The Role of Viral Sequences in Genetic Aberrations and Malignant Transformation

By

LAMECH MALAGHO MWAPAGHA

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of

MEDICAL BIOCHEMISTRY

International Centre for Genetic Engineering and Biotechnology and the
Faculty of Health Sciences, University of Cape Town

September 2014

DEDICATION

To my parents and siblings who have encouraged and stood by me throughout my entire academic journey

To my late uncle and friend Arthur Kinuthia who motivated me in my studies and foresaw the attainment of this degree

To Prof. Revathi Gunturu who initiated this PhD journey

DECLARATION

I Lamech Malagho Mwapagha, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: …… Date: September, 2014

ACKNOWLEDGEMENTS

I would like to thank all the people that have helped and supported me during the entire course of this PhD project; in particular I would like to express my gratitude to:

Prof. Iqbal Parker my PhD advisor for the guidance and motivation he has shown me throughout the course of my PhD studies. Beyond the bench I have learned and benefited from Parker's quick and exacting eye in designing experiments and compiling presentations. This is a skill that will never fail to be useful, and I am forever grateful. Most importantly his intellectual acumen and continuous encouragement has made this an exciting and rewarding learning process. I certainly look forward to future discussions;

My fellow researchers at ICGEB especially from the Cancer Molecular and Cell Biology group, both past and present, for challenging me to exceed my own personal expectations and for sharing with me the occasional small successes that came from unexpected experiments;

Ida van Jaarsveld of CPGR and Gerrit Botha of UCT-CBIO for their help with whole genome sequencing data analysis and Dr. Nicki Tiffin and Darlington Mapiye of SANBI for their help with microarray data analysis;

The University of Cape Town's ICTS High Performance Computing team, for providing the facilities where all the WGS computations were performed: http://hpc.uct.ac.za;

Dr. Kevin Dzobo, Dr. Matjaz Vogelsang and Nokukhanya Mncwabe who proofread my thesis and provided me with useful comments and suggestions that profoundly improved the composition of this thesis;

The International Centre for Genetic Engineering and Biotechnology (ICGEB), which funded me throughout my PhD studies:

The Poliomyelitis Research Foundation (PRF) which also partly supported this project;

This experience would not be complete without thanking my parents (Gregory Mwaisaka and Dorah Mghoi) and siblings (Peter Mwapagha, Maurice Mwapagha and Robert Mwapagha) for their constant encouragement and motivation from the beginning. Thank you for encouraging me to believe in myself. Thank you for your prayers which have carried me through this entire PhD, I will forever be thankful.

Finally, I would like to thank My Almighty God through his son Jesus Christ for the guidance and blessings you have showered on me throughout this PhD, as it was not by my strength or intellect to have reached this far but by your undeserved favour and grace for I know that

"I can do all things through Christ who strengthens me - *Philippians 4:13*".

ABBREVIATIONS

ASR	Age Standardised Rate
Bp	Base Pair
BSA	Bovine Serum Albumin
CASAVA	Consensus Assessment of Sequence and Variation
cDNA	Complementary Deoxyribonucleic Acid
CMV	Cytomegalovirus
COPD	Chronic Obstructive Pulmonary Disease
CPGR	Centre for Proteomic and Genomic Research
DEGs	Differentially Expressed Genes
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxy-Nucleotide Tri-Phosphate
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
GAD	Genetic Association Database
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GO	Gene Ontology
HPV	Human Papillomavirus
IARC	International Agency of Research on Cancer
Indel	Insertion and Deletion
IPA	Ingenuity Pathway Analysis
IPKB	Ingenuity's Pathways Knowledge Base
Kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
MEM	Minimum Essential Medium
NGS	Next Generation Sequencing
ORF	Open Reading Frame

OSCC	Oesophageal Squamous Cell Carcinoma
P/S	Penicillin /Streptomycin
PBS	Phosphate Buffered Saline
PCA	Principle Component Analysis
Rb	Retinoblastoma Protein
RefSeq	Reference Sequence
RMA	Robust Multichip Average
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
RR	Relative Risk
qRT-PCR	Quantitative Reverse Transcriptase -Polymerase Chain Reaction
SANBI	South African National Bioinformatics Institute
SNPs	Single Nucleotide Polymorphisms
SV	Structural Variant
Ti/Tv	Transition/Transversion
VEP	Variant Effect Predictor
WGS	Whole Genome Sequencing
WHO	World Health Organisation

TABLE OF CONTENTS

CHAPTER SIX: MATERIALS AND METHODS

ABSTRACT

Cancer is a leading cause of death worldwide and viral infections such as HBV/HCV and HPV have been known to be responsible for up to 20% of cancers in low- and middle-income countries. Approximately 500,000 of these deaths are due to oesophageal squamous cell carcinoma (OSSC) alone, one of the major cancers in Eastern and Southern Africa, Latin America and Asia. Previous studies have shown HPV DNA to be integrated in nearly 40% of oesophageal tumours whereas it was present in only 3% of normal healthy asymptomatic individuals, implicating it as a possible risk factor.

The aim of this study was to compare the roles and effects of the E6 gene from the low risk HPV11 and high risk HPV18 on the cellular gene expression profile in order to identify genes required for the initiation of cellular transformation and also to identify genomic alterations associated with oesophageal squamous cell carcinoma. The low risk HPV11 and high risk HPV18 E6 genes were cloned into an adenoviral vector and used to transform normal HaCaT cells to induce high-level expression of these genes in these cells. RNA isolated from these cells was used in DNA microarray analysis to elucidate the signalling pathways and networks that are differentially regulated by E6 during the early stages of transformation. Significantly differentially expressed genes (≥2.0/≤-2.0) were selected for further analysis.

The E6 gene from both HPV11 and HPV18 was capable of inducing cellular transformation via the dysregulation of signalling pathways and networks and by exploiting the NFκB signalling pathway in predisposing normal cells toward malignancy. HPV11E6 was shown to have a much reduced effect compared to HPV18E6, probably due to its involvement in the degradation of p53. In order to investigate whether any other integrated viral sequences may be involved in the aetiology of the disease, whole genome sequencing (WGS) was performed on the DNA isolated from tumour biopsies of four patients with histologically confirmed diagnosis of OSSC.

Paired-end sequencing was performed on the Illumina HiSeq2000, with 2x100 bp reads. Reads were aligned to the Homo sapiens reference genome (NCBI37) using ELAND and CASAVA software. Structural variants (SV) reported from the alignment were collated with

gene loci, (upstream, downstream and overlapping) using the variant effect predictor of ENSEMBL. The affected genes were subsequently cross-checked against the genetic association database for disease and cancer associations. All the reads were aligned to a customised BWA human reference viral genome in order to identify integrated viruses present in the tissue. In addition to viral insertions, several deletions (e.g. *BCLAF1*) were shown to be early events in the progression of OSSC via their ability in induce dysregulation of apoptosis and the induction of inflammation.

CHAPTER ONE
LITERATURE REVIEW

1.1 Oesophageal Cancer

Oesophageal cancer begins to grow in the lining of the oesophagus, and can penetrate the wall of the oesophagus to spread to other parts of the body via the lymphatic system (Fig. 1.1). There are two major types of oesophageal cancer: (i) Oesophageal adenocarcinoma (OAC), which occurs at the gastro-esophageal junction and is mainly associated with gastric reflux; (ii) Oesophageal squamous cell carcinoma (OSCC), which is prevalent in the developing countries of Africa, Latin America and Asia, occurs in the upper part of the oesophagus that is lined with squamous cells and is mainly associated predominantly with smoking, consumption of alcohol and very hot liquids (Hendricks & Parker, 2002; Mqoqi et al., 2004). The different sites of oesophageal cancer stem from the different aetiologies; the sequelae of gastro-esophageal reflux disease involve mostly the distal part of the oesophagus (Meining et al., 2004), whereas the effects of the nutritional changes on the other hand are present in the upper and middle third of the oesophagus (Siewert & Ott, 2007).

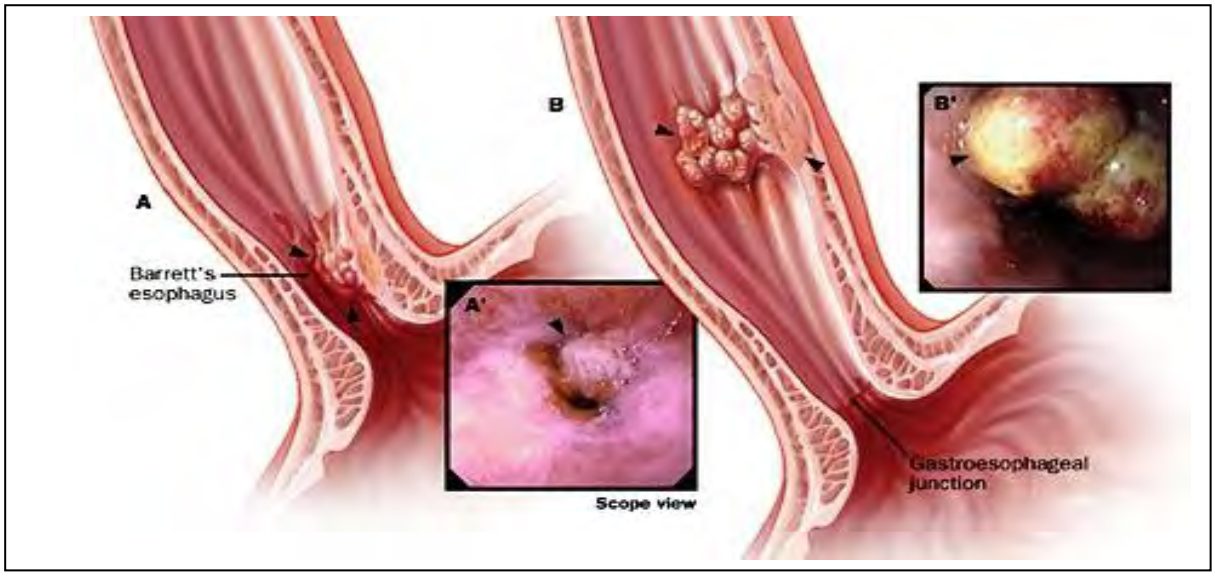


Figure 1.1 Location of the frequently occurring subtypes of oesophageal cancer. A: Oesophageal adenocarcinoma (OAC) occurs mainly in the lower third of the oesophagus and at the gastro-oesophageal junction. **B:** Oesophageal squamous cell carcinoma (OSCC) is generally diagnosed in the upper and middle third of the oesophagus. (Source: http://www.hopkins-gi.org)

1.2 Epidemiology of Oesophageal Cancer

Oesophageal cancer is the eighth most common cancer in the world with an estimated 456,000 new cases in 2012 (3.2% of the total), and the sixth most leading cause of cancer-related mortality death worldwide with an estimated 400,000 deaths (4.9% of the total) (Jemal et al., 2009; Ferlay et al., 2010; Melhado et al., 2010; Ferlay et al., 2014). It is also the fifth most common cancer in the developing world, with approximately 450,000 newly diagnosed patients each year (Parkin et al., 1993; Enzinger & Mayer, 2003; Pohl & Welch, 2005; Lepage et al., 2008; Pennathur et al., 2013). The incidence of oesophageal cancer varies widely according to geographical region and racial background, with greater than 100-fold differences observed between high incidence areas such as China and Iran and low incidence areas such as Western Africa (hul Park & Moon, 2008; Melhado et al., 2010).

These wide variations in incidence are often observed between areas in close geographical proximity (Zaridze et al., 1992). The male to female incidence rate ratio also varies widely with ratios greater than 20:1 in France to near equality or even higher female cases in high incidence areas such as Iran (Melhado et al., 2010). Worldwide, a higher incidence of oesophageal cancer is seen in men with an average 3–4 fold increased rate for OSCC and a 7–10 fold increased rate for OAC compared to women (Vizcaino et al., 2002).

The incidence rates of oesophageal cancer vary internationally with more than 15-fold in men (ASR 22.3 per 100 000 in Southern Africa compared to 1.4 in Western Africa), and almost 20-fold in women (ASR 11.7 per 100 000 in Southern Africa compared to 0.6 in Micronesia and Polynesia). The highest mortality rates are found in both sexes in Eastern and Southern Africa and in Eastern Asia (Figure 1.2, Globocan 2008). There is also a marked regional variation in incidence (MuÑoz & CastellsaguÉ, 1994). Mortality rates closely follow those of incidence. Mortality rates are elevated in Eastern Asia (14.1 per 100,000) and Southern Africa (12.8 per 100,000) in men, and in Eastern Africa (7.3 per 100,000) and Southern Africa (6.2 per 100,000) in women, this maybe due to the relatively late stage of diagnosis and the poor efficacy of treatment (Berrino et al., 1995; Ferlay et al., 2014).

A unique epidemiological feature of oesophageal cancer is its uneven geographic distribution, with high incidence areas within demarcated geographic confines (Kamangar et al., 2006). These geographic ‘hot spots’ include areas in Northern Iran, Kazakhstan, South Africa, Northern China, Southern and Eastern Africa where annual incidence rates can exceed more

than 100 cases per 100 000 population annually (Umar & Fleischer, 2008; Eslick, 2009; Pennathur et al., 2013). In contrast, the incidence rates in most of Western and Middle Africa and Central America are much lower (Jemal et al., 2009).

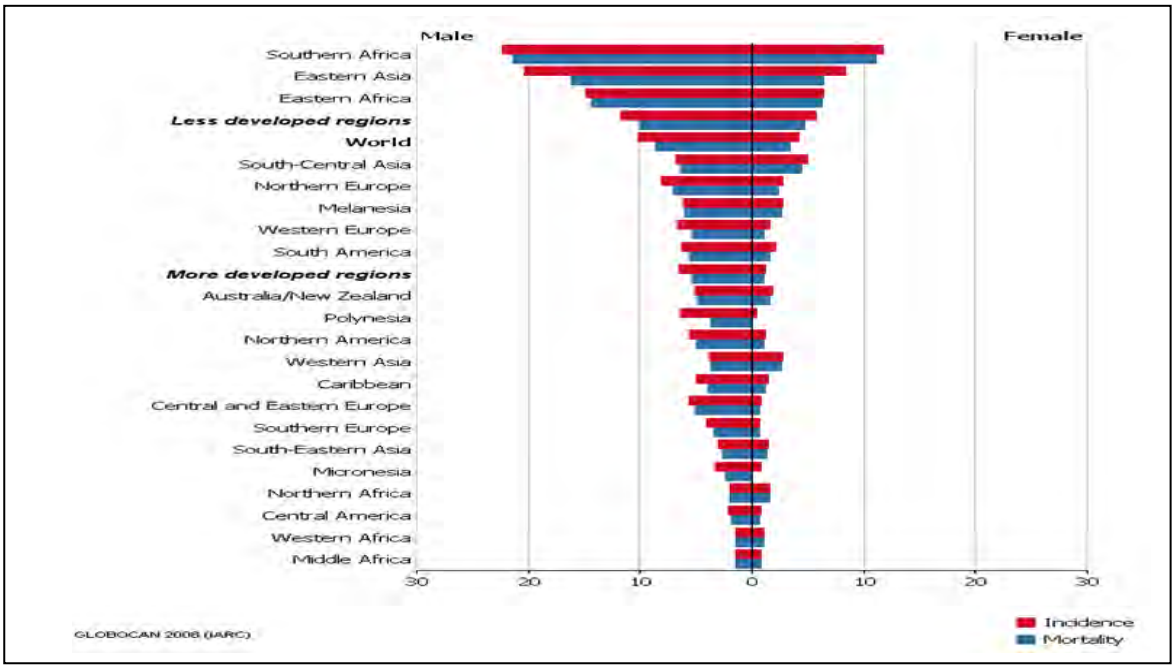


Figure 1.2 Age-standardized oesophageal cancer incidence rates by sex and world area. Incidence rates vary internationally by nearly 16-fold, with the highest rates found in Southern and Eastern Africa and Eastern Asia and lowest rates observed in Western and Middle Africa and Central America in both males and females. Oesophageal cancer is 3 to 4 times more common among males than females. Values for each world area correspond to incidence of oesophageal cancer per 100,000 individuals among males (left) and females (Right). (Source: GLOBOCAN 2008)

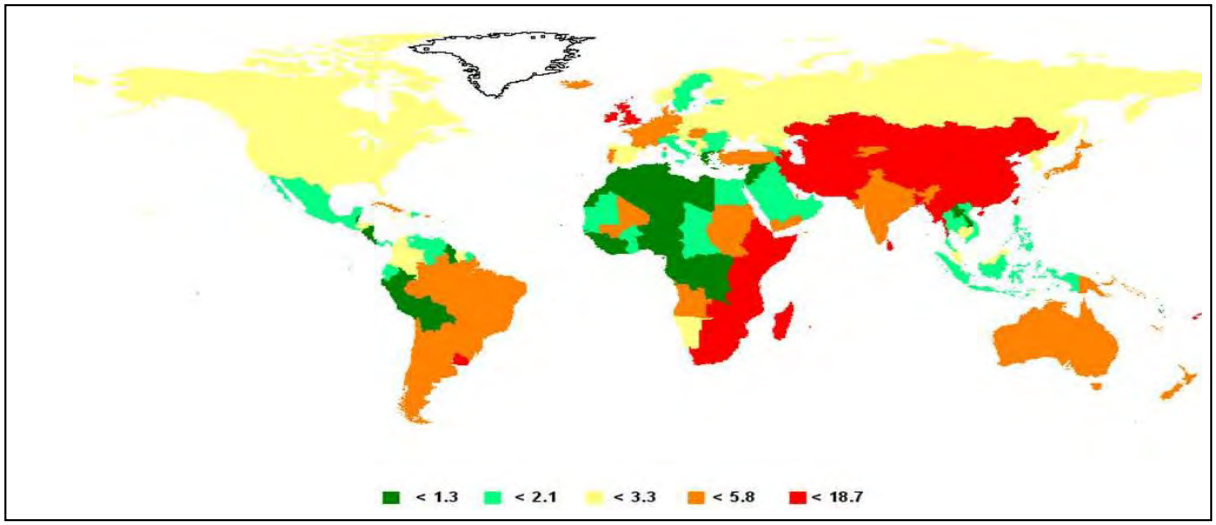


Figure 1.3 Worldwide age-standardized incidence rates (per 100,000 individuals) of oesophageal cancer in both sexes. Incidence rate groups are represented by colours on the world map, ranging from green (low incidence areas) to yellow (medium incidence areas) and red (high incidence areas). The incidence of OSCC is the highest in developing countries such as Southern Africa, China, and South America. However, high incidence areas are also observed in some European countries such as the United Kingdom, France, and Portugal. Source: (Ferlay et al., 2010; Ferlay et al., 2014)

There is extreme geographic variation in the incidence of oesophageal carcinoma with differences of as much as 300-fold between areas of greatest and least prevalence within a given country or between countries (Ferlay et al., 2010; Ferlay et al., 2014), Fig. 1.3.

On the African continent, a number of reports have documented a very high incidence of OSCC in South Africa with a few reports regarding the occurrence of OSCC in Central and Eastern Africa (Hendricks & Parker, 2002; Matsha et al., 2007; Leichman & Thomas, 2011). In parts of central and western Kenya, OSCC ranks among the most common cancers while in Uganda and Tanzania it is less common (Wakhisi et al., 2005).

1.3 Aetiology of OSCC

1.3.1 Alcohol

The mechanism of effect of alcohol is uncertain; it may act as a direct irritant to the oesophageal epithelium, may increase susceptibility to other carcinogens or may contribute to dietary deficiencies that predispose subjects to OSCC (Holmes & Vaughan, 2007). Many case-control studies have shown both independent and synergistic effects of alcohol and smoking (Crew & Neugut, 2004). Alcohol and tobacco use accounts for the majority of the excess risk of developing OAC in US black males (Melhado et al., 2010). Alcohol consumption peaked in the United States in 1980 to 1981 and recent declines in alcohol consumption correlates with declines in OAC incidence (Brown & Devesa, 2002). In South Africa, alcohol consumption has been strongly associated with increased risk of developing OSCC in the rural areas, particularly where the consumption of homemade traditional maize beer is common (Sammon, 1992; Matsha et al., 2006).

In the United Kingdom a case-control study conducted on alcohol, smoking and body mass index (BMI) as risk factors for both OSCC and OAC, confirmed an association of alcohol and OSCC estimating an odds ratio (OR) of 3.39 for subjects consuming more than 34 Units of alcohol per day with a positive dose-response relationship (Lindblad et al., 2005). A case-control study in Italy and Switzerland showed OSCC with family history of cancer as the primary exposure and smoking and alcohol together as secondary exposures, for those without a family history of oesophageal cancer the OR for current smokers consuming ≥49 drinks per week was 15.5 compared with non smokers consuming less alcohol (Garavello et al., 2005). In the high risk areas of the central Asian oesophageal cancer belt, smoking and alcohol consumption play a less significant role in the pathogenesis of OSCC (Melhado et al.,

2010). Alcohol consumption is rare in Golestan, Iran and is therefore unlikely to be a major cause of OSCC (Brown & Devesa, 2002; Pourshams et al., 2005).

1.3.2 Smoking

In Africa, as in other parts of the world with a high incidence of oesophageal cancer, smoking and alcohol consumption feature as the most common risk factors for oesophageal cancer (Gamliel, 2000; Pacella-Norman et al., 2002). Prospective epidemiologic data shows that smoking greatly increases the risk of OSCC, with smokers having a 5-fold higher risk than non-smokers and with risk for heavy smokers increasing nearly 10-fold (Blot & McLaughlin, 1999). The risk of OSCC decreases relatively rapidly after smoking cessation, with substantial declines within 5 to 10 years (Crew & Neugut, 2004). A study in Taiwan also found an increased risk of OSCC with smoking; however, it was found to be a smaller risk increase than for alcohol consumption. The OR for current smokers was 4.2 and for former smokers 3.4, as compared to subjects who had never smoked (Lee et al., 2005).

1.3.3 Chronic Inflammation

Epidemiologic studies support the thesis that acute inflammation contributes to the regression of cancer in contrast to chronic inflammatory diseases, which are frequently associated with increased risk of cancers (Balkwill & Mantovani, 2001; Lu et al., 2006). In China, 65% of males and 63.5% of females from a high risk population undergoing endoscopy had chronic oesophagitis (Munoz et al., 1982). In a similar study in northern Iran, chronic oesophagitis was seen in 80% of subjects, with a high incidence even in young patients (Crespi et al., 1979). In the Transkei, 24% of adults from a high risk area had oesophagitis detected by oesophageal cytology smears; significantly greater than in an adjacent low risk area (Jaskiewicz et al., 1987). A study carried out in a high risk region of China, followed a cohort of patients with cytological diagnoses of normal oesophageal epithelium, oesophagitis and oesophageal dysplasia over 13.5 years and found that oesophagitis conferred no greater risk of invasive OSCC than normal oesophageal epithelium. The risk of subsequent oesophageal cancer increased with the degree of dysplasia, with high grade dysplasia and carcinoma *in situ* having a similar risk, bringing into question the role of chronic inflammation in the pathogenesis of OSCC (Wang et al., 2005).

Caustic injury, which causes intense inflammation and stricturing, was also shown to increase the risk of OSCC approximately 1000-fold, with the time from ingestion to presentation

varying. Since accidental ingestion often occurs in children, the resulting tumours tend to be observed in a younger age group than other oesophageal cancers (Appelqvist & Salmo, 1980; Kim et al., 2001; Kochhar et al., 2006).

1.3.4 Genetic Predisposition

Epidemiologic evidence shows environmental risk factors for OSCC; however, only a subset of exposed individuals will develop OSCC, suggesting that genetic factors may be involved in its pathogenesis (Melhado et al., 2010). OSCC is a frequent form of cancer that shows striking variations in geographic distribution, reflecting exposure to specific environmental factors that are still poorly defined. It develops as the result of a sequence of histopathological changes that typically involve oesophagitis, atrophy, mild to severe dysplasia, carcinoma *in situ* and finally invasive cancer (Mandard et al., 2000).

Genetic changes associated with the development of OSCC include mutations in the *p53* gene, disruption of cell-cycle control in G1 by several mechanisms (inactivation of p16MTS1, amplification of Cyclin D1, alterations of retinoblastoma (RB)), activation of oncogenes (e.g., *EGFR*, *c-MYC*) and inactivation of several tumour suppressor genes. Loss of heterozygosity on chromosome 17q25 has been linked with tylosis, a rare autosomal dominant syndrome associated with high predisposition to OSCC. Microsatellite marker loss of heterozygosity(LOH) studies have also shown that allelic losses on chromosomes 1p, 3p, 5q, 9,11q, 13q, 17 and 18q are frequent in OSCC (Mandard et al., 2000).

Early epidemiologic studies suggested increased susceptibility amongst Turkmen compared to Persians in the high risk Caspian littoral of Iran. These studies further proposed that the high incidence of OSCC along the trade route may be due to high penetrance of susceptibility genes, spread by Turkmen and similar groups who had a nomadic lifestyle and migrated from East Asia, inhabiting much of the Silk route most likely as a result of trading activity (Akbari et al., 2006).

A number of the alcohol dehydrogenases (ADH), that oxidize ethanol to acetaldehyde that is in turn oxidized to aldehyde by aldehyde dehydrogenase (ALDH), have been implicated in the pathogenesis of OSCC (Lao-Sirieix et al., 2010). Acetaldehyde has been demonstrated to be carcinogenic and genetic variations resulting in functional differences in ADH and ALDH activity that lead to increased levels of acetaldehyde in alcohol drinkers may be an important risk factor (Druesne-Pecollo et al., 2009). The *ALDH2* gene encodes the ability to metabolize

acetaldehyde with the *ALDH2**2 allele producing an inactive protein unable to metabolize acetaldehyde; however, individuals who are *ALDH2**2*2 homozygotes are protected from OSCC whilst *ALDH2**1*2 heterozygotes have increased risk (Melhado et al., 2010). In *ALDH2**1*2 heterozygotes the increased risk was modified by alcohol consumption with no appreciable increase in risk among non-drinkers and with moderately increased risk in moderate drinkers (OR 2.49; 95% CI 1.39–4.49) and the highest risk (OR 7; 95% CI 3.07–13.6) was found in heavy drinkers (Lewis & Smith, 2005; Dong et al., 2008).

Xenobiotics are activated to carcinogens by phase 1 enzymes and subsequently detoxified by phase 2 enzymes such as cytochrome p450 1A1 (*CYT1A1*) and glutathione *S*-transferase M1 (*GSTM1*) respectively (Zhuo et al., 2009). Evidence suggests that genetic polymorphisms of these genes may influence the balance between metabolic activation and detoxification of toxicants and are thus related to individual susceptibility to oesophageal cancer. But whether a *CYP1A1* or *GSTM1* polymorphism is a risk factor for oesophageal cancer remains largely uncertain (Hiyama et al., 2007; Matejcic et al., 2011).

1.3.5 Infectious Factors

Microorganisms have an aetiological role in oesophageal carcinogenesis either by producing carcinogens or promoters, or by acting directly on host cells with the chronic inflammatory processes caused by infectious microorganisms resulting in sustained regenerative proliferation of cells and increase in risk of cancer development (Chang et al., 1992). Syrjanen and colleagues found condyloma-like lesions in the specimens of oesophageal cancer providing a link with HPV infection for the first time (Syrjanen et al., 1982). Some studies have also shown HPV induced cytopathic changes leading to infection of the oesophagus and confirmed by immunohistochemical techniques demonstrating HPV antigens in squamous cell papilloma (Syrjanen, 1982; Chang et al., 1990).

Subsequently, DNA hybridization studies and polymerase chain reaction techniques disclosed HPV DNA sequences in both benign and malignant oesophageal lesions (Winkler et al., 1985; Kulski et al., 1986). By also using the filter in situ hybridization (FISH) method mixed with DNA probe containing HPV types 11, 16, and 18, HPV was also found in 22.0% (2/9) of patients without cytologic atypia, 50.0% (3/6) with mild dysplasia, 80.6% (25/31) with moderate dysplasia, 67.9% (19/28) with severe dysplasia and 66.7% (4/6) with invasive OSCC (Galloway & McDougoll, 1989; Chang et al., 1990). A study done in Transkei, South

Africa, a high risk area for oesophageal cancer, also implicated HPV 11 and HPV 39 as aetiological agents in OSCC (Matsha et al., 2002).

1.3.6 Other Factors

Increased risk of OSCC has been shown with exposure to radiation, both in patients exposed to therapeutic radiation and in atomic bomb survivors (Blot & McLaughlin, 1999). Occupational exposures including asbestos, perchloroethylene and combustion products have also been associated with OSCC in some studies (Crew & Neugut, 2004). Experimental induction of OSCC in mice has been reported with an incidence related to the intensity of the radiation administered (Gates & Warren, 1968). The effects of radiation appear to differ according to age at the time of exposure, dose, duration of radiation and the tissue irradiated (Mullen et al., 1979).

An association case-control study of occupational exposures and the risk of both OSCC and OAC in Sweden found significant associations for OSCC with three occupational groups: concrete and construction (OR=2.2), food and tobacco processing (OR=5.1), and hotel and restaurant workers (OR=3.9). The hypothesis advanced was that among concrete and construction workers, airborne particulates may be trapped in airways and swallowed, thereby exposing the oesophageal mucosa to carcinogens. Hotel and restaurant workers on the other hand may be exposed to passive smoke, frying fumes, or the study results could reflect residual confounding by individual consumption of tobacco and alcohol (Jansson et al., 2006).

1.4 Human Papillomavirus (HPVs)

1.4.1 HPV Classification

Human Papillomaviruses are epitheliotropic non-enveloped small double-stranded circular DNA viruses that show a tropism for squamous epithelium (Sapp & Bienkowska-Haba, 2009). There are nearly 150 different HPV types recognized, of which 120 HPV types are fully sequenced (Moody & Laimins, 2010). HPVs have been traditionally referred to as ‘‘types’’, with a type representing a cloned full-length HPV genome whose L1 nucleotide sequence is at least 10% dissimilar from that of any other HPV type (Rautava & Syrjänen, 2012). The HPV types have also been grouped into mucosal or cutaneous types, based on their tropism for specific epithelial sites. Mucosal HPV types found preferentially in precancerous and cancerous lesions have been designated as ‘high-risk’ types and these

include types 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 70. Mucosal HPVs found in benign genital warts and other non-malignant lesions are generally labelled as 'low-risk' types, the most important ones being HPV 6, 11, 42, 43, and 44 (Zur Hausen, 2002). Most genital and mucosal HPVs belong to the alpha genus while the skin types are included in the beta genus (Fig. 1.4).

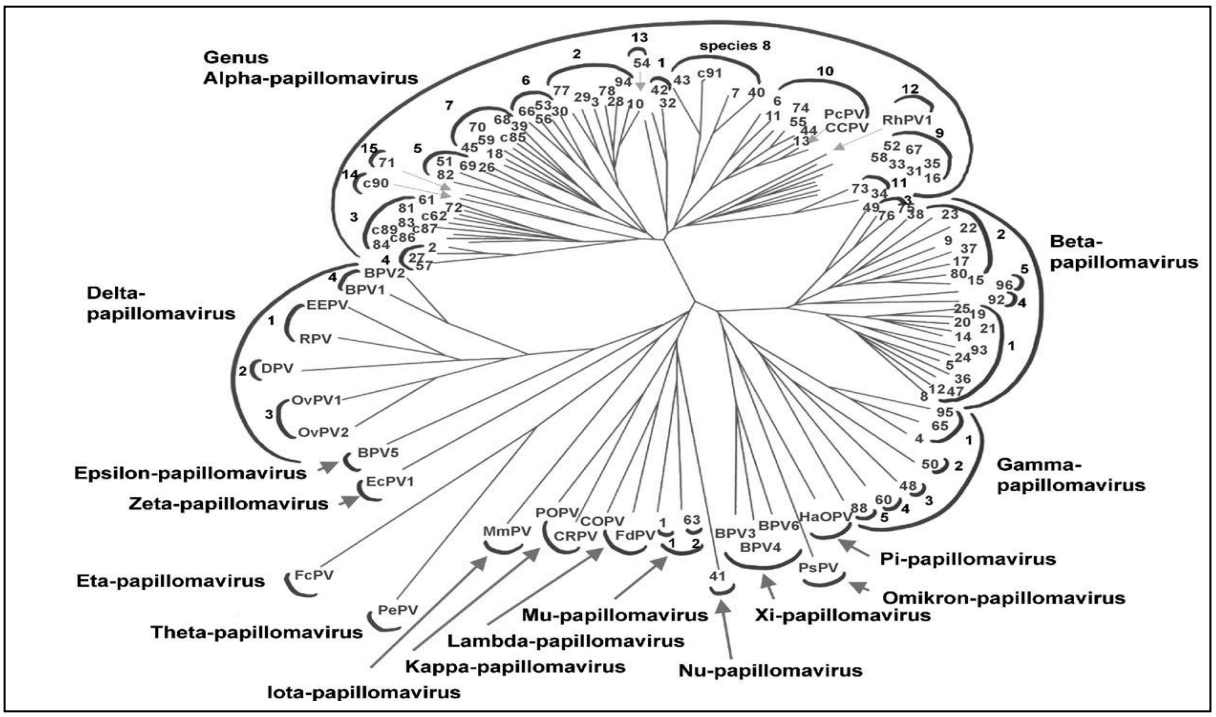


Figure 1.4: Phylogenetic tree containing the sequences of 118 papillomavirus types. The numbers at the ends of each of the branches identify an HPV type; c-numbers refer to candidate HPV types. All other abbreviations refer to animal papillomavirus types. The outermost semicircular symbols identify papillomavirus genera, e.g. the genus alpha-papillomavirus. The number at the inner semicircular symbol refers to papillomavirus species (De Villiers et al., 2004).

The alpha papillomaviruses also include cutaneous viruses such as HPV2 that cause common warts and are only very rarely associated with cancers (De Villiers et al., 2004). The beta papillomaviruses are associated with cutaneous infections in humans and exist in latent form in the general population. Although in immune-compromised individuals and in patients suffering from the inherited disease *Epidermodysplasia Verruciformis* (EV), they can spread unchecked and become associated with the development of non-melanoma skin cancer (Harwood & Proby, 2002; Pfister, 2003). The additional genera, gamma, mu and nu include also HPV types associated with skin papillomas and skin warts and generally do not progress to cancer (Pfister, 1992; Shamanin et al., 1994; De Villiers et al., 2004).

1.4.2 HPV Genome

The circular, double-stranded DNA genomes of all HPVs are approximately 8kb in size. HPVs are small non-enveloped viruses (~55nm in diameter) with an icosahedral capsid (Longworth & Laimins, 2004). Molecular cloning and sequencing of the papillomaviruses have revealed a genomic organization typical of all members of the HPV family, with 8 or 9 open reading frames (ORFs) found on the same DNA strand. These ORFs encode for proteins categorised into early and late proteins according to their expression during the virus life cycle (Longworth & Laimins, 2004; Tommasino, 2014).

The HPV genome can be divided into three main regions: (i) a coding region containing the early genes E1, E2, E4, E5, E6, and E7 that are responsible for transcription, viral genome replication and transformation of the host cell; (ii) a region containing the late genes the major (L1) and minor (L2) capsid proteins which encode the structural capsid protein. The L1 protein is the primary structural element, with infectious virions containing 360 copies of the protein organised into 72 capsomeres (Modis et al., 2002); (iii) a non-coding region, termed the long control region (LCR) sometimes referred to as the upstream regulatory region (URR), which is localized between ORFs L1 and E6 and contains transcriptional, post-transcriptional and replicative cis-regulatory elements (Graham, 2008; Thierry, 2009) see Fig. 1.5.

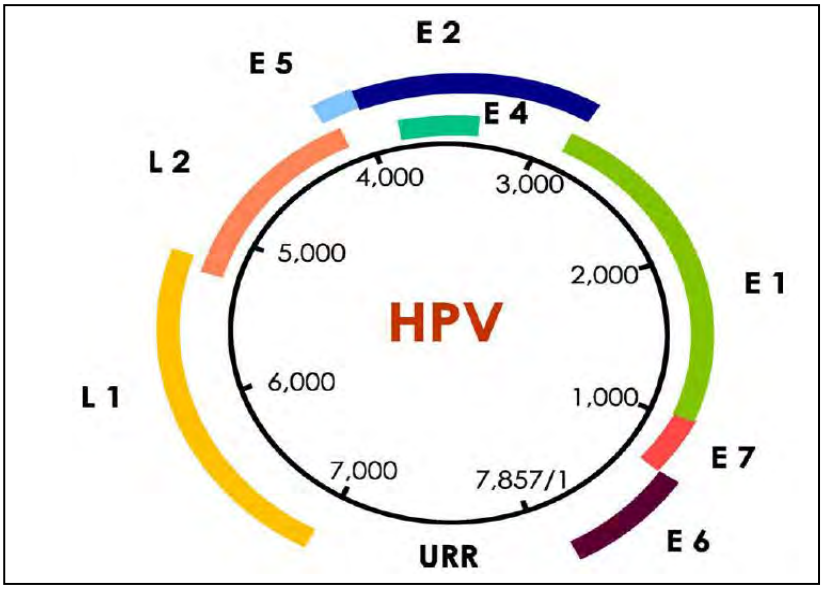


Figure 1.5: Genomic organization of the HPV genome. Schematic representation of the HPV circular genome showing the location and arrangement of the early (E) and late genes (L1 and L2) as well as the upstream regulatory region (URR) adapted from (Munoz et al., 2006).

1.4.3 High and low risk HPV infections

The continuous expression of both the HPV E6 and E7 protein is required for the maintenance of cervical cancers, with the two viral proteins exerting distinct effects on cell survival, proliferation and integration and rearrangement of foreign DNA (Kessis et al., 1996; Duensing & Munger, 2002; DeFilippis et al., 2003; Jabbar et al., 2012). The high-risk E6 proteins have the ability to form a tripartite complex with p53 and the cellular ubiquitin ligase E6-associated protein (E6AP) leading to proteosome-mediated p53 degradation (Huibregtse et al., 1993a). The Low-risk E6 proteins on the other hand binds p53 with a lower affinity than the highrisk types and has no significant ability to bind E6AP to stimulate p53 degradation (Storey et al., 1998). This loss of the p53-mediated DNA damage response in cells expressing high-risk HPV E6 predisposes the host cell to cancer progression.

The high-risk E6 proteins also have a C-terminal PDZ-binding domain and are capable of binding and stimulating the degradation of several cellular targets that contain PDZ motifs, such as human homologue of Dlg (hDlg) and human homologue of the Drosophila scribble protein (hSrib), that are thought to be involved in the regulation of cell growth and attachment (Zeitler et al., 2004). PDZ binding is thought to be a conserved feature of all high-risk E6 proteins, and it has been implicated in the loss of cell polarity seen in HPV associated cervical cancers. It has also been shown to be necessary for the stimulation of epithelial hyperplasia in transgenic animals (Nakagawa & Huibregtse, 2000; Nguyen et al., 2003b).

There are discordant observations that both the high-risk and low-risk forms of E6 are capable of binding and inactivating p73 (Marin et al., 1998; Park & Androphy, 2002). In addition, high-risk E6s are also capable of activating NF-κB, although the mechanism is not known there is evidence that it may depend upon the PDZ binding motif (Nees et al., 2001; Havard et al., 2005; Yuan et al., 2005; James et al., 2006; D'Costa et al., 2012). The E6 proteins from both low- and high-risk mucosal types are also thought to be able to inhibit the interferon response causing the down-regulation of multiple interferon responsive genes (Nees et al., 2001).

1.4.4 Molecular mechanisms of HPV induced carcinogenesis

1.4.4.1 HPV induced Genomic instability

Genomic instability has been observed in multiple epithelial cells expressing the E6 and E7 protein (Smith et al., 1989; Hashida & Yasumoto, 1991; Reznikoff et al., 1994; White et al.,

1994; Steenbergen et al., 1998; Duensing & Munger, 2002). Although the E6 and E7 proteins are necessary for maintenance of the transformed phenotype, they are incapable of directly transforming cells. The High-risk E6 and E7 are thought to independently induce genomic instability in normal cells, which is a characteristic of high-risk HPV-induced malignancies (White et al., 1994). The ability of E6 to induce genomic instability reflects its ability to inhibit the function of p53 (Havre et al., 1995), that leads to the disruption of normal DNA repair processes resulting in genetic change. The genomic instability induced by E7 reflects its effect on centrosome biogenesis and the consequent defects in segregation of daughter chromosomes during cell division while the induction of genetic instability is thought to be an early event in HPV-induced cancers, occuring before integration of the virus into host chromosomes (Duensing et al., 2001). These events are not limited to high-risk E6 and E7 proteins but are also seen in cells expressing their low-risk counterparts (Rihet et al., 1996).

1.4.4.2 HPV induced DNA damage response

E6 and E7 have been shown to independently induce DNA damage and increase the frequency of foreign DNA integration into the host genome (Kessis et al., 1996; Duensing & Munger, 2002). This is thought to reflect the activity of both E6 and E7 in inhibiting p53-mediated cell-cycle arrest and correlates with the ability of E6 to bind and inactivate p53 and E7 ability to not only disrupt the function of the cell-cycle regulator pRb but also to inactivate p21 (Funk et al., 1997; Song et al., 1998). Abrogation of the DNA damage responses may contribute to the accumulation of genetic alterations leading to malignant progression (Moody & Laimins, 2010).

1.4.4.3 HPV induced Immortalization

The E6 and E7 proteins of mucosal high-risk HPVs, which act independently or synergistically, have transforming activity in tissue culture and are capable of immortalizing multiple cell types including cervical, human foreskin keratinocytes and mammary epithelial cells (Hawley-Nelson et al., 1989; Hudson et al., 1990; Halbert et al., 1991; Wazer et al., 1995). The E5 protein can also enhance the immortalization of keratinocytes in combination with both E6 and E7 (Stöppler et al., 1996). The immortalization potential of E6 in mammary epithelial cells and keratinocytes has not only been linked to its inactivation of p53 but also to its ability to induce the expression of telomerase (Dalal et al., 1996; Kiyono et al., 1998; McMurray & McCance, 2004). The use of K14E6/E7 transgenic mouse model has also

shown the persistence of cervical cancers in the cervicovaginal junction and outer cervix than in the cervix proper following exogenous estrogen treatment and withdrawal. Thus, suggesting that the outer cervix and cervicovaginal region may be more proliferative and more sensitive to estrogen than the cervix proper, explaining the rapid onset of recurrent disease (Spurgeon et al., 2014).

1.4.4.4 HPV induced Cell Proliferation and differentiation

HPV infections induce cell proliferation within the differentiating layers of the epithelium that are otherwise devoid of replicating cells (Zur Hausen, 2002). High risk E6 has been shown to induce suprabasal DNA synthesis, a p53-independent activity that correlates with its ability to bind PDZ-domain proteins (Song et al., 1999; Nguyen et al., 2003a). Depending on the titre and nature of the infecting papillomavirus type, the time between initial infection and the appearance of productive papillomas can vary with latency occurring when inoculating titres are low (Campo, 1995; Zhang et al., 1999). E6 PDZ binding is also thought to mediate suprabasal cell proliferation (Nguyen et al., 2003a; Nguyen et al., 2003b) and may contribute to the development of metastatic tumours by disrupting normal cell adhesion.

1.4.4.5 HPV induced Immune Response

Interferon (IFN) is activated following viral infection and HPV proteins act at several levels to interfere with this response (Beglin et al., 2009). In cells in which HPV has integrated, both the E6 and E7 are expressed at higher levels and have been implicated in causing resistance to the effects of interferon (Nees et al., 2001; Beglin et al., 2009). Both the low and high risk E6 proteins are capable of inhibiting the interferon response causing the down regulation of multiple interferon responsive genes (Nees et al., 2001). The high risk E6 and E7 repress *STAT1*, a key transcription factor that regulates the IFN response with the high risk E6 directly binding to the IFN regulatory factor 3 (IRF3) that inhibits transactivation of IFN transcription (Ronco et al., 1998). They both also target the double-stranded RNA Protein Kinase R (PKR) inhibiting protein synthesis, through phosphorylation of the eIF2a proteins. E6 in addition blocks PKR kinase activity by relocalizing it to cytoplasmic P-bodies, which are sites of mRNA storage and degradation (Hebner et al., 2006).

1.4.4.6 HPV induced Apoptosis

High-risk HPV E6 proteins are capable of extending the proliferative capacity of infected cells by inhibiting p53 dependent growth and blocking apoptosis in response to aberrant proliferation through several mechanisms, resulting in the induction of genomic instability and the accumulation of cellular mutations (Moody & Laimins, 2010). They also inhibit apoptotic signaling in response to growth suppressive cytokines through interaction with the tumour necrosis factor (TNF)-α receptor TNFR1, FAS-associated protein with death domain (FADD) and caspase 8 (Filippova et al., 2002; Filippova et al., 2004; Garnett et al., 2006). Both the low and high risk mucosal E6s can also bind to the pro-apoptotic proteins Bax and Bak leading to their degradation (Thomas & Banks, 1999). By utilizing the apoptotic signalling pathways high risk E6 are able to promote replication in differentiating cells (Moody et al., 2007).

1.4.5 Association of HPV in OSCC

Following the observation of characteristic cytopathic changes of HPV infection usually seen in condylomatous lesions, in both benign oesophageal epithelial tissue and malignant oesophageal tumours, a link between HPV and OSCC was made (Syrjanen et al., 1982). An increase in OSCC has also been observed in HIV-infected individuals (Rastegar, 2010; Stebbing et al., 2010), while HPV infection has been found in chronic oesophageal irritation and oesophageal cancer (Syrjanen, 1982; Chang, 1990; Chang et al., 1992). The infection of the oesophagus by HPV was also confirmed by immunohistochemical techniques demonstrating HPV structural antigens L1 and L2 in squamous cell papilloma (Hille et al., 1986; Hale et al., 1989). The HPV DNA sequences in both benign and malignant oesophageal lesions have also being disclosed through DNA hybridization studies and polymerase chain reaction techniques (Syrjanen, 1982; Winkler et al., 1985; Kulski et al., 1986; Chang et al., 1990; Benamouzig et al., 1992).

A clear trend exists between HPV detection rates and geographical area corresponding to OSCC incidence rates, with studies conducted in high-risk regions for OSCC incidence reporting higher rates of HPV in OSCC specimens than studies carried out in low-risk regions (Matsha et al., 2007; Antonsson et al., 2010; Zhang et al., 2010; Cui et al., 2011; Malik et al., 2011). HPV infection rates have also been found to be higher in developing countries than in developed nations (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 1995). HPV positivity prevalence in oesophageal cancer globally has ranged from

0–71% in oesophageal tumour samples examined (Sur & Cooper, 1998) with the wide range being attributed to differences in the sensitivity of detection methods used and in the tumour samples examined (Hendricks & Parker, 2002).

HPV was also detected in 26–71% of oesophageal squamous carcinoma samples from South African patients with the low-risk HPV types 6,11 and 39 being reported more frequently in oesophageal tumour tissues from high-risk regions than low-risk areas (Williamson et al., 1991; Cooper et al., 1995; Lavergne & de Villiers, 1999; Matsha et al., 2002; Matsha et al., 2006; Liyanage et al., 2013b; Schafer et al., 2013). A meta-analysis of studies demonstrated that HPV increased the risk of OSCC by three-fold thus providing the strongest evidence of a potential role for HPV in the aetiology of OSCC (Liyanage et al., 2013a).

1.4.6 The HPV E6 Protein

HPV E6 is a cysteine-rich protein of approximately 150 amino acids whose major structural characteristic is the presence of two zinc-binding regions, referred to as E6N and E6C. Both domains contain two cysteine motifs Cys-X-X-Cys that are conserved in the E6 proteins of all HPV types (Tommasino, 2014). Although it is not clear whether the Cys-X-X-Cys motifs are important for the ability of E6 to bind DNA non-specifically, the importance of these motifs for HPV E6 proteins has been implicated in transformation, transcriptional activation, immortalization and association with cellular proteins (Mallon et al., 1987; Grossman et al., 1989; Vousden et al., 1989; Kanda et al., 1991; Foster et al., 1994; Rapp & Chen, 1998) see Fig. 1.6.

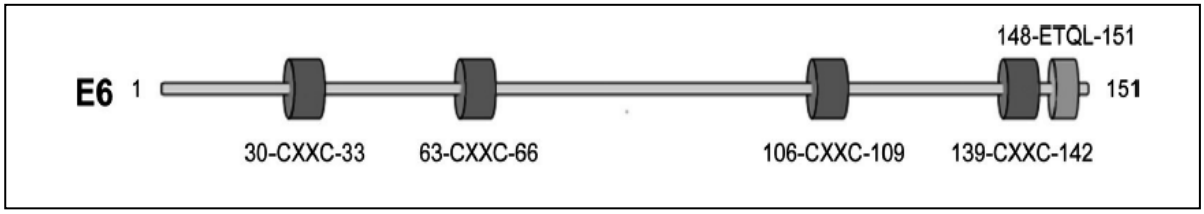


Figure 1.6: Basic structure of HPVE6 Protein. Schematic representation of HPV16 E6 showing the position of amino acids and the CXXC motifs that form zinc complexes important for the E6 structure and function. The consensus PDZ-binding motif (ETQL) is shown positioned at the C terminus, adapted from (Tommasino, 2014).

The E6 protein is regarded as a multifunctional protein in that it acts as a transcriptional transactivator by being coupled to its ability to transform and immortalize cells. In addition, the high risk E6 is capable of inducing immortalization of primary cell cultures, while the low risk E6 augments the ability of E7 to immortalize cells. Furthermore, the E6 modulates normal cellular processes by binding to cellular proteins E6AP and E6BP. The E6-E6AP complex can bind *in vitro* to p53 (Huibregtse et al., 1993a) although the deregulation of the

p53 growth suppressor function appears to be separate from the E6 involvement in the degradation of p53 (Lechner et al., 1992; Lechner & Laimins, 1994). The E6 also targets the degradation of p53 (Hubbert et al., 1992; Scheffner et al., 1992; Band et al., 1993; Foster et al., 1994; Mayers & Androphy, 1995) in a process that requires a ternary complex between E6, E6AP and p53 (Scheffner et al., 1994), while the low-risk HPVs, such as HPV-6 and HPV-11, do not have the capacity to degrade p53.The High risk HPVE6 and to a lesser extent low risk HPVE6 is also able to inhibit the expression of p53-regulated genes by directly interacting with the transcriptional co-activators CBP and p300 (Patel et al., 1999).

Both the low risk and high risk E6 are capable of inhibiting the apoptotic response via the degradation of Bak, a member of the Bcl-2 family, through the interaction with E6AP and the ubiquitin proteasome pathway (Thomas & Banks, 1998; Jackson et al., 2000; Underbrink et al., 2008). Furthermore, high risk E6 activates host-cell telomerase, which might be associated with the loss of alleles in chromosomes 3p and 10p in-turn representing a late event in the steps towards immortalization (Klingelhutz et al., 1996; Steenbergen et al., 1996). The high risk E6 also induces hTERT transcription during cell immortalization through the maintenance of telomerase activity (Liu et al., 2008).

HPV E6 are capable of regulating the expression of miRNAs in cells (Martinez et al., 2008; Wang et al., 2009; McKenna et al., 2010; Wald et al., 2011). High risk HPVE6 down regulates both miR-34a which is involved in targeting cell cycle control genes (Wang et al., 2009) and miR-218 which is involved in regulating the expression of LAMB3 (Martinez et al., 2008), that may play a role in cervical cancer cell growth. High risk HPVE6 also down regulates miR-23b which may also be important for regulating cell migration by causing the up regulation of urokinase plasminogen activator gene (Yeung et al., 2011). It is likely that HPVE6 regulates miRNAs through its interaction with transcriptional factors and signaling proteins (Vande Pol & Klingelhutz, 2013).

1.5 Genomic Alterations in Cancer

Cancers arise as a result of changes having occurred in the DNA sequence of cancer cells genome (Yang et al., 2010). Almost all cancer genomes contain nucleotide sequence changes compared with the germline of the cancer patient (Vogelstein & Kinzler, 1993). These variations include the genomic alterations that promote cancer, termed drivers and alterations present in the cancer genome that do not promote cancer, referred to as passengers (Davies et al., 2005; Stratton et al., 2009). It has been shown that the accumulation of genetic variations

can give rise to tumourigenesis in three types of genes namely stability genes, oncogenes and tumour-suppressor genes (Vogelstein & Kinzler, 2004). The human cells have been known to protect themselves against any lethal effects of cancer causing genetic mutations however; it is the defectiveness of a subset of genes that give rise to the development of cancer (Vogelstein & Kinzler, 2004; Yeang et al., 2008). Thus, the accumulation of genetic alterations in the tumour suppressor genes and oncogenes is thought to play a key role in cancer development (Nowell, 1976; Donahue et al., 2006). The most common genomic alterations in the cancer genome include nucleotide substitution mutations, Insertion/Deletions (InDels), copy number variations (CNV), chromosomal rearrangements and nucleic acids of oncogenic viruses (Weir et al., 2004; Chin & Gray, 2008; Stratton et al., 2009). All these acquired changes may influence the rate of occurrence of genetic alterations (Balmain et al., 2003). Given that InDels are more likely to be drivers of mutation, a substantial number of InDels in the coding sequence (CDS) of cancer genes is not surprising , since Indels are often deleterious as evidenced by their frequent association with human disease (Zoghbi & Orr, 2000; Kondrashov & Rogozin, 2004; Yang et al., 2010). InDels have also been shown to be mutational components of gene and pseudogene evolution that are important in the long-term evolution of genome size (Petrov et al., 2000; Nedelcu, 2001; Gregory, 2004). In addition, InDels have also been associated with apoptosis mechanisms, phenotypic change and activity of surface antigen (Burch et al., 1997; Ferro & Pfeffer, 2001; Rocha & Blanchard, 2002). Thus, the abundance of InDels, which have a comparatively greater influence than SNPs, is consistent with the abnormal function and devastating characteristics of cancer (Yang et al., 2010).

1. 5.1 Genomic Alterations in OSCC

Genomic alterations contribute to human tumorigenesis by altering the expression levels of critical oncogenes and tumour suppressor genes (TSG) (Ying et al., 2012). The understanding of the genomic abnormalities in OSCC is limited to studies of small cohorts, hence the need to extensively identify genomic abnormalities underlying OSCC in-order to elucidate it's molecular basis and guide the development of effective targeted therapies (Luo et al., 2006; Hirasaki et al., 2007; Bass et al., 2009; Hu et al., 2009; Beroukhim et al., 2010; Hu et al., 2010).

Previous reports have shown OSCC to have a higher somatic mutational rate to breast carcinoma and glioblastoma multiforme, but less than head and neck squamous cell

carcinoma (HNSCC), oesophageal adenocarcinoma (OAC) and lung squamous cell carcinoma (SQCC) (Stransky et al., 2011; Agrawal et al., 2012; Cancer Genome Atlas Research Network, 2012; Dulak et al., 2013). The transitions (Ti) were the most common mutations followed by transversions (Tv) in OSCC (Song et al., 2014). Compared with OAC, HNSCC and lung (SQCC) the overall pattern of the mutation spectrum in OSCC was similar to that of HNSCC but different from that of lung SQCC which is dominated by transversions (Stransky et al., 2011; Dulak et al., 2013).

Sequencing of OSCC samples has identified several recurrently mutated OSCC relevant genes and led to the conclusion that OSCC cases from different geographic regions might differ genetically (Agrawal et al., 2012; Gao et al., 2014; Song et al., 2014). Some of the most frequent copy number variations that might contribute to OSCC carcinogenesis involved 3q26, 4q, 5p, 8q24.3, 7q, 9p21, 10q21, 11q13.3, 18p11.3 (Yang et al., 2000; Janssen et al., 2002; Miller et al., 2003; Luo et al., 2006; Law et al., 2007; Yang et al., 2008; Bass et al., 2009; Hu et al., 2009; Lin et al., 2014).While the regions known to have high level deletions containing candidate TSGs were; 4q21.23-21.3, 8p22, 10p11, 13q31.1, 14q32, 16q22-23, 18q11-23 (Sun et al., 2005; Ying et al., 2006; Ying et al., 2012).

Cancer genes known to be frequently mutated in OSCC clustered in several cancer pathways including cell cycle, apoptosis and DNA damage control pathways (Vogelstein et al., 2013). Eight significantly mutated genes have been reported in OSCC, six of which are well known OSCC implicated genes (*TP53, RB1, CDKN2A, PIK3CA, NOTCH1* and *NFE2L2*). The *ADAM29* gene has not been previously described in OSCC but is frequently mutated in melanoma and *FAM135B* gene that has not been previously linked to cancer (Wei et al., 2011; Song et al., 2014). Other genes which have also been frequently mutated on OSCC include; *MSH3, ADH4, EGFR, ERBB4* and *PTEN* (Han et al., 1998; Chou et al., 2002; Yin et al., 2003; Matsumoto et al., 2005; Guo et al., 2006; Hanawa et al., 2006; Lee et al., 2006; Hiyama et al., 2007; Mir et al., 2008; Vogelsang et al., 2012; Vogelsang et al., 2014).

A recent study aimed at creating a comprehensive catalogue of cancer genes revealed 33 novel candidate genes related to immune evasion, cell proliferation, protein homeostasis, genome stability, chromatin regulation, RNA processing and apoptosis. The loss-of-function mutations in these genes would be expected to contribute to oncogenesis, some of these genes included *RHEB, PCBP1, SOS1, ELF3, TRIM23, ALPK2* and *BCLAF1* (Lawrence et al., 2014).

The analyses of genomic alterations in multiple tumour types has led to two observations: (i) there is a substantial difference in genomic alterations from tumours originating in the same tissue (Cancer Genome Atlas Network, 2012) and (ii) similar patterns of genomic alteration are observed in tumours from different tissues of origin (Cancer Genome Atlas Research Network, 2013). These phenomena of cross cancer similarity and intracancer heterogeneity represents an opportunity in designing new therapeutic protocols based on the genomic traits of tumours (Ciriello et al., 2013; Garraway et al., 2013; Garraway, 2013).

1.5.2 Viral Integration in Cancer

Viruses have been shown to cause cellular transformation by the expression of viral oncogenes and genomic integration in altering the activity of cellular tumour suppressors and by inducing inflammation that promotes oncogenesis (Tang et al., 2013; Lin et al., 2014). Although human oncogenic viruses belong to different virus families and utilize diverse strategies to contribute to cancer development, they have been shown to share many common features. These have included their ability to establish long-term persistent infections while evading the host immune response (McLaughlin-Drubin & Munger, 2008). Despite the viral aetiology of several cancers, viruses may often contribute to, but are thought not to be sufficient for carcinogenesis and other co-factors such as chronic inflammation, host cellular mutations and the host immunity must be involved in the transformation process. Thus, a key feature of the oncogenic viruses is the long-term interactions between virus and host as this leads to molecular events that may enable the eventual virus mediated carcinogenesis (Parsonnet, 1999).

Six human viruses have been considered by the International Agency for Research on Cancer (IARC) as being carcinogenic based on sufficient evidence supporting their aetiologic association with human cancers (burkitt lymphoma, melanoma, cancer of the conjunctiva, OSCC, cervical cancer, anogenital cancers and head and neck cancers). These are Epstein-Barr virus, Hepatitis B Virus(HBV), several types of human papilloma viruses (HPV), human T-cell lymphotropic virus type 1, Hepatitis C Virus (HCV), and Kaposi's sarcoma-associated herpesvirus (Parkin, 2006). Other viruses with potential roles in human malignancies include; Human endogenous retrovirus (HERV)-K113 and K115, human mammary tumor virus (HMTV), Torque teno virus (TTV) and Merkel cell polyomavirus (MCPyV) (Nishizawa et al., 1997; Gifford & Tristem, 2003; Burmeister et al., 2004; Holland & Pogo, 2004; Kremsdorf et al., 2006; Park et al., 2006; Lee & Lee, 2007; Houben et al., 2010; Kofman et

al., 2011; Banks et al., 2012; Zwolińska et al., 2013; Hampras et al., 2014; Miner et al., 2014). However, although viruses are widespread in the general population, only a small fraction of infected people develop neoplastic disease, indicating the crucial role of some presently unknown genetic factors (Mazzaro et al., 2005). Some viruses have been known to integrate into the host genome thereby interrupting gene functions or inducing chromosomal instability, while other viruses rarely integrate into a host genome. Thus, detecting the existence of viruses and their integration sites in host genomes is paramount in understanding their molecular mechanisms in disease development (Feitelson & Lee, 2007; Lace et al., 2011; Jiang et al., 2012; McLaughlin-Drubin et al., 2012; Sung et al., 2012; Arzumanyan et al., 2013; Lasithiotaki et al., 2013; Panagiotakis et al., 2013; Pascale et al., 2013).

Next Generation Sequencing (NGS), involving the use of advanced bioinformatic analysis pipelines has allowed for the unbiased discovery of known as well as novel viruses integrated within several human tissue types (Feng et al., 2008; Daly et al., 2011; Yu et al., 2012; Chiu, 2013; DeBoever et al., 2013; Khoury et al., 2013; Naeem et al., 2013; Schelhorn et al., 2013; Zhao et al., 2013). A perfect example of the NGS approach has been in the determination of the full viral structure and human integration sites of MCPyV (Duncavage et al., 2011). The use of unmapped (nonhuman) sequencing reads from NGS of cancer specimens has enabled the identification of viral integration sites (Samuels et al., 2013). This was shown during alignment, in the case of paired-end read data a good indicator for a possible integration site is when discordant pairs can be detected whereby one read is aligned to the viral genome and its mate is aligned to the human genome (Jiang et al., 2012; Chen et al., 2013; Cimino et al., 2014). Thus, integration of the viral genome into the host cell chromosome suggests that persistent viral infection is vital for malignant cell transformation and carcinogenesis (Zhang et al., 2011; Zhang et al., 2011).

1.6 Objective of this study

To determine the role of viral sequences on the initiation of cellular transformation and the identification of genomic alterations associated with malignant transformation.

1.6.1 Aims of this study

This study aims to better understand the aetiology and molecular mechanisms involved in the development of OSCC. This will be achieved in three ways;

1. Determining the role and effects of HPV-11E6 and HPV-18E6 in the transformation of HaCaT cells from the normal to the malignant state.
2. Evaluating and comparing the effects of both HPV11E6 and HPV18E6 on gene expression patterns during the early stages of transformation.
3. Identifying structural variants, primarily large structural variants (insertions and deletions) and integrated viral sequences that may be associated with oesophageal squamous cell carcinoma (OSCC).

CHAPTER TWO
Expression of HPVE6 in Human Keratinocytes

2.1 Introduction

Human papillomaviruses (HPV) have been implicated as causative agents in a variety of human squamous cell carcinomas, including those of the skin, cervix, anogenital region, upper respiratory tract, and digestive track (Syrjänen et al., 1982; Matsha et al., 2002). Currently nearly 150 different HPV types are recognized of which 120 have been fully sequenced, some of which are frequently associated with cancers and are considered high risk HPV (types 16 and 18) whereas others give rise to warts and benign lesions and are considered low risk (types 6, 11, and 33) (Matsha et al., 2002; Rautava & Syrjänen, 2012).

Much basic and clinical research has focused on the high risk HPV types than on the low risk types despite the substantial financial burden imposed by low risk HPV types on the health systems (Hu & Goldie, 2008). Although most studies are also usually performed using either the E6 protein from the high risk HPV's or most frequently involve concurrent expression of both the E6 and E7oncogenes the molecular mechanisms by which the low risk HPV E6 protein may contribute to malignant progression are not fully understood (Reznikoff et al., 1994; Zhang et al., 2007; Pim & Banks, 2010).

The high risk HPV E6 and E7 function as oncoproteins, with the primary activity of the high-risk form of E6 being to target *p53* for degradation, via complex formation with the cellular ubiquitin ligase E6-AP (Scheffner et al., 1990; Werness et al., 1990; Huibregtse et al., 1991; Huibregtse et al., 1993a; Laimins, 1993; Scheffner et al., 1993; Scheffner et al., 1994; Huibregtse & Beaudenon, 1996; Kubbutat & Vousden, 1996; Mantovani & Banks, 1999). The inactivation of *p53* in turn compromises the integrity of the cellular genome and lead to DNA damage and chromosomal instability, resulting in increased cell proliferation and tumour development (Schaeffer et al., 2004; Cheng et al., 2007; Cooper et al., 2007). As a consequence of the E6-mediated *p53* inactivation, *p21* gene transcription is also inhibited

(Cho et al., 2002). Although degradation of p53 by E6 is specific to high risk HPV types, the low risk forms of E6 may also bind to p53 although at a low affinity (Crook et al., 1991; Li & Coffino, 1996). On the other hand the high risk E7 acts by binding to members of the Rb tumour suppressor protein family and inhibits their ability to modulate the function of E2F transcription factors while the low-risk E7 proteins bind to Rb with a 10-fold lower affinity (Dyson et al., 1989; Munger et al., 1989; Heck et al., 1992).

Various cell types have been used in investigating the underlying mechanisms of HPV induced malignant progression such as the prostate gland (Weijerman et al., 1994), cervix (Woodworth et al., 1988) and ovaries (Tsao et al., 1995). The normal human keratinocyte (HaCaT) cell line offers a suitable model to study regulatory mechanisms in the differentiation of human epidermal cells and provides a valuable model system for the study of the role of oncogenes and other factors in the process of malignant transformation of human epithelial cells (Boukamp et al., 1988).

A major barrier to understanding the progression from initial infection to cellular transformation has been the lack of *in vitro* models that allow infection, replication and persistence of the viral genome in epithelial cells (Lee et al., 2004). The use of recombinant adenoviral vectors has been shown to hold great promise for gene transfer in basic research as well as for clinical treatment of many diseases (Kozarsky & Wilson, 1993; Kay & Woo, 1994). They have been shown to be able to transduce foreign genes efficiently into both cultured cells and many target organs *in vivo*, without its viral genome undergoing frequent rearrangements. Futhermore, the inserted foreign genes are generally maintained without change through successive rounds of viral replication (Mizuguchi & Kay, 1998; Vorburger & Hunt, 2002).

This chapter describes the construction of a recombinant adenovirus capable of high level transient expression of the biologically active HPV E6 from the low risk HPV11 and high risk HPV18 and to compare the roles and effects of these two HPVE6 proteins in inducing cellular transformation on normal human keratinocyte (HaCaT) cells.

2.2 Results

2.2.1 Construction of Recombinant Adenovirus Vector

The assembly and production of recombinant adenovirus was accomplished in three stages (Fig. 2.1). Firstly a mammalian expression cassette with E6 constructs from either HPV11 or

HPV18 was excised from pcDNA3.1 and cloned into a pShuttle2 vector. After amplification in *Escherichia Coli DH5α*, the expression cassette was excised from pShuttle2 and ligated to Adeno-X Viral DNA (the adenoviral genome). Lastly, the recombinant Adeno-X vector was packaged into infectious adenovirus by transfecting human embryonic kidney (HEK) 293 cells and harvested by lysing the transfected cells. HaCaT cell lines were then infected with the recombinant adenovirus inorder to transiently express the gene of interest as described in sections 6.12.

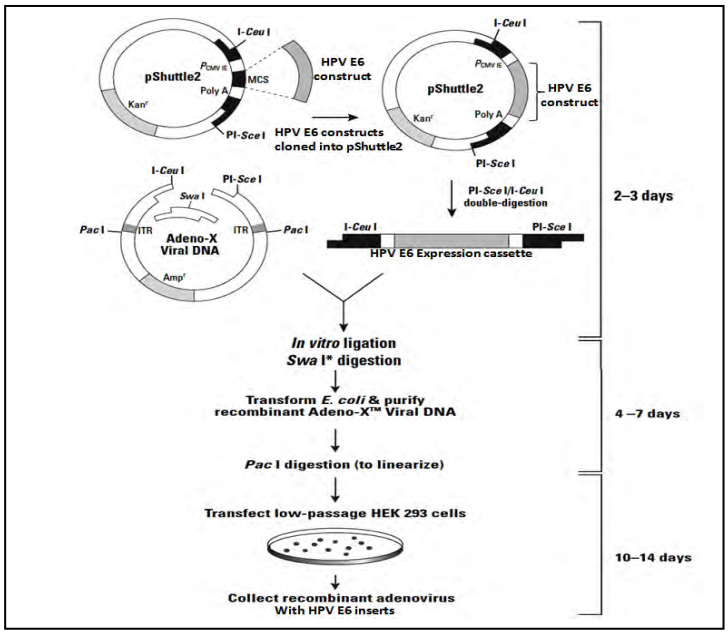


Fig. 2.1 Construction of recombinant adenovirus. The pShuttle2 allowed for the cloning of HPVE6 constructs into the mammalian expression cassette, which consisted of the human cytomegalovirus immediate early promoter/enhancer ($P_{CMV\ IE}$), a multiple cloning site (MCS), and the SV40 polyadenylation signal (SV40 poly A). Following amplification in *Escherichia Coli* DH5α the expression cassette which was flanked by unique I-*Ceu* I and PI-*Sce* I restriction sites was excised and then "shuttled" into Adeno-X Viral DNA by *in vitro* ligation. The products were then treated with *Swa*I to linearize non-recombinant Adeno-HPVE6 DNA and amplified in *Escherichia Coli.* DNA was isolated and digested with *Pac*I to expose the inverted terminal repeats (ITRs) located at either end of the genome. Finally the recombinant Adeno-HPVE6 constructs were packaged into infectious adenovirus by transfecting human embryonic kidney (HEK) 293 cells and harvested by lysing the transfected cells (Clontech Expression, 2012).

2.2.1.1 Preparation of Ultra-Competent *Escherichia Coli* DH5α Strain

The method of Inoue et al., (1990) was used for the preparation of ultracompetent *Escherichia col*i DH5α, with a slight modification of the Hanahan, (1983) technique. This method offers a simple and efficient method for plasmid transfection, enabling improved transformation efficiencies when cloning large plasmids with an extremely high transforming frequency (1-3 x 10^9 cfu/ug of plasmid DNA) as described in section 6.1. Transformation efficiency was defined as the number of colony forming unit (cfu) produced by 1 µg of plasmid DNA, and was measured by performing a control transformation reaction using a

known quantity of DNA, then calculating the number of cfu formed per microgram of DNA (Tu et al., 2005).

2.2.1.2 Construction of Recombinant pShuttle2

pcDNA3.1 vector containing the E6 genes from either HPV11 or HPV18 were kindly donated by Dr. Lawrence Banks (ICGEB Trieste). The E6 inserts were released by digestion with *Nhe*I and *Not*I and cloned into the multiple cloning site (MCS) of the pShuttle2 vector as shown in Fig. 2.1. The pShuttle2 vector was then propagated in *Escherichia Coli DH5α*, and the transformation efficiencies of the pShuttle calculated as shown in Table 2.1.

Table 2.1: The transformation efficiency of respective pShuttle2 per microgram DNA

Plasmid	DNA Conc.	CFU/Plate	Dilution factor	Transformation efficiency
pShuttle2 11E6	0.35ng/ul	444	1000	$1.3x10^9$
pShuttle2 18E6	0.27ng/ul	292	1000	$1.1x10^9$

2.2.1.2.1 Confirmation of E6 inserts sequence identities

After the recombinant plasmids had been propagated in *Escherichia Coli DH5α*, they were isolated as described in section 6.2 and sequenced using the universal T7 primers. A BLASTn search showed 100% sequence identity to the HPV11E6 (Accession number: FN870475.1) and HPV18E6 (Accession number: EF422110.1) sequences on the NCBI nucleotide database.

2.2.1.2.2 Release of the E6 inserts from the recombinant pShuttle2

The restriction endonucleases *PI-Sce*I / *I-Ceu*I were used to excise the newly constructed expression cassette from the recombinant pShuttle2 E6 constructs (Fig. 2.1), and were then "shuttled" into Adeno-X viral DNA by *in vitro* ligation, to generate a circular recombinant E1/E3-deleted adenoviral genome that carries a ColE1 origin of replication and an ampicillin resistance marker for propagation and selection in *Escherichia Coli DH5α. Swa*I digestion was performed to linearize any non-recombinant (i.e. self-ligated) pAdeno-X DNA.

2.2.1.3 Analysis of Recombinant Adeno-HPVE6 Constructs

The identity of the recombinant pAdeno-HPV (large plasmid >32kb) E6 constructs were determined by restriction enzyme digestion and PCR analysis. In restriction analysis the presence of the expression cassette was verified by digestion with *Xho*I and analysis done on

1 % agarose gels (Fig. 2.2A). The pAdeno-HPVE6 constructs were screened for the presence of pShuttle2 derived expression cassettes using the Adeno-X forward and reverse PCR primers, to specifically amplify a 287-bp (Fig. 2.2B) fragment that spanned the *ICeu*I ligation site in pAdeno-X (Fig. 2.1). Only recombinant pAdeno-HPVE6 templates were amplified since non-recombinants lacked the shuttle sequence needed for annealing with the reverse primer. The E6 constructs were also verified using the HPV11E6 and HPV18E6 specific primers (Fig. 2.2C) designed to give different size fragments for the two genes.

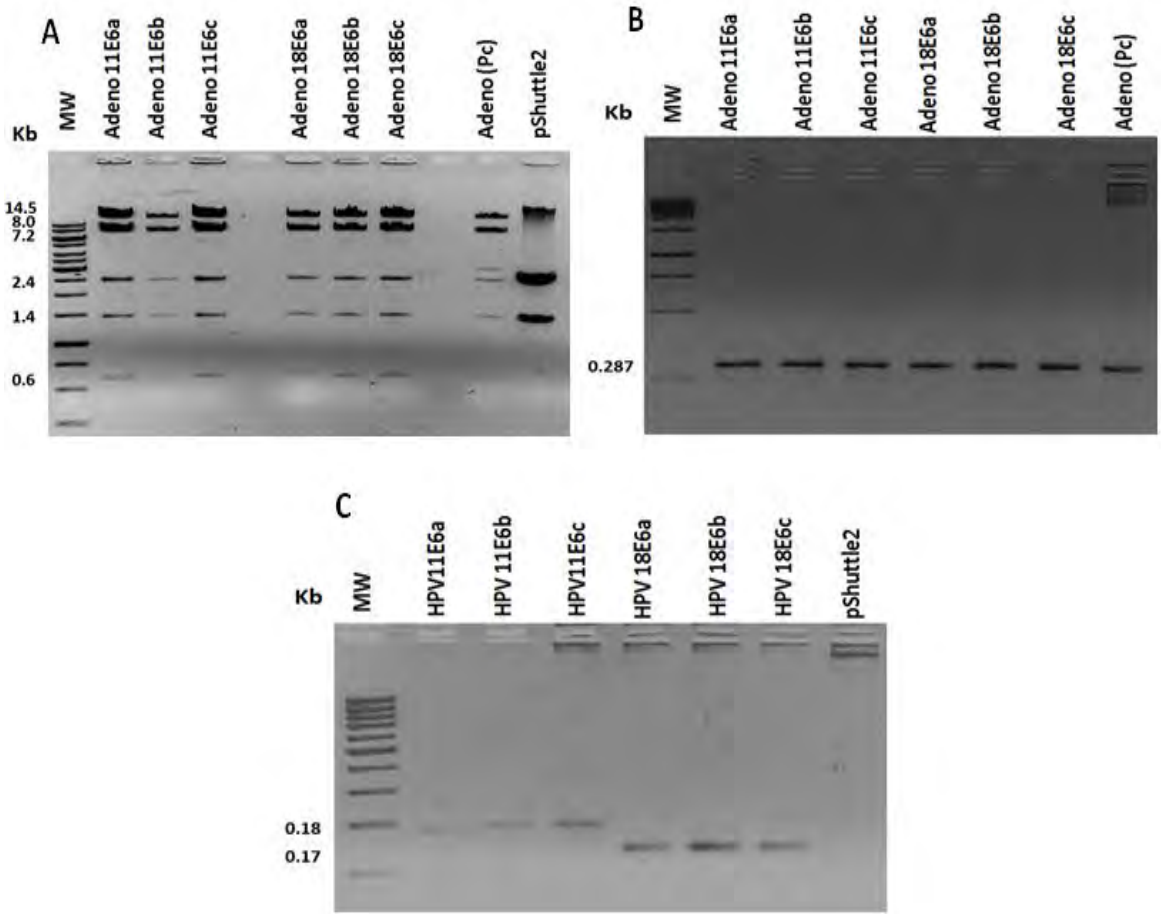


Fig. 2.2 Confirmation of HPVE6 Constructs. **A:** Verification of expression cassette following digestion with *Xho*I of both the HPVE6 constructs an Adeno-X positive control (Pc) and the pShuttle2 vector negative control (Nc). **B:** Verification of recombinant Adeno-HPVE6 constructs by PCR with the Adeno-X forward and reverse PCR primers which specifically amplified a 287-bp sequence that spanned the *ICeu*I ligation site. **C:** Verification of HPV E6 constructs using HPV11E6 and HPV18E6 specific primers with pShuttle vector as a negative control (Nc). In each case three different clones of each construct were selected for analysis (a, b, c).

2.2.1.4 Propagation and purification of recombinant AdenoHPVE6

Before Adeno-HPVE6 constructs could be packaged, the recombinant plasmids were digested with *Pac*I to expose the inverted terminal repeats (ITRs) located at either end of the genome (Fig. 2.1). The ITRs contain the origins of adenovirus DNA replication and had to be

positioned at the termini of the linear Adenovirus DNA molecule to support the formation of the replication complex (Tamanoi & Stillman, 1982). The recombinant Adeno-HPVE6 constructs were then packaged into infectious adenovirus by transfecting human embryonic kidney (HEK) 293 cells (Fig. 2.3A). After overnight incubation the cells displayed a 50% cytopathic effect (CPE) (Fig. 2.3B) and were harvested by lysing with three consecutive freeze thaw cycles to prepare a primary adenoviral stock. The titer of the primary adenoviral amplification stock (Fig. 2.4) was determined using the End-point dilution assay to measure the number of infectious viral particles as described in section 6.7.

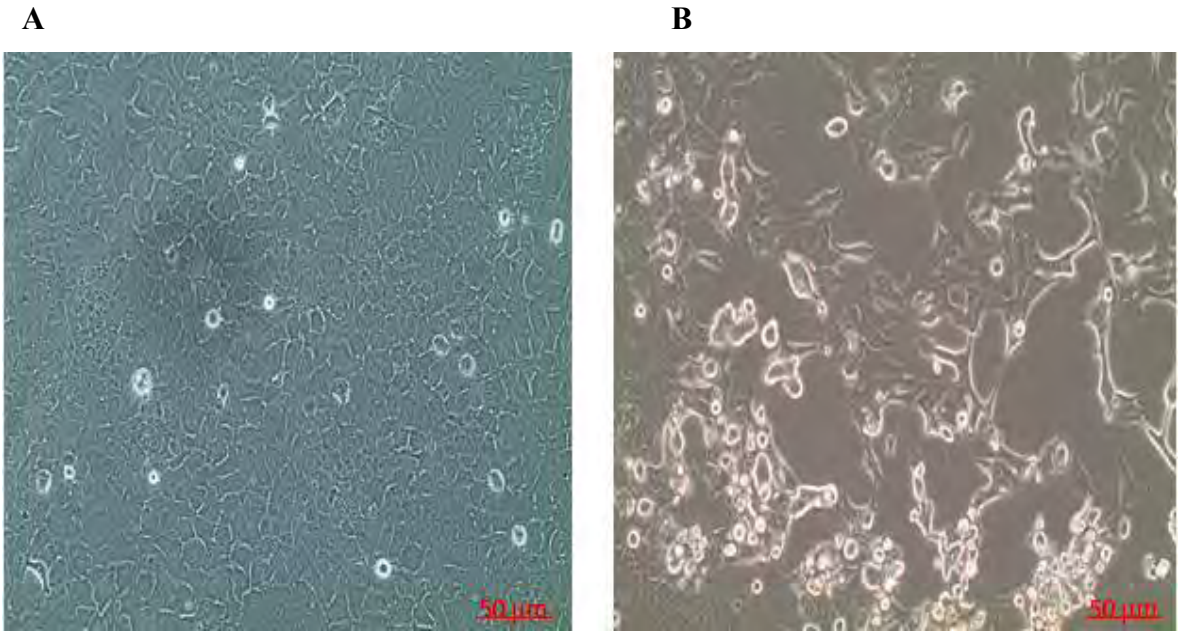


Fig. 2.3 Production of recombinant adenovirus. HEK293 cells at a density of (~10^6 cells) were seeded in a 10cm culture plate and incubated at 37°c in a humidified atmosphere maintained at 5% CO_2 until 70% confluent; **A:** HEK293 cells before packaging of recombinant Adeno-X vector into infectious adenovirus. **B:** HEK293 cells transfected with *Pac*I digested Adeno-HPVE6 showing 50% CPE after overnight incubation. CPE was visualized using the Olympus CKY41 light microscope and the cells harvested by lysing with three consecutive freeze thaw cycles to generate a primary adenoviral stock.

2.2.1.5 Purification of Recombinant Adenovirus

Purification and concentration of the recombinant adenovirus was done using the Adeno-X Maxi Purification Kit (Clontech Purification, 2013), as it is a faster and safer alternative to cesium chloride gradient centrifugation. The kit also allowed for the purification of high yields of recombinant adenovirus directly from the cell pellet.

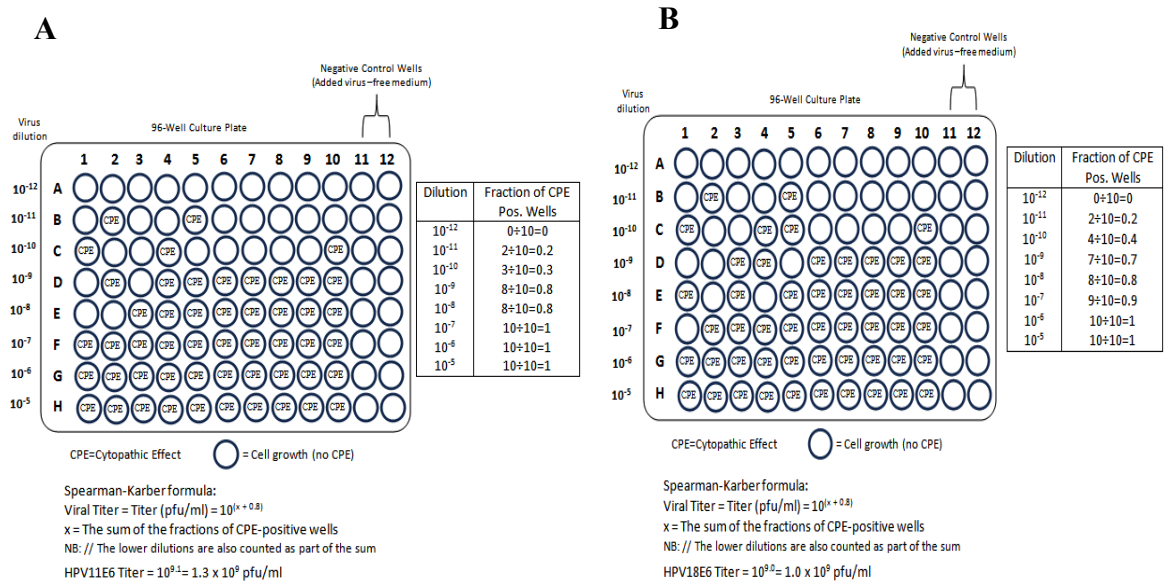


Fig. 2.4 End-point dilution assay. HEK 293 cells were plated in a 96-well plates at a density (~10^4 cells per well) in 100 µl of growth medium and incubated overnight. Serial dilutions of respective HPVE6 virus was then added to each well in columns 1–10 and virus free growth medium was also added to wells in columns 11–12 which served as controls for the viability of non-infected cells. The plates were then covered and incubated in a humidified CO_2 (5%) incubator for 10 days at 37°C. CPE was visualized using the Olympus CKY41 light microscope and the viral titer determined by calculating the fraction of CPE-positive wells in each row as described in section 6.7. **A:** Titer determination of the primary adenoviral amplification stock of HPV11E6. **B:** Titer determination of the primary adenoviral amplification stock of HPV18E6.

A multiplicity of infection (MOI.) of 1 was calculated from the titer of the primary stock (Fig. 2.4) which was used to determine the optimal amount of virus required to seed into cell culture dishes (Table. 2.2).

Table 2.2: Optimal amount of virus required to seed culture dishes for amplification and purification of adenovirus

Adeno-x plasmid	Primary Stock titer	Cells used to determine MOI	MOI-1	Quantity of virus per dish
HPV11E6	$1.3x10^9$	$3.57x10^5$	0.3µl	12µl
HPV 18E6	$1.0x10^9$	$3.57x10^5$	0.4µl	16µl

To produce greater quantities of high titer adenovirus, seven 150 mm HEK 293 cell culture dishes with $1.46x10^7$ cells/dish were each seeded with the calculated quantity of adenoviral E6 constructs (Table. 2.2) until a 50% cytopathic effect (CPE) was observed and harvested by lysing with three consecutive freeze thaw cycles.

2.2.1.6 Determination of Recombinant Adenovirus titer

Once the recombinant adeno-HPVE6 constructs were harvested, purification was done using the Adeno-X Maxi Purification Kit (Clontech Purification, 2013) to remove any traces of growth medium that might interfere with the absorbance at 260 nm. The titer of the harvested recombinant adeno-HPVE6 constructs (Table 2.3) was then determined using the OD_{260} assay (Clontech Expression, 2012).

Table 2.3: Viral titer of recombinant AdenoHPVE6 constructs

Recombinant Adenovirus	Dilution	Virus Stock	0.1% SDS	Conc. ng/µl	A260	A280	260/280	260/230	Viral Titer (opu/ml)
HPV 11E6	1:10	50 µl	450 µl	13.2	0.263	0.139	1.90	0.87	$2.9x10^{12}$
	1:25	20 µl	480 µl	8.3	0.167	0.076	2.20	0.74	$4.6x10^{12}$
	1:50	10 µl	490 µl	6.2	0.124	0.063	1.97	0.61	$6.8x10^{12}$
HPV 18E6	1:10	50 µl	450 µl	13.9	0.279	0.120	2.32	0.90	$3.1x10^{12}$
	1:25	20 µl	480 µl	7.9	0.157	0.067	2.34	0.81	$4.3x10^{12}$
	1:50	10 µl	490 µl	6.5	0.129	0.055	2.35	0.68	$7.1x10^{12}$

Viral titer (opu/ml) = OD_{260} x viral dilution x 1.1 x 10^{12}
Opu = optical particle unit

Performed End-point dilution assay (Fig. 2.4) to determine the working concentration (live virus)

HPV11E6 = $2.9x10^{7.0}$ pfu/ml

HPV18E6 = $3.1x10^{7.0}$ pfu/ml

2.2.1.7 HPV E6 Constructs Confirmation

In order to verify that the encapsidated adenoviral genome contained a full length copy of the HPV E6 construct, PCR analysis was done using the HPV11E6 and HPV18E6 specific primers (Fig. 2.2C) as a quick and efficient way of evaluating the constructs. The insert identities were confirmed by sequencing the E6 constructs using specific promoter CMV primers. A BLASTn search showed 100% nucleotide sequence identity to the HPV11E6 (Accession number: FN870475.1) and HPV18E6 (Accession number: EF422110.1) sequences on the NCBI database. The sequences were interpreted using the Chromas software version 2.01 (McCarthy, 2004) as shown on Fig. 2.5.

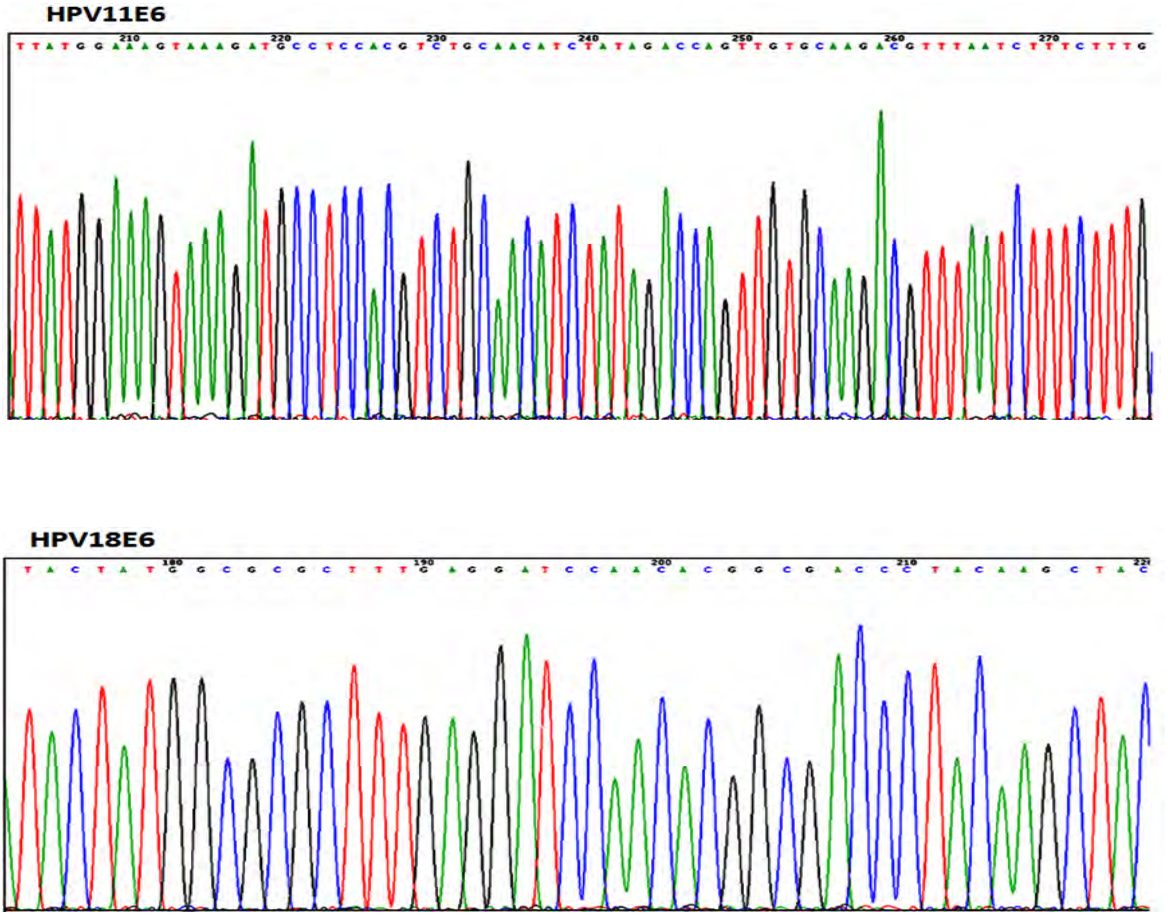


Fig. 2.5: Partial image of NCBI HPVE6 Nucleotide sequence quality confirmation. A: Nucleotide sequence identity of part of the HPV11E6 insert in adenoHPV11E6 using specific CMV promoter primers shows an excellent quality of the sequence using the chromas 2.01 software. **B:** Nucleotide sequence identity of part of HPV18E6 insert in adenoHPV18E6 using specific CMV promoter primers shows an excellent quality of the sequence using the chromas 2.01 software

2.2.1.8 Infection of target cells (HaCaT)

2.2.1.8.1 Determination of multiplicity of infection (MOI)

MOI is typically used to determine the probability and distribution of infectious events in a population of cells. The MOI needed to efficiently transmit a gene of interest into a particular host cell depends on the biological properties of the target cell line and therefore, must be determined empirically. An excessively high MOI can be toxic to cells however; an extremely low MOI may not allow the accurate evaluation of the effect of the transgene in the infected cell line (Buttgereit et al., 2000). In order to achieve a 100% viral infectivity rate in the HaCaT cells, a MOI of 10 and 100 were calculated using the respective viral titers from Table 2.3 and the quantity of virus required to achieve the respective MOI's shown on Table

2.4. A time dependent study (Fig. 2.6) showed that cells infected for 48 hours with MOI 100 gave the highest viral infectivity.

Table 2.4: Quantity of virus per MOI

Cells Per Well	MOI	HPV11E6	HPV18E6
5.0×10^5	10	0.17µl	0.16µl
5.0×10^5	100	1.7µl	1.6µl

2.2.1.8.2 Polycations augment Adenovirus mediated gene transfer

We observed that by infecting HaCaT cells for 48 hours with MOI 100 alone did not achieve a 100% infectivity rate. Thus, we opted to employ the use of polycations specifically DEAE-dextran which has been shown to improve infection efficiency. It was reported that MOI was not an accurate predictor of the efficiency of infection and an increase in MOI did not necessarily improve transduction efficiency (Deyrieux & Wilson, 2007). Thus adenovirus adsorption kinetics was thought to be affected by a number of factors such as inoculum volume, adsorption time, virus concentration and the potential implications of charge interactions for cell binding and entry of adenoviruses (Arcasoy et al., 1997; Nyberg-Hoffman et al., 1997). Gene transfer was shown to be augmented by the addition of polycations such as polybrene, poly-l-lysine, DEAE-dextran and protamine by enabling the association between the negatively charged eukaryotic membrane and the adenovirus capsid also carrying a net negative surface charge (Fasbender et al., 1997; Lanuti et al., 1999). The addition of polybrene (6–8 µg/mL) was reported to dramatically improve the transduction efficiency of adenoviral infection on HaCaT cells to between 80–90% (Jacobsen et al., 2006; Deyrieux & Wilson, 2007).

This study employed four techniques for infecting HaCaT cells; (i) cells were plated then infected through spin inoculation only, (ii) the virus and cells were mixed together before plating, (iii) cells were plated and infected with the virus with the aid of DEAE dextran and spin inoculation and (iv) the virus and cells were mixed together with the DEAE dextran before plating. The third technique using spin inoculation with 7.5µg/ml DEAE-dextran increased the transduction efficiency compared to the other techniques as shown by the increased levels of HPV18E6 protein using western blot analysis (Fig. 2.7). HeLa cells and uninfected HaCaT cells acted as positive and negative controls respectively. The same results were also observed for HPV11E6, although not with western blot analysis due to lack of HPV 11E6 antibodies but with all the other experiments where HPV11E6 was used for infecting cells.

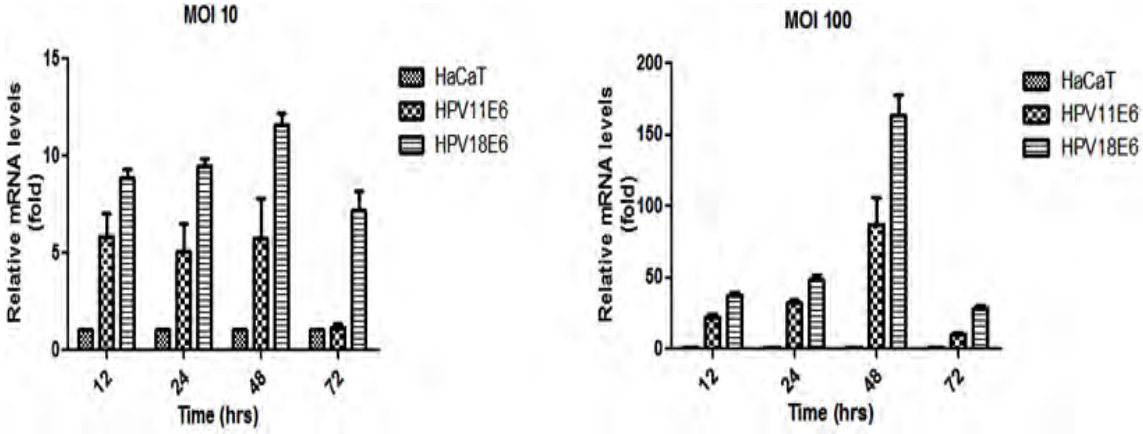


Fig. 2.6: qRT-PCR determination of MOI. HaCaT cells were infected with the indicated HPVE6 constructs with the aid of DEAE dextran and spin infection at MOI 10 and MOI 100 and the levels of mRNA expression determined at various time points as described in section 6.13. Infection of cells with MOI 100 for 48hours showed the highest viral infectivity

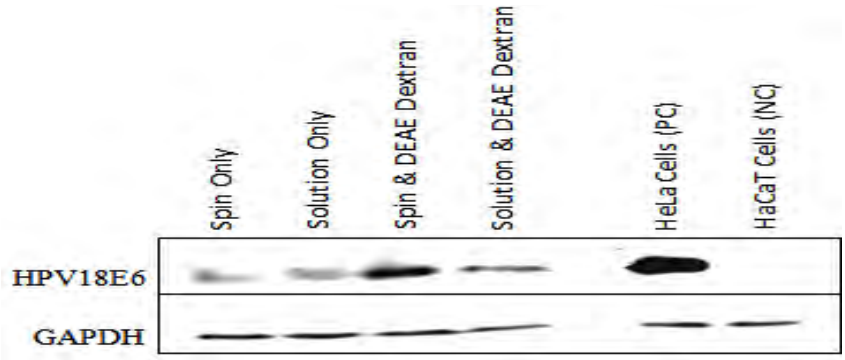


Fig. 2.7: Transduction Efficiency. HaCaT cells were infected with HPV18E6 using four different methods. Firstly cells were plated after reaching 70% confluence and infected through spin inoculation only, in the second instance the virus and cells were mixed together before plating, thirdly cells were plated after reaching 70% confluence and infected with the virus with the aid of DEAE dextran and spin inoculation and finally the virus and cells were mixed together with the help of DEAE dextran before plating. HeLa cells and HaCaT uninfected cells acted as positive and negative controls respectively. Incubation was done for 48hrs at 37°c after which 100µg of Protein was loaded on a 12% SDS-Page Gel. The results showed the addition of DEAE dextran with spin inoculation gave a higher expression of the protein compared to the other techniques as described in section 6.12.

2.2.2 Confirmation of HPV E6 Expression

Confirmation of the levels of HPV11E6 and HPV18E6 was achieved by measuring both mRNA by qRT-PCR (Fig. 2.8) and protein by western blotting (Fig. 2.9) while confocal microscopy (Fig. 2.10) showed the protein localization as described below.

2.2.2.1 HPV E6 mRNA expression

Expression of HPV E6 mRNA in HaCaT cells was confirmed by qRT-PCR, using either HPV11E6 or HPV18E6 specific primers, which firstly, demonstrated the expression of both HPV11E6 and HPV18E6 in HaCaT cells and also confirmed the specificity of the products in both cell lines (Fig. 2.8A). Secondly, the qRT-PCR showed that both p21 and p53 transcripts

were not affected in cells infected with HPV11E6 but were downregulated in cells infected with HPV18E6 (Fig. 2.8B). HeLa cells were used in both cases as positive control for HPV18E6 mRNA.

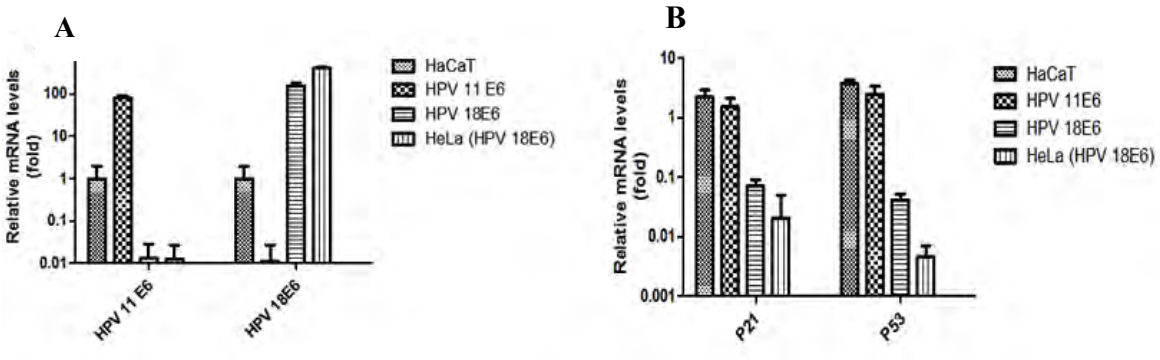


Fig. 2.8: qRT-PCR confirmation of HPV11E6 and HPV18E6 mRNA levels. HaCaT cells were infected with either HPV11E6 or HPV18E6 and the levels of the E6, P21 and P53 mRNA measured by qRT-PCR. **A:** HPV11E6 and HPV18E6 specific primers were used to detect the respective HPVE6. HeLa cells were used as positive control for HPV18E6 mRNA as described in section 6.17. **B:** Transcription of p21 and p53 genes following infection of HaCaT cells with the respective HPVE6 constructs. Both p21 and p53 were not affected in cells infected with HPV11E6 but were downregulated in cells infected with HPV18E6 and in HeLa cells. The figure represents the average of three replicates for each infection.

2.2.2.2 Effect of HPV11E6 and HPV18 E6 expression on p53 and p21

Western blotting analysis confirmed the presence of HPV18E6, p21 and p53 proteins following the infection of HaCaT cells with either HPV11E6 or HPV18E6 (Fig. 2.9). The p21 and p53 protein levels were downregulated in cells expressing HPV18E6 compared to those expressing HPV11E6 and in the HaCaT control cells.

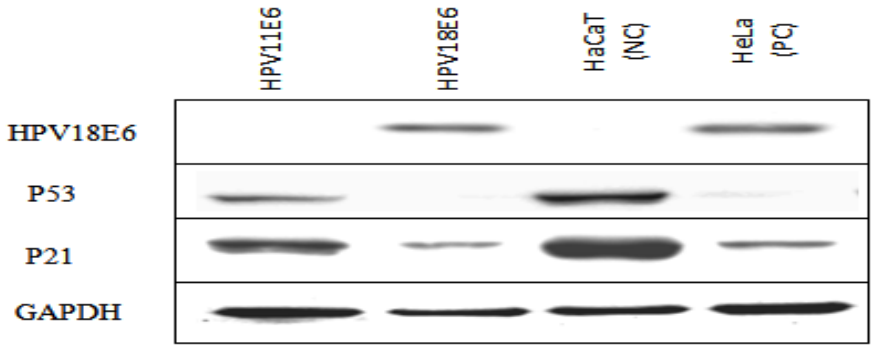


Fig. 2.9: HPV11E6 and HPV18E6 Protein detection. HaCaT cells were infected with either HPV11E6 or HPV18E6 and the protein levels of the HPV18E6, P21 and P53 were measured. Western blotting using HPV18E6 antibodies showed the presence of HPV18E6 in HaCaT cells infected with the HPV18E6 constructs and in HeLa cells which are known to constitutively express the HPV18E6 protein. p21 and p53 protein levels were downregulated in cells expressing HPV18E6 compared to those expressing HPV11E6 and the HaCaT control cells.

2.2.2.3 Confocal Microscopy

The localization of HPV18E6 and p53 in HaCaT cells, expressing either HPV11E6 or HPV18E6, was investigated through confocal microscopy, using the respective specific primary antibodies as described in section 6.19. HPV18E6 is primarily localized in the nucleus in HaCaT cells expressing HPV18E6 and in HeLa cells. Whereas, HPV18E6 is not detected in HaCaT control cells and in HaCaT cells infected with HPV11E6. In HPV11E6 expressing cells p53 was observed to be primarily localised to the nucleus with some diffused staining in the cytoplasm, whereas in HPV18E6 expressing cells, p53 was observed at lower levels than either in the control or HPV11E6 expressing cells. There also appears to be some increased cytoplasmic staining (degradation products) in the latter.

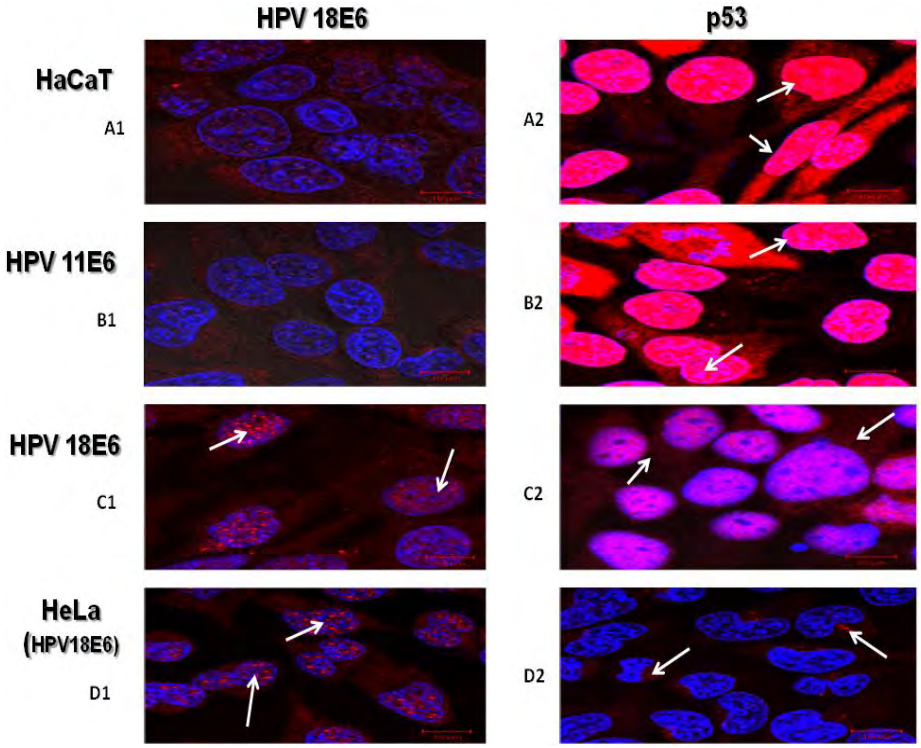


Fig. 2.10: HPV E6 Protein Localization. HaCaT cells were plated on a microscope cover slip placed in a 6-well cell culture dish at a density (~10^5 cells per well) in triplicate. After overnight incubation cells were infected with the respective HPVE6 virus at MOI 100 and incubated in a humidified CO_2 (5%) incubator for 48 hours at 37°C. The cells were then incubated in appropriate antibodies fixed and stained before being analysed using a confocal microscope (Zeiss LSM 510 Meta with NLO) as described in section 6.19. **A1:** HaCaT control cells incubated with HPV18E6 specific primary antibody showing absence of HPV18E6. **A2:** HaCaT control cells incubated with p53 primary antibody showing nuclear and cytoplasmic localization of p53. **B1:** HaCaT cells expressing HPV11E6 incubated with HPV18E6 specific primary antibody showing lack of HPV18E6. **B2:** HaCaT cells expressing HPV11E6 incubated with p53 primary antibody showing nuclear and cytoplasmic localization of p53. **C1:** HaCaT cells expressing HPV18E6 incubated with HPV18E6 specific primary antibody showing nuclear localization of HPV18E6. **C2:** HaCaT cells expressing HPV18E6 incubated with p53 primary antibody showing degradation of p53 in the nucleus and cytoplasmic p53 degradation products **D1:** HeLa cells expressing HPV18E6 incubated with HPV18E6 specific primary antibody showing nuclear localization of HPV18E6. **D2:** HeLa cells expressing HPV18E6 incubated with p53 primary antibody showing absence of p53in the nucleus and cytoplasmic p53 degradation products.

2.2.3 Characterisation of the effects of HPV E6 on HaCaT cells

2.2.3.1 Morphology

The HPV E6 gene from both the constructs did not visibly modify the morphology of HaCaT cells (Fig. 2.11). HaCaT cells expressing both HPV11E6 and HPV18E6 showed typical epithelial cell morphology similar to the control cells.

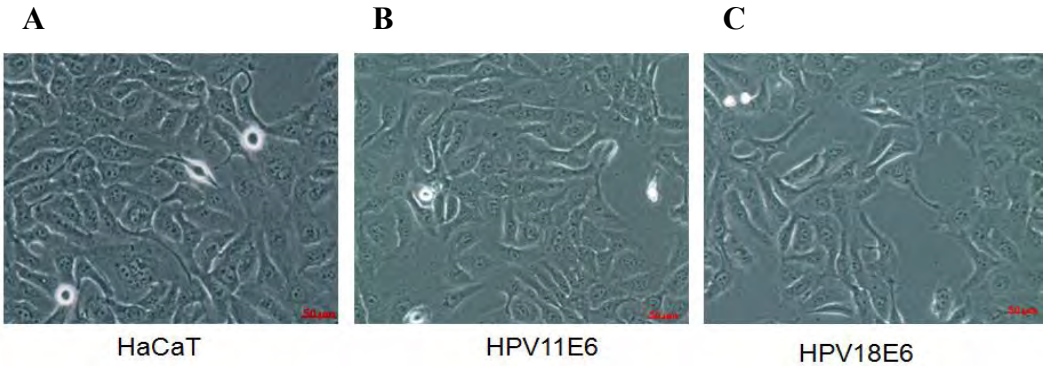


Fig. 2.11: Effects of HPVE6 on Cellular Morphology. HaCaT cells were plated in a 10cm culture dish at a density (~10^6 cells) until 70% confluent. The cells were then infected with the respective HPVE6 at MOI 100 and incubated in a humidified CO_2 (5%) incubator for 48 hours at 37°C. The cells were visualized using an Olympus CKY41 light microscope. **A:** HaCaT control cells. **B:** HaCaT cells expressing HPV11E6. **C:** HaCaT cells expressing HPV18E6

2.2.3.2 Effects of HPVE6 on the Cell Cycle

The effect of HPV E6 expression on the cell cycle (Fig. 2.12) showed that the cell cycle distribution in HaCaT cells expressing HPV11E6 was comparable to the control HaCaT cells, i.e. no difference in the cell cycle phases. However HaCaT cells expressing HPV18E6 showed a significant increase ($p<0.05$) in the percentage of cells in the G2-M phase compared to the control HaCaT cells, leading to the deregulation of the cell cycle.

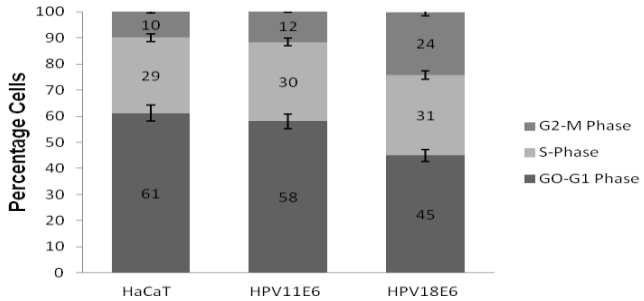


Fig. 2.12: Effects of HPV11E6 and HPV18E6 on the Cell Cycle. HaCaT cells were plated in triplicate on three clones from each HPVE6 virus in a 10cm culture dish at a density (~10^6 cells) until 70% confluent. The cells were then infected with the respective HPVE6 at MOI 100 and incubated in a humidified CO_2 (5%) incubator for 48 hours at 37°C. Cells were then harvested and flow cytometric cell cycle analysis was carried out using a BD FACS ™ Calibur flow cytometer (BD Biosciences, CA, USA). For each sample, at least 20,000 cells were measured for DNA content. Analysis of the results was performed using ModFit LT software (Verity Software House) to calculate the percentage of cells in each stage of the cell cycle as described in section 6.20. HaCaT cells expressing HPV18E6 showed a significantly higher number of cells in the G2-M phase of the cell cycle compared to HaCaT cells expressing HPV11E6 and HaCaT control cells.

2.2.3.3 Cellular Transformation

Cellular transformation induced by HPV11E6 and HPV18E6 was assessed through the use of the soft agar colony formation assay. This is a common method for monitoring anchorage independent growth (one of the hallmarks of cell transformation) and is considered a reliable and commonly used *in vitro* assay for detecting cellular transformation of cells. For this assay, HaCaT cells expressing the respective HPVE6 proteins were cultured in soft agar medium for 21 days and the colonies were analyzed morphologically using crystal violet staining for the presence of pink colonies indicating cellular transformation (Fig. 2.13). While no colony formation was visible in the control HaCaT cell, few colonies were visible in cells expressing HPV11E6 and many more were visible in cells expressing HPV18E6. This showed the ability of both HPV11E6 and HPV18E6 to induce cellular transformation with HPV18E6 having a much stronger effect.

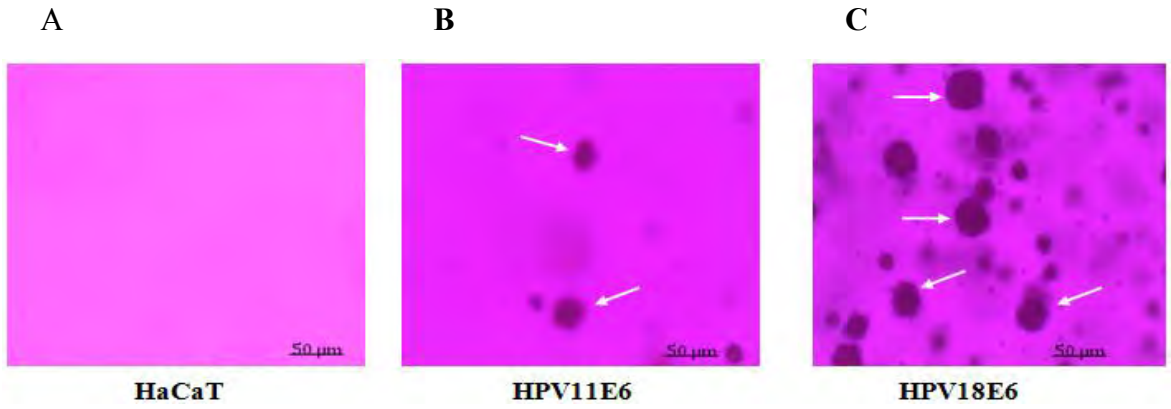


Fig. 2.13: Colony formation by HPV E6. HaCaT cells were plated on a microscope cover slip placed in a 6-well cell culture dish at a density (~10^5 cells per well) in triplicate. After overnight incubation the cells were infected with the respective HPVE6 at MOI 100 and incubated in a humidified CO_2 (5%) incubator for 48 hours at 37°C as described in section 6.21. Cells were then stained with 0.5 ml of 0.005% Crystal Violet for more than 1 hour and pink colonies were counted using an Olympus CKY41 light microscope **A:** No colony formation was visible for the HaCaT control. **B:** Few and tiny colony formation visible from the HaCaT cells expressing HPV11E6**. C:** Many and large colony formation visible from the HaCaT cells expressing HPV18E6.

2.3 Discussion

While considerable information is available on the diverse set of functions of high risk HPVE6, little is known about the function(s) of low risk HPVE6 proteins and the possible role that they may play in the process of cellular transformation. The aim of this chapter was to compare the effects of the E6 proteins from HPV11 (low risk) and HPV18 (high risk) and to evaluate the effect of these two types of E6 proteins on cellular gene expression and transformation. The HPV11E6 and HPV18 E6 genes were cloned into an adenoviral vector to induce, high level transient expression of these genes in the HaCaT cells. RNA isolated from

these cells was later used in DNA microarray analysis to study and compare the effects of the E6 gene on cellular gene expression patterns during the early stages of cellular transformation.

Historically the detection of HPV E6 proteins in cells has proved problematic, largely because of a lack of high avidity E6 antisera and the low levels of E6 expression in cells (Slebos et al., 1995). In order to overcome this problem studies have used adenoviral vectors to infect a broad range of human cells to yield high levels of gene transfer (Vorburger & Hunt, 2002). Although efficient adenovirus mediated gene transfer into cultured epithelial and endothelial cells has been reported, these cells are resistant to adenovirus infection and require high MOI's and a relatively long physical contact time between the adenovirus and target cell for significant gene transfer (Arcasoy et al., 1997). One way to compensate for the low efficiency of viral transduction is the use of high MOI's, but this approach has the drawback of inducing strong inflammatory responses. Jacobsen et al., (2006) showed that a more suitable way to significantly increase the adenovirus transduction efficiency was by the improvement of the viral adsorption by polycations such as DEAE dextran.

The lack of HPV11E6 specific antibodies has seen most studies employing indirect identification methods such as a green fluorescent protein (GFP) tag on HPV11E6 fusion proteins (Sun et al., 2008) or haemagglutinin (HA) tag (Guccione et al., 2002) in an effort to show the expression of HPV11E6. Although Watson et al., (2003) showed that the use of N- or C-terminal epitope tagged forms of E6 may not give a true reflection of E6's biology as the presence of tags could affect the function of the oncoprotein. In this study some of the downstream effects of HPV11E6 were measured indirectly by determining the levels of p21 and p53 proteins. Expression of the p21 and p53 genes in cells expressing HPV11E6 was comparable to the HaCaT control cells as opposed to cells expressing HPV18E6 which showed a marked reduction of these proteins. These findings demonstrated that cells expressing HPV18E6 were capable of binding and degrading p53, leading to the inactivation of the p21 gene which is thought to synergize with tumour suppressors and antagonize oncogenes to protect against cancer (Abbas & Dutta, 2009).This was consistent with other studies that have shown p21 gene transcription to be inhibited as a consequence of the HPVE6 mediated p53 degradation, and its reduced expression has also been reported in invasive squamous cell carcinoma (Cho et al., 2002; Cheng et al., 2007; Huang et al., 2010; Tagle et al., 2014).

High risk E6 proteins have been shown to work via several mechanisms to interfere with p53 function. They recruit the cellular E3 ubiquitin ligase E6-associated protein (E6AP) to a trimeric complex with p53, leading to the subsequent ubiquitination and degradation of p53 (Scheffner et al., 1990; Crook et al., 1991; Huibregtse et al., 1991). Interestingly, low risk E6 proteins are also able to form complexes with E6AP (Oh et al., 2004; Brimer et al., 2007); however this union does not result in the degradation of p53 (Li & Coffino, 1996) suggesting that other cellular factors may be targets for the low risk E6-E6AP complex. In this study the increased levels of p53 and p21 proteins in cells expressing low risk HPVE6 type, compared with those expressing high risk HPVE6 type could be attributed to a number of factors. It could represent a stress response to viral infection (Prives & Hall, 1999), and the p53 expression could also play a role in controlling the initial burst of amplificational replication that occurs after viral entry (Cox, 1997). Alternatively the elevation of p53 could also occur as a result of p53 stabilization by binding with E6 protein, particularly with low risk types in which degradation does not occur (Giannoudis & Herrington, 2000).

To confirm that the HPV18 E6 induced degradation of p53 levels, confocal microscopy was also used to determine whether there was a shift in the localisation of p53 in cells expressing E6 from a low-risk compared to a high-risk HPV type. In HPV11E6 expressing cells p53 was observed to be primarily localised to the nucleus with some diffused staining in the cytoplasm, whereas in HPV18E6 expressing cells, p53 was observed at lower levels than either in the control or HPV11E6 expressing cells. These results may explain the reduced effect of HPV11E6 compared to HPV18E6 in inducing cellular transformation even though it has been shown to able to bind and inactivate but not degrade p53. These results are consistent with studies by Stewart et al., (2005) who showed that high risk HPV18 E6 but not low risk HPV11E6 was able to accumulate predominantly in the nucleus. They also showed that both HPV18E6 and p53 when co-expressed shuttled from the nucleus to the cytoplasm where the p53 is degraded.

Morphological changes in epithelial cells as a result of HPV E6 oncoproteins have been reported after many passages (Shen et al., 2002). A previous study by Kaczkowski et al., (2012) demonstrated that cultures transfected with HPV types; HPV11, HPV16 or HPV45 grew much slower than control transfected cells. This slow growth rate may be attributed to death of part of the cell population as a result of HPV killing the cells or cells switching on the apoptosis programme in response to HPV. This effect was however not observed in the

current study as the morphology of cells expressing HPVE6 was unchanged compared to the HaCaT control cells.

Normal mammalian cells exhibit G1 and G2 cell cycle arrest in response to DNA damage, in order to allow cells time to repair the damage before initiating replicative DNA synthesis or mitosis (Painter & Young, 1980; Denekamp, 1986). In this study, cells expressing HPV11E6 did not disrupt the cell cycle, as opposed to cells expressing HPV18E6 which showed an increase in percentage of cells in the G2/M phase. This was hypothesised to be due to inactivation/degradation of p53 by HPV18E6 that inturn resulted in G2/M delay. This interaction between oncogenic HPVE6 and p53 suggested a plausible mechanism for the enhanced accumulation of genetic alterations. This was necessary for cellular transformation that occurred with a much stronger effect in high risk HPV as opposed to low risk HPV infection. The HPV infected cells expressing E6 were predisposed to acquisition of genetic alterations in a manner similar to that observed in cells with mutant p53 (Kessis et al., 1993).

The interrelation between the increased G2/M population of cells and HPV18E6 expression in the present study was not readily apparent as the ability of HPVE6 to bind to other cellular proteins apart from p53 limits the interpretation of these findings (Chen et al., 1995; Imai et al., 1997; Kiyono et al., 1997; Gao et al., 1999; Taylor et al., 1999). This was also supported by Hamada et al., (2009) who applied kinetic mathematical models to show that, regardless of the DNA damage level, there were delays in the G2/M progression and that apoptosis was only induced in situations where extensive DNA damage occurred.

Cell transformation studies provide information about the involvement of oncogenes and other genetic changes in the initiation and progression of cancer. Transformation of normal cells by viral genes is a frequent approach to inducing early events in the development of a malignant phenotype. In the current study, cellular transformation was shown through monitoring of anchorage independent growth on HaCaT cells. Growth of cells in soft agar showed that the HaCaT control cell did not induce colony formation while cells expressing HPV11E6 were able to form colonies, albeit very few and small compared to cells expressing the high risk HPV18E6. Consistent with these findings (Boukamp et al., 1988; Boukamp et al., 1990; Ren et al., 2006) have also shown that mock transfected HaCaT cells did not form soft agar colonies whereas, tissue culture studies have demonstrated that HPVE6 and E7 genes exhibit immortalizing properties in human foreskin keratinocytes (Hawley-Nelson et al., 1989; Hudson et al., 1990; Woodworth et al., 1990; Griep et al., 1993) and alter

keratinocyte differentiation (Munger et al., 1989; Hudson et al., 1990; Woodworth et al., 1990). The ability of cells expressing the low risk HPVE6 to form colonies may be attributed to the overall similarity of the high risk HPV E6 proteins to the low risk HPV E6 proteins as it seemed likely that they possessed common signalling pathways in inducing cellular transformation (Huibregtse & Beaudenon, 1996; Oh et al., 2004).

Taken together, these findings significantly extend the understanding of HPVE6 mediated p53 degradation and show that both the low risk HPV11E6 and high risk HPV18E6 proteins were implicated in cellular transformation. Although HPV *per se* cannot account for cellular transformation (Schlegel et al., 1988), it was hypothesized that both the HPVE6 proteins might play a role in the transformation of HaCaT cells through the activation or inactivation of other cellular processes apart from the binding and inhibition of p53. Similar results were also reported by Chen et al., (1993) who postulated the functional inactivation of p53 as an effector mechanism for HPV mediated cellular transformation, but that it is not sufficient to confer cellular transformation on its own.

To the best of our knowledge this is the first study to have shown early stages of HPVE6 infection especially the ability of the low risk HPV11E6 in enabling cellular transformation. Although other studies have also speculated on the ability of HPV11E6 in achieving cellular transformation this study entailed the use of an adenoviral vector with the aid of polycations to highly express HPVE6 in an effort to identify genes associated with the initiation of cellular transformation which has not been shown before.

CHAPTER THREE
Effects of HPVE6 on Cellular Gene Expression Patterns

3.1 Introduction

Human papillomaviruses (HPV) mediated carcinogenesis is mainly due to the oncogenic activities of the viral proteins E6 and E7. The classical function of E6 is to induce the degradation of p53 via direct binding to the ubiquitin ligase E6AP, inhibiting p53-dependent signaling upon stress stimuli and contributing to tumorigenesis (Scheffner et al., 1990; Werness et al., 1990; Scheffner et al., 1993). On the other hand the E7 oncoprotein acts by binding to members of the Rb tumour suppressor protein family and inhibits their ability to modulate the function of E2F transcription factors (Dyson et al., 1989; Zur Hausen, 2002).

The E6 proteins from both low risk and high risk HPV also regulate the immune response by causing the down regulation of multiple interferon responsive genes (Nees et al., 2001). They both can bind to Tyk2 of the Jak-Stat pathway (Li et al., 1999). The high risk E6 also binds to IRF3 and inhibits its ability to activate interferon-responsive genes (Ronco et al., 1998). The ability of the HPVE6 to interact with both p53 and p300/CBP likely also plays a role in interferon response regulation (Hebner et al., 2006).

The inhibition of the apoptotic signaling has also been reported. Both low and high risk HPVE6s bind and degrade the pro-apoptotic protein Bak (Thomas & Banks, 1999) independent of its p53 degradation function, suggesting that E6's interaction with Bak appears to be conserved between both the low and high risk HPVE6s types (Klingelhutz & Roman, 2012). The high risk HPV E6s can also bind to procaspase 8 which can prevent E6-expressing cells from responding to apoptotic stimuli (Tungteakkhun & Duerksen-Hughes, 2008). E6 from both low risk and high risk HPVs also bind to and degrade the human mini-chromosome maintenance 7 (hMCM7) protein leading to chromosomal abnormalities in HPV infected cells, although binding of high risk E6 proteins appears to be stronger than that of low risk E6 proteins (Kukimoto et al., 1998; Howie et al., 2009).

The introduction of DNA microarray technology has had a dramatic impact on cancer research, allowing researchers to analyze the expression of thousands of genes and relate gene expression patterns to clinical phenotypes (Alizadeh et al., 2000; García-Escudero & Paramio, 2008). The power of DNA microarray has been demonstrated in the studies of a wide variety of malignancies including cancers of prostate, breast, liver, oesophagus, pancreas, ovary, stomach lung, and head and neck (Tonin et al., 2001; Al Moustafa et al., 2002; Han et al., 2002; Luo et al., 2002). These microarray studies have revealed large sets of genes that are differentially expressed between tumour and normal cells, including those genes known to be important for neoplastic transformation (Bhattacharjee et al., 2001; Garber et al., 2001; Belbin et al., 2002; Guo, 2003).

This chapter assesses the biological changes in the transformed HaCaT cells described in chapter two. The analysis compares the regulation of molecular pathways by genes being uniquely expressed or shared by both the low risk and high risk HPVE6 types and their implication in oesophageal cancer progression.

3.2 Results

3.2.1 Assessment of RNA integrity

The assessment of RNA integrity is a critical first step in obtaining meaningful gene expression data as intact RNA is a key element for successful microarray or qRT-PCR analyses (Lightfoot, 2002; Mueller et al., 2004). RNA was extracted as three biological replicates from HaCaT control cells and HaCaT cells expressing either HPV11E6 or HPV18E6. The RNA concentration and the optical density 260/280 of the samples were quantified using the NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA), values of 1.8–2.0 were considered indicative of relatively good quality RNA (Bojarczuk et al., 2014). RNA integrity was then determined on an Agilent 2100 Bioanalyzer, to assess the degradation of the samples prior to its use in the downstream applications. This provided an electrophoretic tracing and a RNA integrity number (RIN) for judging RNA quality (Schroeder et al., 2006). RNA integrity was determined by samples having a RIN of between 7.0 and 10.0 (Kawasaki, 2006), (Table 3.1).

Table 3.1: RNA integrity assessment

Sample	RIN[a]	260/280[b]	260/230[c]	Comment
HaCaT-A	9.2	1.99	2.11	Pass
HaCaT-B	9.6	1.99	2.0	Pass
HaCaT-C	9.3	1.99	2.1	Pass
HPV11E6-A	9.8	1.97	2.23	Pass
HPV11E6-B	9.9	1.97	2.19	Pass
HPV11E6-C	10.0	1.87	2.23	Pass
HPV18E6-A	9.8	1.95	2.21	Pass
HPV18E6-B	9.9	1.96	2.16	Pass
HPV18E6-C	9.7	1.94	2.24	Pass

a = Measure of the degradation of RNA samples based on a numbering system from 1 to 10, with 1 being the most degraded and 10 being the most intact
b = 1.8-2.0 (indicative of relatively good quality RNA)
c = 1.8-2.2 (secondary measure indicative of relatively good quality RNA)

3.2.2 DNA Microarray analysis

Once the samples had passed the quality control they were used for DNA microarray analysis at the Centre for Proteomic and Genomic Research (CPGR). Human gene ST 2.0 arrays were used for gene expression on an affymetrix platform (Affymetrix, 1999) as depicted in Fig.3.1. Data analysis was done using the Partek® Genomics Suite™ 6.6 software (Partek Inc., 2014).

The raw data was first normalized using Robust Multi-array Average (RMA); a normalisation method used to account for technical variation between the arrays and consisted of three steps: a background adjustment, quantile normalization and summarisation (Bolstad et al., 2003; Irizarry et al., 2003).

After normalization a preliminary data analysis on the samples was done using Principal Components Analysis (PCA) to explore whether differences or similarities existed between the samples. There was a clear difference between HPVE6 expressing cells and the uninfected HaCaT control cells, while there was some similarity between the HPV11E6 and HPV18E6 expressing cells as shown by the overlap of genes (Fig. 3.2A).

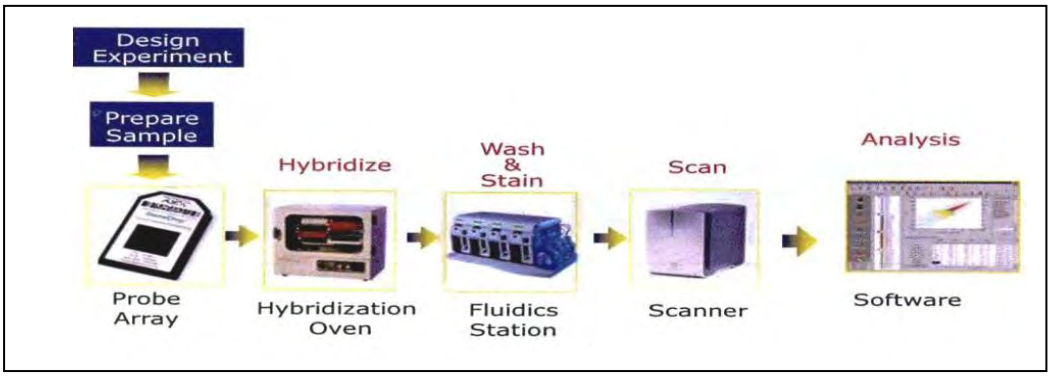


Fig. 3.1 DNA microarray work flow on an affymetrix platform. The DNA samples were first fragmented into single stranded DNA and labelled after which a hybridization mix cocktail was added before being loaded into the human gene ST 2.0 array. The chip Array was then incubated in a hybridization oven for 17 hours followed by washing and staining in the fluidics station 450/250. The array was then scanned using the GeneChip® Scanner 3000 which computed the probe intensity data as .CEL files that were then analysed using the Partek® Genomics Suite™ 6.6 software (Affymetrix, 1999).

Histogram plots were also used to determine whether genes that are expected to be differentially regulated have larger or smaller intensity values so as to view the distribution of the intensities and identify any outliers. All the samples were shown to follow the same distribution pattern indicating that there are no obvious outliers in the data (Fig. 3.2B).

A **B**

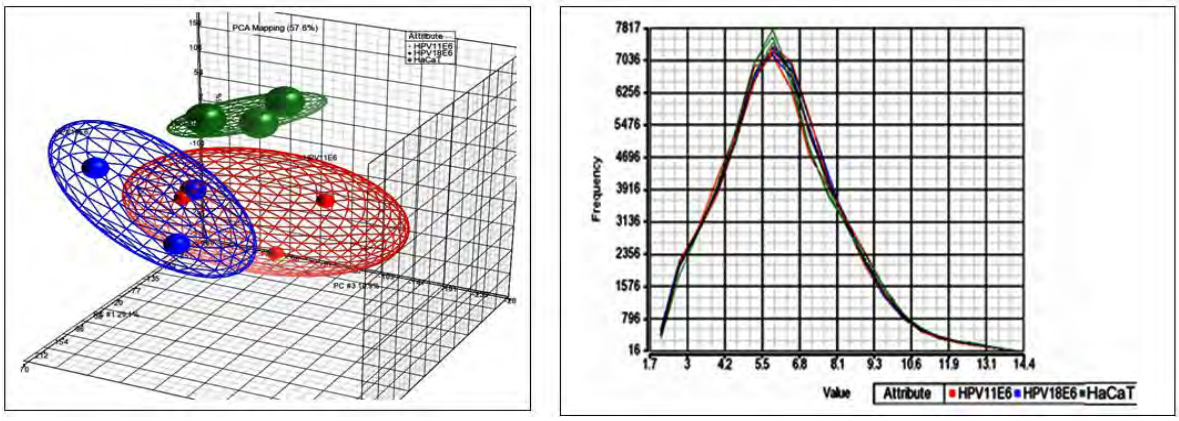


Fig. 3.2: Exploratory data analysis of HaCaT cells expressing HPVE6: A. PCA Analysis: This plots one dot for each of the samples, HaCaT cells expressing HPVE6 show some resemblance to each other and not with the control cell. **B. Histogram Plots:** This plots one line for each of the samples with the intensity of the probes graphed on the X-axis and the frequency of the probe intensity on the Y-axis. In this dataset, all the samples follow the same distribution pattern indicating that there are no obvious outliers in the data. *Green:* HaCaT control cells; *Red:* HPV11E6; *Blue:* HPV18E6

Analysis of variance (ANOVA) was used to identify the differentially expressed genes and a list of genes was generated that was significantly different between the samples. Global differences in gene expression patterns induced by HPV11E6 and HPV18E6 are demonstrated in Fig.3.3 as a heat map of fold changes compared to the HaCaT control. A clear difference is seen with HPV18E6 having induced a higher number of up-regulated

genes compared to HPV11E6. The heat map represents a combination of the standardized gene expression level of each gene in each sample. A list of the most significantly differentially expressed genes was then generated using volcano plots set at a fixed fold-change cut-off of between –2.0 and 2.0 with a significance FDR of ≤ 0.05 (Wen et al., 1998; Tamayo et al., 1999). From the volcano plots HaCaT cells expressing HPV18E6 had a higher number of expressed genes compared to HaCaT cells expressing HPV11E6 Fig. 3.4.

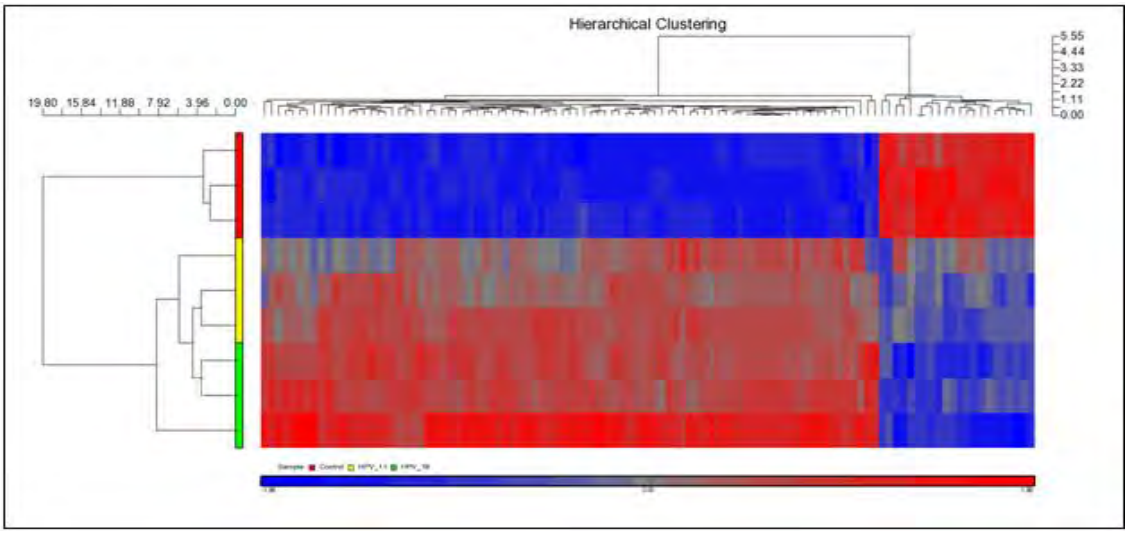


Fig. 3.3: Hierarchical Clustering of HaCaT cells expressing HPVE6. The resulting graph illustrates the standardized gene expression level of each gene in each sample with all the gene intensities being adjusted to a mean of zero and the standard deviation to one. Up-regulated genes have positive values and displayed in red. Down-regulated genes have negative values and displayed as blue. Each gene is represented in one column, and each sample is represented in one row. HPV18E6 dataset showed a greater number of upregulated genes compared HPV11E6.The left dendogram side bars represent; *Red:* HaCaT control cells; *Yellow:* HPV11E6; *Green:* HPV18E6

A **B**

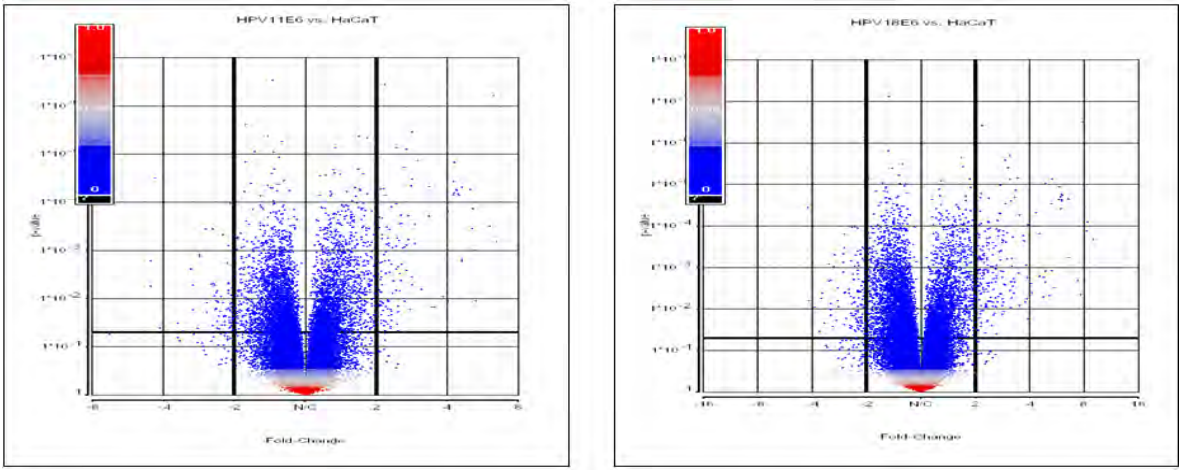


Fig. 3.4: Volcano plots of HaCaT cells expressing HPVE6 vs HaCaT control cells. In the plot, each dot represents a gene. The X-axis represents the fold change and the Y-axis represents the p-values. The genes up-regulated are on the right side and the down-regulated genes are on the left side. The genes become more statistically significant with increasing Y-axis. **A:** HPV11E6 dataset showed a lower number of expressed genes compared **B:** HPV18E6.

Once the list of the most significantly differentially expressed genes was generated, a Venn diagram (Pirooznia et al., 2007) was used to identify genes that were uniquely induced by either HPV11E6 or HPV18E6 and genes that were shared between the two datasets. A total of 3210 genes were significantly differentially expressed, out of these genes 1416 were uniquely induced by HPV18E6 while HPV11E6 uniquely induced 923 genes (Fig. 3.5). An overlap of 871 genes was induced by both HPVs. This result was consistent with the PCA result (Fig. 3.2A) which showed both HPV11E6 and HPV18E6 shared some differentially regulated genes when compared to control HaCaT cells. The results generated from the Venn diagram (Fig. 3.5) were used for the subsequent gene ontologies, pathway and network analysis.

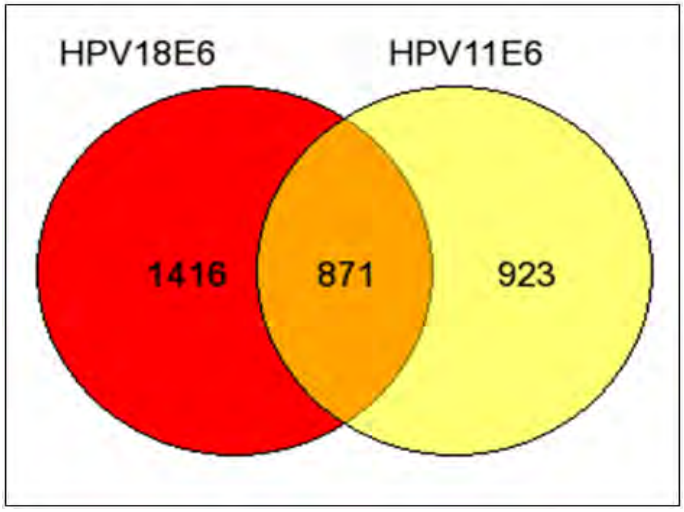


Fig. 3.5: Venn diagram of significantly differentially expressed genes (both up- and down-regulated). There was a distinct difference in the gene expression patterns of HPV11E6 and HPV18E6. HPV18E6 was shown to uniquely express 1416 genes (*Red*) while HPV11E6 uniquely expressed 923 genes (*Yellow*). There was an overlap of 871 genes expressed by both HPVs (*Orange*)

3.2.3 Validation of Microarray data

The large numbers of gene expression measurements obtained in a typical microarray experiment; do yield large numbers of false positive or negative results (Tan, 2004). Thus microarray results require verification and validation by a complementary gene expression profiling method (Quellhorst et al., 2005). The data generated in the microarray analysis was validated by qRT-PCR on selected up- and down-regulated genes from both the HPV11E6 and HPV18E6 datasets (Table 3.2 & 3.3).

Table 3.2: HPV11E6 Top 10 up- and down-regulated genes based on either the extent of up or down-regulation or on their known involvement in carcinogenesis

Up-regulated			Down-regulated		
Gene Symbol	**Gene Name**	**Fold Change**	**Gene Symbol**	**Gene Name**	**Fold Change**
PLAT	Plasminogen Activator, Tissue	4.3	STARD7	StAR-Related Lipid Transfer Domain Containing 7	-4.1
TNFSF10	Tumour Necrosis Factor (Ligand) Superfamily, Member 10	4.2	BRCA1	Breast Cancer 1, Early Onset	-3.1
TP53INP1	Tumour Protein P53 Inducible Nuclear Protein 1	4.1	OR2M7	Olfactory Receptor, Family 2, Subfamily M, Member 7	-3.0
GADD45A	Growth Arrest And DNA-Damage-Inducible, Alpha	3.8	IFI6	Interferon, Alpha-Inducible Protein 6	-2.4
BUB1	Budding Uninhibited By Benzimidazoles 1	3.5	MCM2	Minichromosome Maintenance Complex Component 2	-2.3
BHLHE41	Basic Helix-Loop-Helix Family, Member E41	3.4	TMEM27	Transmembrane Protein 27	-2.2
FBP1	Fructose-1,6-Bisphosphatase 1	3.3	NOX1	NADPH Oxidase 1	-2.1
LXN	Latexin	3.1	JUN	Jun Proto-Oncogene	-2.1
EPPK1	Epiplakin 1	3.0	SNAR-E	Small ILF3/NF90-Associated RNA E	-2.1
PIGR	Polymeric Immunoglobulin Receptor	2.9	CXCL11	Chemokine (C-X-C Motif) Ligand 11	-2.1

Table 3.3: HPV18E6 Top 10 up- and down-regulated genes based on either the extent of up or down-regulation or on their known involvement in carcinogenesis

Up-regulated			Down-regulated		
Gene Symbol	**Gene Name**	**Fold Change**	**Gene Symbol**	**Gene Name**	**Fold Change**
CEACAM5	Carcinoembryonic antigen related cell adhesion molecule 5	22.7	ANKRD1	Ankyrin Repeat Domain 1 (Cardiac Muscle)	-4.8
CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1	13.5	BCLAF1	BCL2-Associated Transcription Factor 1	-4.5
MMP12	Matrix metalloproteinase-12	11.7	FST	Follistatin	-4.0
GNE	Glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	10.2	TRAJ59	T Cell Receptor Alpha Joining 59 (Non-Functional)	-4.0
S100A7	S100 Calcium Binding Protein A7	10.0	PTHLH	Parathyroid Hormone-Like Hormone	-3.9
PNLIPRP3	Pancreatic Lipase-Related Protein 3	9.9	IFI44L	Interferon-Induced Protein 44-Like	-3.7
FABP4	Fatty Acid Binding Protein 4, Adipocyte	9.6	THBS1	Thrombospondin 1	-3.3
SERPINB4	Serpin Peptidase Inhibitor, Clade B (Ovalbumin), Member 4	9.0	ODC1	Ornithine Decarboxylase 1	-3.2
AKR1C2	Aldo-Keto Reductase Family 1, Member C2	8.4	ZIC1	Zinc finger of the cerebellum Family Member 1	-3.2
TRIM31	Tripartite Motif Containing 31	8.4	TGFB2	Transforming Growth Factor, Beta 2	-3.1

The qRT-PCR results showed that trends in expression change for the selected genes were consistent with the microarray results, verifying that the microarray results were accurate and reliable (Fig. 3.6).

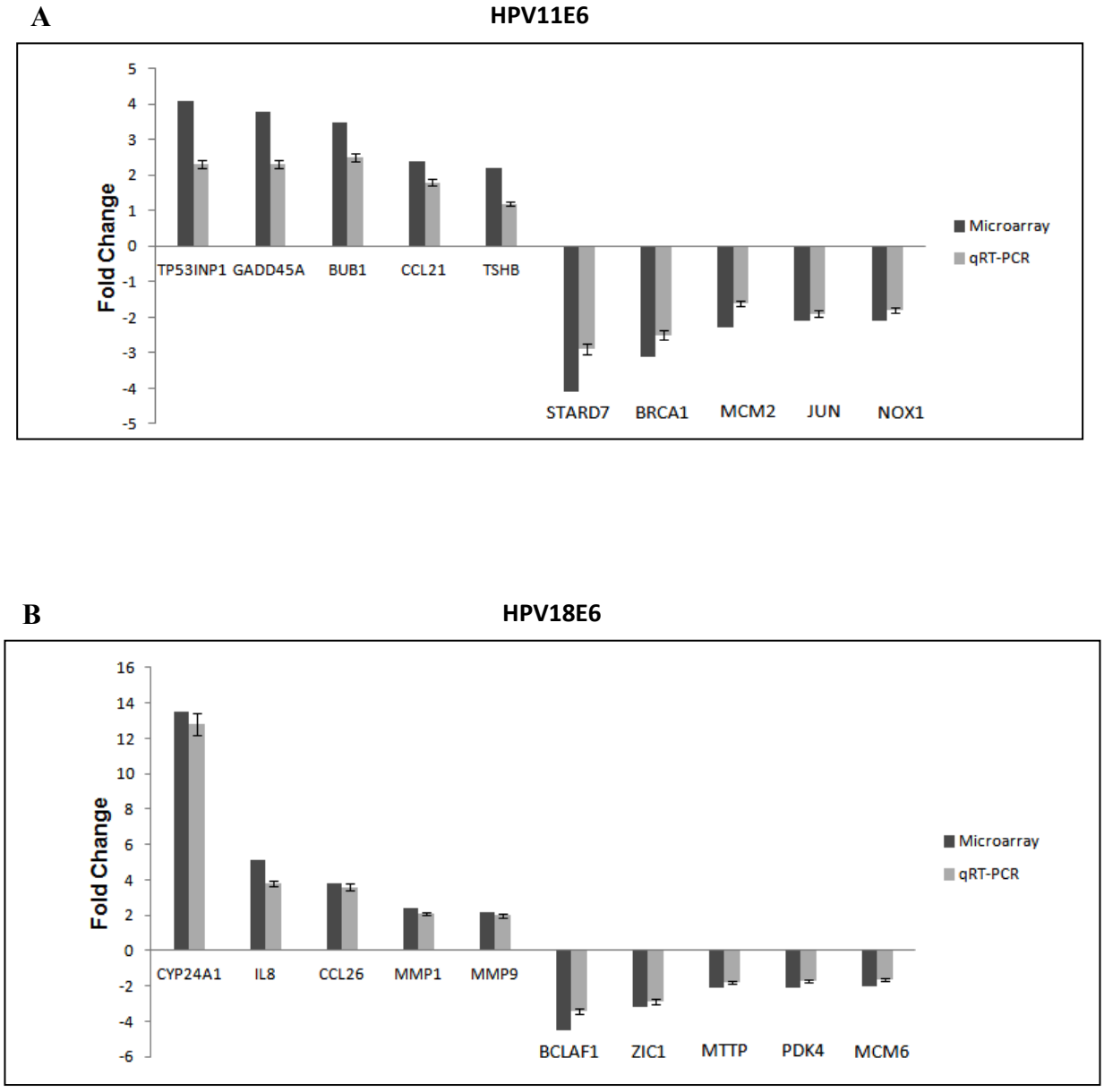


Fig. 3.6: qRT-PCR validation of microarray data. Selected genes with altered expression in response to HPVE6 in **A:** HPV11E6 & **B:** HPV18E6. Fold change shown was normalized to GAPDH mRNA levels and qRT-PCR figures represent the average of three replicates for each infection with the error bars indicating standard deviation.

An in depth analysis of the functional significance of gene expression was undertaken by examining genes that were both common and unique to HPV11E6 and HPV18E6. This was done by analysing the gene ontologies, signalling pathways and networks in which the significantly differentially expressed genes were involved.

3.2.4 Gene Ontology

To determine whether the differentially expressed genes were biologically meaningful in relation to the disease phenotype, the GOTree Machine (GOTM) (Zhang et al., 2004) was utilized to annotate gene functions and classify them into functional categories. Gene ontology was used to identify biological processes that were enriched for genes differentially expressed in both HPV11E6 and HPV18E6. Gene ontology (GO) processes that were present in both HPV11E6 and HPV18E6 would indicate in which way they were similar in their effects on cellular gene expression, while GO processes that were unique to either HPV11E6 or HPV 18E6 would provide insight into the distinct manner that the low risk and high risk HPVE6 genes affect cellular genes. By using all the genes in the human genome (Lander et al., 2001; ENCODE Project Consortium, 2004; Clamp et al., 2007; Pertea & Salzberg, 2010; Wu et al., 2011) as the reference gene set, this study integrated GO categories that were enriched in the set of 3210 genes (Fig. 3.5). Raw p-values generated from GOTM were adjusted using the false discovery methods and the overall significance level was set at $\alpha < 0.05$.

The results revealed numerous GO functions of differentially expressed genes (DEGs), HPV18E6 seen to exhibit more DEGs compared to HPV11E6 and that participated in a variety of functions associated with the greater cancer phenotype. These included metabolism, response to stress, cell proliferation, protein binding, cell cycle control and cell communication. Importantly, these DEGs were also involved in cell shape and cytoskeletal re-organization; all well known features of cancer cells. The results in Fig. 3.7A revealed numerous biological processes unique to HaCaT cells expressing HPV11E6 with 536 genes being involved in the metabolic processes. Further resolution of significantly enriched biological process (Fig. 3.7B) revealed response to stimulus and stress. Biological processes unique HaCaT cells expressing HPV18E6 also had 797 genes (Fig. 3.8A) being associated with the metabolic processes with further resolution of significantly enriched biological process revealing response to stimulus and stress, cell communication, cell cycle and cell signalling (Fig. 3.8B). A look at the commonly DEGs by both HPV11E6 and HPV18E6 also

had 504 genes (Fig. 3.9A) being associated with the metabolic processes with a further resolution of significantly enriched biological process being involved in response to stimulus and stress, cell communication and signal transduction. The regulation of the metabolic process as the most regulated biological process by the HPVE6 DEGs was anticipated since tumours are known to be metabolically active and develop a highly specialized metabolic profile (Griffin & Shockcor, 2004; Zhou et al., 2009; Fulda et al., 2010), the regulation of important functions such as cell cycle, cell communication and signalling suggested an elevated cell proliferation capacity as a result of HPVE6 gene expression. Since these DEGs appeared to have key enriched GO functions fitting the cancer hallmarks the next step was to map these genes to known pathways so as to gain more insight into the underlying biology of these DEGs since pathway analysis reduced complexity and has increased explanatory power as opposed to the GO analysis.

3.2.5 Pathway Analysis

Pathway analysis is a computational approach used to investigate network behaviour as a system (Viswanathan et al., 2008) and it has become the first choice for extracting and explaining the underlying biology for high-throughput molecular measurements (Khatri et al., 2012). To determine the possible mechanisms underlying the significantly differentiated genes a detailed pathway analysis was performed using the Ingenuity™ Pathway Analysis (IPA). This transformed the significantly differentiated genes into a set of relevant networks based on structured content from Ingenuity's Pathways Knowledge Base (IPKB). The results are presented graphically based on scoring of the ratio of significant genes present in the canonical pathway to the total number of genes in the canonical pathway. The threshold level was set at $p < 0.05$. The Kyoto Encyclopedia of Genes and Genomes (KEGG) software was also used to integrate current knowledge on molecular interaction networks in biological processes. These analyses determined which pathways were dysregulated at the early stages of HPVE6 gene expression in HaCaT cells.

The results from the pathway analysis corresponded with the GO enrichment analysis results and showed a common trend in the regulation of pathways involved with signalling as a result of HPVE6 expression. Calcium signalling came out as the top significantly enriched pathway induced by HPV11E6 (Fig. 3.10A). This pathway has previously been associated with enhanced cell proliferation and impaired apoptosis (Capiod et al., 2008).

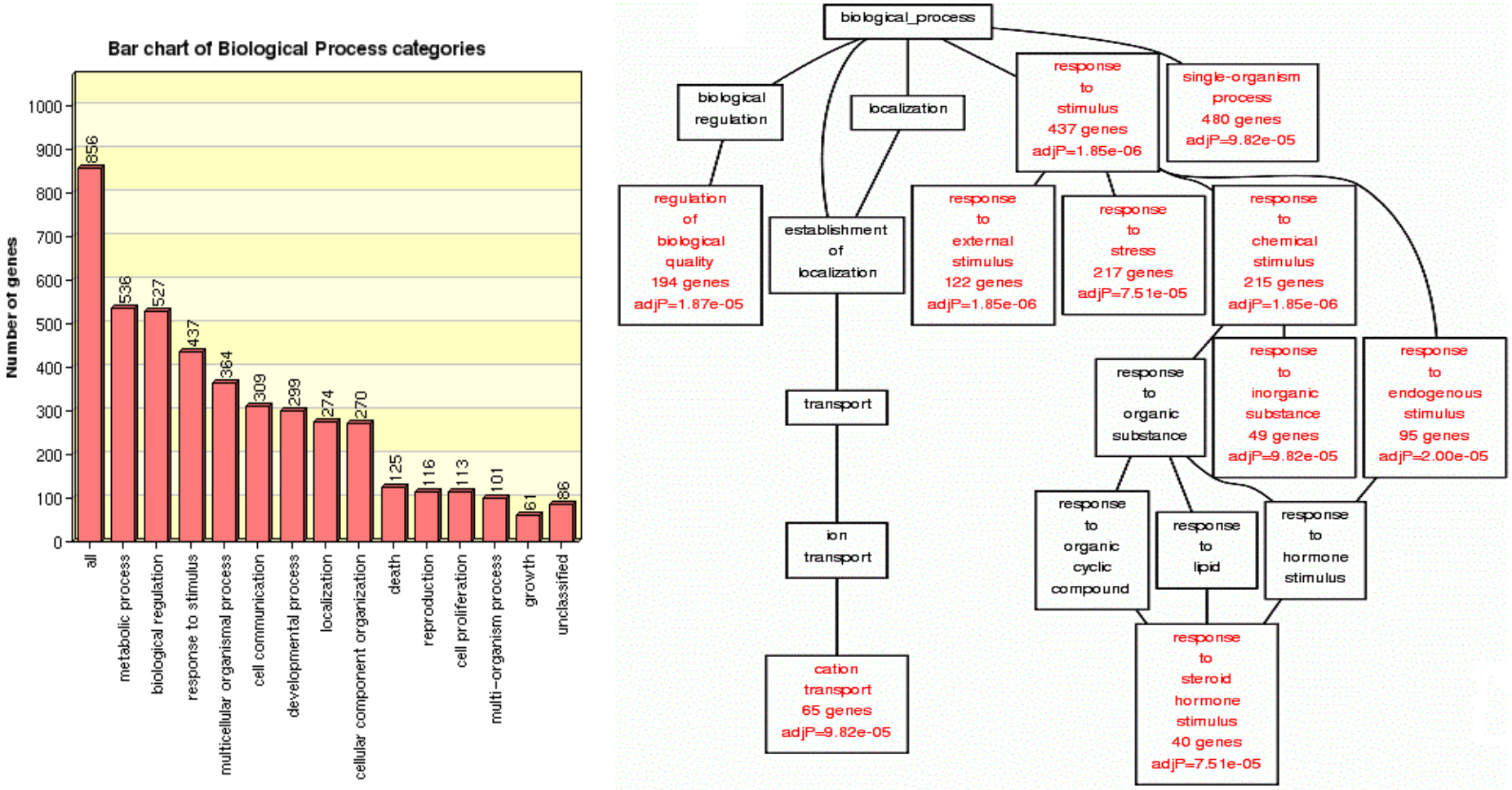


Fig. 3.7: Biological Processes by HaCaT cells expressing HPV11E6. A: Graphical representation enriched biological processes of DEGs. **B:** A directed acyclic graph showing a further resolution of the enriched biological processes of the DEGs (names in red) and their ancestor categories. Metabolic processes involved in response to stimulus specifically response to chemical stimulus, inorganic substance and endogenous stimulus were seen to be highly activated

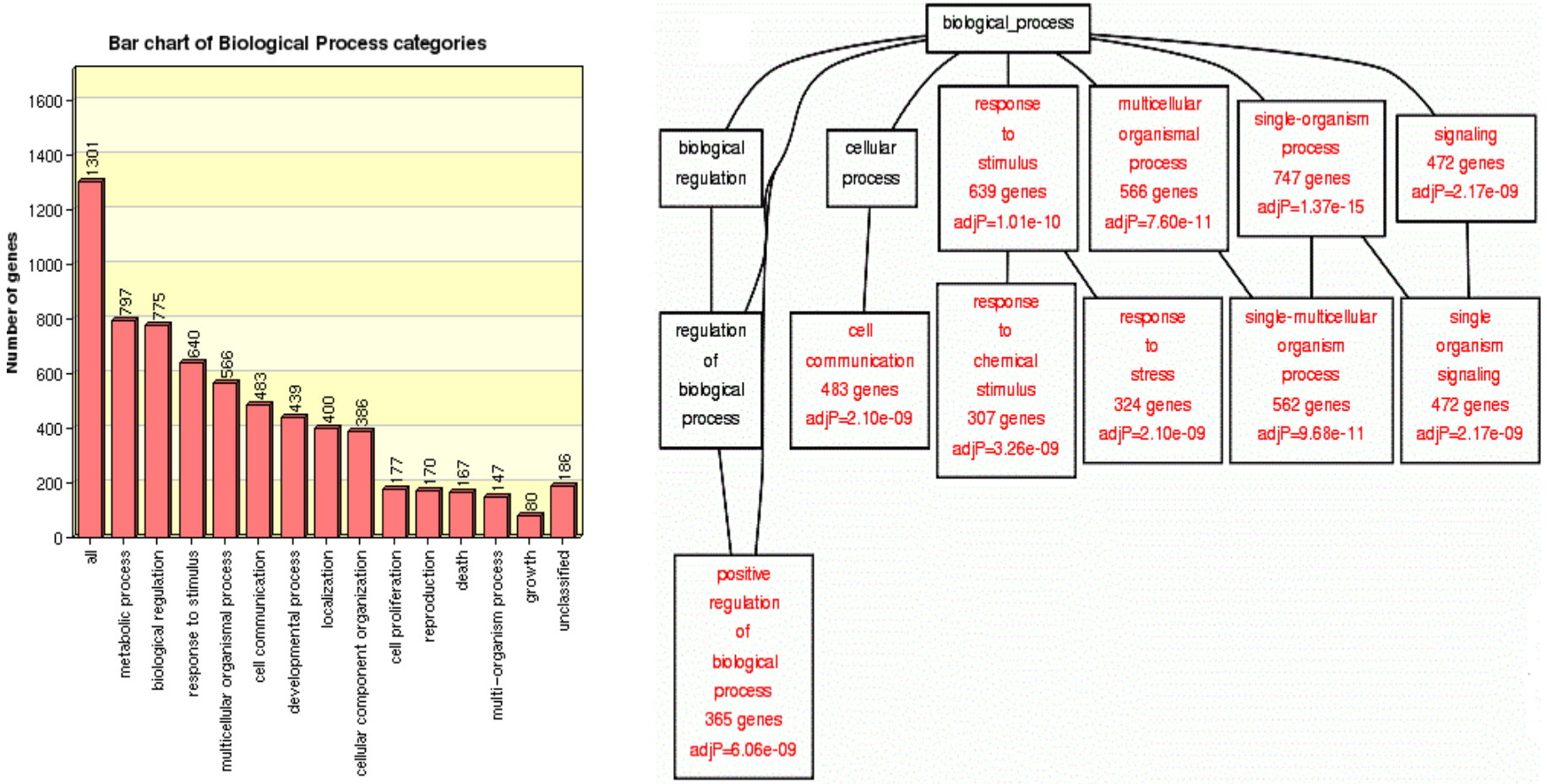


Fig. 3.8: Biological Processes by HaCaT cells expressing HPV18E6. A: Graphical representation enriched biological processes of DEGs. **B:** A directed acyclic graph showing a further resolution of the enriched biological processes of the DEGs (names in red) and their ancestor categories. Metabolic processes involved in response to stimulus specifically response to chemical stimulus, response to stress, cell communication, cell cycle and signalling were seen to be highly activated.

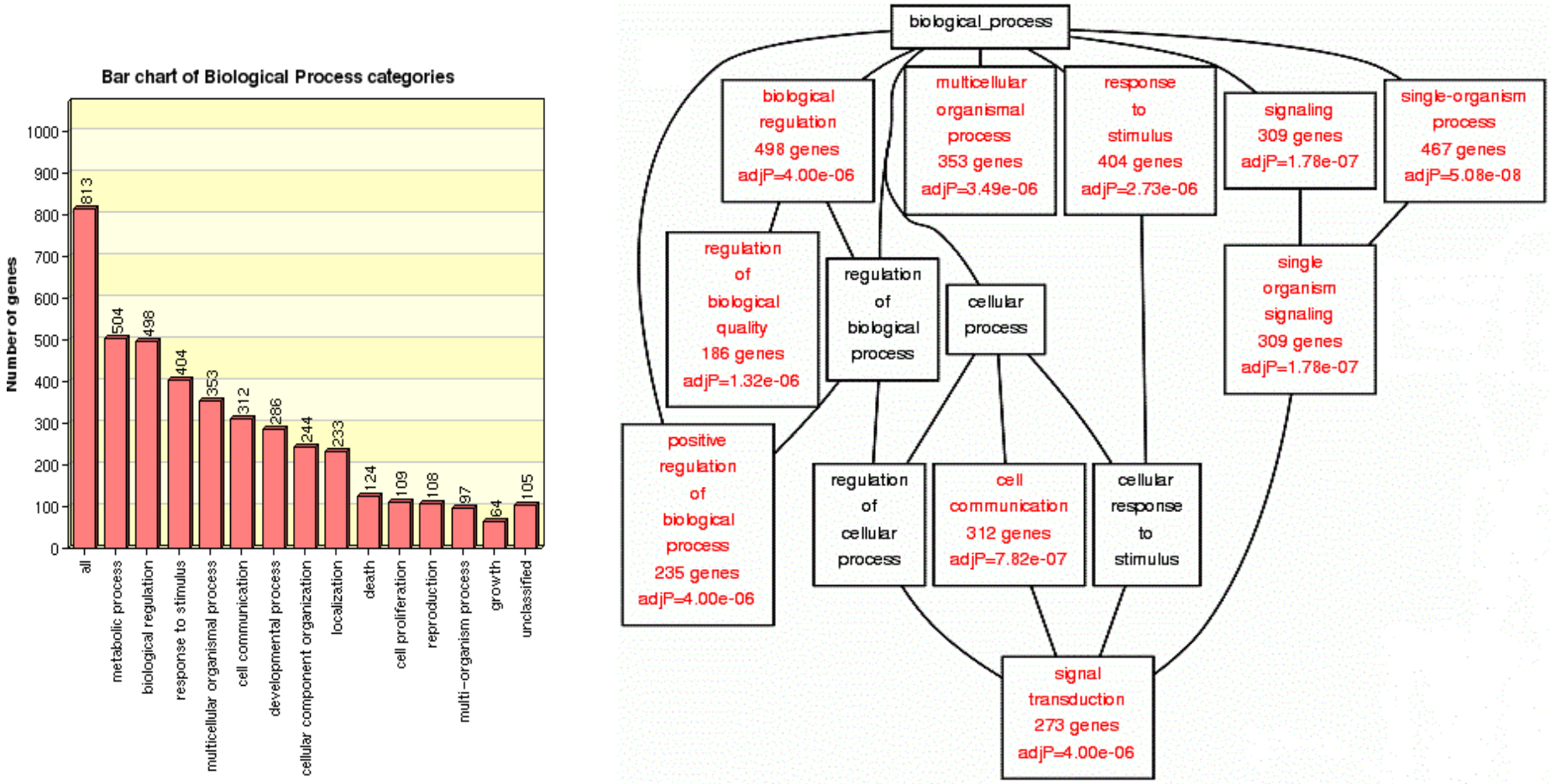


Fig. 3.9: Biological Processes by HaCaT cells expressing both HPV11E6 & HPV18E6. A: Graphical representation enriched biological processes of DEGs. **B:** A directed acyclic graph showing a further resolution of the enriched biological processes of the DEGs (names in red) and their ancestor categories. Metabolic processes involved in response to stimulus specifically response to chemical stimulus, cell communication, signalling and signal transduction were seen to be highly activated.

Other pathways of importance that were also dysregulated by HPV11E6 included *JNK* signalling, *mTOR* signalling and *GADD45* signalling that have previously been associated with viral infections and growth and development (Shuai & Liu, 2003; Alekseev et al., 2009; Laplante & Sabatini, 2012).

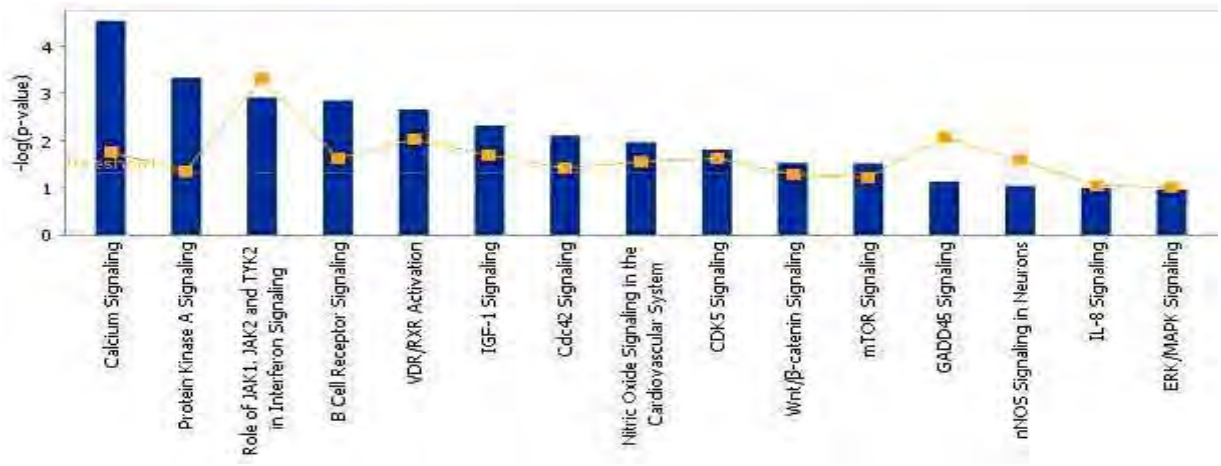
HPV18E6 induced the molecular mechanisms of cancer as the top dysregulated canonical pathway (Fig. 3.10B), while the DEGs shared by both HPV11E6 and HPV18E6 induced *PI3K/AKT* signaling as the top most regulated pathway (Fig. 3.10C).

HPV18E6 also induced other major signalling pathway involved in inter- and intra-cellular communication that lead to malignant phenotypes including the *Ras/integrin* signaling, *AKT* signaling, *TGF-β/BMP* signaling and *WNT* signalling (Fig 3.11). Other pathways of importance that were also dysregulated by HPV18E6 included *CCR3* signaling in eosinophils that has previously been associated with an infiltration of eosinophils in oesophageal squamous cell carcinoma patients (Fukuchi et al., 2011), Integrin signalling was also another pathway that has previously been associated with mutant *p53* in metastasis (Muller et al., 2009; Selivanova & Ivaska, 2009).

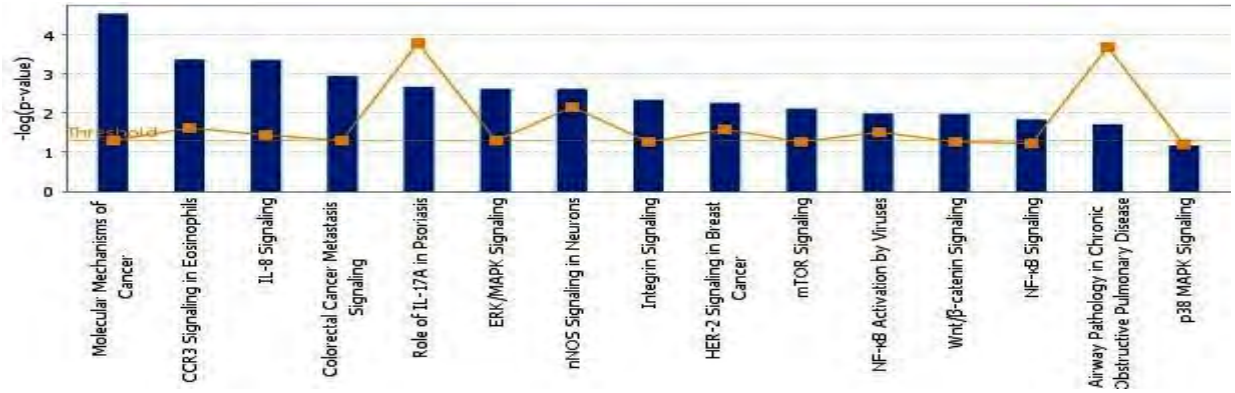
Airway Pathology in Chronic Obstructive Pulmonary Disease (COPD) canonical pathway was one of the interesting pathways that were dysregulated by HPV18E6 a pathway associated with cigarette smoking (Fig. 3.12), which is one of the predisposing factors of OSCC. Attention is also drawn to the up-regulation of *IL8, MMP1* and *MMP9* that have been shown to work together in conferring oesophageal tumorigenesis (Mukherjee et al., 2010; Shiau et al., 2013).

Other pathways also induced by the shared DEGs of both HPV11E6 and HPV18E6 as confirmed by the KEGG cancer pathway included the Cell cycle regulation, molecular mechanisms of cancer and *TGF-β* signalling Fig. 3.13.

A. HPV11E6



B. HPV18E6



C. HPV11E6 & HPV18E6

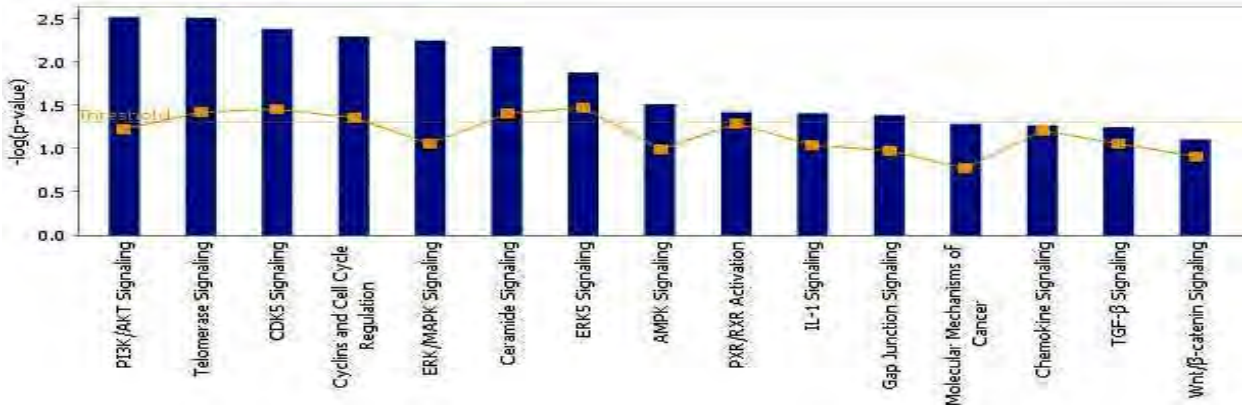


Fig. 3.10: IPA Canonical Pathway Analysis. Top pathways in which significant differentially expressed genes were enriched, following the expression of, **A:** HPV11E6; **B:** HPV18E6; **C:** HPV11E6 & HPV18E6 in HaCaT cells. Bars represent -log (p-value) for disproportionate representation of affected genes in the selected pathway, yellow boxed line represents the ratio of affected genes to the total number of genes in a pathway.

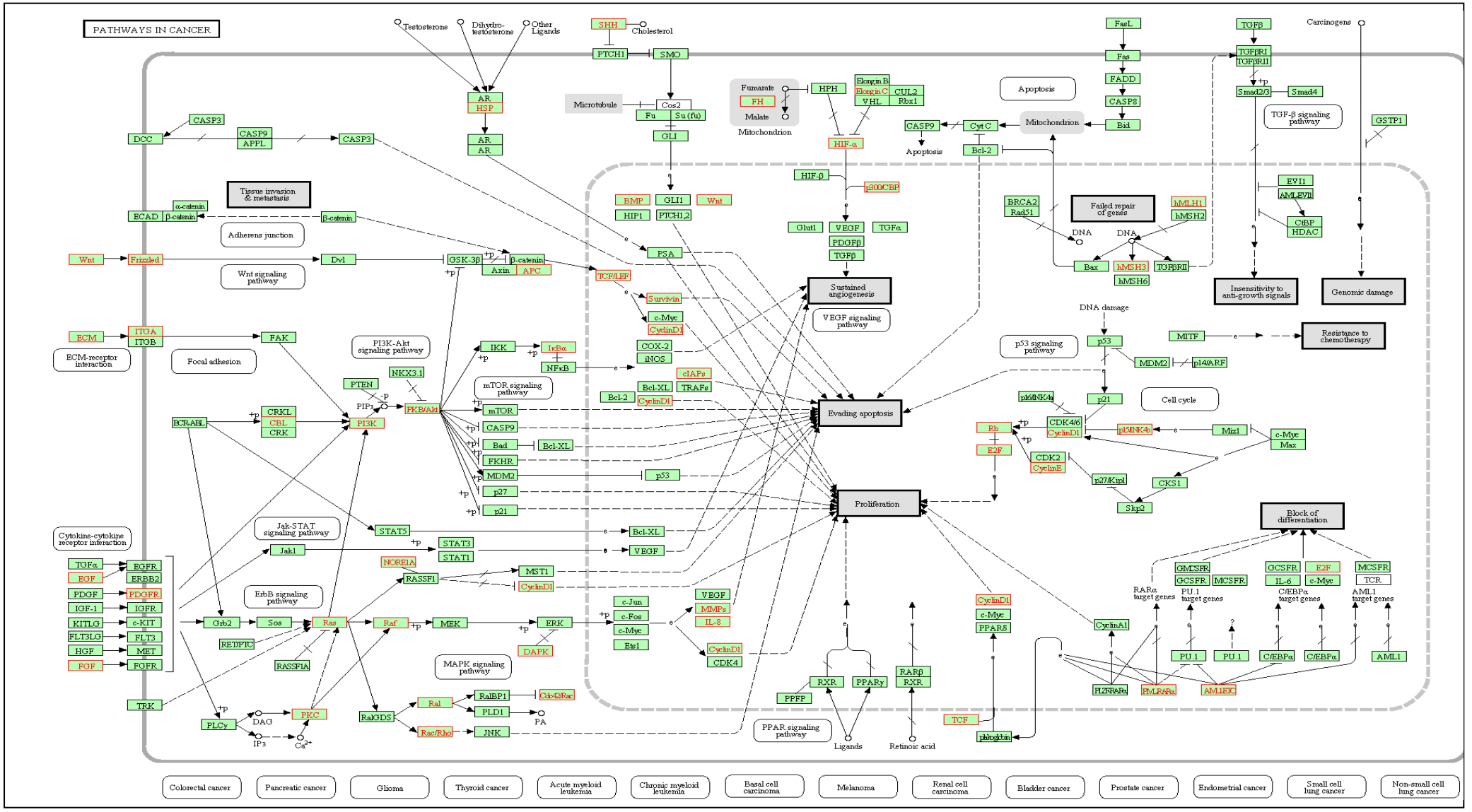
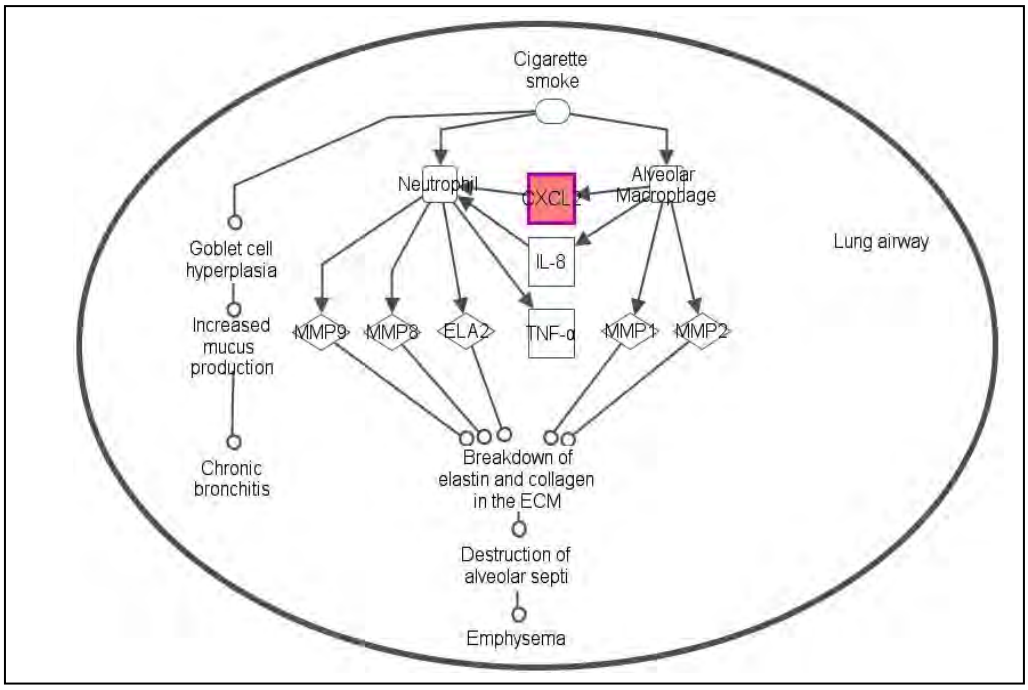


Fig. 3.11: Effect of significantly differentiated genes (Red) on different signaling pathways and their involvement at different stages of oncogenesis following the expression of HPV18E6 as represented by the KEGG Cancer pathway**.** Most of the DEGs were seen to regulate the cell cycle, apoptosis and the *WNT* pathway.

A. HPV11E6



B. HPV18E6

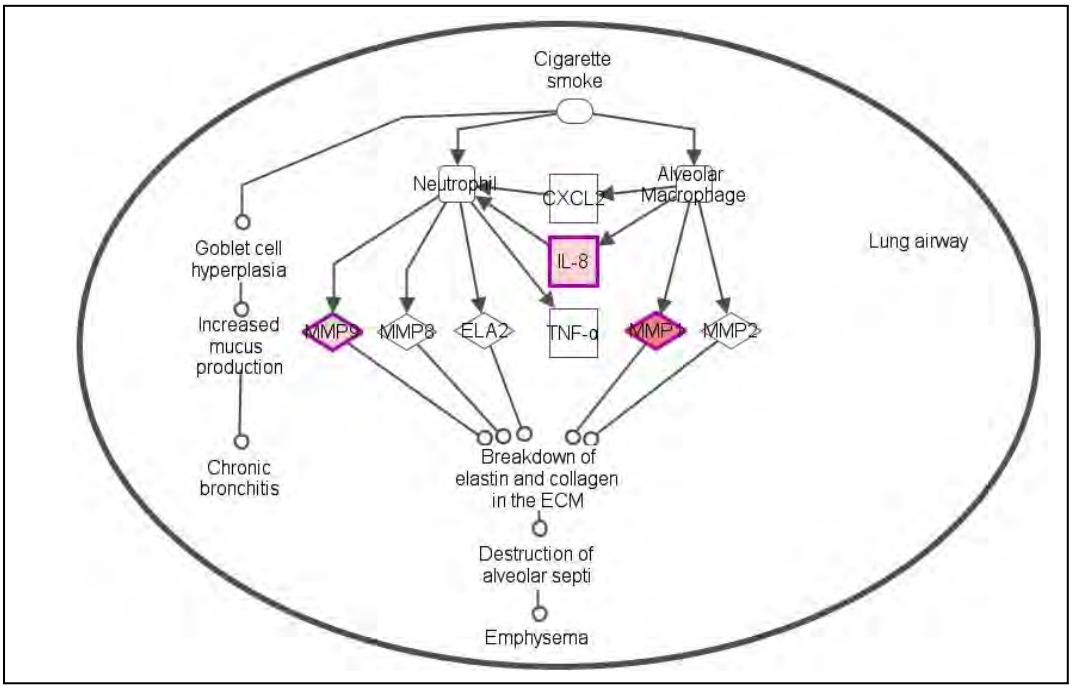


Fig. 3.12: Chronic Obstructive Pulmonary Disease (COPD) Canonical Pathway. Showing DEGs, following the expression of, **A:** HPV11E6 induced the upregulation of *CXCL2* gene only. **B:** HPV18E6 induced the upregulation of *IL8, MMP1* and *MMP9* genes. Red coloured genes indicate increased expression while the uncoloured genes were not identified as differentially expressed by HPVE6 but were rather integrated into the computationally generated pathway on the basis of the evidence stored in the IPA knowledge memory indicating a relevance to this pathway.

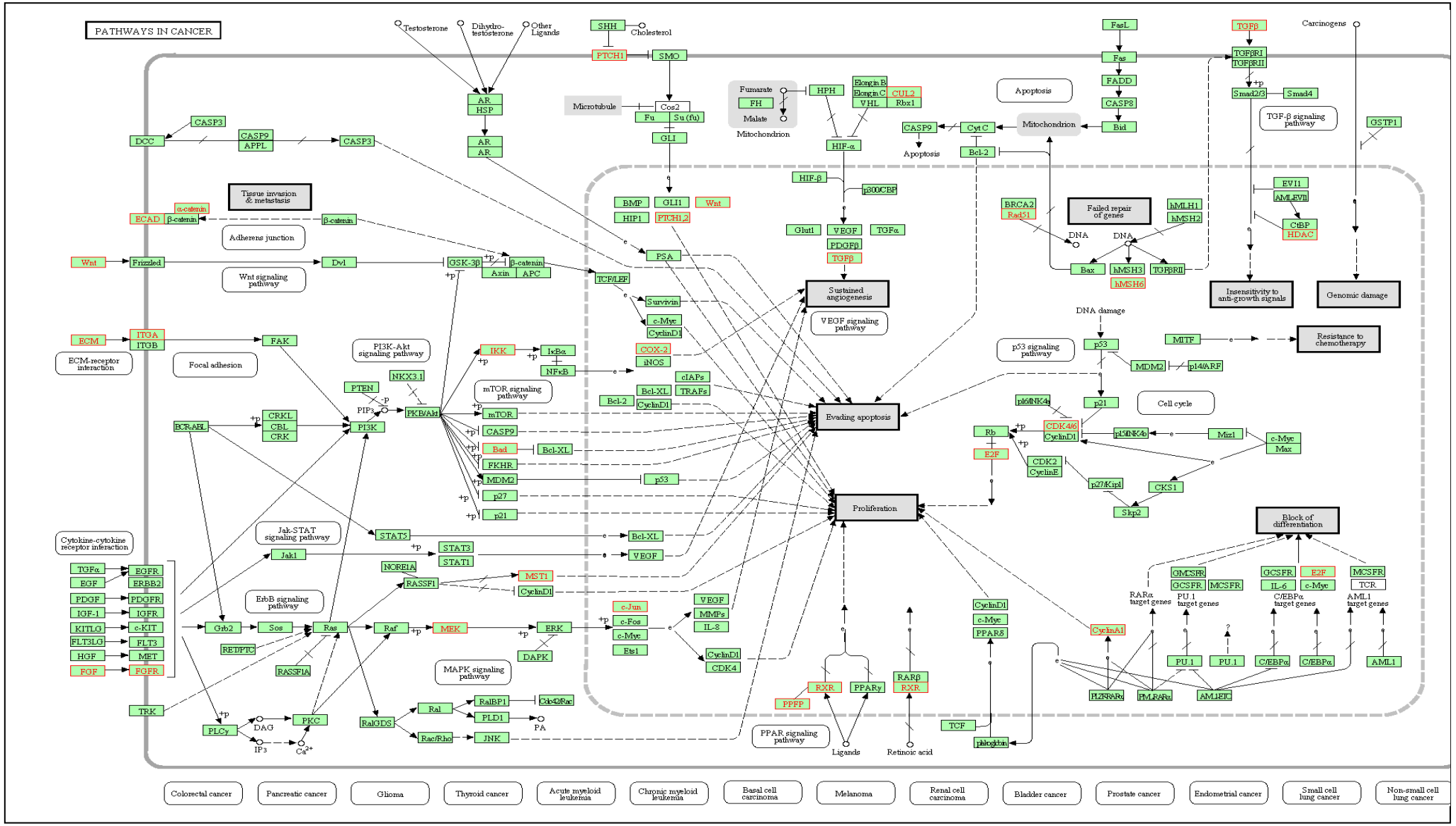


Fig. 3.13: Effect of significantly differentiated genes (Red) on different signaling pathways and their involvement at different stages of oncogenesis following the expression of both HPV11E6 & HPV18E6 as represented by the KEGG Cancer pathway**.** Most of the DEGs were seen to regulate the cell cycle and *WNT* pathway

3.2.6 Network Analysis

The IPA gene network system was used to generate networks to show how the significantly differentially expressed genes co-operate at the molecular level following HPVE6 expression. This process makes the assumption that the highly interconnected networks are likely to represent significant biological function and these are ranked on the basis of their scores. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible genes in the network, as well as the total number of analyzed network eligible genes and the total number of genes in the Ingenuity Knowledge Base that could potentially be included in networks (Tanaka et al., 2014).

The three top networks with the highest score following the expression of HPV11E6 (Table. 3.4) included gene products associated with *NFkB* and *PI3K complex*. While the top networks dysregulated following the expression of HPV18E6 (Table. 3.5) included gene products associated with *TNF*, a cytokine implicated in cancer and autoimmune diseases (Locksley et al., 2001) and *NFkB*, known to play a critical role in carcinogenesis (Tian et al., 2006). The top networks dysregulated following the expression of the shared DEGs by both HPV11E6 and HPV18E6 (Table 3.6) included gene products associated with the *NFkB complex.*

The expression of HPVE6 resulted in a significant up-regulation of *NFkB*, as well as the downstream genes and it was apparent that this pathway was induced by HPVE6. Even though these genes did not appear on the list of the selected top 10 up- or – down-regulated genes, the observed fold-changes were significantly higher than in the control cell lines. For this reason, further computational and experimental exploration of the regulation of *NFkB* was carried out, to elucidate the consequences of the upregulation of this inflammatory pathway by HPVE6.

The expression of HPV11E6 (Fig. 3.14) showed most of the genes that regulate *NFkB* are down-regulated, and are mostly involved in cellular movement and inflammatory response. Whereas, following the expression of HPV18E6 (Fig. 3.15), most of the genes that regulate *NFkB* are up-regulated, and are mostly important in encoding cellular growth, cell cycle and inflammatory response. The shared genes that regulate *NFkB* following the expression of both HPV11E6 and HPV18E6 (Fig. 3.16) are also mostly up-regulated and are shown to be important in cellular movement, cell death and cell cycle.

Table 3.4: Most significant three gene networks of DEGs following expression of HPV11E6

DEGs in Network	Score[a]	Focus DEGs[b]	Top Functions
CASR,CCL27,CD46,CDK5,CYR61,ERBB2,ERK,ERK1/2,Focal adhesion kinase, GAB1, Hsp70, Hsp90, HTR7, ITGB1, KPNA2, LIMK1, MET, MMP19, MYCN, NBN, NOV, NQO1, P38 MAPK, PACRG, PDGF, Pkc(s), PPP1CC, PTGES, PTMA, PXN, RBP1, RGS4, SERPINE1, VEGFC, XRCC3	26	27	Cellular Movement, Cellular Development, Cellular Growth and Proliferation
ADD2,Akt,AMIGO2,Ap1,CCL5,COL2A1,DDIT3,DRD2,Histone h3,HLA-A,IFN Beta, IFNG, IL18, IL1RAP, Interferon alpha, LDL, LGMN, MAP2K4, MICU1,MYD88, NFkB (complex), OAS1, PI3K(complex),PI3K(family),PRKCD, RARRES3, Ras, RHOB,RNA polymerase II, RPL12, SNW1, TCR, TNFRSF10B,XRCC5	18	22	Cell Death and Survival, Inflammatory Disease, Inflammatory Response
ABL1,ATXN1,BAG1,BCL2L1,BRCA1,EGFR,EPO,GH1,GHR,GPX1,GRIN1,GZMA,IGF1R,IL15,IL15RA,IL1B,JAK2,MOAP1,NR3C1,PDCD10,PPP1R15A,PPP2CA,PPP2R1B,PPP3R1,PRKCD,PRLR,RIN1,SBDS,SDCBP,SOX4,STAT5A,STAT5B,STK25,STK17B,TNFRSF1A	13	18	Cell Death and Survival, Cellular Growth and Proliferation, Cellular

a – Numerical value used to rank networks according to their degree of relevance between the network eligible genes in the network, DEGS in the network and the total number of genes in the Ingenuity Knowledge Base that could potentially be included in networks

b - Number of DEGS associated with the network

Table 3.5: Most significant three gene networks of DEGs following expression of HPV18E6

DEGs in Network	Score[a]	Focus DEGs[b]	Top Functions
BCL2L2,CBX5,CHRNA5,CREBBP,Cyclin A, CYP2B6, CYP2C19,DKK1,E2F2,E2F4,EFNA1,ENC1,HAS2,Histone h3,Histone h4, IL3RA, JMJD6, LRP5, MLL, MYBL2,NR1I3, NR3C1,PIK3CB,PRKAR1B,PRKCB,RB1,RBL1, SON, SUMO2, THBD, TNFSF9, TNFSF11,TYMS,USP4	27	31	Cellular Growth and Proliferation, Cell Death and Survival, Cellular Development
Akt,AXL,BTG2,CAT,DES,ERBB3,FPR2,IFN Beta, IFNA2,IL8,IL32,IL1RL1,IL1RN,IRAK1,IRAK2,IRF1,IRF2, LTF,MMP9,NFkB (complex), NFKBIA, PI3K (family), Ras, RIPK2, RRM2, RUSC2, SHH, SUMO4, TICAM1,TNFAIP3, TNFSF4, TNFSF18,TRIP6,UBD,USP18	25	30	Cellular Development, Cellular Growth and Proliferation, Hematological System Development and Function
ABL1, ARNTL, BBC3, CASP2, CDKN1A, CDKN2A, CREBBP,E2F1,FOXO3, HIVEP1,LIG1, MDM2, MYC, NVL,PHF1,PIM1,PTEN,RB1,REL,RGS3,RPL5,RPL11,SAT1,SFN,SNAI1,SYNE2,TERT,TP73,TP53I3,TRIM32,U2AF2	8	17	Cell Cycle, Cellular Growth and Proliferation, Cancer

a – Numerical value used to rank networks according to their degree of relevance between the network eligible genes in the network, DEGS in the network and the total number of genes in the Ingenuity Knowledge Base that could potentially be included in networks

b - Number of DEGS associated with the network

Table 3.6: Most significant three gene networks of DEGs following expression of both HPV11E6 & HPV18E6

DEGs in Network	**Score**[a]	**Focus DEGs**[b]	**Top Functions**
ADCY,ADRB2,Ap1,APOH,ARF6,CCL2,CDH1,CPT1A,CYP1B1,ELF3,ERK,FGFR1,GATA4,GNAI2, GNAS,HDAC3,HNF4G,ITGA2,ITGB4,JUN,LDL,MAP3K1,MAP3K8,Mapk,MCL1,Mek,MUC4,N-cor,NFkB(complex),P38MAPK,Pkc(s),PTGS2, RPL30, SDC4, USP11	22	25	Cellular Movement, Cell Death and Survival, Cancer
ABL1,ANKRD1,ARHGDIB,CARM1,CCL19,CLU, CSNK1A1,CTSD,DBP,E2F4,EIF4G1,HEXB,HEY1, HIST1H3A, MREG,NDRG1, NFKB2, ORAI1, PCGF2, PPP1R13L, PTTG1, PYCARD, RELA, SMAD3,TAP2,TGFB2,TMPO,TNIP2,TOP1,TOP2A ,TP53,TP53BP2,TWIST1,USP2,WASF1	10	16	Cell Death and Survival, Gene Expression, Cellular Development
ACACA,AGTR1,BSN,CKAP4,CNGA3,*CYP1A1*,DIO1,EPO,FASN,Fcgr3,FOXO3,GATA3,HDAC1,HTR2B,IL6,IL7,IL1B,KDM5B,KIF2C,MT1X,NAA10, ORM1,PHF21A,PLS3,PRKAA2,PRKAB1,PRKAG1 ,RAB11FIP5,RPS6KA1,S1PR2,SCNN1A,SGK1,SREBF1,TF,TREM1	9	15	Cell Cycle, Hematopoiesis, Cardiovascular System Development and Function

a – Numerical value used to rank networks according to their degree of relevance between the network eligible genes in the network, DEGS in the network and the total number of genes in the Ingenuity Knowledge Base that could potentially be included in networks

b - Number of DEGS associated with the network

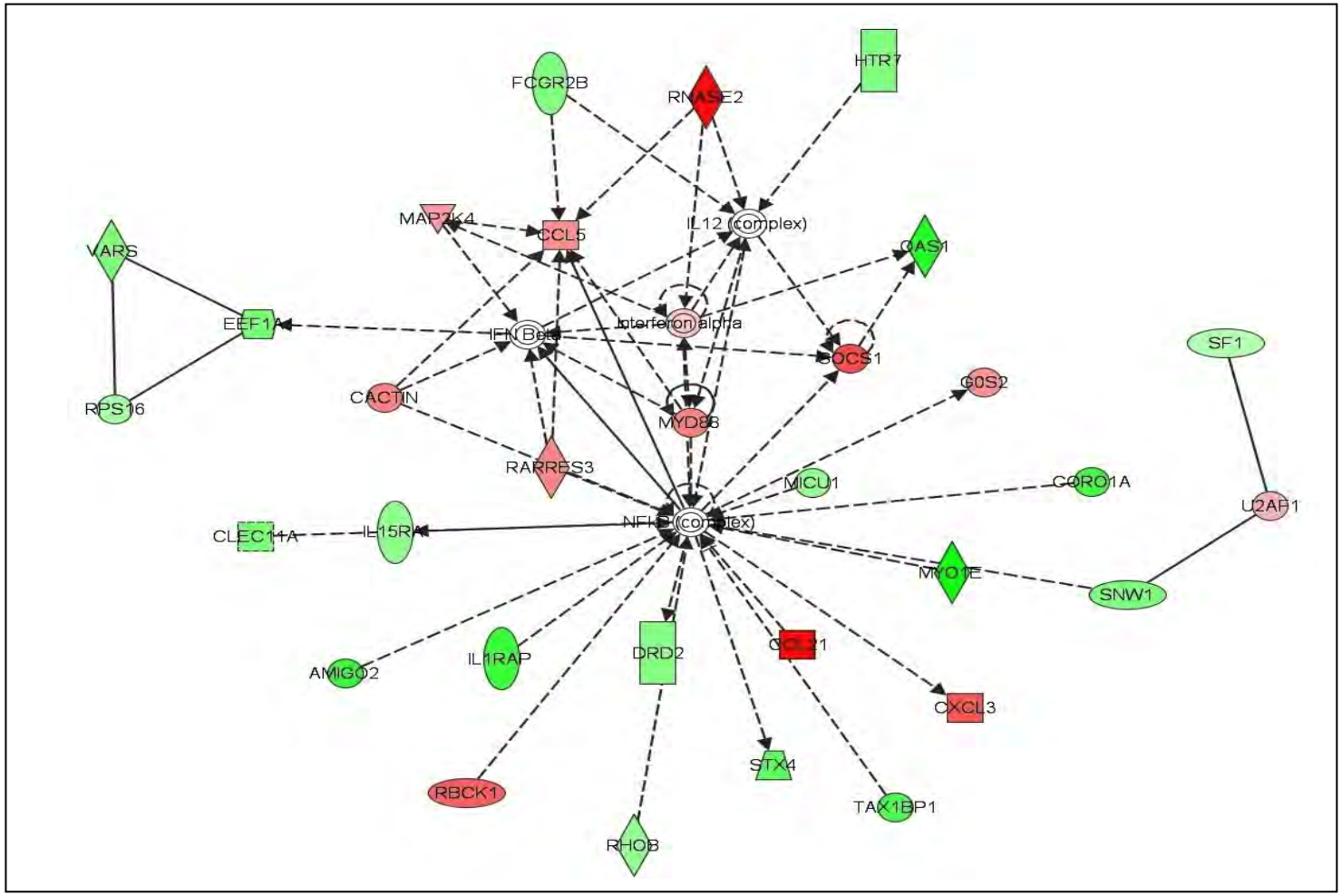


Fig. 3.14: Network of *NFκB* interacting genes. IPA was used to construct a network of DEGs identified around *NFκB complex* following the expression HPV11E6 in HaCaT cells. The fold changes for the genes were superimposed onto this network showing most of the genes being down-regulated. The genes in this network are important in cellular movement and inflammatory response. Straight line (direct interaction) and broken line (indirect interaction) and the intensity of the colour indicating the degree of regulation i.e. Green indicated reduced expression, red indicated increased expression while the uncoloured genes were not identified as differentially expressed by HPVE6 but were rather integrated into the computationally generated networks on the basis of the evidence stored in the IPA knowledge memory indicating a relevance to this network.

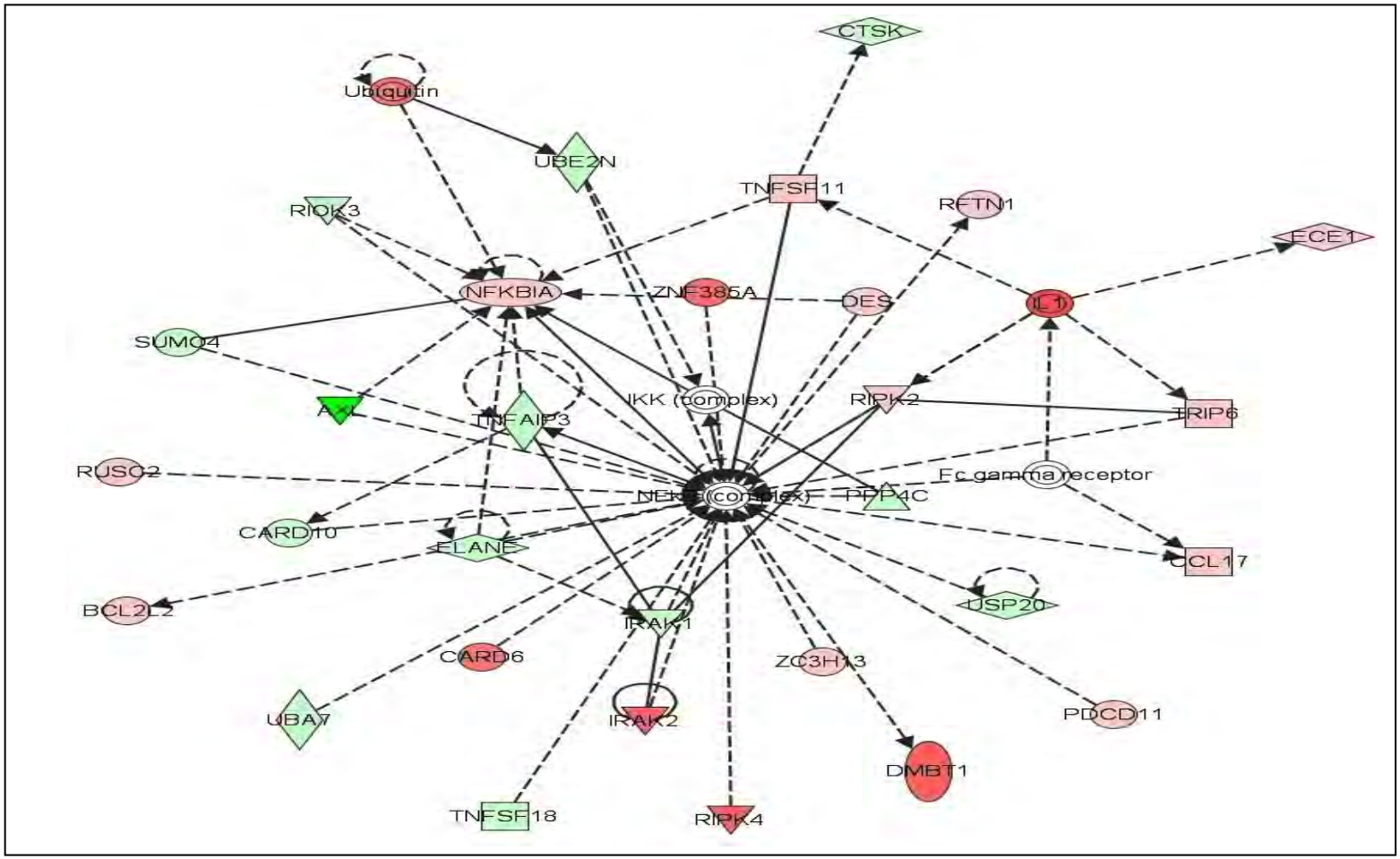


Fig. 3.15: Network of *NFκB* interacting genes. IPA was used to construct a network of DEGs identified around *NFκB complex* following the expression HPV18E6 in HaCaT cells. The fold changes for the genes were superimposed onto this network showing most of the genes being up-regulated. The genes in this network are important in cellular growth, cell cycle and inflammatory response. Straight line (direct interaction) and broken line (indirect interaction) and the intensity of the colour indicating the degree of regulation i.e. Green indicated reduced expression, red indicated increased expression while the uncoloured genes were not identified as differentially expressed by HPVE6 but were rather integrated into the computationally generated networks on the basis of the evidence stored in the IPA knowledge memory indicating a relevance to this network.

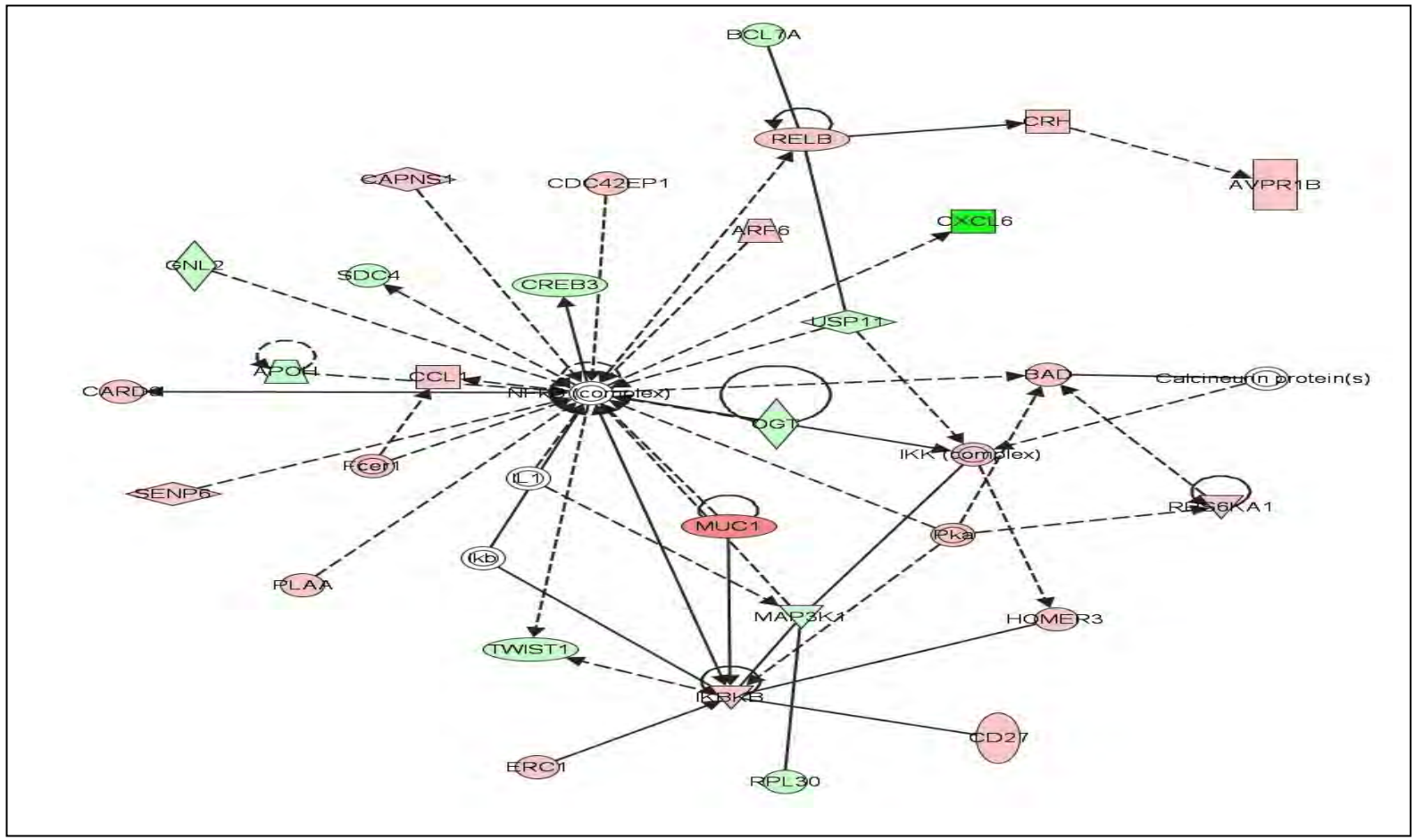


Fig. 3.16: Network of *NFκB* interacting genes. IPA was used to construct a network of shared DEGs identified around *NFκB complex* following the expression of both HPV11E6 and HPV18E6 in HaCaT cells. The fold changes for the genes were superimposed onto this network showing most of the genes being up-regulated. The genes in this network are important in cellular movement, cell death and cell cycle. Straight line (direct interaction) and broken line (indirect interaction) and the intensity of the colour indicating the degree of regulation i.e. Green indicated reduced expression, red indicated increased expression while the uncoloured genes were not identified as differentially expressed by HPVE6 but were rather integrated into the computationally generated networks on the basis of the evidence stored in the IPA knowledge memory indicating a relevance to this network.

3.3 Discussion

DNA microarray analysis was used to explore the gene expression patterns in HaCaT cells during the early stages of the E6 gene expression, from either the low risk HPV11 or the high risk HPV18. The microarray data was first normalised so as to minimize the effects that were caused by technical variations; this allowed the data to be comparable in order to find the actual biological changes (Smyth & Speed, 2003; Fujita et al., 2006).

The gene expression pattern in the HPVE6 expressing cells was shown to be significantly different using PCA and hierarchical clustering analysis with some similarity between the HPV11E6 and HPV18E6 induced patterns as shown by the overlap of genes with the PCA. This was consistent with previous studies that showed an overall similarity in the high-risk HPV18E6 and low-risk HPV11E6 induced proteins due to probable common biochemical activities (Huibregtse & Beaudenon, 1996; Oh et al., 2004). Other studies also suggested that low-risk forms of E6 bind to p53 with a low affinity (Werness et al., 1990; Huibregtse et al., 1993b; Li & Coffino, 1996). The *in vivo* and *in vitro* functional inactivation of p73 by both high-risk HPV18E6 and low-risk HPV11E6 has also been demonstrated previously (Park et al., 2001).These results suggested that both the high risk HPV18E6 and low risk HPV11E6 possibly share some common biological pathways capable of deregulating gene expression and suggested a role for HPV11E6 in cellular transformation.

It is notable that the "metabolic process" was the most significantly enriched gene ontology function in the biological processes having the highest number of DEGs affected by HPVE6. This possibly may be as a result of increased energy metabolism that could enhance HPV replication since HPV replicates in terminally differentiating cells that are likely to have a low nutrient supply (Vande Pol & Klingelhutz, 2013). Importantly, these DEGs were also involved in the deregulation of cell shape, cell proliferation, transport and cytoskeletal re-organization; all of which are well known features of cancer cells. Previous studies have shown that the process of cellular transformation is closely linked to dysregulated proliferation, leading to metabolic transformation (El Mjiyad et al., 2011). Although the specific mechanisms of transformation vary among viruses, the net effect is often the same as with DNA damaging effects inducing cellular mutations and viral oncoproteins leading to constitutive activation or stabilization of oncogenes or inactivation of tumour suppressor genes to facilitate cellular transformation (Noch & Khalili, 2012).

The DEGs induced by both HPV18E6 and HPV11E6 were involved in response to cell signalling, cell communication and signal transduction processes. This result meant that HPV11E6 induced genes had the ability to influence processes known to induce cellular transformation. Thus, these activities may reflect common pathways being utilized by both types of low and high risk E6 viruses during their productive life cycles.

The course of cellular transformation begins with dysregulation of cell signaling processes induced by genetic, environmental factors or can be triggered by oncogenes encoded by viruses that are known to induce malignant transformation (Noch & Khalili, 2012). HPVE6 and Herpes Simplex Virus 1 (HSV1) are known to suppress *IFN* signaling in order to allow more efficient replication (Yokota et al., 2004). In this study HPV11E6 was shown to regulate a number of signalling pathways such as the *JAK1*, *JAK2* and *TYK2* interferon signalling. Li et al., (1999) also reported that *Tyrosine kinase 2* (*Tyk2*) associates with E6, with the Tyk2/E6 interaction being stronger with HPV18E6 than HPV11E6, this was because HPV18E6 binds to the Interferon Regulatory Factor 3 (IRF3) and inhibits its ability to activate interferon-responsive genes (Ronco et al., 1998). A possible explanation for the difference between the low and high risk HPVE6 may lie within the transactivation capacity of HPVE6, which is mediated through interactions with the transcriptional machinery.

The expression of HPV18E6 in this study lead to deregulation of pathways that were not affected by expression of HPV11E6; with the most significant being the Molecular Mechanisms of Cancer pathway, *CCR3* Signalling in Eosinophilis, *IL8* Signalling and the Airway Pathology in Chronic Obstructive Pulmonary Disease *(COPD) Pathway.* The regulation of the cancer pathway as the most significantly regulated pathway upon HPV18E6 expression was not unexpected, since genetic alterations such as mutations or deletions of genes are known to drive tumour progression by generating mutations in oncogenes and tumour suppressor genes (Fearon & Vogelstein, 1990). HPV18E6 is known to account for at most 16-30% of cervical cancers and is considered a major cause of other cancers (Ajay et al., 2012; Yang et al., 2014). HPV18E6 is also known to cause transformation, immortalization and promote carcinogenesis primarily by binding to *p53*, resulting in deregulation of cell cycle checkpoints (Scheffner et al., 1993; Boyer et al., 1996; Bischof et al., 2005). This was evident in the current study where genes such as *CyclinD1, CyclinE, Rb, E2F, MMPs, IL8, Ras, HIF-α, P300CBP* and *WNT* were up-regulated following the expression of HPV18E6 and these were involved in the dysregulation of biological processes

implicated with cellular transformation especially cell proliferation, apoptosis, cell cycle and angiogenesis known hallmarks of cancer (Hanahan & Weinberg, 2011). The expression of HPV18E6 also lead to the upregulation of *CCL26* and *CCR3* genes that have both being implicated with the dysregulation of *CCR3* signalling in eosinophilis pathway. It was hypothesised that these genes dysregulated this pathway through the recruitment of eosinophils to sites of HPV18E6 infection leading to tissue damage by the release of reactive oxygen species (ROS) and toxic granule proteins (Kampen et al., 2000; Sabroe et al., 2000). This was consistent with data on an oesophageal squamous cell carcinoma (OSCC) patient following oesophagectomy that also showed a marked eosinophil infiltration to the oesophagus (Fukuchi et al., 2011).

The *COPD* Pathway was also one of the interesting pathways that were significantly dysregulated by the expression of HPV18E6 and not by HPV11E6, since the *COPD* pathway was not directly related to HPV infection or oesophageal cancer but it showed an up-regulation of *IL8, MMP1* and *MMP9* genes that are implicated in OSCC. MMPs have been shown to be elevated in many cancers including OSCC (Shima et al., 1992; Mathew et al., 2002; Ishibashi et al., 2004; Samantaray et al., 2004; Gu et al., 2005; Li et al., 2009; Mukherjee et al., 2010), and have been closely associated with persistent inflammation and cancer progression (Balkwill & Mantovani, 2001; Bingle et al., 2002; Coussens & Werb, 2002; Owen et al., 2004; Björklund & Koivunen, 2005). Gene profiling studies have also associated *MMP1* and *MMP2* involvement with the ability of human breast cancer cells to metastasize spontaneously from the primary mammary tumour to the lungs of immunodeficient mice (Minn et al., 2005). IL-8 has been associated with metastasis in several cancers, its expression correlates with angiogenesis and tumorigenicity in numerous xenograft and orthotopic *in vivo* models (Deryugina & Quigley, 2006). Other studies have also reported the overexpression of IL8 in human prostate cancer cells, with concomitant up-regulated *MMP9* expression and collagenase activities promoting tumour cell growth and metastasis (Inoue et al., 2000). It is possible from these results, that HPV18E6 expression may potentiate cellular transformation of HaCaT cells via the deregulation of *MMP* and *IL8* genes; this was supported by our earlier results presented in chapter two.

The most significantly dysregulated pathways as a result of DEGs shared by both HPV11E6 and HPV18E6 included the *PI3K/AKT* signalling, telomerase signalling and the cell cycle regulation. Previous studies have shown that the *PI3K/AKT* pathway is overactive in many

types of tumour cells and exhibits significant impact on glycolysis by stimulating expression of key glycolytic enzymes (Sabbatini & McCormick, 1999; Elstrom et al., 2004). Thus it may be hypothesised that HPVE6 mediated cellular transformation through this pathway may lead to concomitant increases in glucose uptake and glycolytic flux, further strengthening the cellular transformation potential (Noch & Khalili, 2012).

It is known that integrins are expressed at low levels in normal epithelial tissues, but that they are upregulated during wound healing and in squamous cell carcinoma (Breuss et al., 1995). Integrins also promote the invasive growth of carcinomas that have become resistant to the anti-proliferative effect of *TGF-β* by enabling matrix degradation through the recruitment of activated *MMP2* (Guo & Giancotti, 2004). It is possible that DEGs shared by both the datasets deregulate the *PI3K/AKT* pathway by activating matrix-degrading proteases through elevation of integrins. Roymans & Slegers, (2001), also demonstrated that an increase in copy number of specific phosphatidylinositol 3-kinase genes or molecular mutation of these genes, lead to a gain of function reflected by a higher or a constitutive expression of the phosphatidylinositol 3-kinase lipid kinase activity and to the oncogenic transformation of many cell types.

The deregulation of the telomerase signalling pathway by the shared DEGs could be due to the DNA damage response system following the expression of HPVE6 although its role in cellular transformation could not be ascertained. Other studies have shown that the high risk HPV E6 was capable of stimulating telomerase activity independent of the effects of E6 on p53 (Klingelhutz et al., 1996; Gewin & Galloway, 2001). Keratinocytes expressing wild type HPV16E6 were shown to have an extended lifespan but did not become immortal (Klingelhutz et al., 1994), indicating that telomerase activation and E6 mediated degradation of p53 are insufficient for their immortalization. This was also supported by the studies of Klingelhutz et al., (1996) who showed that telomerase activation was an intrinsic but insufficient, component of cellular transformation by HPVE6.

The deregulation of the cyclins and cell cycle regulation pathways by the shared HPVE6 DEGs presented in this study is consistent with other studies that have shown that cell cycle progression is regulated by biochemical reactions that involving cyclins and cyclin-dependent kinases. This regulation prevented the replication of damaged or altered DNA (Southern & Herrington, 1998). The high risk HPV18 E6 also abrogates multiple cell cycle checkpoints that contribute to genomic instability (Hanahan & Weinberg, 2011). The E6 proteins of the

high-risk HPVs are capable of binding to cell cycle regulatory proteins, and interfere with both the G1/S and G2/M cell cycle checkpoints more effectively than the E6 proteins of the low-risk HPVs (Southern & Herrington, 1998; Cho et al., 2002; Brenna & Syrjänen, 2003; Davy & Doorbar, 2007). This was confirmed by the upregulation of G2/M specific genes *E2F1 and CDKN2B* by HPV18E6 (see Fig. 3.11). The inability of the low risk as opposed to the high-risk HPV types to induce immortalization and transformation reside in their abilities to interact with the cell cycle components that result in the loss of multiple cell cycle checkpoints responsible for maintaining host genome fidelity (Syrjänen & Syrjänen, 1999; Southern & Herrington, 2000). It was therefore postulated in this study that the expression of the shared HPVE6 DEGs may promote cellular transformation through the deregulation of the cell cycle process.

The Molecular mechanisms of cancer pathway was also one of the interesting canonical pathways deregulated by the shared HPVE6 DEGs although its expression was much lower than when it was induced by the HPV18E6 DEGs. The up-regulated genes as a result of the shared HPVE6 DEGs expression included *TGFβ, E2F, BAD* and *COX2* that are involved in the deregulation of the cell proliferation and sustained angiogenesis biological processes (Li & Neaves, 2006; Greenhough et al., 2009; Santarpia et al., 2010; Koontongkaew, 2013). The results from this study clearly show that the expression of HPV11E6 had a much weaker effect towards dysregulating cancer signalling pathways as opposed to HPV18E6. This being consistent with the ability of HPV18E6 to bind to more cellular proteins such as the *E6AP, E6BP, paxillin, p300/CBP* and *PDZ* domain proteins such as *hDLG, MUPP1, MAGI1* and *hScrib,* compared to HPV11E6 that was shown to be able to bind to fewer cellular protein such as *zyxin, GPS2, Bak,* and *MCM7* (Chen et al., 1995; Kiyono et al., 1997; Kuhne & Banks, 1998; Kukimoto et al., 1998; Gardiol et al., 1999; Patel et al., 1999; Thomas & Banks, 1999; Zimmermann et al., 1999; Glaunsinger et al., 2000; Lee et al., 2000; Pim et al., 2000; Degenhardt & Silverstein, 2001a; Degenhardt & Silverstein, 2001b; Park et al., 2001; Gardiol et al., 2002; Oh et al., 2004). It can thus be hypothesised that the expression of HPVE6 predisposes normal cells toward malignancy.

In order to show how the significantly differentiated genes worked together at the molecular level following HPVE6 expression, gene networks were generated using the IPA gene network system on the assumption that the highly interconnected networks were likely to represent significant biological function. The network analysis assembled the top three

networks following HPVE6 expression. The networks with the highest score in DEGs by HPV11E6 included gene products associated with *ERK/MAPK* and the *NFκB* (*complex*), whereas the networks with the highest score in DEGs by HPV18E6 included gene products associated with *TNF* and the *NFκB (complex)* while with the shared DEGs dataset the highest scoring network included gene products associated with the *NFκB* (*complex*). The top functions of these networks included cellular movement, cellular development, cellular growth, cell proliferation, cancer, inflammatory response and the cell cycle. These results thus, validated the pathway analysis results that showed the deregulation of the inflammatory and cell cycle pathways through HPVE6 expression. It is possible that the expression of HPVE6 can target different genes to achieve the same goal of lowering the capability of the host cell to repair its DNA leading to cellular transformation.

The over expression and activation of cellular oncogenes by the *ERK/MAPK* cascade have been reported in many human carcinomas (Loda et al., 1996; Mishima et al., 1998; Lessard et al., 2001). It was also suggested that *ERK1* expression is an early marker of cervical carcinogenesis as well as other human epithelial neoplasias (Loda et al., 1996; Branca et al., 2004). *TNF* is a multifunctional proinflammatory cytokine involved in the regulation of a wide spectrum of biological processes including cell proliferation and differentiation and has also been implicated in cancer and autoimmune diseases (Locksley et al., 2001). The expression of *TNF* by infiltrating mononuclear cells correlates with spontaneous regression of papillomas (Hagari et al., 1995).

From the top networks in our study *NFκB complex* was associated with the most gene products following HPVE6 expression and included mostly genes that encoded growth factors and cytokines. This was confirmed in other studies that showed *NFκB* signaling as a key event in virus induced carcinogenesis (Hussain et al., 2011). *NFκB* functions as an oncogene through its ability to stimulate cell proliferation and survival and is constitutively activated in several human cancers (Li et al., 2005; Karin, 2006), including cancer of the cervix (Nair et al., 2003; Prusty et al., 2005; Branca et al., 2006). Other studies have also shown that *NFκB* activation is a frequent occurrence in squamous cell carcinomas and there is strong evidence that it is important for transformation of epithelial cells (Huber et al., 2004).

From the network of genes regulating the *NFκB* pathway in the current study, the expression of HPV18E6 had more genes being up-regulated while the expression of HPV11E6 had more

genes being down-regulated, though they both shared genes that had more up- than down-regulated genes. These results highlighted the differences between the expression of HPVE6 and indicated that *NFkB* activity is increased immediately upon infection by HPVE6, serving as a mechanism to alert the immune system to the presence of an infection. It is tempting to speculate that the expression of both the low risk HPV11E6 and the high risk HPV18E6 exploited the *NFκB* signalling pathway in achieving cellular transformation, although to a lesser extent with the low risk HPV11E6. A more indepth analysis of structural variants and other viruses apart from HPV that maybe involved in the progression of OSCC is presented in chapter 4.

CHAPTER FOUR
Genomic Alterations in OSCC

4.1 Introduction

Cancer involves many genome alterations such as DNA sequence changes, copy number aberrations, chromosomal rearrangements and modification in DNA methylation that together drive the development and progression of malignancies (McLendon et al., 2008). Single nucleotide polymorphisms (SNPs) were initially believed to be the main contributing factor to genetic variation (International HapMap Consortium, 2005). However, there has been growing appreciation for the effect of larger-scale rearrangements, such as structural variants (SVs) (Sharp et al., 2006; Scherer et al., 2007; Xi et al., 2010).

Structural variants (SVs) include large deletions, insertions, inversions, duplications and translocations that can lead to abnormal cell growth and proliferation. These are important hallmarks of cancer responsible for the creation of regulatory changes leading to activation or over-expression of oncogenes and inactivation of tumour suppressor genes (Albertson et al., 2003; Mijušković et al., 2012). With the sequencing of human genomes being routine (1000 Genomes Project Consortium, 2010) this has widened the spectrum of structural variants (SVs) and copy number variants (CNVs) from originally being defined as greater than 1 kb in size (Feuk et al., 2006) to also include much smaller events of >50 bp in length. The only challenge being to discover the full extent of structural variation in order to genotype it routinely so as to understand its effects on human disease, complex traits and evolution (Alkan et al., 2011). Rapidly accumulating evidence indicates that structural variants can comprise millions of nucleotides of heterogeneity within every genome and are likely to make an important contribution to human diversity and disease susceptibility (Feuk et al., 2006).

Viruses have also been instrumental in the evolution of cancer (Butel, 2000). The link between human cancers with viral origin was made nearly half a century ago with the discovery of the association between Epstein-Barr virus (EBV) and Burkitts Lymphoma (Javier & Butel, 2008). Since then it has become apparent that a number of viruses e.g.

Merkel cell polyomavirus (MCPyV) (Nishizawa et al., 1997; Gifford & Tristem, 2003; Burmeister et al., 2004; Holland & Pogo, 2004; Kremsdorf et al., 2006; Park et al., 2006; Lee & Lee, 2007; Houben et al., 2010; Kofman et al., 2011; Banks et al., 2012; Zwolińska et al., 2013; Hampras et al., 2014; Miner et al., 2014) play significant roles in the multistage development of human cancers.

Six human viruses have been considered by the International Agency for Research on Cancer (IARC) as being carcinogenic based on sufficient evidence supporting their aetiologic association with human cancers: they are Epstein-Barr virus, Hepatitis B Virus(HBV), several types of human papilloma viruses (HPV), human T-cell lymphotropic virus type 1, Hepatitis C Virus (HCV), and Kaposi's sarcoma-associated herpesvirus (Parkin, 2006). However, even though these infectious agents are widespread in the general population, only a very small fraction of infected people develop neoplastic disease, indicating the crucial role of some presently unknown, genetic factors (Mazzaro et al., 2005). Other infectious agents other than viruses e.g. *Helicobacter pylori* and liver flukes have also been known to be carcinogenic (Sripa et al., 2007; Selgrad et al., 2008).

Approximately 15-20% of cancers are associated with viral infections (zur Hausen, 2001; Parkin, 2006). Oncogenic viruses can contribute to different steps of the carcinogenic process, and the association of a virus with a given cancer can range anywhere from 15 to 100% (Parkin, 2006). Furthermore, many other viral agents have been classified as possibly carcinogenic to humans and others have been occasionally found in human tumours suggesting that this figure may be an underestimation of virus involvement in the aetiology of human cancer. Thus, defining the architecture of a specific cancer genome is therefore essential not only as a first step towards understanding the biology of the tumour and mechanisms of oncogenesis, but also clinically towards designing effective personalised therapies (Druker et al., 2001a; Druker et al., 2001b).

With the completed sequence of the human genome and continuing improvement of high-throughput genomic technologies, it is now feasible to contemplate comprehensive surveys of human cancer genomes (McLendon et al., 2008). Recent advances in high throughput sequencing technology have made it possible to study whole genomes at unprecedented high resolution and relatively low cost (Bentley et al., 2008; Mardis, 2008b). Next-generation sequencing (NGS) technology has made whole genome sequencing (WGS) possible at an individual level (Mardis, 2008b; Rusk & Kiermer, 2008; Metzker, 2010). WGS has revealed

numerous single nucleotide variations (SNVs), de novo mutations and somatic mutations in cancer genomes that had not been previously reported (Beroukhim et al., 2010; Lee et al., 2010; Pleasance et al., 2010; Roach et al., 2010; Fujimoto et al., 2012; Shigemizu et al., 2013).

Some of the applications of NGS include analysis of gene expression and alternative splicing (RNA-seq), DNA–protein interactions, de novo genome sequencing, metagenomics and identification of genetic variants within and across populations (Alkan et al., 2009; Schmidt et al., 2009; Wang et al., 2009; Li et al., 2010). Many computational methods are also available for analysing genetic variants using NGS data. These variants include single-nucleotide polymorphisms (SNPs) and structural variants such as copy numbers, insertions, deletions, tandem duplications, inversions and translocations (Ruffalo et al., 2012). Characterization of such variants is useful in many applications, including genome-wide association studies, identification of driver mutations in cancer and comparative genomics (Hudson, 2008; Mardis, 2008a; Meyerson et al., 2010).

This chapter aims to identify structural variants, primarily large insertions and deletions and viral sequence integration that may be associated with the aetiology of oesophageal squamous cell carcinoma (OSCC) in the South African population.

4.2 Results

4.2.1 WGS Quality Control

This was a pilot study aimed at identifying structural variants and viral sequences associated with OSCC. The whole genome of four patients with oesophageal squamous cell carcinoma (OSCC) was sequenced and compared to the human reference genome (NCBI37), to do a preliminary screen for insertions and deletions. The samples were taken from the black and mixed ancestry population groups since these are the populations most affected by OSCC in South Africa. Quality control (QC) measures were carried out to validate the integrity of the samples for whole genome sequencing (WGS) using the NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA). All samples passed the QC for DNA concentration and quality and conformed to the requirements for the downstream application of WGS as shown in Table. 4.1.

Table 4.1: QC for the four patient samples submitted for WGS

Sample ID	Sample Description	Concentration (ng/µl)	A260/280
B381	Black population	42.23	1.92
B450	Black population	47.26	1.85
M456	Mixed ancestry	76.46	1.89
M478	Mixed ancestry	89.09	1.91

DNA concentration and the optical density 260/280 of the samples was quantified using the NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA), values of 1.8–2.0 are considered indicative of relatively good quality DNA as described in section 6.23.3.

4.2.2 Structural Variants in OSCC

Paired-end sequencing was performed on the Illumina HiSeq2000 (Wang et al., 2010). This generated 4.27 billion reads of 100 bp covering at least 95% of the human reference genome. Quality control indicators such as the duplication rate, coverage and mean depth of coverage indicated that the experiments were successful in acquiring high-quality sequence reads (Table 4.2). The minimum depth of coverage was 33.7, and the library preparation was successful across all samples without being compromised by PCR amplification bias (Aird et al., 2011) after having obtained a 98% unique read proportion (Non-N reference) to the human reference genome (NCBI37).

Table 4.2: Sequence coverage and quality indicators for whole genome sequencing of four tumour samples

Sequence coverage/quality measures	B381	B450	M456	M478
Reference genome length	3,095,693,981	3,095,693,981	3,095,693,981	3,095,693,981
Proportion duplicate reads	0.02	0.02	0.02	0.02
Proportion unique reads	0.98	0.98	0.98	0.98
Mean depth of coverage	40.7	35.0	33.7	34.2
Median fragment length (bp)	320	301	311	314
Fragment length standard deviation (bp)	63	62	62	65

All the four samples were aligned onto the human reference genome (NCBI37) using ELAND and CASAVA software. Duplication rate of the reads was only 2% while 98% were unique reads allowing for a greater confidence in variant calling. Following paired end sequencing, all the sample achieved a median targeted fragment length of > 300bp and fragment length standard deviation around the median of >60bp.

The samples were genotyped on an Illumina HumanOmni2.5-8 array (Soorya et al., 2013), and had a high quality score with a percentage base calling of > Q30 indicating a smaller probability of error. Fig. 4.1A, shows that sample B381 had 89.0% of filtered and aligned reads with a mean depth of coverage of 40.7. While Fig 4.1B, shows that sample B450 had 90.7% of filtered and aligned reads with a mean depth of coverage of 35.0. In Fig. 4.2A, sample M456 had 89.3% of filtered and aligned reads with a mean depth of coverage of 33.6.

While in Fig 4.2B sample M478 had 88.3% of filtered and aligned reads with a mean depth of coverage of 34.2. Additionally, at least 96% of genes (including introns and untranslated regions) were covered at 10X coverage. Thus the sequencing phase of this experiment achieved high quality reads that allowed for greater confidence in the variant calling procedure and results.

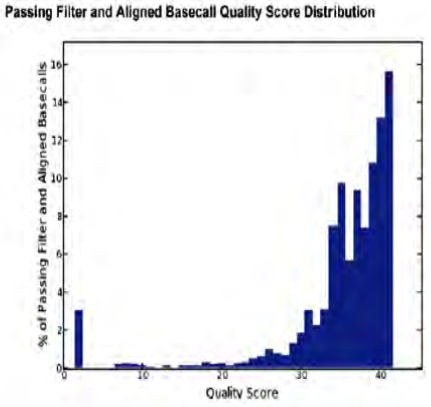
A B381

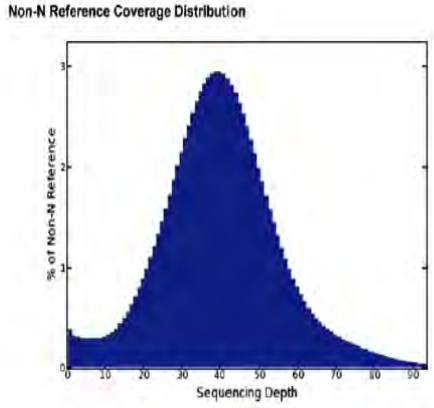
Data Volume

	Yield (Gigabases)	% Bases >=Q30
Passing Filter	137.5	88.3%
Passing Filter and Aligned	122.3	89.0%

Coverage Summary

	Mean Coverage	% Positions >=1X Coverage	% Positions >=10X Coverage
Non-N Reference	40.7	98.0%	95.2%
Genes	41.4	98.7%	96.9%
Exons	40.7	97.9%	96.6%





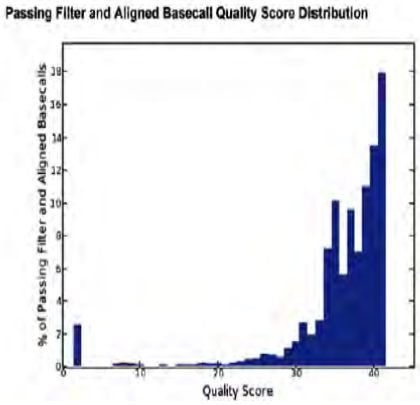
B B450

Data Volume

	Yield (Gigabases)	% Bases >=Q30
Passing Filter	116.3	90.2%
Passing Filter and Aligned	104.0	90.7%

Coverage Summary

	Mean Coverage	% Positions >=1X Coverage	% Positions >=10X Coverage
Non-N Reference	35.0	98.3%	96.0%
Genes	34.9	98.9%	97.1%
Exons	34.1	98.0%	96.8%



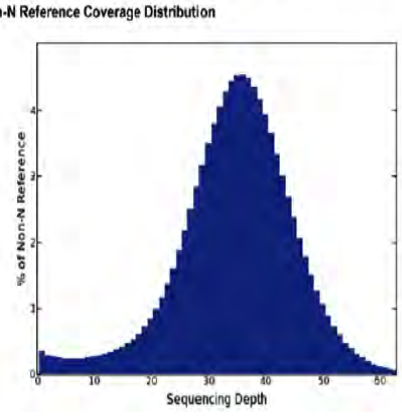


Fig. 4.1: Data volume and coverage summaries. The samples had a high quality score with a percentage base calling of > Q30 indicating a smaller probability of error following alignment to the human reference genome (NCBI37). The passing filters gave the yield in Gigabases of data from the samples input into the NCBI37 reference genome, while the passing filter and aligned reported only the subset of data in Gigabases that aligned to the NCBI37 reference genome. **A:** Sample B381 had 89.0% of filtered and aligned reads with a mean depth of coverage of 40.7. **B:** Sample B450 had 90.7% of filtered and aligned reads with a mean depth of coverage of 35.0

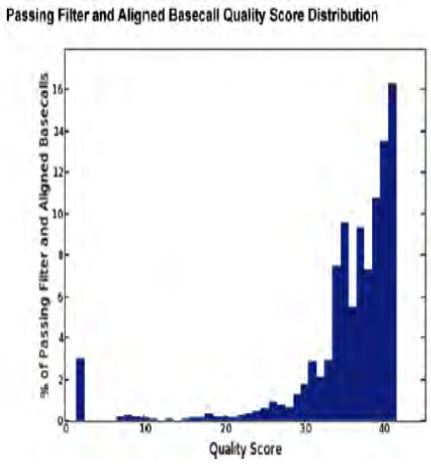
A M456

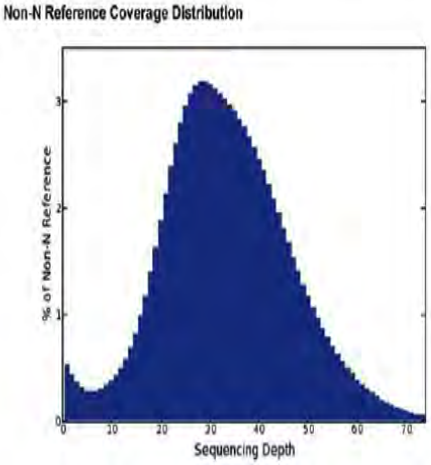
Data Volume

	Yield (Gigabases)	% Bases >=Q30
Passing Filter	113.2	88.5%
Passing Filter and Aligned	100.2	89.3%

Coverage Summary

	Mean Coverage	% Positions >=1X Coverage	% Positions >=10X Coverage
Non-N Reference	33.6	97.8%	94.7%
Genes	34.3	98.6%	96.5%
Exons	33.9	97.8%	96.2%





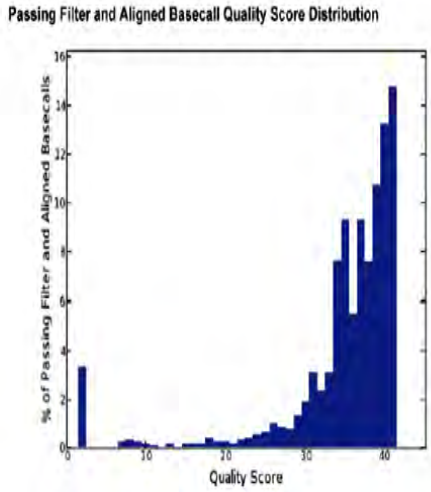
B M478

Data Volume

	Yield (Gigabases)	% Bases >=Q30
Passing Filter	114.5	87.5%
Passing Filter and Aligned	101.4	88.3%

Coverage Summary

	Mean Coverage	% Positions >=1X Coverage	% Positions >=10X Coverage
Non-N Reference	34.2	98.3%	96.9%
Genes	34.4	98.9%	97.1%
Exons	33.8	98.1%	96.8%



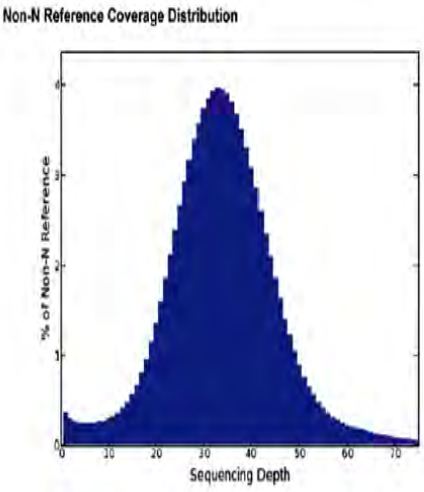


Fig. 4.2: Data volume and coverage summaries. The samples had a high quality score with a percentage base calling of > Q30 indicating a smaller probability of error following alignment to the human reference genome (NCBI37). The passing filters gave the yield in Gigabases of data from the samples input into the NCBI37 reference genome, while the passing filter and aligned reported only the subset of data in Gigabases that aligned to the NCBI37 reference genome. **A:** Sample M456 had 89.3% of filtered and aligned reads with a mean depth of coverage of 33.6. **B:** Sample M478 had 88.3% of filtered and aligned reads with a mean depth of coverage of 34.2.

4.2.2.1 Alignment and Variant Calling

Illumina's proprietary alignment algorithm ELAND and CASAVA software were used to align the reads to the NCBI37 human reference genome (Kidd et al., 2010). Because we did not include matched normal samples, variants with global allele frequency >10% in either exome variant server (EVS) or thousand genomes (TGP) were filtered out to reduce background noise (Bailey et al., 2001; Bailey et al., 2002). Structural variants (SV) reported from the alignment were collated with gene loci (upstream, downstream and overlapping) using the variant effect predictor of Ensembl (McLaren et al., 2010). A large number of SNP effects from all the four samples was observed (Table 4.3), having 98% mean array agreement between the samples and the reference genome NCBI37 of which 38% were associated with Genes.

Table 4.3: Summary of SNP consequences across all four tumour samples

Variant Effects	B381	B450	M456	M478
% Array Agreement	98.47	99.24	97.61	99.22
SNP Total	4,919,446	5,334,173	4,185,808	5,018,129
% in Genes	38.9	38.9	38.8	39.1
% in Exons	1.6	1.6	1.6	1.6
Ti/Tv	2.1	2.1	2.1	2.1
Novel	591,283	617,474	476,394	656,181
Ti/Tv novel	1.5	1.6	1.4	1.7
Synonymous	14,266	15,310	11,750	14,562
Missense	13,349	14,556	11,243	13,660
Nonsense	140	144	118	135
Indel Total	1,308,522	1,358,218	1,153,901	1,303,928
Frameshift indel	417	431	378	446
Splice Region (<=5b From Splice Site)	1,819	1,892	1,450	1,857

Samples were aligned to the NCBI37 Homo sapiens reference genome using ELAND and CASAVA, and the SNP consequences reported from the alignment were collated with gene loci using the variant effect predictor of Ensembl. In all 4 samples there was a 98% mean array agreement between the genotyping and sequencing SNPs to the reference genome NCBI37 of which 38% were associated with Genes. The affected genes were subsequently cross-checked against the genetic association database for disease and cancer associations.

The transition/transversion (Ti/Tv) ratio has been used in multiple studies for assessing the specificity of SNP calls, and has an empirical Ti/Tv ratio of ~2.1 for genome-wide variants (Li & Durbin, 2010; DePristo et al., 2011; Liu et al., 2012; Guo et al., 2013a; Guo et al., 2014). Ti/Tv ratio is computed as the number of transition SNPs divided by the number of transversion SNPs (Guo et al., 2014). Typically, the Ti/Tv ratio is lower in novel SNPs than in known SNPs because of residual false positives and a relative deficit of transitions due to sequencing context bias (DePristo et al., 2011; Zhang et al., 2014).

In all the samples the Ti/Tv ratio was 2.1, further validating the accuracy of SNP calling. The SNP sequence variation in the chromosome showed 90% call accuracy to the reference genome (NCBI37) across all the four samples, which was replicated with the autosomal, mitochondrial and the X-specific sites (Table 4.4).

Table 4.4: Summary of the chromosomal origin of SNP calls across all four tumour samples

Sequence variation	**B381**	**B450**	**M456**	**M478**
Total sites called (out of 3,095,693,981)	2,803,227,246	2,786,676,594	2,797,772,217	2,787,310,210
Autosomal sites				
Total sites called (out of 2,881,033,286)	2,641,282,521	2,639,594,675	2,638,753,835	2,640,156,272
Proportion called	0.91	0.91	0.91	0.91
Called heterozygous	4,071,708	4,374,861	2,991,031	4,136,946
Proportion heterozygous	0.0015	0.0016	0.0011	0.0015
X-specific sites				
Total sites called (out of 152,301,523)	143,955,244	144,539,991	143,955,546	144,602,308
Proportion called	0.94	0.94	0.94	0.94
Called heterozygous	17,022	163,845	15,722	176,523
Proportion heterozygous	0.00011	0.0011	0.00011	0.0012
Y-specific sites				
Total sites called (out of 56,404,529)	15,449,993	0	12,538,545	0
Proportion called	0.27	0	0.22	0
Called heterozygous	11,860	0	9,117	0
Proportion heterozygous	0.00076	0	0.00072	0
Mitochondrial sites				
Total sites called (out of 16,569)	16,566	16,567	16,565	16,565
Proportion called	0.99	0.99	0.99	0.99
Called heterozygous	1	1	0	0
Proportion heterozygous	0.00006	0.00006	0	0

Samples were aligned to the NCBI37 Homo sapiens reference genome using ELAND and CASAVA, and the chromosomal origin of SNP reported using the variant effect predictor of Ensembl. The samples showed an excellent agreement with the sequence-based SNP calls at (>90%) in all the sites generating high-quality and reliable sequencing data as described in section 6.23.3.

The most prevalent structural variants were insertions, deletions and copy number variation (Table 4.5). Very few insertions were observed in all the four samples compared to deletions and copy number variations. This study concentrated on identifying large structural variants (>300bp) and the role that they could play in OSCC.

Table 4.5: Summary of structural variants detected in all four tumour samples

Structural Variants	B381	B450	M456	M478
Insertions	8	3	4	3
Deletions	2,876	2,558	2,478	2,556
Copy number variations	1,748	992	1,465	1,165

Only structural variants which were >300bp following alignment to the human reference genome (NCBI37) using ELAND and CASAVA were collated and called with gene loci, (upstream, downstream and overlapping) using the Variant Effect Predictor of Ensembl as described in section 6.23.3.

4.2.2.2 Effects associated with insertions

No common insertions were identified in all four samples. Five insertions were associated with genes (Table 4.6), but they lacked adequate evidence to justify an association with OSCC. The evaluation of insertions was not explored further since these insertions were identified as intronic variants that were associated with diseases such as sinusitis, Diabetes Mellitus, Type 1, congenital diaphragmatic hernia, Alpers syndrome and Coronary Artery disease while none was associated with the progression of OSCC.

Table 4.6: Description of Genes associated with insertions common to all four tumours, showing five genes that were not associated with OSCC

Genes (HGNC)	Location	Variant Consequence	Disease Associated
IQGAP2	Chr.5	intron_variant	Intellectual disability and sinusitis
LRRC7	Chr.1	intron_variant	Diabetes Mellitus, Type 1
CCDC102B	Chr.18	intron_variant	Congenital diaphragmatic hernia
FARS2	Chr.6	intron_variant	Alpers syndrome
ZNF385D	Chr.3	intron_variant	Coronary Artery Disease

Insertions associated with genes were cross-checked against the genetic association database for disease and cancer associations as described in section 6.23.3.

4.2.2.3 Effects associated with deletions

A total of 697 deletions were observed in all the four samples of which 349 were associated with genes and 127 were associated with different types of cancer (Table 4.7). Only twelve genes were shown to have deletions in the coding region or were associated with oesophageal cancer (Table 4.8). Among the twelve genes; *ADH4*, *EGFR, MSH3* and *BCLAF1* have previously been associated with cancer (Kuwano et al., 2005; Ungerback et al., 2012; Kato et al., 2013; Lawrence et al., 2014; Vogelsang et al., 2014), but only the *BCLAF1* deletion was identified as a possible interesting structural variant in OSCC.

The *BCLAF1* deletion had previously been shown to be associated with Non-Hodgkin's lymphoma (Kasof et al., 1999) and recently with colorectal cancer (Lawrence et al., 2014). In this study the *BCLAF1* deletion spanned 53.20% of the protein coding region as opposed to the other structural variants and was thought to have significant implications for the progression of OSCC.

Table 4.7: Summary of deletions common to all four tumour samples, showing the breakdown of the total number of deletions described in the samples

Deletions	**Common in all four samples**
Number of common deletions	697
Deletions associated with genes	349
Deletions associated with cancer	127

Deletions associated with genes were cross-checked against the genetic association database for disease and cancer associations as described in section 6.23.3.

Table 4.8: Description of Gene deletions common to all four tumour samples

Genes (HGNC)	**Location**	**Variant Consequence**	**MeSH Disease Terms**	**Percentage Protein Coding overlap**
NXPE1	Chr.11	non-coding exon variant	Ulcerative colitis and Inflammatory bowel disease	0.66
DLEU1	Chr.13	non-coding exon variant	Chronic lymphocytic leukemia	0.32
TMBIM4	Chr.12	3'UTR variant	Venezuelan hemorrhagic fever and Bolivian hemorrhagic fever	0.73
ADH4	Chr.4	upstream_gene_variant	Adenocarcinoma, Oesophageal neoplasm and Parkinson Disease	0.12
HYDIN	Chr.16	coding sequence variant	Primary ciliary dyskinesia 5 and Hydin-related primary ciliary dyskinesia	0.49
KCNJ12	Chr.17	Frameshift variant 3'UTR	Hypokalemic periodic paralysis	3.53
BCLAF1	Chr.6	coding sequence variant	Non-Hodgkin's Lymphoma Colorectal cancer	53.20
CACNA1B	Chr.9	coding sequence variant	Pulmonary aspergilloma	0.34
EGFR	Chr.7	intron_variant	Oesophageal neoplasms, colonic neoplasms and cervical squamous cell carcinoma.	0.24
FAM118A	Chr.22	coding sequence variant	Cardiovascular Diseases and Diabetes Mellitus, Type 2	1.70
OR6K5P	Chr.1	Noncoding exon variant	Neuronitis	25.35
MSH3	Chr.5	intron_variant	Adenocarcinoma, Oesophageal neoplasm and Colonic Neoplasms	0.31

Percentage overlap of the protein coding region was reported from the alignment to the human reference genome (NCBI37), and collated with gene loci using the Variant Effect Predictor of Ensembl as described in section 6.23.3.

4.2.2.4 BCLAF1 Validation

In order to ascertain the role of *BCLAF1* deletion on the progression of OSCC we checked the frequency of *BCLAF1* deletion by PCR of thirty DNA tissue biopsies from OSCC patients and their normal blood DNA samples (Fig. 4.3). The PCR reaction was optimized and gave rise to PCR products of 2045 bp and 690 bp for non-deletion and deletion alleles respectively, while the presence of both fragments defined the heterozygous deletion (Figure 4.4). The frequency of the *BCLAF1* deletion in OSCC samples was 40% with 33% (10/30) being homozygous deletions and 7% (2/30) heterozygous deletions, while 60% had no deletions. There were no deletions observed in the respective patient's blood samples used as control (Figure 4.5).

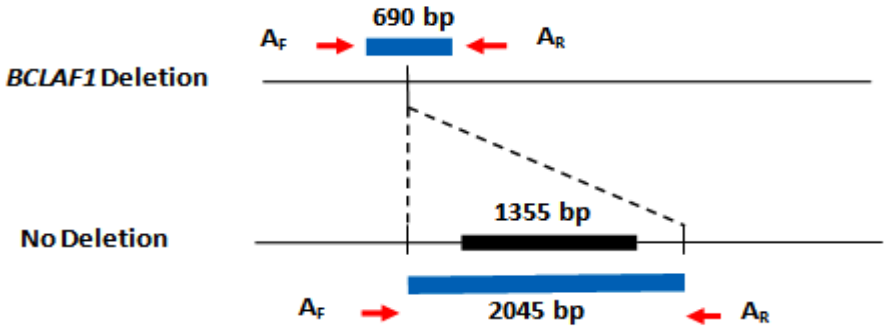


Figure 4.3 PCR assay for *BCLAF1* deletion. A two primer set was used to show the presence of *BCLAF1* deletion. Primers AF and AR amplify two distinct fragments: a 2045 bp fragment detecting the presence of the *BCLAF1* gene and a 690 bp fragment generated by the deletion of *BCLAF1* gene. According to the presence or absence of these fragments, three results could be inferred. The sole presence of the 2045 bp or 690 bp fragment showed no deletion or homozygous deletion, while the contemporary presence of both fragments showed heterozygous deletion. Solid lines represent genomic sequences; the black bar represents the size of *BCLAF1* deletion (1355bp) while the red arrows indicate forward (F) and reverse (R) primers. Expected PCR products are drawn as small blue bars. Results from the two PCR assays for *BCLAF1* deletion are shown in Fig. 4.4 & 4.5.

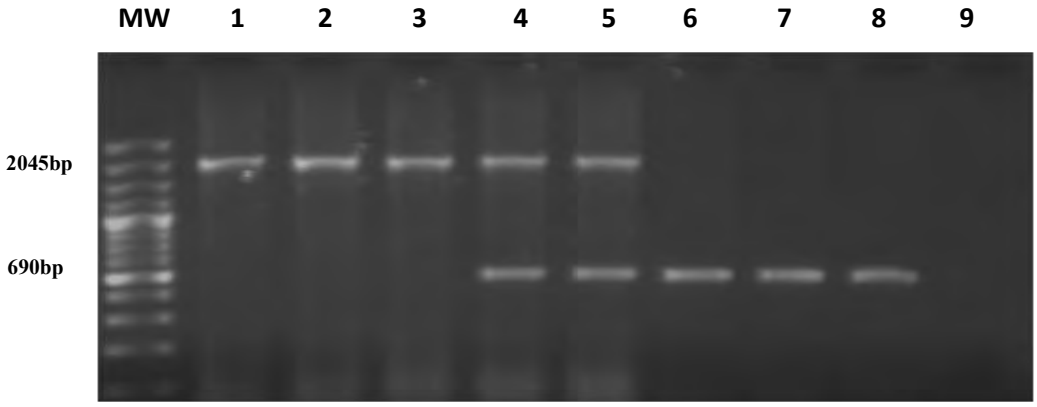


Figure 4.4 Deletion of the *BCLAF1* gene. PCR amplification of tumour biopsy DNA samples using the primers described in appendix A resulted in three possible outcomes: the sole presence of either a 2045 bp band (lanes 1–3) or a 690 bp band (lanes 6-8) defined no deletion and homozygous deletion, respectively, while the presence of both fragments defined the heterozygous deletion (lanes 4&5) with lane 9 being the water blank. MW: Molecular weight marker (Fermentas).

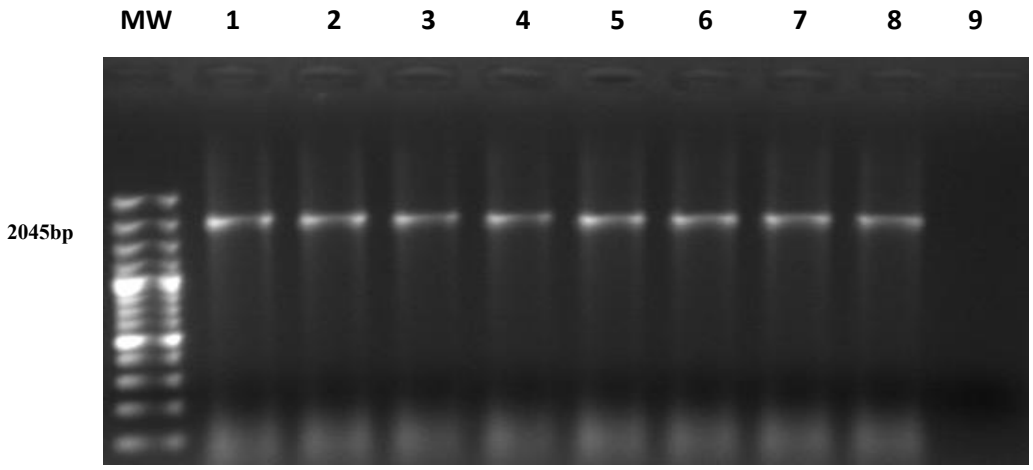


Figure 4.5 No Deletion of the *BCLAF1* gene. PCR amplification of blood DNA samples using the primers described in appendix A, resulted in the sole presence of a 2045 bp band (lanes 1–8) defining the absence of deletion with lane 9 being the water blank. MW: Molecular weight marker (Fermentas).

4.2.3 Viral Integration in OSCC

Having previously described on the role of HPVs in causing cellular transformation (chapter three), we also searched for the presence of other viruses that may be implicated in the aetiology of OSCC. Reads from the WGS that didn't map to the human reference genome were probed for viral association with OSCC. The program DeconSeq (Schmieder & Edwards, 2011) was used to align the unmapped reads against complete set of NCBI RefSeq Viral Genomes (build 64) using the Burrows Wheeler Aligner (BWA) (Li & Durbin, 2010). DeconSeq allows detecting of plausible matches based on a user specified match identity and sequence coverage. For our detection we set the identity match to a minimum of 95% and sequence coverage to a minimum of 90%. Based on the detection criteria Table 4.9 shows the number of reads per sample that has viral origin. From the DeconSeq results a list of the most to least frequent viral hits was built. The BWA database indexing, alignment and filtering were done on the UCT HPC cluster (http://hex.uct.ac.za).

Table 4.9: Total number of unmapped reads for each sample that has viral origin. Based on a minimum of 95% identity match and a minimum of 90% sequence coverage on BWA alignments against NCBI's RefSeq Viral Genomes (build 64).

Sample	Viral Reads (100bp)
B381	1236585
B450	912766
M456	961224
M478	1079614

4.2.4 Detection of viral integration in OSCC

From the list of viral sequences identified in all the four samples, 22 viral sequences were selected which were common and of human origin in all the four samples (Table 4.10). In order to detect the viral integration sites we used BWA to align all the reads from the WGS to the human reference genome (hg19, UCSC) and the 22 viral genomes.

Table 4.10: Selected human viral sequences used in building a BWA database to detect human viral integration

NCBI_RefSeq	Virus Name
GI:9626597/NC_001457.1	Human papillomavirus type 4, complete genome
GI:962739/NC_001596.1	Human papillomavirus type 9, complete genome
GI:9626069/NC_001357.1	Human papillomavirus 18, complete genome
GI:9628574/NC_001694.1	Human papillomavirus 61, complete genome
GI:310698439/NC_001526.2	Human papillomavirus type 16
GI:9629378/NC_001806.1	Human herpesvirus type 1
GI:9629267/NC_001798.1	Human herpesvirus 2
GI:9625875/NC_001348.1	Human herpesvirus 3
GI:139424470/NC_009334.1	Human herpesvirus 4
GI:82503188/NC_007605.1	Human herpesvirus 4 type 1, complete genome
GI:224020395/NC_001664.2	Human herpesvirus 6A
GI:51874225/NC_001716.2	Human herpesvirus 7
GI:139472801/NC_009333.1	Human herpesvirus 8
GI:157781212/NC_009823.1	Hepatitis C virus genotype 2
GI:157781216/NC_009824.1	Hepatitis C virus genotype 3
GI:157781214/NC_009827.1	Hepatitis C virus genotype 6
GI:155573622/NC_006273.2	Human herpesvirus 5 strain Merlin
GI:56160839/AC_000017.1	Human adenovirus type 1
GI:56160492/AC_000007.1	Human adenovirus 2
GI:12175745/NC_002645.1	Human coronavirus 2
GI:9630376/NC_001897.1	Human parechovirus 1
GI:548558394/NC_022518.1	Human endogenous retrovirus K113 complete genome

All the discordant reads were extracted from the complete mapping and only those matches where one of the read pairs mapped to the human genome and the other read pair mapped to any of the viral genome were selected (Jiang et al., 2012). In order to achieve a high confidence in viral integration, we further filtered the selected viral reads to only those that had 100% match in both the BWA alignment score and CIGAR string to the reference human genome and human viral sequence. Following this further filtering only 2 viral sequences that had varying integration sites between the four samples were observed; sample B381 had one integration site observed for Human Endogenous Retrovirus K113 (HERV-K113) and another for two Human herpesvirus 6A (HHV-6A); sample B450 had one integration site

observed for HERV-K113 and two integration sites observed for HHV-6A; sample M456 had four integration sites observed for HERV-K113 and seven integration sites observed for eight HHV-6A; sample M478 had seven integration sites observed for nine HERV-K113 and one integration site observed for HHV-6A (Table 4.11). A Circster figure showing the chromosomal integration sites of the filtered viral sequences for the four patient samples was generated on Galaxy (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010) as shown on Fig 4.6.

Table 4.11: Chromosomal integration sites of filtered viral sequences identified in all four samples

Samples	Viral Sequences	Integration Sites
B381	Human endogenous retrovirus K113 complete genome	4p16.1
	Human herpesvirus 6A	13q12.12*
	Human herpesvirus 6A	13q12.12*
B450	Human herpesvirus 6A	7p22.1
	Human herpesvirus 6A	11q22.1
	Human endogenous retrovirus K113 complete genome	19p13.11
M456	Human herpesvirus 6A	1q24.1
	Human herpesvirus 6A	1q32.1
	Human herpesvirus 6A	2p16.3
	Human herpesvirus 6A	2q31.1
	Human endogenous retrovirus K113 complete genome	3q28
	Human herpesvirus 6A	3q28
	Human endogenous retrovirus K113 complete genome	8p23.1
	Human herpesvirus 6A	9q22.31
	Human herpesvirus 6A	9q34.3
	Human endogenous retrovirus K113 complete genome	19p13.11
	Human endogenous retrovirus K113 complete genome	20p12.1
	Human herpesvirus 6A	21p11.1
M478	Human herpesvirus 6A	1p31.1
	Human endogenous retrovirus K113 complete genome	3q28*
	Human endogenous retrovirus K113 complete genome	3q28*
	Human endogenous retrovirus K113 complete genome	4p16.1
	Human endogenous retrovirus K113 complete genome	6q14.2
	Human endogenous retrovirus K113 complete genome	8p23.3
	Human endogenous retrovirus K113 complete genome	15q21.3*
	Human endogenous retrovirus K113 complete genome	15q21.3*
	Human endogenous retrovirus K113 complete genome	19p13.11
	Human endogenous retrovirus K113 complete genome	Xq13.2

*Multiple viruses integrated at the same site

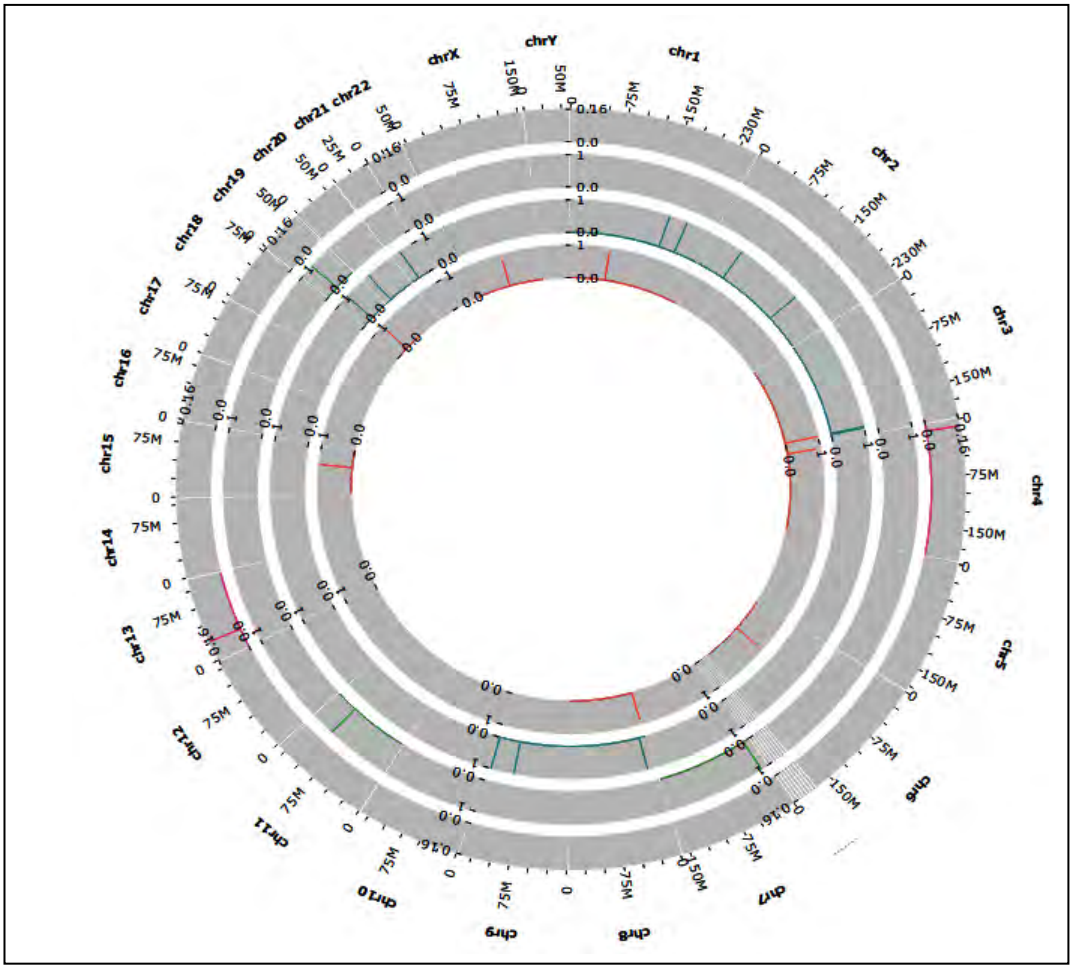


Fig.4.6: Circster plot showing chromosomal integration sites of the filtered viral sequences from the four patient samples**.** *Red spike* **(B381):** Shows the presence of 3 viral integration sites, *Green spike* **(B450):** Shows the presence of 3 viral integration sites, *Blue spike* **(M456):** Shows the presence of 12 viral integration sites and *Orange spike* **(M478):** Shows the presence of 10 viral integration sites

4.3 Discussion

Cancer represents a special opportunity not only to study disease genesis and progression but also to study mechanisms of cellular regulation, as specific genomic alterations can be identified in cancer compared to normal cells (Strausberg & Simpson, 2009). Many genetic alterations that disrupt the normal balance during cell proliferation, survival and differentiation are associated with chromosomal abnormalities (Mitelman et al., 2007).

In this pilot study, four DNA samples from patients with OSCC were sequenced primarily to identify large structural variants (insertions and deletions) and viral sequences that may be associated with the progression of OSCC in the South African population. A total of 4.27 billion reads of 100 bp were generated covering 95% of the human reference genome

(NCBI37) at a mean read depth of 33.7. This allowed for reliable calling of SNPs and structural variants across the human reference genome (NCBI37), and corresponded with previous studies that also showed that an average read depth that exceeded 30× was the de facto standard for achieving reliable variant calling across 95% of the reference genome (Bentley et al., 2008; Wang et al., 2008; Ahn et al., 2009; Yoon et al., 2009; Sims et al., 2014).

The transition/transversion ratio is also a critical metric for assessing the specificity of SNP calling (DePristo et al., 2011) and has been used as a quality control parameter for assessing the overall SNP quality by multiple studies with the expected Ti/Tv ratios in whole-genome sequencing being 2.1 thus, the higher Ti/Tv ratio generally indicated higher accuracy in variant calling (DePristo et al., 2011; Liu et al., 2012; Guo et al., 2013a; Guo et al., 2013b). This corresponded with our results having a Ti/Tv ratio of 2.1 across all four samples, further confirming the reliability of our SNPs and structural variant calling.

Structural variants (SVs) in the human genome play an important role in the genetics of complex diseases (Iafrate et al., 2004; Sebat et al., 2004; Feuk et al., 2006). Although our analysis reduced the number of variants of specific interest from the pool of structural variants in this study to a select few, large numbers of potentially functional alterations that could contribute to OSCC susceptibility were not analysed further for functional consideration at this time with the potential role of CNV and insertional changes not being considered. From the twelve potential genes common in all four samples only three genes have been previously shown to be associated with OSCC; *ADH4*, *EGFR*, *MSH3* and one novel gene *BCLAF1* that has not been reported in OSCC.

A reduction in *ADH4* expression in poorly differentiated OSCC was shown to contribute to the maintenance of the tumour by inhibiting the retinoic acid signaling pathway (Han et al., 1998; Chou et al., 2002; Yin et al., 2003; Matsumoto et al., 2005; Lee et al., 2006), while mutations in *EGFR* have also been shown to occur in primary OSCC (Guo et al., 2006; Hanawa et al., 2006; Mir et al., 2008). Other studies by (Vogelsang et al., 2012; Vogelsang et al., 2014) also showed that *MSH3* polymorphism, together with exposure to tobacco smoke increased the risk of oesophageal carcinogenesis. Although these genes (*ADH4*, *EGFR* and *MSH3*) had previously been associated to OSCC, we did not analyze them further for functional significance in this study due to the low protein coding overlapping regions.

The *BCLAF1* deletion on the other hand, is an interesting structural variant (SV) candidate associated with OSCC, due to its identification as a coding variant with an overlap of 53.20% in the protein coding region. Although the exact molecular role of *BCLAF1* remains to be defined, *BCLAF1* has been shown to have functional connections and mutation patterns consistent with the known hallmarks of cancer (Hanahan & Weinberg, 2011; Imielinski et al., 2012) including cell proliferation, apoptosis, genome stability, chromatin regulation, immune evasion, RNA processing and protein homeostasis (Lawrence et al., 2014) which suggested *BCLAF1* contribution to the progression of OSCC.

BCLAF1 has also been shown to directly interact with and activate *TP53* gene transcription and the silencing of *BCLAF1* gene reduces *TP53* dependent apoptosis thus, providing evidence that *TP53* is a downstream target of *BCLAF1* (Liu et al., 2007). This is also supported by other recent studies that identified *TP53* as among the significantly mutated genes in OSCC (Lin et al., 2014; Song et al., 2014). These studies were thus confirming the DNA microarray gene expression results in chapter two of this thesis which showed the down-regulation of *TP53* gene by the high risk HPV18E6.

In order to ascertain the presence of *BCLAF1* deletion in OSCC, a PCR assay was done on a larger sample size of tumour DNA and their respective normal blood samples. Deletions were specific for the tumour DNA as there were no deletions observed in their respective patient's blood samples used as control. These results demonstrated that *BCLAF1* deletion may be an important determinant for the progression of OSCC. According to our knowledge this is the first study linking *BCLAF1* deletion in OSCC. It is possible that the link between *BCLAF1* deletion and the progression of OSCC is through the down-regulation of *TP53* and the deregulation of cellular processes involved in carcinogenesis. This hypothesis was also supported by studies that showed the role of *BCLAF1* as a tumour suppressor and to promote the stability of genes required for DNA repair and maintenance of genomic stability (Savage et al., 2014).

Viruses have been known to cause cellular transformation by the expression of viral oncogenes, by genomic integration to alter the activity of cellular proto-oncogenes or tumour suppressors and by inducing inflammation that promotes oncogenesis (Tang et al., 2013). Thus, the study of oncogenic viruses has been invaluable to the discovery and analysis of key cellular pathways that are commonly rendered dysfunctional during carcinogenesis (McLaughlin-Drubin & Munger, 2008). This study utilised a discovery approach of analysing

unmapped reads, generated from WGS of the four samples to detect viral integration within the host genome and their association with OSCC (Cimino et al., 2014).

This study showed the Human Herpes virus 6A (HHV6A) and Human endogenous retrovirus K113 (HERV-K113) as the two viral sequences to be integrated into the chromosomes of all the four sequenced human genomes, these findings are not contradictory to others since both have previously been shown to be integrated in the human genome in cases of autoimmune diseases (Moyes et al., 2005; Amirian et al., 2014).

HHV6A has been shown to persist in latent state by naturally integrating into the human genome, where it has been reported to escape immune surveillance by expressing a number of proteins that counteract the host's immune system (Arena et al., 1999; Lusso, 2006; Hall et al., 2008; Nacheva et al., 2008; Flamand et al., 2010). The reactivation of HHV6A has been observed in AIDS and cancer patients (Lusso & Gallo, 1995; Michalek et al., 1999). In this study HHV6A is integrated in several chromosome specific sites with some of these sites being consistent with other studies that also identified 9q34.3 as a HHV6A chromosome specific integration site (Nacheva et al., 2008; Morissette & Flamand, 2010). Previous studies have also reported on the role of HHV6A specifically its ORF-1 gene in its ability to bind and inactivate p53, while its other gene product U95 has been postulated to contribute to cancer by binding and deregulating the nuclear factor-kappa B (*NFkB*) (Doniger et al., 1999; Ablashi et al., 2010; Kofman et al., 2011). We thus hypothesise that the integration of HHV6A could have oncogenic properties that modify the expression of host cell proteins that are involved in apoptotic signaling and cell growth leading to OSCC.

The human genome harbours about 8% of endogenous retroviral sequences known to originate from germ cell infections by exogenous retroviruses during evolution (Zwolińska et al., 2013). While most of those sequences have been shown to be inactive because of accumulation of mutations, HERV-K113 is capable of being transcribed and translated leading to malignant transformation of cells (Hanke et al., 2013; Downey et al., 2014). Previous studies have reported chromosomal integration sites 8p23.1 and 19p13.11for HERV-K113 (Turner et al., 2001; Kurth & Bannert, 2010; Zwolińska et al., 2013). This study also showed several chromosomal integration sites of HERV-K113 including 8p23.1 and 19p13.11. HERV-K113 is known to inhibit tumor suppressor genes and promote cell growth in lymphoma, breast, and skin cancer (Burmeister et al., 2004; Contreras-Galindo et al., 2008; Golan et al., 2008; Wang-Johanning et al., 2008; Kurth & Bannert, 2010).We hypothesise

that the integration of HERV-K113 may be responsible for modifying the expression of tumour suppressor genes involved in cell growth regulation and tumourigenesis leading to oncogenic transformation of the host cells.

These results highlight the role played by *BCLAF1* deletion, HHV6A and HERV-K113 in enabling genomic instability of the host cell and support a model wherein the frequent assault of the human genome by widespread genomic alterations significantly widens oncogenic opportunities in patients with OSCC.

CHAPTER FIVE
FINAL CONCLUSION

5.1 Conclusion

In this study, a normal human keratinocyte (HaCaT) cell line was used as a model to further our understanding on the early stages of HPVE6 expression through global gene expression profiling and also to identify structural variants and viral sequence integration associated with the initiation and progression of OSCC.

The first objective was aimed at constructing a recombinant adenovirus capable of high level transient expression of the biologically active HPV E6 from the low risk HPV11 and high risk HPV18 and to ascertain their roles in inducing cellular transformation. HPVE6 appeared to serve as an initiation event capable of inducing cellular transformation by both HPV11E6 and HPV18E6 although HPV11E6 achieved this at a much lower rate compared to HPV18E6, this may have been attributed to the differential ability of HPVE6 proteins to bind to and degrade p53. Thus, although HPVE6 infection alone may not be sufficient for the process of cellular transformation, expression of the HPVE6 may facilitate cellular transformation by the activation or inactivation of other cellular processes apart from the binding and inhibition of p53.

The second objective was aimed at explaining the biological changes observed in HaCaT cells infected with HPVE6, and to assess the global effects of HPVE6 expression. This enabled us to gain a broader view of the overall effects that the E6 gene has on gene expression and to provide a first step towards elucidating the pathways and networks involved in HPV-induced oesophageal carcinogenesis. Gene ontology analysis revealed that most of the DEGs were part of the metabolic pathways and further inspection of this process showed that these DEGs were also involved in the deregulation of cell shape, cell proliferation, transport and cytoskeletal re-organization all of which are well known features of cancer cells. The up-regulation of metabolic pathway in this study may therefore be an early indication of cellular transformation since tumour cells are known to be metabolically active and thus metabolic pathways could be potential sources of therapeutic targets.

It was evident that HPV18E6 was capable of dysregulating more cancer signalling pathways compared to HPV11E6 thus, highlighting the differences between the expressions of HPVE6. This can be attributed to the ability of HPV18E6 being able to bind to more cellular proteins that are involved in cellular transformation. From the network of genes the expression of both the low risk HPV11E6 and the high risk HPV18E6 exploited the *NFκB* signalling pathway in predisposing normal cells toward malignancy, although to a lesser extent with the low risk HPV11E6.

The third objective was aimed at identifying structural variants, primarily large insertions and deletions and viral sequence integration sites that may be associated with the aetiology of OSCC. From the results the *BCLAF1* deletion and the integration of both HHV6A and HERV-K113 in the tumour tissue were implicated as aetiological agents associated with OSCC pathogenesis. This could be due to their involvement in the down-regulation of *p53* leading to the deregulation of cellular processes involved in cell proliferation and genomic instability, which are known hallmarks of cancer.

Although it is unlikely that the role played by viruses and structural variants alone may be sufficient for the process of malignant transformation, this study contributes towards the understanding of the mechanisms involved in genomic alterations. It also provides for a better understanding of the underlying mechanisms in tumourigenesis which may ultimately lead to the development of more effective therapeutic interventions for OSCC.

5.2 Future Work

Future plans include examining the effect of inhibiting some of the dysregulated signalling pathways in order to understand the roles being played by the respective DEGs by HPVE6. It would also be interesting to investigate the action of the NFκB signaling pathway and to evaluate its potential role as a specific target for therapeutic strategies in OSCC with high NFκB activity.

Sequence large numbers of tumour and normal DNA from patients to get more detailed information on integration sites and use PCR and other techniques to validate data and improve the precision of mapping so as to achieve high quality reads allowing for a greater confidence in variant calling.

Having identified *BCLAF1* deletion as novel structural variants implicated in OSCC it is important to look at the downstream effects of *BCLAF1* deletion following *BCLAF1* knockout studies in cells using siRNA. PCR and Sanger sequencing will also be done to validate the HHV6A and HERV-K113 integrated viral sequences sites in both tumour and normal samples.

CHAPTER SIX
MATERIALS AND METHODS

6.1 Preparation of ultracompetent Escherichia coli DH5α

The method of Inoue et al., (1990) was used for the preparation of ultracompetent *Escherichia col*i DH5α, with a slight modification of the Hanahan, (1983) technique. Glycerol stocks of *E.coli* DH5α were streaked on a Luria Bertani (LB) plate and incubated overnight at 37°C, in a humidified atmosphere of 5% CO_2. Prepared a starter culture by inoculating 25ml of Luria Bertani (LB) broth with a single colony of *E.coli* DH5α in a 250-ml flask the starter culture was then incubated for 6-8 hours at 37ºC with vigorous shaking (250-300 rpm) in a controlled environment incubator shaker (New Brunswick Scientific, Inc. G25; Edison, NJ, USA). 10 ml of the starter culture was then inoculated in 1-liter flask, containing 250 ml of LB and incubated overnight at 18-22ºC with moderate shaking. The OD_{600} was then read on a spectrophotometer and monitored every 45 minutes until the OD_{600} reached 0.55.The culture vessel was then transferred to an ice water bath for 10 minutes and harvested by centrifugation at 2500g for 10 minutes at 4ºC. The medium was decanted and the centrifuge tube placed upside down on a stack of paper towels for 2 minutes after which a pipette was used to remove any drops of remaining medium adhering to walls of the centrifuge bottle or trapped in its neck. The cells were then gently re-suspended by swirling in 20 ml of ice-cold Inoue transformation buffer (section 6.23) with 1.5 ml of DMSO added, the cells were mixed by swirling and stored in ice for 10 minutes. Aliquots of the suspensions were quickly dispensed into chilled, sterile 1.5 ml microfuge tubes which were tightly closed and immediately snap-frozen by being immersed in a bath of liquid nitrogen. The tubes with the competent cells were then stored at -80ºC until needed.

6.2 Large scale production of recombinant pShuttle2 vector

Competent cells (50 µl) were transformed by adding 50 ng of the empty pShuttle2 vector, pShuttle2-11E6 or pShuttle2-18E6 to the cells, mixing gently, and incubating on ice for 30 minutes. The cells were heat shocked for 50 seconds at 42°C, and placed on ice for 5 minutes. 1 ml of LB broth was added to the cells, which were incubated at 37°C for 1 hour

with shaking. Thereafter, 50 µl of the transformed bacteria was plated onto LB agar containing 50 µg/ml kanamycin (Sigma-Aldrich, Steinheim, Germany) and incubated 37°C for 12 to 15 hours. Bacterial colonies containing the transformed plasmid were then picked into 5ml of LB broth containing 50 µg/ml of kanamycin, and incubated for 10 to 12 hours at 37°C with shaking. A 100 µl aliquot of the overnight culture was used to inoculate 100 ml of LB containing 50 µg/ml kanamycin, and incubated overnight at 37°C with shaking. Large-scale preparation of plasmid DNA was performed using the QIAGEN Endofree Plasmid Maxi Kit (QIAGEN, Hilden, Germany). This kit employed a modified alkaline lysis procedure followed by binding of plasmid DNA to QIAGEN Anion-Exchange Resin under appropriate low-salt and pH conditions. RNA, proteins, dyes, and low-molecular weight impurities were removed by a medium-salt wash and the plasmid DNA was eluted in a high salt buffer and then concentrated and desalted by isopropanol precipitation

6.3 Construction of Recombinant Adenoviral DNA

The recombinant pShuttle2 plasmid DNA with inserts of either HPV11E6 or HPV18E6 were excised with unique restriction endonucleases *PI-Sce*I and *I-Ceu*I and then "shuttled" into Adeno-X Viral DNA by means of an invitro ligation (Clontech Expression, 2012). Once the ligation was completed the products were treated with *Swa*I to linearize non-recombinant pAdeno-X DNA.

6.4 Transformation of DH5α E.coli with recombinant Adeno-X^{TM} DNA

DH5α *E.coli* (50 µl) were transformed by adding 50 ng of the ligated product this was mixed gently and incubated on ice for 30 minutes. The cells were heat shocked for 50 seconds at 42°C, and placed on ice for 5 minutes. 1 ml of Luria broth (LB) was added to the cells, which were incubated at 37°C for 1 hour with shaking. Thereafter, 50 µl of the transformed bacteria was plated onto LB agar plate (100 µg/ml ampicillin) and incubated 37°C overnight. The colonies were checked for recombinant pAdeno-X DNA by PCR using the Adeno-X forward and reverse PCR Primers (Appendix A). Bacterial colonies containing the transformed plasmid were then picked into 5ml of LB broth containing 100 µg/ml of ampicillin, and incubated at 37°C overnight with continous shaking. A 100 µl aliquot of the overnight culture was used to inoculate 100 ml of LB containing 100 µg/ml ampicillin, and incubated overnight at 37°C with continous shaking. Large-scale preparation of plasmid DNA was performed using the Endofree Plasmid Maxi Kit (QIAGEN, Hilden, Germany) as described earlier.

6.5 Analyzing Recombinant Adeno-X^{TM} DNA

Following purification the identity and integrity of the respective recombinant Adeno-X plasmids were confirmed through PCR analysis, by screening the pAdeno-X DNA for the presence of pShuttle2-derived expression cassettes using the Adeno-X Forward PCR Primer and Reverse PCR Primer. Since these primers specifically amplified the 287-bp sequence that spanned the *I-Ceu*I ligation site in pAdeno-X, only the recombinant pAdeno-X templates were amplified as the non-recombinants lacked the Shuttle sequence needed for annealing with the reverse primer. Presence of the pAdeno-X DNA was also verified by digestion with *Xho*I.

6.6 Transfecting HEK 293 Cells with PacI-Digested Adeno-X DNA

Before Adeno-X E6 constructs could be packaged, the recombinant plasmids were digested with *Pac*I to expose the inverted terminal repeats (ITRs) located at either end of the genome. The ITRs contained the origins of adenovirus DNA replication that had to be positioned at the termini of the linear Adenovirus DNA molecule to support the formation of the replication complex (Tamanoi & Stillman, 1982). HEK293 cells were plated at 50-70% confluence, approximately 1 x 10^6 cells, in a 100 mm cell culture dish, and incubated overnight at 37°C, in a humidified atmosphere of 5 % CO_2. The cells were then washed twice with 1x sterile PBS and then transfected using Lipofectamine LTX and PLUS reagents (Invitrogen, CA, USA). Plasmid DNA from the large scale preparation; 10 µg from Adeno-11E6 and Adeno -18E6 respectively, was added to 750 µl of MEM, together with 15 µl of PLUS reagent, and incubated for 10 minutes at room temperature. To 720 µl serum free MEM, 30 µl of Lipofectamine LTX was added, and this was added to the DNA mixture, and incubated for 30 minutes at room temperature. The entire transfection mixture was added drop wise to the cells and the plates were incubated at 37°C, in a humidified CO_2 (5%) incubator for 6 hours, after which the medium was removed and replaced with fresh 10ml complete MEM.

6.7 End-point dilution assay

HEK293 cells were plated in four 96-well plates carefully seeded at the same density (~10^4 cells per well) in 100 µl of growth medium and incubated overnight. Serial dilutions of the respective HPVE6 virus were prepared by first making a 1:100 dilution i.e (10 µl virus stock to 990 µl sterile growth medium) then 1:10 serial dilutions were prepared from the 1:100 dilution by transferring 100 µl diluted HPVE6 virus to 900 µl sterile growth medium. The 96-

well culture plates were then removed from the incubator and the wells inspected to ensure that the cells had attached to form an even monolayer. Respective 100 µl diluted HPVE6 virus was then added to each well in columns 1–10 (Fig. 2.4), and 100 µl of virus free growth medium was also added to wells in columns 11–12 which served as controls for the viability of non-infected cells this was done in duplicate for each virus. The plates were then covered and incubated in a humidified CO_2 (5%) incubator for 10 days at 37°C. Using a microscope each well was then checked for cytopathic effect (CPE) while comparing the duplicate plates for similar results. Finally only one plate for each of the respective HPVE6 viruses was used to determine the viral titer by calculating the fraction of CPE-positive wells in each row (Fig. 2.4).

6.8 Amplifying Recombinant Adenovirus (Preparing High-Titer Stocks)

Following the determination of adenoviral titre using the End-point dilution assay (Clontech Expression, 2012), a multiplicity of infection (M.O.I.) of 1 was determined. This was used to quantify the optimal amount of virus required to seed seven 150 mm HEK 293 cell culture dishes in order to produce a greater quantity of high titer adenovirus. Each dish was seeded with $1.46x10^7$ cells with the calculated quantity of adenoviral E6 constructs until a 50% cytopathic effect (CPE) was observed and harvested by lysing with three consecutive freeze thaw cycles. This was then purified using the Adeno-X Maxi purification kit (Clontech Purification, 2013).

6.9 Evaluation of the Recombinant Virus (Confirmation of Constructs)

PCR analysis was done to verify that the encapsidated adenoviral genome contained a functional copy of the HPV E6 constructs, using the Adeno-X forward and reverse PCR primers and also by using the HPV11E6 and HPV18E6 specific primers. PCR as carried out in a 2720 Thermal cycler (Applied Biosciences, CA, USA) using primers specific to the HPV11E6 and HPV18E6 genes. PCR reactions were carried out using 1 x PCR buffer (KAPA Biosystems, Cape Town, South Africa), 1.5 mM $MgCl_2$, 0.4 uM dNTPs, 0.5 uM of each primer, 1U KAPA Taq DNA Polymerase and 5 ng plasmid DNA. Primers sequences and melting temperatures are described in Appendix A. Sequencing reactions and electrophoresis were performed by Macrogen Inc. (Seoul, South-Korea) using universal T7 primers and a BLASTn search was conducted against the database of nucleotide sequences provided by the National Centre for Biotechnology Information (NCBI) internet site

(http://ncbi.blast.nlm.nih.gov/BLAST/) to confirm sequence identity. The sequences were interpreted using the Chromas software version 2.01 (McCarthy, 2004)

6.10 Preparation of agarose gel for electrophoresis

For the separation of the amplified PCR products we used a 1% agarose gel containing 0.5 µg of Ethidium Bromide (EthBr)/ ml of gel solution from stock solution (10 mg/ml). Thus, we added 5 µl of stock solution for every 100 ml of gel solution. First a running buffer was made by adding 25 µl (10 mg/ml) of EthBr into 500 ml of 1X TBE (Tris-Borate-EDTA) solution. Then to make 500ml of 1% agarose gel we added 5 g of agarose powder to 500 ml of 1X TBE and the solution was dissolved by warming it up in a microwave for 5-10 min. When the agarose was completely dissolved it was allowed to cool down to 40-50°C and 25 µl of EthBr was added and mixed properly. The gel was poured into the gel tank, combs were inserted and allowed to set, being careful to remove all the bubbles present on the gel surface. When the gel had set the comb was removed and the gel placed into the electrophoresis tank filled with 500 ml of running buffer. The total volume of each PCR product (10 µl) was then mixed with 2 µl of 6X DNA loading dye and then loaded into the wells of the gel. As marker, we used GeneRuler DNA Ladder Mix from Fermentas. The electrophoresis chamber was then connected to the battery by the +/– conductors and the running was carried out for 1 hour at 100volts. The gels were then digitally photographed using a UV transilluminator (BioSpectrumTM 500 Imaging System, Upland, CA, USA) with VisionWorks LS Image Acquisition and Analysis Software (Version 6.8).

6.11 Maintenance of cells in culture

HaCaT Cells were cultured at 37°C in a humidified atmosphere of 5 % CO_2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % heat inactivated fetal calf serum (FCS), 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. The medium for HaCaT cells was changed every 2 to 3 days, and the cells were sub-cultured every 5 to 7 days, or when 70% confluent, by discarding the used medium and washing the cells with Phosphate Buffered Saline (PBS). Cells were trypsinized with 0.05% trypsin-EDTA and once they had detached, the trypsin-EDTA was inactivated by adding 5 ml of complete DMEM. Cells were collected by centrifugation at 4000 rpm for 5 minutes, resuspended in DMEM and replated at a dilution of 1:3. Preparation of frozen stocks involved the steps described above, however, instead of replating the cells, they were resuspended in cell freezing medium (70 % DMEM, 20 % FCS and 10 % DMSO) by gently pippetting twice. The cells were aliqouted

into 2 ml ampoules at about 1 x 10^6 cells/ml and stored at -80^0C for 48 hours before being transferred to a liquid nitrogen tank. Every two months the cells were checked for mycoplasma contamination. When required, frozen stocks were thawed at 37°C in a water bath, and directly plated into 10 ml of supplemented DMEM. The cells were incubated overnight at 37°C, in a humidified atmosphere of CO_2 (5%) after which the medium was replaced with fresh DMEM. The techinique for maintaining the Minimum essential medium for HEK293 cells was also the same as DMEM.

6.12 Infection of HaCaT with Adenovirus-E6 Constructs

Using a 6 well plate 2.0x10^5 HaCaT cells were plated into each well and incubated overnight, until 70% confluent. The media was removed and the cells washed once with 1x PBS. 1ml of DMEM media was mixed with respective E6 construct (MOI 100) and1µL DEAE-dextran (7.5µg/ml) and was added gently drop-wise onto each well. The plate was then sealed with parafilm and incubated at 37°c in a humidified atmosphere of CO_2 (5%) for 1 hour; the plates were then centrifuged for 1 hour at 2000rpm set at 37°c. A fresh 1ml of DMEM media mixed with respective E6 construct (MOI 100) and1µL DEAE-dextran (7.5µg/ml) and was added gently drop-wise onto each well making the a total final volume of 2 ml. The plate was then sealed with parafilm and incubated at 37°c in a humidified atmosphere of 5 % CO_2 for 1 hour; the plates were then again centrifuged for 1 hour at 2000rpm set at 37°c. The plates were then incubated at 37°c in a humidified atmosphere of CO_2 (5%) for 48 hours. Finally, DNA, RNA and Protein extraction was done to enable the confirmation of the expression of HPV11E6 and HPV18E6 at both an mRNA level through qRT-PCR and at a protein level through western blotting and confocal microscopy.

6.13 Determination of Multiplicity of infection (MOI) and Time point.

Using a 6 well plate 2.0x10^5 HaCaT cells were plated into each well as three biological replicates for each time point and incubated overnight, until 70% confluent. The cells were then infected with HPV11E6 and HPV18E6 and incubated at various time points i.e. 12 hours, 24 hours, 48 hours and 72 hours at 37°C, CO_2 (5%), in a humidified atmosphere a control uninfected plate was also included. After each time point cells were harvested, RNA extracted and cDNA prepared which was used for qRT-PCR (section 6.16). The MOI and time point which gave the highest viral infectivity based on mRNA levels was selected.

6.14 DNA Extraction

6.14.1 DNA Extraction from Agarose Gel

A known amount of plasmid DNA was digested with the required restriction enzymes for at least 1 hr, depending on the enzyme being used, and the digested material resolved on 1% agarose gels containing 0.5 µg/ml ethidium bromide. DNA molecular weight markers were routinely included in all agarose gels to determine the approximate sizes of the DNA bands. The DNA fragment of interest was excised from the gel, excess agarose was removed and gel pieces transferred to a microfuge tube containing 500 µl of Promega DNA binding solution. The DNA was extracted using the Promega gel extraction kit (Promega, 2010). Following electrophoresis, the DNA band was excised from gel and placed in a 1.5ml microcentrifuge tube. 10µl Membrane Binding Solution per 10mg of gel slice was added vortexed and incubated at 50–65°C until gel slice was completely dissolved. An equal volume of membrane binding solution was added to the dissolved gel slice and transferred to a minicolumn assembly. The mixture was then centrifuged at 16,000 ×g for 1 minute and the flow through discarded. The preparation was washed by adding 700µl Membrane Wash Solution (ethanol added) and centrifuged at 16,000 ×g for 1 minute and discarded the flow through, this step was reapeated by 500µl Membrane Wash Solution and centrifuge at 16,000 ×g for 5 minutes and the flow through discarded. The minicoloumn was then recentrifuged for 1 minute to allow evaporation of any residual ethanol.The Minicolumn was then carefully transferred to a clean 1.5ml microcentrifuge tube. 50µl of Nuclease-Free Water was then added to the minicolumn and incubated at room temperature for 1 minute, then centrifuged at 16,000 ×g for 1 minute. The minicolumn was then discarded and DNA stored at –20°C.

6.15 RNA Extraction

RNA extraction from cell cultures was performed using Qiazol Lysis Reagent (QIAGEN, Hilden, Germany). Three clones each from the three cell lines, HaCaT control, HPV11E6 and HPV18E6 were harvested using trypsin-EDTA as described above. The cells were collected by centrifugation at 4000 rpm for 5 minutes, resuspended in 1 ml PBS and transferred to a 1.5 ml microcentrifuge tube. Cells were collected by centrifugation at 4000 rpm for 5 minutes. The supernatant was discarded, and the cell pellet resuspended in 1 ml Qiazol. The cells were lysed by aspirating with a pipette, and incubated for 5 minutes at room temperature. Chloroform, 200µl per 1ml of Qiazol, was added shaken vigorously (not vortexed) for 15 seconds to mix well and incubated for 3 minutes at room temperature. The samples were then

centrifuged at 14000 rpm for 15 minutes at 4°c. The upper (clear) aqueous layer was transferred to a fresh tube and 0.5 ml of isopropanol was added and mixed thoroughly by shaking for 15 seconds this was then incubated at room temperature for 10 minutes. The samples were then centrifuged at 14000rpm for 20 minutes at 4°c to pellet the RNA. The supernatant was carefully removed. 1ml 75% DEPC-ethanol was added and vortexed on low for 5-10 seconds to wash the pellet thoroughly, this was then centrifuged at 12000 rpm for 5 minutes at 4°c to re-pellet the RNA. The supernatant was carefully removed and the pellet air dried at room temperature for 10 minutes. The pellet was then dissolved in 30µl DEPC-water by gentle pipetting and incubated at 55°c for 5 minutes. RNA was quantitated on a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA) and the integrity checked by electrophoresis on 1% agarose formaldehyde gels containing 0.5 µg/ml ethidium bromide and stored at -80°C until needed.

6.15.1 Preparation of a 1% formaldehyde agarose gel

Agarose formaldehyde mini gels were prepared by boiling 0.5 g high-gelling agarose in 45 ml distilled water mixed with 5 ml of 10 x MOPS. After the solution had cooled down to approximately 65^{o}C, 2.7 ml of 27 % formaldehyde was added. After cooling further to 50^{o}C, ethidium bromide was added to the solution and mixed properly. The gel was poured into the gel tank, combs were inserted and allowed to set in the fume hood. MOPS (1x) was added to the tank and used as the running buffer.

6.15.2 RNA Visualization

1 µg of RNA was mixed with RNA loading buffer and the mixture added to the formaldehyde agarose gel wells. Running standards were loaded beside the samples. Electrophoresis was done at 60V for 45 minutes. The gels were then digitally photographed using a UV transilluminator (BioSpectrumTM 500 Imaging System, Upland, CA, USA) with VisionWorks LS Image Acquisition and Analysis Software (Version 6.8).

6.16 cDNA synthesis

Conversion of mRNA to cDNA was performed using the ImProm-II™ Reverse Transcription System (Promega, WI, USA), following manufacturer's instructions. Briefly, 2 µg template mRNA, oligo dT and dH_2O up to 10 µl, were added together and incubated for 5 minutes at room temperature, 10 minutes at 70°C and 5 minutes at 4°C. $MgCl_2$, buffer, dNTPs, RNAsin and reverse transcriptase were added and incubated for 90 minutes at 42°C, followed by 10

minutes at 70°C to inactivate the reverse transcriptase. The cDNA was stored at -20°C until needed.

6.17 Quantitative Real-Time PCR (qRT-PCR)

Quantitative real-time PCR was performed on a Roche Lightcycler 480 II. To a mixture of SYBR Green PCR Master Mix (KAPA SYBR Fast qPCR Kit, KAPA Biosystems), 20 uM of forward and reverse primers, as well as 2 µl of cDNA was added, in a total volume of 20 µl. GAPDH was used as a housekeeping gene. Primers and melting temperatures are shown in Appendix A.

6.18 Western Blotting

HaCaT cells were infected as explained earlier with HPV11E6 and HPV18E6, a control uninfected plate was also included. The cells incubated for 48 hours at 37°C, CO_2 (5%) in a humidified atmosphere. Before harvesting, the spent media was aspirated, and the cells rinsed twice with ice cold PBS. Cold cell lysis buffer (Cell Signalling, MA, USA), approximately 100 µl for a 100 mm cell culture dish and 50 µl for a 6-well dish, was added, and incubated with the cells for 5 minutes on ice. A cell scraper was used to harvest the cells, which were transferred to a 1.5 ml microcentrifuge tube and sonicated for 10 seconds (Heat System-Ultrasonics, Inc, Plainview, New York). The lysate was centrifuged for 15 minutes at 14000 rpm, at 4°C. The supernatant was transferred to a fresh tube, leaving behind the pellet of insoluble material and cell debris. Protein was quantitated using the Pierce BCA (Bicinchoninic Acid) Protein Assay (Thermo Scientific, IL, USA) according to manufacturer's instruction. Briefly, a standard curve was constructed using Bovine Serum Albumin (BSA) diluted in dH_2O to concentrations between 0 and 2000 µg/ml. Protein samples were diluted 10x in 25 µl dH_2O. The assay was conducted in a 96-well plate, where 200 µl working solution, prepared by adding reagent A (sodium carbonate, sodium bicarbonate, BCA detection reagent, sodium tartrate in 0.1 M NaOH) and reagent B (4% CuSO4.5H20) together in a ratio of 50:1. The standards and samples were added to the reaction mix, and incubated at 37°C for 30 minutes, where after the absorbance was measured at 595 nm. The equation of the slope of the standard curve was used to determine the concentration of the protein samples. Equal protein concentrations (100µg/µl) were separated electrophoretically on a 12% SDS-PAGE (Laemmli, 1970) at 180 V. A PageRuler ™ Plus Prestained Protein Ladder (Fermentas, MD, USA) was also loaded in order to visualize the size of the protein of interest. The proteins were then transferred

electrophoretically onto nitrocellulose membrane in a transfer tank (Biorad, CA, USA) for 1.5 hours at 100 V, according to a standard protocol (Towbin et al., 1979). Gels were stained using coomassie blue staining solution (50% methanol, 10% glacial acetic acid, 0.25% coomassie blue). Nitrocellulose membranes were rinsed three times in TBS (50mM Tris pH7.5, 150mM NaCl) containing 0.1% Tween-20 (TBST) and blocked in 5% non-fat milk in TBST for 1 hour with shaking at room temperature to reduce non-specific binding. Primary antibodies against HPV18E6, p53, p21 and GAPDH were diluted 1:1000 in blocking solution. Membranes were incubated with the indicated primary antibody (Table 6.1) at 4°C overnight with shaking. Membranes were washed 3 times for 15 min in TBST and then incubated for 1 hour in 5% non-fat milk blocking solution. Horseradish peroxidase-conjugated secondary antibodies were diluted 1:5000 in 5% blocking solution. Membranes were incubated with secondary antibody for 1 hour at room temperature with shaking. Protein bands were detected using the Lumiglo Reserve chemiluminescent substrate system detection kit (KPL, Gaithersburg, MD, USA) and the UVP BioSpectrum chemiluminescent detection system (BioSpectrumTM 500 Imaging System, Upland, CA, USA) with VisionWorks LS Image Acquisition and Analysis Software (Version 6.8). Membranes were then stripped of primary and secondary antibodies by washing with 1M glycine pH 2.5 for 10 min with shaking, after which 1ml Tris pH 7.5 was added and membranes were washed three times for 10 min in TBST. Membranes were incubated with blocking solution for 1 hour before being re-probed with a different primary antibody.

Table 6.1: Antibody conditions for western blotting

Protein detected	**Primary Antibody**	**Secondary Antibody**
GAPDH	Rabbit anti-GAPDH (Santa Cruz) 1:1000 dilution 5 % non-fat dried milk powder in 0.1% TBST	Goat anti- rabbit HRP-conjugate (Bio rad) 1:5000 dilution 5% non-fat dried milk powder in 0.1% TBST
HPV18E6	Mouse anti-HPV E6 type 18 (Arbovita) 1:1000 dilution 5 % non-fat dried milk powder in 0.1% TBST	Goat anti-mouse HRP-conjugate (Bio rad) 1:5000 dilution 5% non-fat dried milk powder in 0.1% TBST
P53	Mouse anti-P53 (Santa Cruz) 1:1000 dilution 5 % non-fat dried milk powder in 0.1% TBST	Goat anti-mouse HRP-conjugate (Bio rad) 1:5000 dilution 5% non-fat dried milk powder in 0.1% TBST
P21	Rabbit anti-P21 (Santa Cruz) 1:1000 dilution 5 % non-fat dried milk powder in 0.1% TBST	Goat anti- rabbit HRP-conjugate (Bio rad) 1:5000 dilution 5% non-fat dried milk powder in 0.1% TBST

6.19 Double Immunofluorescence for Confocal Microscopy

A microscope cover slip was placed inside a 6-well cell culture dish, and 1.0 x 10^5 HaCaT cells were seeded onto the cover-slips in triplicate. The cells were infected as explained earlier with HPV11E6 and HPV18E6 with some being left uninfected to act as controls. The plates were incubated for 48hours at 37°C, 5% CO_2, in a humidified atmosphere. After incubation, media was removed and the cells rinsed twice in pre-warmed 1x PBS. The cover-slips with cells were then fixed by being immersed in absolute ethanol and incubated at -20°c for 5 minutes. The cells were then fixed in 4% Paraformaldehyde in PBS for 5 minutes at room temperature and washed in PBS, 3 times for 10 minutes. Cells were then incubated in Blocking Solution (1% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS) for 1 hour at room temperature to reduce non-specific staining. The cells were then incubated with the appropriate pair of primary antibodies (Table 6.2) diluted 1:100 in blocking solution, this was then sealed in a humidified chamber and incubated overnight at 4^{o}C. The cells were then washed in PBS, 3 times for 10 minutes and incubated in secondary antibodies diluted 1:500 in blocking solution. After which the cells were incubated for 90 minutes in a humidified chamber in the dark and then washed in PBS, 3 times for 10 minutes at room temperature. Counterstaining was done with a DAPI Nuclear Stain (0.5 µg/ml in PBS) diluted 1:100 in blocking solution for 10mins and the cells washed again in PBS, 3 times for 10 minutes at room temperature. The cells were then mounted onto microscope slides with mowiol (Sigma-Aldrich, Steinheim, Germany) containing n-propylgallate as anti-fading agent. The slides were then stored in the dark at room temperature for the mowiol to set and then stored in the dark at 4°C until viewing. The slides were finally analysed using a confocal microscope (Zeiss LSM 510 Meta with NLO).

Table 6.2: Antibody conditions for Double Immunofluorescence

Protein	**Primary Antibody**	**Secondary Antibody**
HPV18E6	Mouse anti-HPV E6 type 18 (Arbovita) 1:100 dilution 1 % BSA	Sheep anti-Mouse cy3 (sigma) 1:500 dilution 1 % BSA
P53	Mouse anti-P53 (santa cruz) 1:100 dilution 1 % BSA	Goat anti-Mouse cy3 (sigma) 1:500 dilution 1 % BSA

6.20 Flow cytometry

HaCaT cells were infected as explained earlier with HPV11E6 and HPV18E6, a control uninfected plate was also included. The cells incubated for 48 hours at 37°C, CO_2 (5%) in a humidified atmosphere. After incubation, the cells were harvested as described previously. The cell pellet was resuspended in PBS, and washed by centrifugation at 4000 rpm for 5 minutes and resuspended in 1 ml of 70% ethanol. Before analysis, samples were centrifuged at 4000 x g for 5 minutes. The pellet was washed twice in PBS by centrifugation at 4000 x g for 5 minutes. The pellet was resuspended in 100 µl PBS containing 100 µg/ml Ribonuclease-A (RNase), and left at room temperature for 5 minutes. Finally, 500 µl of propidium iodide (50 µg/ml in PBS) was added to the cells, which were stored in the dark at 4°C until flow cytometric analysis was performed (1 – 2 hours). The experiment was performed in triplicate on three clones from each HPVE6 virus. Flow cytometric cell cycle analysis was carried out using a BD FACS ™ Calibur flow cytometer (BD Biosciences, CA, USA). For each sample, at least 20 000 cells were measured for DNA content. Analysis of the results was performed using ModFit LT software (Verity Software House) to calculate the percentage of cells in each stage of the cell cycle.

6.21 Soft agar assay

HaCaT cells were infected as explained earlier with HPV11E6 and HPV18E6, a control uninfected plate was also included. The cells incubated for 48 hours at 37°C, CO_2 (5%) in a humidified atmosphere. The used media was then discarded and the cells washed with Phosphate Buffered Saline (PBS), trypsinized with 0.05% trypsin-EDTA and once detached, the trypsin-EDTA was inactivated by adding 5 ml of complete DMEM. Cells were collected by centrifugation at 4000 rpm for 5 minutes and resuspended in 3ml DMEM. A base agar was prepared by mixing 1% agarose gel (40°c) with 2x DMEM (20% FCS and 2% P/S). 1ml of the mixture was added to each well of a 6 well plate and swirled to cover the whole surface. The plates were then stored at 4°c for upto one week sealed in a parafilm or let to set at room temperature for 30 minutes before using them. Once the base agar had set a top agar base was prepared by melting 0.7% DNA agarose in microwave and cooling it to 40°c. The HPVE6 infected and HaCaT control cells were counted to 2,500 cells/plate. By using 15,000/tube, there was enough to plate a six well agar plate from each original tissue culture plate, the volume was adjusted so that the cell count = 150,000 cells/ml. A mixture of 0.1ml cell suspension, 3ml of 2x DMEM (20% FCS and 2% P/S) and 3ml of 0.7% DNA agar was

added into a 10ml tube and mixed gently. Then 1ml of the mixture was added gently onto the warmed base agar in triplicate and incubated at 37°C in 5% CO_2 humidified incubator for 10 to 30 days with feeding of the cells with 1ml DMEM 1-2 times per week. After 21 days the cells were stained with 0.5 ml of 0.005% crystal violet for more than 1 hour and the pink colonies were counted using an Olympus CKY41 light microscope set at 10 x magnification.

6.22 Microarray analysis

RNA samples extracted from HaCaT cells infected as explained earlier with HPV11E6 and HPV18E6 with a control uninfected plate in triplicate were analysed on the agilent bioanalyzer using the total RNA Nano Assay. An internal quality control sample control RNA-50ng/µl was included as internal reference. The NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA) was used to determine sample concentration and the presence of possible contaminants followed by the Agilent Bioanalyzer Nano Assay to assess the integrity of the samples prior to downstream application(Agilent, 2014). The prepared samples were then hybridized overnight to the Affymetrix Human Gene ST 2.0 arrays. Following hybridization, the arrays were washed and stained using the GeneChip Fluidics Station 450 and scanned using the GeneChip® Scanner 3000 7G. An initial analysis was carried out to determine the quality of the data using expression console, which included a range of quality assessment metrics with this software being used primarily in determining the quality of the data and the identification of any possible outliers. Raw intensity files were normalised using the Robust Multichip Average (RMA) method for affymetrix microarray chips (Affymetrix, 1999). Analysis of Variance (ANOVA) was used in identifying the differentially expressed genes having a cut-off fold change of between –2.0 to 2.0 and significance FDR of ≤ 0.05, this was achieved through the use of the Partek® Genomics Suite™ 6.6 software(Partek Inc., 2014).

6.22.1 Venn diagram

A web-based Venn diagram (Pirooznia et al., 2007) was used to calculate and identify genes that were uniquely induced by either HPV11E6 or HPV18E6 and genes that were shared between the two datasets.

6.22.2 Gene Ontology (GO) analysis

A web-based GOTree Machine (GOTM) (Zhang et al., 2004) that identifies enriched GO terms in ranked lists of genes was used for GO analysis of the differentially expressed genes.

The output of the enrichment analysis was visualised as a bar graphs and acyclic graphs, providing a view of the relations between GO terms.

6.22.3 Pathway analysis

The Ingenuity™ Pathway Analysis (IPA) system (Ingenuity, 2014) was used to identify the top most significantly regulated pathways from Venn diagram list, and identified genes that were uniquely induced by either HPV11E6 or HPV18E6 and genes that were shared between the two datasets. The results were presented graphically based on scoring of the ratio of significant genes present in the canonical pathway to the total number of molecules in the canonical pathway. The threshold level was set at $p = 0.05$. The Kyoto Encyclopedia of Genes and Genomes (KEGG) software (Kanehisa et al., 2004) was also used due to its capability to integrated current knowledge on molecular interaction networks in biological processes.

6.22.4 Network analysis

The Ingenuity™ Pathway Analysis (IPA) system (Ingenuity, 2014) was also used to generate networks which were assembled based on gene connectivity with other genes based on the assumption that the more connected a gene, the more influence it had and the more important it was. Networks were assembled using decreasingly connected molecules from data set of respective HPVE6 DEGs. A maximum of 35 genes comprised a network which was annotated with high-level functional categories. Focus molecules with the most interactions to other focus molecules were connected together to form a network with non-focus molecules and dataset genes from the knowledge base were also added to fill areas lacking connectivity, the resulting networks were scored and then sorted based on the p-score.

6.23 Whole genome sequencing (WGS)

6.23.1 DNA Extraction from Blood

Genomic DNA was extracted according to the protocol from (Gustafson et al., 1987)**,** which consisted of the following steps; Whole blood samples were defrosted and transferred to a sterile polypropylene 50 ml tube and diluted with 40-45 ml of phosphate buffered saline (PBS) solution and then centrifuged at 7000 rpm for 15 min. The supernatant was then poured off and the pellet suspended in 25 ml of Sucrose Triton X-100 Lysing Buffer. The homogenized solution was spun for 5 min at 6000 rpm and the supernatant poured off. The

pellet was then resuspended in 3 ml of T20E5 solution, 200 µl of 10% Sodium dodecyl sulfate (SDS) to a final concentration of 1% and 100 µl of proteinase K, 10 mg/ml to a final concentration of 250 µg/ml. This solution was incubated at 45°C overnight. 4 ml of saturated NaCl was added and mixed vigorously for 15 sec. and the samples spun for 40 min at 7000 rpm to allow for precipitation of proteins. The supernatant containing the DNA was transferred to a 12 ml tube and absolute ethanol was added in order to precipitate the DNA. A white isolated lump made of precipitated DNA was then transferred to a 2 ml tube and washed in 1 ml of 70% ethanol. The tubes were then finally centrifuged at 7000 rpm for 5 min and pellets (DNA) were dissolved in 400 µl of 1X TE buffer. The DNA samples were then left at 4°C overnight or longer, until the pellet was completely dissolved in TE buffer. DNA was then quantitated on a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA) and the samples stored at –20°C until needed.

6.23.2 DNA Extraction from Tissue Biopsy

Frozen tissue was first thawed on ice and placed into a sterile microcentrifuge tube containing 0.5 ml of QIAzol Lysis Reagent™. The sample volume was not to exceed 10 % of the volume of the QIAzol Lysis Reagent™. For efficient disruption of tissue samples, the size of the sample was not supposed to be greater than half the diameter of the homogeniser probe, and the sample was into smaller pieces if necessary. The tip of the tissue pestle disposable probe was then placed into the QIAzol Lysis Reagent™ and the pestle operated the pestle at full speed until the tissue lysate was uniformly homogeneous (1-5min). An additional 0.5 ml of QIAzol Lysis Reagent™ reagent was added and mixed by vortexing for 30 sec. The homogenized samples were then incubated for 5 minutes at room temperature (20°-25°C) to permit the complete dissociation of nucleoprotein complexes. 0.2 ml of chloroform per 1ml of QIAzol Lysis Reagent™ was then added to the sample tubes, capped and shaken vigorously by hand for 15 sec. After which the samples were incubated at room temperature for 2 to 3 min. The samples were then centrifuged at 12 000 x g for 15 minutes at 4° C. Following centrifugation, the mixture separated into 3 phases: a colourless upper aqueous phase containing RNA, a white middle interphase containing DNA and a lower red organic phenol phase containing protein. The upper aqueous phase to a new tube and stored at 4°C for later RNA extraction. 0.3 ml of 100% ethanol was the added to the interphase and phenol phase, and carefully the samples were mixed by inversion and incubated at room temperature for 2–3 min. The samples were then centrifuged at 14000 x g for 5 min at 4° C to sediment DNA, and the phenol/ethanol supernatant was discarded. 1 ml sodium citrate solution was

then added to the DNA pellet and the samples Incubated at room temperature for 30 min, with mixing by inversion every 5 min. The samples were then centrifuged at 14,000 x g for 5 min at 4°C, and the supernatant carefully removed. 1.5 ml of 75% ethanol was then added to the DNA pellet and Incubated at room temperature for 20 min, with mixing by inversion every 5 min. The samples were then centrifuged at 12,000 x g for 5 min at 4° C and ethanol supernatant was completely removed before the DNA pellet was air dried for 5-15 min. The DNA pellet was dissolved by adding 200 ul of 8 mM NaOH to the pellet and the solution vortexed for 15 sec and before being incubated for 30 min at 60°C. The samples were then centrifuged at 14,000 x g for 10 min at room temperature to remove insoluble material, and the supernatant transferred to a new tube. The pH of 8 mM NaOH was approximately 9. For storage, the pH of the DNA sample solution was adjusted to pH 7–8 by addition of 23.4 ul of 0.1 M HEPES and 2.2 µl 100 mM EDTA (final concentration 1 mM). DNA was then quantitated on a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA) and the samples stored at –20°C until needed.

6.23.3 DNA concentration measurement

DNA concentration was determined by NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific, IL, USA). For each sample, 2 µl of dissolved DNA was loaded on the spectrophotometer, which provided the DNA concentration in ng/µl, the 260/230 ratio for organic chemicals and solvent contamination (important to have this ratio above 1.6 for PCR), and the 260/280 ratio for protein contamination (ratios above 1.8 were considered relatively good quality). The samples were then stored at 4°C prior PCR.

6.23.4 Variant Calling

Four oesophageal squamous cell carcinoma (OSCC) DNA biopsy samples were submitted to CPGR for whole genome sequencing. Quality control (QC) measures were carried out to validate the integrity of the samples for whole genome sequencing with the Quant-iT™ PicoGreen® dsDNA kit (Bhat et al., 2010). Paired-end sequencing was performed on the Illumina HiSeq2000, with 2x100 bp reads (Wang et al., 2010). Reads were aligned to the NCBI37 Homo sapiens reference genome using ELAND and CASAVA (Kidd et al., 2010). Structural variants (SV) reported from the alignment were collated with gene loci, (upstream, downstream and overlapping) using the Variant Effect Predictor of Ensembl (McLaren et al., 2010). The affected genes were subsequently cross-checked against the genetic association database for disease and cancer associations (Becker et al., 2004).

6.23.5 Determination of *BCLAF1* deletion

The *BCLAF1* deletion was determined by PCR assay as previously described (Matejcic et al., 2011). A two primer set (Table 6.1) was used to identify both the deletion and non-deletion in a single reaction. Briefly, 50 ng of genomic DNA was added to buffer containing 1.5 mM MgCl2, 2 mM dNTPs, 10 pmol of each oligonucleotide primer and 0.5 U of HotStart Taq polymerase (Roche). Fragments were amplified in a final volume of 25 µl. The PCR reaction was carried out in a Veriti Thermal Cycler (Applied Biosystems) and thermal cycling conditions consisted of an initial hot start of 10 min at 95°C followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec, and extension at 72°C for 3 min, with a final extension step at 72°C for 7 min. PCR products were separated by electrophoresis on 1% agarose gels in Tris-Borate-EDTA buffer. The gels were stained with ethidium bromide and digitally photographed using a UV transilluminator (BioSpectrumTM 500 Imaging System, Upland, CA, USA) with VisionWorks LS Image Acquisition and Analysis Software (Version 6.8). The resulting bands in the gel were distinguished by comparing their position with that of standard bands from the GeneRuler DNA Ladder Mix marker (Fermentas).

6.23.6 Determination of viral sequence integration sites

Reads from the WGS that didn't map to the human reference genome were probed for viral association with OSCC. The program DeconSeq (Schmieder & Edwards, 2011) was used to align the unmapped reads against complete set of NCBI RefSeq Viral Genomes (build 64) using the Burrows Wheeler Aligner (BWA) (Li & Durbin, 2010). DeconSeq allows detecting of plausible matches based on a user specified match identity and sequence coverage. For our detection we set the identity match to a minimum of 95% and sequence coverage to a minimum of 90%. From the DeconSeq results a list of the most to least frequent viral hits were built. The BWA database indexing, alignment and filtering were done on the UCT HPC cluster (http://hex.uct.ac.za). From the list of viral sequences identified in all the four samples, 22 viral sequences were selected which were common and of human origin in all the four samples. In order to detect the viral integration sites we used BWA to align all the reads from the WGS to the human reference genome (hg19, UCSC) and the 22 viral genomes. All the discordant reads were extracted from the complete mapping and only those matches where one of the read pairs mapped to the human genome and the other read pair mapped to any of the viral genome were selected (Jiang et al., 2012). In order to achieve a high confidence in viral integration, we further filtered the selected viral reads to only those that

had 100% match in both the BWA alignment score and CIGAR string to the reference human genome and human viral sequence. A Circster image was then generated on Galaxy to visualize the chromosomal integration sites of the filtered viral sequences for the four patient samples (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010).

6.24 Solutions & Buffers

Acrylamide/Bisacrylamide Solution (30%)
29 g Acrylamide
1 g Bisacrylamide
Made to 100 ml with distilled water and stored in dark at 4°C

7.5 M Ammonium acetate, pH 7.4
Dissolve 57.81 g ammonium acetate in 60 ml of distilled water, adjust pH to 7.4 with ammonia solution, make up to 100 ml with distilled water and sterilise by filtration using a filter of 0.45 μm pore size

2 M Ammonium acetate, pH 7.4
Dissolve 15.416 g ammonium acetate in 60 ml of distilled water, adjust pH to 7.4 with ammonia solution, make up to 100 ml with distilled water and sterilise by filtration using a filter of 0.45 μm pore size

10 % Ammonium persulfate (APS)
Dissolve 100 mg of APS in 1 ml of distilled water, store at 4°C

Ampicillin Stock Soln (10 mg/ml)
Dissolve 500 mg Ampicillin in 50 ml of distilled water. Sterilise by filtration through a filter of pore size 0.22 μm and store at -80°C. Dilute to a working solution of 50 μg/ml.

Coomassie blue staining solution
0.5 g Coomassie BB
500 ml methanol
100 ml acetic acid
400 ml distilled water
DEPC-Treated water (0.01 %)
Add 10 μl of DEPC to distilled water (100 ml) and stir at room temperature. Leave overnight and then autoclave

DEAE-dextran (Stock); Sigma Aldrich (D9885-10G)
Reconstitute 5 grams of DEAE-Dextran in one litre of distilled sterile water.
Filter through a 0.45-micron filter and aliquot 10 ml in 15 ml conical tubes.
Store tubes at - 80°C

Destaining Solution (10 % Acetic acid (v/v))
100 ml acetic acid
100 ml methanol
800 ml distilled water

6X DNA loading dye
Weigh 25 mg bromophenol blue, 25 mg xylene cyanol FF, and 4 gm Sucrose
Dissolve them in 9 ml distilled water to make up 10 ml solution
Store the solution in small aliquots (1 ml) at -20°C for long time.

0.5 M EDTA, pH 8.0
Add 37.22 g of Na2EDTA.2H2O to 140 ml of distilled water. Adjust the pH to 8.0 with 10 N NaOH then add distilled water to a final volume of 200 ml. Autoclave and store at room temperature.

70 % Ethanol
70 ml of absolute ethanol
Add 30 ml of distilled water

Ethidium Bromide
Stock solution 10 mg/ml
Final concentration 0.5 mg/ml
Add 10 µl of stock solution (10 mg/ml) to 200 ml dH2O to get a final concentration of 0.5 mg/ml
Store the solution in a light-proof container at 4°C for long time.

75 % DEPC- Ethanol
75 ml absolute ethanol
25 ml DEPC-treated water

Fixing Solution
100 ml acetic acid
500 ml methanol
400 ml distilled water

Freezing down media
70 % Fetal bovine serum
20 % DMEM media
10 % Dimethylsulphoxide (DMSO)
Store at 4^0C

Inoue transformation buffer
The Inoue transformation buffer was prepared by dissolving all the solutes listed below in 800 ml of distilled H_2O and then 20 ml of 0.5 M PIPES (pH 6.7) was added. Adjust the volume of the Inoue transformation buffer to 1 litre with distilled H_2O

Reagent	Amount per litre	Final Concentration
$MnCl_2.4H_20$	10.88g	55mM
$CaCl_2.2H_2O$	2.20g	15 mM
KCl	18.65g	250mM
PIPES(0.5M,Ph 6.7)	20ml	10mM
H_2O	to 1 litre	

The Inoue transformation buffer was then sterilized by filtration through a pre-rinsed 0.45-mm Nalgene filter. Divide into aliquots (10 ml) and store at -20ºC

Luria Bertani (LB) Broth (with Ampicillin)
10g Bacto Tryptone
5g Bacto Yeast Extract
10g Nacl
Make to 1 litre with DH_2O
pH 7.0
Autoclave and allow to cool to about 40^0C then add Ampicillin to a final concentration of 100 µg/ml.

Luria Bertani (LB) Plates (with Ampicillin)
10g Bacto Tryptone
5g Bacto Yeast Extract
10g Nacl
Make to 1 litre with DH_2O
pH 7.0
Autoclave and allow to cool to about 40^0C then add Ampicillin to a final concentration of 100 µg/ml. Pour into 100 mm dishes and let the agar harden. Store at 4^0C

10 X Loading Buffer
0.25 % bromophenol blue in 40 % glycerol (v/v)

10 x MOPS Buffer
0.4M MOPS, pH 7.0
0.1M Sodium acetate

5 % non-fat milk blocking solution
Dissolve 2.5 g non-fat milk powder in 30 ml TBST; make up to 50 ml with TBS-T

1X PBS (pH 7.4)
154 mM NaCl
2.7 mM KCl
5.6 mM Na2HPO4 or Na2HPO42H2O
1mM KH2PO4
Adjust pH to 7.4 using concentrated HCl or NaOH if necessary
Dissolve into 1 liter distilled water and autoclave.

Phenol: Chloroform: Isoamyl alcohol (25:24:1)
Mix one part of phenol and one part of chloroform: isoamyl alcohol (24:1)

5 x Protein loading dye
200 µl 5 x protein loading buffer
50 µl of β-mercaptoethanol, 10 µl of bromophenol blue

RIPA Buffer
200 µl RIPA solution
2 µl of 2 mg/ml pepstatin
2 µl of 10 mg/ml aprotinin
2 µl of 100 µM PMSF

RNA Loading Buffer (25 ml)
1 mM EDTA, pH 8.0
0.25 % Bromophenol Blue (w/v)
0.25% Xylene Cyanol
50% Glycerol (v/v)
50 µl DEPC
Add distilled water to a final volume of 25 ml, Stir overnight and autoclave

10 % SDS
Dissolve 10 g sodium dodecyl sulphate in 80 ml distilled water (heat to 80°C), make up to 100 ml with distilled water

10 × Running buffer
29 g Tris
144 g Glycine
10 g SDS
Made to 1 Litre with distilled water

SDS-PAGE Running Buffer (5 x)
0.25 M Tris-HCl (pH 8.3)
1.92 M Glycine
0.5 % SDS
SDS/Tris Buffer
5% SDS (w/v)
0.5 M Tris-HCl (pH 7.4)
0.125 M EDTA

Stripping Solution
1.4 ml β-mercaptoethanol
40 ml 10 % SDS
12.5 ml Tris-HCl, pH 6.7
Make up to 200 ml with distilled water

TAE Electrophoresis Buffer (1 x) (pH 8.3)
90 mM Tris
90 mM Boric Acid, 2.5 mM EDTA. Adjust pH to 8.3.

TBS-Tween
50 ml 1M Tris base
30 ml 5M NaCl
250 µl Tween-20
Make up to 1litre with distilled water

1X TBE
0.445M Tris (pH 8.0)
0.445M boric acid
0.01M EDTA (pH 8.0))
Add to 1 liter distilled water and autoclave.

TE buffer (10x) (pH 8.0)
100 mM Tris-Cl
10 mM EDTA
Adjust pH to 8.0

10×Transfer buffer
144 g glycine
38 g Tris
Make up to 1 litre with distilled water.

Tris Buffered Saline (TBS)
25 mM Tris (pH 7.5)
0.9% NaCl (w/v)

1M Tris pH 7.5 (per litre)
Add 121 g Tris base to 800 ml of distilled water
Adjust pH with concentrated HCl to pH 7.5 and add distilled water up to one litre. Autoclave

1M Tris pH 8.8 (per litre)
Add 121 g Tris base to 800 ml of distilled water
Adjust pH with concentrated HCl to pH 8.8 and then add distilled water up to one litre. Autoclave

1 X Trypsin-EDTA solution
0.05% (w/v) trypsin, 0.53 mM EDTA
Prepare in a calcium and magnesium-free 1 x PBS

APPENDIX-A

Table A1: Primer sequences and annealing temperatures

Primer	Sequence	Annealing temperature
BCLAF1	5'-CGCGTCGAAGGTAGCTCTAT-3' 5'-TTGGAGCGACCCATTTCTTTT-3'	60°C
BRCA1	5'-GGCTATCCTCTCAGAGTGACATTTTA-3' 5'-GCTTTATCAGGTTATGTTGCATGGT-3'	60°C
BUB1	5'-TCATTCATGGAGACATTAAAC-3' 5'-CTGAGCATCTCAACACACTG-3'	60°C
CCL21	5'-CAAGCTTAGGCTGCTCCATC-3' 5'-TCAGTCCTCTTGCAGCCTTT-3'	60°C
CCL26	5'- AACTCCGAAACAATTGTACTCAGCTG- 3' 5'-GTAACTCTGGGAGGAAACACCCTCTCC-3'	60°C
CYP24A1	5'-CAA ACCGTG GAAGGCTATC-3' 5'-AGTCTTCCCCTTCCAGGATCA-3'	60°C
GADD45A	5'-GAAGTCCGCGGCCAGGACACAGTTCC-3' 5'-GGTCCCCGCCGGGCTGTCACTCGG-3'	60°C
GAPDH	5'- GGCTCTCCAGAACATCATCC - 3' 5'- GCCTGCTTCACCACCTTC- 3'	60°C
HPV-11 E6	5'- AAGATGCCTCCACGTCTGCAA- 3' 5'- CTTGCAGTTCTAGCAACAGGC- 3'	60°C
HPV-18E6	5'- CCAGAAACCGTTGAATCCAG - 3' 5'- GAGTCGTTCCTGTCGTGCTC - 3'	60°C
IL8	5'-ACTGAGAGTGATTGAGAGTGGAC-3' 5'-AACCCTCTGCACCCAGTTTTC-3'	60°C
JUN	5'-CCT TGAAAGCTCAGAACTCGG AG-3' 5'-TGCTGCGTTAGCATG AGTTGG C-3'	60°C
MCM2	5'-GCTCTGGCCCTG TTTGGA-3' 5'-GAAGATGGCACGGCTAGACAC T-3'	60°C
MCM6	5'-CTCTGAATGCCAGGACATCCA T-3' 5'-CATTGCATTCATCCACCAGAA-3'	60°C
MMP1	5'-GGAGGGGATGCTCATTTTGATG-3' 5'-TAGGGAAGCCAAAGGAGCTGT-3'	60°C
MMP9	5'-GGCCAACTACGACACCGACGAC-3' 5'-CGCCGCCACGAGGAACAAAC-3'	60°C
MTTP	5'-GGACTT TTTGGATTTCAAAAGTGAC-3' 5'-GGAGAAACGGTCATAATTGTG-3'	60°C
NOX1	5'-CACGCTGAGAAAGCCATTGGATCAC-3' 5'-GGATGGAATAGGCTGGAGAGAACA-3'	60°C
P21	5'- ACCTCACCTGCTCTGCTGC-3' 5'- ATTAGGGCTTCCTCTTGGAGA-3'	65°C
P53	5'- CATCATCACACTGGAAGACTCC-3' 5'- CAGTGCTCGCTTAGTGCTCC-3'	65°C
PDK4	5'-CCCGCTGTCCATGAAGCAGC-3' 5'-CCAATGTGGCTTGGGTTTCC-3'	60°C
STARD7	5'-GGTAATCAAGCTGGAGGTGATTG-3' 5'-GAGTACATTGGATAAGGAAAATGGGT-3'	60°C
TP53INP1	5'-TGAACACATTTGCCTTGTGAA-3' 5'-GGCAAAGTGCTGTGCTGTT-3'	60°C
TSHB	5'-GGCAAACTGTTTCTTCCCAA-3' 5'-TCTGTGGCTTGGTGCAGTAG-3'	60°C
ZIC1	5'-AAACTGGTTAACCACATCCGC-3' 5'-CTCAAACTCGCACTTGAAGG-3'	60°C
BCLAF1 Deletion	5'-TGTGGGGAAAGGCAGAAAGG-3' 5'-CATCCACCATAAGCCGTGTA-3'	60°C
T7-Promoter	5'-TAATAC GACTCACTATAGGG-3'	DNA Plasmid Primers
CMV-Promoter	5'-GGCGGTAGGCGTGTA-3'	

APPENDIX-B

Table A2: Selected viral sequences identified from un-mapped reads (50). Based on a minimum of 95% identity match and a minimum of 90% sequence coverage on BWA alignments against NCBI's RefSeq Viral Genomes (build 64)

NCBI_RefSeq	Virus
NC_001694.1	Human papillomavirus 61, complete genome
NC_001493.2	Ictalurid herpesvirus 1 strain Auburn 1, complete genome
NC_022518.1	Human endogenous retrovirus K113 complete genome
NC_001782.1	Saccharomyces cerevisiae killer virus M1, complete genome
NC_004102.1	Hepatitis C virus genotype 1, complete genome
NC_001348.1	Human herpesvirus 3
NC_007346.1	Emiliania huxleyi virus 86, complete genome
NC_002229.3	Gallid herpesvirus 2, complete genome
NC_009334.1	Human herpesvirus 4
NC_001479.1	Encephalomyocarditis virus, complete genome
NC_009127.1	Cyprinid herpesvirus 3, complete genome
NC_001716.2	Human herpesvirus 7, complete genome
NC_009333.1	Human herpesvirus 8
NC_002641.1	Meleagrid herpesvirus 1, complete genome
NC_001806.1	Human herpesvirus 1, complete genome
NC_000898.1	Human herpesvirus 6B, complete genome
NC_001596.1	Human papillomavirus type 9, complete genome
NC_001664.2	Human herpesvirus 6A, complete genome
NC_014637.1	Cafeteria roenbergensis virus BV-PW1, complete genome
NC_007605.1	Human herpesvirus 4 type 1, complete genome
NC_009824.1	Hepatitis C virus genotype 3
NC_014567.1	Saimiriine herpesvirus 1, complete genome
NC_016447.1	Aotine herpesvirus 1 strain S34E, complete genome
NC_001847.1	Bovine herpesvirus 1, complete genome
NC_001526.2	Human papillomavirus type 16, complete genome
NC_001798.1	Human herpesvirus 2, complete genome
NC_001611.1	Variola virus, complete genome
NC_009827.1	Hepatitis C virus genotype 6
NC_007821.1	Enterobacteria phage WA13, complete genome
NC_002169.1	Spodoptera exigua MNPV, complete genome
NC_001897.1	Human parechovirus, genome
NC_002645.1	Human coronavirus 2
NC_001357.1	Human papillomavirus - 18, complete genome
NC_001987.1	Ateline herpesvirus 3 complete genome
NC_001457.1	Human papillomavirus type 4, complete genome
NC_006273.2	Human herpesvirus 5 strain Merlin
NC_001623.1	Autographa californica nucleopolyhedrovirus, complete genome
NC_009823.1	Hepatitis C virus genotype 2, complete genome
AC_000017.1	Human adenovirus type 1
NC_019445.1	Escherichia phage TL-2011b, complete genome
NC_002188.1	Fowlpox virus, complete genome
NC_008349.1	Plutella xylostella multiple nucleopolyhedrovirus, complete genome
NC_001954.1	Enterobacteria phage If1, complete genome
NC_011357.1	Stx2-converting phage 1717, complete prophage genome
NC_005294.1	Streptococcus prophage EJ-1, complete genome
AC_000007.1	Human adenovirus 2
NC_017972.1	Pseudomonas phage Lu11, complete genome
NC_001655.1	Hepatitis GB virus B, complete genome
NC_001330.1	Enterobacteria phage alpha3, complete genome
NC_001697.1	Haemophilus phage HP1, complete genome

REFERENCES:

1000 Genomes Project Consortium 2010. A map of human genome variation from population-scale sequencing. *Nature.* 467(7319): 1061-1073.

Abbas, T. & Dutta, A. 2009. p21 in cancer: intricate networks and multiple activities. *Nature Reviews Cancer.* 9(6): 400-414.

Ablashi, D.V., Devin, C.L., Yoshikawa, T., Lautenschlager, I., Luppi, M., Kühl, U. & Komaroff, A.L. 2010. Review Part 3: Human herpesvirus-6 in multiple non-neurological diseases. *Journal of Medical Virology.* 82(11): 1903-1910.

Affymetrix Ed. 1999. *Affymetrix Microarray Suite User Guide*. version 4 ed. Santa Clara, CA: Affymetrix.

Agilent 2014. *Agilent RNA 6000 Nano Kit, Quick Start Guide.* Available: http://www.genomics.agilent.com/en/product.jsp?cid=AG-PT-105&tabId=AG-PR-1172&_requestid=323496.

Agrawal, N., Jiao, Y., Bettegowda, C., Hutfless, S.M., Wang, Y., David, S., Cheng, Y., Twaddell, W.S., Latt, N.L., Shin, E.J. et al. 2012. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discovery.* 2(10): 899-905.

Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C. et al. 2009. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Research.* 19(9): 1622-1629.

Aird, D., Ross, M.G., Chen, W., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. & Gnirke, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12(2): R18.

Ajay, A.K., Meena, A.S. & Bhat, M.K. 2012. Human papillomavirus 18 E6 inhibits phosphorylation of p53 expressed in HeLa cells. *Cell & Bioscience.* 2(1): 1-11.

Akbari, M.R., Malekzadeh, R., Nasrollahzadeh, D., Amanian, D., Sun, P., Islami, F., Sotoudeh, M., Semnani, S., Boffeta, P. & Dawsey, S.M. 2006. Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. *International Journal of Cancer.* 119(5): 1047-1051.

Al Moustafa, A.E., Alaoui-Jamali, M.A., Batist, G., Hernandez-Perez, M., Serruya, C., Alpert, L., Black, M.J., Sladek, R. & Foulkes, W.D. 2002. Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene.* 21(17): 2634-2640.

Albertson, D.G., Collins, C., McCormick, F. & Gray, J.W. 2003. Chromosome aberrations in solid tumors. *Nature Genetics.* 34(4): 369-376.

Alekseev, O.M., Richardson, R.T., Alekseev, O. & O'Rand, M.G. 2009. Analysis of gene expression profiles in HeLa cells in response to overexpression or siRNA-mediated depletion of NASP. *Reprod Biol Endocrinol.* 7(1): 45.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 403(6769): 503-511.

Alkan, C., Coe, B.P. & Eichler, E.E. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics.* 12(5): 363-376.

Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., Kitzman, J.O., Baker, C., Malig, M. & Mutlu, O. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genetics.* 41(10): 1061-1067.

Amirian, E.S., Bondy, M.L., Mo, Q., Bainbridge, M.N. & Scheurer, M.E. 2014. Presence of viral DNA in whole-genome sequencing of brain tumor tissues from the cancer genome atlas. *Journal of Virology.* 88(1): 774-13.

Antonsson, A., Nancarrow, D.J., Brown, I.S., Green, A.C., Drew, P.A., Watson, D.I., Hayward, N.K., Whiteman, D.C. & Australian Cancer Study 2010. High-risk human papillomavirus in esophageal squamous cell carcinoma. *Cancer Epidemiology, Biomarkers & Prevention.* 19(8): 2080-2087.

Appelqvist, P. & Salmo, M. 1980. Lye corrosion carcinoma of the esophagus. A review of 63 cases. *Cancer.* 45(10): 2655-2658.

Arcasoy, S.M., Latoche, J.D., Gondor, M., Pitt, B.R. & Pilewski, J.M. 1997. Polycations increase the efficiency of adenovirus-mediated gene transfer to epithelial and endothelial cells in vitro. *Gene Therapy.* 4(1): 32-38.

Arena, A., Liberto, M.C., Iannello, D., Capozza, A.B. & Foca, A. 1999. Altered cytokine production after human herpes virus type 6 infection. *The New Microbiologica.* 22(4): 293-300.

Arzumanyan, A., Reis, H.M. & Feitelson, M.A. 2013. Pathogenic mechanisms in HBV-and HCV-associated hepatocellular carcinoma. *Nature Reviews Cancer.* 13(2): 123-135.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. & Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science (New York, N.Y.).* 297(5583): 1003-1007.

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. & Eichler, E.E. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Research.* 11(6): 1005-1017.

Balkwill, F. & Mantovani, A. 2001. Inflammation and cancer: back to Virchow? *The Lancet.* 357(9255): 539-545.

Balmain, A., Gray, J. & Ponder, B. 2003. The genetics and genomics of cancer. *Nature Genetics.* 33238-244.

Band, V., Dalal, S., Delmolino, L. & Androphy, E.J. 1993. Enhanced degradation of p53 protein in HPV-6 and BPV-1 E6-immortalized human mammary epithelial cells. *The EMBO Journal.* 12(5): 1847-1852.

Banks, L., Pim, D. & Thomas, M. 2012. Human tumour viruses and the deregulation of cell polarity in cancer. *Nature Reviews Cancer.* 12(12): 877-886.

Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P. & Gat-Viks, I. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature Genetics.* 41(11): 1238-1242.

Becker, K.G., Barnes, K.C., Bright, T.J. & Wang, S.A. 2004. The genetic association database. *Nature Genetics.* 36(5): 431-432.

Beglin, M., Melar-New, M. & Laimins, L. 2009. Human papillomaviruses and the interferon response. *Journal of Interferon & Cytokine Research.* 29(9): 629-635.

Belbin, T.J., Singh, B., Barber, I., Socci, N., Wenig, B., Smith, R., Prystowsky, M.B. & Childs, G. 2002. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Research.* 62(4): 1184-1190.

Benamouzig, R., Pigot, F., Quiroga, G., Validire, P., Chaussade, S., Catalan, F. & Couturier, D. 1992. Human papillomavirus infection in esophageal squamous-cell carcinoma in western countries. *International Journal of Cancer.* 50(4): 549-552.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L. & Bignell, H.R. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 456(7218): 53-59.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J. & Urashima, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature.* 463(7283): 899-905.

Berrino, F., Sant, M., Verdecchia, A., Capocaccia, R., Hakulinen, T., Est ve , J. & World Health Organization 1995. *Survival of cancer patients in Europe: the Eurocare study.* (IARC Scientific Publication No 132). Lyon, France: IARC Scientific Publications.

Bhat, S., Curach, N., Mostyn, T., Bains, G.S., Griffiths, K.R. & Emslie, K.R. 2010. Comparison of methods for accurate quantification of DNA mass concentration with traceability to the international system of units. *Analytical Chemistry.* 82(17): 7185-7192.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M. et al. 2001. Classification of human lung carcinomas

by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America.* 98(24): 13790-13795.

Bingle, L., Brown, N. & Lewis, C. 2002. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *The Journal of Pathology.* 196(3): 254-265.

Bischof, O., Nacerddine, K. & Dejean, A. 2005. Human papillomavirus oncoprotein E7 targets the promyelocytic leukemia protein and circumvents cellular senescence via the Rb and p53 tumor suppressor pathways. *Molecular and Cellular Biology.* 25(3): 1013-1024.

Björklund, M. & Koivunen, E. 2005. Gelatinase-mediated migration and invasion of cancer cells. *Biochimica Et Biophysica Acta (BBA)-Reviews on Cancer.* 1755(1): 37-69.

Blankenberg, D., Kuster, G.V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. & Taylor, J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology.* 19.10. 1-19.10. 21.

Blot, W.J. & McLaughlin, J.K. 1999. The changing epidemiology of esophageal cancer. *Seminars in Oncology.* 26(5 Suppl 15): 2-8.

Bojarczuk, K., Siernicka, M., Dwojak, M., Bobrowicz, M., Pyrzynska, B., Gaj, P., Karp, M., Giannopoulos, K., Efremov, D.G., Fauriat, C. et al. 2014. B-cell receptor pathway inhibitors affect CD20 levels and impair antitumor activity of anti-CD20 monoclonal antibodies. *Leukemia.* 28(5): 1163-1167.

Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England).* 19(2): 185-193.

Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A. & Fusenig, N.E. 1988. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *The Journal of Cell Biology.* 106(3): 761-771.

Boukamp, P., Stanbridge, E.J., Foo, D.Y., Cerutti, P.A. & Fusenig, N.E. 1990. c-Ha-ras oncogene expression in immortalized human keratinocytes (HaCaT) alters growth potential in vivo but lacks correlation with malignancy. *Cancer Research.* 50(9): 2840-2847.

Boyer, S.N., Wazer, D.E. & Band, V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Research.* 56(20): 4620-4624.

Branca, M., Ciotti, M., Santini, D., Bonito, L.D., Benedetto, A., Giorgi, C., Paba, P., Favalli, C., Costa, S., Agarossi, A. et al. 2004. Activation of the ERK/MAP kinase pathway in cervical intraepithelial neoplasia is related to grade of the lesion but not to high-risk human papillomavirus, virus clearance, or prognosis in cervical cancer. *American Journal of Clinical Pathology.* 122(6): 902-911.

Branca, M., Giorgi, C., Ciotti, M., Santini, D., Di Bonito, L., Costa, S., Benedetto, A., Bonifacio, D., Di Bonito, P., Paba, P. et al. 2006. Upregulation of nuclear factor-kappaB (NF-kappaB) is related to the grade of cervical intraepithelial neoplasia, but is not an independent predictor of high-risk human papillomavirus or disease outcome in cervical cancer. *Diagnostic Cytopathology.* 34(8): 555-563.

Brenna, S.M.F. & Syrjänen, K.J. 2003. Regulation of cell cycles is of key importance in human papillomavirus (HPV)-associated cervical carcinogenesis. *São Paulo Medical Journal.* 121(3): 128-132.

Breuss, J.M., Gallo, J., DeLisser, H.M., Klimanskaya, I.V., Folkesson, H.G., Pittet, J.F., Nishimura, S.L., Aldape, K., Landers, D.V. & Carpenter, W. 1995. Expression of the beta 6 integrin subunit in development, neoplasia and tissue repair suggests a role in epithelial remodeling. *Journal of Cell Science.* 108(6): 2241-2251.

Brimer, N., Lyons, C. & Vande Pol, S.B. 2007. Association of E6AP (UBE3A) with human papillomavirus type 11 E6 protein. *Virology.* 358(2): 303-310.

Brown, L.M. & Devesa, S.S. 2002. Epidemiologic trends in esophageal and gastric cancer in the United States. *Surgical Oncology Clinics of North America.* 11(2): 235-256.

Burch, C.L., Danaher, R.J. & Stein, D.C. 1997. Antigenic variation in Neisseria gonorrhoeae: production of multiple lipooligosaccharides. *Journal of Bacteriology.* 179(3): 982-986.

Burmeister, T., Ebert, A.D., Pritze, W., Loddenkemper, C., Schwartz, S. & Thiel, E. 2004. Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls. *AIDS Research & Human Retroviruses.* 20(11): 1223-1229.

Butel, J.S. 2000. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. *Carcinogenesis.* 21(3): 405-426.

Buttgereit, P., Weineck, S., Ropke, G., Marten, A., Brand, K., Heinicke, T., Caselmann, W.H., Huhn, D. & Schmidt-Wolf, I.G. 2000. Efficient gene transfer into lymphoma cells using adenoviral vectors combined with lipofection. *Cancer Gene Therapy.* 7(8): 1145-1155.

Campo, M.S. 1995. Infection by bovine papillomavirus and prospects for vaccination. *Trends in Microbiology.* 3(3): 92-97.

Cancer Genome Atlas Network 2012. Comprehensive molecular portraits of human breast tumours. *Nature.* 490(7418): 61-70.

Cancer Genome Atlas Research Network 2012. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 489(7417): 519-525.

Cancer Genome Atlas Research Network 2013. Integrated genomic characterization of endometrial carcinoma. *Nature.* 497(7447): 67-73.

Capiod, T., Shuba, Y., Skryma, R. & Prevarskaya, N. 2008. Calcium signalling and cancer cell growth. In *Calcium Signalling and Disease.* Springer. 405-427.

Chang, F., Shen, Q., Zhou, J., Wang, C., Wang, D., Syrjänen, S. & Syrjänen, K. 1990. Detection of human papillomavirus DNA in cytologic specimens derived from esophageal precancer lesions and cancer. *Scandinavian Journal of Gastroenterology.* 25(4): 383-388.

Chang, F. 1990. Role of papillomaviruses. *Journal of Clinical Pathology.* 43(4): 269-276.

Chang, F., Syrjanen, S., Wang, L. & Syrjanen, K. 1992. Infectious agents in the etiology of esophageal cancer. *Gastroenterology.* 103(4): 1336-1348.

Chen, J.J., Reid, C.E., Band, V. & Androphy, E.J. 1995. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. *Science (New York, N.Y.).* 269(5223): 529-531.

Chen, T.M., Pecoraro, G. & Defendi, V. 1993. Genetic analysis of in vitro progression of human papillomavirus-transfected human cervical cells. *Cancer Research.* 53(5): 1167-1171.

Chen, Y., Yao, H., Thompson, E.J., Tannir, N.M., Weinstein, J.N. & Su, X. 2013. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics (Oxford, England).* 29(2): 266-267.

Cheng, Y.W., Wu, M.F., Wang, J., Yeh, K.T., Goan, Y.G., Chiou, H.L., Chen, C.Y. & Lee, H. 2007. Human papillomavirus 16/18 E6 oncoprotein is expressed in lung cancer and related with p53 inactivation. *Cancer Research.* 67(22): 10686-10693.

Chin, L. & Gray, J.W. 2008. Translating insights from the cancer genome into clinical practice. *Nature.* 452(7187): 553-563.

Chiu, C.Y. 2013. Viral pathogen discovery. *Current Opinion in Microbiology.* 16(4): 468-478.

Cho, N.H., Kim, Y.T. & Kim, J.W. 2002. Alteration of cell cycle in cervical tumor associated with human papillomavirus: cyclin-dependent kinase inhibitors. *Yonsei Medical Journal.* 43(6): 722-728.

Chou, C.F., Lai, C.L., Chang, Y.C., Duester, G. & Yin, S.J. 2002. Kinetic mechanism of human class IV alcohol dehydrogenase functioning as retinol dehydrogenase. *The Journal of Biological Chemistry.* 277(28): 25209-25216.

Cimino, P.J., Zhao, G., Wang, D., Sehn, J.K., Lewis Jr, J.S. & Duncavage, E.J. 2014. Detection of viral pathogens in high grade gliomas from unmapped next-generation sequencing data. *Experimental and Molecular Pathology.* 96(3): 310-315.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N. & Sander, C. 2013. Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics.* 45(10): 1127-1133.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K. & Lander, E.S. 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences of the United States of America.* 104(49): 19428-19433.

Clontech Expression 2012. *Adeno-X™ Expression System 1 User Manual.* Available: http://www.clontech.com/ZA/Products/Viral_Transduction/Adenoviral_Vector_Systems/Adenoviral_Expression_System_1?sitex=10023:22372:US.

Clontech Purification 2013. *Adeno-X™ Maxi Purification Kit User Manual.* Available: http://www.clontech.com/ZA/Products/Viral_Transduction/Adenoviral_Transduction_Tools/Adenovirus_Purification?sitex=10023:22372:US.

Contreras-Galindo, R., Kaplan, M.H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., Giusti, F., Dosik, M.H., Hayes, D.F., Gitlin, S.D. et al. 2008. Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *Journal of Virology.* 82(19): 9329-9336.

Cooper, K., Taylor, L. & Govind, S. 1995. Human papillomavirus DNA in oesophageal carcinomas in South Africa. *The Journal of Pathology.* 175(3): 273-277.

Cooper, B., Brimer, N. & Vande Pol, S.B. 2007. Human papillomavirus E6 regulates the cytoskeleton dynamics of keratinocytes through targeted degradation of p53. *Journal of Virology.* 81(22): 12675-12679.

Coussens, L.M. & Werb, Z. 2002. Inflammation and cancer. *Nature.* 420(6917): 860-867.

Cox, L.S. 1997. Multiple pathways control cell growth and transformation: overlapping and independent activities of p53 and p21Cip1/WAF1/Sdi1. *The Journal of Pathology.* 183(2): 134-140.

Crespi, M., Grassi, A., Amiri, G., Munoz, N., Aramesh, B., Mojtabai, A. & Casale, V. 1979. Oesophageal lesions in northern Iran: a premalignant condition? *The Lancet.* 314(8136): 217-221.

Crew, K.D. & Neugut, A.I. 2004. Epidemiology of upper gastrointestinal malignancies. *Seminar in Oncology.* 31(4): 450-464.

Crook, T., Vousden, K.H. & Tidy, J.A. 1991. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. *Cell.* 67(3): 547-556.

Cui, M.C., Li, Y., He, X., Wang, X.L., Wang, L.D. & Liu, H.T. 2011. Study of human papillomavirus in biopsy tissue specimens of esophageal carcinomas in Linzhou city. *Chinese Journal of Experimental and Clinical Virology.* 25(1): 39-41.

D'Costa, Z.J., Leong, C., Shields, J., Matthews, C. & Hibma, M.H. 2012. Screening of drugs to counteract human papillomavirus 16 E6 repression of E-cadherin expression. *Investigational New Drugs.* 30(6): 2236-2251.

Dalal, S., Gao, Q., Androphy, E.J. & Band, V. 1996. Mutational analysis of human papillomavirus type 16 E6 demonstrates that p53 degradation is necessary for immortalization of mammary epithelial cells. *Journal of Virology.* 70(2): 683-688.

Daly, G.M., Bexfield, N., Heaney, J., Stubbs, S., Mayer, A.P., Palser, A., Kellam, P., Drou, N., Caccamo, M. & Tiley, L. 2011. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS One.* 6(12): e28879.

Davies, H., Hunter, C., Smith, R., Stephens, P., Greenman, C., Bignell, G., Teague, J., Butler, A., Edkins, S., Stevens, C. et al. 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Research.* 65(17): 7591-7595.

Davy, C. & Doorbar, J. 2007. G2/M cell cycle arrest in the life cycle of viruses. *Virology.* 368(2): 219-226.

De Villiers, E., Fauquet, C., Broker, T.R., Bernard, H. & zur Hausen, H. 2004. Classification of papillomaviruses. *Virology.* 324(1): 17-27.

DeBoever, C., Reid, E.G., Smith, E.N., Wang, X., Dumaop, W., Harismendy, O., Carson, D., Richman, D., Masliah, E. & Frazer, K.A. 2013. Whole Transcriptome Sequencing Enables Discovery and Analysis of Viruses in Archived Primary Central Nervous System Lymphomas. *PloS One.* 8(9): e73956.

DeFilippis, R.A., Goodwin, E.C., Wu, L. & DiMaio, D. 2003. Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *Journal of Virology.* 77(2): 1551-1563.

Degenhardt, Y.Y. & Silverstein, S. 2001a. Interaction of zyxin, a focal adhesion protein, with the e6 protein from human papillomavirus type 6 results in its nuclear translocation. *Journal of Virology.* 75(23): 11791-11802.

Degenhardt, Y.Y. & Silverstein, S.J. 2001b. Gps2, a protein partner for human papillomavirus E6 proteins. *Journal of Virology.* 75(1): 151-160.

Denekamp, J. 1986. Cell kinetics and radiation biology. *International Journal of Radiation Biology.* 49(2): 357-380.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A. & Hanna, M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics.* 43(5): 491-498.

Deryugina, E.I. & Quigley, J.P. 2006. Matrix metalloproteinases and tumor metastasis. *Cancer and Metastasis Reviews.* 25(1): 9-34.

Deyrieux, A.F. & Wilson, V.G. 2007. In vitro culture conditions to study keratinocyte differentiation using the HaCaT cell line. *Cytotechnology.* 54(2): 77-83.

Donahue, S.L., Lin, Q., Cao, S. & Ruley, H.E. 2006. Carcinogens induce genome-wide loss of heterozygosity in normal stem cells without persistent chromosomal instability. *Proceedings of the National Academy of Sciences of the United States of America.* 103(31): 11642-11646.

Dong, L.M., Potter, J.D., White, E., Ulrich, C.M., Cardon, L.R. & Peters, U. 2008. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *Jama.* 299(20): 2423-2436.

Doniger, J., Muralidhar, S. & Rosenthal, L.J. 1999. Human cytomegalovirus and human herpesvirus 6 genes that transform and transactivate. *Clinical Microbiology Reviews.* 12(3): 367-382.

Downey, R.F., Sullivan, F.J., Wang-Johanning, F., Ambs, S., Giles, F.J. & Glynn, S.A. 2014. Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? *International Journal of Cancer.*

Druesne-Pecollo, N., Tehard, B., Mallet, Y., Gerber, M., Norat, T., Hercberg, S. & Latino-Martel, P. 2009. Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. *The Lancet Oncology.* 10(2): 173-180.

Druker, B.J., Sawyers, C.L., Kantarjian, H., Resta, D.J., Reese, S.F., Ford, J.M., Capdeville, R. & Talpaz, M. 2001a. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *New England Journal of Medicine.* 344(14): 1038-1042.

Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R. & Ohno-Jones, S. 2001b. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *New England Journal of Medicine.* 344(14): 1031-1037.

Duensing, S., Duensing, A., Crum, C.P. & Munger, K. 2001. Human papillomavirus type 16 E7 oncoprotein-induced abnormal centrosome synthesis is an early event in the evolving malignant phenotype. *Cancer Research.* 61(6): 2356-2360.

Duensing, S. & Munger, K. 2002. The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability. *Cancer Research.* 62(23): 7075-7082.

Dulak, A.M., Stojanov, P., Peng, S., Lawrence, M.S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S.E. & Shefler, E. 2013. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genetics.* 45(5): 478-486.

Duncavage, E.J., Magrini, V., Becker, N., Armstrong, J.R., Demeter, R.T., Wylie, T., Abel, H.J. & Pfeifer, J.D. 2011. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *The Journal of Molecular Diagnostics.* 13(3): 325-333.

Dyson, N., Howley, P.M., Munger, K. & Harlow, E. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science (New York, N.Y.).* 243(4893): 934-937.

El Mjiyad, N., Caro-Maldonado, A., Ramirez-Peinado, S. & Munoz-Pinedo, C. 2011. Sugar-free approaches to cancer cell killing. *Oncogene.* 30(3): 253-264.

Elstrom, R.L., Bauer, D.E., Buzzai, M., Karnauskas, R., Harris, M.H., Plas, D.R., Zhuang, H., Cinalli, R.M., Alavi, A., Rudin, C.M. et al. 2004. Akt stimulates aerobic glycolysis in cancer cells. *Cancer Research.* 64(11): 3892-3899.

ENCODE Project Consortium 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.).* 306(5696): 636-640.

Enzinger, P.C. & Mayer, R.J. 2003. Esophageal cancer. *New England Journal of Medicine.* 349(23): 2241-2252.

Eslick, G.D. 2009. Epidemiology of esophageal cancer. *Gastroenterology Clinics of North America.* 38(1): 17-25.

Fasbender, A., Zabner, J., Chillon, M., Moninger, T.O., Puga, A.P., Davidson, B.L. & Welsh, M.J. 1997. Complexes of adenovirus with polycationic polymers and cationic lipids increase the efficiency of gene transfer in vitro and in vivo. *The Journal of Biological Chemistry.* 272(10): 6479-6489.

Fearon, E.R. & Vogelstein, B. 1990. A genetic model for colorectal tumorigenesis. *Cell.* 61(5): 759-767.

Feitelson, M.A. & Lee, J. 2007. Hepatitis B virus integration, fragile sites, and hepatocarcinogenesis. *Cancer Letters.* 252(2): 157-170.

Feng, H., Shuda, M., Chang, Y. & Moore, P.S. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science (New York, N.Y.).* 319(5866): 1096-1100.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. & Bray, F. 2014. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer.*

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. & Parkin, D.M. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer.Journal International Du Cancer.* 127(12): 2893-2917.

Ferro, P. & Pfeffer, U. 2001. Are there CAG repeat expansion-related disorders outside the central nervous system? *Brain Research Bulletin.* 56(3): 259-264.

Feuk, L., Carson, A.R. & Scherer, S.W. 2006. Structural variation in the human genome. *Nature Reviews Genetics.* 7(2): 85-97.

Filippova, M., Parkhurst, L. & Duerksen-Hughes, P.J. 2004. The human papillomavirus 16 E6 protein binds to Fas-associated death domain and protects cells from Fas-triggered apoptosis. *The Journal of Biological Chemistry.* 279(24): 25729-25744.

Filippova, M., Song, H., Connolly, J.L., Dermody, T.S. & Duerksen-Hughes, P.J. 2002. The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. *The Journal of Biological Chemistry.* 277(24): 21730-21739.

Flamand, L., Komaroff, A.L., Arbuckle, J.H., Medveczky, P.G. & Ablashi, D.V. 2010. Review, part 1: Human herpesvirus-6-basic biology, diagnostic testing, and antiviral efficacy. *Journal of Medical Virology.* 82(9): 1560-1568.

Foster, S.A., Demers, G.W., Etscheid, B.G. & Galloway, D.A. 1994. The ability of human papillomavirus E6 proteins to target p53 for degradation in vivo correlates with their ability to abrogate actinomycin D-induced growth arrest. *Journal of Virology.* 68(9): 5698-5705.

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M. & Miya, F. 2012. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics.* 44(7): 760-764.

Fujita, A., Sato, J.R., Rodrigues Lde, O., Ferreira, C.E. & Sogayar, M.C. 2006. Evaluating different methods of microarray data normalization. *BMC Bioinformatics.* 7:469.

Fukuchi, M., Sakurai, S., Suzuki, M., Naitoh, H., Tabe, Y., Fukasawa, T., Kiriyama, S., Yokobori, T. & Kuwano, H. 2011. Esophageal squamous cell carcinoma with marked eosinophil infiltration. *Case Reports in Gastroenterology.* 5(3): 648-653.

Fulda, S., Gorman, A.M., Hori, O. & Samali, A. 2010. Cellular stress responses: cell survival and cell death. *International Journal of Cell Biology.* 2010(2010): 214074.

Funk, J.O., Waga, S., Harry, J.B., Espling, E., Stillman, B. & Galloway, D.A. 1997. Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes & Development.* 11(16): 2090-2100.

Galloway, D.A. & McDougoll, J.K. 1989. Human papillomaviruses and carcinomas. *Advances in Virus Research.* 37:125-171.

Gamliel, Z. 2000. Incidence, epidemiology, and etiology of esophageal cancer. *Chest Surgery Clinics of North America.* 10(3): 441-450.

Gao, Y., Chen, Z., Li, J., Hu, X., Shi, X., Sun, Z., Zhang, F., Zhao, Z., Li, Z. & Liu, Z. 2014. Genetic landscape of esophageal squamous cell carcinoma [in press]. *Nature Genetics.*

Gao, Q., Srinivasan, S., Boyer, S.N., Wazer, D.E. & Band, V. 1999. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. *Molecular and Cellular Biology.* 19(1): 733-744.

Garavello, W., Negri, E., Talamini, R., Levi, F., Zambon, P., Dal Maso, L., Bosetti, C., Franceschi, S. & La Vecchia, C. 2005. Family history of cancer, its combination with smoking and drinking, and risk of squamous cell carcinoma of the esophagus. *Cancer Epidemiology, Biomarkers & Prevention.* 14(6): 1390-1393.

Garber, M.E., Troyanskaya, O.G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G.D., Perou, C.M., Whyte, R.I. et al. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences of the United States of America.* 98(24): 13784-13789.

García-Escudero, R. & Paramio, J.M. 2008. Gene expression profiling as a tool for basic analysis and clinical application of human cancer. *Molecular Carcinogenesis.* 47(8): 573-579.

Gardiol, D., Galizzi, S. & Banks, L. 2002. Mutational analysis of the discs large tumour suppressor identifies domains responsible for human papillomavirus type 18 E6-mediated degradation. *The Journal of General Virology.* 83(Pt 2): 283-289.

Gardiol, D., Kuhne, C., Glaunsinger, B., Lee, S.S., Javier, R. & Banks, L. 1999. Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome-mediated degradation. *Oncogene.* 18(40): 5487-5496.

Garnett, T., Filippova, M. & Duerksen-Hughes, P. 2006. Accelerated degradation of FADD and procaspase 8 in cells expressing human papilloma virus 16 E6 impairs TRAIL-mediated apoptosis. *Cell Death & Differentiation.* 13(11): 1915-1926.

Garraway, L.A. 2013. Genomics-driven oncology: framework for an emerging paradigm. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* 31(15): 1806-1814.

Garraway, L.A., Verweij, J. & Ballman, K.V. 2013. Precision oncology: an overview. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* 31(15): 1803-1805.

Gates, O. & Warren, S. 1968. Radiation-induced experimental cancer of the esophagus. *The American Journal of Pathology.* 53(5): 667-685.

Gewin, L. & Galloway, D.A. 2001. E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. *Journal of Virology.* 75(15): 7198-7201.

Giannoudis, A. & Herrington, C.S. 2000. Differential expression of p53 and p21 in low grade cervical squamous intraepithelial lesions infected with low, intermediate, and high risk human papillomaviruses. *Cancer.* 89(6): 1300-1307.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J. et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research.* 15(10): 1451-1455.

Gifford, R. & Tristem, M. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes.* 26(3): 291-315.

Glaunsinger, B.A., Lee, S.S., Thomas, M., Banks, L. & Javier, R. 2000. Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. *Oncogene.* 19(46): 5270-5280.

Goecks, J., Nekrutenko, A. & Taylor, J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11(8): R86.

Golan, M., Hizi, A., Resau, J.H., Yaal-Hahoshen, N., Reichman, H., Keydar, I. & Tsarfaty, I. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia.* 10(6): 521-IN2.

Graham, S.V. 2008. Papillomavirus 3' UTR regulatory elements. *Frontiers in Bioscience : A Journal and Virtual Library.* 135646-5663.

Greenhough, A., Smartt, H.J., Moore, A.E., Roberts, H.R., Williams, A.C., Paraskeva, C. & Kaidi, A. 2009. The COX-2/PGE2 pathway: key roles in the hallmarks of cancer and adaptation to the tumour microenvironment. *Carcinogenesis.* 30(3): 377-386.

Gregory, T.R. 2004. Insertion–deletion biases and the evolution of genome size. *Gene.* 324:15-34.

Griep, A.E., Herber, R., Jeon, S., Lohse, J.K., Dubielzig, R.R. & Lambert, P.F. 1993. Tumorigenicity by human papillomavirus type 16 E6 and E7 in transgenic mice correlates with alterations in epithelial cell growth and differentiation. *Journal of Virology.* 67(3): 1373-1384.

Griffin, J.L. & Shockcor, J.P. 2004. Metabolic profiles of cancer cells. *Nature Reviews Cancer.* 4(7): 551-561.

Grossman, S.R., Mora, R. & Laimins, L.A. 1989. Intracellular localization and DNA-binding properties of human papillomavirus type 18 E6 protein expressed with a baculovirus vector. *Journal of Virology.* 63(1): 366-374.

Gu, Z., Li, J., Li, M., Gu, J., Shi, X., Ke, Y. & Chen, K. 2005. Matrix metalloproteinases expression correlates with survival in patients with esophageal squamous cell carcinoma. *American Journal of Gastroenterology.* 100(8): 1835-1843.

Guccione, E., Massimi, P., Bernat, A. & Banks, L. 2002. Comparative analysis of the intracellular location of the high- and low-risk human papillomavirus oncoproteins. *Virology.* 293(1): 20-25.

Guo, M., Liu, S., Herman, J.G., Zhuang, H. & Lu, F. 2006. Brief Communication Gefitinib-Sensitizing Mutation in Esophageal Carcinoma Cell Line Kyse450. *Cancer Biology & Therapy.* 5(2): 152-155.

Guo, W. & Giancotti, F.G. 2004. Integrin signalling during tumour progression. *Nature Reviews Molecular Cell Biology.* 5(10): 816-826.

Guo, Y., Zhao, S., Sheng, Q., Ye, F., Li, J., Lehmann, B., Pietenpol, J., Samuels, D.C. & Shyr, Y. 2014. Multi-perspective quality control of Illumina exome sequencing data using QC3. *Genomics.* 103(5-6): 323-328.

Guo, Q.M. 2003. DNA microarray and cancer. *Current Opinion in Oncology.* 15(1): 36-43.

Guo, Y., Li, J., Li, C.I., Shyr, Y. & Samuels, D.C. 2013a. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics (Oxford, England).* 29(9): 1210-1211.

Guo, Y., Ye, F., Sheng, Q., Clark, T. & Samuels, D.C. 2013b. Three-stage quality control strategies for DNA re-sequencing data [in press]. *Briefings in Bioinformatics.*

Gustafson, S., Proper, J.A., Bowie, E. & Sommer, S.S. 1987. Parameters affecting the yield of DNA from human blood. *Analytical Biochemistry.* 165(2): 294-299.

Hagari, Y., Budgeon, L.R., Pickel, M.D. & Kreider, J.W. 1995. Association of tumor necrosis factor-alpha gene expression and apoptotic cell death with regression of Shope papillomas. *The Journal of Investigative Dermatology.* 104(4): 526-529.

Halbert, C.L., Demers, G.W. & Galloway, D.A. 1991. The E7 gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. *Journal of Virology.* 65(1): 473-478.

Hale, M.J., Liptz, T.R. & Paterson, A.C. 1989. Association between human papillomavirus and carcinoma of the oesophagus in South African blacks. A histochemical and immunohistochemical study. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde.* 76(7): 329-330.

Hall, C.B., Caserta, M.T., Schnabel, K., Shelley, L.M., Marino, A.S., Carnahan, J.A., Yoo, C., Lofthus, G.K. & McDermott, M.P. 2008. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. *Pediatrics.* 122(3): 513-520.

Hamada, H., Tashima, Y., Kisaka, Y., Iwamoto, K., Hanai, T., Eguchi, Y. & Okamoto, M. 2009. Sophisticated framework between cell cycle arrest and apoptosis induction based on p53 dynamics. *PLoS One.* 4(3): e4795.

Hampras, S.S., Giuliano, A.R., Lin, H.Y., Fisher, K.J., Abrahamsen, M.E., McKay-Chopin, S., Gheit, T., Tommasino, M. & Rollison, D.E. 2014. Natural history of polyomaviruses in men: The HIM Study. *The Journal of Infectious Diseases.*

Han, C., Liao, C., Wu, C., Hwong, C., Lee, A. & Yin, S. 1998. Contribution to first-pass metabolism of ethanol and inhibition by ethanol for retinol oxidation in human alcohol dehydrogenase family. *European Journal of Biochemistry.* 254(1): 25-31.

Han, H., Bearss, D.J., Browne, L.W., Calaluce, R., Nagle, R.B. & Von Hoff, D.D. 2002. Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. *Cancer Research.* 62(10): 2890-2896.

Hanahan, D. & Weinberg, R.A. 2011. Hallmarks of cancer: the next generation. *Cell.* 144(5): 646-674.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology.* 166(4): 557-580.

Hanawa, M., Suzuki, S., Dobashi, Y., Yamane, T., Kono, K., Enomoto, N. & Ooi, A. 2006. EGFR protein overexpression and gene amplification in squamous cell carcinomas of the esophagus. *International Journal of Cancer.* 118(5): 1173-1180.

Hanke, K., Chudak, C., Kurth, R. & Bannert, N. 2013. The Rec protein of HERV-K (HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). *International Journal of Cancer.* 132(3): 556-567.

Harwood, C.A. & Proby, C.M. 2002. Human papillomaviruses and non-melanoma skin cancer. *Current Opinion in Infectious Diseases.* 15(2): 101-114.

Hashida, T. & Yasumoto, S. 1991. Induction of chromosome abnormalities in mouse and human epidermal keratinocytes by the human papillomavirus type 16 E7 oncogene. *The Journal of General Virology.* 72(7): 1569-1577.

Havard, L., Rahmouni, S., Boniver, J. & Delvenne, P. 2005. High levels of p105 (NFKB1) and p100 (NFKB2) proteins in HPV16-transformed keratinocytes: role of E6 and E7 oncoproteins. *Virology.* 331(2): 357-366.

Havre, P.A., Yuan, J., Hedrick, L., Cho, K.R. & Glazer, P.M. 1995. p53 inactivation by HPV16 E6 results in increased mutagenesis in human cells. *Cancer Research.* 55(19): 4420-4424.

Hawley-Nelson, P., Vousden, K.H., Hubbert, N.L., Lowy, D.R. & Schiller, J.T. 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *The EMBO Journal.* 8(12): 3905-3910.

Hebner, C.M., Wilson, R., Rader, J., Bidder, M. & Laimins, L.A. 2006. Human papillomaviruses target the double-stranded RNA protein kinase pathway. *The Journal of General Virology.* 87(11): 3183-3193.

Heck, D.V., Yee, C.L., Howley, P.M. & Munger, K. 1992. Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses. *Proceedings of the National Academy of Sciences of the United States of America.* 89(10): 4442-4446.

Hendricks, D. & Parker, M.I. 2002. Oesophageal cancer in Africa. *IUBMB Life.* 53(4-5): 263-268.

Hille, J., MARGOLlUS, K.A., Markowitz, S. & Isaacson, C. 1986. Human papillomavirus infection to oesophageal carcinoma in South Africans. *SAMJ.* 69417.

Hirasaki, S., Noguchi, T., Mimori, K., Onuki, J., Morita, K., Inoue, H., Sugihara, K., Mori, M. & Hirano, T. 2007. BAC clones related to prognosis in patients with esophageal squamous carcinoma: an array comparative genomic hybridization study. *The Oncologist.* 12(4): 406-417.

Hiyama, T., Yoshihara, M., Tanaka, S. & Chayama, K. 2007. Genetic polymorphisms and esophageal cancer risk. *International Journal of Cancer.* 121(8): 1643-1658.

Holland, J.F. & Pogo, B.G. 2004. Mouse mammary tumor virus-like viral infection and human breast cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* 10(17): 5647-5649.

Holmes, R.S. & Vaughan, T.L. 2007. Epidemiology and pathogenesis of esophageal cancer. 17(1): 2-9.

Houben, R., Shuda, M., Weinkam, R., Schrama, D., Feng, H., Chang, Y., Moore, P.S. & Becker, J.C. 2010. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. *Journal of Virology.* 84(14): 7064-7072.

Howie, H.L., Katzenellenbogen, R.A. & Galloway, D.A. 2009. Papillomavirus E6 proteins. *Virology.* 384(2): 324-334.

Hu, D. & Goldie, S. 2008. The economic burden of noncervical human papillomavirus disease in the United States. *American Journal of Obstetrics and Gynecology.* 198(5): 500. e1-500. e7.

Hu, N., Clifford, R.J., Yang, H.H., Wang, C., Goldstein, A.M., Ding, T., Taylor, P.R. & Lee, M.P. 2010. Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics.* 11(1): 576.

Hu, N., Wang, C., Ng, D., Clifford, R., Yang, H.H., Tang, Z.Z., Wang, Q.H., Han, X.Y., Giffen, C., Goldstein, A.M. et al. 2009. Genomic characterization of esophageal squamous cell carcinoma from a high-risk population in China. *Cancer Research.* 69(14): 5908-5917.

Huang, L., Seow, K., Lee, C., Lin, Y., Pan, H. & Chen, H. 2010. Decreased p21 expression in HPV-18 positive cervical carcinomas. *Pathology & Oncology Research.* 16(1): 81-86.

Hubbert, N.L., Sedman, S.A. & Schiller, J.T. 1992. Human papillomavirus type 16 E6 increases the degradation rate of p53 in human keratinocytes. *Journal of Virology.* 66(10): 6237-6241.

Huber, M.A., Beug, H. & Wirth, T. 2004. Epithelial-mesenchymal transition: NF-kappaB takes center stage. *Cell Cycle.* 3(12): 1477-1480.

Hudson, M.E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources.* 8(1): 3-17.

Hudson, J.B., Bedell, M.A., McCance, D.J. & Laiminis, L.A. 1990. Immortalization and altered differentiation of human keratinocytes in vitro by the E6 and E7 open reading frames of human papillomavirus type 18. *Journal of Virology.* 64(2): 519-526.

Huibregtse, J.M. & Beaudenon, S.L. 1996. Mechanism of HPV E6 proteins in cellular transformation. 7(6): 317-326.

Huibregtse, J.M., Scheffner, M. & Howley, P.M. 1991. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *The EMBO Journal.* 10(13): 4129-4135.

Huibregtse, J.M., Scheffner, M. & Howley, P.M. 1993a. Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53. *Molecular and Cellular Biology.* 13(2): 775-784.

Huibregtse, J.M., Scheffner, M. & Howley, P.M. 1993b. Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. *Molecular and Cellular Biology.* 13(8): 4918-4927.

hul Park, J.C. & Moon, C. 2008. Gene therapy for esophageal cancer. *Cancer Therapy.* 635-46.

Hussain, I., Fathallah, I., Accardi, R., Yue, J., Saidj, D., Shukla, R., Hasan, U., Gheit, T., Niu, Y., Tommasino, M. et al. 2011. NF-kappaB protects human papillomavirus type 38 E6/E7-immortalized human keratinocytes against tumor necrosis factor alpha and UV-mediated apoptosis. *Journal of Virology.* 85(17): 9013-9022.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. & Lee, C. 2004. Detection of large-scale variation in the human genome. *Nature Genetics.* 36(9): 949-951.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans 1995. Human papillomaviruses. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans / World Health Organization, International Agency for Research on Cancer.* 641-378.

Imai, T., Matsuda, K., Shimojima, T., Hashimoto, T., Masuhiro, Y., Kitamoto, T., Sugita, A., Suzuki, K., Matsumoto, H., Masushige, S. et al. 1997. ERC-55, a binding protein for the papilloma virus E6 oncoprotein, specifically interacts with vitamin D receptor among nuclear receptors. *Biochemical and Biophysical Research Communications.* 233(3): 765-769.

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M. & Sivachenko, A. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell.* 150(6): 1107-1120.

Ingenuity 2014. *Ingenuity Pathway Analysis.* Available: http://www.ingenuity.com/products/ipa.

Inoue, H., Nojima, H. & Okayama, H. 1990. High efficiency transformation of Escherichia coli with plasmids. *Gene.* 96(1): 23-28.

Inoue, K., Slaton, J.W., Eve, B.Y., Kim, S.J., Perrotte, P., Balbay, M.D., Yano, S., Bar-Eli, M., Radinsky, R., Pettaway, C.A. et al. 2000. Interleukin 8 expression regulates tumorigenicity and metastases in androgen-independent prostate cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* 6(5): 2104-2119.

International HapMap Consortium 2005. A haplotype map of the human genome. *Nature.* 437(7063): 1299-1320.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. & Speed, T.P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England).* 4(2): 249-264.

Ishibashi, Y., Matsumoto, T., Niwa, M., Suzuki, Y., Omura, N., Hanyu, N., Nakada, K., Yanaga, K., Yamada, K. & Ohkawa, K. 2004. CD147 and matrix metalloproteinase-2 protein expression as significant prognostic factors in esophageal squamous cell carcinoma. *Cancer.* 101(9): 1994-2000.

Jabbar, S.F., Park, S., Schweizer, J., Berard-Bergery, M., Pitot, H.C., Lee, D. & Lambert, P.F. 2012. Cervical cancers require the continuous expression of the human papillomavirus type 16 E7 oncoprotein even in the presence of the viral E6 oncoprotein. *Cancer Research.* 72(16): 4008-4016.

Jackson, S., Harwood, C., Thomas, M., Banks, L. & Storey, A. 2000. Role of Bak in UV-induced apoptosis in skin cancer and abrogation by HPV E6 proteins. *Genes & Development.* 14(23): 3065-3073.

Jacobsen, F., Hirsch, T., Mittler, D., Schulte, M., Lehnhardt, M., Druecke, D., Homann, H., Steinau, H. & Steinstraesser, L. 2006. Polybrene improves transfection efficacy of recombinant replication-deficient adenovirus in cutaneous cells and burned skin. *The Journal of Gene Medicine.* 8(2): 138-146.

James, M.A., Lee, J.H. & Klingelhutz, A.J. 2006. Human papillomavirus type 16 E6 activates NF-kappaB, induces cIAP-2 expression, and protects against apoptosis in a PDZ binding motif-dependent manner. *Journal of Virology.* 80(11): 5301-5307.

Janssen, J.W., Imoto, I., Inoue, J., Shimada, Y., Ueda, M., Imamura, M., Bartram, C.R. & Inazawa, J. 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): 460-464.

Jansson, C., Plato, N., Johansson, A.L., Nyren, O. & Lagergren, J. 2006. Airborne occupational exposures and risk of oesophageal and cardia adenocarcinoma. *Occupational and Environmental Medicine.* 63(2): 107-112.

Jaskiewicz, K., Venter, F.S. & Marasas, W.F. 1987. Cytopathology of the esophagus in Transkei. *Journal of the National Cancer Institute.* 79(5): 961-967.

Javier, R.T. & Butel, J.S. 2008. The history of tumor virology. *Cancer Research.* 68(19): 7693-7706.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J. & Thun, M.J. 2009. Cancer statistics, 2009. *CA: A Cancer Journal for Clinicians.* 59(4): 225-249.

Jiang, Z., Jhunjhunwala, S., Liu, J., Haverty, P.M., Kennemer, M.I., Guan, Y., Lee, W., Carnevali, P., Stinson, J., Johnson, S. et al. 2012. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Research.* 22(4): 593-601.

Kaczkowski, B., Rossing, M., Andersen, D.K., Dreher, A., Morevati, M., Visser, M.A., Winther, O., Nielsen, F.C. & Norrild, B. 2012. Integrative analyses reveal novel strategies in HPV11,-16 and -45 early infection. *Scientific Reports.* (2): 515.

Kamangar, F., Dores, G.M. & Anderson, W.F. 2006. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* 24(14): 2137-2150.

Kampen, G.T., Stafford, S., Adachi, T., Jinquan, T., Quan, S., Grant, J.A., Skov, P.S., Poulsen, L.K. & Alam, R. 2000. Eotaxin induces degranulation and chemotaxis of eosinophils through the activation of ERK2 and p38 mitogen-activated protein kinases. *Blood.* 95(6): 1911-1917.

Kanda, T., Watanabe, S., Zanma, S., Sato, H., Furuno, A. & Yoshiike, K. 1991. Human papillomavirus type 16 E6 proteins with glycine substitution for cysteine in the metal-binding motif. *Virology.* 185(2): 536-543.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research.* 32(Database issue): D277-80.

Karin, M. 2006. Nuclear factor-kappaB in cancer development and progression. *Nature.* 441(7092): 431-436.

Kasof, G.M., Goyal, L. & White, E. 1999. Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins. *Molecular and Cellular Biology.* 19(6): 4390-4404.

Kato, H., Arao, T., Matsumoto, K., Fujita, Y., Kimura, H., Hayashi, H., Nishiki, K., Iwama, M., Shiraishi, O. & Yasuda, A. 2013. Gene amplification of EGFR, HER2, FGFR2 and MET in esophageal squamous cell carcinoma. *International Journal of Oncology.* 42(4): 1151-1158.

Kawasaki, E.S. 2006. The end of the microarray Tower of Babel: will universal standards lead the way? *Journal of Biomolecular Techniques.* 17(3): 200.

Kay, M.A. & Woo, S.L. 1994. Gene therapy for metabolic disorders. *Trends in Genetics : TIG.* 10(7): 253-257.

Kessis, T.D., Connolly, D.C., Hedrick, L. & Cho, K.R. 1996. Expression of HPV16 E6 or E7 increases integration of foreign DNA. *Oncogene.* 13(2): 427-431.

Kessis, T.D., Slebos, R.J., Nelson, W.G., Kastan, M.B., Plunkett, B.S., Han, S.M., Lorincz, A.T., Hedrick, L. & Cho, K.R. 1993. Human papillomavirus 16 E6 expression disrupts the p53-mediated cellular response to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America.* 90(9): 3988-3992.

Khatri, P., Sirota, M. & Butte, A.J. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Computational Biology.* 8(2): e1002375.

Khoury, J.D., Tannir, N.M., Williams, M.D., Chen, Y., Yao, H., Zhang, J., Thompson, E.J., TCGA Network, Meric-Bernstam, F., Medeiros, L.J. et al. 2013. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *Journal of Virology.* 87(16): 8916-8926.

Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K. & Eichler, E.E. 2010. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell.* 143(5): 837-847.

Kim, Y.T., Sung, S.W. & Kim, J.H. 2001. Is it necessary to resect the diseased esophagus in performing reconstruction for corrosive esophageal stricture? *European Journal of Cardio-Thoracic Surgery.* 20(1): 1-6.

Kiyono, T., Foster, S.A., Koop, J.I., McDougall, J.K., Galloway, D.A. & Klingelhutz, A.J. 1998. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature.* 396(6706): 84-88.

Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, Y., Akiyama, T. & Ishibashi, M. 1997. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proceedings of the National Academy of Sciences of the United States of America.* 94(21): 11612-11616.

Klingelhutz, A.J. & Roman, A. 2012. Cellular transformation by human papillomaviruses: lessons learned by comparing high-and low-risk viruses. *Virology.* 424(2): 77-98.

Klingelhutz, A.J., Barber, S.A., Smith, P.P., Dyer, K. & McDougall, J.K. 1994. Restoration of telomeres in human papillomavirus-immortalized human anogenital epithelial cells. *Molecular and Cellular Biology.* 14(2): 961-969.

Klingelhutz, A.J., Foster, S.A. & McDougall, J.K. 1996. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature.* 380(6569): 79-82.

Kochhar, R., Sethy, P.K., Kochhar, S., Nagi, B. & Gupta, N.M. 2006. Corrosive induced carcinoma of esophagus: report of three patients and review of literature. *Journal of Gastroenterology and Hepatology.* 21(4): 777-780.

Kofman, A., Marcinkiewicz, L., Dupart, E., Lyshchev, A., Martynov, B., Ryndin, A., Kotelevskaya, E., Brown, J., Schiff, D. & Abounader, R. 2011. The roles of viruses in brain tumor initiation and oncomodulation. *Journal of Neuro-Oncology.* 105(3): 451-466.

Kondrashov, A.S. & Rogozin, I.B. 2004. Context of deletions and insertions in human coding sequences. *Human Mutation.* 23(2): 177-185.

Koontongkaew, S. 2013. The tumor microenvironment contribution to development, growth, invasion and metastasis of head and neck squamous cell carcinomas. *Journal of Cancer.* 4(1): 66.

Kozarsky, K.F. & Wilson, J.M. 1993. Gene therapy: adenovirus vectors. *Current Opinion in Genetics & Development.* 3(3): 499-503.

Kremsdorf, D., Soussan, P., Paterlini-Brechot, P. & Brechot, C. 2006. Hepatitis B virus-related hepatocellular carcinoma: paradigms for viral-related human carcinogenesis. *Oncogene.* 25(27): 3823-3833.

Kubbutat, M.H. & Vousden, K.H. 1996. Role of E6 and E7 oncoproteins in HPV-induced anogenital malignancies. 7(5): 295-304.

Kuhne, C. & Banks, L. 1998. E3-ubiquitin ligase/E6-AP links multicopy maintenance protein 7 to the ubiquitination pathway by a novel motif, the L2G box. *The Journal of Biological Chemistry.* 273(51): 34302-34309.

Kukimoto, I., Aihara, S., Yoshiike, K. & Kanda, T. 1998. Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. *Biochemical and Biophysical Research Communications.* 249(1): 258-262.

Kulski, J., Demeter, T., Sterrett, G. & Shilkin, K. 1986. Human papilloma virus DNA in oesophageal carcinoma. *The Lancet.* 328(8508): 683-684.

Kurth, R. & Bannert, N. 2010. Beneficial and detrimental effects of human endogenous retroviruses. *International Journal of Cancer.* 126(2): 306-314.

Kuwano, H., Kato, H., Miyazaki, T., Fukuchi, M., Masuda, N., Nakajima, M., Fukai, Y., Sohda, M., Kimura, H. & Faried, A. 2005. Genetic alterations in esophageal cancer. *Surgery Today.* 35(1): 7-18.

Lace, M.J., Anson, J.R., Klussmann, J.P., Wang, D.H., Smith, E.M., Haugen, T.H. & Turek, L.P. 2011. Human papillomavirus type 16 (HPV-16) genomes integrated in head and neck cancers and in HPV-16-immortalized human keratinocyte clones express chimeric virus-cell mRNAs similar to those found in cervical cancers. *Journal of Virology.* 85(4): 1645-1654.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227(5259): 680-685.

Laimins, L.A. 1993. The biology of human papillomaviruses: from warts to cancer. *Infectious Agents and Disease.* 2(2): 74-86.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M. & FitzHugh, W. 2001. Initial sequencing and analysis of the human genome. *Nature.* 409(6822): 860-921.

Lanuti, M., El Kouri, C., Force, S., Chang, M., Amin, K., Xu, K., Blair, I., Kaiser, L. & Albelda, S. 1999. Use of protamine to augment adenovirus-mediated cancer gene therapy. *Gene Therapy.* 6(9): 1600-1610.

Lao-Sirieix, P., Caldas, C. & Fitzgerald, R.C. 2010. Genetic predisposition to gastro-oesophageal cancer. *Current Opinion in Genetics & Development.* 20(3): 210-217.

Laplante, M. & Sabatini, D.M. 2012. mTOR signaling in growth control and disease. *Cell.* 149(2): 274-293.

Lasithiotaki, I., Antoniou, K., Derdas, S., Sarchianaki, E., Symvoulakis, E., Psaraki, A., Spandidos, D., Stathopoulos, E., Siafakas, N. & Sourvinos, G. 2013. The presence of Merkel cell polyomavirus is associated with deregulated expression of BRAF and Bcl-2 genes in non-small cell lung cancer. *International Journal of Cancer.* 133(3): 604-611.

Lavergne, D. & de Villiers, E. 1999. Papillomavirus in esophageal papillomas and carcinomas. *International Journal of Cancer.* 80(5): 681-684.

Law, F., Chen, Y., Wong, K., Ying, J., Tao, Q., Langford, C., Lee, P., Law, S., Cheung, R. & Chui, C. 2007. Identification of a novel tumor transforming gene GAEC1 at 7q22 which encodes a nuclear protein and is frequently amplified and overexpressed in esophageal squamous cell carcinoma. *Oncogene.* 26(40): 5877-5888.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S. & Getz, G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 505(7484): 495-501.

Lechner, M.S. & Laimins, L.A. 1994. Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *Journal of Virology.* 68(7): 4262-4273.

Lechner, M.S., Mack, D.H., Finicle, A.B., Crook, T., Vousden, K.H. & Laimins, L.A. 1992. Human papillomavirus E6 proteins bind p53 in vivo and abrogate p53-mediated repression of transcription. *The EMBO Journal.* 11(8): 3045-3052.

Lee, A. & Lee, C. 2007. Oncogenesis and transforming viruses: the hepatitis B virus and hepatocellularcarcinoma–the etiopathogenic link. *Front Biosci.* 12(1): 234-245.

Lee, C., Lee, J., Wu, D., Hsu, H., Kao, E., Huang, H., Wang, T., Huang, M. & Wu, M. 2005. Independent and combined effects of alcohol intake, tobacco smoking and betel quid chewing on the risk of esophageal cancer in Taiwan. *International Journal of Cancer.* 113(3): 475-482.

Lee, S., Chiang, C., Lee, S., Hsia, Y., Chuang, T., Lin, J., Liang, S., Nieh, S. & Yin, S. 2006. Immunochemical features in the classification of human alcohol dehydrogenase family. *Alcohol.* 39(1): 13-20.

Lee, W., Jiang, Z., Liu, J., Haverty, P.M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K.P. & Bhatt, D. 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature.* 465(7297): 473-477.

Lee, J.H., Yi, S.M., Anderson, M.E., Berger, K.L., Welsh, M.J., Klingelhutz, A.J. & Ozbun, M.A. 2004. Propagation of infectious human papillomavirus type 16 by using an adenovirus and Cre/LoxP mechanism. *Proceedings of the National Academy of Sciences of the United States of America.* 101(7): 2094-2099.

Lee, S.S., Glaunsinger, B., Mantovani, F., Banks, L. & Javier, R.T. 2000. Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. *Journal of Virology.* 74(20): 9680-9693.

Leichman, L. & Thomas, C.R.,Jr 2011. Squamous cell cancer of the esophagus: the forgotten one. *Gastrointestinal Cancer Research : GCR.* 4(1): 22-23.

Lepage, C., Rachet, B., Jooste, V., Faivre, J. & Coleman, M.P. 2008. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *The American Journal of Gastroenterology.* 103(11): 2694-2699.

Lessard, J.L., Robinson, R.A. & Hoffman, H.T. 2001. Differential expression of ras signal transduction mediators in verrucous and squamous cell carcinomas of the upper aerodigestive tract. *Archives of Pathology & Laboratory Medicine.* 125(9): 1200-1203.

Lewis, S.J. & Smith, G.D. 2005. Alcohol, ALDH2, and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiology, Biomarkers & Prevention.* 14(8): 1967-1971.

Li, Y., Ma, J., Guo, Q., Duan, F., Tang, F., Zheng, P., Zhao, Z. & Lu, G. 2009. Overexpression of MMP-2 and MMP-9 in esophageal squamous cell carcinoma. *Diseases of the Esophagus.* 22(8): 664-667.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* 26(5): 589-595.

Li, L. & Neaves, W.B. 2006. Normal stem cells and cancer stem cells: the niche matters. *Cancer Research.* 66(9): 4553-4557.

Li, Q., Withoff, S. & Verma, I.M. 2005. Inflammation-associated cancer: NF-kappaB is the lynchpin. *Trends in Immunology.* 26(6): 318-325.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K. et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research.* 20(2): 265-272.

Li, S., Labrecque, S., Gauzzi, M.C., Cuddihy, A.R., Wong, A.H., Pellegrini, S., Matlashewski, G.J. & Koromilas, A.E. 1999. The human papilloma virus (HPV)-18 E6 oncoprotein physically associates with Tyk2 and impairs Jak-STAT activation by interferon-alpha. *Oncogene.* 18(42): 5727-5737.

Li, X. & Coffino, P. 1996. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *Journal of Virology.* 70(7): 4509-4516.

Lightfoot, S. 2002. Quantitation comparison of total RNA using the Agilent 2100 bioanalyzer, ribogreen analysis, and UV spectrometry. *Agilent Application Note, Publication Number.*

Lin, D., Hao, J., Nagata, Y., Xu, L., Shang, L., Meng, X., Sato, Y., Okuno, Y., Varela, A.M. & Ding, L. 2014. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nature Genetics.* 46(5): 467-473.

Lindblad, M., Rodríguez, L.A.G. & Lagergren, J. 2005. Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes & Control.* 16(3): 285-294.

Liu, X., Roberts, J., Dakic, A., Zhang, Y. & Schlegel, R. 2008. HPV E7 contributes to the telomerase activity of immortalized and tumorigenic cells and augments E6-induced hTERT promoter function. *Virology.* 375(2): 611-623.

Liu, H., Lu, Z.G., Miki, Y. & Yoshida, K. 2007. Protein kinase C delta induces transcription of the TP53 tumor suppressor gene by controlling death-promoting factor Btf in the apoptotic response to DNA damage. *Molecular and Cellular Biology.* 27(24): 8480-8491.

Liu, Q., Guo, Y., Li, J., Long, J., Zhang, B. & Shyr, Y. 2012. Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. *BMC Genomics.* 13S8.

Liyanage, S.S., Rahman, B., Ridda, I., Newall, A.T., Tabrizi, S.N., Garland, S.M., Segelov, E., Seale, H., Crowe, P.J. & Moa, A. 2013a. The aetiological role of human papillomavirus in oesophageal squamous cell carcinoma: a meta-analysis. *PloS One.* 8(7): e69238.

Liyanage, S.S., Segelov, E., Garland, S.M., Tabrizi, S.N., Seale, H., Crowe, P.J., Dwyer, D.E., Barbour, A., Newall, A.T. & Malik, A. 2013b. Role of human papillomaviruses in esophageal squamous cell carcinoma. *Asia-Pacific Journal of Clinical Oncology.* 9(1): 12-28.

Locksley, R.M., Killeen, N. & Lenardo, M.J. 2001. The TNF and TNF receptor superfamilies-integrating mammalian biology. *Cell.* 104(4): 487-501.

Loda, M., Capodieci, P., Mishra, R., Yao, H., Corless, C., Grigioni, W., Wang, Y., Magi-Galluzzi, C. & Stork, P.J. 1996. Expression of mitogen-activated protein kinase phosphatase-1 in the early phases of human epithelial carcinogenesis. *The American Journal of Pathology.* 149(5): 1553-1564.

Longworth, M.S. & Laimins, L.A. 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiology and Molecular Biology Reviews : MMBR.* 68(2): 362-372.

Lu, H., Ouyang, W. & Huang, C. 2006. Inflammation, a key event in cancer development. *Molecular Cancer Research : MCR.* 4(4): 221-233.

Luo, J., Dunn, T., Ewing, C., Sauvageot, J., Chen, Y., Trent, J. & Isaacs, W. 2002. Gene expression signature of benign prostatic hyperplasia revealed by cDNA microarray analysis. *The Prostate.* 51(3): 189-200.

Luo, M.L., Shen, X.M., Zhang, Y., Wei, F., Xu, X., Cai, Y., Zhang, X., Sun, Y.T., Zhan, Q.M., Wu, M. 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): 11690-11699.

Lusso, P. 2006. HHV-6 and the immune system: mechanisms of immunomodulation and viral escape. *Journal of Clinical Virology.* 37S4-S10.

Lusso, P. & Gallo, R.C. 1995. Human herpesvirus 6 in AIDS. *Immunology Today.* 16(2): 67-71.

Malik, S.M., Nevin, D.T., Cohen, S., Hunt, J.L. & Palazzo, J.P. 2011. Assessment of immunohistochemistry for p16INK4 and high-risk HPV DNA by in situ hybridization in esophageal squamous cell carcinoma. *International Journal of Surgical Pathology.* 19(1): 31-34.

Mallon, R.G., Wojciechowicz, D. & Defendi, V. 1987. DNA-binding activity of papillomavirus proteins. *Journal of Virology.* 61(5): 1655-1660.

Mandard, A., Hainaut, P. & Hollstein, M. 2000. Genetic steps in the development of squamous cell carcinoma of the esophagus. *Mutation Research/Reviews in Mutation Research.* 462(2): 335-342.

Mantovani, F. & Banks, L. 1999. The interaction between p53 and papillomaviruses. 9(6): 387-395.

Mardis, E.R. 2008a. The impact of next-generation sequencing technology on genetics. *Trends in Genetics.* 24(3): 133-141.

Mardis, E.R. 2008b. Next-generation DNA sequencing methods. *Annu.Rev.Genomics Hum.Genet.* 9:387-402.

Marin, M.C., Jost, C.A., Irwin, M.S., DeCaprio, J.A., Caput, D. & Kaelin, W.G.,Jr 1998. Viral oncoproteins discriminate between p53 and the p53 homolog p73. *Molecular and Cellular Biology.* 18(11): 6316-6324.

Martinez, I., Gardiner, A., Board, K., Monzon, F., Edwards, R. & Khan, S. 2008. Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells. *Oncogene.* 27(18): 2575-2582.

Matejcic, M., Li, D., Prescott, N.J., Lewis, C.M., Mathew, C.G. & Parker, M.I. 2011. Association of a deletion of GSTT2B with an altered risk of oesophageal squamous cell carcinoma in a South African population: a case-control study. *PloS One.* 6(12): e29366.

Mathew, R., Khanna, R., Kumar, R., Mathur, M., Shukla, N.K. & Ralhan, R. 2002. Stromelysin-2 overexpression in human esophageal squamous cell carcinoma: potential clinical implications. *Cancer Detection and Prevention.* 26(3): 222-228.

Matsha, T., Donninger, H., Erasmus, R., Hendricks, D., Stepien, A. & Parker, M. 2007. Expression of p53 and its homolog, p73, in HPV DNA positive oesophageal squamous cell carcinomas. *Virology.* 369(1): 182-190.

Matsha, T., Brink, L., van Rensburg, S., Hon, D., Lombard, C. & Erasmus, R. 2006. Traditional home-brewed beer consumption and iron status in patients with esophageal cancer and healthy control subjects from Transkei, South Africa. *Nutrition and Cancer.* 56(1): 67-73.

Matsha, T., Erasmus, R., Kafuko, A.B., Mugwanya, D., Stepien, A., Parker, M.I. & CANSA/MRC Oesophageal Cancer Research Group 2002. Human papillomavirus associated with oesophageal cancer. *Journal of Clinical Pathology.* 55(8): 587-590.

Matsumoto, M., Yokoyama, H., Suzuki, H., Shiraishi-Yokoyama, H. & Hibi, T. 2005. Retinoic acid formation from retinol in the human gastric mucosa: role of class IV alcohol dehydrogenase and its relevance to morphological changes. *American Journal of Physiology.Gastrointestinal and Liver Physiology.* 289(3): G429-33.

Mayers, G. & Androphy, E. 1995. The E6 protein. *Human Papillomaviruses.* 47-57.

Mazzaro, C., Tirelli, U. & Pozzato, G. 2005. Hepatitis C virus and non-Hodgkin's lymphoma 10 years later. *Digestive and Liver Disease.* 37(4): 219-226.

McCarthy, C. 2004. *Chromas 2.01.* Available: http://technelysium.com.au/?page_id=13.

McKenna, D.J., McDade, S.S., Patel, D. & McCance, D.J. 2010. MicroRNA 203 expression in keratinocytes is dependent on regulation of p53 levels by E6. *Journal of Virology.* 84(20): 10644-10652.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. & Cunningham, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England).* 26(16): 2069-2070.

McLaughlin-Drubin, M.E., Meyers, J. & Munger, K. 2012. Cancer associated human papillomaviruses. *Current Opinion in Virology.* 2(4): 459-466.

McLaughlin-Drubin, M.E. & Munger, K. 2008. Viruses associated with human cancer. *Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease.* 1782(3): 127-150.

McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N. & Aldape, K. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 455(7216): 1061-1068.

McMurray, H.R. & McCance, D.J. 2004. Degradation of p53, not telomerase activation, by E6 is required for bypass of crisis and immortalization by human papillomavirus type 16 E6/E7. *Journal of Virology.* 78(11): 5698-5706.

Meining, A., Ott, R., Becker, I., Hahn, S., Muhlen, J., Werner, M., Hofler, H., Classen, M., Heldwein, W. & Rosch, T. 2004. The Munich Barrett follow up study: suspicion of Barrett's oesophagus based on either endoscopy or histology only--what is the clinical significance? *Gut.* 53(10): 1402-1407.

Melhado, R.E., Alderson, D. & Tucker, O. 2010. The changing face of esophageal cancer. *Cancer.* 2(3): 1379-1404.

Metzker, M.L. 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics.* 11(1): 31-46.

Meyerson, M., Gabriel, S. & Getz, G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics.* 11(10): 685-696.

Michalek, J., Horvath, R., Benedik, J. & Hrstkova, H. 1999. Human herpesvirus-6 infection in children with cancer. *Pediatric Hematology-Oncology.* 16(5): 423-430.

Mijušković, M., Brown, S.M., Tang, Z., Lindsay, C.R., Efstathiadis, E., Deriano, L. & Roth, D.B. 2012. A streamlined method for detecting structural variants in cancer genomes by short read paired-end sequencing. *PloS One.* 7(10): e48314.

Miller, C.T., Aggarwal, S., Lin, T.K., Dagenais, S.L., Contreras, J.I., Orringer, M.B., Glover, T.W., Beer, D.G. & Lin, L. 2003. Amplification and overexpression of the dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 (DYRK2) gene in esophageal and lung adenocarcinomas. *Cancer Research.* 63(14): 4136-4143.

Miner, A.G., Patel, R.M., Wilson, D.A., Procop, G.W., Minca, E.C., Fullen, D.R., Harms, P.W. & Billings, S.D. 2014. Cytokeratin 20-negative Merkel cell carcinoma is infrequently associated with the Merkel cell polyomavirus. *Modern Pathology.*

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L. & Massagué, J. 2005. Genes that mediate breast cancer metastasis to lung. *Nature.* 436(7050): 518-524.

Mir, M.M., Dar, N.A., Salam, I. & Shah, Z.A. 2008. Mutations in epidermal growth factor receptor gene in esophageal squamous cell carcinoma patients in kashmir- a high incidence area of India. *International Journal of Health Sciences.* 2(2): 17-25.

Mishima, K., Yamada, E., Masui, K., Shimokawara, T., Takayama, K., Sugimura, M. & Ichijima, K. 1998. Overexpression of the ERK/MAP kinases in oral squamous cell carcinoma. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc.* 11(9): 886-891.

Mitelman, F., Johansson, B. & Mertens, F. 2007. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer.* 7(4): 233-245.

Mizuguchi, H. & Kay, M.A. 1998. Efficient construction of a recombinant adenovirus vector by an improved in vitro ligation method. *Human Gene Therapy.* 9(17): 2577-2583.

Modis, Y., Trus, B.L. & Harrison, S.C. 2002. Atomic model of the papillomavirus capsid. *The EMBO Journal.* 21(18): 4754-4762.

Moody, C.A. & Laimins, L.A. 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nature Reviews Cancer.* 10(8): 550-560.

Moody, C.A., Fradet-Turcotte, A., Archambault, J. & Laimins, L.A. 2007. Human papillomaviruses activate caspases upon epithelial differentiation to induce viral genome amplification. *Proceedings of the National Academy of Sciences of the United States of America.* 104(49): 19541-19546.

Morissette, G. & Flamand, L. 2010. Herpesviruses and chromosomal integration. *Journal of Virology.* 84(23): 12100-12109.

Moyes, D.L., Martin, A., Sawcer, S., Temperton, N., Worthington, J., Griffiths, D.J. & Venables, P.J. 2005. The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. *Genomics.* 86(3): 337-341.

Mqoqi, N., Kellett, P., Sitas, F. & Jula, M. 2004. Incidence of histologically diagnosed cancer in South Africa, 1998-1999. In *National Cancer Registry.* Johannesburg, South Africa: National Health Laboratory Service.

Mueller, O., Lightfoot, S. & Schroeder, A. 2004. RNA integrity number (RIN)– Standardization of RNA Quality Control. *Agilent Application Note, Publication.* 1-8.

Mukherjee, S., Roth, M.J., Dawsey, S.M., Yan, W., Rodriguez-Canales, J., Erickson, H.S., Hu, N., Goldstein, A.M., Taylor, P.R. & Richardson, A.M. 2010. Increased matrix metalloproteinase activation in esophageal squamous cell carcinoma. *J Transl Med.* 8(1): 91-97.

Mullen, J.L., Reichman, J. & Rosato, E.F. 1979. Multiple carcinomas following therapy for Hodgkin disease. *Journal of Surgical Oncology.* 11(1): 75-78.

Muller, P.A., Caswell, P.T., Doyle, B., Iwanicki, M.P., Tan, E.H., Karim, S., Lukashchuk, N., Gillespie, D.A., Ludwig, R.L. & Gosselin, P. 2009. Mutant p53 drives invasion by promoting integrin recycling. *Cell.* 139(7): 1327-1341.

Munger, K., Phelps, W.C., Bubb, V., Howley, P.M. & Schlegel, R. 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of Virology.* 63(10): 4417-4421.

Munoz, N., Grassi, A., Qiong, S., Crespi, M., Guo Qing, W. & Zhang Cai, L. 1982. Precursor lesions of oesophageal cancer in high-risk populations in Iran and China. *The Lancet.* 319(8277): 876-879.

MuÑoz, N. & CastellsaguÉ, X. 1994. Epidemiology of oesophageal cancer. *European Journal of Gastroenterology & Hepatology.* 6(8): 649-655.

Munoz, N., Castellsagué, X., de González, A.B. & Gissmann, L. 2006. HPV in the etiology of human cancer. *Vaccine.* 24S1-S10.

Nacheva, E.P., Ward, K.N., Brazma, D., Virgili, A., Howard, J., Leong, H.N. & Clark, D.A. 2008. Human herpesvirus 6 integrates within telomeric regions as evidenced by five different chromosomal sites. *Journal of Medical Virology.* 80(11): 1952-1958.

Naeem, R., Rashid, M. & Pain, A. 2013. READSCAN: a fast and scalable pathogen discovery program with accurate genome relative abundance estimation. *Bioinformatics (Oxford, England).* 29(3): 391-392.

Nair, A., Venkatraman, M., Maliekal, T.T., Nair, B. & Karunagaran, D. 2003. NF-kappaB is constitutively activated in high-grade squamous intraepithelial lesions and squamous cell carcinomas of the human uterine cervix. *Oncogene.* 22(1): 50-58.

Nakagawa, S. & Huibregtse, J.M. 2000. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Molecular and Cellular Biology.* 20(21): 8244-8253.

Nedelcu, A.M. 2001. Complex patterns of plastid 16S rRNA gene evolution in nonphotosynthetic green algae. *Journal of Molecular Evolution.* 53(6): 670-679.

Nees, M., Geoghegan, J.M., Hyman, T., Frank, S., Miller, L. & Woodworth, C.D. 2001. Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-kappaB-responsive genes in cervical keratinocytes. *Journal of Virology.* 75(9): 4283-4296.

Nguyen, M.L., Nguyen, M.M., Lee, D., Griep, A.E. & Lambert, P.F. 2003a. The PDZ ligand domain of the human papillomavirus type 16 E6 protein is required for E6's induction of epithelial hyperplasia in vivo. *Journal of Virology.* 77(12): 6957-6964.

Nguyen, M.M., Nguyen, M.L., Caruana, G., Bernstein, A., Lambert, P.F. & Griep, A.E. 2003b. Requirement of PDZ-containing proteins for cell cycle regulation and differentiation in the mouse lens epithelium. *Molecular and Cellular Biology.* 23(24): 8970-8981.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications.* 241(1): 92-97.

Noch, E. & Khalili, K. 2012. Oncogenic viruses and tumor glucose metabolism: like kids in a candy store. *Molecular Cancer Therapeutics.* 11(1): 14-23.

Nowell, P.C. 1976. The clonal evolution of tumor cell populations. *Science (New York, N.Y.).* 194(4260): 23-28.

Nyberg-Hoffman, C., Shabram, P., Li, W., Giroux, D. & Aguilar-Cordova, E. 1997. Sensitivity and reproducibility in adenoviral infectious titer determination. *Nature Medicine.* 3(7): 808-811.

Oh, S.T., Longworth, M.S. & Laimins, L.A. 2004. Roles of the E6 and E7 proteins in the life cycle of low-risk human papillomavirus type 11. *Journal of Virology.* 78(5): 2620-2626.

Owen, J.L., Iragavarapu-Charyulu, V. & Lopez, D.M. 2004. T cell-derived matrix metalloproteinase-9 in breast cancer: friend or foe? *Breast Disease.* 20(1): 145-153.

Pacella-Norman, R., Urban, M., Sitas, F., Carrara, H., Sur, R., Hale, M., Ruff, P., Patel, M., Newton, R. & Bull, D. 2002. Risk factors for oesophageal, lung, oral and laryngeal cancers in black South Africans. *British Journal of Cancer.* 86(11): 1751-1756.

Painter, R.B. & Young, B.R. 1980. Radiosensitivity in ataxia-telangiectasia: a new explanation. *Proceedings of the National Academy of Sciences of the United States of America.* 77(12): 7315-7317.

Panagiotakis, G.I., Papadogianni, D., Chatziioannou, M.N., Lasithiotaki, I., Delakas, D. & Spandidos, D.A. 2013. Association of human herpes, papilloma and polyoma virus families with bladder cancer. *Tumor Biology.* 34(1): 71-79.

Park, J., Kim, E., Lee, J., Sin, H., Namkoong, S. & Um, S. 2001. Functional inactivation of p73, a homolog of p53 tumor suppressor protein, by human papillomavirus E6 proteins. *International Journal of Cancer.* 91(6): 822-827.

Park, N.H., Song, I.H. & Chung, Y.H. 2006. Chronic hepatitis B in hepatocarcinogenesis. *Postgraduate Medical Journal.* 82(970): 507-515.

Park, R.B. & Androphy, E.J. 2002. Genetic analysis of high-risk e6 in episomal maintenance of human papillomavirus genomes in primary human keratinocytes. *Journal of Virology.* 76(22): 11359-11364.

Parkin, D., Pisani, P. & Ferlay, J. 1993. Estimates of the worldwide incidence of eighteen major cancers in 1985. *International Journal of Cancer.* 54(4): 594-606.

Parkin, D.M. 2006. The global health burden of infection-associated cancers in the year 2002. *International Journal of Cancer.* 118(12): 3030-3044.

Parsonnet, J. 1999. Microbes and malignancy: infection as a cause of human cancers. In Oxford University Press, USA. 89-106.

Partek Inc. 2014. *Gene Expression analysis using Partek® software.* 6.6. St. Louis: Partek Inc.

Pascale, M., Pracella, D., Barbazza, R., Marongiu, B., Roggero, E., Bonin, S. & Stanta, G. 2013. Is Human Papillomavirus Associated with Prostate Cancer Survival? *Disease Markers.* 35(6): 607-613.

Patel, D., Huang, S., Baglia, L.A. & McCance, D.J. 1999. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *The EMBO Journal.* 18(18): 5061-5072.

Pennathur, A., Gibson, M.K., Jobe, B.A. & Luketich, J.D. 2013. Oesophageal carcinoma. *The Lancet.* 381(9864): 400-412.

Pertea, M. & Salzberg, S.L. 2010. Between a chicken and a grape: estimating the number of human genes. *Genome Biol.* 11(5): 206.

Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L. & Shaw, K.L. 2000. Evidence for DNA loss as a determinant of genome size. *Science (New York, N.Y.).* 287(5455): 1060-1062.

Pfister, H. 1992. Human papillomaviruses and skin cancer. *Seminars in Cancer Biology.* 3(5): 263-271.

Pfister, H. 2003. Chapter 8: Human papillomavirus and skin cancer. *Journal of the National Cancer Institute.Monographs.* 31(31): 52-56.

Pim, D. & Banks, L. 2010. Interaction of viral oncoproteins with cellular target molecules: infection with high-risk vs low-risk human papillomaviruses. *Apmis.* 118(6-7): 471-493.

Pim, D., Thomas, M., Javier, R., Gardiol, D. & Banks, L. 2000. HPV E6 targeted degradation of the discs large protein: evidence for the involvement of a novel ubiquitin ligase. *Oncogene.* 19(6): 719-725.

Pirooznia, M., Nagarajan, V. & Deng, Y. 2007. GeneVenn - A web application for comparing gene lists using Venn diagrams. *Bioinformation.* 1(10): 420-422.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M., Ordóñez, G.R. & Bignell, G.R. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature.* 463(7278): 191-196.

Pohl, H. & Welch, H.G. 2005. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *Journal of the National Cancer Institute.* 97(2): 142-146.

Pourshams, A., Saadatian-Elahi, M., Nouraie, M., Malekshah, A., Rakhshani, N., Salahi, R., Yoonessi, A., Semnani, S., Islami, F. & Sotoudeh, M. 2005. Golestan cohort study of oesophageal cancer: feasibility and first results. *British Journal of Cancer.* 92(1): 176-181.

Prives, C. & Hall, P.A. 1999. The p53 pathway. *Journal of Pathology.* 187(1): 112-126.

Promega 2010. *Technical Bulletin, Wizard® SV Gel and PCR Clean-Up System-A9282.* Available: https://www.promega.com/resources/protocols/technical-bulletins/101/wizard-sv-gel-and-pcr-cleanup-system-protocol/.

Prusty, B.K., Husain, S.A. & Das, B.C. 2005. Constitutive activation of nuclear factor -kB: preferntial homodimerization of p50 subunits in cervical carcinoma. *Frontiers in Bioscience : A Journal and Virtual Library.* 10:1510-1519.

Quellhorst, G., Han, Y. & Blanchard, R. 2005. *Validating Microarray Data Using RT 2 Real-Time PCR.* SABiosciences.

Rapp, L. & Chen, J.J. 1998. The papillomavirus E6 proteins. *Biochimica Et Biophysica Acta.* 1378(1): F1-19.

Rastegar, D.A. 2010. HIV infection and esophageal cancer. *Archives of Internal Medicine.* 170(10): 916-917.

Rautava, J. & Syrjänen, S. 2012. Biology of human papillomavirus infections in head and neck carcinogenesis. *Head and Neck Pathology.* 6(1): 3-15.

Ren, Q., Kari, C., Quadros, M.R., Burd, R., McCue, P., Dicker, A.P. & Rodeck, U. 2006. Malignant transformation of immortalized HaCaT keratinocytes through deregulated nuclear factor kappaB signaling. *Cancer Research.* 66(10): 5209-5215.

Reznikoff, C.A., Belair, C., Savelieva, E., Zhai, Y., Pfeifer, K., Yeager, T., Thompson, K.J., DeVries, S., Bindley, C. & Newton, M.A. 1994. Long-term genome stability and minimal genotypic and phenotypic alterations in HPV16 E7-, but not E6-, immortalized human uroepithelial cells. *Genes & Development.* 8(18): 2227-2240.

Rihet, S., Lorenzato, M. & Clavel, C. 1996. Oncogenic human papillomaviruses and ploidy in cervical lesions. *Journal of Clinical Pathology.* 49(11): 892-896.

Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M. et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science (New York, N.Y.).* 328(5978): 636-639.

Rocha, E.P. & Blanchard, A. 2002. Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. *Nucleic Acids Research.* 30(9): 2031-2042.

Ronco, L.V., Karpova, A.Y., Vidal, M. & Howley, P.M. 1998. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes & Development.* 12(13): 2061-2072.

Roymans, D. & Slegers, H. 2001. Phosphatidylinositol 3-kinases in tumor progression. *European Journal of Biochemistry.* 268(3): 487-498.

Ruffalo, M., Koyuturk, M., Ray, S. & LaFramboise, T. 2012. Accurate estimation of short read mapping quality for next-generation genome sequencing. *Bioinformatics (Oxford, England).* 28(18): i349-i355.

Rusk, N. & Kiermer, V. 2008. Primer: Sequencing--the next generation. *Nature Methods.* 5(1): 15.

Sabbatini, P. & McCormick, F. 1999. Phosphoinositide 3-OH kinase (PI3K) and PKB/Akt delay the onset of p53-mediated, transcriptionally dependent apoptosis. *The Journal of Biological Chemistry.* 274(34): 24263-24269.

Sabroe, I., Peck, M.J., Van Keulen, B.J., Jorritsma, A., Simmons, G., Clapham, P.R., Williams, T.J. & Pease, J.E. 2000. A small molecule antagonist of chemokine receptors CCR1 and CCR3. Potent inhibition of eosinophil function and CCR3-mediated HIV-1 entry. *The Journal of Biological Chemistry.* 275(34): 25985-25992.

Samantaray, S., Sharma, R., Chattopadhyaya, T., Gupta, S.D. & Ralhan, R. 2004. Increased expression of MMP-2 and MMP-9 in esophageal squamous cell carcinoma. *Journal of Cancer Research and Clinical Oncology.* 130(1): 37-44.

Sammon, A.M. 1992. A case-control study of diet and social factors in cancer of the esophagus in transkei. *Cancer.* 69(4): 860-865.

Samuels, D.C., Han, L., Li, J., Quanghu, S., Clark, T.A., Shyr, Y. & Guo, Y. 2013. Finding the lost treasures in exome sequencing data. *Trends in Genetics.* 29(10): 593-599.

Santarpia, L., Nicoloso, M. & Calin, G.A. 2010. MicroRNAs: a complex regulatory network drives the acquisition of malignant cell phenotype. *Endocrine-Related Cancer.* 17(1): F51-75.

Sapp, M. & Bienkowska-Haba, M. 2009. Viral entry mechanisms: human papillomavirus and a long journey from extracellular matrix to the nucleus. *FEBS Journal.* 276(24): 7206-7216.

Savage, K.I., Gorski, J.J., Barros, E.M., Irwin, G.W., Manti, L., Powell, A.J., Pellagatti, A., Lukashchuk, N., McCance, D.J. & McCluggage, W.G. 2014. Identification of a BRCA1-mRNA Splicing Complex Required for Efficient DNA Repair and Maintenance of Genomic Stability. *Molecular Cell.* 54(3): 445-459.

Schaeffer, A.J., Nguyen, M., Liem, A., Lee, D., Montagna, C., Lambert, P.F., Ried, T. & Difilippantonio, M.J. 2004. E6 and E7 oncoproteins induce distinct patterns of chromosomal aneuploidy in skin tumors from transgenic mice. *Cancer Research.* 64(2): 538-546.

Schafer, G., Kabanda, S., van Rooyen, B., Marusic, M.B., Banks, L. & Parker, M.I. 2013. The role of inflammation in HPV infection of the Oesophagus. *BMC Cancer.* 13185-2407-13-185.

Scheffner, M., Romanczuk, H., Münger, K., Huibregtse, J., Mietz, J. & Howley, P. 1994. Functions of human papillomavirus proteins. *Curr Top Microbiol Immunol.* (186): 83-99.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D. & Howley, P.M. 1993. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell.* 75(3): 495-505.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J. & Howley, P.M. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell.* 63(6): 1129-1136.

Scheffner, M., Munger, K., Huibregtse, J.M. & Howley, P.M. 1992. Targeted degradation of the retinoblastoma protein by human papillomavirus E7-E6 fusion proteins. *The EMBO Journal.* 11(7): 2425-2431.

Schelhorn, S., Fischer, M., Tolosi, L., Altmüller, J., Nürnberg, P., Pfister, H., Lengauer, T. & Berthold, F. 2013. Sensitive detection of viral transcripts in human tumor transcriptomes. *PLoS Computational Biology.* 9(10): e1003228.

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E. & Feuk, L. 2007. Challenges and standards in integrating surveys of structural variation. *Nature Genetics.* 39S7-S15.

Schlegel, R., Phelps, W.C., Zhang, Y.L. & Barbosa, M. 1988. Quantitative keratinocyte assay detects two biological activities of human papillomavirus DNA and identifies viral types associated with cervical carcinoma. *The EMBO Journal.* 7(10): 3181-3187.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J. & Odom, D.T. 2009. ChIP-seq: Using high-throughput sequencing to discover protein–DNA interactions. *Methods.* 48(3): 240-248.

Schmieder, R. & Edwards, R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS One.* 6(3): e17288.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. & Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology.* 7:3.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. et al. 2004. Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.).* 305(5683): 525-528.

Selgrad, M., Malfertheiner, P., Fini, L., Goel, A., Boland, C.R. & Ricciardiello, L. 2008. The role of viral and bacterial pathogens in gastrointestinal cancer. *Journal of Cellular Physiology.* 216(2): 378-388.

Selivanova, G. & Ivaska, J. 2009. Integrins and Mutant p53 on the Road to Metastasis. *Cell.* 139(7): 1220-1222.

Shamanin, V., Glover, M., Rausch, C., Proby, C., Leigh, I.M., zur Hausen, H. & de Villiers, E.M. 1994. Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Research.* 54(17): 4610-4613.

Sharp, A.J., Cheng, Z. & Eichler, E.E. 2006. Structural variation of the human genome. *Annu.Rev.Genomics Hum.Genet.* 7:407-442.

Shen, Z., Xu, L., Chen, M., Shen, J., Cai, W. & Zeng, Y. 2002. Progressive transformation of immortalized esophageal epithelial cells. *World Journal of Gastroenterology.* 8(6): 976-981.

Shiau, M., Fan, L., Yang, S., Tsao, C., Lee, H., Cheng, Y., Lai, L. & Chang, Y. 2013. Human papillomavirus up-regulates MMP-2 and MMP-9 expression and activity by inducing interleukin-8 in lung adenocarcinomas. *PloS One.* 8(1): e54423.

Shigemizu, D., Fujimoto, A., Akiyama, S., Abe, T., Nakano, K., Boroevich, K.A., Yamamoto, Y., Furuta, M., Kubo, M. & Nakagawa, H. 2013. A practical method to detect SNVs and indels from whole genome and exome sequencing data. *Scientific Reports.* 3(2161).

Shima, I., Sasaguri, Y., Kusukawa, J., Yamana, H., Fujita, H., Kakegawa, T. & Morimatsu, M. 1992. Production of matrix metalloproteinase-2 and metalloproteinase-3 related to malignant behavior of esophageal carcinoma. A clinicopathologic study. *Cancer.* 70(12): 2747-2753.

Shuai, K. & Liu, B. 2003. Regulation of JAK–STAT signalling in the immune system. *Nature Reviews Immunology.* 3(11): 900-911.

Siewert, J.R. & Ott, K. 2007. Are squamous and adenocarcinomas of the esophagus the same disease? 17(1): 38-44.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. & Ponting, C.P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics.* 15(2): 121-132.

Slebos, R.J., Kessis, T.D., Chen, A.W., Han, S.M., Hedrick, L. & Cho, K.R. 1995. Functional consequences of directed mutations in human papillomavirus E6 proteins: abrogation of p53-mediated cell cycle arrest correlates with p53 binding and degradation in vitro. *Virology.* 208(1): 111-120.

Smith, P.P., Bryant, E.M., Kaur, P. & McDougall, J.K. 1989. Cytogenetic analysis of eight human papillomavirus immortalized human keratinocyte cell lines. *International Journal of Cancer.* 44(6): 1124-1131.

Smyth, G.K. & Speed, T. 2003. Normalization of cDNA microarray data. *Methods.* 31(4): 265-273.

Song, Y., Li, L., Ou, Y., Gao, Z., Li, E., Li, X., Zhang, W., Wang, J., Xu, L. & Zhou, Y. 2014. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature.* 509(7498): 91-95.

Song, S., Gulliver, G.A. & Lambert, P.F. 1998. Human papillomavirus type 16 E6 and E7 oncogenes abrogate radiation-induced DNA damage responses in vivo through p53-dependent and p53-independent pathways. *Proceedings of the National Academy of Sciences of the United States of America.* 95(5): 2290-2295.

Song, S., Pitot, H.C. & Lambert, P.F. 1999. The human papillomavirus type 16 E6 gene alone is sufficient to induce carcinomas in transgenic animals. *Journal of Virology.* 73(7): 5887-5893.

Soorya, L., Kolevzon, A., Zweifach, J., Lim, T., Dobry, Y., Schwartz, L., Frank, Y., Wang, A.T., Cai, G. & Parkhomenko, E. 2013. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Mol Autism.* 4(1): 18.

Southern, S. & Herrington, C. 2000. Disruption of cell cycle control by human papillomaviruses with special reference to cervical carcinoma. *International Journal of Gynecological Cancer.* 10(4): 263-274.

Southern, S.A. & Herrington, C.S. 1998. Differential cell cycle regulation by low- and high-risk human papillomaviruses in low-grade squamous intraepithelial lesions of the cervix. *Cancer Research.* 58(14): 2941-2945.

Spurgeon, M.E., Chung, S. & Lambert, P.F. 2014. Recurrence of Cervical Cancer in Mice after Selective Estrogen Receptor Modulator Therapy. *The American Journal of Pathology.* 184(2): 530-540.

Sripa, B., Kaewkes, S., Sithithaworn, P., Mairiang, E., Laha, T., Smout, M., Pairojkul, C., Bhudhisawasdi, V., Tesana, S. & Thinkamrop, B. 2007. Liver fluke induces cholangiocarcinoma. *PLoS Medicine.* 4(7): e201.

Stebbing, J., Krown, S.E., Bower, M., Batra, A., Slater, S., Serraino, D., Dezube, B.J., Dhir, A.A. & Pantanowitz, L. 2010. Primary esophageal carcinoma in the era of highly active antiretroviral therapy. *Archives of Internal Medicine.* 170(2): 203-207.

Steenbergen, R.D., Hermsen, M.A., Walboomers, J.M., Meijer, G.A., Baak, J., Meijer, C.J. & Snijders, P.J. 1998. Non-random allelic losses at 3p, 11p and 13q during HPV-mediated immortalization and concomitant loss of terminal differentiation of human keratinocytes. *International Journal of Cancer.* 76(3): 412-417.

Steenbergen, R.D., Walboomers, J.M., Meijer, C.J., van der Raaij-Helmer, E.M., Parker, J.N., Chow, L.T., Broker, T.R. & Snijders, P.J. 1996. Transition of human papillomavirus type 16 and 18 transfected human foreskin keratinocytes towards immortality: activation of telomerase and allele losses at 3p, 10p, 11q and/or 18q. *Oncogene.* 13(6): 1249-1257.

Stewart, D., Ghosh, A. & Matlashewski, G. 2005. Involvement of nuclear export in human papillomavirus type 18 E6-mediated ubiquitination and degradation of p53. *Journal of Virology.* 79(14): 8773-8783.

Stöppler, M.C., Straight, S.W., Tsao, G., Schlegel, R. & Mccance, D.J. 1996. The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA. *Virology.* 223(1): 251-254.

Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I.M., Matlashewski, G. & Banks, L. 1998. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. *Nature.* 393(6682): 229-234.

Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A. et al. 2011. The mutational

landscape of head and neck squamous cell carcinoma. *Science (New York, N.Y.).* 333(6046): 1157-1160.

Stratton, M.R., Campbell, P.J. & Futreal, P.A. 2009. The cancer genome. *Nature.* 458(7239): 719-724.

Strausberg, R. & Simpson, A. 2009. Whole-genome cancer analysis as an approach to deeper understanding of tumour biology. *British Journal of Cancer.* 102(2): 243-248.

Sun, X., Frierson, H.F., Chen, C., Li, C., Ran, Q., Otto, K.B., Cantarel, B.M., Vessella, R.L., Gao, A.C. & Petros, J. 2005. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nature Genetics.* 37(4): 407-412.

Sun, L., Zhang, G., Lei, T., Huang, C., Song, T. & Si, L. 2008. Two different HPV-11E6 fusion proteins trap p53 in the cytoplasm and induce apoptosis. *Cancer Biology & Therapy.* 7(12): 1909-1915.

Sung, W., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N. & Tennakoon, C. 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nature Genetics.* 44(7): 765-769.

Sur, M. & Cooper, K. 1998. The role of the human papilloma virus in esophageal cancer. *Pathology.* 30(4): 348-354.

Syrjänen, K., Syrjänen, S. & Pyrhönen, S. 1982. Human papilloma virus (HPV) antigens in lesions of laryngeal squamous cell carcinomas. *ORL.* 44(6): 323-334.

Syrjänen, S.M. & Syrjänen, K.J. 1999. New concepts on the role of human papillomavirus in cell cycle regulation. *Annals of Medicine.* 31(3): 175-187.

Syrjanen, K., Pyrhonen, S., Aukee, S. & Koskela, E. 1982. Squamous cell papilloma of the esophagus: a tumour probably caused by human papilloma virus (HPV). *Diagnostic Histopathology / Published in Association with the Pathological Society of Great Britain and Ireland.* 5(4): 291-296.

Syrjanen, K.J. 1982. Histological changes identical to those of condylomatous lesions found in esophageal squamous cell carcinomas. 52(4): 283-292.

Tagle, D.K.J., Sotelo, D.H., Illades-Aguiar, B., Leyva-Vazquez, M.A., Alfaro, E.F., Coronel, Y.C., del Moral Hernández, O. & Romero, Luz del Carmen Alarcón 2014. Expression of E6, p53 and p21 proteins and physical state of HPV16 in cervical cytologies with and without low grade lesions. *International Journal of Clinical and Experimental Medicine.* 7(1): 186.

Tamanoi, F. & Stillman, B.W. 1982. Function of adenovirus terminal protein in the initiation of DNA replication. *Proceedings of the National Academy of Sciences of the United States of America.* 79(7): 2221-2225.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. & Golub, T.R. 1999. Interpreting patterns of gene expression with self-organizing maps:

methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences of the United States of America.* 96(6): 2907-2912.

Tan, P. 2004. *Experimental validation of Microarray data.* Singapore: Dept of Cellular and Molecular Research, National Cancer Centre/Defence Medical and Environmental Research Institute.

Tanaka, Y., Tsuda, S., Kunikata, H., Sato, J., Kokubun, T., Yasuda, M., Nishiguchi, K.M., Inada, T. & Nakazawa, T. 2014. Profiles of Extracellular miRNAs in the Aqueous Humor of Glaucoma Patients Assessed with a Microarray System. *Scientific Reports.* 4.

Tang, K., Alaei-Mahabadi, B., Samuelsson, T., Lindh, M. & Larsson, E. 2013. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nature Communications.* 4(2513).

Taylor, W.R., DePrimo, S.E., Agarwal, A., Agarwal, M.L., Schonthal, A.H., Katula, K.S. & Stark, G.R. 1999. Mechanisms of G2 arrest in response to overexpression of p53. *Molecular Biology of the Cell.* 10(11): 3607-3622.

Thierry, F. 2009. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology.* 384(2): 375-379.

Thomas, M. & Banks, L. 1998. Inhibition of Bak-induced apoptosis by HPV-18 E6. *Oncogene.* 17(23): 2943-2954.

Thomas, M. & Banks, L. 1999. Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. *The Journal of General Virology.* 80(6): 1513-1517.

Tian, F., Zang, W.D., Hou, W.H., Liu, H.T. & Xue, L.X. 2006. Nuclear factor-kB signaling pathway constitutively activated in esophageal squamous cell carcinoma cell lines and inhibition of growth of cells by small interfering RNA. *Acta Biochimica Et Biophysica Sinica.* 38(5): 318-326.

Tommasino, M. 2014. The human papillomavirus family and its role in carcinogenesis. *Semin Cancer Biol.* (26): 13-21.

Tonin, P.N., Hudson, T.J., Rodier, F., Bossolasco, M., Lee, P.D., Novak, J., Manderson, E.N., Provencher, D. & Mes-Masson, A. 2001. Microarray analysis of gene expression mirrors the biology of an ovarian cancer model. *Oncogene.* 20(45): 6617-6626.

Towbin, H., Staehelin, T. & Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America.* 76(9): 4350-4354.

Tsao, S., Mok, S.C., Fey, E.G., Fletcher, J.A., Wan, T.S., Chew, E., Muto, M.G., Knapp, R.C. & Berkowitz, R.S. 1995. Characterization of human ovarian surface epithelial cells immortalized by human papilloma viral oncogenes (HPV-E6E7 ORFs). *Experimental Cell Research.* 218(2): 499-507.

Tu, Z., He, G., Li, K.X., Chen, M.J., Chang, J., Chen, L., Yao, Q., Liu, D.P., Ye, H. & Shi, J. 2005. An improved system for competent cell preparation and high efficiency plasmid transformation using different Escherichia coli strains. *Electronic Journal of Biotechnology.* 8(1): 113-120.

Tungteakkhun, S.S. & Duerksen-Hughes, P.J. 2008. Cellular binding partners of the human papillomavirus E6 protein. *Archives of Virology.* 153(3): 397-408.

Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M.I., Kidd, K.K. & Lenz, J. 2001. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Current Biology.* 11(19): 1531-1535.

Umar, S.B. & Fleischer, D.E. 2008. Esophageal cancer: epidemiology, pathogenesis and prevention. *Nature Clinical Practice Gastroenterology & Hepatology.* 5(9): 517-526.

Underbrink, M.P., Howie, H.L., Bedard, K.M., Koop, J.I. & Galloway, D.A. 2008. E6 proteins from multiple human betapapillomavirus types degrade Bak and protect keratinocytes from apoptosis after UVB irradiation. *Journal of Virology.* 82(21): 10408-10417.

Ungerback, J., Belenki, D., Jawad ul-Hassan, A., Fredrikson, M., Fransen, K., Elander, N., Verma, D. & Soderkvist, P. 2012. Genetic variation and alterations of genes involved in NFkappaB/TNFAIP3- and NLRP3-inflammasome signaling affect susceptibility and outcome of colorectal cancer. *Carcinogenesis.* 33(11): 2126-2134.

Vande Pol, S.B. & Klingelhutz, A.J. 2013. Papillomavirus E6 oncoproteins. *Virology.* 445(1): 115-137.

Viswanathan, G.A., Seto, J., Patil, S., Nudelman, G. & Sealfon, S.C. 2008. Getting started in biological pathway construction and analysis. *PLoS Computational Biology.* 4(2): e16.

Vizcaino, A.P., Moreno, V., Lambert, R. & Parkin, D.M. 2002. Time trends incidence of both major histologic types of esophageal carcinomas in selected countries, 1973–1995. *International Journal of Cancer.* 99(6): 860-868.

Vogelsang, M., Paccez, J.D., Schäfer, G., Dzobo, K., Zerbini, L.F. & Parker, M.I. 2014. Aberrant methylation of the MSH3 promoter and distal enhancer in esophageal cancer patients exposed to first-hand tobacco smoke. *Journal of Cancer Research and Clinical Oncology.* 1-9.

Vogelsang, M., Wang, Y., Veber, N., Mwapagha, L.M. & Parker, M.I. 2012. The cumulative effects of polymorphisms in the DNA mismatch repair genes and tobacco smoking in oesophageal cancer risk. *PloS One.* 7(5): e36962.

Vogelstein, B. & Kinzler, K.W. 1993. The multistep nature of cancer. *Trends in Genetics.* 9(4): 138-141.

Vogelstein, B. & Kinzler, K.W. 2004. Cancer genes and the pathways they control. *Nature Medicine.* 10(8): 789-799.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A.,Jr & Kinzler, K.W. 2013. Cancer genome landscapes. *Science (New York, N.Y.).* 339(6127): 1546-1558.

Vorburger, S.A. & Hunt, K.K. 2002. Adenoviral gene therapy. *The Oncologist.* 7(1): 46-59.

Vousden, K.H., Androphy, E.J., Schiller, J.T. & Lowy, D.R. 1989. Mutational analysis of bovine papillomavirus E6 gene. *Journal of Virology.* 63(5): 2340-2342.

Wakhisi, J., Patel, K., Buziba, N. & Rotich, J. 2005. Esophageal cancer in north rift valley of western Kenya. *African Health Sciences.* 5(2): 157-163.

Wald, A.I., Hoskins, E.E., Wells, S.I., Ferris, R.L. & Khan, S.A. 2011. Alteration of microRNA profiles in squamous cell carcinoma of the head and neck cell lines by human papillomavirus. *Head & Neck.* 33(4): 504-512.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J. & Zhang, J. 2008. The diploid genome sequence of an Asian individual. *Nature.* 456(7218): 60-65.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics.* 10(1): 57-63.

Wang, G.Q., Abnet, C.C., Shen, Q., Lewin, K.J., Sun, X.D., Roth, M.J., Qiao, Y.L., Mark, S.D., Dong, Z.W., Taylor, P.R. et al. 2005. Histological precursors of oesophageal squamous cell carcinoma: results from a 13 year prospective follow up study in a high risk population. *Gut.* 54(2): 187-192.

Wang, X., Wang, H.K., McCoy, J.P., Banerjee, N.S., Rader, J.S., Broker, T.R., Meyers, C., Chow, L.T. & Zheng, Z.M. 2009. Oncogenic HPV infection interrupts the expression of tumor-suppressive miR-34a through viral oncoprotein E6. *RNA (New York, N.Y.).* 15(4): 637-647.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., Chen, X. & Li, Y. 2010. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (Ipomoea batatas). *BMC Genomics.* 11726-2164-11-726.

Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J.B., Yan, P., Sastry, K.J., Piyathilake, C.J., Hunt, K.K. & Johanning, G.L. 2008. Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. *Cancer Research.* 68(14): 5869-5877.

Watson, R.A., Thomas, M., Banks, L. & Roberts, S. 2003. Activity of the human papillomavirus E6 PDZ-binding motif correlates with an enhanced morphological transformation of immortalized human keratinocytes. *Journal of Cell Science.* 116(24): 4925-4934.

Wazer, D.E., Liu, X.L., Chu, Q., Gao, Q. & Band, V. 1995. Immortalization of distinct human mammary epithelial cell types by human papilloma virus 16 E6 or E7.

Proceedings of the National Academy of Sciences of the United States of America. 92(9): 3687-3691.

Wei, X., Moncada-Pazos, A., Cal, S., Soria-Valles, C., Gartner, J., Rudloff, U., Lin, J.C., Rosenberg, S.A., López-Otín, C. & Samuels, Y. 2011. Analysis of the disintegrin-metalloproteinases family reveals ADAM29 and ADAM7 are often mutated in melanoma. *Human Mutation.* 32(6): E2148-E2175.

Weijerman, P.C., Konig, J.J., Wong, S.T., Niesters, H.G. & Peehl, D.M. 1994. Lipofection-mediated immortalization of human prostatic epithelial cells of normal and malignant origin using human papillomavirus type 18 DNA. *Cancer Research.* 54(21): 5579-5583.

Weir, B., Zhao, X. & Meyerson, M. 2004. Somatic alterations in the human cancer genome. *Cancer Cell.* 6(5): 433-438.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L. & Somogyi, R. 1998. Large-scale temporal gene expression mapping of central nervous system development. *Proceedings of the National Academy of Sciences of the United States of America.* 95(1): 334-339.

Werness, B.A., Levine, A.J. & Howley, P.M. 1990. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science (New York, N.Y.).* 248(4951): 76-79.

White, A.E., Livanos, E.M. & Tlsty, T.D. 1994. Differential disruption of genomic integrity and cell cycle regulation in normal human fibroblasts by the HPV oncoproteins. *Genes & Development.* 8(6): 666-677.

Williamson, A.L., Jaskiesicz, K. & Gunning, A. 1991. The detection of human papillomavirus in oesophageal lesions. *Anticancer Research.* 11(1): 263-265.

Winkler, B., Capo, V., Reumann, W., Ma, A., Porta, R.L., Reilly, S., Green, P.M., Richart, R.M. & Crum, C.P. 1985. Human papillomavirus infection of the esophagus. A clinicopathologic study with demonstration of papillomavirus antigen by the immunoperoxidase technique. *Cancer.* 55(1): 149-155.

Woodworth, C.D., Bowden, P.E., Doniger, J., Pirisi, L., Barnes, W., Lancaster, W.D. & DiPaolo, J.A. 1988. Characterization of normal human exocervical epithelial cells immortalized in vitro by papillomavirus types 16 and 18 DNA. *Cancer Research.* 48(16): 4620-4628.

Woodworth, C.D., Waggoner, S., Barnes, W., Stoler, M.H. & DiPaolo, J.A. 1990. Human cervical and foreskin epithelial cells immortalized by human papillomavirus DNAs exhibit dysplastic differentiation in vivo. *Cancer Research.* 50(12): 3709-3715.

Wu, D., Irwin, D.M. & Zhang, Y. 2011. De novo origin of human protein-coding genes. *PLoS Genetics.* 7(11): e1002379.

Xi, R., Kim, T.M. & Park, P.J. 2010. Detecting structural variations in the human genome using next generation sequencing. *Briefings in Functional Genomics.* 9(5-6): 405-415.

Yang, L., Yang, H., Wu, K., Shi, X., Ma, S. & Sun, Q. 2014. Prevalence of HPV and variation of HPV 16/HPV 18 E6/E7 genes in cervical cancer in women in South West China. *Journal of Medical Virology.* 86(11): 1926-1936.

Yang, Y., Chu, J., Luo, M., Wu, Y., Zhang, Y., Feng, Y., Shi, Z., Xu, X., Han, Y. & Cai, Y. 2008. Amplification of PRKCI, located in 3q26, is associated with lymph node metastasis in esophageal squamous cell carcinoma. *Genes, Chromosomes and Cancer.* 47(2): 127-136.

Yang, H., Zhong, Y., Peng, C., Chen, J.Q. & Tian, D. 2010. Important role of indels in somatic mutations of human cancer genes. *BMC Medical Genetics.* 11128-2350-11-128.

Yang, Z.Q., Imoto, I., Fukuda, Y., Pimkhaokham, A., Shimada, Y., Imamura, M., Sugano, S., Nakamura, Y. & Inazawa, J. 2000. Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. *Cancer Research.* 60(17): 4735-4739.

Yeang, C.H., McCormick, F. & Levine, A. 2008. Combinatorial patterns of somatic gene mutations in cancer. *FASEB Journal.* 22(8): 2605-2622.

Yeung, C.A., Tsang, T., Yau, P. & Kwok, T. 2011. Human papillomavirus type 16 E6 induces cervical cancer cell migration through the p53/microRNA-23b/urokinase-type plasminogen activator pathway. *Oncogene.* 30(21): 2401-2410.

Yin, S., Chou, C., Lai, C., Lee, S. & Han, C. 2003. Human class IV alcohol dehydrogenase: kinetic mechanism, functional roles and medical relevance. *Chemico-Biological Interactions.* 143:219-227.

Ying, J., Li, H., Cui, Y., Wong, A., Langford, C. & Tao, Q. 2006. Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter. *Leukemia.* 20(6): 1173-1175.

Ying, J., Shan, L., Li, J., Zhong, L., Xue, L., Zhao, H., Li, L., Langford, C., Guo, L. & Qiu, T. 2012. Genome-wide screening for genetic alterations in esophageal cancer by aCGH identifies 11q13 amplification oncogenes associated with nodal metastasis. *PloS One.* 7(6): e39797.

Yokota, S., Yokosawa, N., Okabayashi, T., Suzutani, T., Miura, S., Jimbow, K. & Fujii, N. 2004. Induction of suppressor of cytokine signaling-3 by herpes simplex virus type 1 contributes to inhibition of the interferon signaling pathway. *Journal of Virology.* 78(12): 6282-6286.

Yoon, S., Xuan, Z., Makarov, V., Ye, K. & Sebat, J. 2009. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research.* 19(9): 1586-1592.

Yu, G., Greninger, A.L., Isa, P., Phan, T.G., Martínez, M.A., de la Luz Sanchez, Maria, Contreras, J.F., Santos-Preciado, J.I., Parsonnet, J. & Miller, S. 2012. Discovery of a novel polyomavirus in acute diarrheal samples from children. *PloS One.* 7(11): e49449.

Yuan, H., Fu, F., Zhuo, J., Wang, W., Nishitani, J., An, D.S., Chen, I.S. & Liu, X. 2005. Human papillomavirus type 16 E6 and E7 oncoproteins upregulate c-IAP2 gene expression and confer resistance to apoptosis. *Oncogene.* 24(32): 5069-5078.

Zaridze, D.G., Basieva, T., Kabulov, M., Day, N.E. & Duffy, S.W. 1992. Oesophageal cancer in the Republic of Karakalpakstan. *International Journal of Epidemiology.* 21(4): 643-648.

Zeitler, J., Hsu, C.P., Dionne, H. & Bilder, D. 2004. Domains controlling cell polarity and proliferation in the Drosophila tumor suppressor Scribble. *The Journal of Cell Biology.* 167(6): 1137-1146.

Zhang, H., Jin, Y., Chen, X., Jin, C., Law, S., Tsao, S. & Kwong, Y. 2007. Papillomavirus type 16 E6/E7 and human telomerase reverse transcriptase in esophageal cell immortalization and early transformation. *Cancer Letters.* 245(1): 184-194.

Zhang, P., Nouri, M., Brandsma, J.L., Iftner, T. & Steinberg, B.M. 1999. Induction of E6/E7 expression in cottontail rabbit papillomavirus latency following UV activation. *Virology.* 263(2): 388-394.

Zhang, Q., Zhang, D., Shen, Z., Xu, L., Li, E. & Au, W.W. 2011. Infection and integration of human papillomavirus in esophageal carcinoma. *International Journal of Hygiene and Environmental Health.* 214(2): 156-161.

Zhang, Y., Li, B., Li, C., Cai, Q., Zheng, W. & Long, J. 2014. Improved Variant Calling Accuracy by Merging Replicates in Whole-Exome Sequencing Studies. *BioMed Research International.* (2014): 319534.

Zhang, B., Schmoyer, D., Kirov, S. & Snoddy, J. 2004. GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC Bioinformatics.* (18): 5-16.

Zhang, D., Zhang, Q., Zhou, L., Huo, L., Zhang, Y., Shen, Z. & Zhu, Y. 2010. Comparison of prevalence, viral load, physical status and expression of human papillomavirus-16, -18 and -58 in esophageal and cervical cancer: a case-control study. *BMC Cancer.* 10650-2407-10-650.

Zhang, K., Li, J.T., Li, S.Y., Zhu, L.H., Zhou, L. & Zeng, Y. 2011. Integration of human papillomavirus 18 DNA in esophageal carcinoma 109 cells. *World Journal of Gastroenterology : WJG.* 17(37): 4242-4246.

Zhao, G., Krishnamurthy, S., Cai, Z., Popov, V.L., da Rosa, Amelia P Travassos, Guzman, H., Cao, S., Virgin, H.W., Tesh, R.B. & Wang, D. 2013. Identification of Novel Viruses Using VirusHunter--an Automated Data Analysis Pipeline. *PloS One.* 8(10): e78470.

Zhou, X., Spangle, J.M. & Münger, K. 2009. Expression of a viral oncoprotein in normal human epithelial cells triggers an autophagy-related process. *Autophagy.* 5(4): 578-579.

Zhuo, W., Zhang, Y., Wang, Y., Zhuo, X., Zhu, B., Cai, L. & Chen, Z. 2009. Association Studies of CYP1A1 and GSTM1 Polymorphisms with Esophageal Cancer Risk: Evidence-based Meta-analyses. *Archives of Medical Research.* 40(3): 169-179.

Zimmermann, H., Degenkolbe, R., Bernard, H.U. & O'Connor, M.J. 1999. The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *Journal of Virology.* 73(8): 6209-6219.

Zoghbi, H.Y. & Orr, H.T. 2000. Glutamine repeats and neurodegeneration. *Annual Review of Neuroscience.* 23(1): 217-247.

zur Hausen, H. 2001. Viruses in human cancers. *Curr.Sci.* 81523-527.

Zur Hausen, H. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nature Reviews Cancer.* 2(5): 342-350.

Zwolińska, K., Knysz, B., Gąsiorowski, J., Pazgan-Simon, M., Gładysz, A., Sobczyński, M. & Piasecki, E. 2013. Frequency of human endogenous retroviral sequences (HERV) K113 and K115 in the Polish population, and their effect on HIV infection. *PloS One.* 8(10): e77820.