1* (Figure 4.14). These genes are all involved in cell cycle control and particularly *SMAD4*, *SMAD6* and *TGFBI* genes are involved in the TGF-β pathway (Attisano et al., 2001). The FHIT protein is involved in Wnt inhibition, which regulates expression of the *CCND1* explaining the interaction observed between these genes (Weiske et al., 2007).

Table 4.9 Biological pathways associated with genes amplified in this study.

DNA metabolism	Transcription	Proliferation, Growth and Maintenance	Cell death	Cell signalling	Chromatin structure/nuclear organisation	RNA processing	Cell motility and attachment	Development
MCM2	*TGFBI*	*EYA2*	*BAD*	*FGF3*	*MCM2*	*SF1*	*KCNMA1*	*ELAVL1*
KCNMA1	*LITAF*	*FGF4*		*FGF4*	*TERT*	*ELAVL1*	*TGFBI*	*EYA2*
TERT	*BAD*	*CCND1*		*KCNMFGFA1*				*FGF19*
BAD	*ESRRA*	*PSCA*		*MAP2K7*				*FGF3*
ELAVL1	*HES6*	*FGF3*		*PER2*				*KCNMA1*
ESRRA	*KLF2*	*MAP2K7*		*PSCA*				
EYA2	*LITAF*	*MCM2*						
HES6	*MCM2*	*TERT*						
NCOR2	*NCOR2*	*TGFBI*						
PER2	*PER2*							
SF1	*SF1*							
TERT								
12	**11**	**9**	**1**	**6**	**2**	**2**	**2**	**5**

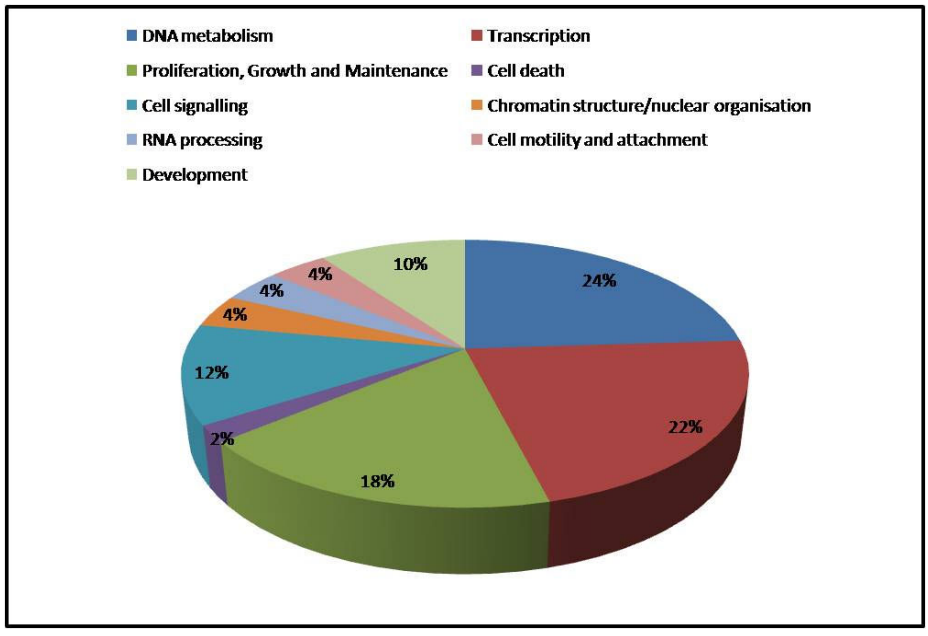

Figure 4.11 Pie chart showing the proportion of the 20 candidate genes involved in each biological process listed. This data was established using the EASE 2.0 software.

Table 4.10 Biological pathways associated with deleted genes in this study.

Cell growth/maintenance	Nucleic acid metabolism	Cell communication	transcription	DNA damage response/DNA repair	Apoptosis
ADAMTS1	*FHIT*	*ADAMTS1*	*HMBG2*	*RAD18*	*FHIT*
DCC	*HMGB2*	*ASB5*	*LZTS1*	*RFC1*	*CASP3*
FHIT	*LZTS1*	*EPHA3*	*RFC1*	*XRCC4*	*DCC*
KIT	*RAD18*	*HUNK*	*ZNF248*		
MSR1	*RFC1*	*KIT*			
RASA1	*WRN*	*PCSK1*			
RFC1	*XRCC4*	*PDGFRL*			
SPTA1	*ZNF248*	*RASA1*			
8	**8**	**8**	**4**	**3**	**3**

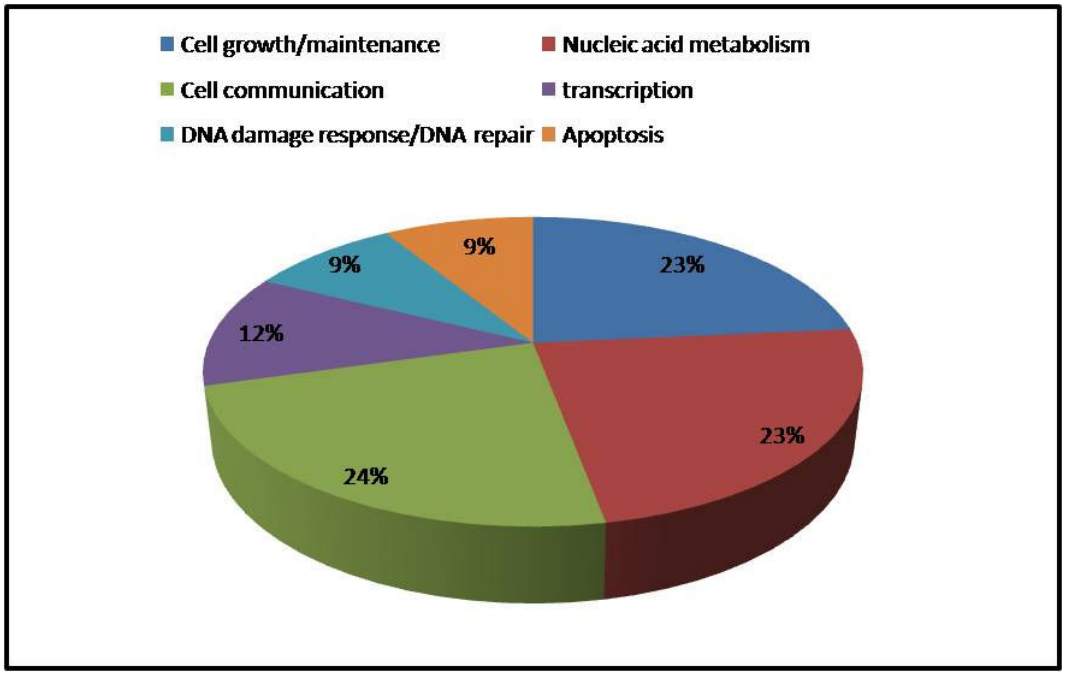

Figure 4.12 Pie chart showing the proportion of deleted genes associated with the listed biological pathways.

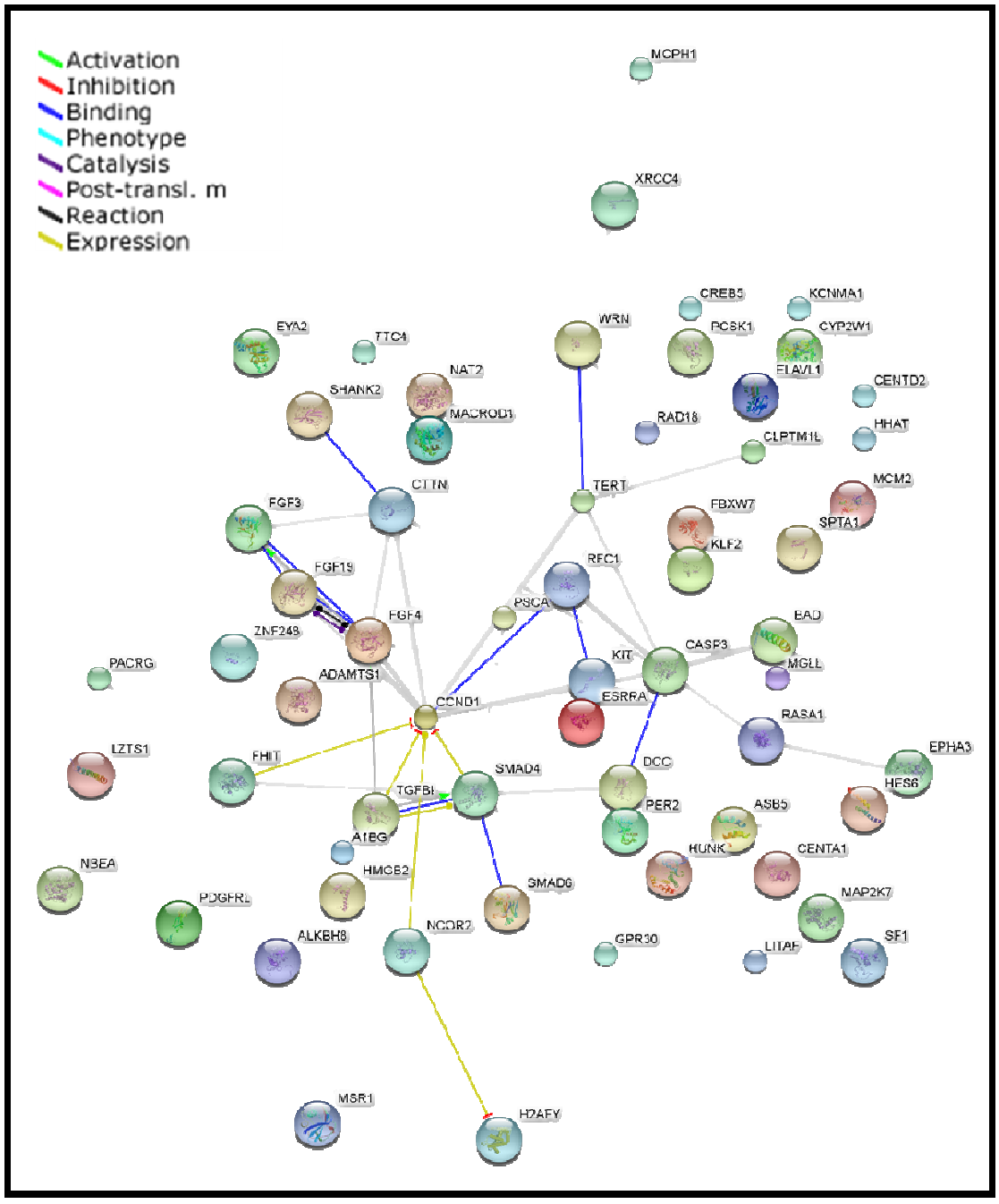

Figure 4.13 Predicted functional interactions for all potential oncogenes and tumour suppressor genes amplified or deleted in this study using the STRING 8.3 software (http://string.embl.de, accessed July 2010). The colour of the respective lines listed in the key relates to the type of relationship linking the proteins.

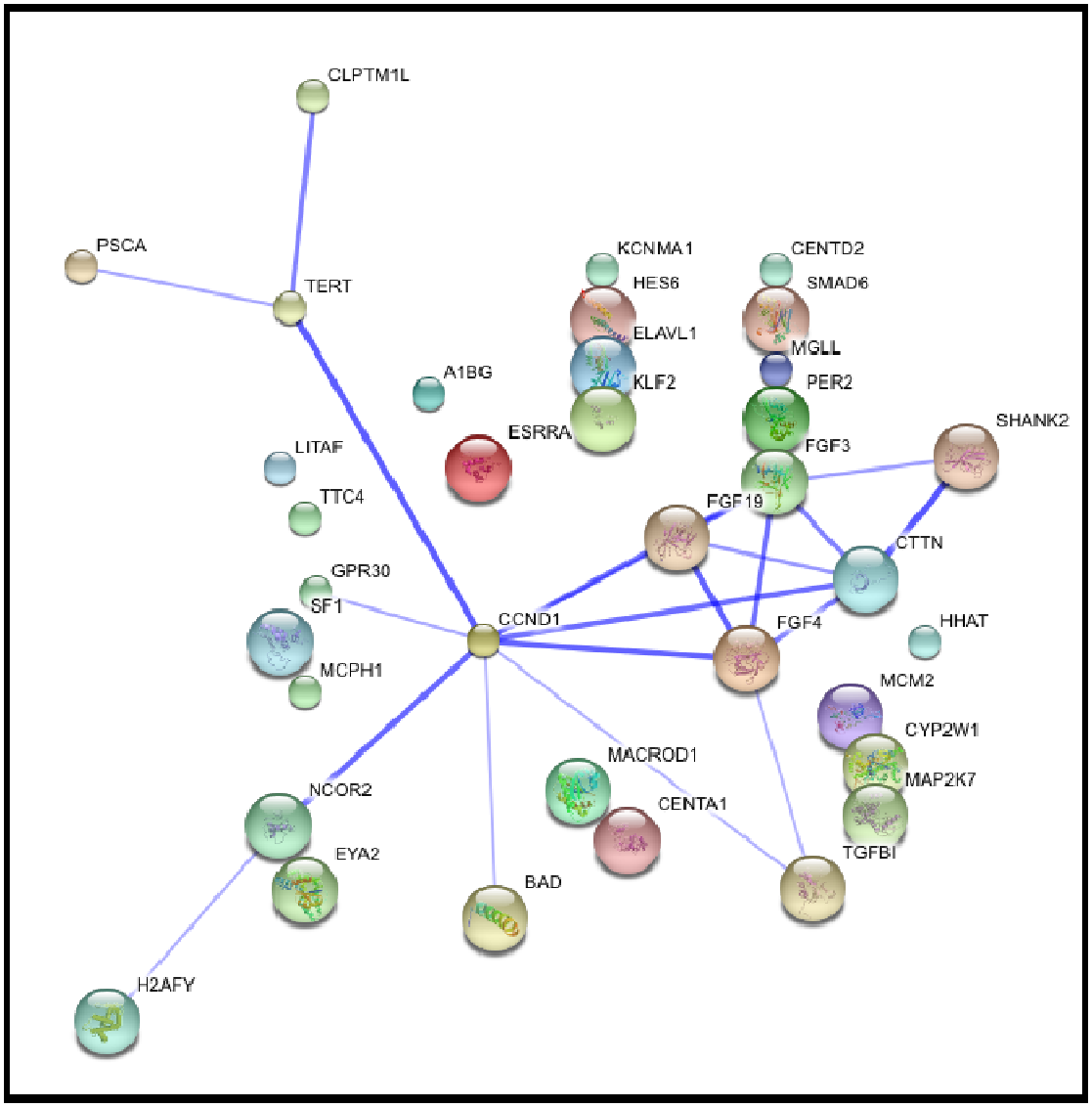

Figure 4.14 Predicted functional interactions for the genes in amplified regions as determined by the STRING 8.3 software. This network is the confidence view; with the line linking proteins being thicker depending on the combined confidence score for those two proteins. The thickness of the line represents the confidence score based on literature and experimental evidence. Confidence was set to medium confidence of 0.400.

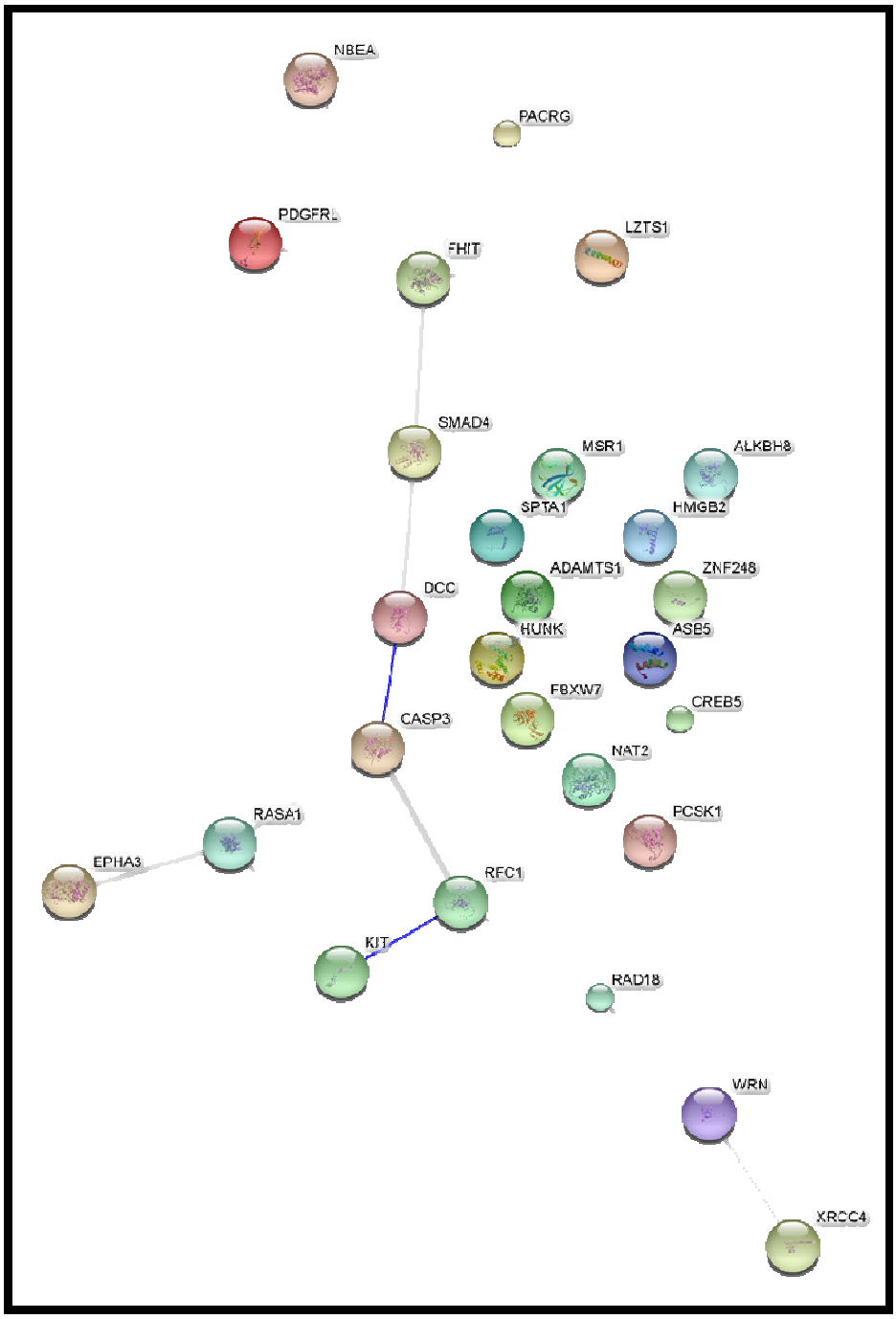

Figure 4.15 Predicted functional interactions for the genes in deleted regions as determined by STRING 8.3 software. This network is the confidence view. The confidence limit was set to medium confidence of 0.400.

4.4 Discussion

4.4.1 OSCC demographics in SA

In cancer registries, including the most recent PROMEC and MRC cancer registry report from the Eastern Cape of SA, the incidence of OSCC is said to be higher in males than in females with a prevalence of 31.3 and 18 per 100 000 individuals respectively with the age range between 65-69 years (Somdyala et al., 2010) i.e. a male to female gender ratio of 1.74:1. In contrast, the data from the cohort collected from the Nelson Mandela Academic Hospital in Umtata showed a gender ratio of 1:0.97 and a lower age of incidence at about 58.86 years compared to the report of Somdyala et al (2010). Although no definitive conclusion can be drawn from this observation, this warrants proper epidemiology studies to be conducted in Umtata where nearly as many females as males seem to be affected.

4.4.2 SNP arrays as a method for copy number analysis for FFPE samples

FFPE samples provide an important source of genomic material for research, and the availability of histological information allows for microdissection of samples with a high content of tumour cells. The challenges in using formalin fixed paraffin embedded tissues result from the variation in fixation methodologies. The major factors that affect sample quality are the time a sample is fixed in formalin and years of storage, which both have damaging effects on the DNA. The major effects of prolonged fixation are the cross-links formed between DNA strands, which inhibit downstream PCR techniques. These factors are not under the control of the researcher and good quality array results can be difficult to obtain. Analysis of SNP array technology results obtained from FFPE samples has been evaluated previously; it is found that larger fragments usually fail to amplify in the PCR step of the assay, which reduces the call rates (Jacobs et al., 2007). A multiplex PCR is

recommended to pre-determine the performance of samples in PCR and exclude poor samples. The average call rates detected in a published study were 79.84% and 75.17% for Nsp and Sty respectively (Jacobs et al., 2007). The average call rates for the samples in the present study were 70.05% and 73.38% for Nsp and Sty arrays respectively, with highest call rates obtained of 79.86% and 83.53% for Nsp and Sty respectively. Call rates for fresh tissue or blood are usually greater than 93% and the study by Jacobs et al (2007) determined that excluding SNPs residing on fragments of sizes larger than 700bp improved the call rates of FFPE samples such that they were concordant with those of fresh samples. Adding this step reduced the signal to noise ratio and although the number of SNPs available for copy number analysis was reduced to 308,788, there was sufficient data for copy number analysis. In addition, the FFPE samples generally do not yield as much PCR product as fresh samples and it was determined here that an additional three separate reactions were needed to make up the 90µg required for hybridisation to an array. Another study found good concordance between results from fresh samples and matching FFPE samples using the Affymetrix® 10K arrays (E. R. Thompson et al., 2005). Since Partek® Genomics Suite has a function to remove SNPs on large fragments from the data, it was used to analyse the copy number changes in the FFPE samples.

Only 51 out of 117 samples, for which DNA was extracted, could be analysed by SNP array technology, showing that the quality of FFPE samples remains an issue. More than double the required number of samples for analysis needs to be collected due to the poor sample quality of DNA obtained from FFPE samples. In instances where there is not enough DNA material, newer or alternative methods may be more suitable, for example, oligonucleotide arrays designed specifically for copy number detection. For example, the Agilent 4x44K

array platform was suggested to outperform SNP based analysis (Nasri et al., 2010). Next generation sequencing has also been suggested for analysis of FFPE samples and generated reproducible results comparable to those obtained from array-CGH (H. M. Wood et al., 2010). More recently, molecular inversion probes have been applied to copy number analysis in FFPE samples. This technique involves probes with two specific homology sequences that leave a 1bp gap when hybridised to the genome, they have sequence tags, which are read on a tag array. Molecular ligation is performed and then exonucleases eliminate the linear probes and two PCR primers (common to all probes) are used to amplify circularised sequences and introduce a fluorescent label (Yuker Wang et al., 2005). The sample is then hybridised to the tag array. The technology only requires sequences that are ~ 40bp in size, which is advantageous for samples with degraded DNA (Yuker Wang et al., 2005). Some of these methods are likely to improve analysis of copy number analysis changes in FFPE samples in the future.

Depsite the challenges in obtaining results on arrays with FFPE samples, 51 samples gave good quality results for analysis of copy number changes. The analysis of copy number changes in this OSCC cohort identified 20 regions of high amplification, 25 regions of moderate amplification, one homozygous deletion and 55 regions of hemizygous deletion. These regions encompassed 36 potential oncogenes and 27 potential tumour suppressor genes respectively.

4.4.3 Similarities between the five cell lines and the Eastern Cape patient cohort.

The similarity between the cell lines and the cohort is important for future studies as cell lines could be used as a model for functional analysis. The most striking similarity was a common deletion affecting the *EPHA3* gene on 3p11.1-p12.1. The 3p11 deletion observed in

the cohort coincided with the translocation breakpoints and deletions detected in the four cell lines WHCO1, WHCO3, WHCO6 and SNO. Amplification of the 11q13.3 region, which affected the *CCND1, CTTN, FGF3, FGF4* and *FGF19* genes and the 8q24.21 region, involving the c-*MYC* gene, were also common to both data sets.

Additional regions shared by the cell lines and the FFPE patient's cohort included deletion of chromosome 1p31.1 and involved the anti-proliferative gene, *IFI44*. This gene product binds and depletes GTP, resulting in cell cycle arrest (Hallen et al., 2007). Also commonly deleted was a 3p26.3 chromosomal region with a minimum region of overlap 966545-2566394bp. No genes previously described in cancer are in this region. Both data sets had a deletion affecting the 3p14.2 region harbouring the important tumour suppressor gene, *FHIT*. Chromosome 4p14-q35.2 deletion and 4q31.3 overlapped both data sets and involved the *SMAD1* gene and *SFRP2* genes respectively. Chromosome 6q26 deletion, affecting the *PACRG* gene, was also common to both data sets. Chromosome 8p22 was deleted in the cohort and the nearest deletion detected in the cell lines was 8p23.3-q21. Chromosome 18q21.1-q21.2, which was deleted in the cell lines, overlapped with the 18q21.1-q21.2 region in the cohort and involved the *DCC* and *SMAD4* genes.

Genes that were commonly affected by copy number changes in both cohorts need to be assessed by expression studies and could then be studied in the cell lines to evaluate their role in OSCC cell biology. The relevance of these genes in cancer is discussed further.

4.4.4 Comparison of this study's results to published literature on copy number changes

Previous international studies have highlighted amplification of chromosomes 1p36.13-p36.21, 2q31-q32, 3q28, 5p15.2, 6p25.3, 7p14, 8q24, 11q12.3, 20p11.1 and 20q12.1 as the

most common aberrations detected in OSCC (C. C. Yen et al., 2001; C. C. Yen et al., 2005; Xu Zhang et al., 2008; Chen, et al., 2008; Y. R. Qin et al., 2008; N. Hu et al., 2006; Carneiro, et al., 2008; Chattopadhyay et al., 2010).

The 11q and 8q amplicons are the most frequently reported amplicons which were also detected in this study (Chattopadhyay et al., 2010; Carneiro et al., 2008; Y. R. Qin et al., 2008; N. Hu et al., 2006; C. C. Yen et al., 2001; Du Plessis et al., 1999; Chen et al., 2008). Due to the recurrent involvement of these aberrations in different studies and including this one, they are likely to represent key abnormalities in OSCC carcinogenesis. In addition, amplification of 3q, 5p and 20q are also common, but the variability in the reported sub-bands is high. Chromosome 3q amplification is reported to affect the *SOX2* gene (3q26) in 15% of OSCC cases in the USA (Bass et al., 2009). This gene was not detected in the current study, again highlighting population or geographic differences in OSCC carcinogenesis. Chromosome 5p15.33 amplification may target the *TERT* oncogene, which is known to be involved in the maintenance of telomeres and has been implicated in multiple malignancies (C. G. Hsu et al., 2007; X. Kang et al., 2009; Metzger et al., 2004). Interestingly, *TERT* gene over expression results in the immortalization of oesophageal epithelial cells (H. Zhang et al., 2006). The 5p15.33 gain is shown to be the most frequent event in early stages of lung cancer (J. U. Kang et al., 2008) and it would be interesting to test for this abnormality in early lesions of OSCC in SA.

Chromosome 20q13.3 amplified in the present study was similar to the most frequently amplified region, 20q13.1, reported in other studies (Y. Fujita et al., 2003; C. C. Yen et al., 2001; Hirasaki et al., 2007). Gain of a whole chromosome 20, iso 20q or duplication of 20q was also the most persistent event in HPV16E6/E7 immortalized oesophageal cells (H. Zhang

et al., 2006). These findings suggest that 20q gain could be a non-random, early event contributing to cell immortalization. Chromosome 20q amplification has been associated with a poor prognosis in OSCC (Y. Fujita et al., 2003; C. C. Yen et al., 2003) as well as in various tumour types (Carneiro et al., 2008; Chen et al., 2008b; C. C. Yen et al., 2001; Hirasaki et al., 2007; Derek J Nancarrow et al., 2008). The 20q13.3 region amplified in this study harbours the *EYA2* gene, which is a transcriptional co-activator, up-regulated in ovarian carcinoma and promotes tumour growth (L. Zhang et al., 2005). Ovarian carcinoma patients with elevated expression of *EYA2* are shown to have a poor survival rate (L. Zhang et al., 2005). The *EYA2* gene is a likely target gene for amplification on chromosome 20q and may contribute to OSCC carcinogenesis.

The epidermal growth factor (*EGFR*) gene is commonly over expressed in OSCC, sometimes due to amplification of the gene on chromosome 7p14, which is a frequently reported amplicon in the literature (Hanawa et al., 2006; Gongyuan Zhang et al., 2010). Amplification of the *EGFR* gene was not detected in this cohort indicating that gene amplification of *EGFR* is not important in OSCC in SA. However, expression of this gene needs to be assessed in fresh samples from South Africa to establish if *EGFR* over expression occurs. The non-receptor tyrosine kinase gene, *TNK2*, located in the amplified 3q29 region, is known to bind EGFR and maintain EGFR on the cell surface. Knockdown of TNK2 by siRNA results in fewer EGFR molecules being expressed on the cell surface (Howlin et al., 2008). In the 5 OSCC cell lines *EGFR* is known to be over expressed (R Veale, 1989). It is possible that *TNK2* gene amplification may be one of the mechanisms contributing to EGFR protein expression on the cell surface in OSCC cell lines. It would be of value to assess *EGFR* expression in patient's samples.

Five novel or rarely described regions of amplification were seen in this study. Chromosome 1p amplification has also been reported in the literature with the most common region at 1p36. In this study a novel region, chromosome 1p32.3, was amplified. Chromosome 19p and 19q amplification have only been reported once before in OSCC by Carneiro et al (2008) (Carneiro et al., 2008). Here, 19q13.43 amplification was detected in 33 % of cases and was the region with the highest copy number change affecting one potential gene (the *A1BG* gene) not reported before in OSCC. Other novel regions of amplification included the 2p21 and 4q35.1 regions, which had no genes previously reported in cancer.

The deletions previously reported in OSCC in the literature most frequently affected chromosome 1p, 4p14-p15.3, 3p14.2, 3p26.3, 3p24.2, 5q33, 8p23.2, 9p21.3 and 18q21.2-q22 (C. C. Yen et al., 2001; C. C. Yen et al., 2005; Xu Zhang et al., 2008; Jing Chen, 2008; Y. R. Qin et al., 2008; N. Hu et al., 2006; Carneiro et al. 2008; Chattopadhyay et al., 2010).

Two homozygous deletions were detected and included the *BCL8* gene on chromosome 15q11.2 and the *CNN1* gene on chromosome 19p13.2. The *BCL8* gene has been implicated in translocations detected in diffuse-large B-cell lymphomas (DLBCL). Translocations involving the gene results in ectopic expression of *BCL8* and supports a role in DLBCL (Dyomin et al. 1997). There has been no evidence of *BCL8* disruption in OSCC and the exact function of the gene is unknown. Functional studies would be important to establish how, if at all, deletion of this gene could contribute to cacinogenesis in OSCC. The *CNN1* gene is a putative tumour suppressor gene in OSCC. *CNN1* is an actin binding protein and over expression of *CNN1* was shown to decrease cell proliferation in a uterine leiomyosarcoma cell line (Horiuchi et al. 1999). The gene is also down-regulated in rats with pre-cancerous lesions of the esophagus induced by zinc deficiency (Horiuchi et al. 1999). Dietary

deficiencies are associated with OSCC carcinogenesis and the *CNN1* gene deletions could be important in the carcinogenic process of OSCC. Both these regions were not affected by hemizygous deletion; this does not exclude the possibility of inactivating mutations in samples that did not have homozygous deletion in these two genes.

Similar regions of deletion, previously reported in the OSCC studies mentioned above, affecting chromosomes 4p (4p14-q35.2) and 3p26 (3p25.2-p26.3) were detected here. The commonly reported 3p14.3 (*FHIT*), 5q22.2 (*APC)*, 9p21.3 (*CDKN2A* and *B*) and 18q21.2-q22 (*DCC* and *SMAD4*) deletions in the above studies were also deleted in this study, highlighting them as important tumour supressors.

Chromosome 6q26 (*PACRG*) deletion has not been reported before in OSCC copy number studies. The high frequency (66.7%) of 6q26 deletion and the small size of the region suggests that this is a key target tumour suppressor gene in this study.

Chromosome 8p deletions are frequently reported in a variety of cancers (Di Benedetto et al., 2006; J. M. Flanagan et al., 2004; Kanae Ono et al., 2003) and an 8p loss was previously reported in a small cohort of OSCC from SA by Du Plessis et al (1999). These authors used conventional CGH and found that 8p deletion was found exclusively in black male patients from a cohort that included 29 males and females patients of black, colored and Caucasian ethnic groups (Du Plessis et al., 1999). The 8p deletions detected in this study were not exclusive to male patients but were present in both male and female patients. The 8p22 chromosome band has also been reported to be deleted in prostate, breast, lymphomas, hepatocellular and colorectal cancer (Bova et al., 1996; Di Benedetto et al., 2006; J. M. Flanagan et al., 2004; Thomassen et al., 2009; Ying et al., 2007). Chromosome 8p deletion is therefore an important region, requiring further investigation in OSCC.

4.4.5 *EPHA3* as a candidate gene in OSCC

Three of the cell lines had deletions affecting chromosome 3p11.2, harbouring the *EPHA3* gene, which consists of 17 exons. Two cell lines had deletions of sequences spanning exons 1-3, one cell line had deletions of both exons 1-3 and exons 11-17 and one cell line had a deletion in the region 5' to the *EPHA3* gene. In addition, the expression of the gene was found to be lower in the four OSCC cell lines that had an *EPHA3* deletion compared to normal oesophageal tissue (chapter 3, section 3.3). Hemizygous deletion of 3p11.1-12.1, which overlaps *EPHA3,* was detected in 74.5% of the patients by array copy number analysis. This was confirmed in 8 of these cases by FISH analysis. This gene is therefore of great interest and is suspected to be a candidate gene in OSCC. To interpret the possible role that deletion of *EPHA3* may have in OSCC, its function in normal and other cancerous tissues needs to be understood.

Structure and function of EPHA receptors

The *EPHA3* protein is an ephrin receptor tyrosine kinase. Eph receptors form the largest subfamily of receptor tyrosine kinases (Pasquale, 2010). Ephrins are the activating ligands for these receptors; they are tethered to the membrane by a glycosol phosphotidyl inositol moiety and are also capable of activating signalling within a cell (Binns et al., 2000). There are two classes of Eph receptors, A and B. The ephrin ligands are also classed as A or B as they usually bind specifically to either A or B class Eph receptors (Pasquale 2010). When ephrin receptors interact with an ephrin on an opposing cell, activation of signalling by the ephrin protein in that opposing cell can occur (reverse signal) and at the same time activation of the Eph receptor by the ephrin ligand induces signalling within the cell (foward signal) (Figure 4.16) (Pasquale 2010). The Eph receptors and ephrins that interact with

opposing cells, termed a *trans* interaction, results in the above mentioned bi-directional signalling. Eph receptors and ephrins can interact on the same cell, termed a *cis* interaction, which inhibits this bi-directional signalling (Figure 4.16) (Arvanitis & Davy, 2008). The bi-directional signalling of Eph receptors and ephrins regulates cytoskeletal structure and cell adhesion (Binns et al., 2000).

The EphaA receptors consist of an extracellular N-terminal domain, which binds with a ligand, a juxtamembrane domain, tyrosine kinase domain and a sterile alpha motif (SAM) (Binns et al., 2000). Binding of ephrin ligands to the extracellular domain, results in oligomerisation, which causes autophosphorylation of tyrosine residues, causing direct stimulation of catalytic activity of the tyrosine kinase. The oligomerisation also activates autophosphorylation within the juxtamembrane domain, which creates docking sites for SH2 domain containing proteins, activating these signalling molecules (Binns et al., 2000). In this way, the Eph/ephrin signalling interacts with other signalling pathways in mammalian cells, for example, MAPK, Akt, PI3K, FAK, JNK and ZO-1 (Arvanitis & Davy, 2008). These signalling interactions result in modulation of cell adhesion, cell sorting and patterning, cell migration, fusion, craniofacial development and platelet aggregation (Arvanitis & Davy, 2008). The deregulation of the Eph/ephrin signalling can therefore also contribute to carcinogenesis (Surawska et al., 2004).

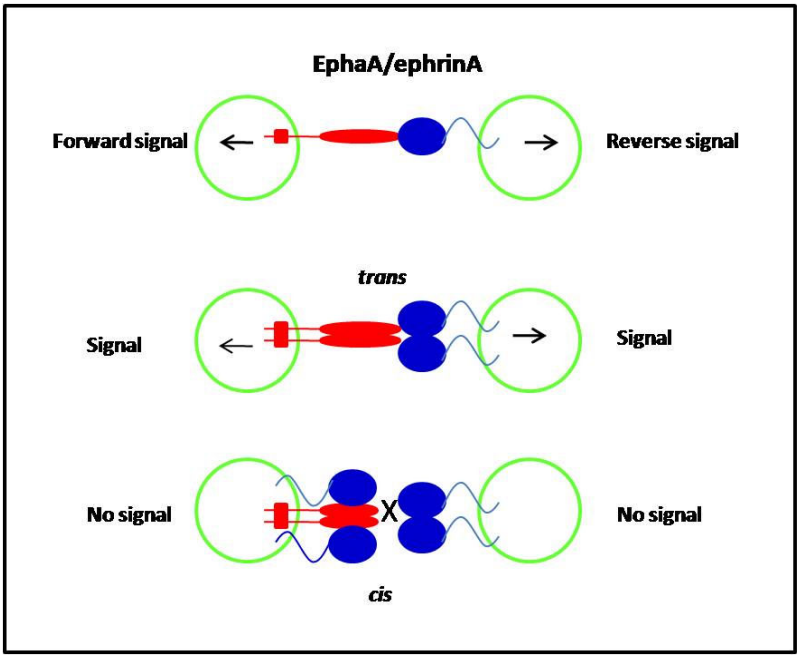

Figure 4.16 Schematic diagram of EphA/ephrin signalling adapted from Arvanitis *et al* (2008). Interaction between EphA receptors and ephrins on neighbouring cells results in signalling in both directions, reverse signalling activated by ephrin stimulation and forward signalling is activated by EphA receptor stimulation. Activation across cells is termed *trans* interaction. *Cis* interaction is when co-expressed EphA receptors and ephrins on the same cell interact and this inhibits signalling (Arvanitis & Davy 2008).

Role of Epha/ephrin signalling in cancer

The most important role for EphA proteins seems to be in cellular adhesion and de-adhesion. Cell adhesion is mediated through contact points or focal adhesions. These focal adhesions often bring activated receptor tyrosine kinases and their substrates into contact (Surawska et al. 2004). For example, EphA2 protein association with Fak protein (focal adhesion kinase) is constitutional and results in cell adhesion. Stimulation of EphA2 protein by ephrinA1 results in dissociation of the EphA2/Fak complex, which leads to cell detachment (Surawska et al., 2004). Contrastingly it was also shown that ephrin stimulation

triggers FAK phosphorylation, cell spreading and formation of adhesive structures (Surawska et al., 2004). The explanation for the contrasting functionality is proposed to be that Eph and ephrin density and composition on cells determines the signalling and cellular response. This in turn determines whether there is detachment or attachment of cells (Fox & Kandpal, 2004). This was illustrated by the study by Fox et al (2004).

The Eph receptors and ligands are involved in the correct positioning of epithelial cells in breast tissue (Fox & Kandpal, 2004). Fox et al (2004) studied the expression profiles of 13 Eph receptors and their 8 ligands in breast carcinoma cell lines. The study showed that up regulation of some of the Ephrins and EphA receptors contribute to tumourigenesis and down regulation of others plays a role in invasiveness (Fox & Kandpal, 2004). The up regulation of *EPHA2*, 7 and *10* specifically were related to tumourigenesis and invasiveness while down regulation of *EPHA3* was one of the factors specifically contributing to invasiveness (Fox & Kandpal, 2004). The Eph/ephrin network is therefore complex and the mechanisms of the signalling network are not fully understood and contradicting roles for *EPHA3* in cancer are reported in the literature.

Further evidence of *EPHA3* acting as a tumour suppressor is that soluble forms of EphA3 appear to inhibit tumour angiogenesis and tumour progression. Soluble EphA receptors delivered by transgene or osmotic mini pumps into pancreatic islet cell carcinoma results in inhibition of angiogenesis and reduced tumour size (N. Cheng et al., 2003). The soluble forms of EphA receptors were used in treatment of 4TI mouse mammary adenocarcinoma cell lines transplanted into Balb/c mice. This resulted in decreased vascular density and tumour growth (D. M. Brantley et al., 2002). This supports the idea that Eph receptors and their ligands function in tumour progression (D. M. Brantley et al., 2002).

One example of the *EPHA3* gene acting in an oncogenic way is that over expression of EphA3 protein results in resistance of prostate cancer cells to p53 mediated damage induced apoptosis. Down regulation of EphA3 by sh-RNA results in sensitization of prostate cancer cells to DNA damage induced apoptosis (T. Jiang et al., 2004). EphA3 appears to retain p53 in the cytosol, possibly by masking p53s nuclear localization signal thus inhibiting p53 (T. Jiang et al., 2004). EphA3 therefore has inhibitory effects on both the transcriptional activity and mitochondrial functions of p53 (T. Jiang et al., 2004).

The evidence for *EPHA3* acting as a tumour suppressor in the literature indicates that deletion of *EPHA3* may contribute to OSCC, perhaps in a similar mechanism to breast cancer as described by Fox et al (2004). Further investigation is required of the full Eph/ephrin complement in order to assess the exact contributions these proteins have in carcinogenesis. It is hypothesized that lowered expression of *EPHA3* may contribute to cell detachment, migration and metastasis.

4.4.6 Common copy number aberrations and potential pathways affected

The results of the analysis of biological pathway and functional interaction with the EASE and STRING software programs showed that the genes potentially affected by amplification were mostly involved in DNA metabolism (24%), transcription (22%), proliferation (18%) and cell signalling (12%). The deleted regions affected genes involved in cell communication (24%), nucleic acid metabolism (23%) and RNA transcription (12%). All of these pathways may be up-regulated or down-regulated resulting in uncontrolled cell growth in OSCC or any cancer. Potential mechanisms of these contributions are discussed.

Chromosome 11q amplification

There were three sub-amplicons on chromosome 11q. The *CTTN* gene is part of the most frequent (66.7%) and most amplified (57 copies) region on chromosome 11q13.3 (70061246-70310057bp). The two sub amplicons flanking the *CTTN* locus, 11q13.3 (68884395-70061246bp) and 11q13.3-q13.4 (70310057-71374778bp), were less amplified at 14 copies and in a lower number of samples (37.2% and 56.8% respectively). These two regions include the oncogenes, *CCND1*, *FGF3*, *FGF4*, *FGF19* and *SHANK2*. In light of the high level of amplification of the *CTTN* locus, the *CTTN* gene is likely the main target of amplification on chromosome 11q13 and an important oncogene. It has been shown that gene copy number in 11q13 amplicons is variable and that distinct cores (groups of genes) had differing levels of amplification (Sugahara et al. 2011). The combination of these cores, one of which included the *CTTN* gene, was specifically associated with lymph node metastasis in oral squamous carcinoma (Sugahara et al. 2011). The sub-amplicon detected on 11q in this study may also hold distinct groups of genes that are important in OSCC carcinogenesis. The *CTTN* gene is over expressed in OSCC pre-cancerous lesions, as well as in carcinogen induced murine OSCC supporting a role for *CTTN* in OSCC carcinogenesis (N. Y. Hsu et al. 2008). CTTN is an actin-binding protein and is involved in cell motility (P. Hofman et al. 2008), and has a role in actin reorganization (Du et al. 1998). CTTN facilitates cell motility by preventing anoikis through the PI3-Akt pathway (Du et al. 2009).

The *SHANK2* gene on the subamplicon 11q (70310057-71374778bp) encodes a protein involved in the regulation of trans-epithelial absorption (Han et al. 2006). The *SHANK2* gene is amplified in other cancers including oesophageal cancer (Carneiro, Isinger, Karlsson, J. Johansson, Jönsson, et al. 2008). The SHANK2 protein is also a ligand for the CTTN protein

through the SH3 binding domain of cortactin (Du et al. 1998). As co-amplified genes, *SHANK2* and *CTTN* over expression together may have a compound effect on carcinogenesis. In addition to these genes on chromosome 11q13, the FGF genes are also implicated in cancer and although *CTTN* is probably a more important target due the high level of amplification, the other genes in 11q13 may still contribute to OSCC.

In the literature, the FGF proteins (*FGF3*, *FGF4*, and *FGF19*) are reported to induce the Wnt pathway, which in turn is responsible for *CCND1* induced expression (Masuko Katoh & Masaru Katoh, 2006). CTTN was previously shown to down regulate the expression of cyclin–dependant kinase inhibitors; knockdown of CTTN by sh-RNA results in increased binding of the p21, p27 and p57 proteins to cyclin D1, inhibiting cells from progressing into the S-phase (Croucher et al., 2010). These reported interactions between the *FGF3*, *FGF4*, *FGF19*, *CCND1*, *CTTN* and *SHANK2* genes may be at play in OSCC in SA and all contribute to providing cell growth advantage. However, one or more of these genes may have a stronger impact on carcinogenesis, while others may be passenger genes within an amplicon. As could be seen in the relative quantification by RT-PCR in the cell lines (chapter 3), where the *FGF4* gene was not over expressed and is likely a passaneger in the 11q amplicon.

Chromosome 19p amplification

The most highly amplified region was on chromosome 19p13.11-p13.12, where an average of 111 copies was detected in 33.3% of samples. The region has 16 genes, none with clear involvement in cancer except the *KLF2* gene. This gene encodes a transcription factor. The *KLF2* gene has not been implicated in OSCC before and it is unclear whether it is oncogenic. Although it is over expressed in prostate cancer samples and neuroblastoma cell lines, it is also reported to have tumour suppressor function (Duhagon et al. 2010;Cotterman &

Knoepfler 2009). The KLF2 protein causes quiescence in T-cells by down-regulating *C-MYC* expression (Buckley et al. 2001). In addition to the other genes located in this region, many microRNAs have been mapped here and perhaps amplification of these may have a role in OSCC carcinogenesis.

The 270.81kb region from 58858174-59128983bp at 19q13.43 encompasses a target gene, the *A1BG* (alpha-IB-glycoprotein precursor) gene. This gene is over expressed in pancreatic cancer and was identified as a potential biomarker in pancreatic cancer as well as hepatocellular and bladder cancer (M. Tian et al., 2008; S. Y. Yoon et al., 2006; Kreunin et al., 2007). Higher levels of the A1BG protein are detected in the serum of patients with gastrointestinal tumours such as colorectal and gastric tumours, highlighting this gene as a biomarker for non-invasive screening procedures (Szilvás et al., 1998).

The high level of amplification on chromosome 19, suggests that novel candidates located in these sub-bands may be of interest in OSCC carcinogenesis. Amplification of the above mentioned candidate genes should be verified in more samples by an alternative method and investigation into their role in OSCC is warranted.

Amplifications and deletions affecting the TGF-β* and *Wnt signalling pathways

Transforming growth factor-beta signaling (TGF-β) disruption is also implied by the aberrations detected in this study. TGF-β signalling can have a tumour suppressive role or promote tumour growth depending on the balance of molecules within the pathway (Ikushima & Miyazono 2010). The *SMAD6* oncogene was amplified on chromosome 15q22.31 (average copy of 26.9 copies) in 25.5% of samples. The *SMAD6* gene encodes a protein involved in the TGF-β pathway and usually inhibits the apoptotic function of the

TGF-β pathway (Attisano, Silvestri, Izzi, & Labbé, 2001). Increased expression of *SMAD6* increased cell growth in pancreatic cancer (Kleeff et al., 1999). In contrast, *SMAD6* over expression was found in early OSCC tumours and was inversely correlated with depth of invasion (Osawa, Nakajima, Kato, Fukuchi, & Kuwano, 2004). In addition, in oral squamous cell carcinoma, *SMAD6* over expression was associated with a better outcome in conjunction with a loss of *SMAD2* expression (Mangone et al. 2010). It is suggested that expression of *SMAD6* inhibits the pro-oncogenic role of the TGF-β pathway in later carcinogenesis (Mangone et al. 2010). It is clear from the contrasting results on *SMAD6* expression that further investigation of the role of the SMAD6 protein in OSCC is required.

The Smad 4 protein (encoded by *SMAD4*) is activated by TGF-β and together with other proteins activates transcription of genes involved in cell cycle regulation (Attisano, Silvestri, Izzi, & Labbé, 2001). The *SMAD4* gene was part of the deleted region on chromosome 18q21.1-q21.2 in 70.6% of cases. Loss of *SMAD4* expression was previously reported in 51.2% of OSCC patients and correlated with depth of invasion (Fukuchi et al. 2002). Hence, loss of expression of *SMAD4* may cumulatively add to decreased cell cycle regulation within the TGF-β signaling pathway.

In addition to the *SMADs*, the *FBXW7* gene also involved in the TGF-β pathway, was targeted by deletion of chromosome 4p14-q35.2. The FBXW7 protein targets proteins for degradation and specifically the TGIF1 (TGF-β induced factor 1) protein. When TGIF1 is not degraded, accumulation of phosphorylated TGIF results in suppression of TGF-β dependant transcription and de-regulated cell growth and migration (Bengoechea-Alonso & Ericsson 2010). The *MYC* gene on 8q24 was moderately amplified (8.5 copies) in 21.57% of the samples, which may result in over expression of *MYC*. The MYC protein is also targeted for

degradation by the FBWX7 protein (Y. Fujii et al. 2006). Consequently, deletions of the *FBXW7* gene could contribute to the the loss of apoptosis normally induced by TGF-β and also contribute to the accumulation of the MYC oncoprotein. A graphic representation of the potential interactions of the above mentioned genes in TGF-β signalling can be seen in figure 4.17.

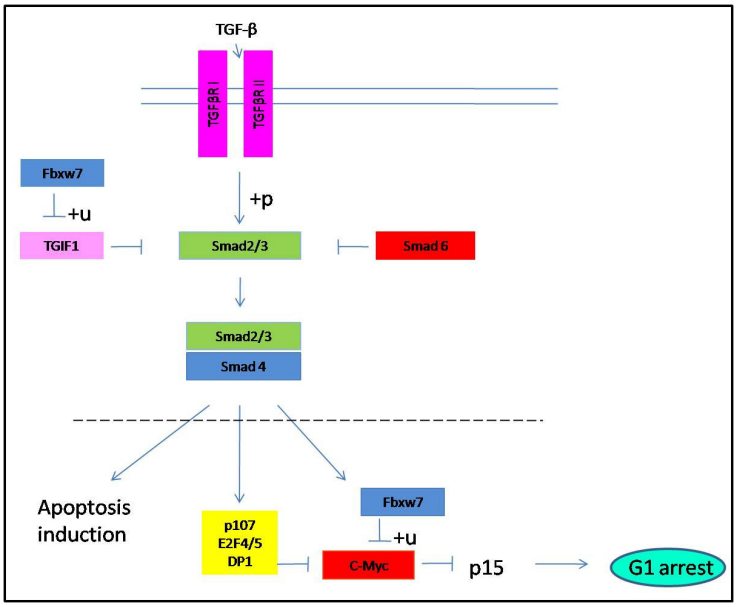

Figure 4.17. Schematic representations of TGF-β signalling. The red coloured blocks are genes amplified in the OSCC cohort and blue blocks are genes deleted in this cohort. The potential interaction of these amplified and deleted genes may contribute to inactivation of the TGF-β pathway and induction of apoptosis. Smad 6 inhibits the activity of smad2/3 and smad4, which would normally activate apoptosis through activation of transcription. If TGIF1 is not targeted for degradation by the Fbxw7 protein, it will inhibit the smad2/3 and 4 proteins again inhibiting apoptosis.

The *FHIT* gene was part of a deleted region on 3p detected in 21.6% of samples. This gene is an important known tumour suppressor. The mode of *FHIT* tumour suppressor function has

been debated but it has been shown that it likely regulates the Wnt pathway (Weiske et al. 2007). The FHIT protein binds to the LEF-1/TCF/β-catenin complex and binding of Fhit to β-catenin is essential in the repressive function of Fhit on transcription through the LEF-1/TCF/β-catenin complex (Weiske et al., 2007). Previous studies of FHIT knock out, resulted in increased TCF/β-catenin mediated transcription of genes such as *CCND1* (Weiske et al., 2007). FHIT may also regulate *CCND1* expression by down-regulating cyclophilin A, a promoter of Cyclin D1 synthesis (Shuho Semba & Kay Huebner, 2006). Loss of *FHIT* may therefore contribute to the deregulation of the Wnt pathway and transcription of genes downstream of this pathway.

The gene, *APC*, is a key regulator of β-catenin degregadation and therefore downregulation of the Wnt pathway. The gene was included in the deleted region on chromosome 5, 5q21.2-q23.3. The *SFRP2* gene deletion, on chromosome 4q31.3-q35.1, was detected in 74.5% of samples. This gene is also a regulator of the Wnt pathway as described in chapter 2. The *SFRP2* gene encodes a frizzled-related Wnt inhibitor (Kongkham et al., 2010). The fibroblast growth factor pathway, involving the *FGF3*, *FGF4* and *FGF19* genes may also feed back into Wnt signalling, by inhibiting GSK-3β targeting of β-catenin for degradation (Masuko Katoh & Masaru Katoh, 2006). In a scenario where the FGF genes are overexpressed and stimulate the Wnt pathway, over expression of the c-*MYC* and *CCND1* genes may occur as they are transcriptional targets of the TCF/β-catenin complex, which would not be inhibited if β-catenin is not degraded. Graphic representation of the Wnt pathway and the gene interactions can be seen in figure 4.18. A recent publication found that treatment of oesophageal carcinoma cell lines with methylselenic acid could reduce the β-catenin levels and inhibit the TCF/β-catenin pathway, which would suggest that this is one

manner in which selenium reduces oesophageal cancer risk (W. Zhang et al., 2010). This effect was reversed by over expressing β-catenin (W. Zhang et al., 2010). This would suggest that a selenium nutrient deficiency, commonly associated with OSCC, may act on the Wnt/β-catenin pathway.

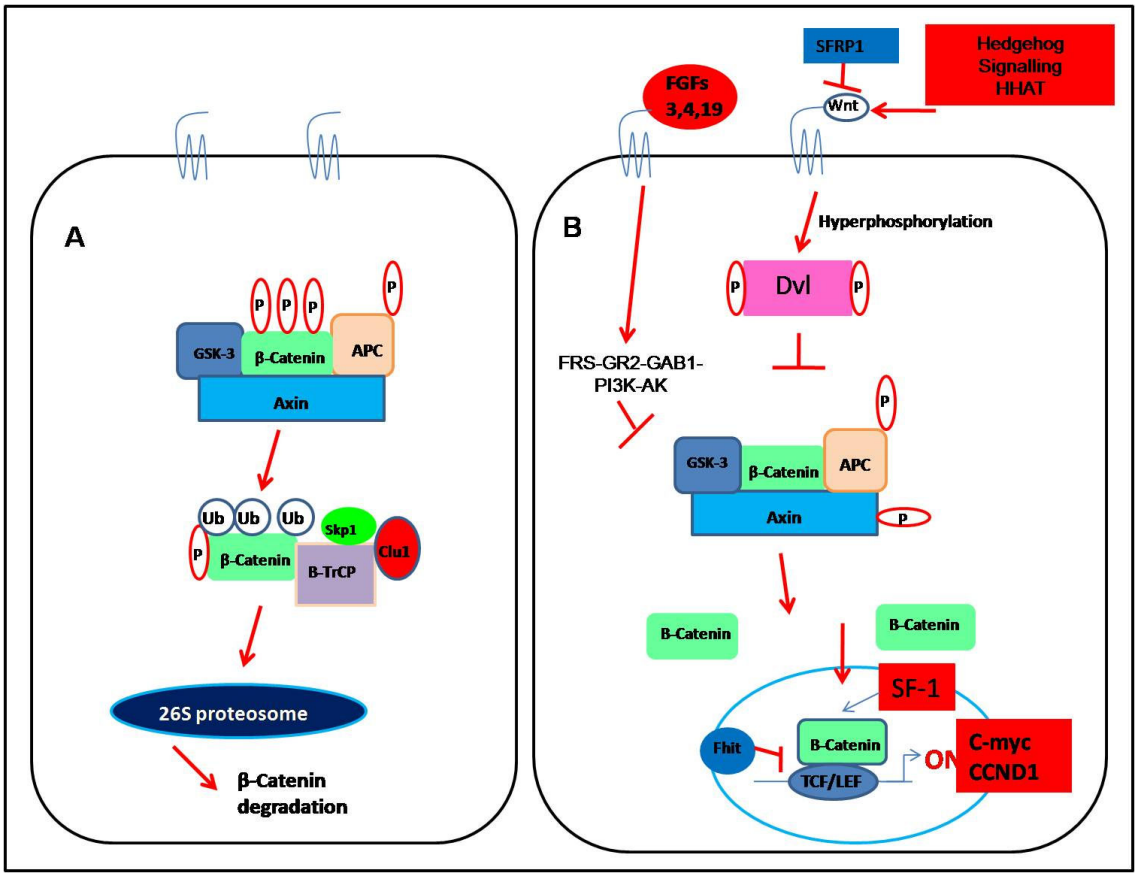

Figure 4.18. Schematic representations of the Wnt, FGF and Hedgehog pathway interactions. A In the non activated state, β-catenin is targeted for degradation. **B** Stimulation of the Wnt pathway results in activation of β-catenin/TCF/LEF transcription of genes. Blue blocks indicate genes deleted in this cohort and red blocks indicate genes amplified in this cohort.

Genes with aberrant copy number involved in genomic stability

Genes, in regions of deletions in this study, are implicated in DNA repair pathways and genomic instability. These genes included *WRN*, *ATM*, *XRCC4* and *RAD18*.

The *WRN* gene (deleted on 8p21.3-p22 in 21.57% of cases) expresses a protein involved in processing stalled replication forks and recovery from replication stress (Franchitto et al., 2008). The WRN protein is phosphorylated under conditions of replication arrest or DNA damage, causing stalling of the replication fork (Pichierri et al., 2003). In addition WRN is important for maintaining stability of fragile sites. WRN accumulates at the sites of aphidicolin induced replication delay and cells lacking WRN showed increased incidence of gaps and breaks in the common fragile sites including FRA3B (Pirzio et al., 2008). Cells showing loss of this protein generally show premature senescence, genomic instability and telomere loss (F. J. Liu et al., 2010). Wrn depleted cells are more sensitive to DNA damage (Mao et al., 2010).

The *ATM* gene, deleted on 11q22.3 in 49% of cases, is a critical gene involved in the repair of double-stranded breaks and DNA damage checkpoints (K. Suzuki et al., 2010). The ATM protein is also involved in regulating WRN; the ATM protein is required to maintain WRN protein in a phosphorylated status (Pichierri et al., 2003) and together they assist in the prevention of double stranded breaks (Ammazzalorso et al., 2010). The *RAD18* gene was in the deleted region on 3p25.3-p26.3 in 66.7% of cases; this gene is involved in DNA damage responses during replication stress and suppresses the formation of gross chromosomal rearrangements (Putnam et al., 2010). RAD18 protein accumulates at sites of DNA damage and promotes homologous recombination (Jun Huang et al., 2009). The *XRCC4* gene on chromosome 5q14.1-q14.2, which was deleted in 54.9% of cases and encodes a ligase that is

essential for canonical non-homologous end-joining repair of double stranded breaks, specifically related to radiation induced DNA damage (Ferguson & Alt, 2001). This enzyme was shown to suppress the occurrence of translocations in mammalian cells exposed to radiation (Simsek & Jasin, 2010). Deletion of this gene may also contribute to genetic instability in OSCC.

The combination of deletions affecting these genes involved in repair and maintaining genomic stability may contribute to the high number of rearrangements and genomic instability seen in OSCC. Specifically, they may contribute to the deletion or amplification of tumour suppressor genes/oncogenes residing in areas of chromosome instability like the *FHIT*, *BCL6*, *C-MYC* and *PACRG* genes residing in the FRA3B, FRA3C, FRA8C and FRA6E fragile sites respectively.

Potential markers for prognosis, diagnosis and treatment

Comments on prognostic markers cannot be made in this study due to the lack of clinical data but genes with reported involvement in disease progression and metastasis in cancer are worth mentioning. *C-MYC* gene expression is associated with a poor prognosis, correlating with depth of invasion and short survival and could therefore be used as a marker for prognosis (Wei Wang et al., 2010). Although this gene was not highly amplified in the cohort, high expression was detected in the cell lines and the expression of C-*MYC* should be investigated further in OSCC samples.

Five genes potentially affected here could contribute to invasiveness and metastasis. The *MACROD1* gene (amplified on 11q13.1 in 19.6% of cases) is also known as the leukaemia

related protein 16 (*LRP16*). Over expression of this gene is detected in gastric cancers and is associated with lymphoid invasion and metastasis (Y. Z. Li et al., 2009).

Higher expression of *KCNMA1* (amplified on 10q21.3 in 19.6% of cases) is reported in metastatic breast cancer and siRNA knockdown of *KCNMA1* resulted in reduced invasion (Khaitan et al., 2009).

The *PSCA* gene (amplified on 8q24.3 in 70.6% of cases) is over expressed in prostate cancer and is associated with metastasis (Ananias et al., 2009; Raff et al., 2009). The *PSCA* gene is also over expressed in pancreatic cancer and can be detected in blood of these patients; it could therefore be used as a marker for detection (Grubbs et al., 2006).

Genes involved in drug metabolism or resistance were *CYP2W1* (amplified on 7p in 41.2% of cases), *LZTS1* (deleted on 8p12-p21.2 in 19.6% of cases) and *KLF2* on 19p13.43 (amplified in 66.7% of cases). The *CYP2W1*gene encodes a cytochrome p450 enzyme that is expressed in colon and adrenal cancers and could be a potential drug target (Karlgren et al., 2006).The *LZTS1* gene loss is shown to predict treatment response to taxane-based therapy in ovarian carcinoma (D. Califano et al., 2010), it is also highlighted as a candidate tumour suppressor in oesophageal squamous carcinoma (Chattopadhyay et al., 2010).

Some of these genes have not been described in OSCC before and may be worth investigating in a study of OSCC samples from different stages of disease where accurate clinical data is available.

4.4.7 Summary

In summary, this chapter has established the genomic copy number map in OSCC from patients in the Eastern Cape and has highlighted genes potentially contributing to the

uncontrolled cell growth and genomic instability which is characteristic of OSCC. The detection of *EPHA3* deletions possibly implicates a new signalling network (Eph/Ephrin) that may contribute to OSCC development. DNA repair pathways were highlighted by the potential disruption of the homologous recombination (loss of *RAD18*) and non homologous end joining repair (loss of *XRCC4*) pathways, which would fit with the high level of genomic instability observed in OSCC. There were key pathways highlighted by the abnormalities detected in this study. These included the Wnt signalling (*FHIT*, *SFRP2*, *CCND1*, C-*MYC*), the FGF signalling (*FGF3, FGF4, FGF19*) and TGF-β signalling (*SMAD4*, *SMAD6*, *FBXW7*) pathways, which may contribute to uncontrolled cell growth. There is cross talk between these pathways, which may additively contribute to OSCC carcinogenesis. Future studies will be aimed at investigating the most commonly rearranged genes/ pathways in depth. These results opened new avenues for the investigation of OSCC in SA, and are likely to be valuable in developing an understanding of OSCC carcinogenesis in the future.

CHAPTER 5

Final conclusions and future directions

5.1 Conclusions

This study is the first in South Africa to look at whole genome copy number aberrations in OSCC using microarray technology and the first comprehensive molecular cytogenetic analysis of South African OSCC cell lines. Twenty nine key candidate genes have been highlighted by this study. The *KLF2, A1BG, SHANK2, CCND1, FGF3, FGF4, FGF19, CTTN, C-MYC, MACROD1, KCNMA1, PSCA, CYP2W1, BCL6, TNK2* and *SMAD6* genes were in regions of amplification. The *SMAD4, FBXW7, FHIT, EPHA3, SFRP2, LZTS1, WRN, ATM, APC, RAD18, XRCC4* and *PACRG* genes are potential tumour suppressor genes that were affected by deletions in this study.

The *EPHA3* gene is a novel gene described in OSCC. This gene locus was deleted, in a large proportion of samples and cell lines and *EPHA3* gene expression appears to be lost in cell lines with deletions. A hypothesis is proposed that the Eph/ephrin protein network, ephrin receptors and their ligands, are at play in the aggressive and invasive nature of OSCC. Deletion of the *EPHA3* gene may contribute to cell detachment and invasiveness. This hypothesis needs to be tested by analysing the expression of Ephrin receptors and their ligands in fresh OSCC samples. Ideally, a cell line model in normal oesophageal epithelium and oesophageal carcinoma cell lines would facilitate functional studies of the interplay of these molecules in oesophageal epithelium and whether they function in OCC carcinogenesis. In addition, *EPHA3* mutations have been detected in various tumours (Wood et al. 2006) and in view of this, investigation of the prescence of *EPHA3* mutations should be done on these patients and the five OSCC cell lines.

The genes identified in copy number aberrations in the tumour samples appear to be involved in four main signalling pathways, the Wnt, TGF-β, Hedgehog and Fibroblast growth

factor pathways. These pathways are all interconnected and the combined deregulation of molecules in these pathways is likely to contribute to the aggressive nature of OSCC. The *CCND1, CTTN, FGF3, FGF4, FGF19, C-MYC, FHIT, SPRP2, EPHA3, SMAD4* and *DCC* genes were affected in the 5 cell lines, which would suggest that these aberrations have been carried through numerous passages and would indicate that they are essential gene aberrations required for maintaining permanent cell survival. The fact that aberrations affecting these genes were detected in the cohort supports their potential importance. Expression studies in the cell lines confirmed that gene amplification of the *CCND1*, c-*MYC* and *FGF19* genes correlated with over expression of these genes making them likely targets of amplification. In addition, the *CCND1* and C-*MYC* genes were over expressed in cell lines showing polyploidy. Other mechanisms in addition to amplification may also contribute to over expression of these genes. For example, the loss of the *FBXW7* gene, which targets the C-MYC protein for degradation, could contribute to the accumulation of C-MYC protein. Loss of FHIT protein contributes to the deregulation of the Wnt pathway, which results in transcription of *CCND1* and C-*MYC*. Fhit also regulates the cyclophilin A protein which promotes Cyclin D1 synthesis. In light of this, cases where amplification of *CCND1* and C-*MYC* were not observed may still have over expression of these genes.

The genomic instability of this cancer was recognised in the complexity of genetic aberrations as seen in this study. This could be explained by deletions of genes that are important for normal functioning of DNA repair and control of genomic stability such as the *WRN, ATM, RAD18* and *XRCC4* genes. The genes *FHIT*, *BCL6, C-MYC* and *PACRG* all reside at fragile sites. Fragile sites are prone to deletion and rearrangement, especially in DNA repair

deficient cells, and are sensitive to environmental carcinogen exposure, the main risk factor associated with OSCC.

5.2 Future directions

The genes that are of specific importance going forward are the *EPHA3, CCND1, CTTN, FGF3, FGF4, FGF19, C-MYC, FHIT, SFRP2, SMAD4, APC, DCC, WRN, ATM, PACRG, RAD18* and *XRCC4* genes. It is envisaged that an expression study of these genes in a cohort of at least 50 fresh OSCC samples will be done to confirm the association between amplification/deletion and deregulated expression. A second objective would be to isolate the functional roles of deregulated candidates in normal and cancerous oesophageal cell line models, either by knock down of over expressed genes or re-introduction of deleted genes. A full scale study of this kind for the Ephrin receptor/ligand network is of particular interest as it has not been studied in OSCC. Studies of this nature could contribute to establishing how these genes contribute to OSCC carcinogenesis and assist in identifying targets for therapy.

Markers for early detection of OSCC have not yet been isolated and implemented. The difficulty in obtaining early disease specimens for research is owing to the late presentation of patients to clinics and hospitals. If early and late stage disease specimens could be obtained, some of the aberrations detected in this study could be analysed for their stage of involvement. In particular, it would be hypothesized that genes involved in DNA repair would be deregulated first, so loss of *WRN*, *ATM*, *XRCC4* and *RAD18* would be expected to be lost early in disease pathogenesis. Loss of DNA repair and genome stability functions would facilitate deletion of genes in genetically instable regions early on and then

amplification of chromosomal regions targeting genes that facilitate cell proliferation such as the *CCND1*, *FGF3*, *FGF4*, *FGF19*, *CTTN* and C-*MYC* genes may occur. A suggested method to study early events would be to analyse tissue blocks by immunohistochemistry to determine the expression patterns of the above mentioned candidate genes in samples with various stages of disease that can be identified by a pathologist. This may help identify those genes that are deregulated early in the development of the disease and which of them could be used as markers for early screening. Overall, this study has contributed to the literature on DNA copy number aberrations in OSCC and has generated some hypotheses and strategies for investigation of OSCC pathogenesis going forward.

REFERENCES

Abeysinghe, S.S. et al., 2003. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Human mutation*, 22(3), pp.229-44.

Albertson, D.G. & Pinkel, D., 2003. Genomic microarrays in human genetic disease and cancer. *Human molecular genetics*, 12 Spec No, pp.R145-52.

Ammazzalorso, F. et al., 2010. ATR and ATM differently regulate WRN to prevent DSBs at stalled replication forks and promote replication fork recovery. *The EMBO journal*, 29(18), pp.3156-69.

Ananias, H.J.K. et al., 2009. Expression of the gastrin-releasing peptide receptor, the prostate stem cell antigen and the prostate-specific membrane antigen in lymph node and bone metastases of prostate cancer. *The Prostate*, 69(10), pp.1101-1108.

Antonsson, A. et al., 2010. High-Risk Human Papillomavirus in Esophageal Squamous Cell Carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, 19(8), pp.2080-2087.

Arai, H. et al., 2003. Detection of amplified oncogenes by genome DNA microarrays in human primary esophageal squamous cell carcinoma: comparison with conventional comparative genomic hybridization analysis. *Cancer genetics and cytogenetics*, 146(1), pp.16-21.

Arvanitis, D. & Davy, A., 2008. Eph/ephrin signaling: networks. *Genes & development*, 22(4), pp.416-29.

Attisano, L. et al., 2001. The transcriptional role of Smads and FAST (FoxH1) in TGFbeta and activin signalling. *Molecular and cellular endocrinology*, 180(1-2), pp.3-11.

Bailey, L., 2003. Folate, methyl-related nutrients, alcohol, and the MTHFR 677C-->T polymorphism affect cancer risk: intake recommendations. *The Journal of Nutrition,*133, pp.3748-3753.

Bartsch, Helmut & Nair, J., 2006. Chronic inflammation and oxidative stress in the genesis and perpetuation of cancer: role of lipid peroxidation, DNA damage, and repair. *Langenbeck's archives of surgery / Deutsche Gesellschaft für Chirurgie*, 391(5), pp.499-510.

Bass, A.J. et al., 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature genetics*, 41(11), pp.1238-42.

Bayani, J. et al., 2007. Genomic mechanisms and measurement of structural and numerical instability in cancer cells. *Seminars in Cancer Biology*, 17(1), pp.5-18.

Bedoya, F. et al. 2009. Mutation in CNR1 gene and VEGF expression in esophageal cancer. *Tumori*, 95(1), pp.68-75.

Beers, E.H. van et al., 2006. A multiplex PCR predictor for aCGH success of FFPE samples. *British journal of cancer*, 94(2), pp.333-7.

Bengoechea-Alonso, M.T. & Ericsson, J., 2010. Tumor suppressor Fbxw7 regulates TGFβ signaling by targeting TGIF1 for degradation. *Oncogene*, 29(38), pp.5322-8.

Bennett, W.P. et al., 1997. p53 mutations in esophageal tumors from a high incidence area of China in relation to patient diet and smoking history. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 6(11), pp.963-6.

Beroukhim, R. et al., 2007. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), pp.20007-12.

Bey, E. et al., 1976. Carcinoma of the esophagus in Africans: establishment of a continuously growing cell line from a tumor specimen. *In vitro*, 12(2), pp.107-14.

Bianco, S. et al., 2009. Modulating estrogen receptor-related receptor-alpha activity inhibits cell proliferation. *The Journal of Biological Chemistry*, 284(35), pp.23286-23292.

Bignell, G.R. et al., 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome research*, 14(2), pp.287-95.

Binns, K.L. et al., 2000. Phosphorylation of tyrosine residues in the kinase domain and juxtamembrane region regulates the biological and catalytic activities of Eph receptors. *Molecular and cellular biology*, 20(13), pp.4791-805.

Bokobza, S.M. et al., 2009. Reduced expression of BMPR-IB correlates with poor prognosis and increased proliferation of breast cancer cells. *Cancer genomics & proteomics*, 6(2), pp.101-8.

Boone, J. et al., 2009. Targets for molecular therapy in esophageal squamous cell carcinoma: an immunohistochemical analysis. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 22(6), pp.496-504.

Bova, G.S. et al., 1996. Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. *Genomics*, 35(1), pp.46-54.

Brantley, D.M. et al., 2002. Soluble Eph A receptors inhibit tumor angiogenesis and progression in vivo. *Oncogene*, 21(46), pp.7011-26.

Brevern, M.C. von et al., 1996. Circulating anti-p53 antibodies in esophageal cancer patients are found predominantly in individuals with p53 core domain mutations in their tumors. *Cancer research*, 56(21), pp.4917-21.

Brooks, P.J., Goldman, D. & Li, T.-K., 2009. Alleles of alcohol and acetaldehyde metabolism genes modulate susceptibility to oesophageal cancer from alcohol consumption. *Human genomics*, 3(2), pp.103-5.

Buckley, A et al. 2001. Transcription factor LKLF is sufficient to program T cell quiescence via a c-Myc--dependent pathway. *Nature immunology*, 2(8), pp. 698-704.

Burger, M.J. et al., 2002. Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer. *International journal of cancer. Journal international du cancer*, 100(2), pp.228-37.

Burkhardt, L. et al., 2010. Gene amplification in ductal carcinoma in situ of the breast. *Breast cancer research and treatment*, 123(3), pp.757-65.

Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of molecular endocrinology*, 25(2), pp.169-93.

Califano, D. et al., 2010. FEZ1/LZTS1 protein expression in ovarian cancer. *Journal of cellular physiology*, 222(2), pp.382-6.

Camps, J. et al., 2009. Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. *Genes, chromosomes & cancer*, 48(11), pp.1002-17.

Carneiro, A. et al., 2008. Prognostic impact of array-based genomic profiles in esophageal squamous cell cancer. *BMC Cancer*, 8, p.98.

Casson, A.G. et al., 1998. p53 alterations in oesophageal cancer: association with clinicopathological features, risk factors, and survival. *Molecular pathology : MP*, 51(2), pp.71-9.

Chan, J.Y. et al., 2008. Resveratrol displays converse dose-related effects on 5-fluorouracil-evoked colon cancer cell apoptosis: the roles of caspase-6 and p53. *Cancer biology & therapy*, 7(8), pp.1305-12.

Chang, G et al., 2010. Hypoexpression and epigenetic regulation of candidate tumor suppressor gene CADM-2 in human prostate cancer. *Clinical cancer research,* 16(22), pp. 5390-401.

Chang, Y.C. et al., 2010. Molecular cytogenetic characterization of esophageal cancer detected by comparative genomic hybridization. *Journal of clinical laboratory analysis*, 24(3), pp.167-74.

Chang, Z.W. et al., 2009. [Characterization of the changes in comparative genomic hybridization in esophageal cancer patients with family history]. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*, 29(6), pp.1166-9.

Chattopadhyay, I. et al., 2010. Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India. *Mutation Research*, 696(2), pp.130-138.

Chelule, P.K. et al., 2006. The frequency of cytochrome P450 2E1 polymorphisms in Black South Africans. *Disease markers*, 22(5-6), pp.351-4.

Chen, Jing et al., 2008. Genomic profiling of 766 cancer-related genes in archived esophageal normal and carcinoma tissues. *International Journal of Cancer. Journal International du Cancer*, 122(10), pp.2249-2254.

Chen, L.C., Dollbaum, C. & Smith, H.S., 1989. Loss of heterozygosity on chromosome 1q in human breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 86(18), pp.7204-7.

Chen, L. et al., 2009. Down-regulation of tumor suppressor gene FEZ1/LZTS1 in breast carcinoma involves promoter methylation and associates with metastasis. *Breast cancer research and treatment*, 116(3), pp.471-8.

Cheng, N. et al., 2003. Inhibition of VEGF-dependent multistage carcinogenesis by soluble EphA receptors. *Neoplasia (New York, N.Y.)*, 5(5), pp.445-456.

Chetty, R & Chetty, S., 1997. Cyclin D1 and retinoblastoma protein expression in oesophageal squamous cell carcinoma. *Molecular pathology : MP*, 50(5), pp.257-60.

Chetty, R & Simelane, S., 1999. p53 and cyclin A protein expression in squamous carcinoma of the oesophagus. *Pathology oncology research : POR*, 5(3), pp.193-6.

Cheung, L.C.M. et al., 2007. Establishment and characterization of a new xenograft-derived human esophageal squamous cell carcinoma cell line HKESC-4 of Chinese origin. *Cancer genetics and cytogenetics*, 178(1), pp.17-25.

Chung, Y. et al., 2007. Altered E-cadherin expression and p120 catenin localization in esophageal squamous cell carcinoma. *Annals of surgical oncology*, 14(11), pp.3260-7.

Cleary, S.P. et al., 2010. Cigarette smoking, genetic variants in carcinogen-metabolizing enzymes, and colorectal cancer risk. *American journal of epidemiology*, 172(9), pp.1000-14.

Clifford, N. et al., 2008. The EphA3 receptor is expressed in a subset of rhabdomyosarcoma cell lines and suppresses cell adhesion and migration. *Journal of cellular biochemistry*, 105(5), pp.1250-9.

Cotterman, R. & Knoepfler, P.S., 2009. N-Myc regulates expression of pluripotency genes in neuroblastoma including lif, klf2, klf4, and lin28b. *PLoS ONE*, 4(6), p.e5799.

Crevel, G., Bennett, D. & Cotterill, S., 2008. The human TPR protein TTC4 is a putative Hsp90 co-chaperone which interacts with CDC6 and shows alterations in transformed cells. *PloS one*, 3(3), p.e0001737.

Croucher, D.R. et al., 2010. Cortactin modulates RhoA activation and expression of Cip/Kip cyclin-dependent kinase inhibitors to promote cell cycle progression in 11q13-amplified head and neck squamous cell carcinoma cells. *Molecular and cellular biology*, 30(21), pp.5057-70.

Da, C.L. et al., 2009. Significance and relationship between Yes-associated protein and survivin expression in gastric carcinoma and precancerous lesions. *World journal of gastroenterology : WJG*, 15(32), pp.4055-61.

Dandara, C. et al., 2006. Gene-environment interaction: the role of SULT1A1 and CYP3A5 polymorphisms as risk modifiers for squamous cell carcinoma of the oesophagus. *Carcinogenesis*, 27(4), pp.791-7.

Davies, H. et al., 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer research*, 65(17), pp.7591-5.

De Groot, K.M., 2006. Cancer of the oesophagus -- quo vadis? *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, 96(3), pp.197-198.

Deblois, G. et al., 2009. Genome-wide identification of direct target genes implicates estrogen-related receptor alpha as a determinant of breast cancer heterogeneity. *Cancer research*, 69(15), pp.6149-57.

Dennis, G. et al., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology*, 4(5), p.P3.

Di Benedetto, M. et al., 2006. Mutation analysis of the 8p22 candidate tumor suppressor gene ATIP/MTUS1 in hepatocellular carcinoma. *Molecular and cellular endocrinology*, 252(1-2), pp.207-15.

Ding, J.hua et al., 2010. Alcohol dehydrogenase-2 and aldehyde dehydrogenase-2 genotypes, alcohol drinking and the risk for esophageal cancer in a Chinese population. *Journal of human genetics*, 55(2), pp.97-102.

Domanowski G. “Pathology: Squamous cell carcinoma.”2009. <http://www.emedicine.medscape.com> [Accessed 21 April 2009].

Dragovich, T. & Campen, C., Anti-EGFR-Targeted Therapy for Esophageal and Gastric Cancers: An Evolving Concept. *Journal of oncology*, 2009, p.804108.

Du Plessis, L. et al., 1999. Mapping of novel regions of DNA gain and loss by comparative genomic hybridization in esophageal carcinoma in the Black and Colored populations of South Africa. *Cancer Research*, 59(8), pp.1877-1883.

Du, X.L. et al., 2009. Calreticulin promotes cell motility and enhances resistance to anoikis through STAT3-CTTN-Akt pathway in esophageal squamous cell carcinoma. *Oncogene*, 28(42), pp.3714-22.

Du, Y. et al., 1998. Identification of a novel cortactin SH3 domain-binding protein and its localization to growth cones of cultured neurons. *Molecular and cellular biology*, 18(10), pp.5838-51.

Duhagon, M.A. et al., 2010. Genomic profiling of tumor initiating prostatospheres. *BMC Genomics*, 11, p.324.

Duin, M. van et al., 2007. High-resolution array comparative genomic hybridization of chromosome 8q: evaluation of putative progression markers for gastroesophageal junction adenocarcinomas. *Cytogenetic and genome research*, 118(2-4), pp.130-7.

Duthie, SJ., 2010. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J Inherit Metab Dis.* 34(1), pp. 101-109.

Dyomin, V et al. 1997. BCL8, a novel gene involved in translocations affecting band 15q11-13 in diffuse large-cell lymphoma. *Proceedings of the national academy of science*, 94(11), pp. 5728-32.

Fantozzi, I. et al., 2008. Overexpression of cortactin in head and neck squamous cell carcinomas can be uncoupled from augmented EGF receptor expression. *Acta oncologica (Stockholm, Sweden)*, 47(8), pp.1502-12.

Ferguson, D.O. & Alt, F.W., 2001. DNA double strand break repair and chromosomal translocation: lessons from animal models. *Oncogene*, 20(40), pp.5572-9.

Filardo, E.J., Quinn, J.A. & Sabo, E., 2008. Association of the membrane estrogen receptor, GPR30, with breast tumor metastasis and transactivation of the epidermal growth factor receptor. *Steroids*, 73(9-10), pp.870-873.

Flanagan, J.M. et al., 2004. Mapping of a candidate colorectal cancer tumor-suppressor gene to a 900-kilobase region on the short arm of chromosome 8. *Genes, chromosomes & cancer*, 40(3), pp.247-60.

Fox, B.P. & Kandpal, R.P., 2004. Invasiveness of breast carcinoma cells and transcript profile: Eph receptors and ephrin ligands as molecular markers of potential diagnostic and prognostic application. *Biochemical and biophysical research communications*, 318(4), pp.882-92.

Franchitto, A. et al., 2008. Replication fork stalling in WRN-deficient cells is overcome by prompt activation of a MUS81-dependent pathway. *The Journal of cell biology*, 183(2), pp.241-52.

Freier, K. et al., 2003. Tissue microarray analysis reveals site-specific prevalence of oncogene amplifications in head and neck squamous cell carcinoma. *Cancer research*, 63(6), pp.1179-82.

Freier, K. et al., 2006. Recurrent coamplification of cytoskeleton-associated genes EMS1 and SHANK2 with CCND1 in oral squamous cell carcinoma. *Genes, chromosomes & cancer*, 45(2), pp.118-25.

Fujii, S. et al., 1998. Quantitative analysis of the cyclin expression in human esophageal cancer cell lines. *Journal of experimental & clinical cancer research : CR*, 17(4), pp.491-6.

Fujii, Yo et al., 2006. Fbxw7 contributes to tumor suppression by targeting multiple proteins for ubiquitin-dependent degradation. *Cancer science*, 97(8), pp.729-36.

Fujita, Y. et al., 2003. Chromosome arm 20q gains and other genomic alterations in esophageal squamous cell carcinoma, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. *Hepato-gastroenterology*, 50(54), pp.1857-63.

Fukuchi, M. et al., 2002. Decreased Smad4 expression in the transforming growth factor-beta signaling pathway during progression of esophageal squamous cell carcinoma. *Cancer*, 95(4), pp.737-43.

Fukuoka, T., Hibi, K. & Nakao, A., 2006. Aberrant methylation is frequently observed in advanced esophageal squamous cell carcinoma. *Anticancer research*, 26(5A), pp.3333-5.

Gamieldien, W. et al., 1998. p53 and p16/CDKN2 gene mutations in esophageal tumors from a high-incidence area in South Africa. *International journal of cancer. Journal international du cancer*, 78(5), pp.544-9.

Ganesh, B., Talole, S.D. & Dikshit, R., 2009. Tobacco, alcohol and tea drinking as risk factors for esophageal cancer: A case-control study from Mumbai, India. *Cancer Epidemiol*, 33(6), pp.431-434.

Gelderblom, W C et al., 1991. Toxicity and carcinogenicity of the Fusarium moniliforme metabolite, fumonisin B1, in rats. *Carcinogenesis*, 12(7), pp.1247-51.

Gen, Y. et al., 2010. SOX2 identified as a target gene for the amplification at 3q26 that is frequently detected in esophageal squamous cell carcinoma. *Cancer genetics and cytogenetics*, 202(2), pp.82-93.

Gholamin, M. et al., 2009. Overexpression and interactions of interleukin-10, transforming growth factor beta, and vascular endothelial growth factor in esophageal squamous cell carcinoma. *World journal of surgery*, 33(7), pp.1439-45.

Giaginis, C. et al., 2009. Clinical significance of MCM-2 and MCM-5 expression in colon cancer: association with clinicopathological parameters and tumor proliferative capacity. *Digestive Diseases and Sciences*, 54(2), pp.282-291.

Gibault, L. et al., 2005. Diffuse EGFR staining is associated with reduced overall survival in locally advanced oesophageal squamous cell cancer. *British journal of cancer*, 93(1), pp.107-15.

Gisselsson, D et al., 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp.5357-5362.

Gisselsson, David, 2005. Mitotic instability in cancer: is there method in the madness? *Cell cycle (Georgetown, Tex.)*, 4(8), pp.1007-10.

Glatt, H. et al., 2001. Human cytosolic sulphotransferases: genetics, characteristics, toxicological aspects. *Mutation research*, 482(1-2), pp.27-40.

Gotoh, M. et al., 2007. Epidermal growth factor receptor is a possible predictor of sensitivity to chemoradiotherapy in the primary lesion of esophageal squamous cell carcinoma. *Japanese journal of clinical oncology*, 37(9), pp.652-7.

Grubbs, E.G. et al., 2006. Utilizing quantitative polymerase chain reaction to evaluate prostate stem cell antigen as a tumor marker in pancreatic cancer. *Annals of Surgical Oncology : the Official Journal of the Society of Surgical Oncology*, 13(12), pp.1645-1654.

Gu, Z.D. et al., 2005. Matrix metalloproteinases expression correlates with survival in patients with esophageal squamous cell carcinoma. *The American journal of gastroenterology*, 100(8), pp.1835-43.

Guan, J.ting et al., 2009. [Expression of AIF and caspase-3 in colorectal carcinoma, adenoma, and normal mucosa and its correlation]. *Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery*, 12(4), pp.391-4.

Guo, F.J. et al., 2010. Expression and functional characterization of platelet-derived growth factor receptor-like gene. *World Journal of Gastroenterology : WJG*, 16(12), pp.1465-1472.

Guo, M. et al., 2006. Accumulation of promoter methylation suggests epigenetic progression in squamous cell carcinoma of the esophagus. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12(15), pp.4515-22.

Guo, X.Q. et al., 2007. DNA methylation and loss of protein expression in esophageal squamous cell carcinogenesis of high-risk area. *Journal of experimental & clinical cancer research : CR*, 26(4), pp.587-94.

Guo, Y.M. et al., 2008. Genetic polymorphisms in cytochrome P4502E1, alcohol and aldehyde dehydrogenases and the risk of esophageal squamous cell carcinoma in Gansu Chinese males. *World journal of gastroenterology : WJG*, 14(9), pp.1444-9.

Hallen, L.C. et al., 2007. Antiproliferative activity of the human IFN-alpha-inducible protein IFI44. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research*, 27(8), pp.675-80.

Hamada, M. et al., 1996. Suppressed apoptotic induction in esophageal squamous cell carcinomas expressing extensive p53 protein. *Japanese journal of clinical oncology*, 26(6), pp.398-404.

Han, W. et al., 2006. Shank2 associates with and regulates Na+/H+ exchanger 3. *The Journal of biological chemistry*, 281(3), pp.1461-9.

Hanawa, M. et al., 2006. EGFR protein overexpression and gene amplification in squamous cell carcinomas of the esophagus. *International journal of cancer. Journal international du cancer*, 118(5), pp.1173-80.

Hang, B., 2010. Formation and repair of tobacco carcinogen-derived bulky DNA adducts. *Journal of nucleic acids*. p p.709521.

Hashimoto, N. et al., 2007. Clinicopathologic significance of expression of cyclooxygenase-2 in human esophageal squamous cell carcinoma. *Hepato-gastroenterology*, 54(75), pp.758-60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17591056.

Haverty, P.M. et al., 2009. High-resolution analysis of copy number alterations and associated expression changes in ovarian tumors. *BMC medical genomics*, 2, p.21.

Hayashi, Hideko et al., 2006. Centaurin-alpha1 is a phosphatidylinositol 3-kinase-dependent activator of ERK1/2 mitogen-activated protein kinases. *The Journal of biological chemistry*, 281(3), pp.1332-7.

He, J., Zhang, R.G. & Zhu, D., 1995. [Clinical significance of c-myc gene in esophageal squamous cell carcinoma]. *Zhonghua yi xue za zhi*, 75(2), pp.94-6, 127.

Hellemans, J. et al., 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology*, 8(2), pp.19.

Hendricks, Denver & Parker, M Iqbal, 2002. Oesophageal cancer in Africa. *IUBMB life*, 53(4-5), pp.263-8.

Henrichsen, C et al., 2009. Copy number variants, diseases and gene expression. *Human Molecular Genetics*, 18(R-1), pp R1-8.

Higuchi, R. et al., 1992. Simultaneous amplification and detection of specific DNA sequences. *Bio/technology (Nature Publishing Company)*, 10(4), pp.413-7.

Hirasaki, S. et al., 2007. BAC clones related to prognosis in patients with esophageal squamous carcinoma: an array comparative genomic hybridization study. *The oncologist*, 12(4), pp.406-17.

Hofman, P. et al., 2008. Prognostic significance of cortactin levels in head and neck squamous cell carcinoma: comparison with epidermal growth factor receptor status. *British journal of cancer*, 98(5), pp.956-64.

Horiuchi, A et al. (1999). Possible role of calponin h1 as a tumor suppressor in human uterine leiomyosarcoma.*Journal of the National Cancer Institute*, 91(9), pp. 790-6.

Hosack, D.A. et al., 2003. Identifying biological themes within lists of genes with EASE. *Genome biology*, 4(10), p.R70.

Hoshino, M. et al., 2007. Nuclear expression of phosphorylated EGFR is associated with poor prognosis of patients with esophageal squamous cell carcinoma. *Pathobiology : journal of immunopathology, molecular and cellular biology*, 74(1), pp.15-21.

Howard, P.C. et al., 2001. Fumonisin b1 carcinogenicity in a two-year feeding study using F344 rats and B6C3F1 mice. *Environmental health perspectives*, 109 Suppl , pp.277-82.

Howlin, J., Rosenkvist, J. & Andersson, T., 2008. Research article TNK2 preserves epidermal growth factor receptor expression on the cell surface and enhances migration and invasion of human breast cancer cells. *Breast Cancer Research*, 10(2), pp.1-13.

Hsu, C.G. et al., 2007. Concurrent high expression of human telomerase reverse transcriptase and human nonmetastatic clone 23 in high-grade squamous intraepithelial neoplasia and squamous cell carcinoma of uterine cervix. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, 17(4), pp.851-7.

Hsu, K.F. et al., 2009. Cortactin, fascin, and survivin expression associated with clinicopathological parameters in esophageal squamous cell carcinoma. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 22(5), pp.402-8.

Hsu, N.Y. et al., 2008. Cortactin overexpression in the esophageal squamous cell carcinoma and its involvement in the carcinogenesis. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 21(5), pp.402-8.

Hu, H., Zhang, S. & Zhu, S., 2009. Influence of aspirin and cigarette smoke extract on the expression of cyclin D1 and effects of cell cycle in esophageal squamous cell carcinoma cell line. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 22(4), pp.310-6.

Hu, N. et al., 2006. Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetrix GeneChip Mapping 10 K array. *BMC Genomics*, 7, p.299.

Hu, Xiaolan et al., 2009. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. *Molecular cancer research : MCR*, 7(4), pp.511-22.

Hu, Y.C. et al., 2002. Establishment, characterization, karyotyping, and comparative genomic hybridization analysis of HKESC-2 and HKESC-3: two newly established human esophageal squamous cell carcinoma cell lines. *Cancer genetics and cytogenetics*, 135(2), pp.120-7.

Huang, Jian et al., 2009. ARHI, as a novel suppressor of cell growth and downregulated in human hepatocellular carcinoma, could contribute to hepatocarcinogenesis. *Molecular carcinogenesis*, 48(2), pp.130-40.

Huang, Jun et al., 2009. RAD18 transmits DNA damage signalling to elicit homologous recombination repair. *Nature cell biology*, 11(5), pp.592-603.

Huang, X.P. et al., 2006. Negative implication of C-MYC as an amplification target in esophageal cancer. *Cancer genetics and cytogenetics*, 165(1), pp.20-4.

Hupé, P. et al., 2004. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics (Oxford, England)*, 20(18), pp.3413-22.

Ikushima, H & Miyazono, K. 2010. TGFbeta signalling: a complex web in cancer progression. *Nature Reviews in Cancer*, 10(6), pp. 415-24.

Imoto, I. et al., 2002. Expression of cIAP1, a target for 11q22 amplification, correlates with resistance of cervical cancers to radiotherapy. *Cancer research*, 62(17), pp.4860-6.

Inomata, M. et al., 1998. Amplification and overexpression of cyclin D1 in aggressive human esophageal cancer. *Oncology reports*, 5(1), pp.171-6.

Isaacson, C., 2005. The change of the staple diet of black South Africans from sorghum to maize (corn) is the cause of the epidemic of squamous carcinoma of the oesophagus. *Medical hypotheses*, 64(3), pp.658-60.

Islami, F. et al., 2009. Oesophageal cancer in Golestan Province, a high-incidence area in northern Iran - a review. *European Journal of Cancer*, 45(18), pp.3156-3165.

Jacobs, S. et al., 2007. Genome-Wide , High-Resolution Detection of Copy Number , Loss of Heterozygosity , and Genotypes from Formalin-Fixed , Paraffin-Embedded Tumor Tissue Using Microarrays. *Breast Cancer Research*, (6), pp.2544-2551.

Janssen, J.W. et al., 2000. Concurrent activation of a novel putative transforming gene, myeov, and cyclin D1 in a subset of multiple myeloma cell lines with t(11;14)(q13;q32). *Blood*, 95(8), pp.2691-8.

Janssen, J.W.G. et al., 2002. MYEOV: a candidate gene for DNA amplification events occurring centromeric to CCND1 in breast cancer. *International journal of cancer. Journal international du cancer*, 102(6), pp.608-14.

Janssen, J.W.G. et al., 2002. MYEOV, a gene at 11q13, is coamplified with CCND1, but epigenetically inactivated in a subset of esophageal squamous cell carcinomas. *Journal of human genetics*, 47(9), pp.460-4.

Jaskiewicz, K. & De Groot, K.M., 1994. p53 gene mutants expression, cellular proliferation and differentiation in oesophageal carcinoma and non-cancerous epithelium. *Anticancer research*, 14(1A), pp.137-40.

Jelski, W. et al., 2009. The activity of class I, II, III, and IV alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) in esophageal cancer. *Digestive diseases and sciences*, 54(4), pp.725-30.

Jensen, L.J. et al., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, 37(Database issue), pp.D412-6.

Jensen, S.A. et al., 2008. Prognostic significance of numeric aberrations of genes for thymidylate synthase, thymidine phosphorylase and dihydrofolate reductase in colorectal cancer. *Acta oncologica (Stockholm, Sweden)*, 47(6), pp.1054-61.

Jiang, T. et al., 2004. Bi-directional regulation between tyrosine kinase Etk/BMX and tumor suppressor p53 in response to DNA damage. *The Journal of biological chemistry*, 279(48), pp.50181-9.

Jiang, W et al., 1992. Amplification and expression of the human cyclin D gene in esophageal cancer. *Cancer research*, 52(10), pp.2980-3.

Jiang, W et al., 1993. Altered expression of the cyclin D1 and retinoblastoma genes in human esophageal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 90(19), pp.9026-30.

Jin, Y et al., 2000. Centromere cleavage is a mechanism underlying isochromosome formation in skin and head and neck carcinomas. *Chromosoma*, 109(7), pp.476-81.

Jin, Yuesheng et al., 2004. Cytogenetic and fluorescence in situ hybridization characterization of clonal chromosomal aberrations and CCND1 amplification in esophageal carcinomas. *Cancer genetics and cytogenetics*, 148(1), pp.21-8.

Johansson, B & Mertens, F, 1988. Frequency and distribution of mitomycin C-induced structural chromosome aberrations in lymphocytes from non-Hodgkin lymphoma patients. *Cytogenetics and cell genetics*, 48(2), pp.79-83.

Jones, G.J. et al., 1993. Amplification and expression of the TGF-alpha, EGF receptor and c-myc genes in four human oesophageal squamous cell carcinoma lines. *Bioscience reports*, 13(5), pp.303-12.

Kadowaki, T. et al., 1994. E-cadherin and alpha-catenin expression in human esophageal cancer. *Cancer research*, 54(1), pp.291-6.

Kanao, H. et al., 2005. Overexpression of LAMP3/TSC403/DC-LAMP promotes metastasis in uterine cervical cancer. *Cancer research*, 65(19), pp.8640-5.

Kang, J.U. et al., 2008. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer genetics and cytogenetics*, 182(1), pp.1-11.

Kang, X. et al., 2009. Regulation of the hTERT promoter activity by MSH2, the hnRNPs K and D, and GRHL2 in human oral squamous cell carcinoma cells. *Oncogene*, 28(4), pp.565-74.

Karlgren, M. et al., 2006. Tumor-specific expression of the novel cytochrome P450 enzyme, CYP2W1. *Biochemical and Biophysical Research Communications*, 341(2), pp.451-458. Available at: http://view.ncbi.nlm.nih.gov/pubmed/16426568.

Katoh, Masaru, 2007. Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis. *Stem cell reviews*, 3(1), pp.30-8

Katoh, Masuko & Katoh, Masaru, 2006. Cross-talk of WNT and FGF signaling pathways at GSK3beta to regulate beta-catenin and SNAIL signaling cascades. *Cancer biology & therapy*, 5(9), pp.1059-64.

Kawano, Y & Kypta R. 2003. Secreted antagonists of the Wnt signalling pathway. *Journal of cell science*, 116(pt13), pp. 2627-34.

Keightley, M.C., Sales, K.J. & Jabbour, H.N., 2010. PGF2α-F-prostanoid receptor signalling via ADAMTS1 modulates epithelial cell invasion and endothelial cell function in endometrial cancer. *BMC cancer*, 10, p.488.

Khaitan, D. et al., 2009. Role of KCNMA1 gene in breast cancer invasion and metastasis to brain. *BMC cancer*, 9, p.258.

Khushalani, N., 2008. Cancer of the esophagus and stomach. *Mayo Clinic proceedings. Mayo Clinic*, 83(6), pp.712-22.

Kinoshita T, et al. 2011. Decreased expression and aberrant hypermethylation of the SFRP genes in human gastric cancer. *Hepato-gastroenterology*, 58(107-108), pp. 1051-6.

Kim, M.R. et al., 2007. Pro-apoptotic PUMA and anti-apoptotic phospho-BAD are highly expressed in colorectal carcinomas. *Digestive diseases and sciences*, 52(10), pp.2751-6.

Kim, Y.T. et al., 2009. Aberrant promoter CpG island hypermethylation of the adenomatosis polyposis coli gene can serve as a good prognostic factor by affecting lymph node metastasis in squamous cell carcinoma of the esophagus. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 22(2), pp.143-50.

Kitahara, K. et al., 1996. Expression of cyclin D1, CDK4 and p27KIP1 is associated with the p16MTS1 gene status in human esophageal carcinoma cell lines. *Journal of experimental therapeutics & oncology*, 1(1), pp.7-12.

Kitamura, A. et al., 2001. Reduced Fhit expression occurs in the early stage of esophageal tumorigenesis: no correlation with p53 expression and apoptosis. *Oncology*, 61(3), pp.205-11.

Kleeff, J. et al., 1999. Smad6 suppresses TGF-beta-induced growth inhibition in COLO-357 pancreatic cancer cells and is overexpressed in pancreatic cancer. *Biochemical and biophysical research communications*, 255(2), pp.268-73.

Kleinsasser, N.H. et al., 2006. [Does nicotine add to the carcinogenic strain of tobacco smoke?]. *HNO*, 54(5), pp.369-72, 374-5.

Kongkham, P.N. et al., 2010. The SFRP family of WNT inhibitors function as novel tumor suppressor genes epigenetically silenced in medulloblastoma. *Oncogene*, 29(20), pp.3017-24.

Koshiol, J. et al., 2010. No role for human papillomavirus in esophageal squamous cell carcinoma in China. *International journal of cancer. Journal international du cancer*, 127(1), pp.93-100.

Kreunin, P. et al., 2007. Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *Journal of proteome research*, 6(7), pp.2631-9.

Kuo, K.T. et al., 2009. Prognostic role of PGE2 receptor EP2 in esophageal squamous cell carcinoma. *Annals of surgical oncology*, 16(2), pp.352-60.

Kuroki, T. et al., 2003. Allele loss and promoter hypermethylation of VHL, RAR-beta, RASSF1A, and FHIT tumor suppressor genes on chromosome 3p in esophageal squamous cell carcinoma. *Cancer research*, 63(13), pp.3724-8.

Lam, A.K., 2000. Molecular biology of esophageal squamous cell carcinoma. *Critical reviews in oncology/hematology*, 33(2), pp.71-90.

Lam, K Y et al., 1997. Pre-operative chemotherapy for squamous cell carcinoma of the oesophagus: do histological assessment and p53 overexpression predict chemo-responsiveness? *European journal of cancer (Oxford, England : 1990)*, 33(8), pp.1221-5.

Lam, K Y et al., 1995. Expression of p53 in oesophageal squamous cell carcinoma in Hong Kong Chinese. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 21(3), pp.242-7.

Lam, K Y et al., 1997. Prevalence and predictive value of p53 mutation in patients with oesophageal squamous cell carcinomas: a prospective clinico-pathological study and

survival analysis of 70 patients. *International journal of cancer. Journal international du cancer*, 74(2), pp.212-9.

Lawler, K. et al., 2009. Shear stress induces internalization of E-cadherin and invasiveness in metastatic oesophageal cancer cells by a Src-dependent pathway. *Cancer science*, 100(6), pp.1082-7.

Lee, D. et al., 2010. HMGB2 stabilizes p53 by interfering with E6/E6AP-mediated p53 degradation in human papillomavirus-positive HeLa cells. *Cancer letters*, 292(1), pp.125-32.

Lee, E.J. et al., 2008. CpG island hypermethylation of E-cadherin (CDH1) and integrin alpha4 is associated with recurrence of early stage esophageal squamous cell carcinoma. *International journal of cancer. Journal international du cancer*, 123(9), pp.2073-9.

Lee, K.H. et al., 2009. Lack of death receptor 4 (DR4) expression through gene promoter methylation in gastric carcinoma. *Langenbeck's archives of surgery / Deutsche Gesellschaft für Chirurgie*, 394(4), pp.661-70.

Lee, Y.J. et al., 2010. Variable inhibition of thrombospondin 1 against liver and lung metastases through differential activation of metalloproteinase ADAMTS1. *Cancer research*, 70(3), pp.948-56.

Li, D.-M. et al., 2005. Expression of human chorionic gonadotropin, CD44v6 and CD44v4/5 in esophageal squamous cell carcinoma. *World journal of gastroenterology : WJG*, 11(47), pp.7401-4.

Li, D.P. et al., 2008. Genetic polymorphisms of alcohol metabolising enzymes: their role in susceptibility to oesophageal cancer. *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 46(3), pp.323-8.

Li, D., Dandara, C. & Parker, M Iqbal, 2005. Association of cytochrome P450 2E1 genetic polymorphisms with squamous cell carcinoma of the oesophagus. *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 43(4), pp.370-5.

Li, Y.Z., Zhao, P. & Han, W.-D., 2009. Clinicopathological significance of LRP16 protein in 336 gastric carcinoma patients. *World journal of gastroenterology : WJG*, 15(38), pp.4833-7.

Lieber, M.R. et al., 2003. Mechanism and regulation of human non-homologous DNA end-joining. *Nature reviews. Molecular cell biology*, 4(9), pp.712-20.

Lin, De-Chen et al., 2010. [Expression of cell cycle related proteins cyclin D1, p53 and p21WAF1/Cip1 in esophageal squamous cell carcinoma]. *Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji*, 32(5), pp.455-60.

Liu, C., et al. 2005. Modulation of gene expression in precancerous rat esophagus by dietary zinc deficit and replenishment. *Cancer research*, 65(17), pp. 7790-9.

Liu, C.X. et al., 2007. gamma-Secretase-mediated release of the low density lipoprotein receptor-related protein 1B intracellular domain suppresses anchorage-independent growth of neuroglioma cells. *The Journal of biological chemistry*, 282(10), pp.7504-11.

Liu, F.J., Barchowsky, A. & Opresko, P.L., 2010. The Werner syndrome protein suppresses telomeric instability caused by chromium (VI) induced DNA replication stress. *PloS one*, 5(6), p.e11152.

Liu, N. et al., 2010. The Fbw7/human CDC4 tumor suppressor targets proproliferative factor KLF5 for ubiquitination and degradation through multiple phosphodegron motifs. *The Journal of biological chemistry*, 285(24), pp.18858-67.

Liu, P. et al., 2009. Expression of VEGF-C correlates with a poor prognosis based on analysis of prognostic factors in 73 patients with esophageal squamous cell carcinomas. *Japanese journal of clinical oncology*, 39(10), pp.644-50.

Liu, R., Yin, L.-hong & Pu, Y.-pu, 2007. Association of combined CYP2E1 gene polymorphism with the risk for esophageal squamous cell carcinoma in Huai'an population, China. *Chinese medical journal*, 120(20), pp.1797-802.

Liu, S. et al., 2009. Expression and functional roles of Smad1 and BMPR-IB in glioma development. *Cancer investigation*, 27(7), pp.734-40.

Liu, W.K et al., 2005. The relationship between HPV16 and expression of CD44v6, nm23H1 in esophageal squamous cell carcinoma. *Archives of virology*, 150(5), pp.991-1001. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15645378.

Liu, Wen-Kang et al., 2009. The relationship between cyclooxygenase-2, CD44v6, and nm23H1 in esophageal squamous cell carcinoma. *Onkologie*, 32(10), pp.574-8.

Liu, W. et al., 2008. Homozygous deletions and recurrent amplifications implicate new genes involved in prostate cancer. *Neoplasia (New York, N.Y.)*, 10(8), pp.897-907.

Lo, A.W.I. et al., 2002. DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. *Neoplasia (New York, N.Y.)*, 4(6), pp.531-8.

Lo, P.H.Y. et al., 2010. Extracellular protease ADAMTS9 suppresses esophageal and nasopharyngeal carcinoma tumor formation by inhibiting angiogenesis. *Cancer research*, 70(13), pp.5567-76.

Lu, X.M. et al., 2005. Relationship between genetic polymorphisms of metabolizing enzymes CYP2E1, GSTM1 and Kazakh's esophageal squamous cell cancer in Xinjiang, China. *World journal of gastroenterology : WJG*, 11(24), pp.3651-4.

Lukyanova, N.Y. et al., 2009. Molecular profile and cell cycle in MCF-7 cells resistant to cisplatin and doxorubicin. *Experimental oncology*, 31(2), pp.87-91.

Lum, L. & Beachy, P.A., 2004. The Hedgehog response network: sensors, switches, and routers. *Science (New York, N.Y.)*, 304(5678), pp.1755-9.

Luo, Man-Li et al., 2006. Amplification and overexpression of CTTN (EMS1) contribute to the metastasis of esophageal squamous cell carcinoma by promoting cell migration and anoikis resistance. *Cancer research*, 66(24), pp.11690-9.

Mangone, F, et al. (2010). Smad2 and Smad6 as predictors of overall survival in oral squamous cell carcinoma patients. *Molecular cancer*, 9, pp. 106.

Mao, F.J. et al., 2010. The Human WRN and BLM RecQ Helicases Differentially Regulate Cell Proliferation and Survival after Chemotherapeutic DNA Damage. *Cancer research*, 70(16), pp.6548-55.

Marasas, W.F., Rensburg, S.J. van & Mirocha, C.J., 1979. Incidence of Fusarium species and the mycotoxins, deoxynivalenol and zearalenone, in corn produced in esophageal cancer areas in Transkei. *Journal of agricultural and food chemistry*, 27(5), pp.1108-12.

Maréchal, R. et al., 2010. Deoxycitidine kinase is associated with prolonged survival after adjuvant gemcitabine for resected pancreatic adenocarcinoma. *Cancer*, 116(22), pp.5200-6.

Matsha, T. et al., 2007. Expression of p53 and its homolog, p73, in HPV DNA positive oesophageal squamous cell carcinomas. *Virology*, 369(1), pp.182-90.

Matsha, T. et al., 2002. Human papillomavirus associated with oesophageal cancer. *Journal of clinical pathology*, 55(8), pp.587-90.

Matsha, T. et al., 2006. Self-induced vomiting -- risk for oesophageal cancer? *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 96(3), pp.209-12.

McCabe, M & Dlamini Z., 2005. The molecular mechanisms of oesophageal cancer. *International Immunopharmacology*, 5 (7-8), pp.1113-30.

McClintock, B., 1941. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics*, 26(2), pp.234-82.

McKay, J.D. et al., 2008. Lung cancer susceptibility locus at 5p15.33. *Nature genetics*, 40(12), pp.1404-6.

Menin, C. et al., 2000. Anomalous transcripts and allelic deletions of the FHIT gene in human esophageal cancer. *Cancer genetics and cytogenetics*, 119(1), pp.56-61.

Mestre-Escorihuela, C. et al., 2007. Homozygous deletions localize novel tumor suppressor genes in B-cell lymphomas. *Blood*, 109(1), pp.271-80.

Metzger, R. et al., 2004. Molecular biology of esophageal cancer. *Onkologie*, 27(2), pp.200-6.

Miller, S.E. & Veale, R.B., 2001. Environmental modulation of alpha(v), alpha(2) and beta(1) integrin subunit expression in human oesophageal squamous cell carcinomas. *Cell biology international*, 25(1), pp.61-9.

Millonig, G. et al., 2011. Ethanol-mediated carcinogenesis in the human esophagus implicates CYP2E1 induction and the generation of carcinogenic DNA-lesions. *International journal of cancer. Journal international du cancer*, 128(3), pp.533-40.

Mitelman, Felix, Johansson, Bertil & Mertens, Fredrik, 2007. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer*, 7(4), pp.233-45.

Miyazaki, S. et al., 1992. Analysis of c-myc oncogene in human esophageal carcinoma: immunohistochemistry, in situ hybridization and northern and Southern blot studies. *Anticancer research*, 12(5), pp.1747-55.

Montesano, R., Hollstein, M. & Hainaut, P., 1996. Genetic alterations in esophageal cancer and their relevance to etiology and pathogenesis: a review. *International journal of cancer. Journal international du cancer*, 69(3), pp.225-35.

Mori, M et al., 2000. Altered expression of Fhit in carcinoma and precarcinomatous lesions of the esophagus. *Cancer research*, 60(5), pp.1177-82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10728669.

Morita, M. et al., 2006. Expression of FHIT in esophageal epithelium and carcinoma: reference to drinking, smoking and multicentric carcinogenesis. *Anticancer research*, 26(3B), pp.2243-8.

Moriwaki, Y. et al., 2001. Mycobacterium bovis Bacillus Calmette-Guerin and its cell wall complex induce a novel lysosomal membrane protein, SIMPLE, that bridges the missing link between lipopolysaccharide and p53-inducible gene, LITAF(PIG7), and estrogen-inducible gene, EET-1. *The Journal of biological chemistry*, 276(25), pp.23065-76.

Moss, A.C. et al., 2006. ETV4 and Myeov knockdown impairs colon cancer cell line proliferation and invasion. *Biochemical and biophysical research communications*, 345(1), pp.216-21.

Muro, K. et al., 1996. Association of p53 protein expression with responses and survival of patients with locally advanced esophageal carcinoma treated with chemoradiotherapy. *Japanese journal of clinical oncology*, 26(2), pp.65-9.

Myllykangas, S., Böhling, T. & Knuutila, S., 2007. Specificity, selection and significance of gene amplifications in cancer. *Seminars in cancer biology*, 17(1), pp.42-55.

Nagayama, K. et al., 2007. Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. *Genes, chromosomes & cancer*, 46(11), pp.1000-10.

Naidoo, R et al., 2005. Aberrations in the mismatch repair genes and the clinical impact on oesophageal squamous carcinomas from a high incidence area in South Africa. *Journal of clinical pathology*, 58(3), pp.281-4.

Nair, K S, Naidoo, R & Chetty, R, 2005. Expression of cell adhesion molecules in oesophageal carcinoma and its prognostic value. *Journal of clinical pathology*, 58(4), pp.343-51.

Nair, Kriebashne S, Naidoo, Richard & Chetty, Runjan, 2006. Microsatellite analysis of the APC gene and immunoexpression of E-cadherin, catenin, and tubulin in esophageal squamous cell carcinoma. *Human pathology*, 37(2), pp.125-34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16426911.

Nakagawa, H et al., 1995. Human cyclin D1 oncogene and esophageal squamous cell carcinoma. *Cancer*, 76(4), pp.541-9.

Nakakuki, K. et al., 2002. Novel targets for the 18p11.3 amplification frequently observed in esophageal squamous cell carcinomas. *Carcinogenesis*, 23(1), pp.19-24.

Nancarrow, Derek J et al., 2008. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. *Cancer research*, 68(11), pp.4163-72.

Nasri, S. et al., 2010. Oligonucleotide array outperforms SNP array on formalin-fixed paraffin-embedded clinical samples. *Cancer genetics and cytogenetics*, 198(1), pp.1-6.

Noguchi, T. et al., 2003. FHIT expression and hypermethylation in esophageal squamous cell carcinoma. *International journal of molecular medicine*, 11(4), pp.441-7.

Nomura, D.K. et al., 2010. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell*, 140(1), pp.49-61.

Novak, A. & Dedhar, S., 1999. Signaling through beta-catenin and Lef/Tcf. *Cellular and molecular life sciences : CMLS*, 56(5-6), pp.523-37.

Nozoe, T. et al., 2004. Significance of immunohistochemical over-expression of CD44v6 as an indicator of malignant potential in esophageal squamous cell carcinoma. *Journal of cancer research and clinical oncology*, 130(6), pp.334-8.

Okazawa, T. et al.,2008. Expression of vascular endothelial growth factor C is a prognostic indicator in esophageal cancer. *Hepato-gastroenterology*, 55(86-87), pp.1503-8.

Ono, Kanae et al., 2003. Down-regulation of FEZ1/LZTS1 gene with frequent loss of heterozygosity in oral squamous cell carcinomas. *International journal of oncology*, 23(2), pp.297-302.

Or, Y.Y. et al., 2005. Characterization of chromosome 3q and 12q amplicons in nasopharyngeal carcinoma cell lines. *International journal of oncology*, 26(1), pp.49-56.

Osawa, H, et al. 2004. Prognostic value of the expression of Smad6 and Smad7, as inhibitory Smads of the TGF-beta superfamily, in esophageal squamous cell carcinoma. *Anticancer research*, 24(6), pp. 2703-9.

Pacella-Norman, R. et al., 2002. Risk factors for oesophageal, lung, oral and laryngeal cancers in black South Africans. *British journal of cancer*, 86(11), pp.1751-6.

Park, H.L. et al., 2008. DCC promoter hypermethylation in esophageal squamous cell carcinoma. *International journal of cancer. Journal international du cancer*, 122(11), pp.2498-502.

Parthasarathy, S. et al., 2010. Aberrant Expression of EGFR and Its interaction with PKCδ in Inflammation Associated Neoplastic Transformation of Human Esophageal Epithelium in High Risk Populations. *Journal of gastroenterology and hepatology*.

Pasca di Magliano, M. & Hebrok, M., 2003. Hedgehog signalling in cancer formation and maintenance. *Nature reviews. Cancer*, 3(12), pp.903-11.

Pasquale, E. 2010. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Pathology*, 10(3), pp. 165-180.

Pattyn, F. et al., 2003. RTPrimerDB: the real-time PCR primer and probe database. *Nucleic acids research*, 31(1), pp.122-3.

Peng, G. et al., 2009. BRIT1/MCPH1 links chromatin remodelling to DNA damage response. *Nature cell biology*, 11(7), pp.865-72.

Peng, H. et al., 2009. Expression and significance of adenomatous polyposis coli, beta-catenin, E-cadherin and cyclin D1 in esophageal squamous cell carcinoma assessed by tissue microarray. *Ai zheng = Aizheng = Chinese journal of cancer*, 28(1), pp.38-41.

Peters, C.J. et al., 2010. A 4-Gene Signature Predicts Survival of Patients With Resected Adenocarcinoma of the Esophagus, Junction, and Gastric Cardia. *Gastroenterology*.

Picard, F. et al., 2011. Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics (Oxford, England)*.

Pichierri, P., Rosselli, F. & Franchitto, A., 2003. Werner's syndrome protein is phosphorylated in an ATR/ATM-dependent manner following replication arrest and DNA damage induced during the S phase of the cell cycle. *Oncogene*, 22(10), pp.1491-500.

Pinkel, D. & Albertson, D.G., 2005. Comparative genomic hybridization. *Annual review of genomics and human genetics*, 6, pp.331-54.

Pirzio, L.M. et al., 2008. Werner syndrome helicase activity is essential in maintaining fragile site stability. *The Journal of cell biology*, 180(2), pp.305-14.

Puca, R. et al., 2009. HIPK2 modulates p53 activity towards pro-apoptotic transcription. *Molecular cancer*, 8, p.85.

Putnam, C.D., Hayes, T.K. & Kolodner, R.D., 2010. Post-replication repair suppresses duplication-mediated genome instability. *PLoS genetics*, 6, p.e1000933.

Qin, J.M. et al., 2008. Interaction of methylenetetrahydrofolate reductase C677T, cytochrome P4502E1 polymorphism and environment factors in esophageal cancer in Kazakh population. *World journal of gastroenterology : WJG*, 14(45), pp.6986-92.

Qin, Y.R. et al., 2008. Single-nucleotide polymorphism-mass array reveals commonly deleted regions at 3p22 and 3p14.2 associate with poor clinical outcome in esophageal squamous cell carcinoma. *International journal of cancer. Journal international du cancer*, 123(4), pp.826-30.

Qin, Y.R. et al., 2008. Comparative genomic hybridization analysis of genetic aberrations associated with development of esophageal squamous cell carcinoma in Henan, China. *World journal of gastroenterology : WJG*, 14(12), pp.1828-35.

Qin, Y.R. et al., 2005. [Comparative genomic hybridization: the profile of chromosomal imbalances in esophageal squamous cell carcinoma]. *Zhonghua bing li xue za zhi Chinese journal of pathology*, 34(2), pp.80-3.

Quintela-Fandino, M. et al., 2010. HUNK suppresses metastasis of basal type breast cancers by disrupting the interaction between PP2A and cofilin-1. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), pp.2622-7.

Raff, A.B., Gray, A. & Kast, W.M., 2009. Prostate stem cell antigen: a prospective therapeutic and diagnostic target. *Cancer letters*, 277(2), pp.126-32.

Rajkumar, T. & Yamuna, M., 2008. Multiple pathways are involved in drug resistance to doxorubicin in an osteosarcoma cell line. *Anti-cancer drugs*, 19(3), pp.257-65.

Rattner, A, et al. 1997. A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *Proceedings of the national academy of science*, 94(7), pp. 2859-63.

Reich, M. et al., 2006. GenePattern 2.0. *Nature genetics*, 38(5), pp.500-1.

Rensburg, S.J. van, 1987. Oesophageal cancer risk factors common to endemic regions. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, Suppl, pp.9-11.

Roncalli, M. et al., 1998. Cell cycle-related gene abnormalities and product expression in esophageal carcinoma. *Laboratory investigation; a journal of technical methods and pathology*, 78(9), pp.1049-57.

Rosemary Siafakas, A. & Richardson, D.R., 2009. Growth arrest and DNA damage-45 alpha (GADD45alpha). *The international journal of biochemistry & cell biology*, 41(5), pp.986-9.

Roye, G.D. et al., 1996. CD44 expression in dysplastic epithelium and squamous-cell carcinoma of the esophagus. *International journal of cancer. Journal international du cancer*, 69(4), pp.254-8.

Rubie, C. et al., 2005. Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Molecular and cellular probes*, 19(2), pp.101-9.

Sahin, M., Sahin, E. & Gümüslü, S., 2009. Cyclooxygenase-2 in cancer and angiogenesis. *Angiology*, 60(2), pp.242-53.

Sakai, N. et al., 2010. Study of abnormal chromosome regions in esophageal squamous cell carcinoma by comparative genomic hybridization: relationship of lymph node metastasis and distant metastasis to selected abnormal regions. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E*, 23(5), pp.415-21.

Salam, I. et al., 2009. Aberrant promoter methylation and reduced expression of p16 gene in esophageal squamous cell carcinoma from Kashmir valley: a high-risk area. *Molecular and cellular biochemistry*, 332(1-2), pp.51-8.

Sammon, A & Morgan, A. 2002. Dietary fat and salivary prostaglandin E2. *Prostaglandins and other lipid mediators*, 67(2), pp. 137-41.

Sammon, A.M. & Iputo, J.E., 2006. Maize meal predisposes to endemic squamous cancer of the oesophagus in Africa: breakdown of esterified linoleic acid to the free form in stored meal leads to increased intragastric PGE2 production and a low-acid reflux. *Medical hypotheses*, 67(6), pp.1431-6.

Sato, M et al., 1996. Increased expression of integrins by heparin-binding EGF like growth factor in human esophageal cancer cells. *Cancer letters*, 102(1-2), pp.183-91.

Selvarajah, S. et al., 2006. The breakage-fusion-bridge (BFB) cycle as a mechanism for generating genetic heterogeneity in osteosarcoma. *Chromosoma*, 115(6), pp.459-67.

Semba, Shuho & Huebner, Kay, 2006. Protein expression profiling identifies cyclophilin A as a molecular target in Fhit-mediated tumor suppression. *Molecular cancer research : MCR*, 4(8), pp.529-38.

Seta, T. et al., 1998. Expression of p53 and p21WAF1/CIP1 proteins in gastric and esophageal cancers: comparison with mutations of the p53 gene. *Digestive diseases and sciences*, 43(2), pp.279-89.

Sexton, T, et al. 2007. The role of transcription factories in large-scale structure and dynamics of interphase chromatin. *Seminars in cell and developmental biology*, 18(5), pp. 691-7.

Shamma, A. et al., 1998. Effect of cyclin D1 and associated proteins on proliferation of esophageal squamous cell carcinoma. *International journal of oncology*, 13(3), pp.455-60.

Shen, Z.Y. et al., 2001. The genetic events of HPV-immortalized esophageal epithelium cells. *International journal of molecular medicine*, 8(5), pp.537-42.

Shephard, G S et al., 2007. Exposure assessment for fumonisins in the former Transkei region of South Africa. *Food additives and contaminants*, 24(6), pp.621-9.

Shephard, Gordon S et al., 2005. Fumonisin mycotoxins in traditional Xhosa maize beer in South Africa. *Journal of agricultural and food chemistry*, 53(24), pp.9634-7.

Sheyn, I. et al., 1997. Amplification and expression of the cyclin D1 gene in anal and esophageal squamous cell carcinomas. *Human pathology*, 28(3), pp.270-6.

Shi, Z. et al., 2002. [NOEY2 gene mRNA expression in breast cancer tissue and its relation to clinicopathological parameters]. *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 24(5), pp.475-8.

Shimada, K. et al., 2009. A novel human AlkB homologue, ALKBH8, contributes to human bladder cancer progression. *Cancer research*, 69(7), pp.3157-64.

Shimada, Y et al., 1999. Prognostic factors of oesophageal squamous cell carcinoma from the perspective of molecular biology. *British journal of cancer*, 80(8), pp.1281-8.

Shimada, Y et al., 2000. Loss of fragile histidine triad gene expression is associated with progression of esophageal squamous cell carcinoma, but not with the patient's prognosis and smoking history. *Cancer*, 89(1), pp.5-11.

Shinozaki, H. et al., 1996. Cyclin D1 amplification as a new predictive classification for squamous cell carcinoma of the esophagus, adding gene information. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 2(7), pp.1155-61.

Shitara, K et al., 2010. Heavy smoking history interacts with chemoradiotherapy for esophageal cancer prognosis: a retrospective study. *Cancer Science,* 101 (4), pp. 1001-6.

Simsek, D. & Jasin, M., 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nature structural & molecular biology*, 17(4), pp.410-6.

Sitas, Freddy et al., 2007. The relationship between anti-HPV-16 IgG seropositivity and cancer of the cervix, anogenital organs, oral cavity and pharynx, oesophagus and prostate in a black South African population. *Infectious agents and cancer*, 2, p.6.

Smith, A.J. et al., 2009. Expression of the Bcl-2 protein BAD promotes prostate cancer growth. *PloS one*, 4(7), p.e6224.

Somdyala, N.I. et al., 2010. Cancer incidence in a rural population of South Africa, 1998-2002. *International journal of cancer. Journal international du cancer*, 127(10), pp.2420-9.

Sonoda, E. et al., 1998. Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *The EMBO journal*, 17(2), pp.598-608.

Sozzi, G. et al., 1997. Association between cigarette smoking and FHIT gene alterations in lung cancer. *Cancer research*, 57(11), pp.2121-3.

Stein, C.K. et al., 2002. Direct correlation between FRA3B expression and cigarette smoking. *Genes, chromosomes & cancer*, 34(3), pp.333-40.

Stephens, P. et al., 2005. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nature genetics*, 37(6), pp.590-2.

Su, D.M. et al., 2009. Two types of human malignant melanoma cell lines revealed by expression patterns of mitochondrial and survival-apoptosis genes: implications for malignant melanoma therapy. *Molecular cancer therapeutics*.

Sugahara, K, et al. 2011. Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma. *International journal of oncology*, 39(4), pp. 761-9.

Sugimoto, T. et al., 2007. Integrated analysis of expression and genome alteration reveals putative amplified target genes in esophageal cancer. *Oncology reports*, 18(2), pp.465-72.

Sun, G. et al., 2007. Fumonisin B1 contamination of home-grown corn in high-risk areas for esophageal and liver cancer in China. *Food additives and contaminants*, 24(2), pp.181-5.

Surawska, H., Ma, P.C. & Salgia, R., 2004. The role of ephrins and Eph receptors in cancer. *Cytokine & growth factor reviews*, 15(6), pp.419-33.

Suzuki, H., Fujioka, Y. & Nagashima, K., 1998. Cyclin D1 gene amplification and p16 gene deletion in patients with esophageal carcinosarcoma. *Diagnostic molecular pathology : the American journal of surgical pathology, part B*, 7(5), pp.253-9.

Suzuki, K. et al., 2010. Requirement of ATM-dependent pathway for the repair of a subset of DNA double strand breaks created by restriction endonucleases. *Genome integrity*, 1(1), p.4.

Szilvás, A. et al., 1998. [Serum acid alpha-1-glycoprotein (APG) levels in gastrointestinal tumors]. *Orvosi hetilap*, 139(37), pp.2199-202.

Tada, K. et al., 2000. Cytogenetic analysis of esophageal squamous cell carcinoma cell lines by comparative genomic hybridization: relationship of cytogenetic aberrations to in vitro cell growth. *Cancer genetics and cytogenetics*, 117(2), pp.108-12.

Taghavi, N. et al., 2010. p16INK4a hypermethylation and p53, p16 and MDM2 protein expression in esophageal squamous cell carcinoma. *BMC cancer*, 10, p.138.

Takata, M. et al., 1998. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO journal*, 17(18), pp.5497-508.

Takayama, N. et al., 2003. Relationship between the expression of adhesion molecules in primary esophageal squamous cell carcinoma and metastatic lymph nodes. *Anticancer research*, 23(6a), pp.4435-42.

Takeuchi, H. et al., 1997. Altered p16/MTS1/CDKN2 and cyclin D1/PRAD-1 gene expression is associated with the prognosis of squamous cell carcinoma of the esophagus. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 3(12 Pt 1), pp.2229-36.

Tan, W. et al., 2000. Impact of genetic polymorphisms in cytochrome P450 2E1 and glutathione S-transferases M1, T1, and P1 on susceptibility to esophageal cancer among high-risk individuals in China. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 9(6), pp.551-6.

Tanaka, F. et al., 2010. Strong interaction between the effects of alcohol consumption and smoking on oesophageal squamous cell carcinoma among individuals with ADH1B and/or ALDH2 risk alleles. *Gut*, 59(11), pp.1457-64.

Tanaka, H. et al., 1998. Methylation of the 5' CpG island of the FHIT gene is closely associated with transcriptional inactivation in esophageal squamous cell carcinomas. *Cancer research*, 58(15), pp.3429-34.

Thomassen, M., Tan, Q. & Kruse, T.A., 2009. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast cancer research and treatment*, 113(2), pp.239-49.

Thompson, E.R. et al., 2005. Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue. *Human mutation*, 26(4), pp.384-9.

Tian, M. et al., 2008. Proteomic analysis identifies MMP-9, DJ-1 and A1BG as overexpressed proteins in pancreatic juice from pancreatic ductal adenocarcinoma patients. *BMC cancer*, 8, p.241.

Toma, M.I. et al., 2008. Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array. *Neoplasia (New York, N.Y.)*, 10(7), pp.634-42.

Tsuneoka, M. et al., 2004. Mina53 as a potential prognostic factor for esophageal squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 10(21), pp.7347-56.

Uehara, H. et al., 2005. Deficiency of hMLH1 and hMSH2 expression is a poor prognostic factor in esophageal squamous cell carcinoma. *Journal of surgical oncology*, 92(2), pp.109-15.

Umar, S.B. & Fleischer, D.E., 2008. Esophageal cancer: epidemiology, pathogenesis and prevention. *Nature clinical practice. Gastroenterology & hepatology*, 5(9), pp.517-26.

Unoki, M. & Nakamura, Y., 2001. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene*, 20(33), pp.4457-65.

Vandesompele, J. et al., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7), p.RESEARCH0034.

Vasavi, M et al., 2010. Microsatellite instability analysis and its correlation with hMLH1 repair gene hypermethylation status in esophageal pathologies including cancers. *Cancer biomarkers : section A of Disease markers*, 7(1), pp.1-10.

Vasavi, Mohan et al., 2006. DNA methylation in esophageal diseases including cancer: special reference to hMLH1 gene promoter status. *Tumori*, 92(2), pp.155-62.

Veale R, 1989. Increased single class low-affinity egf receptors expressed by human oesophageal squamous carcinoma cell lines. *S Afr Med J Sci*, 85, pp.375-379.

Vos, M. et al., 2003. Polymorphisms and mutations found in the regions flanking exons 5 to 8 of the TP53 gene in a population at high risk for esophageal cancer in South Africa. *Cancer genetics and cytogenetics*, 140(1), pp.23-30.

Wagata, T. et al., 1993. Loss of 17p, mutation of the p53 gene, and overexpression of p53 protein in esophageal squamous cell carcinomas. *Cancer research*, 53(4), pp.846-50.

Wang, A.-H. et al., 2004. Genetic susceptibility and environmental factors of esophageal cancer in Xi'an. *World journal of gastroenterology : WJG*, 10(7), pp.940-4.

Wang, B. et al., 2006. A growth-related oncogene/CXC chemokine receptor 2 autocrine loop contributes to cellular proliferation in esophageal cancer. *Cancer research*, 66(6), pp.3071-7.

Wang, B. et al., 2009. A key role for early growth response-1 and nuclear factor-kappaB in mediating and maintaining GRO/CXCR2 proliferative signaling in esophageal cancer. *Molecular cancer research : MCR*, 7(5), pp.755-64.

Wang, Jianming et al., 2008. Aberrant DNA methylation of P16, MGMT, and hMLH1 genes in combination with MTHFR C677T genetic polymorphism in esophageal squamous cell carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 17(1), pp.118-25.

Wang, M. et al., 1997. Inhibition of tumor growth and metastasis of human breast cancer cells transfected with tissue inhibitor of metalloproteinase 4. *Oncogene*, 14(23), pp.2767-74.

Wang, P. et al., 2003. The POU homeodomain protein OCT3 as a potential transcriptional activator for fibroblast growth factor-4 (FGF-4) in human breast cancer cells. *Biochemical Journal*, 375(1), p.199.

Wang, Wei, Xue, L. & Wang, Pengju, 2010. Prognostic value of beta-catenin, c-myc, and cyclin D1 expressions in patients with esophageal squamous cell carcinoma. *Medical oncology (Northwood, London, England)*.

Wang, Xueqian et al., 2010. Detection of HPV DNA in esophageal cancer specimens from different regions and ethnic groups: a descriptive study. *BMC cancer*, 10, p.19.

Wang, Yong et al., 2008. [Expression of minichromosome maintenance protein 2 in colonic adenocarcinoma, adenoma and normal colonic mucosa and its clinical significance]. *Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery*, 11(5), pp.465-8.

Wang, Yuker et al., 2005. Allele quantification using molecular inversion probes (MIP). *Nucleic acids research*, 33(21), p.e183.

Weinshilboum, R.M. et al., 1997. Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 11(1), pp.3-14.

Weir, B.A. et al., 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature*, 450(7171), pp.893-8.

Weiske, J., Albring, K.F. & Huber, O., 2007. The tumor suppressor Fhit acts as a repressor of beta-catenin transcriptional activity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), pp.20344-9.

Weterman, M.A. et al., 1996. Fine mapping of the 1q21 breakpoint of the papillary renal cell carcinoma-associated (X;1) translocation. *Human genetics*, 98(1), pp.16-21.

Willem, P., Brown, J. & Schouten, J., 2006. A novel approach to simultaneously scan genes at fragile sites. *BMC cancer*, 6, p.205.

Wood, H.M. et al., 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic acids research*, 38(14), p.e151.

Wood, L.D. et al., 2006. Somatic mutations of GUCY2F, EPHA3, and NTRK3 in human cancers. *Human mutation*, 27(10), pp.1060-1.

Wu, M.T. et al., 2003. SULT1A1 polymorphism and esophageal cancer in males. *International journal of cancer. Journal international du cancer*, 103(1), pp.101-4.

Wu, Xianglin et al., 2009. Microcephalin regulates BRCA2 and Rad51-associated DNA double-strand break repair. *Cancer research*, 69(13), pp.5531-6.

Wu, Xiaocheng et al., 2007. Incidence of esophageal and gastric cancers among Hispanics, non-Hispanic whites and non-Hispanic blacks in the United States: subsite and histology differences. *Cancer causes & control : CCC*, 18(6), pp.585-93.

Wu, Y.P. et al., 2006. Identification of chromosome aberrations in esophageal cancer cell line KYSE180 by multicolor fluorescence in situ hybridization. *Cancer genetics and cytogenetics*, 170(2), pp.102-7.

Xiao, S. et al., 1991. Cytogenetic studies of five primary esophageal cancers. *Cancer genetics and cytogenetics*, 55(2), pp.197-205.

Yang, G. et al., 2004. Cystathionine gamma-lyase overexpression inhibits cell proliferation via a H2S-dependent modulation of ERK1/2 phosphorylation and p21Cip/WAK-1. *The Journal of biological chemistry*, 279(47), pp.49199-205.

Yang, S.J. et al., 2010. Relationship between genetic polymorphisms of ALDH2 and ADH1B and esophageal cancer risk: a meta-analysis. *World journal of gastroenterology : WJG*, 16(33), pp.4210-20.

Yang, Y. et al., 2008. Chromosome analysis of esophageal squamous cell carcinoma cell line KYSE 410-4 by repetitive multicolor fluorescence in situ hybridization. *Journal of Genetics and Genomics*, 35(1), pp.11-16.

Yao, P.F. et al., 2006. Evidence of human papilloma virus infection and its epidemiology in esophageal squamous cell carcinoma. *World journal of gastroenterology : WJG*, 12(9), pp.1352-5.

Yen, C.C. et al., 2001. Comparative genomic hybridization of esophageal squamous cell carcinoma: correlations between chromosomal aberrations and disease progression/prognosis. *Cancer*, 92(11), pp.2769-77.

Yen, C.C. et al., 2003. Genotypic analysis of esophageal squamous cell carcinoma by molecular cytogenetics and real-time quantitative polymerase chain reaction. *International journal of oncology*, 23(4), pp.871-81.

Yen, C.C. et al., 2005. Copy number changes of target genes in chromosome 3q25.3-qter of esophageal squamous cell carcinoma: TP63 is amplified in early carcinogenesis but down-regulated as disease progressed. *World journal of gastroenterology : WJG*, 11(9), pp.1267-72.

Ying, J. et al., 2007. Tumor-specific methylation of the 8p22 tumor suppressor gene DLC1 is an epigenetic biomarker for Hodgkin, nasal NK/T-cell and other types of lymphomas. *Epigenetics : official journal of the DNA Methylation Society*, 2(1), pp.15-21.

Yoon, S.Y. et al., 2006. Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. *International journal of oncology*, 29(2), pp.315-27.

Yu, L. et al., 2009. Prostaglandin E(2) promotes cell proliferation via protein kinase C/extracellular signal regulated kinase pathway-dependent induction of c-Myc expression in human esophageal squamous cell carcinoma cells. *International journal of cancer. Journal international du cancer*, 125(11), pp.2540-6.

Yunis, J.J. & Soreng, A.L., 1984. Constitutive fragile sites and cancer. *Science (New York, N.Y.)*, 226(4679), pp.1199-204.

Zare, M. et al., 2009. Qualitative analysis of Adenomatous Polyposis Coli promoter: hypermethylation, engagement and effects on survival of patients with esophageal cancer in a high risk region of the world, a potential molecular marker. *BMC cancer*, 9, p.24.

Zhang, Gongyuan et al., 2010. Expression of nucleostemin, epidermal growth factor and epidermal growth factor receptor in human esophageal squamous cell carcinoma tissues. *Journal of cancer research and clinical oncology*, 136(4), pp.587-94.

Zhang, G.H., Mai, R.-Q. & Huang, B., 2010. Meta-analysis of ADH1B and ALDH2 polymorphisms and esophageal cancer risk in China. *World journal of gastroenterology : WJG*, 16(47), pp.6020-5.

Zhang, H. et al., 2006. Cytogenetic aberrations in immortalization of esophageal epithelial cells. *Cancer genetics and cytogenetics*, 165(1), pp.25-35.

Zhang, L. et al., 2005. Transcriptional coactivator Drosophila eyes absent homologue 2 is up-regulated in epithelial ovarian cancer and promotes tumor growth. *Cancer research*, 65(3), pp.925-32.

Zhang, W. et al., 2010. beta-Catenin/TCF pathway plays a vital role in selenium induced-growth inhibition and apoptosis in esophageal squamous cell carcinoma (ESCC) cells. *Cancer letters*, 296(1), pp.113-22.

Zhang, Xu et al., 2008. [Analyzing the profile of chromosomal imbalances in esophageal atypical hyperplasia and early stage esophageal squamous cell carcinoma by 250K Snp Array]. *Zhonghua yi xue za zhi*, 88(37), pp.2636-41.

Zhao, Xiaojun et al., 2005. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer research*, 65(13), pp.5561-70.

Zheng, J., Hou, P. & Zhou, C., 1996. [Loss of CDKN2 gene and amplification of cyclin D1 gene in human esophageal cancer]. *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 18(6), pp.408-11.

Zhong, X. et al., 2010. NAT2 polymorphism and gastric cancer susceptibility: a meta-analysis. *Archives of medical research*, 41(4), pp.275-80.

Zhu, S., Zhai, F. & Yin, W., 1998. [Effect of cyclin D1 protein expression on the prognosis of esophageal squamous-cell carcinoma treated with preoperative radiotherapy]. *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 20(2), pp.129-31.

APPENDIX A

Cell culture and banding buffers

0,075M KCl

2,8g KCl (Merck)

500ml distilled water

Incubate at 37°C

Fixative

3 parts methanol (Merck)

1 part glacial acetic acid (Merck)

Keep ice-cold

Stock Trypsin

2.5g Trypsin (Difco)

200ml distilled water

Working Trypsin solution

2ml of Stock trypsin in 48ml 0.9% Saline

Wright's stain

0.2% Wright's (Sigma®) stain in 100% methanol

Filtered through Whatman no 1 filter paper.

Stock Gurr buffer

Dissolve one Gurr buffer ampoule of pH 6.8 (Gibco) in 100ml Deionised water

Working solution- Gurr

Make up a 5% solution

DNA extraction buffers

10X PBS

80g NaCl

2g KCl

14,4g Na_2PO_4

2,4g KH_2PO_4

800ml distilled water

pH to 7,4 with HCl

Adjust to 1litre

1X PBS

20ml 10X PBS

180ml distilled water

0.4X SSC/0.3% tween

1ml 20X SSC pH 7

47,5ml distilled water

150µl Tween® 20 (Merck Laboratory Supplies)

Lysis buffer

50mM Tris HCl

1mM EDTA

0.5% tween® 20

2% Agarose gel

2g multi-purpose agarose (Bioline)

100ml 1x TAE buffer

10µl Gel Red™ (Biotium, CA)

Heat until clear

Pours 3 gels

1x TAE buffer

40mM Tris-acetate (pH 7.6) (Merck)

1mM Na_2EDTA (Merck)

BAC culture

LB top agar – per litre

10g Bacto® tryptone (DIFCO)

5g Bacto® yeast extract (DIFCO)

5g AAR® NaCl (SMM Instruments)

Autoclave

Frozen Stock cultures

Glycerol solution

65% Glycerol (vol/vol) (Merck® Laboratory Supplies)

0.1M $MgSO_4$ (Merck® Laboratory Supplies)

0.025M Tris.Cl, pH 8 (Merck® Laboratory Supplies)

Glycerol Stock

850ul of BAC culture

150ul Glycerol solution

Immediately freeze at -70°C

Probe labelling

10X Nick translation buffer

0,5M Tris-HCl pH8.0 (Merck® Laboratory Supplies)

50mM $MgCl_2$ (Merck® Laboratory Supplies)

0,5mg/ml Bovine Serum Albumin (Boerhinger Mannheim)

0,1M β-mercaptoethanol

0.1ml β-mercaptoethanol (BDH)

14,4ml double-distilled water

10x nucleotide stock - SpectrumOrange

0.5mM dATP (Promega)

0.5mM dGTP (Promega)

0.5mM dCTP (Promega)

0.25mM dTTP (Promega)

0.25mM SpectrumOrange d-UTP (Abbott Laboratories)

10x nucleotide stock – SpectrumGreen

0.5mM dATP (Promega)

0.5mM dGTP (Promega)

0.5mM dCTP (Promega)

0.25mM dTTP (Promega)

0.25mM SpectrumGreen d-UTP (Abbott Laboratories)

DNase I solution

3mg DNase I (Roche)

0.5ml of 0.3M NaCl

0.5ml glycerol

Store at -20˚C

Hybridisation buffer

50% deionised formamide (Appendix C) (Saarchem)

2x SSC

10% dextran sulphate (Sigma®)

50mM sodium dihydrogen orthophosphate (Saarchem)

pH to 7,0 with disodium hydrogen orthophosphate (Saarchem)

Store at -20˚C

FISH solutions

Denaturing buffer

35ml deionised formamide (Saarchem)

5ml phosphate buffer

5ml 20x SSC

5ml distilled water

pH to 7 with concentrated HCl

Deionised formamide

1 spatula full Analytical grade mixed bed resin AG 50-X8 (BioRad®) for every 100ml formamide (Biorad)

Place on stirrer for 2 hrs

Filter with Whatman® No1 filter paper

Store at 4ºC

Phosphate buffer

Solution A: KH_2PO_4 (Saarchem) – 4,5g per 500ml, pH to 4,51

Solution B: Na_2HPO_4, $2H_2O$ (Saarchem) – 5,94g per 500ml, pH to 8,97

For 100ml add 41,3ml solution A and 58,7ml solution B

PH to 7.0

Autoclave

50% formamide

20ml 20x SSC

80ml distilled water

100 ml formamide

pH to 7 with HCl

20x SSC

3M NaCl (SMM Chemicals)

0.3M sodium citrate (SMM Chemicals)

Adjust to pH 7.0 with concentrated HCl

Autoclave and store at room temperature

2X SSC with tween

2X SSC

0.05% Tween® 20 (Merck)

DAPI (4', 6-diamino-2-phenylindole) stock solution

0,2mg/ml DAPI (Merck)

2x SSC

DAPI working solution

100 ml 2x SSC

100µl DAPI (0.2µg/ml)

DAPI wash solution

2x SSC: 5ml 20x SSC in 50ml

25µl Tween® 20 (Merck Laboratory Supplies) in 50ml

1N HCl

31ml of 37% HCl (Merck Laboratory Supplies)

Distilled water up to 1000ml

1% Formaldehyde

13.5ml of 37% formaldehyde (Merck Laboratory Supplies)

450ml distilled water

Ph to 7.0 with NaOH

Make up to 500ml, autocalve

Store at 4°C

APPENDIX B

RQ-PCR optimisation and experiment set up

Selection and assessment of stability of reference genes for RQ-PCR analysis.

Internal control genes or reference genes are required for accurate normalisation of expression data. The mRNA for these genes should not be variable in levels of expression between cancerous and normal tissues nor under experimental conditions, i.e. stably expressed or constantly expressed (Vandesompele et al., 2002). The number of reference genes used for normalisation depends on the number of target genes being assessed, the more target genes, the more reference genes are required. A minimum of three genes is recommended where possible (Vandesompele et al., 2002). The stability of reference genes also needs to be assessed experimentally and so reference genes for this study were selected based on the article by Rubie et al, 2005. This paper describes work done to establish a list of reference genes that are stably expressed in various cancerous and corresponding normal tissues, particularly OSCC. Based on their experimental results, three genes were chosen, a highly expressed gene (*ACTB*), a median expressed gene (*YWHAZ*) and a low expressed gene (*PGK1*). These genes appeared to be stably expressed between normal oesophagus and cancerous tissue (Rubie et al., 2005).

The stability of the reference genes was also examined in-house by performing an RQ-PCR experiment using 5 normal oesophageal tissue samples and 2 normal blood samples with 3 replicates each. The run data was imported into Qbase PLUS (Hellemans et al., 2007) and the GNorm methodology was used to establish the reference gene stability (Vandesompele et al., 2002). The reference gene stability was established by comparing the expression ratios of two genes at a time (pairwise variation), the increasing variation in this ratio means that there is decreased stability of expression for one of the genes, this stability value is

delineated *M*. The lower the *M* value the more stable the expression of that gene (Vandesompele et al., 2002). The recommendation is that the *M* values and CV values (coefficient of variation) calculated in Qbase PLUS (Hellemans et al., 2007) are below 0.5 and 25% respectively and the closer the *M* value is to 0.1 the better the stability (Vandesompele et al., 2002).

Results of the reference gene stability test

The reference genes should be stably expressed across all samples being analysed in an experiment so that accurate normalisation factors can be calculated for correction of the relative quantities of each gene being investigated in each sample. It has been established that a minimum of 3 reference genes is required for accurate normalisation and no specific gene is stable in all tissues (Vandesompele et al., 2002). The three genes chosen for this study were selected based on the study by Rubie et al (2005), which looked at reference gene stability in normal and cancerous oesophageal tissue. The three most stable genes were *ACTB*, *PGK1* and *YWHAZ*. Seven different samples, 5 from presumed normal oesophagus and 2 peripheral blood samples from healthy volunteers were used to establish the stability of expression of the 3 chosen reference genes, *ACTB*, *PGK1* and *YWHAZ*. The three genes were also run in all the experiments to verify their stability in the cell lines and samples. To establish the stability of reference genes, the co efficient of variability (CV) and stability values (*M*) were calculated in Qbase PLUS. The results can be seen in table B1. The *M* values should be as close as possible to 0.1 with an acceptable value less than 0.5 (Vandesompele et al., 2002). The reference genes had relatively low expression stability; the gene that had the most stable expression was the *YWHAZ* gene (figure B1). The normalisation factors for each sample decreased when all 3 reference genes were used for

calculation of the normalisation factors. When *YWHAZ* was used alone, a higher normalisation factor was calculated per sample. Although one gene is not ideal, the recommendation in GNorm was to use the gene with the lowest *M* value due to the instability of the remaining two genes. Therefore the *YWHAZ* gene was used to normalise the RQ-PCR data for the experiments.

Table B1. Reference gene stability evaluation based on M and CV values

Reference gene	M value	CV
ACTB	0.584	0.252
PGK1	0.64	0.327
YWHAZ	0.488	0.153
Average	0.571	0.224

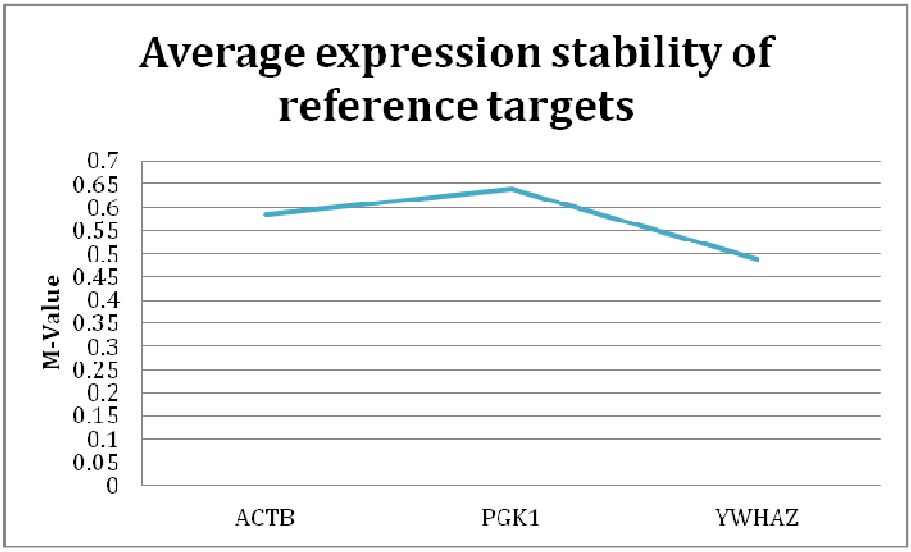

Figure B1. Graphic represenstation of the M-values calculated for each reference gene in Qbase PLUS. *YWHAZ* had the lowest M value (0.488) and was therfore the most stably expressed gene in this experiment.

Melt curve and gel analysis of PCR products

Melt curves were established and compared to the resolved bands by gel electrophoresis to insure the correct products with only one melt peak were amplified (figures B2 B3 and B4).

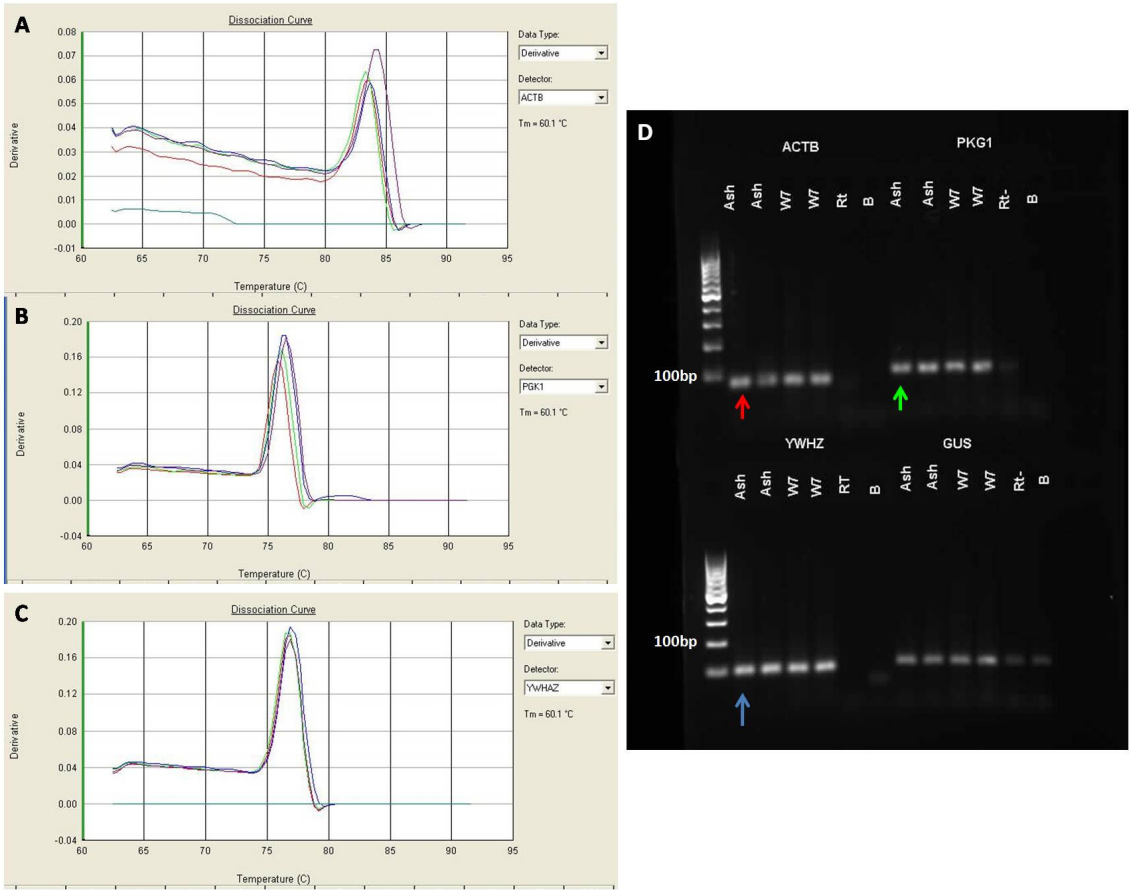

Figure B2. Dissociation curves and 2% gel electrophoresis showing the correct size products detected for the reference genes, *ACTB*, *PGK1* and *YWHAZ* in replicates of a blood control sample. The dissociation curve for *ACTB* (A) showed a melting temperature of approximately 84°C, corresponding to the expected product size of approximately 87bp (red arrow) (D). The dissociation curve for *PGK1* showed an approximate melting temperature of 76°C (B) corresponding to an expected product size of 95bp (green arrow) (D). The melt curve analysis for *YWHAZ* showed a melting temperature of approximately 76°C (C), corresponding to a product size of 94bp (blue arrow) (D).

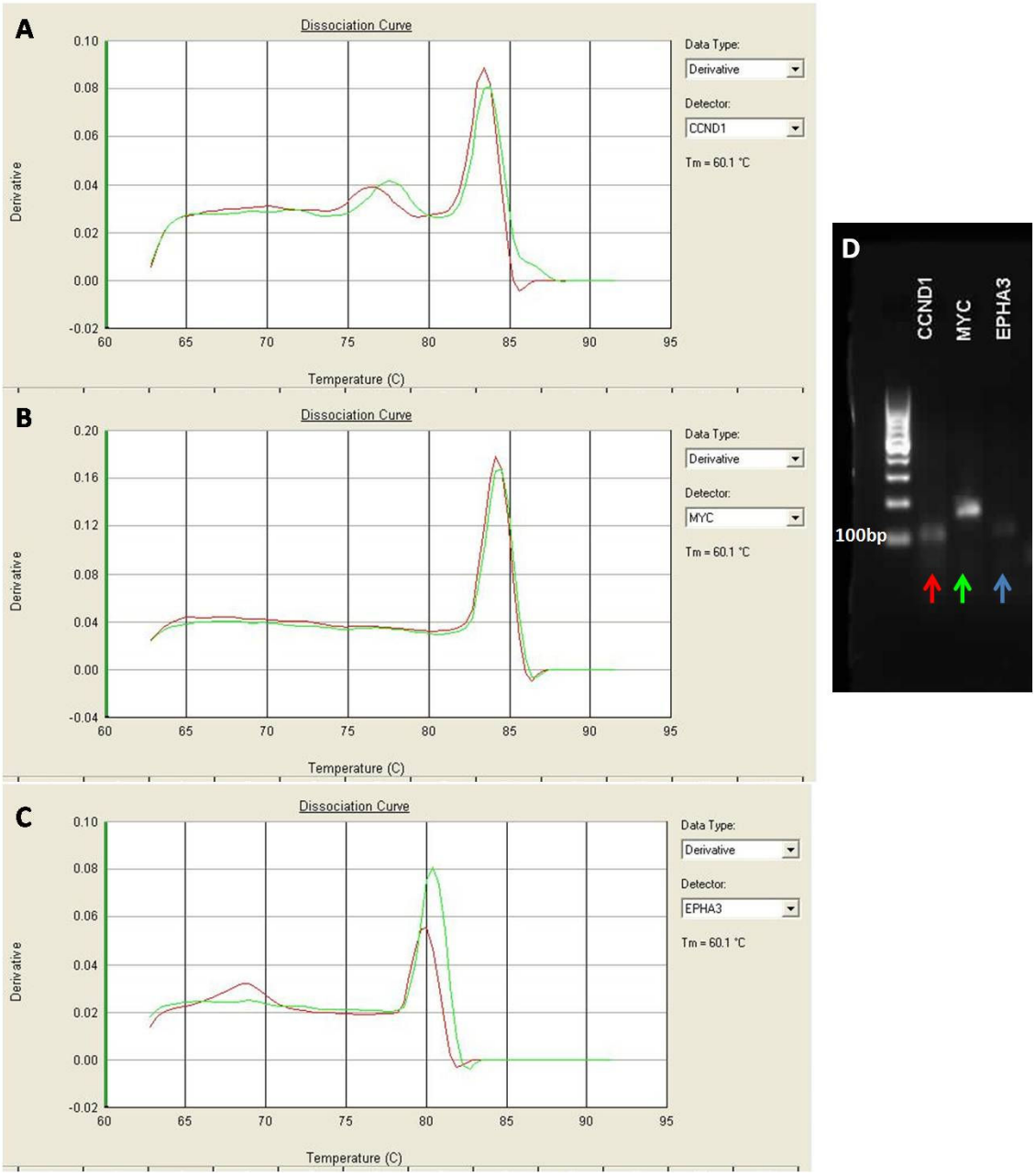

Figure B3. Dissociation curves and 2% gel electrophoresis showing the correct size products detected for the target genes, *CCND1*, *c-MYC* and *EPHA3* in replicates of the positive control cell line, MCF7. The *CCND1* amplicon showed an approximate melting peak at 84°C (A) corresponding to a 124bp product (red arrow) (D). C-*MYC* melts at approximately 85°C, which corresponded to a product size of 147bp (green arrow) (D). *EPHA3* shows very low amplification (D) with a product size of 112bp (blue arrow) and the melting temperature was approximately 80°C (C).

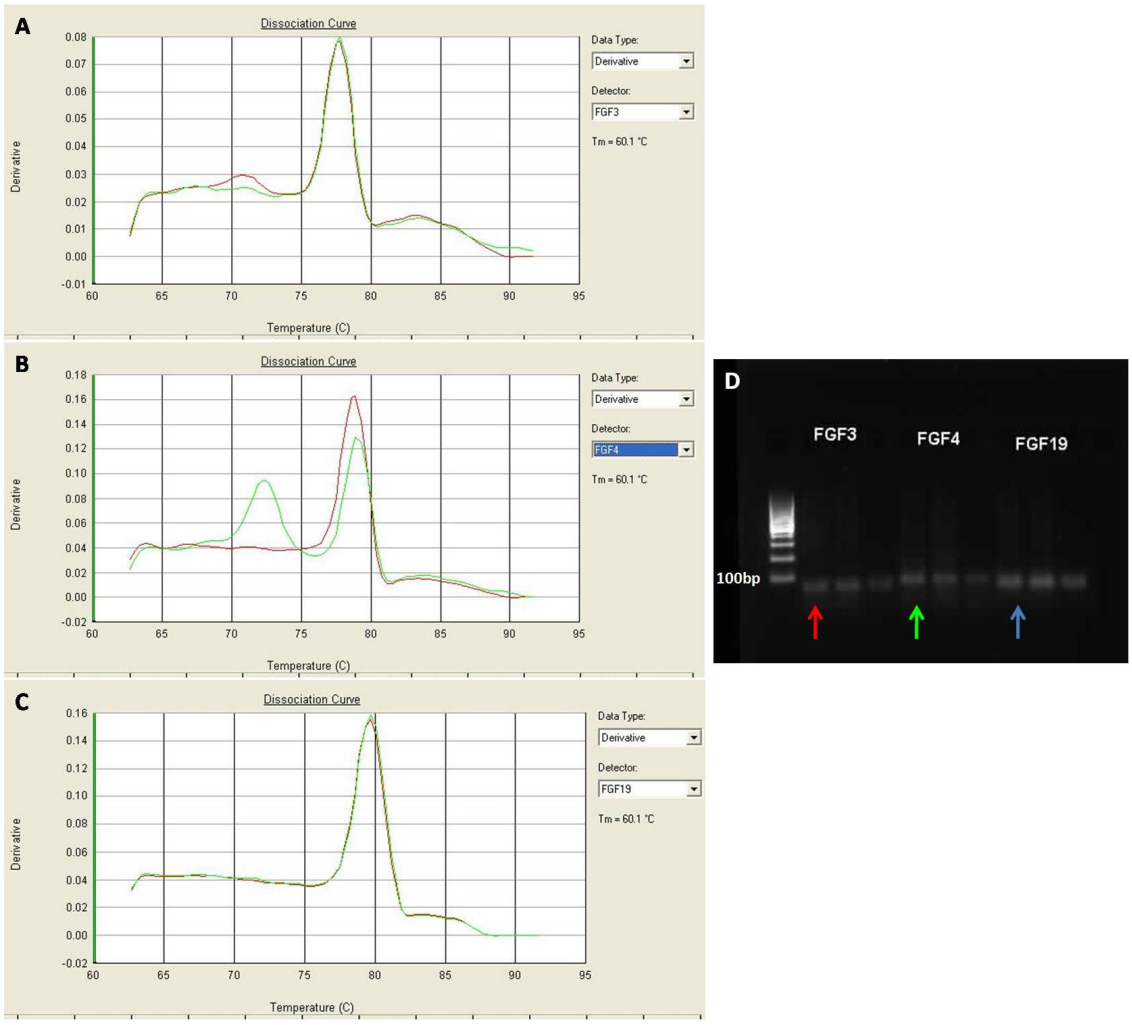

Figure B4. Dissociation curves and 2% gel electrophoresis showing the correct size products detected for the target genes *FGF3*, *FGF4* and *FGF19* in replicates of the MCF7 cell line positive control. The *FGF3* amplicon shows an approximate melting peak at 75.5°C (A) corresponding to 64bp product (red arrow) (D). *FGF4* melts at approximately 80°C, which corresponded to a product size of 83bp (green arrow) (D). *FGF19* showed a product size of 67bp (blue arrow) and the melting temperature was approximately 80°C (C).

RQ-PCR experiment set up

The following tables represent the plate set up for the RQ-PCR experiments in plates 1-9.

Plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A:ACTB	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
B:ACTB	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
C:ACTB	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
D:PKG1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
E:PGK1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
F:PGK1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
G:												
H:												

Plate 2	1	2	3	4	5	6	7	8	9	10	11	12
A:ACTB	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
B:ACTB	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
C:ACTB	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
D:PKG1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
E:PGK1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
F:PGK1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
G:												
H:												

Plate 3	1	2	3	4	5	6	7	8	9	10	11	12
A:YWHAZ	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
B:YWHAZ	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
C:YWHAZ	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
D:EPHA3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
E:EPHA3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
F:EPHA3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
G:												
H:												

Plate 4	1	2	3	4	5	6	7	8	9	10	11	12
A:YWHAZ	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
B:YWHAZ	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
C:YWHAZ	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
D:EPHA3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
E:EPHA3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
F:EPHA3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
G:												
H:												

Plate 5	1	2	3	4	5	6	7	8	9	10	11	12
A: CCND1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
B:CCND1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
C:CCND1	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
D:MYC	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
D:MYC	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA3	WITSCA10	WITSCA10N	
D:MYC	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA4	WITSCA11	WITSCA10N	

Plate 6	1	2	3	4	5	6	7	8	9	10	11	12
A: CCND1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
B:CCND1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
C:CCND1	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
D:MYC	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
D:MYC	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
D:MYC	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
G:												
H:												

Plate 7	1	2	3	4	5	6	7	8	9	10	11	12
A:FGF3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
B:FGF3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
C:FGF3	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	NTC
D:FGF4	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA2	WITSCA9	WITSCA10N	
E:FGF4	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA3	WITSCA10	WITSCA10N	
F:FGF4	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	OC7BT	OC8BT	WITSCA4	WITSCA11	WITSCA10N	
G:												
H:												

Plate 8	1	2	3	4	5	6	7	8	9	10	11	12
A:FGF3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
B:FGF3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
C:FGF3	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				NTC
D:FGF4	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
E:FGF4	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
F:FGF4	WHCO1_A	WHCO3_A	WHCO5_A	WITS35N	WITS36N	WITS37N	WITS39N	WITS40N				
G:												
H:												

Plate 9	1	2	3	4	5	6	7	8	9	10	11	12
A:FGF19	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	WITSCA10N	WITS35N				NTC
B:FGF19	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	WITSCA10N	WITS35N				
C:FGF19	WHCO1	WHCO3	WHCO5	WHCO6	SNO	MCF7	WITSCA10N	WITS35N				
D:												
E:												
F:												
G:												
H:												

Well	Sample Name	Detector	Ct	StdDev Ct	Tm	Well	Sample Name	Detector	Tm
A6	MCF7	ACTB	15.19	0.028	84.4	A1	WHCO1_A	ACTB	83.6
B6	MCF7	ACTB	15.18	0.028	84.4	B1	WHCO1_A	ACTB	83.6
C6	MCF7	ACTB	15.23	0.028	84.7	C1	WHCO1_A	ACTB	84.3
A7	OC7BT	ACTB	16.02	0.032	84.4	A2	WHCO3_A	ACTB	83.9
B7	OC7BT	ACTB	15.96	0.032	84.7	B2	WHCO3_A	ACTB	83.9
C7	OC7BT	ACTB	15.97	0.032	84.7	C2	WHCO3_A	ACTB	84.3
A8	OC8BT	ACTB	16.7	0.068	84.4	A3	WHCO5_A	ACTB	83.9
B8	OC8BT	ACTB	16.57	0.068	84.4	B3	WHCO5_A	ACTB	84.3
C8	OC8BT	ACTB	16.67	0.068	84.7	C3	WHCO5_A	ACTB	84.3
A5	SNO	ACTB	15.19	0.057	84.4	A4	WITS35N	ACTB	84.3
B5	SNO	ACTB	15.1	0.057	84.4	B4	WITS35N	ACTB	84.3
C5	SNO	ACTB	15.09	0.057	84.7	C4	WITS35N	ACTB	84.7
A1	WHCO1	ACTB	15.57	0.096	83.6	A5	WITS36N	ACTB	83.9
B1	WHCO1	ACTB	15.4	0.096	84	B5	WITS36N	ACTB	84.3
C1	WHCO1	ACTB	15.41	0.096	84.4	C5	WITS36N	ACTB	84.7
A2	WHCO3	ACTB	15.53	0.047	84	A6	WITS37N	ACTB	84.3
B2	WHCO3	ACTB	15.56	0.047	84	B6	WITS37N	ACTB	84.3
C2	WHCO3	ACTB	15.47	0.047	84.4	C6	WITS37N	ACTB	84.7
A3	WHCO5	ACTB	15.03	0.055	84	A7	WITS39N	ACTB	84.3
B3	WHCO5	ACTB	14.96	0.055	84	B7	WITS39N	ACTB	84.7
C3	WHCO5	ACTB	14.92	0.055	84.4	C7	WITS39N	ACTB	84.7
A4	WHCO6	ACTB	15.4	0.043	84	A8	WITS40N	ACTB	84.3
B4	WHCO6	ACTB	15.32	0.043	84.4	B8	WITS40N	ACTB	84.3
C4	WHCO6	ACTB	15.4	0.043	84.4	C8	WITS40N	ACTB	84.7
A11	WITSCA10N	ACTB	15.14	0.067	84	A11		ACTB	
B11	WITSCA10N	ACTB	15.23	0.067	84	D1	WHCO1_A	PGK1	76.7
C11	WITSCA10N	ACTB	15.27	0.067	84	E1	WHCO1_A	PGK1	76.7

Table B2 Raw data from RQ-PCR experiments.

A9	WITSCA2	ACTB	15.38	0.025	84	F1	WHCO1_A	PGK1	76.3
B9	WITSCA2	ACTB	15.35	0.025	84.4	D2	WHCO3_A	PGK1	76.7
C9	WITSCA2	ACTB	15.39	0.025	84.7	E2	WHCO3_A	PGK1	76.7
A10	WITSCA9	ACTB	15.49	0.037	84	F2	WHCO3_A	PGK1	76.7
B10	WITSCA9	ACTB	15.42	0.037	84.4	D3	WHCO5_A	PGK1	76.7
C10	WITSCA9	ACTB	15.46	0.037	84.4	E3	WHCO5_A	PGK1	76.7
A12		ACTB	21.59		82.2	F3	WHCO5_A	PGK1	76.7
D6	MCF7	PGK1	20.74	0.211	76.7	D4	WITS35N	PGK1	76.7
E6	MCF7	PGK1	20.97	0.211	76.7	E4	WITS35N	PGK1	77
F6	MCF7	PGK1	20.55	0.211	76.7	F4	WITS35N	PGK1	76.7
D7	OC7BT	PGK1	21.65	0.669	77.1	D5	WITS36N	PGK1	77
E7	OC7BT	PGK1	20.43	0.669	77.4	E5	WITS36N	PGK1	77
F7	OC7BT	PGK1	21.51	0.669	76.7	F5	WITS36N	PGK1	76.7
D8	OC8BT	PGK1	22.3	0.135	76.7	D6	WITS37N	PGK1	77
E8	OC8BT	PGK1	22.04	0.135	76.7	E6	WITS37N	PGK1	77
F8	OC8BT	PGK1	22.13	0.135	76.7	F6	WITS37N	PGK1	77
D5	SNO	PGK1	19.89	0.307	77.1	D7	WITS39N	PGK1	77
E5	SNO	PGK1	20.03	0.307	76.7	E7	WITS39N	PGK1	77
F5	SNO	PGK1	19.44	0.307	76.7	F7	WITS39N	PGK1	77
D1	WHCO1	PGK1	20.08	0.115	76.7	D8	WITS40N	PGK1	77
E1	WHCO1	PGK1	20.28	0.115	76.7	E8	WITS40N	PGK1	76.7
F1	WHCO1	PGK1	20.28	0.115	76.4	F8	WITS40N	PGK1	76.7
D2	WHCO3	PGK1	20.65	0.18	76.7	C11		PGK1	
E2	WHCO3	PGK1	20.33	0.18	76.7				
F2	WHCO3	PGK1	20.63	0.18	76.7				
D3	WHCO5	PGK1	18.82	0.255	76.7				
E3	WHCO5	PGK1	19.02	0.255	76.7				
F3	WHCO5	PGK1	18.52	0.255	76.7				
D4	WHCO6	PGK1	19.82	0.093	76.7				

E4	WHCO6	PGK1	Unknown	19.71	0.093	76.7
F4	WHCO6	PGK1	Unknown	19.63	0.093	76.7
D11	WITSCA10N	PGK1	Unknown	21.17	0.109	76.7
E11	WITSCA10N	PGK1	Unknown	21.32	0.109	76.4
F11	WITSCA10N	PGK1	Unknown	Undetermined		70.2
D9	WITSCA2	PGK1	Unknown	20.39	0.174	76.7
E9	WITSCA2	PGK1	Unknown	20.21	0.174	76.7
F9	WITSCA2	PGK1	Unknown	20.56	0.174	76.7
D10	WITSCA9	PGK1	Unknown	19.56	0.144	76.7
E10	WITSCA9	PGK1	Unknown	19.82	0.144	76.7
F10	WITSCA9	PGK1	Unknown	19.8	0.144	76.7
D12		PGK1	NTC	Undetermined		

Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm	Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm
D6	MCF7	EPHA3	Unknown	33.92	0.505	80.8	D1	WHCO1_A	EPHA3	Unknown	30.8	0.225	80.7
E6	MCF7	EPHA3	Unknown	32.99	0.505	80.8	E1	WHCO1_A	EPHA3	Unknown	30.66	0.225	80.7
F6	MCF7	EPHA3	Unknown	33.1	0.505	80.8	F1	WHCO1_A	EPHA3	Unknown	31.1	0.225	80.7
D7	OC7BT	EPHA3	Unknown	35.39	0.699	80.8	D2	WHCO3_A	EPHA3	Unknown	34.92	0.258	80.7
E7	OC7BT	EPHA3	Unknown	34	0.699	80.8	E2	WHCO3_A	EPHA3	Unknown	34.64	0.258	80.7
F7	OC7BT	EPHA3	Unknown	34.57	0.699	80.8	F2	WHCO3_A	EPHA3	Unknown	35.16	0.258	80.7
D8	OC8BT	EPHA3	Unknown	32.67	0.317	80.8	D3	WHCO5_A	EPHA3	Unknown	39.98	0.475	69.5
E8	OC8BT	EPHA3	Unknown	32.28	0.317	80.8	E3	WHCO5_A	EPHA3	Unknown	39.03	0.475	69.5
F8	OC8BT	EPHA3	Unknown	32.91	0.317	80.4	F3	WHCO5_A	EPHA3	Unknown	39.58	0.475	69.5
D5	SNO	EPHA3	Unknown	36.33	0.392	80.8	D4	WITS35N	EPHA3	Unknown	31.86	0.132	81
E5	SNO	EPHA3	Unknown	35.74	0.392	80.8	E4	WITS35N	EPHA3	Unknown	31.6	0.132	81
F5	SNO	EPHA3	Unknown	36.48	0.392	80.8	F4	WITS35N	EPHA3	Unknown	31.69	0.132	81
D1	WHCO1	EPHA3	Unknown	31.24	1.703	80.4	D5	WITS36N	EPHA3	Unknown	34.13	0.63	80.7
E1	WHCO1	EPHA3	Unknown	32.01	1.703	80.4	E5	WITS36N	EPHA3	Unknown	35.26	0.63	81
F1	WHCO1	EPHA3	Unknown	34.5	1.703	80	F5	WITS36N	EPHA3	Unknown	35.17	0.63	81
D2	WHCO3	EPHA3	Unknown	35.52	0.027	80.4	D6	WITS37N	EPHA3	Unknown	32.95	0.561	81

E2	WHCO3	EPHA3	Unknown	35.47	0.027	80.4	E6	WITS37N	EPHA3	Unknown	32.29	0.561	81
F2	WHCO3	EPHA3	Unknown	35.52	0.027	80.4	F6	WITS37N	EPHA3	Unknown	33.41	0.561	81
D3	WHCO5	EPHA3	Unknown	Undetermined		66	D7	WITS39N	EPHA3	Unknown	32.18	0.136	81
E3	WHCO5	EPHA3	Unknown	Undetermined		69.2	E7	WITS39N	EPHA3	Unknown	31.94	0.136	81
F3	WHCO5	EPHA3	Unknown	Undetermined		80	F7	WITS39N	EPHA3	Unknown	31.94	0.136	81
D4	WHCO6	EPHA3	Unknown	34.57	0.139	80.4	D8	WITS40N	EPHA3	Unknown	33.36	0.076	81
E4	WHCO6	EPHA3	Unknown	34.45	0.139	80.4	E8	WITS40N	EPHA3	Unknown	33.23	0.076	81
F4	WHCO6	EPHA3	Unknown	34.29	0.139	80	F8	WITS40N	EPHA3	Unknown	33.22	0.076	80.7
D11	WITSCA10N	EPHA3	Unknown	38.68	0.901	80.8	C12		EPHA3	NTC	Undetermined		
E11	WITSCA10N	EPHA3	Unknown	39.95	0.901	80.4	A1	WHCO1_A	YWHAZ	Unknown	16.24	0.117	76.7
F11	WITSCA10N	EPHA3	Unknown	Undetermined		68.9	B1	WHCO1_A	YWHAZ	Unknown	16.03	0.117	77.1
D9	WITSCA2	EPHA3	Unknown	34.08	0.413	80.8	C1	WHCO1_A	YWHAZ	Unknown	16.04	0.117	77.1
E9	WITSCA2	EPHA3	Unknown	33.35	0.413	80.8	A2	WHCO3_A	YWHAZ	Unknown	16.03	0.051	77.1
F9	WITSCA2	EPHA3	Unknown	33.38	0.413	80.8	B2	WHCO3_A	YWHAZ	Unknown	15.93	0.051	77.1
D10	WITSCA9	EPHA3	Unknown	33.91	1.15	80.8	C2	WHCO3_A	YWHAZ	Unknown	15.98	0.051	77.1
E10	WITSCA9	EPHA3	Unknown	34.25	1.15	80.8	A3	WHCO5_A	YWHAZ	Unknown	15.22	0.074	76.7
F10	WITSCA9	EPHA3	Unknown	36.05	1.15	80.4	B3	WHCO5_A	YWHAZ	Unknown	15.12	0.074	77.1
D12		EPHA3	NTC	Undetermined			C3	WHCO5_A	YWHAZ	Unknown	15.07	0.074	77.4
A6	MCF7	YWHAZ	Unknown	16.34	0.048	77.2	A4	WITS35N	YWHAZ	Unknown	15.32	0.053	77.1
B6	MCF7	YWHAZ	Unknown	16.26	0.048	77.5	B4	WITS35N	YWHAZ	Unknown	15.22	0.053	77.4
C6	MCF7	YWHAZ	Unknown	16.25	0.048	77.5	C4	WITS35N	YWHAZ	Unknown	15.25	0.053	77.4
A7	OC7BT	YWHAZ	Unknown	17.28	0.208	77.2	A5	WITS36N	YWHAZ	Unknown	16.15	0.103	77.1
B7	OC7BT	YWHAZ	Unknown	17.14	0.208	77.2	B5	WITS36N	YWHAZ	Unknown	15.95	0.103	77.4
C7	OC7BT	YWHAZ	Unknown	17.55	0.208	77.5	C5	WITS36N	YWHAZ	Unknown	16.04	0.103	77.4
A8	OC8BT	YWHAZ	Unknown	18.48	0.271	77.2	A6	WITS37N	YWHAZ	Unknown	15.98	0.073	77.1
B8	OC8BT	YWHAZ	Unknown	18.02	0.271	77.5	B6	WITS37N	YWHAZ	Unknown	15.85	0.073	77.4
C8	OC8BT	YWHAZ	Unknown	18.01	0.271	77.5	C6	WITS37N	YWHAZ	Unknown	15.85	0.073	77.4
A5	SNO	YWHAZ	Unknown	16.48	0.163	77.2	A7	WITS39N	YWHAZ	Unknown	15.52	0.012	77.4
B5	SNO	YWHAZ	Unknown	16.19	0.163	77.5	B7	WITS39N	YWHAZ	Unknown	15.49	0.012	77.4

C5	SNO	YWHAZ	Unknown	16.2	0.163	77.5	C7	WITS39N	YWHAZ	Unknown	15.51	0.012	77.8
A1	WHCO1	YWHAZ	Unknown	16.16	0.069	76.8	A8	WITS40N	YWHAZ	Unknown	15.85	0.054	77.1
B1	WHCO1	YWHAZ	Unknown	16.07	0.069	76.8	B8	WITS40N	YWHAZ	Unknown	15.77	0.054	77.4
C1	WHCO1	YWHAZ	Unknown	16.02	0.069	77.2	C8	WITS40N	YWHAZ	Unknown	15.75	0.054	77.4
A2	WHCO3	YWHAZ	Unknown	16.23	0.131	76.8	A12		YWHAZ	NTC	Undetermined		
B2	WHCO3	YWHAZ	Unknown	15.98	0.131	77.2							
C2	WHCO3	YWHAZ	Unknown	16.04	0.131	77.2							
A3	WHCO5	YWHAZ	Unknown	15.57	0.218	76.8							
B3	WHCO5	YWHAZ	Unknown	15.24	0.218	76.8							
C3	WHCO5	YWHAZ	Unknown	15.16	0.218	77.2							
A4	WHCO6	YWHAZ	Unknown	16.34	0.156	76.8							
B4	WHCO6	YWHAZ	Unknown	16.52	0.156	77.2							
C4	WHCO6	YWHAZ	Unknown	16.21	0.156	77.9							
A11	WITSCA10N	YWHAZ	Unknown	15.99	0.103	76.8							
B11	WITSCA10N	YWHAZ	Unknown	15.79	0.103	77.2							
C11	WITSCA10N	YWHAZ	Unknown	15.85	0.103	77.2							
A9	WITSCA2	YWHAZ	Unknown	16	0.014	77.2							
B9	WITSCA2	YWHAZ	Unknown	16	0.014	77.2							
C9	WITSCA2	YWHAZ	Unknown	16.03	0.014	77.5							
A10	WITSCA9	YWHAZ	Unknown	16.06	0.062	77.2							
B10	WITSCA9	YWHAZ	Unknown	16.01	0.062	77.2							
C10	WITSCA9	YWHAZ	Unknown	16.13	0.062	77.2							
A12		YWHAZ	NTC	Undetermined									
Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm	Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm
A6	MCF7	CCND1	Unknown	31.71	1.392	83.8	A1	WHCO1_A	CCND1	Unknown	34.48	0.07	82.8
B6	MCF7	CCND1	Unknown	29.17	1.392	84.2	B1	WHCO1_A	CCND1	Unknown	34.42	0.07	83.1
C6	MCF7	CCND1	Unknown	29.45	1.392	84.2	C1	WHCO1_A	CCND1	Unknown	34.56	0.07	83.5
A7	OC7BT	CCND1	Unknown	33.49	0.929	83.8	A2	WHCO3_A	CCND1	Unknown	32.81	1.415	82.8
B7	OC7BT	CCND1	Unknown	Undetermined		67.6	B2	WHCO3_A	CCND1	Unknown	30.44	1.415	83.1

C7	OC7BT	CCND1	Unknown	34.8	0.929	83.8	C2	WHCO3_A	CCND1	Unknown	32.98	1.415	83.5
A8	OC8BT	CCND1	Unknown	36.21	1.072	83.5	A3	WHCO5_A	CCND1	Unknown	32.59	0.288	83.1
B8	OC8BT	CCND1	Unknown	35.08	1.072	83.8	B3	WHCO5_A	CCND1	Unknown	33.1	0.288	83.5
C8	OC8BT	CCND1	Unknown	34.07	1.072	84.2	C3	WHCO5_A	CCND1	Unknown	33.07	0.288	83.5
A5	SNO	CCND1	Unknown	31.45	0.329	83.5	A4	WITS35N	CCND1	Unknown	35.76	0.264	83.1
B5	SNO	CCND1	Unknown	30.85	0.329	83.8	B4	WITS35N	CCND1	Unknown	36.07	0.264	83.9
C5	SNO	CCND1	Unknown	31.38	0.329	83.8	C4	WITS35N	CCND1	Unknown	36.28	0.264	83.5
A1	WHCO1	CCND1	Unknown	34.02	0.696	83.5	A5	WITS36N	CCND1	Unknown	34.56	0.923	78.4
B1	WHCO1	CCND1	Unknown	35.09	0.696	83.5	B5	WITS36N	CCND1	Unknown	36.4	0.923	77.7
C1	WHCO1	CCND1	Unknown	33.79	0.696	83.1	C5	WITS36N	CCND1	Unknown	35.59	0.923	83.9
A2	WHCO3	CCND1	Unknown	33.56	0.032	83.1	A6	WITS37N	CCND1	Unknown	34	0.937	83.5
B2	WHCO3	CCND1	Unknown	33.5	0.032	83.1	B6	WITS37N	CCND1	Unknown	35.43	0.937	83.9
C2	WHCO3	CCND1	Unknown	33.52	0.032	83.5	C6	WITS37N	CCND1	Unknown	33.66	0.937	83.9
A3	WHCO5	CCND1	Unknown	33.11	0.466	83.1	A7	WITS39N	CCND1	Unknown	33.46	0.21	83.5
B3	WHCO5	CCND1	Unknown	32.61	0.466	83.5	B7	WITS39N	CCND1	Unknown	33.43	0.21	83.5
C3	WHCO5	CCND1	Unknown	32.18	0.466	83.5	C7	WITS39N	CCND1	Unknown	33.81	0.21	83.9
A4	WHCO6	CCND1	Unknown	32.82	0.772	83.5	A8	WITS40N	CCND1	Unknown	35.25	0.645	83.1
B4	WHCO6	CCND1	Unknown	32.71	0.772	83.8	B8	WITS40N	CCND1	Unknown	34.59	0.645	83.9
C4	WHCO6	CCND1	Unknown	31.43	0.772	83.8	C8	WITS40N	CCND1	Unknown	35.88	0.645	84.2
A11	WITSCA10N	CCND1	Unknown	35.18	0.105	83.5	A12		CCND1	NTC	Undetermined		
B11	WITSCA10N	CCND1	Unknown	35.06	0.105	83.8	D1	WHCO1_A	MYC	Unknown	18.05	0.102	84.2
C11	WITSCA10N	CCND1	Unknown	35.27	0.105	84.2	E1	WHCO1_A	MYC	Unknown	18.08	0.102	84.2
A9	WITSCA2	CCND1	Unknown	33.49	0.147	83.5	F1	WHCO1_A	MYC	Unknown	18.24	0.102	84.2
B9	WITSCA2	CCND1	Unknown	33.2	0.147	83.8	D2	WHCO3_A	MYC	Unknown	18.04	0.268	84.2
C9	WITSCA2	CCND1	Unknown	33.35	0.147	83.8	E2	WHCO3_A	MYC	Unknown	18.54	0.268	84.6
A10	WITSCA9	CCND1	Unknown	34.19	0.14	83.8	F2	WHCO3_A	MYC	Unknown	18.11	0.268	84.2
B10	WITSCA9	CCND1	Unknown	34.29	0.14	83.8	D3	WHCO5_A	MYC	Unknown	17.13	0.109	84.2
C10	WITSCA9	CCND1	Unknown	34.01	0.14	83.5	E3	WHCO5_A	MYC	Unknown	17.23	0.109	84.6
A12		CCND1	NTC	Undetermined			F3	WHCO5_A	MYC	Unknown	17.35	0.109	84.6

D6	MCF7	MYC	Unknown	19.05	0.116	85	D4	WITS35N	MYC	Unknown	18.63	0.081	84.6
E6	MCF7	MYC	Unknown	19.21	0.116	85	E4	WITS35N	MYC	Unknown	18.51	0.081	84.6
F6	MCF7	MYC	Unknown	18.99	0.116	85	F4	WITS35N	MYC	Unknown	18.48	0.081	84.6
D7	OC7BT	MYC	Unknown	21.43	0.162	85	D5	WITS36N	MYC	Unknown	21.14	0.097	84.6
E7	OC7BT	MYC	Unknown	21.22	0.162	84.6	E5	WITS36N	MYC	Unknown	20.96	0.097	84.6
F7	OC7BT	MYC	Unknown	21.11	0.162	85	F5	WITS36N	MYC	Unknown	20.99	0.097	84.6
D8	OC8BT	MYC	Unknown	20.65	0.053	84.6	D6	WITS37N	MYC	Unknown	20.05	1.101	84.6
E8	OC8BT	MYC	Unknown	20.75	0.053	85	E6	WITS37N	MYC	Unknown	21.79	1.101	84.6
F8	OC8BT	MYC	Unknown	20.7	0.053	84.6	F6	WITS37N	MYC	Unknown	19.75	1.101	84.6
D5	SNO	MYC	Unknown	18.7	0.061	85	D7	WITS39N	MYC	Unknown	18.8	0.588	84.6
E5	SNO	MYC	Unknown	18.8	0.061	85	E7	WITS39N	MYC	Unknown	19.14	0.588	84.9
F5	SNO	MYC	Unknown	18.69	0.061	84.6	F7	WITS39N	MYC	Unknown	19.95	0.588	84.6
D1	WHCO1	MYC	Unknown	18.13	0.087	84.6	D8	WITS40N	MYC	Unknown	20.87	0.104	84.6
E1	WHCO1	MYC	Unknown	17.98	0.087	84.2	E8	WITS40N	MYC	Unknown	20.68	0.104	84.6
F1	WHCO1	MYC	Unknown	18.13	0.087	84.2	F8	WITS40N	MYC	Unknown	20.86	0.104	84.6
D2	WHCO3	MYC	Unknown	17.84	0.088	84.6	C12		MYC	NTC	Undetermined		79.5
E2	WHCO3	MYC	Unknown	17.86	0.088	84.6							
F2	WHCO3	MYC	Unknown	18	0.088	84.2							
D3	WHCO5	MYC	Unknown	17.07	0.041	84.6							
E3	WHCO5	MYC	Unknown	17.15	0.041	84.6							
F3	WHCO5	MYC	Unknown	17.1	0.041	84.6							
D4	WHCO6	MYC	Unknown	20.05	0.047	85							
E4	WHCO6	MYC	Unknown	19.97	0.047	84.6							
F4	WHCO6	MYC	Unknown	19.97	0.047	84.6							
D11	WITSCA10N	MYC	Unknown	20.01	0.661	84.6							
E11	WITSCA10N	MYC	Unknown	20.05	0.661	84.6							
F11	WITSCA10N	MYC	Unknown	21.18	0.661	84.2							
D9	WITSCA2	MYC	Unknown	18.57	0.037	84.6							
E9	WITSCA2	MYC	Unknown	18.56	0.037	84.6							

F9	WITSCA2	MYC	Unknown	18.63	0.037	84.6
D10	WITSCA9	MYC	Unknown	20.09	0.052	84.6
E10	WITSCA9	MYC	Unknown	20.13	0.052	84.6
F10	WITSCA9	MYC	Unknown	20.2	0.052	84.2
C12		MYC	NTC	Undetermined		

Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm	Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm
A6	MCF7	FGF3	Unknown	37.23	1.018	77.5	A1	WHCO1_A	FGF3	Unknown	38.96	0.85	77.2
B6	MCF7	FGF3	Unknown	37.67	1.018	77.5	B1	WHCO1_A	FGF3	Unknown	37.75	0.85	73.9
C6	MCF7	FGF3	Unknown	35.73	1.018	77.9	C1	WHCO1_A	FGF3	Unknown	Undetermined		68.4
A7	OC7BT	FGF3	Unknown	39.18	2.079	71.4	A2	WHCO3_A	FGF3	Unknown	37.49	0.358	71.7
B7	OC7BT	FGF3	Unknown	Undetermined		73.2	B2	WHCO3_A	FGF3	Unknown	36.9	0.358	77.6
C7	OC7BT	FGF3	Unknown	36.24	2.079	72.5	C2	WHCO3_A	FGF3	Unknown	37.55	0.358	77.6
A8	OC8BT	FGF3	Unknown	37.36	0.603	71.7	A3	WHCO5_A	FGF3	Unknown	37.78	1.818	77.2
B8	OC8BT	FGF3	Unknown	36.71	0.603	71.7	B3	WHCO5_A	FGF3	Unknown	35.65	1.818	77.6
C8	OC8BT	FGF3	Unknown	36.15	0.603	72.1	C3	WHCO5_A	FGF3	Unknown	34.16	1.818	75.4
A5	SNO	FGF3	Unknown	34.88	0.36	73.2	A4	WITS35N	FGF3	Unknown	37.72	0.455	72.1
B5	SNO	FGF3	Unknown	34.58	0.36	77.5	B4	WITS35N	FGF3	Unknown	37.08	0.455	71.7
C5	SNO	FGF3	Unknown	34.17	0.36	77.9	C4	WITS35N	FGF3	Unknown	Undetermined		68.4
A1	WHCO1	FGF3	Unknown	37.69	0.87	72.9	A5	WITS36N	FGF3	Unknown	36.7	0.878	73.5
B1	WHCO1	FGF3	Unknown	Undetermined		71	B5	WITS36N	FGF3	Unknown	34.98	0.878	75.4
C1	WHCO1	FGF3	Unknown	38.92	0.87	72.1	C5	WITS36N	FGF3	Unknown	36.15	0.878	77.9
A2	WHCO3	FGF3	Unknown	35.68	1.007	71.7	A6	WITS37N	FGF3	Unknown	36.41	0.804	73.5
B2	WHCO3	FGF3	Unknown	36.52	1.007	73.6	B6	WITS37N	FGF3	Unknown	37.64	0.804	77.6
C2	WHCO3	FGF3	Unknown	37.68	1.007	74.3	C6	WITS37N	FGF3	Unknown	37.93	0.804	77.9
A3	WHCO5	FGF3	Unknown	38.85	0.468	77.2	A7	WITS39N	FGF3	Unknown	37.54	0.615	72.8
B3	WHCO5	FGF3	Unknown	37.94	0.468	77.5	B7	WITS39N	FGF3	Unknown	36.37	0.615	77.6
C3	WHCO5	FGF3	Unknown	38.59	0.468	77.5	C7	WITS39N	FGF3	Unknown	37.28	0.615	77.9

A4	WHCO6	FGF3	Unknown	35.79	0.179	79.8	A8	WITS40N	FGF3	Unknown	37.77	1.265	71.7
B4	WHCO6	FGF3	Unknown	35.57	0.179	73.6	B8	WITS40N	FGF3	Unknown	Undetermined		87.6
C4	WHCO6	FGF3	Unknown	35.92	0.179	77.9	C8	WITS40N	FGF3	Unknown	35.99	1.265	76.8
A11	WITSCA10N	FGF3	Unknown	36.64	0.729	72.9	A12		FGF3	NTC	Undetermined		
B11	WITSCA10N	FGF3	Unknown	36.84	0.729	77.2	D1	WHCO1_A	FGF4	Unknown	33.54	0.646	72.1
C11	WITSCA10N	FGF3	Unknown	37.99	0.729	77.5	E1	WHCO1_A	FGF4	Unknown	34.83	0.646	78.7
A9	WITSCA2	FGF3	Unknown	36.78	1.006	77.5	F1	WHCO1_A	FGF4	Unknown	34.19	0.646	72.1
B9	WITSCA2	FGF3	Unknown	36.08	1.006	71.4	D2	WHCO3_A	FGF4	Unknown	35.07	0.681	78.7
C9	WITSCA2	FGF3	Unknown	38.06	1.006	77.5	E2	WHCO3_A	FGF4	Unknown	34.83	0.681	78.7
A10	WITSCA9	FGF3	Unknown	28	0.372	77.5	F2	WHCO3_A	FGF4	Unknown	33.78	0.681	78.7
B10	WITSCA9	FGF3	Unknown	27.41	0.372	77.5	D3	WHCO5_A	FGF4	Unknown	36.12	1.962	72.1
C10	WITSCA9	FGF3	Unknown	27.31	0.372	77.9	E3	WHCO5_A	FGF4	Unknown	39.32	1.962	70.6
A12		FGF3	NTC	Undetermined			F3	WHCO5_A	FGF4	Unknown	35.75	1.962	74.3
D6	MCF7	FGF4	Unknown	34.45	0.302	79	D4	WITS35N	FGF4	Unknown	36.46	2.216	79
E6	MCF7	FGF4	Unknown	34.81	0.302	79	E4	WITS35N	FGF4	Unknown	35.88	2.216	73.9
F6	MCF7	FGF4	Unknown	35.05	0.302	78.6	F4	WITS35N	FGF4	Unknown	39.97	2.216	73.5
D7	OC7BT	FGF4	Unknown	34.94	0.526	73.6	D5	WITS36N	FGF4	Unknown	36.19	0.581	72.1
E7	OC7BT	FGF4	Unknown	34.2	0.526	72.5	E5	WITS36N	FGF4	Unknown	36.3	0.581	72.1
F7	OC7BT	FGF4	Unknown	Undetermined		73.9	F5	WITS36N	FGF4	Unknown	35.24	0.581	79.7
D8	OC8BT	FGF4	Unknown	33.41	0.801	72.5	D6	WITS37N	FGF4	Unknown	33.87	1.521	73.9
E8	OC8BT	FGF4	Unknown	33.42	0.801	72.9	E6	WITS37N	FGF4	Unknown	36.02	1.521	72.5
F8	OC8BT	FGF4	Unknown	34.8	0.801	72.9	F6	WITS37N	FGF4	Unknown	Undetermined		66.6
D5	SNO	FGF4	Unknown	35	0.077	78.6	D7	WITS39N	FGF4	Unknown	34.51	0.645	79
E5	SNO	FGF4	Unknown	35.15	0.077	79	E7	WITS39N	FGF4	Unknown	33.26	0.645	79
F5	SNO	FGF4	Unknown	35.03	0.077	78.6	F7	WITS39N	FGF4	Unknown	34.18	0.645	75.8
D1	WHCO1	FGF4	Unknown	32.79		72.1	D8	WITS40N	FGF4	Unknown	35.01	1.111	75
E1	WHCO1	FGF4	Unknown	Undetermined		71.4	E8	WITS40N	FGF4	Unknown	35.01	1.111	72.1
F1	WHCO1	FGF4	Unknown	Undetermined		68.5	F8	WITS40N	FGF4	Unknown	36.93	1.111	74.3
D2	WHCO3	FGF4	Unknown	34.1	0.736	71.7	C12		FGF4	NTC	Undetermined		72.1

E2	WHCO3	FGF4	Unknown	33.6	0.736	78.6
F2	WHCO3	FGF4	Unknown	35.05	0.736	71.7
D3	WHCO5	FGF4	Unknown	33.42	0.485	78.6
E3	WHCO5	FGF4	Unknown	34.23	0.485	78.6
F3	WHCO5	FGF4	Unknown	34.28	0.485	71.7
D4	WHCO6	FGF4	Unknown	34.13	0.369	78.6
E4	WHCO6	FGF4	Unknown	34.39	0.369	72.5
F4	WHCO6	FGF4	Unknown	33.67	0.369	72.1
D11	WITSCA10N	FGF4	Unknown	33.12	0.386	78.6
E11	WITSCA10N	FGF4	Unknown	33.84	0.386	78.6
F11	WITSCA10N	FGF4	Unknown	33.72	0.386	72.1
D9	WITSCA2	FGF4	Unknown	34.71	0.693	79
E9	WITSCA2	FGF4	Unknown	34.01	0.693	79
F9	WITSCA2	FGF4	Unknown	33.32	0.693	79
D10	WITSCA9	FGF4	Unknown	28.73	0.342	79
E10	WITSCA9	FGF4	Unknown	28.71	0.342	79
F10	WITSCA9	FGF4	Unknown	29.31	0.342	79
D12		FGF4	NTC	Undetermined		

Well	Sample Name	Detector	Task	Ct	StdDev Ct	Tm
A6	MCF7	FGF19	Unknown	34.62	0.451	79.1
B6	MCF7	FGF19	Unknown	33.78	0.451	79.5
C6	MCF7	FGF19	Unknown	33.92	0.451	79.5
A5	SNO	FGF19	Unknown	24.13	0.058	79.5
B5	SNO	FGF19	Unknown	24.01	0.058	79.8
C5	SNO	FGF19	Unknown	24.04	0.058	79.8
A1	WHCO1	FGF19	Unknown	28.21	0.134	79.1
B1	WHCO1	FGF19	Unknown	27.95	0.134	79.5
C1	WHCO1	FGF19	Unknown	28.07	0.134	79.5
A2	WHCO3	FGF19	Unknown	28.9	0.116	79.1
B2	WHCO3	FGF19	Unknown	29.11	0.116	79.1
C2	WHCO3	FGF19	Unknown	29.09	0.116	79.5
A3	WHCO5	FGF19	Unknown	22.25	0.072	79.5
B3	WHCO5	FGF19	Unknown	22.22	0.072	79.5
C3	WHCO5	FGF19	Unknown	22.11	0.072	79.8
A4	WHCO6	FGF19	Unknown	26.98	0.048	79.5
B4	WHCO6	FGF19	Unknown	27.06	0.048	79.5
C4	WHCO6	FGF19	Unknown	27.07	0.048	79.8
A8	WITS35N	FGF19	Unknown	Undetermined		66.7
B8	WITS35N	FGF19	Unknown	Undetermined		66
C8	WITS35N	FGF19	Unknown	Undetermined		66
A7	WITSCA10N	FGF19	Unknown	36.75	0.965	79.1
B7	WITSCA10N	FGF19	Unknown	Undetermined		67.5
C7	WITSCA10N	FGF19	Unknown	35.38	0.965	79.5
A9		FGF19	NTC	Undetermined		

APPENDIX C

500K Assay

StyI and NspI restriction digest

In a final volume of 19.75µl, 1x NE buffer 2/3 (Nsp1 and Sty1 respectively), 1x BSA and 0.5units of Nsp 1 or Sty 1 (New England Biolabs). 250ng of DNA is added for the reaction.

The reaction was incubated in an Applied Biosystems 9700 thermal cycler for 2 hours at 37°C, 65°C for 20 minutes and held at 4°C.

Sty1 and Nsp1 ligation

To the restriction digest in a final volume of 25µl, 1.5uM of Sty1 or Nsp1 adaptor (Affymetrix®), 1x T4 DNA ligase buffer and 64 units of T4 DNA ligase (New England Biolabs) was added. The reaction was incubated at 16°C for 3 hours, 70°C for 20 minutes and held at 4°C.

75µl of molecular grade water (Accugene®, Lonza) was added to the ligation reaction prior to PCR.

Sty1 and Nsp 1 PCR

6 reactions were set up per sample. In a reaction volume of 100µl, 1x Titanium Taq buffer (Clontech), 1M G-C Melt (Clontech), 350µM dNTP mix (Clontech), 4.5µM Primer 002 (Affymetrix), 1x Titanium Taq DNA polymerase (Clontech) was added. 10µl of the ligation mix was added to each reaction. The reaction was performed on the ABI 9700 thermal cycler with the following, 94°C for 3 minutes, 30 cycles of 94°C for 30 seconds, 60°C for 45 seconds, 68°C for 15 seconds and final extension for 7 minutes at 68°C.

3µl of PCR product was resolved on a 2% agarose (Bioline) gel run at 100V for 1 hour. The expected fragments range from 100 to 1100bp. FFPE samples generally amplified to around 700bp (Figure C1).

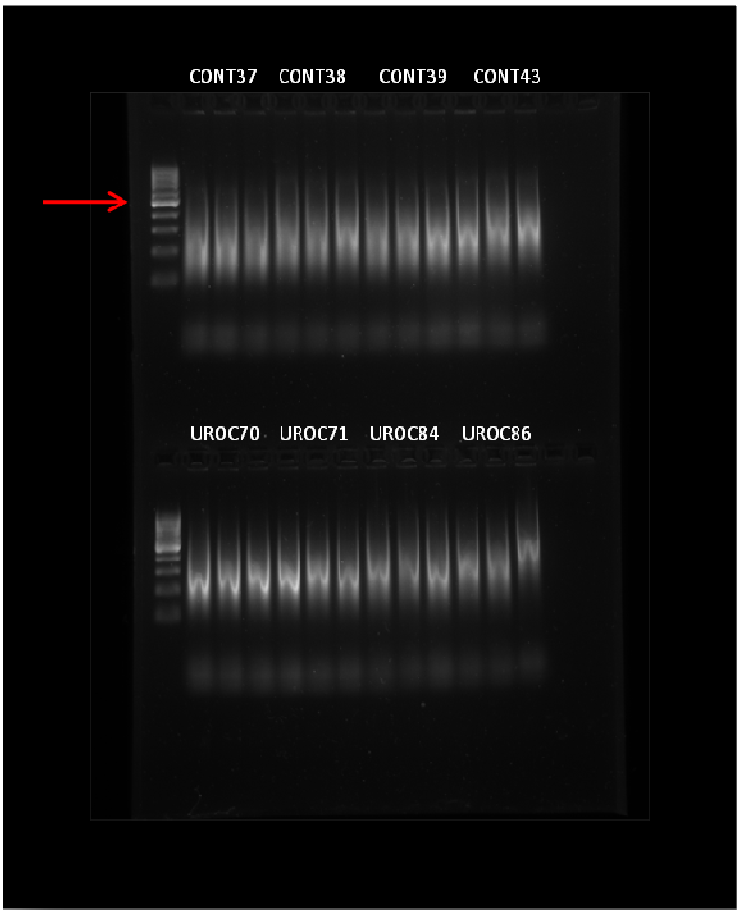

Figure C1. Example of the Nsp1/Sty1 PCR in FFPE samples. The average size amplified was up to 700bp. The red arrow indicates the 500bp mark of the 100bp DNA ladder (Fermentas O'Gene Ruler).

PCR purification

The PCR products were purified by addition of 8ul of 0.1M EDTA (Ambion). Three PCR reactions were mixed together and loaded onto a Clontech DNA amplification clean up plate, which was vacuum pumped at 600 bar for approximately 2 hours until the wells were dry. The wells were washed three times with 50µl molecular grade water (Accugene®, Lonza). The PCR products were then eluted using 45µl of the Clontech RB buffer. The same elution was used for the second set of three PCR reactions for the same sample. The concentration of the PCR product was assessed by making a 100x dilution and reading on the Nanodrop 1000. A total of 90µg in 45µl was used for the remainder of the assay.

Fragmentation

In a final volume of 55µl, 1x fragmentation buffer (Affymetrix®), 0.05 units of fragmentation reagent (Affymetrix®). The reaction was performed on the Applied Biosystems 9700 thermal cycler and was incubated at 37°C for 35 minutes, 95°C for 15 minutes and held at 4°C. The product was immediately assessed for sufficient digestion by resolving 4µl of product on a 2% agarose gel at 100V for 40 minutes (Figure C2).

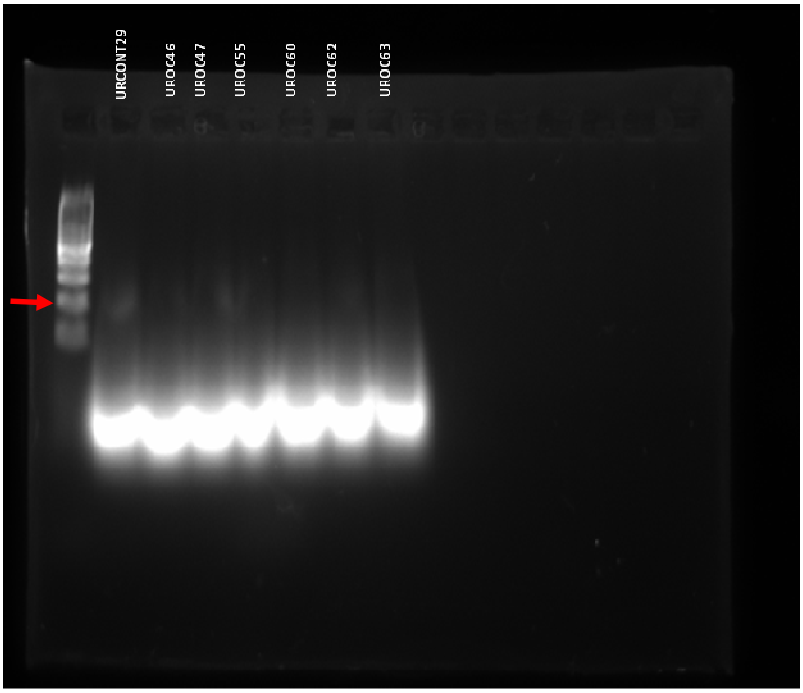

Figure C2. Example of the expected results for the fragmentation step of the 500K assay. The size should be less than 180bp. The red arrow indicates the 200bp mark for the 100bp ladder.

Labelling

In a final volume of 74.5µl, added to the fragmented DNA, 1x TDT buffer (Affymetrix®), 0.857mM labelling reagent (Affymetrix) and 1.5U of TdT (Affymetrix®) was added.

The reaction was performed on an Applied Biosystems 9700 thermal cycler and was incubated at 37°C for 4 hours, 95°C for 15 minutes and held at 4°C. The samples were immediately mixed with hybridisation buffer and stored at -20°C until hybridisation.

Hybridisation

To the labelled DNA in a final volume of 264.5µl, 0.056M MES, 5% DMSO (Sigma®), 2.5x Denhardt's solution (Sigma®), 5.77mM EDTA (Ambion), 0.115mg/ml Herring Sperm DNA (Promega), 1x Oligo control reagent (OCR 0100, Affymetrix®), 11.5µg/ml of Cot1 Human DNA (Invitrogen), 0.01% Tween 20 (Pierce), 2.69M TMACL (Sigma®).

The hybridisation cocktail was heated to 99°C for 10 minutes, placed on ice for 10 seconds, spun down briefly and incubated at 49°C for 1 minute. 200µl of the hybridisation cocktail was loaded into the Affymetrix® 250K Nsp or 250K Sty chip and incubated at 49°C for 18 hours.

Washing

The washing station for Affymetrix® Genechips® is an automated system requiring the stain solutions and the wash buffers. The Mapping 500Kv1_450 protocol was initiated after priming the fluidics station with the Prime_450 protocol.

Buffers

1M Tris
60,57g Tris
(Merck) buffer
400ml H2O
pH8
adjust to 500ml

1M Na2EDTA
93,06g EDTA
(Merck)
400ml H2O
pH 8 with NaOH

12X MES stock (Store at 4 degrees and shield from light)

1.25M MES
0.89M Na+

70,4g MES hydrate (Sigma®)
193,3g MES sodium salt (Sigma®)
800ml H2O
pH 6.5-6.7

Stain buffer

In a final volume of 1188µl, 6x SSPE (Accugene®, Lonza), 0.01% Tween 20 (Pierce) and 1x Denhardt's (Sigma®) solution were added. This buffer was split into two tubes of 594µl, in one tube a final concentration of 10µg/ml of SAPE (Molecular Probes) was added and in the other, 5µg/ml of antibody (Vector laboratories). A third vial containing 830µl of array hold solution was added. These vials are loaded into the appropriate positions on the fluidics station.

Wash buffer A

6X SSPE, 0.01% tween 20 (Pierce)
300ml
20X SSPE (Accugene®, Lonza)
1ml 10% Tween 20
699ml H2O
Filter through 0.2um filter (Nalgene)

Wash buffer B

0.6%SSPE, 0.01%
Tween20
30ml 10XSSPE
1ml 10% tween20
969ml H2O
pH8
Filter through 0.2um
filter

1X array hold buffer
100mM MES (Sigma®)
1M Na+
0.01% Tween 20

8,3ml 12X MES
18,5ml 5M NaCl (Ambion)
0.1ml of 10% Tween20
73.1ml water
Store at 4 degrees
Shield from light

Array assay results

Table C1. Results of the 500K assay for the FFPE OSCC samples

Batch	Sample	Dates		Nsp PCR result	Nsp PCR conc	Nsp Frag result	Nsp Call	Sty PCR result	Sty PCR conc	Sty Frag result	Sty call
		Nsp	Sty								
a	UROC16	31/03/2008-25/03/2008	20/02/2008-25/03/2008	700bp	**70,391ug**	<180bp	**73.89**	700bp	**125,24ug**	<180bp	**74.57**
b	UROC20	14/08/2008-26/08/2008	11/08/2008-26/08/2008	700bp	**61,66ug**	<180bp	**71.42**	700bp	**96,7ug**	<200bp	**71.68**
c	UROC22	13/10/2008-16/10/2008	20/10/2008-23/10/2008	700bp	**139,19ug**	<180bp	**74.37**	700bp	**121.56ug**	<180bp	**74.11**
d	UROC26	03/03/2008-10/03/2008	20/02/2008-03/03/2008	700bp	**103,97ug**	<180bp	**79.25**	700bp	**136,71ug**	<180bp	**79.61**
e	UROC29	11/03/2008-25/03/2008	10/03/2008-17/03/2008	800bp	**69,79ug**	<180bp	**69.2**	700bp	**95,03ug**	<180bp	**67.59**
c	UROC30	13/10/2008-16/10/2008	20/10/2008-23/10/2008	700bp	**150,5ug**	<180bp	**75.69**	700bp	**179,05ug**	<180bp	**80.24**
f	UROC31	12/01/2009-21-01/2009	20/11/2008-27/11/2008	700bp	**62,52ug**	some>200bp	**65.62**	700bp	**130,89ug**	<180bp	**67.45**
g	UROC33	4/11/2008-10/11/2008	10/03/2008-17/03/2008	700bp	**142,56ug**	<180bp	**74.41**	700bp	**102,,6ug**	<180bp	**75.67**
h	UROC37	18/03/2008-25/03/2008	10/03/2008-17/03/2008	700bp	**76,11ug**	<180bp	**74.94**	700bp	**106,34ug**	<180bp	**71.48**
i	UROC40	23/03/2009-26/03/2009	16/04/2009-20/04/2009	700bp	**82,42ug**	<180bp	**69.2**	700bp	**151,96ug**	<180bp	**69.19**
g	UROC42	04/11/2008-10/11/2008	10/03/2008-17/03/2008	700bp	**125,99ug**	<200bp	**71.9**	700bp	**103,03ug**	<180bp	**72.86**
j	UROC46	04/05/2009-08/05/2009	11/05/2009-15/05/2009	>700bp	**102,76ug**	200bp>	**71.6**	>700bp	**132,87ug**	<200bp	**74.97**
j	UROC47	04/05/2009-08/05/2009	11/05/2009-15/05/2009	>700bp	**156,52ug**	200bp>	**74.96**	>700bp	**147,06ug**	<200bp	**74.58**
k	UROC48	09/02/2009-11/02/2009	16/02/2009-18/02/2009	700bp	**98,47ug**	<180bp	**76.12**	700bp	**102,64ug**	<180bp	**75.8**
k	UROC53	09/02/2009-11/02/2009	16/02/2009-18/02/2009	700bp	**78,12ug**	<180bp	**71.47**	700bp	**100,79ug**	<180bp	**73.93**
k	UROC54	09/02/2009-11/02/2009	16/02/2009-18/02/2009	700bp	**73,1ug**	<200bp	**74.13**	700bp	**101,05ug**	<180bp	**70.31**
k	UROC55	09/02/2009-11/02/2009	11/05/2009-15/05/2009	700bp	**90,16ug**	<200bp	**74.65**	>700bp	**147,49ug**	<200bp	**72.47**
l	UROC59	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**68.37ug**	<200bp	**61.84**	>700bp	**132.87ug**	<200bp	**64.03**
j	UROC60	04/05/2009-08/05/2009	11/05/2009-15/05/2009	>700bp	**157,38ug**	200bp>	**71.21**	>700bp	**157,38ug**	<200bp	**75.85**
j	UROC62	04/05/2009-08/05/2009	11/05/2009-15/05/2009	>700bp	**185,94ug**	200bp>	**77.21**	>700bp	**174,58ug**	<200bp	**82.29**
j	UROC63	04/05/2009-08/05/2009	11/05/2009-15/05/2009	>700bp	**140.18ug**	200bp>	**76.49**	>700bp	**175,44ug**	<200bp	**80.13**

m	UROC65	10/07/2009-17/07/2009	27/07/2009-31/07/2009	>700bp	**167.3ug**	<200bp	**60.37**	>700bp	**172.86ug**	<200bp	**66.26**
l	UROC67	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**185.33ug**	<200bp	**75.85**	>700bp	**106.94ug**	<200bp	**70.45**
l	UROC69	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**147.49ug**	<200bp	**64.96**	>700bp	**110.08ug**	<200bp	**66.66**
n	UROC70	02/06/2009-05/06/2009	08/06/2009-11/06/2009	>700bp	**69,96ug**	<200bp	**62.11**	>700bp	**148.92ug**	<200bp	**72.86**
n	UROC71	02/06/2009-05/06/2009	08/06/2009-11/06/2009	>700bp	**112.53ug**	<200bp	**68.65**	>700bp	**170.85ug**	<200bp	**75.73**
n	UROC84	02/06/2009-05/06/2009	08/06/2009-11/06/2009	>700bp	**151.96ug**	<200bp	**75.63**	>700bp	**165.84ng**	<200bp	**83.67**
n	UROC86	02/06/2009-05/06/2009	08/06/2009-11/06/2009	>700bp	**118.58ug**	<200bp	**70.37**	>700bp	**172ug**	<200bp	**83.53**
m	UROC87	10/07/2009-17/07/2009	27/07/2009-31/07/2009	>700bp	**185.33ug**	<200bp	**68.7**	>700bp	**129.56ug**	<200bp	**78.4**
o	UROC98	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	**234.9ug**	<200bp	**64.65**	>700bp	**165.12ug**	<200bp	**79.68**
o	UROC 103	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	**236.8ug**	<200bp	**60.35**	>700bp	**156.2ug**	<200bp	**66.47**
o	UROC124	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	**260.3ug**	<200bp	**65.02**	>700bp	**147.5ug**	<200bp	**66.39**
o	UROC125	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	**197.6ug**	<200bp	**64.13**	>700bp	**176.87ug**	<200bp	**67.66**
p	UROC128	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**74.16ug**	<200bp	**70.46**	>700bp	**129.33ug**	<200bp	**71.5**
p	UROC130	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**99ug**	<200bp	**66.63**	>700bp	**164.5ug**	<200bp	**70.64**
p	UROC131	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**77.1ug**	<200bp	**62.24**	>700bp	**151.9ug**	<200bp	**69.61**
p	UROC132	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**72ug**	<200bp	**65.2**	>700bp	**149.87ug**	<200bp	**73.03**
p	UROC133	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**83.75ug**	<200bp	**62.83**	>700bp	**140.71ug**	<200bp	**67.7**
p	UROC142	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**120.94ug**	<200bp	**66.44**	>700bp	**149.1ug**	<200bp	**72.77**
p	UROC144	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**132ug**	<200bp	**66.75**	>700bp	**156.45ug**	<200bp	**74.16**
q	UROC148	24/12/2009-31/12/2009	24/12/2009-31/12/2009	>700bp	**81.24ug**	<200bp	**65.2**	>700bp	**127.8ug**	<200bp	**69.33**
p	UROC151	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**99.95ug**	<200bp	**67.45**	>700bp	**153.98ug**	<200bp	**70.16**
p	UROC156	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**163.5ug**	<200bp	**72.51**	>700bp	**190.96ug**	<200bp	**72.48**
p	UROC158	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**160.22ug**	<200bp	**72.11**	>700bp	**164.79ug**	<200bp	**72.47**
p	UROC167	18/12/2009-23/12/2009	18/12/2009-23/12/2009	>700bp	**141.8ug**	<200bp	**67.04**	>700bp	**143.52ug**	<200bp	**71.1**
q	UROC171	24/12/2009-31/12/2009	24/12/2009-31/12/2009	>700bp	**81ug**	<200bp	**69.31**	>700bp	**134.62ug**	<200bp	**73.27**
q	UROC172	24/12/2009-31/12/2009	24/12/2009-31/12/2009	>700bp	**115.96ug**	<200bp	**67.63**	>700bp	**155.1ug**	<200bp	**72.08**
x	1708/04	29/01/2010-05/02/2010	29/01/2010-05/02/2010	>700bp	**144.67ug**	<200bp	**73.36**	>700bp	**169.28ug**	<200bp	**78.6**
x	1805/04	29/01/2010-	29/01/2010-	>700bp	**142.62ug**	<200bp	**69.08**	>700bp	**161.92ug**	<200bp	**72.62**

		kj05/02/2010	05/02/2010								
x	2134/04	29/01/2010-05/02/2010	29/01/2010-05/02/2010	>700bp	186.89ug	<200bp	79.86	>700bp	186.98ug	<200bp	82.53
x	2136/04	29/01/2010-05/02/2010	29/01/2010-05/02/2010	>700bp	159.41ug	<200bp	77.94	>700bp	210.51ug	<200bp	81.68
Average							70.05				73.38
o	UROC104	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	254.13ug	<200bp	57.59	>700bp	146.63ug	<200bp	67.41
o	UROC106	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	231.2ug	<200bp	58.23	>700bp	179.74ug	<200bp	66.48
o	UROC108	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	209.84ug	<200bp	57.79	>700bp	178.02ug	<200bp	65.46
o	UROC122	03/08/2009-06/08/2009	06/08/2009-19/08/2009	>700bp	229.7ug	<200bp	58.96	>700bp	189.2ug	<200bp	69.75

The samples highlighted in turquoise were excluded from analysis due to the low call rates

Table C2. Table of results of the 500K assay for the FFPE control samples.

r	URCONT3	29/09/2008-08/10/2008	16/09/2008-08/10/2008	700bp	71,25ug	<180bp	59.42	700bp	82,69ug	<180bp	65.12
c	URCONT4	13/10/2008-16/10/2008	16/09/2008-25/09/2008	700bp	71,81ug	<180bp	64.56	700bp	77,65ug	<180bp	65.94
s	URCONT5	22/09/2008-29/09/2008	16/09/2008-25/09/2008	700bp	116,1ug	<180bp	70.16	700bp	63,6ug	<180bp	70.95
c	URCONT6	13/10/2008-16/10/2008	16/09/2008-08/10/2008	700bp	115,24ug	200bp	73.52	700bp	85,7ug	<180bp	73.11
t	URCONT8	27/11/2008-04/12/2008	04/12/2008-10/12/2008	700bp	141,6ug	<180bp	66.27	700bp	94,77ug	<180bp	71.97
t	URCONT11	27/11/2008-04/12/2008	04/12/2008-10/12/2008	700bp	131,88ug	<180bp	69.27	700bp	103,50ug	<180bp	70.42
t	URCONT13	27/11/2008-04/12/2008	04/12/2008-22/12/2008	700bp	171,7ug	<180bp	69.97	700bp	127,28ug	<180bp	63.79
i	URCONT14	23/03/2009-26/03/2009	16/04/2009-20/04/2009	700bp	118,12ug	<180bp	72.85	700bp	154,32ug	<180bp	71.35
u	URCONT16	05/01/2009-08/01/2009	17/12/2008-22/12/2008	700bp	116,23ug	200bp	67.45	700bp	159,23ug	200bp	62.6
v	URCONT19	19/01/2009-21/01/2009	02/02/2009--4/02/2009	700bp	135,32ug	200bp	73.28	700bp	116,7ug	<200bp	78.16
i	URCONT22	23/03/2009-26/03/2009	16/04/2009-20/04/2009	700bp	82,56ug	200bp	66.74	700bp	139,06ug	200bp	62.97
k	URCONT25	09/02/2009-11/02/2009	16/02/2009-19/02/2009	700bp	86,43ug	<200bp	71.19	700bp	100,79ug	<200bp	69.09
k	URCONT26	09/02/2009-11/02/2009	16/02/2009-19/02/2009	700bp	92,45ug	<200bp	79.45	700bp	104,92ug	<200bp	81.59
k	URCONT27	09/02/2009-11/02/2009	16/02/2009-19/02/2009	700bp	88,58ug	<200bp	76.41	700bp	97,91ug	<200bp	80.07
k	URCONT28	09/02/2009-11/02/2009	16/02/2009-19/02/2009	700bp	95,32ug	<200bp	78.6	700bp	104,92ug	<200bp	78.53
j	URCONT29	04/05/2009-	11/05/2009-	>700bp	137,17ug	200bp>	68.94	>700bp	148,35ug	<200bp	71.71

		08/05/2009	15/05/2009								
l	URCONT32	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**177.16ug**	<200bp	**66.14**	>700bp	**84.14ug**	,200bp	**68.09**
l	URCONT33	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**175.87ug**	<200bp	**66.45**	>700bp	**161.25ug**	<200bp	**65.75**
l	URCONT34	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**184.47ug**	<200bp	**69.61**	>700bp	**177.16ug**	<200bp	**67.72**
l	URCONT35	18/05/2009-22/05/2009	25/05/2009-29/05/2009	700bp	**188.34ug**	<200bp	**68.07**	>700bp	**182.6ug**	<200bp	**73.72**
n	URCONT37	02/06/2009-05/06/2009	08/06/2009-12/06/2009	>700bp	**160.69ug**	<200bp	**67.43**	>700bp	**202.5ug**	<200bp	**69.56**
n	URCONT38	02/06/2009-05/06/2009	08/06/2009-12/06/2009	>700bp	**193.93ug**	<200bp	**69.14**	>700bp	**179.31ug**	<200bp	**80.32**
n	URCONT39	02/06/2009-05/06/2009	08/06/2009-12/06/2009	>700bp	**158,84ug**	<200bp	**70.59**	>700bp	**162.54ug**	<200bp	**75.22**
g	URCONT41	04/11/2008-10/11/2008	20/11/2008-27/11/2008	700bp	**156,39ug**	<180bp	**71.21**	700bp	**149,64ug**	<180bp	**72.1**
g	URCONT42	04/11/2008-10/11/2008	20/11/2008-27/11/2008	700bp	**148,65ug**	<180bp	**74.39**	700bp	**119,25ug**	<180bp	**61.19**
n	URCONT43	02/06/2009-05/06/2009	08/06/2009-12/06/2009	>700bp	**176,3ug**	<200bp	**72.76**	>700bp	**182.46ug**	<200bp	**81.27**
m	URCONT44	10/07/2009-17/07/2009	27/07/2009-31/07/2009	>700bp	**187.65ug**	<200bp	**66.63**	>700bp	**132.44ug**	<200bp	**73.45**
m	URCONT46	10/07/2009-17/07/2009	27/07/2009-31/07/2009	>700bp	**159.5ug**	<200bp	**60.24**	>700bp	**113.52ug**	<200bp	**67.67**
w	URCONT47	03/08/2009-07/08/2009	06/08/2009-19/08/2009	>700bp	**219.3ug**	<200bp	**69.2**	>700bp	**207.8ug**	<200bp	**72.35**
w	URCONT48	03/08/2009-07/08/2009	06/08/2009-19/08/2009	>700bp	**229.5ug**	<200bp	**71.18**	>700bp	**165.8ug**	<200bp	**74.12**
w	URCONT49	03/08/2009-07/08/2009	06/08/2009-19/08/2009	>700bp	**243.7ug**	<200bp	**79.13**	>700bp	**196.5ug**	<200bp	**82.88**
w	URCONT50	03/08/2009-07/08/2009	06/08/2009-19/08/2009	>700bp	**224.2ug**	<200bp	**66.47**	>700bp	**153.1ug**	<200bp	**81.7**
average							**70.2355**				**72.2374**

The sample highlighted in turquoise was excluded from the analysis.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Ms Jacquiline Brown

CLEARANCE CERTIFICATE	M090658
PROJECT	Molecular Profiling of Oesophageal Squamous Cell Carcinoma in the South African Population
INVESTIGATORS	Ms Jacquiline Brown.
DEPARTMENT	Molecular Medicine & Haematology
DATE CONSIDERED	09.06.26
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 09.08.12 CHAIRPERSON .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Dr P Willem

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/WE fully understand the conditions under which I am/ARE authorized to carry out the abovementioned research and I/WE guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/WE undertake to resubmit the protocol to the Committee. I agree to a completion of a yearly progress report.

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

.....

14/10/09