CIS-REGULATION IN HEAD AND NECK SQUAMOUS CELL CARCINOMA (HNSCC): IMPLICATIONS FOR NEW THERAPIES

by

Harmeet Kaur Gill

A thesis submitted to the

UNIVERSITY OF BIRMINGHAM

for the degree of

DOCTOR OF PHILOSOPHY

School of Dentistry

College of Medical and Dental Sciences

University of Birmingham

January 2019

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

HNSCC is the sixth most common cancer in the world and over the last decade its genetic landscape has been extensively investigated. However, due to the heterogenic nature of HNSCC tumours, which is complicated by HPV infection, the effectiveness of current therapies and the development of new therapies have been hindered.

In this thesis, by using a combination of genome-wide methods, the *cis*-regulatory network was investigated in HNSCC cell lines and compared to non-tumourigenic primary oral cells. The analyses revealed a common core regulatory network shared by both the HNSCC and non-tumourigenic cells. The network included pathways, which have previously been found to promote HNSCC such as the EGFR, MAPK, TGF-β and WNT pathways. A novel finding was the involvement of the Hippo pathway and its transcription factor effector, TEAD4. The network appears to be driven primarily by the p63-AP1-TEAD4 transcription factors and in conjunction with the mentioned pathways, appears to form the underlying network required for epithelial function but also holds the potential to become deregulated, leading to HNSCC. Disruption of Hippo-driven gene transcription can reduce HNSCC cell viability therefore pointing to potential therapeutic targets and mechanisms to elucidate HNSCC function.

Dedication

To my parents; without your support, encouragement and sacrifices,

this would have not been possible.

To my incredibly patient husband; I am eternally grateful for your

unwavering support.

A special dedication to Dad - I hope you'd have been proud.

Acknowledgements

First and foremost, I would like to thank my primary supervisor Dr Malgorzata Wiench for all her advice, support and help throughout my PhD. I would also thank my additional supervisors, Dr Sally Roberts and Professor Gabriel Landini for all their help and guidance.

I would especially like to say thank you to Dr Sam Clokie for all his help with the bioinformatics analyses.

Finally, I would like to thank everyone who had any input in my project and allowed me to discuss my thoughts with them (scientific or otherwise), especially Hannah, Fatemeh and Phil.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
ATAC-Seq	Assay for Transposase Accessible Chromatin-sequencing
CAM	Cell Adhesion Molecule
cDNA	Complementary DNA
CEM	Complete E media
CGI	CpG Island
ChIP-Seq	Chromatin ImmunoPrecipitation-Sequencing
CNS	Central Nervous System
CRC	Chromatin Remodelling Complex
DHS	DNase I Hypersensitive Site
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase I-Seq	DNase I-Sequencing
EBV	Epstein Barr Virus
EC	Enhancer Cluster
ECM	Extracellular Matrix
EDTA	Ethylene Diaminetetraacetic Acid
EGF	Epidermal Growth Factor
EMT	Epithelial to Mesenchymal Transition
ENCODE	ENCyclopaedia Of DNA Elements
FACS	Fluorescence-Activated Cell Sorting
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements
FBS	Foetal Bovine Serum
FDA	Food and Drug Administration
FDR	False Discovery Rate
FITC	Fluorescein Isothiocyanate

FPKM	Fragments Per Kilobase exon per Million mapped reads
HAT	Histone Acetyltransferase
HDAC	Histone Deacetyltransferase
HMT	Histone Methyltransferase
HNSCC	Head and Neck Squamous Cell Carcinoma
HOK	Human Oral Keratinocyte
HOMER	Hypergeometric Optimization of Motif EnRichment
HPV	Human papillomavirus
HTLV-1	Human T-lymphotropic infection
KEGG	Kyoto Encyclopedia of Genes and Genomes
miRNA	Micro RNA
NIH	National Institutes of Health
OP	Optimisation
OTK	Oral Tonsil Keratinocyte
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline-Tween 20
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PI	Protease Inhibitor
PIC	Pre-Initiation Complex
PTM	Post-Translational Modification
QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNAPII	RNA Polymerase II
RT	Room Temperature
SDS	Sodium Dodecyl Sulfate
SEM	Standard Error of Mean
shRNA	Short hairpin loop RNA

siRNA	Short interfering RNA
TAD	Topologically Associating Domain
TCGA	The Cancer Genome Atlas
TE	Tris-Ethylene Diaminetetraacetic Acid
TF	TranscriptionFactor
TSS	Transcription Start Site
TTS	Transcription Termination Site
TXN	Transcription
UCSC	University of California Santa Cruz
UTR	Untranslated Region
VP	Verteporfin
WP	WikiPathways

CHAPTER 1: INTRODUCTION

1.1 Epithelial structure

Epithelial tissue is one of the most varied tissues in the body and lines most of the body's surfaces including skin, organs and blood vessels. The tissue's primary function is to protect against dehydration and mechanical, chemical and microbial insults [1]. Two main varieties of epithelial structure exist- simple and stratified. Simple epithelium consists of a single layer of cells allowing for diffusion or filtering of small molecules [1]. Anatomical locations of simple epithelium include, the stomach lining, intestines, circulatory system and lungs, while stratified squamous epithelium can be found in the skin, oral mucosa and oesophagus [1].

The stratified squamous epithelium is a tightly packed, multi-layered and differentiated tissue. The deepest layer of stratified epithelia is the basement membrane, which provides support, allows exchange of nutrients and separates the epithelium from the dermis- a fibrous connective tissue [1, 2] (**Figure 1.1**). Epithelia can be keratinised or non-keratinised; the main difference is at the surface of the structure, where keratinised epithelia possess a layer of dead cells (keratinised layer), while non-keratinised epithelia possess a superficial layer of squamous cells, which have a flatter morphology [1] (**Figure 1.1**).
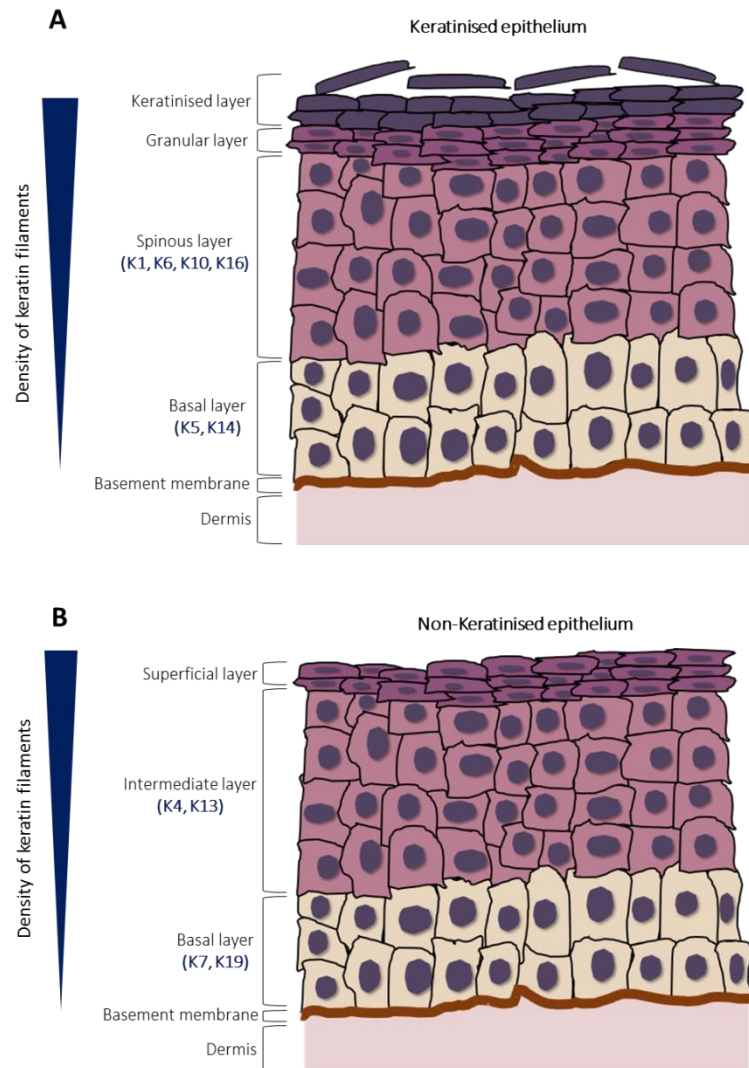
Figure 1.1 The structure of stratified epithelium.
The stratified epithelium is a multi-layered structure and its cells (keratinocytes) express keratin proteins, which form keratin filaments (intermediate filaments). The keratin proteins are characteristic for each type of epithelium (keratinised or non-keratinised) and for the different layers. The dermis (a fibrous connective tissue) is required for mechanical support of the epithelium, but also provides nutrients and a blood supply. Between the dermis and epithelium is the basement membrane, whose primary function is to anchor the basal cells with cell adhesion molecules. **(A)** Keratinised epithelium is composed of five layers; the basal layer, the spinous layer, the granular layer and the keratinised layer. **(B)** The non-keratinised epithelium consists of the basal, intermediate and superficial layers. (Figure recreated and adapted with permission from Roberts S., 2015 [3]).

1.1.1 Oral epithelia

Within the oral cavity, regions involved in mastication, such as gingiva and hard palate, are typically lined by keratinised epithelia [1]. Less rigid regions, such as the mucosal lining, required for speech and swallowing, tend to be covered by non-keratinised epithelia [1].

Keratinised stratified epithelia consist of the basal layer, spinous layer, granular layer and keratinised layer **(Figure 1.1A)**. Layers of non-keratinised epithelia include the basal, intermediate and superficial layers [4] **(Figure 1.1B)**. The basal layer, in order to replace dead cells, provides proliferative and differentiation capacity to the tissue [1, 2]. Upon differentiation, the basal cells migrate upwards and undergo changes to replace cells of the remaining layers [4].

Epithelial cells produce proteins called keratin (also known as cytokeratins), therefore the cells are sometimes referred to as keratinocytes [5]. Keratins are a family of fibrous proteins encoded by 54 different genes, which form heterodimers and polymerise as intermediate filaments to form an essential part of the cytoskeletal network [5]. The functional importance of keratin has been demonstrated in a study in which keratin 6 null mice developed fragile oral mucosal epithelia and were not viable [6]. The basal cells of keratinised epithelia express keratins 5 and 14, while directly above in the spinous layer, keratins 1, 6, 10 and 16 are expressed [4] **(Figure 1.1A)**. The cells of the granular layer in keratinised epithelia are flatter and contain the most dense network of keratin filaments, which offers protection against mechanical and chemical insults [4, 7]. Cells of the keratinised layer possess fewer cell adhesion junctions and no organelles and therefore can be shed from the epithelium [4]. Shed cells are eventually replaced by

3

basal cells as they migrate upwards [4]. Basal cells of non-keratinised epithelia express keratin 7 and 19, while cells of the intermediate layer express keratins 4 and 13 [4] **(Figure 1.1B)**.

The packed cell structure characteristic of epithelia is maintained by cell adhesion molecules (CAMs), which are located either between two cells for cell-cell adhesion or between the cell and extracellular matrix (ECM) (cell-matrix adhesion) [8] **(Figure 1.2)**. There are five main types of CAMs: cadherins, the immunoglobulin superfamily, selectins, mucins and integrins [8] **(Figure 1.2)**. CAMs can assemble with adapter proteins into junctions, of which there are four types: tight junctions, gap junctions, cell-cell and cell-matrix adhesion junctions [8] **(Figure 1.2)**. Tight junctions connect cells and prevent diffusion or passing of molecules, while gap junctions allow the controlled exchange of small molecules between cells [8]. Cell-cell and cell-matrix junctions utilise distinct CAM proteins; cadherins, immunoglobulin superfamily, selectins and mucins, which are involved in cell-cell junctions, while integrins are involved in cell-matrix junctions [8] **(Figure 1.2)**.

Cell-cell adhesion proteins are found throughout epithelial tissue, except in the upper layer of keratinised epithelia [4, 8]. Specialised cell-cell junctions also exist, which consist of clusters of CAMs. For example, adherens junctions consist of cadherin proteins, which bind actin filaments of the cytoskeletal network [8] **(Figure 1.2)**. Actin filaments are composed of F-actin monomers which polymerise to form the filaments [9].

Epithelial basal cells are anchored to the basement membrane by cell-matrix adhesion proteins [2, 8] **(Figure 1.2)**. Integrin proteins are involved in cell-ECM

junctions which can form hemidesmosome junctions, which anchor the epithelium

to the basement membrane via intermediate filaments [2, 8] **(Figure 1.2)**.
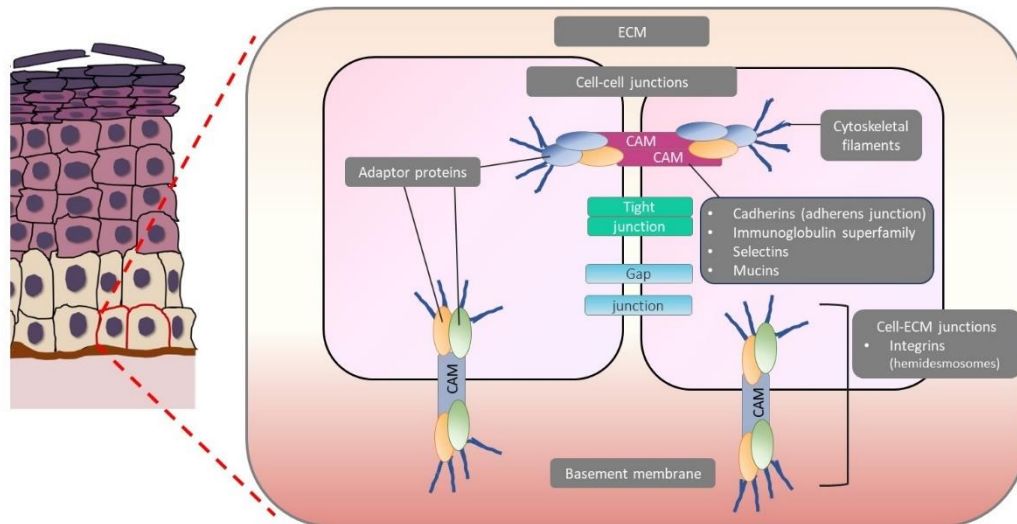
Figure 1.2. Cell adhesion.
Epithelial cells possess a tightly packed structure held together by four main types of junctions: tight junctions (prevent the passing of small molecules between cells), gap junctions (allow controlled passing of small molecules between cells), cell-cell adhesion junctions (anchor adjacent cells or to the cell exterior via the cytoskeleton) (e.g. adherens junction, which bind actin filaments) and cell-matrix junctions (anchor the cell to the extracellular matrix (ECM) or basement membrane via intermediate filaments) (e.g. hemidesmosomes). The junctions consist of cell adhesion molecules (CAMs): cadherins, immunoglobulin superfamily, selectins, mucins (involved in cell-cell adhesion) and integrins (involved in cell-matrix adhesion).

1.2 Molecular mechanisms in the development of epithelia

1.2.1 The role of p63 in epithelial development

Two of the most important features of stratified epithelia are its differentiated state and the maintenance of balance between cell proliferation and cell death. The p63 protein is a transcription factor, encoded by the *TP63* gene. p63 has been identified as being key in epithelial development and the first transcription factor to be expressed by basal cells in the developing epithelium [10, 11]. The importance of p63 function has been demonstrated in mice and humans. For example, p63-null mice have been shown to display incomplete or missing epithelial stratification and truncated limbs [12]. In humans, mutations of *TP63* result in ectodermal dysplasia-associated disorders, leading to cleft lip/palate and skin erosion [12].

The function of p63 can be complicated by alternative splicing events in the *TP63* gene **(Figure 1.3A)**. Due to the presence of alternative promoters, splicing of *TP63* can give rise to two main p63 protein isoforms; p53-like N-terminal transactivation domain (TAp63), while the other isoform lacks this domain (ΔNp63) [12, 13] **(Figure 1.3B and C)**. TAp63 is the first protein of the isoforms to be expressed and is thought to initiate K14 expression in the basal layer [13]. Epidermal-specific deletion of the ΔNp63 isoform in mice leads to skin erosions and impaired terminal keratinocyte differentiation [14, 15]. In addition, further alternative splicing events can give rise to three TAp63 isoforms and three ΔNp63 isoforms: α, β and γ [12, 13] **(Figure 1.3B and C)**.
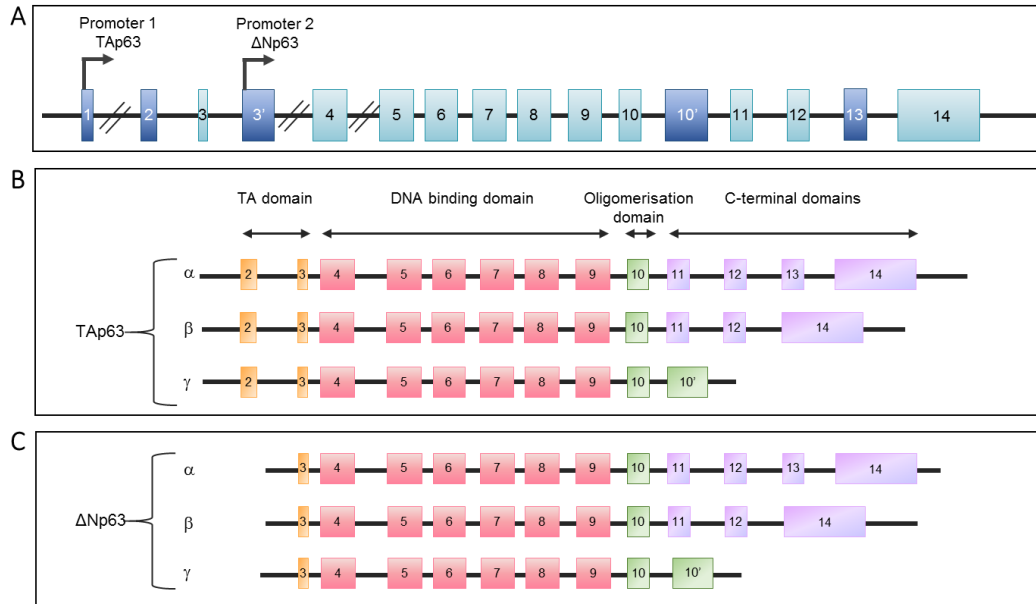
Figure 1.3. Schematic of the *TP63* gene, its splice sites and p63 isoforms.
(A) The *TP63* gene contains 14 exons and two promoters: promoter 1 allows transcription initiation of the *TAp63* isoform, while promoter 2 allows for Δ*Np63* transcription initiation. **(B)** The *TAp63* protein consists of four main domains, the TA (transactivation) domain (yellow exons), DNA binding domain (red exons), oligomerisation domain (green exons) and the c-terminal domain (purple exons). **(C)** The N-terminal TA domain of Δ*Np63* is truncated. (**B and C**) In both protein types, alternative splicing can produce three isoforms (α, β and γ).

The role of p63 in epithelial proliferation has been reported in many studies [13, 16, 17]. For example, the role of ΔNp63 in proliferation of primary human keratinocytes was highlighted after observing hypoproliferation in ΔNp63 knockdown cells [17]. Another study reported that ΔNp63 can maintain keratinocyte cell proliferation by inhibiting expression of the *CDKN1A* (cyclin dependent kinase inhibitor 1A) gene, which encodes the p21 protein [13]. The p21 protein is a cyclin-dependent kinase inhibitor, which inhibits cyclin D/CDK (cyclin dependent kinase) complexes, required for M phase cell cycle progression [13, 16]. p63 can also inhibit p21 expression via inhibition of NOTCH (neurogenic locus notch homolog), an upstream regulator of p21 [13, 16]. NOTCH proteins are a

family of transmembrane receptors involved in the terminal differentiation of keratinocyte cells [18].

1.2.2 The role of the NOTCH signalling pathway in epithelial development

The primary role of the NOTCH pathway is cell-cell signalling mediated by NOTCH receptors [19]. NOTCH receptors include NOTCH1-4 and their ligands include Jagged1/2 and Delta-like ligands1/3/4 [20]. A study by Dotto *et al*. supports this by demonstrating abnormal integrin expression in Notch1 deficient mice; expression was observed in the suprabasal layers of the epidermis [21]. Another study reported that Notch1 overexpression in mice leads to cell cycle arrest and differentiation by induction of p21 expression [22]. Therefore, it has been suggested that *NOTCH1* may be a tumour suppressor gene. Further evidence of this is that *Notch1* knockdown mice display an increase in oncogenic Wnt (wingless)/β-catenin signalling [23]. The role of NOTCH in tumour development will be discussed in more detail later.

1.2.3 The role of the WNT signalling pathway in epithelial development

WNT proteins are a family of 19 secreted glycoproteins, which function as part of the WNT pathway [24]. The WNT pathway can be divided into two pathways: canonical, which involves the β-catenin protein (WNT/β-catenin pathway), and non-canonical, which is β-catenin-independent [24]. The β-catenin protein is a transcription factor, which upon activation of the WNT pathway, translocates to the nucleus to induce transcription of WNT-target genes [24]. The WNT pathway has

a role in many cell processes including proliferation, survival, migration and differentiation [24]. During normal development of oral epithelia, the WNT/β-catenin pathway plays a prominent role in the development of oral organs including teeth, taste buds, lip and palate [25]. In the context of tumour development, aberrant WNT/β-catenin activation can induce transcription of genes involved in cell cycle progression, such as *CCND1*, which encodes cyclin D1 [24].

1.2.4 The role of the EGFR signalling pathway in epithelial development

EGFR (epidermal growth factor receptor) is part of the ERBB family of receptor tyrosine kinases, which includes ERBB-1 (EGFR/HER1), ERBB-2 (HER2), ERBB-3 (HER3) and ERBB-4 (HER4) [26]. EGFR is expressed by epithelial basal cells and expression is lost after maturation and differentiation of cells [27]. Activation of the EGFR signalling pathway results in cell cycle progression, proliferation and inhibition of apoptosis [26] **(Figure 1.4)**. EGFR consists of three domains; an extracellular ligand-binding domain; an intracellular tyrosine kinase domain and a hydrophobic transmembrane domain [26, 28] **(Figure 1.4)**. The hydrophobic domain allows for interaction between receptors, resulting in hetero- or homodimerization, while the intracellular domain is involved in phosphorylation of signalling proteins and EGFR itself **(Figure 1.4B)**.

EGFR ligands include the epidermal growth factor (EGF), transforming growth factor (TGF)-α, amphiregulin (AREG) and epiregulin (EREG) [26, 28]. Upon ligand binding, EGFR homo- or heterodimerisation with other members of its family occurs, resulting in autophosphorylation of the tyrosine kinase domains [26, 28] **(Figure 1.4B)**. This results in activation of two main protein axes: (i) mitogen-

activated protein kinases (MAPK) (RAS, RAF (Raf-1 proto-oncogene, serine/threonine kinase), MEK and ERK (extracellular signal-regulated kinase)), which are involved in cell proliferation, and (ii) activation of PI3K (phosphotidylinositol-3 kinase)-AKT (protein kinase B)-mTOR (mechanistic target of rapamycin), which are involved in cell cycle progression (**Figure 1.4B**).
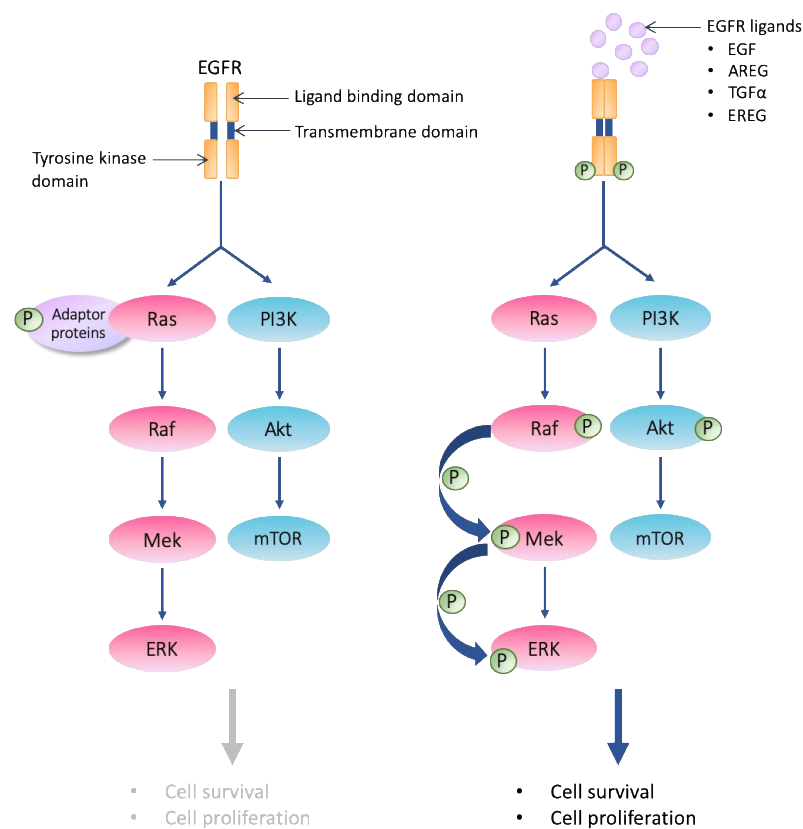
Figure 1.4. EGFR signalling promotes cell survival and proliferation.
(A) The EGF (epidermal growth factor) receptor consists of three domains: ligand binding domain (extracellular), a hydrophobic domain (transmembrane) and tyrosine kinase domain (intracellular). In the absence of EGFR-ligands the pathway is inactive. (B) Upon binding of EGFR ligands (EGF, AREG, TGF-α and EREG), the receptor undergoes homodimerization or heterodimerisation with either member of ERBB2-4. The tyrosine kinase domain undergoes autophosphorylation, leading to dephosphorylation of RAS (resistance to audiogenic seizures) adaptor proteins, initiating a phosphorylation cascade of the remaining MAPK proteins: RAF once phosphorylated, phosphorylates MEK (MAP kinase-ERK kinase), which in turn phosphorylates ERK. This cascade promotes cell proliferation. A second cascade activated by EGFR involves PI3K, which phosphorylates AKT, which activates mTOR to promote cell cycle progression.

1.3 Head and neck squamous cell carcinoma (HNSCC)

Deregulation in epithelial maintenance and development can result in a wide range of disorders, some of which are briefly mentioned in Section 1.2. One potential outcome of epithelial misregulation could be the development of head and neck squamous cell carcinoma (HNSCC).

HNSCC is the sixth most common cancer in the world with a five-year survival rate of 40-50% [29, 30]. Anatomical regions of the head and neck where HNSCC develops include the oral cavity, pharynx (nasopharynx, oropharynx (which includes the base of the tongue and tonsils) and hypopharynx), larynx, paranasal sinuses, nasal cavity and the salivary glands [23] **(Figure 1.5)**. The most common risk factors for HNSCC are exposure to tobacco and alcohol; over 75% of HNSCC cases can be attributed to these risk factors [31, 32]. There also exists a smaller group of HNSCC cases (5-30%) where the patient has no history of smoking [31]. The majority (~90%) of tobacco- and alcohol-associated tumours arise in the larynx and oral cavity (tongue and floor of the mouth) with a smaller proportion of cases developing in the oro- or hypopharynx [33, 34].
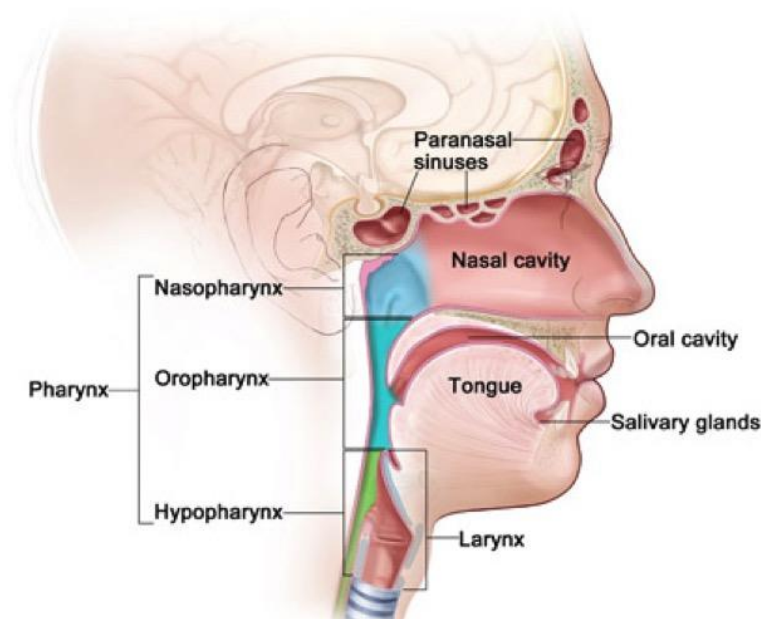
Figure 1.5. Anatomical regions of the head and neck.
Regions of the head and neck include the oral cavity, larynx, pharynx (nasopharynx, oropharynx and hypopharynx), paranasal sinuses, nasal cavity and salivary glands (Figure used with permission from the National Cancer Institute., 2012 [35]).

In addition to alcohol and tobacco, infection by the human papillomavirus (HPV) is a major risk factor. For HNSCC tumours which arise due to HPV are called HPV-positive(+) HNSCC tumours, while those which arise due to factors other than HPV are called HPV-negative(-). The majority of HPV(-) tumours are of non-oropharyngeal origin (94%) [23, 36, 37]. HPV(+) tumours arise either in the oropharynx (45.8%), specifically in the tonsils and base of the tongue, in the oral cavity (24.2%) or larynx (22.1%) [38, 39]. Investigation of the incidence of HPV(+) HNSCC cases in a meta-analysis of 12,163 HNSCC patients, revealed that 41% of patients in Europe were HPV(+); 60% of patients are HPV(+) in USA [40]. While in many developed countries the incidence of HPV(-) tumours is decreasing due to a decline in smoking habits, the incidence of HPV(+) HNSCC cases are increasing [37].

1.3.1 HPV classification

HPVs are small non-enveloped, double-stranded circular DNA viruses, which tend to infect squamous epithelial cells [20]. Over 210 types of HPVs have been identified thus far and can be categorised into five genera: alpha-, beta-, gamma-, mu- and nu-papillomaviruses [41, 42]. Infection by HPV can induce benign papillomas and persistent infection can lead to cell transformation and cancer [20, 41]. HPVs can be further grouped into low-risk and high-risk depending on the virus' ability to induce malignant cell transformation [20]. Overall, cases of HPV-induced HNSCC result from infection by a high-risk α-HPV. An example of a high-risk type is HPV16, which infects mucosal epithelia and is most frequently detected in HNSCC as it is present in more than 90% of oropharyngeal cancers [43].

1.3.2 HPV genome

The HPV genome can exist in two forms, either as an episome, where DNA replication is independent of the chromosomal DNA and is typically associated with non-malignancy, or HPV can be integrated into the host genome, giving rise to cancer [20]. The HPV genome encodes eight main proteins, which can be categorized into early (E) proteins and late (L) proteins. The early proteins include, E1, E2, E5, E6 and E7 and late proteins include, L1 and L2 [20] (**Figure 1.6A and B**). While E4 is named as an early protein, it is expressed later in the HPV life cycle, as discussed next.

The E1 and E2 proteins are expressed in the basal and spinous layers (or intermediate layer) [3] (**Figure 1.6C**). The early protein, E1 is an ATP (adenosine triphosphate)-dependent helicase, which along with E2 initiates viral replication

[20]. E2 is involved in segregation of the viral DNA plasmid during cell division and also functions as a transcription factor to regulate E6 and E7 expression [20, 44].

The E4 protein is expressed during the later stages of the HPV life cycle, in the spinous and granular layers [41]. E4 has been reported to be involved in viral DNA amplification by support of the G2 cell cycle phase [41].

E5 is expressed in low levels and is difficult to detect [45]. However, E5 transcript levels have been measured in HPV(+) HNSCC tumours and has been found to be variable [46]. E5 is an anti-apoptotic protein but also affects cell-matrix adhesion and epithelial differentiation, possibly through upregulation of the PI3K/AKT [47]. Interestingly, E5 expression has been found to correlate with EGFR expression [46].

The E6 and E7 proteins are expressed in the basal and spinous layers [3] (**Figure 1.6C**). High expression levels of E6 and E7 are key for cell transformation in HNSCC, as discussed later. E6 and E7 expression is maintained at very low levels until integration of its genome into the host genome [41]. Alternative mechanisms of increased E6/E7 expression have been reported. For example, knockdown of ERBB3 in oral SCC cell lines has been seen to decrease E6/E7 protein levels [48]. A second mechanism of increased E6/E7 expression has been described in primary human foreskin keratinocytes (HFKs) and cervical cancer cell lines, where the SP1 transcription factor induces HPV16 E6/E7 expression [49, 50]. In HFKs it has further been demonstrated that E6/E7 can induce human telomerase reverse transcriptase (hTERT) expression, which can lead to genomic instability and cell transformation [50].

The late proteins are expressed in the granular and keratinised (or superficial) epithelial layers [3, 20] **(Figure 1.6C)**. These proteins are structural capsid proteins; L1, the late major protein, is synthesised in the cytoplasm and then transported to the nucleus, where it has a role in the packaging of viral particles [20]. The L2 minor capsid protein binds the viral DNA in the nucleus, in order to recruit L1 for new viral particles to be assembled [20].
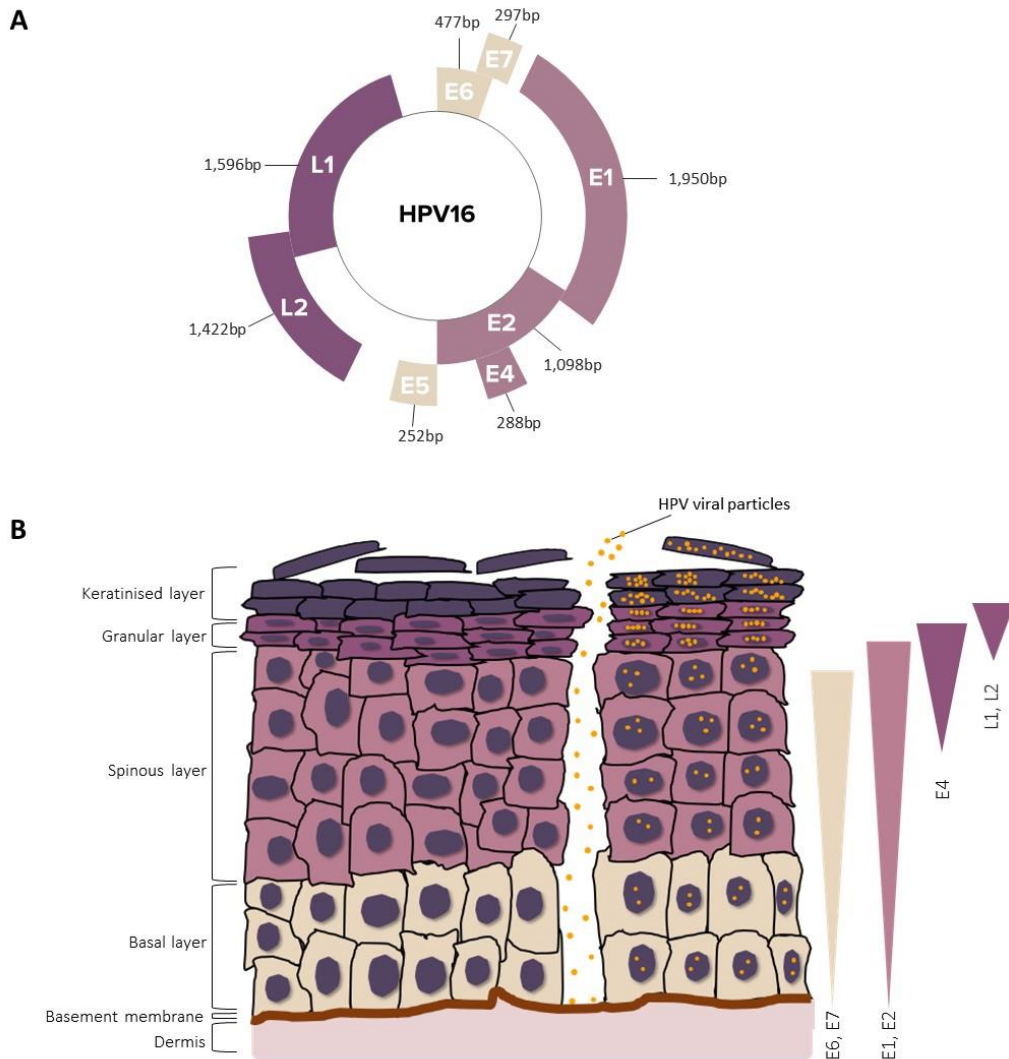
Figure 1.6. HPV16 genome and life cycle.
(A) The size of the HPV16 genome is ~7,900bp. The genome contains six early genes: E1 (1,950bp), E2 (1,098bp), E4 (288bp), E5 (252bp), E6 (477bp) and E7 (297bp) and two late genes: L1 (1,596bp) and L2 (1,422bp). **(B)** HPV can infect epithelia via small abrasions or cuts, where the viral particles infect the basal cells, as they hold the potential to proliferate and migrate upwards through the layers, eventually aiding viral release. The infected basal cells will be forced to proliferate and remain undifferentiated until viral amplification is completed. The E1, E2 and E4 proteins are expressed predominantly in the spinous layer of the epithelium, while the E6 and E7 proteins are expressed primarily in the basal layer [51]. The L1 and L2 proteins are expressed in the granular layer [51]. (Figure A recreated and adapted with permission from Rampias *et al*., 2014 [20]. Figure B recreated and adapted with permission from Roberts, 2015 [3]).

1.3.3 HPV life cycle

The route of HPV infection is not well described in HNSCC, nor are the mechanisms of HPV transformation, however these processes are better understood in cervical cells. Cervical cancer is also a mucosal SCC and nearly all cases (99.7%) are associated with HPV infection [52]. Therefore, cervical SCC can be a useful model to compare with and gain insights into the mechanism of HPV-driven carcinogenesis in HNSCC [52-54]. HPV enters cervical epithelial tissue via small abrasions or lesions in the mucosa [42] **(Figure 1.6B)**. Epithelial basal cells possess the highest proliferative potential; therefore, the virus preferentially infects these cells [4, 42]. Viral DNA enters the nucleus of the undifferentiated basal cell, where E1 and E2 gene transcription and protein synthesis occurs [42]. E1 and E2 proteins can then initiate replication of the viral episome to ~50-100 copies per cell [42]. The E6 and E7 viral oncoproteins are also expressed, which can hijack regulators of the host DNA replication machinery, namely p53 and pRB (retinoblastoma protein), respectively. The combined action of both E6 and E7 forces the host cell to remain in the cell cycle, uncoupling the proliferation and differentiation processes [42]. Although terminal differentiation of the infected cells is delayed by E6 and E7 expression, eventual differentiation in the upper layers of the epithelium results in expression of late genes, including E4, which promotes viral genome amplification [42, 55]. In addition, expression of the L1 and L2 capsid proteins facilitate assembly of infectious virions, which are sloughed off with dead cells on the surface of the skin [42, 55].

1.4 The molecular basis of HPV-induced HNSCC

In normal non-proliferating cells, both pRB and p53 act as tumour suppressor proteins (**Figure 1.7**). pRB is bound to the E2F (E2 factor) transcription factor, inhibiting E2F-induced gene transcription [29]. Upon mitogen signalling, cyclin D1/CDK-4 or -6 hyperphosphorylate pRB, releasing E2F and initiating transcription of E2F target genes, which are involved in the progression of the cell cycle into S phase [29]. E2F target genes encode for proteins such as DNA polymerase α, PCNA (proliferating cell nuclear antigen) and cyclins A and E [56] (**Figure 1.7**). Upon receiving senescence and differentiation signals, the p16^{INK4A} protein (encoded by *CDKN2A* (cyclin dependent kinase inhibitor 2A) inhibits cyclin D1/CDK4- or -6, blocking G1/S progression [29] (**Figure 1.7**). For progression of G2 phase into M phase, the p53 protein plays a central role in the regulation of genes involved in growth arrest, DNA repair and apoptosis [18]. p53 is part of a protein family, which also includes p63 and p73 [12]. One of the roles of p53 is to induce cell cycle arrest or apoptosis upon recognition of DNA damage, and this is mediated by the p21 protein [29] (**Figure 1.7**). p21 inhibits cyclin E/CDK2 complexes, which normally hyperphosphorylate pRB, inhibiting G1/S progression [29]. p21 also inhibits cyclin A/CDK1, which is required for S phase progression [29] (**Figure 1.7**).
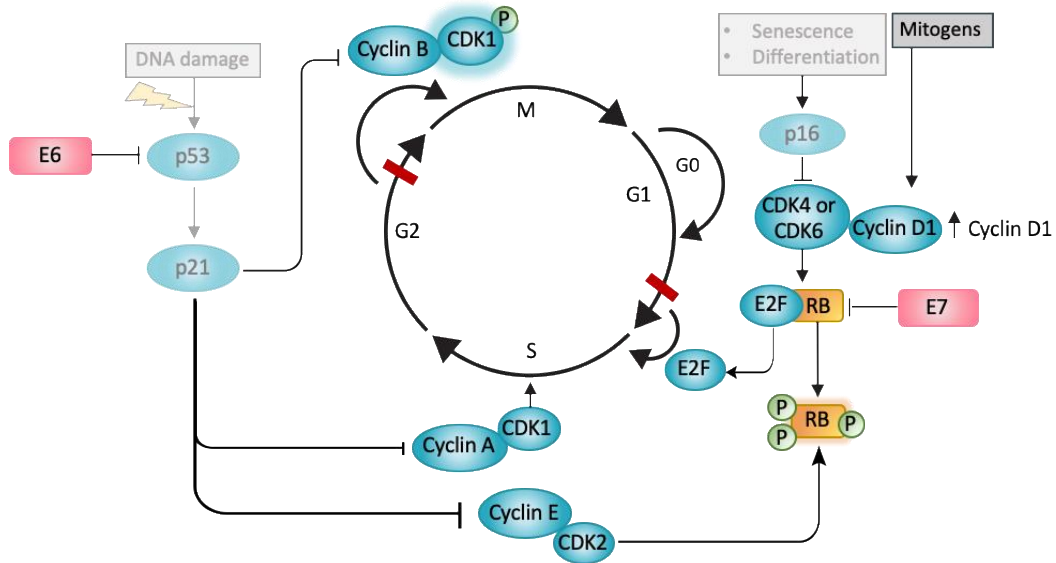
Figure 1.7. Cell cycle regulation.
Progression of the cell cycle depends on two checkpoints (red bars); G1/S and G2/M. The checkpoints are regulated by pRB and p53, respectively. G1/S regulation: upon mitogen signalling, cyclin D1/CDK-4 or -6 complexes hyperphosphorylate pRB, releasing the E2F transcription factor to induce gene transcription of genes involved in S phase progression. The p16^{INK4A} protein (encoded by *CDKN2A*) is involved in inhibition of CDK-4, -6/cyclin D1 complexes, preventing hyperphosphorylation of pRB and stabilising E2F. In HPV(+) HNSCC, the E7 viral oncoprotein binds pRB, which releases E2F allowing transcription induction. In HPV(-) HNSCC, *CDKN2A* is often mutated, resulting in loss of p16^{INK4A}, leading to release of E2F. G2/M regulation: Upon detection of DNA damage, p53 induces the p21 protein, which inhibits cyclin A/CDK1 and cyclin E/CDK2 from hyperphosphorylating pRB, thereby blocking E2F-mediated transcription. p21 also inhibits cyclin B/CDK 1 complex, which is required for progression to M phase. In HPV(+) HNSCC, the E6 viral oncoprotein binds p53, targeting it for ubiquitin-mediated degradation. In HPV(-) HNSCC, the *TP53* gene which encodes for p53, is often mutated, resulting in loss of functional p53. (Figure adapted with permission from Leemans CR *et al.*, 2018 [29]).

Upon HPV infection, actions of the viral oncoproteins E6 and E7 can deregulate the cell cycle by inhibiting p53 and pRB, respectively [42]. In malignant cells the E6 and E7 proteins of high-risk HPVs are considered to be the strongest protein effectors of cancer development due to their ability to transform keratinocytes [42]. E6 acts to limit the pro-apoptotic function of p53 by forming a complex with a host protein called, E6-AP (E6 associated protein) and p53 [20, 57, 58]. The complex functions as a p53-specific ubiquitin ligase to induce p53 degradation, leading to aberrant cell cycle progression [20] **(Figure 1.7)**. Another way in which E6 may contribute to cell transformation is by binding to the hDlg (human disks large homolog 1) protein, which is a PDZ domain-containing protein [59]. The hDlg protein is a homologue of the drosophila dlg protein [59]. Mutations of the *dlg* gene have demonstrated a loss in cell polarity in epithelial cells and neoplastic proliferation [59, 60]. Furthermore, the hDlg protein is thought to have a role in cell-cell adhesion and E6 binding may contribute to this [61].

In addition to E6, the E7 oncoprotein can promote cell cycle progression in several ways. E7 can disrupt the tumour suppressor protein, pRB, leading to disruption of the interaction of pRB with E2F and inducing aberrant transcription of E2F target genes [56, 62] **(Figure 1.7)**. E7 can also target pRB for degradation by hijacking the cullin 2 E3 ubiquitin ligase complex [20]. Another way E7 can promote cell cycle progression is by inhibiting cyclin-dependent kinase inhibitors, such as p21 and as a result, aberrantly activating cyclin-CDK complexes leading to cell cycle progression [63]. In addition, E7 can disrupt histone modifications such as acetylation and methylation by interacting with histone acetyltransferases, deacetylases, methyltransferases and demethylases [20]. The function of histone modifications is discussed in detail later.

20

Integration of the HPV genome can also contribute to malignancy, as the E2 and E5 genomic sequences, which have a role in repressing E6 and E7 expression, are often lost from the episome [20, 41, 64]. A study conducted by The Cancer Genome Atlas of 36 HPV(+) tumours revealed HPV integration in most tumours and association with genomic amplifications of the host genome [36]. However, such integration events were not recurrent and so a single driver could not be identified. Other studies have reported loss of function of tumour suppressor genes as a result of HPV integration, such as the *RAD51* (RAD51 recombinase) gene, which encodes an enzyme involved in DNA repair [65]. In addition, amplification of oncogenes such as *NR4A2* (nuclear receptor subfamily 4 group A member 2) have also been reported [65].

1.5 Genetic alterations in HNSCC

In recent years, genome-wide and high-throughput DNA sequencing of tumours has become more common. These methods allow large cohorts of tumour samples to be investigated for genetic alterations such as gene mutations, amplifications or deletions, gene expression changes and changes in protein function. In HNSCC, the majority of genetic alterations occur in tumour suppressor genes and genes involved in cell survival, proliferation and metastasis.

The largest study of genetic alterations in HNSCC was conducted by The Cancer Genome Atlas (TCGA) consortium in 2015, which consisted of 279 tumours [36]. The study investigated the HPV status, copy number alterations, somatic mutations and gene expression changes in HNSCC tumours. The cohort consisted of tumours derived from the oral cavity (172/279, 62%), oropharynx (33/279, 12%) and larynx (72/279, 26%). The HPV status of the tumours was determined by RNA-sequencing

to investigate expression levels of the viral oncogenes, *E6* and *E7* [36]. The majority of the tumours were HPV(-) (243/279, 87%) and agreeing with previously published data nearly all of the HPV(-) tumours were of non-oropharyngeal origin [23, 36, 37].

The TCGA study was followed up three years later with the Pan-Atlas study, in which the cohort was expanded to 522 HNSCC tumours [66]. In the Pan-Atlas study, 33 different cancer types were studied, including five types of SCCs (HNSCC, lung, oesophageal, cervical and bladder) [66]. The aim of this study was to compare features of SCC with other cancers. Genes found to be significantly altered in SCC included stem cell markers (*SOX2* (SRY-box 2)), mitogenic growth and cell cycle related genes (*PDGFRA* (platelet derived growth factor receptor alpha), *IGF1R* (insulin like growth factor 1 receptor), *CDK6*, *RAC1* (AKT serine/threonine kinase 1), *MAPK1*, *EPHA2* (EPH receptor A2), and *CREBBP* (CREB binding protein), PI3K signalling genes (*AKT1/3*), NF-kB signalling genes (*REL* (REL proto-oncogene) and *TRAF3* (tumour necrosis factor receptor associated factor 3), genes related to squamous differentiation (*FAT* (FAT atypical adherin)-*1/2*, *ROBO* (roundabout)-*1*, *ZNF* (zinc finger protein)-*750*, *JUB* (Ajuba LIM protein), *NOTCH1*, and *TP63*) and immune system related genes (*PD-L1* (programmed cell death ligand 1) and *B2M* (β2 microglobulin)) [66].

Seven genome-wide studies including the two TCGA studies are summarised in **Table 1.1**, showing the most commonly altered genes in HNSCC. Four studies reported genetic alterations while discriminating by HPV status, generating two datasets for HPV(+) and HPV(-) tumours [36, 67-69]. Three studies reported alterations in HNSCC tumours without discriminating for HPV and will be referred

to as HNSCC tumours [66, 70, 71]. Genetic alterations and gene expression changes

identified in all three datasets are discussed below.

Table 1.1. Summary of genetic alterations reported by seven genome-wide studies of HNSCC tumours.

Gene	TCGA, 2015	Seiwert et al., 2015	Stransky et al., 2011	Agrawal et al., 2011	Campbell et al., 2018	Pickering et al., 2013	Pickering et al., 2014
HPV status	HPV(-)			N/A			
Tumour type	HNSCC	HNSCC	HNSCC	HNSCC	HNSCC	OSCC	Tongue
n =	243	69	64	28	522	35	44
TP53	84% (M)	81% (M)	72% (M)	79% (M)	67% (M)	66% (M)	71% (M)
CDKN2A	57% (M, D)	33% (M, D)	17% (M, D)	71% (M)	19% (M)	74% (D); 23% (A)	5% (M); 61% (D)
PIK3CA	34% (M, A)	13% (M, A)	6% (M)	7% (M)	17% (M)	11% (M); 9% (A)	11% (M); 75% (A)
FADD	32% (A)	N/A	0%	0% (M)	0.6% (M)	N/A	56% (A) (CCND1/FADD)
FAT1	32% (M, D)	N/A	14% (M)	0% (M)	21% (M)	34% (M); 14% (D)	16% (M); 41% (D)
CCND1	31% (A)	13% (A)	22% (A)	0% (M)	0.4% (M)	22% of 38 (A)	56% (A) (CCND1/FADD)
NOTCH 1		16% (M)	16% (M)	25% (M)	17% (M)	9% (M)	21% (M)
NOTCH 2	29% (M, D)	N/A	13% (M)	0% (M)	4% (M)	3% (M); 6% (D)	N/A
NOTCH 3		N/A	3% (M)	0% (M)	N/A	N/A	N/A
TP63	19% (A)	N/A	13% (M)	4% (M)	2% (M)	9% (M); 26% (A)	N/A
EGFR	15% (M, A)	12% (A)	3% (M)	0% (M)	3% (M)	17% (A)	40% (A)
CASP8	11%	N/A	9% (M)	4% (M)	10% (M)	9% (M)	9% (M)
HRAS	5% (M)	N/A	5% (M)	7% (M)	6% (M)	9% (M)	N/A
AJUBA	7%	N/A	0%	0% (M)	N/A	N/A	0%
BIRC2	7% (A)	N/A	N/A	N/A	N/A	N/A	N/A
HPV status	HPV(+)						
n =	36	51	10	4			
E6/E7	100%	100%	100%	100%			
TP53	3%	N/A	0% (M)	0% (M)			
CDKN2A	0%	N/A	0% (M)	0% (M)			
PIK3CA	56% (M, A)	22% (M)	2% (M)	25% (M)			
FADD	6%	N/A	0% (M)	0% (M)			
FAT1	3%	N/A	0% (M)	0% (M)			
CCND1	3%	N/A	0% (M)	0% (M)			
NOTCH 1		N/A	1% (M)	0% (M)			
NOTCH 2	17% (M)	N/A	0% (M)	0% (M)			
NOTCH 3		N/A	1% (M)	0% (M)			
TP63	28% (A)	16% (M, A)	0% (M)	0% (M)			
EGFR	6% (M)	N/A	0% (M)	0% (M)			
CASP8	3%	N/A	0% (M)	0% (M)			
HRAS	0%	N/A	1% (M)	0% (M)			
AJUBA	0%	N/A	0% (M)	0% (M)			
E2F1	19% (A)	N/A	0% (M)	0% (M)			
SOX2	N/A	12% (A)	0% (M)	0% (M)			
BIRC2	3% (A)	N/A	N/A	N/A			

The table consists of seven studies in which genetic alterations in HPV(-), HPV(+) and HNSCC tumours are characterised. Genetic alterations include mutations (M), deletions (D) and amplifications (A). Genes in red indicate transcription factor genes. Data unavailable for certain genes is denoted by 'N/A'. Adapted from Beck and Golemis., 2016 [23].

1.5.1 Mutations of the *TP53* gene

The tumour suppressor gene *TP53* was the most commonly mutated gene in the TCGA cohort of HPV(-) tumours, agreeing with the remaining six studies where between 66% and 81% of HPV(-) or HNSCC tumours displayed *TP53* loss-of-function mutations [36, 66-71] **(Table 1.1)**. The *TP53* gene encodes for the p53 tumour suppressor protein, the function of which is described previously in Section 1.4. Mutations of the *TP53* gene were rarely reported in HPV(+) HNSCC tumours; the TCGA reported 1/36 HPV(+) tumours harbouring a *TP53* mutation [36]. The importance of the *TP53* gene has been highlighted by the association between *TP53* mutations and reduced patient survival [36] **(Table 1.1)**.

1.5.2 Amplification of chromosome 3q (*TP63*, *SOX2* and *PIK3CA*)

The *TP63* gene is located on chromosome 3q and is amplified in 19-26% of HPV(-) tumours and 16-28% of HPV(+) tumours [36, 66-70, 72] **(Table 1.1)**. Chromosome 3q also contains the *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) and *SOX2* genes [58, 67, 73]. As a result of focal amplifications, all three genes are amplified in HNSCC; *TP63* and *PIK3CA* amplifications are observed in both HPV(+) and HPV(-) tumour types, while *SOX2* amplifications are observed in HPV(+) tumours [58, 67, 73] **(Table 1.1)**.

The *PIK3CA* gene encodes a kinase catalytic subunit of PI3K called p110α, which is part of the PI3K/AKT/mTOR pathway. Upon ligand activation of EGFR, phosphorylation of p110α occurs, which in turn phosphorylates the serine-threonine kinase, AKT, hyperactivation of which is involved in development and

progression of various cancers [74] **(Figure 1.4)**. The pathway then triggers downstream events associated with tumour cell survival and transformation [74]. For example, the interleukin-6 (IL-6) protein is a mediator of the PI3K/AKT/mTOR pathway and is often overexpressed in cancers, including HNSCC, and is associated with reduced overall survival of HNSCC patients [75].

In HPV(+) HNSCC cells, it has been reported that HPV can bind to heparan-sulfonated proteoglycans to enter keratinocyte cells [76]. It is thought that this stimulates the PI3K/AKT/mTOR pathway via activation of $\alpha6\beta4$ integrins [76]. In addition, it has been reported that the viral oncoproteins E6 and E7 can promote viral survival by activation of AKT and mTOR [77, 78]. Furthermore, the ability of E7 to inactivate pRB has been reported to upregulate AKT protein expression and E7 can promote sustained PI3K/AKT signalling by inhibiting their dephosphorylation [77, 78].

SOX2 is a transcription factor commonly involved in embryonic development and it has been shown to be important in the development and maintenance of squamous epithelia [79]. The *SOX2* gene has been reported as being oncogenic in many SCCs, including oesophageal carcinomas [79]. Interestingly, both SOX2 and p63 transcription factors have been shown to co-localise at intragenic regions, indicating co-regulation between the two transcription factors [79].

1.5.3 Mutations of the *NOTCH* gene family

Inactivating mutations in *NOTCH* (neurogenic locus notch homolog)-*1-3* are observed in both HPV(-) and HPV(+) tumours **(Table 1.1)**. Of the three *NOTCH*

genes, *NOTCH1* exhibits mutations most frequently in HPV(-) HNSCC tumours **(Table 1.1)** [36, 66-71]. NOTCH1 signalling is involved in terminal differentiation of primary keratinocyte cells [80]. In addition to *NOTCH1* mutations, EGFR can also block NOTCH1-mediated keratinocyte differentiation [80]. In addition, p63 is also able to block NOTCH1 signalling in cells of the basal layer, inhibiting p21 and allowing cell cycle progression [13, 16, 18].

1.5.4 Mutations of the *CDKN2A* gene

The second most frequently altered gene in HPV(-) HNSCC tumours is the tumour suppressor gene *CDKN2A*, which displays alterations including mutations and deletions rendering the gene inactive [36, 66-71] **(Table 1.1)**. Between 17% and 97% of HPV(-) tumours display *CDKN2A* alterations and no HPV(+) tumours have been reported to display any such alterations **(Table 1.1)** [36, 66-71]. As discussed earlier, the *CDKN2A* gene encodes for the p16^{INK4A} protein, which is involved in senescence and differentiation signalling **(Figure 1.7)**. Due to the pRB protein being a negative regulator of p16 and pRB being inactive in HPV(+) tumours, the p16 protein is often used as a marker for HPV [36, 67-69, 81, 82]. In addition, it has been shown in cervical carcinoma cell lines that E7 expression can induce expression of the histone demethylase protein, KDM6B which in turn can induce p16^{INK4A} expression [83].

1.5.5 Amplification of chromosome 11q13 (*CCND1*, *FADD*, *CTTN*)

Amplification of the 11q13 chromosome containing the *CCND1*, *FADD* (fas associated via death domain) and *CTTN* (cortactin) genes is frequently observed in

HNSCC tumours and the majority of amplifications are observed in HPV(-) tumours [36, 67-69, 84, 85] **(Table 1.1)**. *CCND1*, which encodes cyclin D1, has been reported to be an oncogene in many cancers including HNSCC, breast, bladder, liver and lung cancers [86]. The *FADD* (fas associated via death domain) gene is an adaptor molecule involved in apoptosis by interaction with cell surface death receptors, inducing caspase recruitment to initiate apoptosis [86]. FADD expression is linked to HNSCC metastasis and decrease in patient survival [86]. Amplification of *CTTN*, which encodes cortactin (an actin binding protein), has been observed in several cancers, including HNSCC and has a role in tumour cell migration and invasion [86].

1.5.6 Amplification of chromosome 11q22 (*BIRC2* and *YAP1*)

The TCGA have reported chromosome 11q22 as being frequently amplified in HNSCC, which has also been observed in lung, cervical and oesophageal SCC [36, 66] **(Table 1.1)**. This chromosome location contains the *BIRC2* (baculoviral IAP repeat-containing protein 2) and *YAP1* (yes-associated protein 1) genes [36]. The *BIRC2* gene encodes for a member of the cellular inhibitor of apoptosis family of proteins (c-IAP1), which inhibits caspase activity including CASP8 (cysteine-dependent aspartate-directed proteases) [23]. The *CASP8* gene, which encodes for the pro-apoptotic enzyme caspase 8, is also mutated in HNSCC more frequently in HPV(-) tumours than in HPV(+) [36, 68, 69] **(Table 1.1)**.

The *YAP1* gene encodes the YAP protein. *YAP1* is often observed as being overexpressed in many cancers, including prostate, ovary and colon [87]. YAP is part of the Hippo signalling pathway, which is involved in the regulation of organ

growth and cell-cell contact signalling [88]. In HNSCC, YAP has been reported to be upregulated compared with surrounding benign tissues, furthermore, YAP expression has been observed more strongly at HNSCC tumour margins [87].

1.5.7 Amplification of the *EGFR* gene

Recurrent focal amplifications of the *EGFR* gene were observed primarily in HPV(-) tumours (3-40%), which has also been confirmed by other studies [36, 67, 89, 90] **(Table 1.1)**. In HNSCC, *EGFR* amplification is commonly observed with *PIK3CA*, *AKT1*, *RPS6KB1* (ribosomal protein S6 kinase B1) and *MYC* (MYC proto-oncogene, bHLH transcription factor) gene amplifications [70]. Amplification of genes of other tyrosine kinase receptors- Erb tyrosine kinases (*ERBB2*, *ERBB3*, *ERBB4*) have also been observed [36]. Despite *EGFR* amplification occurring in 3-40% of the HNSCC tumour cohorts investigated, EGFR protein overexpression is observed in up to 90% of HNSCC tumours [70, 91] **(Table 1.1)**. Therefore, EGFR expression may be increased by an alternative, mutation-independent mechanism.

1.5.8 Mutations of the *AJUBA* and *FAT1* genes

AJUBA (ajuba LIM protein) is altered exclusively in of HPV(-) tumours, while *FAT-1* is altered in 14-57% of HPV(-) tumours and in 3% of HPV(+) tumours [36, 66, 68-71]. Both AJUBA, which is a tumour suppressor protein, and FAT1 are involved in the WNT and Hippo pathways [29]. AJUBA is a LIM domain-containing protein, which forms part of a larger family of LIM proteins [29]. In primary keratinocyte cells, through interaction with the WNT/β-catenin pathway, AJUBA is involved in cell-cell adhesion [29]. The AJUBA protein has also been

reported to associate with LATS (large tumour suppressor kinase)-1/2 of the Hippo pathway [29]. LATS proteins are kinases, which are negative regulators of the Hippo pathway, as they mediate inhibition of the downstream family of transcription factor effectors, TEAD (TEA domain transcription factor)-1-4 [88, 92]. The FAT1 protein is part of the family of cadherin proteins, also involved in cell adhesion [29]. In addition, FAT1 has been linked to the WNT/β-catenin pathway, as FAT1 can bind to β-catenin to inhibit β-catenin-mediated gene transcription [29]. Recently, it has been reported that loss of FAT1 results in YAP activation in HNSCC [93].

Therefore, analysis of genetic alterations in the datasets outlined in Table 1.1 reveal that alterations are observed both in HPV(-) and HPV(+) HNSCC tumours. While the signature of gene alterations differ in HPV(+) tumours compared to HPV(-), these alterations can affect the same pathways in both tumour types (**Figure 1.8**).

In addition to the complicated molecular landscape, there are clinical differences between patients suffering from HPV(-) and HPV(+) HNSCC. For example, patients with HPV(+) HNSCC tumours tend to be younger than those diagnosed with HPV(-) (median age: 53.5 and 57 years old, respectively). Furthermore, prognosis is often poorer for patients suffering from HPV(-) tumours compared to those with HPV(+) tumours [94]. Therefore, the treatment of HNSCC remains challenging due to tumour and clinical heterogeneity.
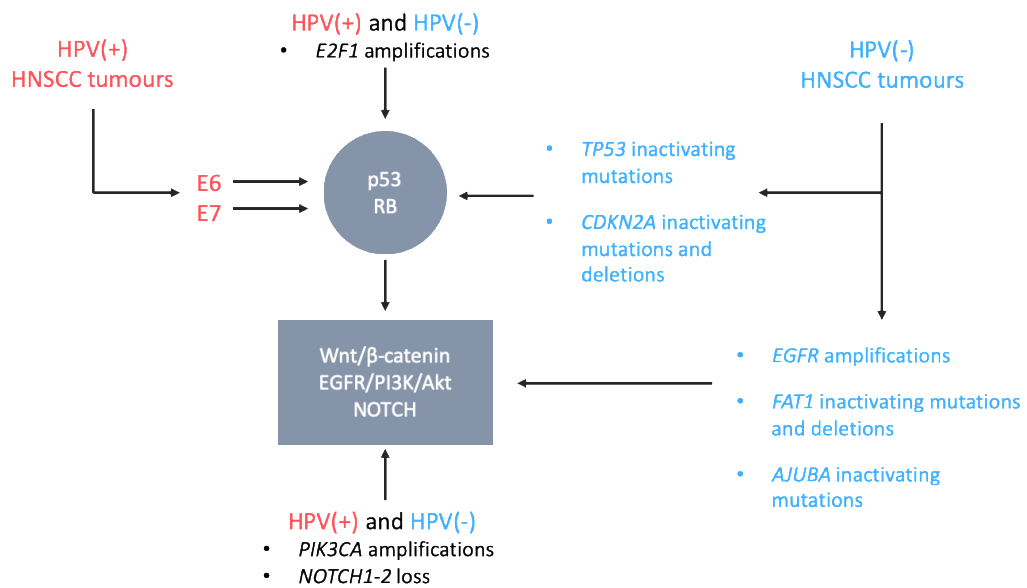
Figure 1.8. Genetic and pathway alterations in HPV(-) and HPV(+) HNSCC tumours. The pathways deregulated in HPV(+) and HPV(-) HNSCC tumours can be common. In HPV(+) tumours, the viral oncoproteins E6 and E7 can hijack the p53 and pRB proteins, respectively, leading to deregulation of the cell cycle. In HPV(-) tumours, inactivating mutations and/or deletions in *TP53* and *CDKN2A* genes can lead to deregulation of the p53 and pRB proteins. *EGFR* amplifications can lead to deregulation of the EGFR/PI3K/AKT pathway. Loss of *FAT1* and *AJUBA* genes through inactivating mutations and deletions can lead to deregulation of the WNT/β-catenin pathway. In both HPV(+) and HPV(-) tumours, *E2F1* amplifications can deregulate p53 and pRB. *PIK3CA* amplifications and loss of *NOTCH1-3* genes can lead to deregulation of the WNT/β-catenin, EGFR/PI3K/AKT and NOTCH pathways.

1.5.9 Detection of HPV

To determine the HPV status of HNSCC tumours, one of the most commonly used methods is immunohistochemical staining for p16^{INK4A}. Levels of p16^{INK4A} can increase as a result of HPV infection, as the inactivated pRB protein is a negative regulator of p16^{INK4A} [63, 95]. In addition, as mentioned earlier, E7 expression can induce KDM6B expression, leading to increased high p16^{INK4A} expression [83]. PCR (polymerase chain reaction) or immunohistochemistry can also be used to detect HPV DNA and determine HPV status [23]. However, the disadvantage of DNA-based methods is that the presence of viral DNA may not mean cell transformation and cancer, as persistent HPV infection is required for this [96]. In fact, nearly all HPV infections (~90%) are cleared in 6-24 months [96, 97]. In addition, the presence of HPV DNA does not mean the virus is driving the tumour, other factors may induce lead to tumour formation. Therefore, detection of HPV DNA can be useful in determining the HPV status of the tumours but is not very predictive of HNSCC development. A more predictive marker of HPV transformation is the detection of RNA of the viral oncoproteins E6/E7, as the presence of RNA indicates a transcriptionally active viral genome [36, 97].

1.6 HNSCC therapies

HNSCC tumours are currently treated with a combination of therapies including surgery, radiation and adjuvant chemotherapy [23]. Early stage tumours are typically treated with surgery and radiation, while more advanced tumours are treated with a combination of all three therapies [23, 98]. Most recent progress in the development of HNSCC therapies include targeted immunotherapies to combat the immunosuppressive nature of HNSCC [99]. However, despite these advances, the five-year HNSCC survival rate reported in 2015 was 25-40% for HPV(-) and 70-80% for HPV(+) tumours and these numbers have not changed in the past three decades [100, 101].

As discussed earlier, recurrent focal amplifications of *EGFR* are observed in HNSCC, with up to 90% of tumours displaying EGFR protein overexpression [36, 102, 103]. The importance of EGFR overexpression in HNSCC is highlighted by its association with reduced overall survival [36]. Therefore, inhibition of EGFR was introduced as one of the therapies for HNSCC patients [104].

1.6.1 Monoclonal antibodies and tyrosine kinase inhibitors as therapies

In 2006, a monoclonal antibody called cetuximab, which inhibits the EGFR was approved by the Food and Drug Administration (FDA) as a therapy for HNSCC [104]. Cetuximab targets the EGFR extracellular domain and blocks EGFR by competition with ligands for the ligand-binding domain **(Figure 1.9)**. Furthermore, in HNSCC cells it has been reported that inhibition by a monoclonal antibody can result in internalisation of the receptor, preventing ligand binding [26]. This can

lead to inhibition of cell growth and induction of apoptosis [28]. A second mechanism of action of cetuximab is antibody-dependent cell-mediated cytotoxicity (ADCC) [105]. ADCC is an EGFR-independent mechanism, which involves the Fc region of the antibody interacting with Fcγ receptors on immune effector cells [105, 106]. This triggers an immune response rendering tumour cells more susceptible to cytotoxicity [105, 106].

A second class of EGFR inhibitors are tyrosine kinase inhibitors, which inhibit the intracellular tyrosine kinase domain **(Figure 1.9)**. An example of this is an experimental drug called, AG1478 (tyrphostin). AG1478 and mAB 806 (an experimental EGFR epitope-binding drug), have been used together to successfully inhibit growth of a xenograft composed of an oral SCC cell line [107]. It was demonstrated that AG1478 can induce a conformational change of EGFR, allowing mAb 806 to bind with greater affinity [107].
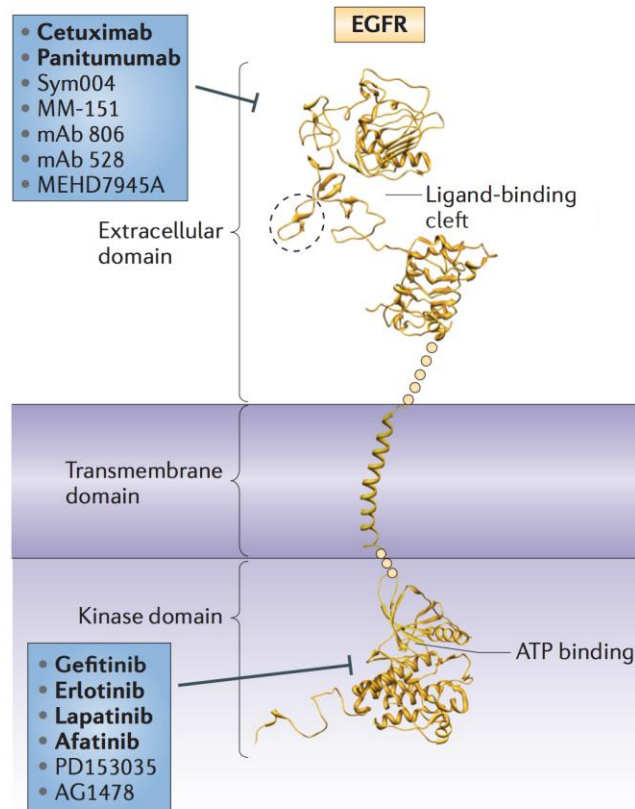
Figure 1.9. EGFR as a therapeutic target.

EGFR contains two main targetable domains, (i) the extracellular domain, which harbours the ligand-binding domain, and (ii) the intracellular domain, which upon ligand binding and receptor activation, is involved in receptor dimerization. Extracellular domain: all therapies of the extracellular domain are monoclonal antibodies. **Cetuximab**, **panitumumab**, are antibodies which block ligand-binding. **Sym004** and **MM-151** consist of a mixture of antibodies, which bind to non-overlapping EGFR epitopes in order to increase efficacy of the therapy. **mAb 806** is EGFR epitope-binding. **mAb 528** inhibits the ligand-binding domain. **MEHD7945A** binds both EGFR and ERBB3 to prevent heterodimerisation. Intracellular domain: all therapies of the intracellular domain are tyrosine kinase inhibitors. **Gefitinib**, **erlotinib** and **PD153035** are reversible tyrosine kinase inhibitors, which compete for the kinase domain with ATP to inhibit receptor activation by phosphorylation. **Lapatinib** inhibits both EGFR and ERBB2 receptors. **Afatinib** binds irreversibly to the kinase domain. **AG1478** is an EGFR-specific molecule, which competes reversibly for the kinase domain and induces a conformational change. Therapies approved for cancer treatment by the FDA are in bold within the figure. (Figure adapted with permission from Tebbutt and Pedersen., 2013 [108]).

1.6.2 Cetuximab resistance and alternative therapies

Despite EGFR overexpression being identified in 90% of HNSCC tumours, only 10-20% of patients respond to the cetuximab treatment [109, 110]. Moreover, it has been demonstrated that cetuximab resistance cannot be predicted by EGFR copy number [110, 111]. This indicates that the EGFR pathway does not function independently and therefore, elucidating co-regulators may offer alternative or complementary therapies to cetuximab.

Mechanisms of cetuximab resistance have been investigated by many groups. For example, it has been demonstrated that *ERBB2* gene overexpression has been associated with resistance to cetuximab [112, 113]. Furthermore, the use of lapatinib, an EGFR and ERBB2 tyrosine kinase inhibitor, in conjunction with chemoradiation has been shown to increase overall survival by 11% [114, 115]. Another mechanism of cetuximab resistance involves epithelial to mesenchymal transition (EMT). During EMT, phenotypic changes that appear to facilitate metastasis include loss of cell junction proteins and has been associated with gefitinib resistance [116]. In addition, upregulation of EMT markers such as vimentin have been observed in cetuximab resistant cell lines [117]. Another mechanism of cetuximab resistance is overexpression of proteins involved in the suppression of immune effector cells. For example, the cytokine TGF-β, which has been shown to be overexpressed in HNSCC cells can also suppress immune effector molecules, thereby supressing the function of cetuximab [105, 118]. Therefore, alternative therapies to target the EGFR pathway could involve targeting the immune system.

In 2016, the first HNSCC treatment since cetuximab, Pembrolizumab (an anti-programmed cell death-1 (PD-1) antibody) was FDA-approved [101]. PD-1 is an immunoreceptor, which upon ligand activation can induce cell survival in HNSCC [101]. Pembrolizumab has shown promise in an ongoing trial, as decreases in HPV(-) and HPV(+) tumour lesion size have been observed in 57% and 62% of 192 patients, respectively [119]. Moreover, recently, the TCGA have highlighted upregulation in RNA levels for PDL (programmed cell death ligand)-1 [66].

Despite the development of a new therapy and extensive investigation of the genetic landscape, between the development and clinical use of cetuximab and pembrolizumab there was a ten-year gap. Therefore, many aspects of HNSCC biology require enhanced investigation. One approach could be to define the *cis*-regulatory network by interrogation of epigenetic gene regulation such as transcription factor binding and chromatin changes in HNSCC.

1.7 Chromatin structure

The human genome consists of 3.2×10^9 nucleotides and the structure of chromatin allows this considerably large structure to be packaged into a relatively smaller nucleus [120] (**Figure 1.10**). This is achieved by DNA being wrapped around proteins called histones to form a nucleosome complex. A nucleosome consists of 147bp of DNA wrapped around a histone octamer [121] (**Figure 1.10**). The octamer consists of four types of histone proteins, H2A, H2B, H3 and H4, which assemble into two histone H2A-H2B dimers and one H3-H4 tetramer [121]. In between each nucleosome is a stretch of DNA called 'spacer DNA', which is bound to the H1 linker histone [121] (**Figure 1.10**). Nucleosomes can be further assembled into a higher order chromatin structure, where the chromatin coils into a '30nm fibre' and

is further packaged to form a chromatid and eventually a chromosome [122]. Apart from the packaging of DNA, a second feature of chromatin structure is its influence on gene transcription by either occluding or exposing DNA sequences to proteins called transcription factors (TFs), which facilitate and coordinate gene transcription [121].
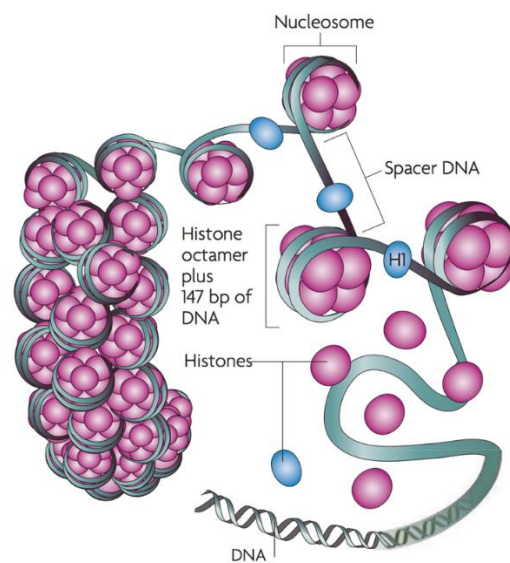
Figure 1.10. Chromatin structure.
Chromatin consists of DNA and its associated proteins. Nucleosomes are 147bp of DNA wrapped around eight histone proteins to form an octamer. An octamer consists of two H2A-H2B dimers and one H3-H4 tetramer. The nucleosomes are linked by the H1 linker histone (spacer histone), which binds to the loops of DNA in between nucleosomes. Nucleosomes can further coil into a '30nm fibre', which is further packaged to form a chromatid and eventually a chromosome (Figure adapted with permission from Figuiredo et al., 2009 [123]).

1.8 *Cis*-regulation

Only ~1.5% of the human genome is involved in encoding an estimated 20,000 genes [124-126]. Therefore, nearly 99% of the genome is involved in functions other than protein coding. It is now known that a part of the non-protein-coding DNA sequences form distinct regions called regulatory elements, such as promoters, enhancers, silencers and insulators [126] **(Figure 1.11)**. These elements have a role in *cis*-regulation of gene expression, which is spatially and temporally coordinated by TF proteins, co-activator proteins, chromatin remodelling complexes and associated epigenetic modifications (DNA methylation and post-translational histone modifications) [127].

1.8.1 Insulator and silencer elements

One type of regulatory element are insulators, which are found at gene boundaries and function to block the neighbouring gene from becoming active [128]. Insulator elements are bound by TF repressor proteins such as, CTCF (CCCTC- binding factor) [129] **(Figure 1.11)**. Recently, a role for insulator elements has been demonstrated in topologically associating domains (TADs) [130-132]. TADs are discrete chromatin domains of several megabases and serve distinct biological functions. For example, developmental genes involved in cell identity may be contained within a TAD and to prevent interaction of the genes with those in other regions, insulator elements maintain the boundary between different TADs [130-132].

A second type of element are silencer elements, which as the name suggests, function to silence gene transcription and can be found as a part of promoter or

enhancer elements [128] **(Figure 1.11)**. Similarly to insulator elements, silencer elements function by facilitating the binding of TF repressors [128]. Repressors may recruit co-factor proteins (co-repressors) to interfere with transcription activating proteins [128].

It is important to mention that other non-coding DNA elements exist apart from the ones mentioned above. For example, repeat elements, which make up nearly 50% of the genome, have a role in genomic maintenance and make up non-coding RNA such as microRNAs, which can repress RNA by blocking their translation or targeting the RNA for degradation [126, 133, 134]. In addition to these elements, promoter and enhancer elements are required for gene function, as discussed later.
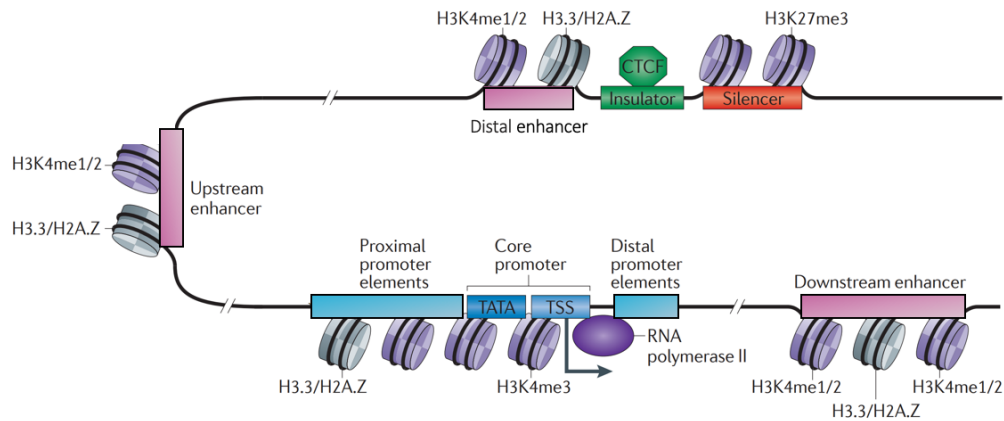
Figure 1.11. Gene structure.
Gene regulation is mediated in part by four regulatory elements: promoters, enhancers, insulators and silencers. Promoters are typically found in proximity to a TSS (transcription start site) of a gene and can consist of three units; the core, proximal and distal promoters. The core promoter contains the TSS and TATA box; the assembly sites for transcription initiation factors, such as RNA polymerase II. The proximal and distal promoters are sites of TF binding, which facilitate binding of transcription initiation machinery. Enhancer elements can facilitate gene transcription by binding to promoters via TFs. Spatially, enhancer elements can be proximal or distal to the promoter element and can be up- or downstream of the promoter. Insulator elements, which often overlap with gene boundaries, restrict transcription by blocking neighbouring genes from being activated. Insulators are bound by TF repressors such as CTCF, which has many functions, among which is aiding insulator function. Silencer elements are also sites of repressor TFs and block transcription. The N-termini of histones (also known as histone tails) are sites of post-translational modifications (PTMs). PTMs include acetylation and methylation of lysine amino acid residues; H3K4me1, -me2 and -me3 are marks of active enhancers; H3K27me3 is a poised enhancer mark; and H3K4me3 is also a mark for active promoter elements. Histone protein variants H3.3 and H2A.Z, along with histone marks of active enhancers, are often found flanking TF binding sites. (Figure adapted with permission from Ong and Corces., 2011[135]).

1.8.2 Transcription factor binding to regulatory elements

Approximately 5-7% of the genome encodes between 1,700 and 1,900 TFs [136, 137]. Regulatory elements harbour TF motifs (short DNA sequences between 6-12bp), which allow TF recognition and interaction with regulatory elements [137]. TFs have been reported to display at least 1,000-fold more affinity for TF motifs compared to other sequences [137]. The motifs can be computationally predicted and classified into families by sequence homology to previously characterised TF DNA-binding domains [137].

Upon binding to the DNA, TFs can either activate or repress transcription [137]. For example, in laryngeal SCC cell lines, it has been shown that ΔNp63 can directly bind to the *YAP1* promoter to inhibit its expression [138, 139]. Furthermore, TF binding can be dependent on other TFs. For example, TF binding investigated in mouse embryonic stem cells revealed regions where Nanog (nanog homeobox), Oct4 (POU5F1- POU class 5 homeobox 1) or Sox2 TFs cluster and preferentially bind together [140]. Depletion of Nanog or Oct4 led to a decrease in the number of TF interactions at these regions [140].

1.8.3 Promoter-enhancer interactions

In order for gene transcription to occur, the promoter element (core promoter) is located near to the gene transcription start site (TSS) and serves as a hub for the assembly of transcription machinery, such as the RNA polymerase II enzyme (RNAP II) and the pre-initiation complex (PIC) [127] **(Figure 1.11)**. This process is facilitated by TFs, as they can directly recruit transcription machinery. For

example, the TBP (TATA-binding protein) TF can directly recruit RNAP II to the promoter [137].

Enhancer elements are required for the activation of promoters to initiate gene transcription. Enhancer elements are sequences typically between 200-500bp and are largely found in intergenic and intronic regions [141]. Rarely, some enhancer elements can overlap with exons of neighbouring genes or with the exons of genes they regulate [142-144]. Enhancers can activate transcription by interacting with one or more promoters, but enhancers can often be several kilobases away from their promoter [145]. One of the most striking examples of this is an enhancer for the *POU3F4* (POU class 3 homeobox 4) gene, which is located 900kb upstream of its target [146]. A mechanism as to how promoter and enhancer interactions may occur is DNA looping, which will be discussed in the next section.

In order for enhancer elements to carry out their function, TF binding is required. TFs can exhibit specificity to enhancer elements, as demonstrated in a study by Visel *et al.*, which revealed that, of 41 TF motifs identified in enhancer elements, at least 90% were unique to its enhancer and were not found at promoter elements [147]. This has been confirmed by another group who investigated TF binding in mice, and found 7/14 TFs (e.g. Oct4, Sox2 and Nanog) bound preferentially to enhancers, while the remainder (e.g. c-Myc), bound promoter elements [148].

Upon TF binding to enhancers, recruitment of other proteins (co-activators) can be required to induce transcription [149]. Co-activators do not harbour specific motifs enabling them to bind regulatory elements, but can recognize different chromatin states (i.e. accessibility) and alter chromatin state through histone modifying enzymatic activity, i.e. histone acetyltransferases (HATs), histone

methyltransferases (HMTs) or ATP-dependent chromatin remodelling complexes, as will be discussed later [150].

Therefore, in order for enhancers to facilitate transcription, two main criteria must be fulfilled; firstly, chromatin decondensation must occur to allow TFs access to enhancers; and secondly, the chromatin structure must be altered, a process known as DNA or chromatin looping, to allow proximity and interaction between enhancers and promoters [145]. There are several mechanisms which can influence chromatin structure, which include DNA looping, chromatin remodelling complexes, histone modifications and DNA methylation, as discussed next.

1.8.4 DNA looping

It is not yet fully understood how enhancers and promoter elements gain proximity. One model is DNA looping, a mechanism first proposed by Rippe *et al*., 1995 and since then many models have been put forward as to how DNA looping may occur [151, 152]. The most commonly reported model involves architectural proteins: CTCF (CCCTC-binding factor), cohesin and the mediator (MED) complex [152, 153] **(Figure 1.12)**. The model suggests that sections of the chromatin at both enhancers and promoters are surrounded by a ring-like cohesin complex, along with MED complexes and CTCF [152] **(Figure 1.12B)**. It is thought that CTCF has a role in separating folding chromatin domains, while also maintaining the position of cohesin [154]. TF binding has also been shown to influence chromatin structure, as recruitment of factors such as Nanog and Oct4 have been observed during DNA looping [140].
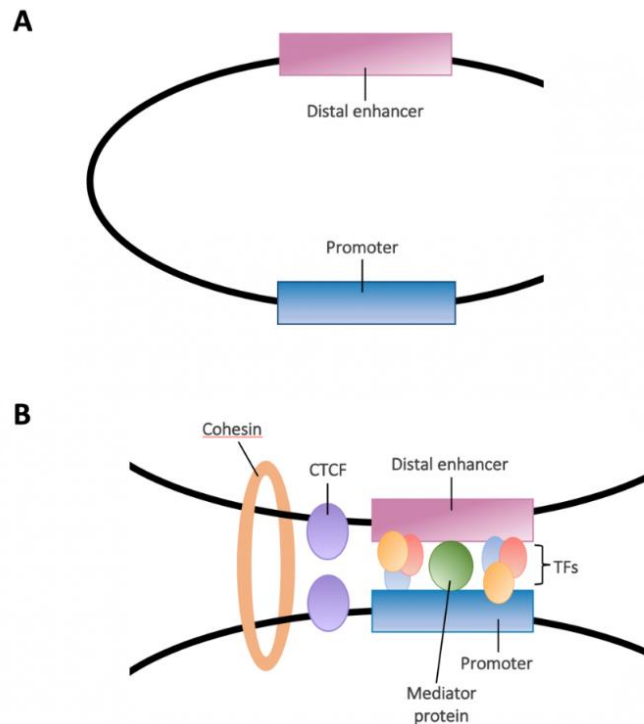
Figure 1.12. DNA looping.
(A) In a transcriptionally inactive state, enhancers can be located several hundred kilobases away from their target promoter. **(B)** Upon binding of TFs to enhancers and promoters, both elements are brought into proximity with each other. It is thought that promoter-enhancer proximity occurs by DNA (chromatin) looping, which involves the architectural proteins: cohesin, mediator (MED) protein and CTCF. Both enhancers and promoters are surrounded by a ring-like cohesin complex, along with MED complexes and CTCF. CTCF may function to separate the folding chromatin domains, while also maintaining the position of cohesin. TF recruitment also occurs during DNA looping and may aid the process.

1.8.5 Chromatin remodelling complexes

One of the mechanisms by which chromatin decondensation can be induced is with the aid of chromatin remodelling complexes (CRCs). CRCs can use ATP energy to reposition nucleosomes, hence exposing TF motifs for TF binding [155]. Four families of ATP-dependent CRCs exist, the SWI/SNF family (switching defective/sucrose non-fermenting), ISWI family (imitation switch), the NuRD family (Mi-2/nucleosome remodelling and histone deacetylation) and the INO80

45

family (inositol requiring 80) [155]. Interestingly, ACTL6A (actin like 6A), a subunit of the SWI/SNF CRCs, is overexpressed in laryngeal SCC cells and has been associated with increased cell proliferation [155].

1.8.6 Histone modifications

In addition to the distinct functions carried out by regulatory elements which are mediated by TFs, histone modifications are also required to aid spatially and temporally coordinated gene regulation.

The N-terminal histone tails can undergo post-translational modifications (PTMs), including addition of acetyl, methyl, phosphate or ubiquitin groups, amongst others [121]. These modifications are added or removed by enzymes specific to the PTM; acetyl groups are transferred by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs), while methyl groups can be transferred by histone methyltransferases (HMTs) and removed by histone demethylases (HMT) [150]. The addition of PTMs occur on specific amino acid residues, for example, acetylation occurs on lysine residues, while methylation occurs on arginine, lysine and histidine residues [121]. It is important to note that this is not an exhaustive list of PTMs and the same PTM can be added on to an amino acid residue multiple times [156]. The common nomenclature for denoting a histone modification is, for example, H3K27Ac, which is acetylation of lysine 27 of histone H3. Methylation is denoted with 'me' and the number of modifications can be denoted by mono-, di- or tri- (me1, me2 and me3, respectively).

PTMs have distinct effects on chromatin structure and therefore on DNA accessibility. For example, the addition of an acetyl group can neutralise the histone's positive charge, relaxing the chromatin, leading to select DNA sequences

46

being made accessible to TFs [157]. An example of a histone modification leading to an open (active) chromatin state is H3K27ac [121]. On the other hand, methylation of histone tails can make DNA inaccessible, for example H3K9me is typically associated with closed, inactive chromatin [121].

Altered histone modifications resulting from expression changes of their respective enzymes have been associated with many cancers, including oral SCC. For example, HDAC6 overexpression has been observed in oral SCC cell lines and is associated with tumour aggressiveness [155]. In a further example, increased HDAC-8 and -9 levels have been associated with increased cell proliferation in oral SCC cells [155].

1.8.7 DNA methylation

DNA can also undergo modifications by addition of methyl groups onto the fifth carbon of the cytosine residue in CpG dinucleotides [158, 159]. Methylation of DNA is largely associated with gene silencing and inaccessible chromatin. For example, methylation has a role in X chromosome inactivation and genetic imprinting [159]. A family of DNA methyltransferase (DNMT) enzymes carry out methylation of CpGs and can be removed either by a family of TET (ten-eleven translocation) enzymes or by a passive process of DNA demethylation, which is dependent on replication [159]. Approximately, 15% of CpG dinucleotides are found in CpG islands (CGIs), which are dense regions of CpGs defined as containing a GC content of more than 50% [159]. CGIs overlap with ~70% of all gene promoter elements in the human genome [159].

Alterations in the pattern of DNA methylation is commonly observed in cancer cells, in the form of, for example, hypermethylation and is frequently found to affect

47

tumour suppressor genes such as *TP53* [160]. Recently, the PanCancer Atlas study, which used TCGA data derived from lung, HNSCC, oesophageal, cervical and bladder tumours to investigate alterations common to SCC, reported hypomethylation of CpGs near the *ΔNp63* TSS, which may be a reason for its overexpression in HNSCC [66].

1.9 Cell-specificity and identification of regulatory elements

As highlighted thus far, *cis*-regulation is a complicated and dynamic process, reliant on coordination between several regulatory elements, DNA/histone modifications and TF proteins. This coordination can be further refined by cell type specificity of enhancers. For example, within the mammalian genome, some 400,000 to 1.4 million putative enhancers have been identified and each cell type can contain between 10,000 and 150,000 enhancers [161, 162]. Therefore, due to their complex nature, both abundance and lack of distinguishing features, genome-wide identification of enhancers has proven difficult. Several criteria are used to identify enhancer elements, which involve interrogation of the local chromatin structure through histone modifications, TF binding, nucleosome occupancy and sensitivity to DNase I digestion [147].

In 2004, the challenge of characterising enhancers was undertaken by the ENCODE (ENCyclopaedia Of DNA Elements) consortium [163]. The pilot ENCODE project involved using genome-wide methods to sequence and annotate ~1% of the human genome, including regulatory elements [163]. With the advance in genome-wide

methods, this study was followed up in 2012 to encompass sequencing data from 147 cell types [164].

One of the features of enhancers highlighted by ENCODE were histone modifications [164]. Histone modifications on enhancer elements can indicate three different enhancer states, active, silent and poised. For example, the histone modifications H3K27ac and H3K4me1 are marks of active enhancers, while H3K27me3 is a mark of silenced enhancers [164-167]. Poised enhancers are largely associated with developmental genes, which are poised for expression upon differentiation [168]. Poised enhancers therefore possess both repressive marks and active marks [167, 169]. To identify putative enhancers, an approach used in many studies is genome-wide chromatin immunoprecipitation, followed by sequencing (ChIP-seq) for enhancer histone marks (H3K4me1 and H3K27ac) [129, 162, 164, 170]. To verify the DNA sequences as enhancers, reporter assays can be conducted where the reporter construct contains the putative enhancer sequence next to a promoter element upstream from a reporter gene [129, 162, 164, 170].

A second feature of enhancers and method of identification is the presence of two histone variants H3.3 and H2A.Z. Nucleosomes consisting of these histone variants are unstable and coupled with active histone marks (H3K27ac and H3K4me1), the nucleosomes offer easier access to occluded TF binding sites [129, 168, 171]. Therefore, the presence of H3.3 and H2A.Z often mark active enhancers [172].

Another characteristic of enhancers was described by Heintzman *et al.,* where p300, a HAT enzyme, was discovered as a marker for enhancers, demonstrating their cell type-specific pattern. In addition, p300 was shown to overlap with DNase I-

hypersensitive sites (DHSs), which are indicative of regulatory elements, as will be discussed later [129, 167].

1.10 Enhancer clusters and stretch enhancers

In addition to typical enhancer elements described above, several studies have reported that enhancers can also be found clustered in 2D space [173-177]. These enhancer types were termed 'super-enhancers' and were first introduced by Whyte *et al*. in 2013 [173, 175]. However, this term implies an enhanced role of these elements in transcriptional regulation, but this is yet to be proven. Therefore, in this thesis, super-enhancers will be called 'enhancer clusters' (ECs).

ECs have been implicated as features of the chromatin, which can reflect cell identity. They were first described in mouse embryonic stem cells and were found to be bound by master TFs, Oct4, Sox2 and Nanog, which have an essential role in development [173, 175]. The criteria for defining an EC was the following: (i) a site bound by all three master TFs, (ii) concatenation of neighbouring enhancers within a 12.5 kb region, and (iii) the concatenated 12.5kb regions were ranked based on Med1 enrichment levels, normalised to the background **(Figure 1.13)**. A ranked score of more than one was defined as an EC and the remaining enhancers were classed as 'normal' [173] **(Figure 1.13)**. The criteria set out by Whyte *et al.* for EC identification has not been adhered to by all groups. For example, instead of using the master TFs as markers of enhancers, alternative enhancer marks have been used such as, H3K27ac [174, 175]. In addition, in the final step, instead of Med1 enrichment, H3K27ac levels or MyoD (myoblast determination protein D) levels (a master TF involved in mouse development) have been used to assess for enhancer enrichment [173, 175].

Other features of ECs, which separate them from 'normal' enhancers, include EC median length, which was reported to be 8,667bp compared to 703bp, respectively. In addition, ECs are more enriched for enhancer features such as DNase I sensitivity, RNA polymerase II, p300, H3K4me1, H3K4me2 and H3K27ac marks [175].

Interestingly, ECs have been associated with known oncogenes, for example, *EGFR* in glioblastoma and pancreatic cancer, *ERBB2* in liver, breast, prostate and colorectal cancers and *NOTCH1* in T-cell leukaemia, lung, prostate and colorectal cancers [175]. Furthermore, an *EGFR* EC was identified in a cervical cancer cell line where it was reported to be activated by E6 expression [178]. Activation of the *EGFR* EC was shown to be induced by E6-mediated degradation of KDM5C (lysine demethylase 5C), a demethylase enzyme. This leads to deregulation of H3K4me1 and H3K4me3 histone marks at the *EGFR* EC [178]. An EC has also been identified for the *KLF* (Krüppel-like factor)-*5* gene in HNSCC tumours, where focal amplifications of the *KLF5* EC have been linked to KLF5 protein overexpression [179]. The same study found that the *KLF5* gene can activate transcription of genes associated with SCC cell identity such as *KRT* (keratin)-*5*, *KRT6A*, *KRT13*, *TP63* and *WNT10A*.

Another type of enhancer has been described by Parker *et al.* called 'stretch-enhancers', named due to their main feature of spanning more than 3kb. Unlike ECs, stretch-enhancers are not concatenated regions [141]. Stretch-enhancers have been associated with pancreatic islet function and linked to type 2 diabetes [177].
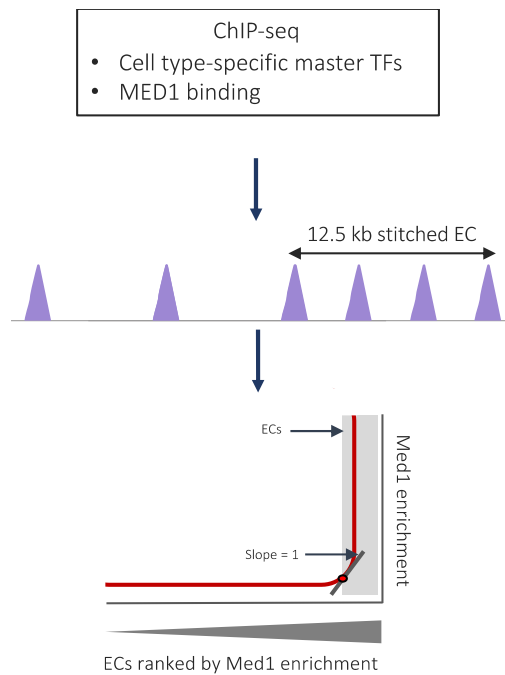
Figure 1.13. Identification of enhancer clusters (ECs).
ChIP-seq was conducted for cell type-specific master TF regulators and for MED1 (Mediator 1). Regions of 12.5kb, which were co-bound by the TFs and exhibited MED1 enrichment (normalised to background MED1 occupancy) were concatenated. Concatenated and individual enhancers were ranked by MED1 enrichment; MED1 ranking was geometrically scaled from 0-1. The point was determined for where a line with a slope of 1 was tangent to the curve. Regions beyond this point were considered to be ECs, while regions below were considered to be 'normal' enhancers.

1.11 Genome-wide methods to identify accessible chromatin

In addition to the abundance of features, regulatory elements are also characterised by accessible chromatin structure. These regions are found at nucleosome-depleted or structurally unstable chromatin regions during chromatin decondensation. Such regions overlap with regulatory elements to facilitate gene transcription [180-182]. Accessible chromatin and therefore DNA become sensitive to nuclease enzymes and can be exploited to identify regulatory elements [183, 184].

Enzymes such as DNase I endonucleases and Tn5 transposases have been used, followed by genome-wide next-generation sequencing, giving rise to techniques such as, DNase I hypersensitivity assay (DNase I-seq) and Assay for Transposase Accessible Chromatin-sequencing (ATAC-seq), respectively. An alternative method to identify open chromatin regions is the FAIRE (formaldehyde-assisted isolation of regulatory element) assay, as discussed later.

1.11.1 DNase I hypersensitivity assay (DNase I)-sequencing

The DNase I assay involves the isolation of nuclei, followed by limited chromatin digestion with the DNase I enzyme, which preferentially digests open regions of the chromatin or regions with unstable nucleosomes, giving rise to DNase I hypersensitive sites (DHSs) [185].

The assay was further advanced with the development of next-generation sequencing to identify DHSs genome-wide, giving rise to DNase I-sequencing (DNase I-seq). The field of next-generation sequencing has been rapidly advancing, and many platforms are available (reviewed in Barba *et al*., 2014). One of the more commonly used platforms is the Illumina HiSeq 2500 sequencer, which has the capacity to sequence the whole human genome in 24 hours [186]. However, to produce meaningful data (i.e. data of high enough resolution to minimise background noise) biological and technical factors must be considered, such as DNA complexity (particularly GC content) and depth of coverage [187]. A deeper depth of coverage increases signal of the data and reduces background, however this has to be compromised with cost as higher coverage is more expensive [186].

DNase I-sequencing (DNase I-seq) has revealed that the genomic distribution of DHSs largely overlaps with intronic and intergenic regions and therefore with

enhancers [142-144]. Another characteristic of DHSs is their reflection of the cell type-specific nature of enhancers. For example, DNase I-seq was performed in 19 cell lines and intergenic DHSs were the most cell type-specific, while DHSs overlapping with TSSs were the least cell specific [188]. Another study reported cell type-specific DHSs to be distal to TSSs, indicating overlap with distal regulatory elements [189].

While the assay is a valuable tool to isolate regulatory networks, a drawback is that it requires between 25-50 million cells per DNase I preparation [190]. Furthermore, the required depth of coverage is 20-50 million reads [191]. A second method to identify open chromatin regions is FAIRE-seq.

1.11.2 FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements)-sequencing

In the FAIRE assay, to preserve DNA-protein interactions, the chromatin is crosslinked with formaldehyde, followed by shearing by sonication and phenol-chloroform separation of the nucleosome-free and nucleosome-bound DNA [192, 193]. Subsequently, qPCR (quantitative polymerase chain reaction) or next-generation sequencing (FAIRE-seq) can be used to identify the DNA sequences. Similarly to the DNase I-seq assay, FAIRE-seq requires cells in the millions (1-50 million) [194]. The advantage of FAIRE-seq is that nuclei isolation is not required as in DNase I-seq [191]. Instead, chromatin crosslinking by formaldehyde is required, which preserves DNA-protein contacts [191]. Many groups have utilized the assay, including the ENCODE Consortium [164]. However, optimization of the fixation step is required and the signal:noise ratio in FAIRE is much lower compared to DNase I-seq [189].

Comparisons of FAIRE-seq and DNase I-seq assays have revealed that there is a discrepancy in the regions identified by both methods. For example, it was found that 30-40% of the top 100,000 peaks (regions) from a total of 870,000 across seven cell lines overlapped, while 80% of the top 10,000 peaks overlapped [189]. This suggests the most stable peaks are more frequently identified by both assays. To investigate this further, ChIP-seq was conducted for CTCF and MYC TFs to identify accessible regions and then compared with peaks identified in DNase I-seq and FAIRE-seq [189]. It was found that over 96% of ChIP peaks overlapped in both assays. A similar result was found when p63 ChIP-seq was conducted; it was found that 80% of peaks overlapped with all three assays [189]. Therefore, the vast majority of open regions are identified by both assays, but DNase I-seq demonstrates a better signal:noise ratio.

1.11.3 ATAC (Assay for Transposase-Accessible Chromatin)-sequencing

More recently, the ATAC-seq assay has been developed allowing sequencing of open regions of just one cell or using a population of between 500 and 50,000 cells, as oppose to the millions of cells needed for DNase I-seq and FAIRE-seq [195]. Therefore, ATAC-seq can be advantageous for the identification of open regions in primary cells and non-malignant cell lines where cells have a limited lifespan or proliferation can be slow. Furthermore, the ATAC-seq protocol can be completed in three hours compared to two overnight incubations for both DNase I-seq and FAIRE-seq [190, 194]. Briefly, the ATAC-seq protocol involves cell lysis, followed by the transposition reaction. The transposition reaction uses the Tn5 transposase enzyme, which probes and digests open chromatin regions, while

inserting adapters required for next-generation sequencing and this is followed by PCR amplification of the DNA and eventual sequencing [181].

The correlation of open regions yielded from ATAC-seq were compared with those from DNase I-seq. Both assays were conducted in a lymphoblastoid cell line and it was found that the specificity and sensitivity was comparable, despite the difference in starting cell number [195]. Furthermore, intensities of peaks displayed high correlation between DNase I-seq and ATAC-seq [195].

Until recently, the major disadvantage of this method was that, due to contamination with mitochondrial DNA, the assay required 60-100 million reads (compared to 20-50 million reads for DNase I- and FAIRE-seq) [195]. While mitochondrial DNA can be computationally filtered out, during sequencing it will deplete the number of reads from the nuclear DNA [195]. However, the assay was improved on by the addition of two detergents (Tween-20 and digitonin) during cell lysis, followed by an extra wash step after lysis and the addition of phosphate buffered saline (PBS) to the transposition reaction [196]. Upon sequencing it was discovered that these changes yielded a three-fold higher percentage of nuclear mapped reads [196].

1.11.4 Chromatin immunoprecipitation (ChIP)-sequencing

The spatial and temporal nature of TF binding can be studied with the aid of chromatin immunoprecipitation (ChIP). Very briefly, this technique involves preserving protein-DNA interactions by performing formaldehyde crosslinking [197]. The DNA is then sheared by sonication yielding fragmented DNA. Next, an antibody, specific to the protein of interest, is used to isolate the DNA sequence to which the protein is bound [197].

ChIP conducted with crosslinked chromatin is also known as X-ChIP, while N (native)-ChIP is conducted on non-crosslinked chromatin and is most commonly used to map histone modifications [198, 199]. N-ChIP offers the advantage of preserving the antibody epitope, which in the presence of formaldehyde may be disrupted or destroyed [199]. On the other hand, an advantage of X-ChIP is that histone organisation remains intact, while in N-ChIP histone rearrangement and protein-TF complex disruption may occur during the procedure [199].

In order to interrogate the DNA sequences yielded from ChIP, the assay can be followed by hybridisation to an array harbouring DNA probes (ChIP-ChIP) or qPCR, which involves using specific primer sites (ChIP-qPCR) [200]. However, both these techniques require prior knowledge of where the protein may bind. An alternative is ChIP-seq, which involves following ChIP with next-generation sequencing and allows genome-wide analysis [200].

1.12 Aims

Great advances have been made in the investigation of the HNSCC genetic landscape. More recently, The Cancer Genome Atlas consortium HNSCC cohort, has greatly contributed to this, as highlighted in the introduction. The findings of these studies have indicated a role for the *cis*-regulatory network in HNSCC development, however this is yet to be put into context. Therefore, the aims of this thesis are the following:

(i) To identify elements of the *cis*-regulatory network in HNSCC cell lines by genome-wide localisation of open chromatin regions, key TFs and signalling pathways.

(ii) To investigate and compare the *cis*-regulatory network in HPV(+) and HPV(-) HNSCC cell lines.

(iii) To compare the *cis*-regulatory network between HNSCC and non-tumourigenic oral keratinocyte cells.

(iv) To pharmacologically manipulate the identified signalling pathways to elucidate new therapeutic targets for HNSCC.

CHAPTER 2: METHODS AND MATERIALS

2.1 Cell culture

2.1.1 General cell culture and origin

Four head and neck squamous cell carcinoma (HNSCC) cell lines were used. Two were negative for the human papillomavirus (HPV) and two were positive for HPV16. The presence or absence of HPV will be denoted by (+) or (-), respectively. The HPV(-) cell lines were, VU40T(-) and SCC040(-), and the two HPV(+) cell lines were VU147(+) and UDSCC2(+) **(Table 2.1)**. The SCC040(-) cells were purchased from the German Culture Collection, DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) (#ACC660) by Dr. S. Roberts (Supervisor, University of Birmingham). The VU40T(-) and VU147(+) cell lines were gifted to Dr. S. Roberts from Prof. H. Joenje (VU University Medical Centre, Netherlands). The UDSCC2(+) cells were obtained from Dr. J.S. Gutkind (National Institute of Dental and Craniofacial Research (NIDCR), National Institutes of Health (NIH), Bethesda).

All HNSCC cell lines were cultured in Dulbecco′s Modified Eagle′s Medium (DMEM) (Sigma-Aldrich, UK), supplemented with 2% L-glutamine (Gibco, UK), 10% (v/v) foetal bovine serum (FBS) (Sigma-Aldrich, UK), 1% (v/v) penicillin/streptomycin (Gibco, UK), 1% (v/v) MEM (minimum essential medium) non-essential amino acids (Gibco, UK) and 1% (v/v) sodium pyruvate (Gibco, UK).

Table 2.1. Cell lines and primary cell.

Cell line	Origin	Age	Sex	Stage	Smoking/ alcohol	HPV status
OKF6/ hTERT	Floor of mouth	28	M	N/A	-/-	-
HOK	Oral mucosa	20 week gestation	Foetal			
OTK	Tonsil	Unknown	F	N/A	N/A	N/A
NIH J2-3T3	Mouse embryo fibroblast	N/A				
VU40T(-)	Tongue	65	F	T3N0	-/-	(-)
SCC040(-)	Tongue	50	M	T2N2	-/+	(-)
VU147(+)	Floor of mouth	58	M	T4N2	+/+	HPV16
UDSCC2(+)	Hypopharynx	58	M	T1N2	Unknown	HPV16

The table outlines cell line information including, anatomical origin, age and sex of cell donor, stage of cancer (if applicable; not applicable is denoted by 'N/A'), smoking and alcohol habits of donor and HPV status of cells (if applicable). OKF6 [201-203], VU40T(-) [204], SCC040(-) [205, 206], VU147(+) [207], UDSCC2(+) [208].

A hTERT (human telomerase reverse transcriptase) immortalised human oral keratinocyte cell line, OKF6 (obtained from Prof. Paul Cooper, University of Birmingham) was used **(Table 2.1)**. OKF6 cells were cultured in keratinocyte-serum free medium (SFM) (Gibco, UK) and supplemented with 5ng/ml epidermal growth factor (EGF) 1-53 (Gibco, UK), 50µg/ml bovine pituitary extract (BPE) (Gibco, UK), 0.4mM calcium chloride (Sigma-Aldrich, UK), 1% (v/v) penicillin/streptomycin (Gibco, UK).

Two primary cell types were also used, human oral keratinocyte (HOK) and human oral tonsil keratinocyte (OTK) cells **(Table 2.1)**. The cells were cultured in Oral Keratinocyte Medium (OKM), supplemented with 1% (v/v) penicillin/streptomycin and 1% (v/v) oral keratinocyte growth supplement (OKGS). The HOK cells and all reagents used to culture them were purchased from Caltag Medsystems, UK (HOK

cells' catalogue number: 2610). The OTK cells were a gift from Dr. S. Roberts (ethical approval number: REC 06/Q1702/45.

The OTK cells were cultured in 90mm dishes using Complete E media (CEM) which consisted of DMEM (Gibco, UK), supplemented with 32% (v/v) Ham's F12, (Gibco, UK), 0.1% (v/v) hydrocortisone (Sigma-Aldrich, UK), 10% (v/v) Hyclone FCS (Thermo Fisher Scientific, UK), 0.5% (v/v) EGF (epidermal growth factor) (BD Biosciences, UK), 2% (v/v) L-glutamine (Gibco, UK), 0.1% (v/v) cholera toxin A (MP Biomedicals, USA), 0.2% (v/v) insulin (Sigma-Aldrich, UK), 0.2% (v/v) transferrin (Sigma-Aldrich, UK), 4nM tri-iodo–lL-thyronine (Sigma-Aldrich, UK) and 36μM adenine (Sigma-Aldrich, UK). Irradiated NIH J2-3T3 cells were used as a feeder layer for the maintenance of OTK cells **(Table 2.1)**. The J2-3T3 cells were cultured in DMEM, supplemented with 2% (v/v) L-glutamine and 10% (v/v) adult calf serum (Gibco, UK).

All cell lines were maintained in 5% CO_2 at 37ºC and were confirmed negative for mycoplasma throughout the study, as described in Section 2.1.4.

2.1.2 Cell harvesting and subculturing

The HNSCC cells were cultured to approximately 70% confluency before being subcultured. To subculture, the cells were washed with PBS (Sigma-Aldrich, UK) and trypsinised using 0.25% trypsin-EDTA (ethylene diaminetetraacetic acid) (Thermo Fisher Scientific, UK). The cells were incubated at 37ºC until approximately 80% of cells had detached. The trypsin was neutralised with an equal volume of serum-containing medium. To subculture, the VU40T(-) and SCC040(-) cells were reseeded at approximately 30-40% cell density, and the UDSCC2(+) and VU147(+) cells were reseeded at approximately 40-50% cell density.

The OKF6 cells were cultured to 80% confluency before being subcultured as described above, with the added step of pelleting the cells at 300 xg for 5mins after neutralising the trypsin. The cells were resuspended in medium and reseeded at a confluency of 60-70%. To subculture, the cells were reseeded at approximately 60% cell density.

The HOK cells were cultured to 80% confluency before subculturing. To subculture, the cells were washed with PBS and trypsinised with 2ml of 0.25% trypsin-EDTA in 8ml of PBS. The cells were incubated for 1-5mins at 37ºC until they appeared round and were transferred to a centrifuge tube containing 5ml of FBS. The flask was transferred back into the incubator for 1-2mins to detach any remaining cells. The trypsin in the flask was neutralised with 10% FBS in PBS and the cells were transferred to the centrifuge tube. To wash off any remaining cells, 5ml of 10% FBS in PBS was used and transferred to the centrifuge tube. The cells were centrifuged for 5mins at 300 xg, the supernatant was discarded and the cells resuspended in an appropriate volume of medium. To subculture cells, the cells were counted (Section 2.1.3) and seeded into poly-L-lysine (PLL) coated flasks at a density of 5,000 cells/cm^2. The PLL was used at a density of 2µg/cm^2.

To subculture OTK cells, the feeder layer (γ-irradiated J2-3T3 cells) was first removed by adding 0.5mM of EDTA in PBS and incubating the cells until detached. The dish was washed with the solution to dislodge any remaining J2-3T3 cells and the suspension was discarded. The dish was further washed twice with 5ml of PBS to remove all cells of the feeder layer and was viewed under a light microscope to ensure all J2-3T3 cells had been removed. Next, the OTK cells were removed by adding 2ml of the TrypLE trypsin solution (Gibco, UK). The dish was incubated at 37ºC until approximately 90% of OTK cells had detached. The trypsin was

neutralised with CEM and the dish was washed with a further 10ml of CEM. The suspension was transferred to a centrifuge tube and centrifuged for 538 xg for 5mins and the supernatant was discarded. The cells were resuspended in CEM, counted and seeded at a density of $2x10^5$ per dish onto $2x10^6$ cells of irradiated J2-3T3 cells, which had been seeded at least 24hrs earlier. The cells were irradiated with 30 Grays of caesium-137 and were stored in CEM at 4ºC for up to 5 days.

J2-3T3 cells were grown to 80% confluency before being subcultured. To subculture, the cells were washed twice with 5ml of PBS and were incubated with TrypLE for 1-5mins until the cells became non-adherent. The trypsin was neutralised with media and cells were washed off with the solution. The cells were centrifuged for 300 xg for 5mins and the supernatant was discarded. The cells were resuspended in appropriate media, counted and seeded at a density of $1x10^4$/ml.

2.1.3 Cell counting

To plate the required number of cells, or to calculate the proportion of viable cells by trypan blue staining, cells were harvested and counted using a haemocytometer. Ten microlitres of the cell suspension was loaded into the haemocytometer. To assess cell viability, trypan blue solution (Sigma-Aldrich, UK) was used and prepared by diluting 1:1 with PBS; the cells were diluted 1:10 in the solution. The cell membrane of non-viable cells can breakdown, allowing penetration of trypan blue. Therefore, non-viable cells appear blue, while viable cells appear clear. To determine viability, the ratio of clear cells to blue cells was calculated.

2.1.4 Mycoplasma testing

Cells were tested for mycoplasma infection every six months. Mycoplasma are a bacteria which can infect cells in culture. Infection can disrupt cell function in many ways including altering DNA, RNA and protein levels [209].

The MycoAlert mycoplasma detection kit was used as per manufacturer's instructions (Lonza, USA). The kit utilises mycoplasma specific enzymes, which catalyse the conversion of ADP (adenosine diphosphate) to ATP (adenosine triphosphate). Therefore, in the presence of mycoplasma, ATP will be produced. The levels of ATP were measured before and after cell lysis. Samples with a ratio of less than 0.9 (after lysis:before lysis) were considered as non-contaminated.

The MycoAlert reagent and substrates were supplied as lyophilised pellets and were dissolved in 600µl of MycoAlert assay buffer. The reagent and substrate were stored at 4°C. For use, the solutions were equilibrated to room temperature (RT) for 15mins prior to using. Cells were cultured for at least 48hrs in antibiotic-free medium before taking a 200µl aliquot. The medium was centrifuged at 200 *xg* for 5mins, 100µl of supernatant medium was transferred to a new centrifuge tube and fresh medium was transferred to a second tube as a negative control. Next, an equal volume of MycoAlert reagent was added to each sample. The solution was incubated in the dark at RT for 5mins. Luminescence was measured to obtain a reading of ATP before conversion. Next, 100µl of MycoAlert Substrate was added and the samples were incubated for 10mins in the dark. The luminescence was measured after lysing mycoplasma and a ratio of the reading before and after was calculated.

2.1.5 HPV detection

To assess the HPV status of the HNSCC cell lines, PCR (polymerase chain reaction) was conducted to amplify the HPV16 oncogene *E6*. RNA was isolated from three biological replicates of each HNSCC cell line and from the non-tumourigenic OKF6 cell line, followed by cDNA synthesis (Section 2.13.1-2.13.2). PCR was conducted, as outlined in **Tables 2.2 and 2.3**. The OKF6 cell line served as a negative control. The HPV16 E6 genomic sequence (NC_001526) was obtained from the GenBank database to design the PCR primers (forward 5' AGGACCCACAGGAGCGACCC 3', reverse 5' ACGTCGCAGTAACTGTTGCTTGCA 3'). PCR was conducted using a TProfessional Standard PCR Thermocycler (Biometra, Germany).

The PCR product was run on a 2% agarose gel, as described in Section 2.3. The expected size of the PCR product was 112bp and the observed product was just above the 100bp marker, indicating amplification of the *E6* transcript **(Figure 2.1)**. The presence of the *E6* transcript was verified in the three biological replicates of the HPV(+) cell lines, UDSCC2(+) and VU147(+). An *E6* amplicon was not detected in the HPV(-) cell lines (VU40T(-) and SCC040(-)) or in OKF6 cells. Therefore, it was concluded that only UDSCC2(+) and VU147(+) cells contain *E6* transcripts and are HPV16 positive, confirming previously published findings [207, 208] **(Table 2.1)**.

Table 2.2. PCR reaction setup.

PCR reagent	Final concentration
GoTaq (Promega, UK)	1x
cDNA	1 µl of 1:4 diluted in nuclease-free water
Forward primer	62nM
Reverse primer	62nM
Water	up to 12µl

Table 2.3. PCR parameters

Cycle Step	Temperature (ºC)	Time (Secs)	Number of cycles
Initial denaturation	95	300	
Denaturation	95	10	x30
Annealing	60	10	
Final extension	72	60	
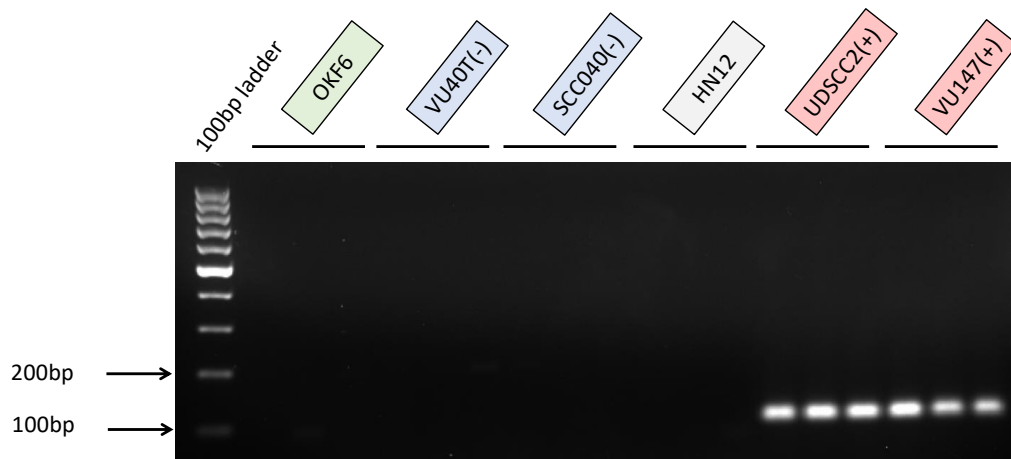
Figure 2.1. PCR of HPV16 E6 gene.
PCR was conducted using primers designed to the HPV16 oncogene, *E6* (112bp). Three biological replicates were assessed for each cell line. The non-tumourigenic OKF6 cell line, VU40T(-) and SCC040(-) were negative for the *E6* amplicon. The HPV(+) cell lines, UDSCC2(+) and VU147(+) were positive for *E6* transcripts. (HN12 is a cell line which was not investigated in this project).

2.2 General methods

2.3 Agarose gel electrophoresis

To separate DNA fragments by size, agarose gel electrophoresis was used. The agarose gel was made by mixing agarose powder (Sigma-Aldrich, UK) in 0.75x TBE (Tris/Borate/EDTA) and boiling the solution until clear. The percentage of agarose gel depended on fragment size; fragments larger than 500bp or smaller than 500bp were separated out on 1% and 2% agarose gels, respectively. To visualise the DNA, 1x ethidium bromide, a DNA intercalator, was used to stain the DNA (Sigma-Aldrich, UK). Ethidium bromide is visible under ultraviolet light and so the presence of DNA can be detected.

The gel was poured into a casting tray with a comb to form wells and left until it had polymerised. The comb was removed, and the PCR product was loaded into the well. A marker for DNA size was also loaded; 1μg of a 100bp or 1kb ladder (Fermentas, UK). The gel was run at 80V for either ~1hr for a 1% gel or ~2hrs for a 2% gel. The gel was visualised using the Gel Doc™ XR system (Bio-Rad, UK).

2.4 Quantification of nucleic acids

The quality and purity of nucleic acids were assessed using the UV-based NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, UK). The output was nucleic acid concentration in ng/μl. The purity and quality of the nucleic acids was determined by the ratio of the absorbance wavelengths: 260/280 and 260/230. The ratio of 260/280 of ~1.8 for DNA and ~2.0 for RNA and the ratio of 260/230 of ~2.0 for DNA and RNA indicated 'pure' nucleic acid preparations [210]

2.5 QPCR

QPCR (quantitative polymerase chain reaction) was used for relative quantification of gene expression levels and to assess DHSs (DNase I hypersensitive sites) and ChIP (chromatin immunoprecipitation) enrichment. QPCR was conducted using the LightCycler 480 II instrument (Roche, UK). The qPCR reaction composition and conditions are shown in **Tables 2.4 and 2.5**, respectively. Primers were designed using the '*in silico* PCR' tool in UCSC (University of California Santa Cruz) and the Primer3 software [211, 212]. Primers were designed to produce a product between 90-150bp. A standard curve of three standards were prepared and run for each primer set. To prepare the standard curve, a 10x serial dilution was prepared. The first standard was prepared by taking an equal volume of each experimental sample.

QPCR results were calculated and shown as relative to a housekeeping gene for gene expression assays (Section 2.13.3), as percentage of input for chromatin immunoprecipitation-sequencing (ChIP-seq) samples (Section 2.8.4) and relative to a DHS-negative region for DNase I-seq samples (Section 2.7.7). In experiments involving time course or different drug treatments, results were calculated and shown relative to the vehicle sample (Section 2.16).

Table 2.4: qPCR reagents.

qPCR reagent	Final concentration
x2 LightCycler 480 SYBR Green I Master mix (Roche)	1x
DNA	N
Forward primer	62nM
Reverse primer	62nM
Water	up to 12µl

N= x10 or x50; all DNA samples were diluted x10, except library DNA which was diluted x50. This is due to the library preparation method involving a PCR step, which increases DNA concentration.

Table 2.5: qPCR parameters.

Cycle step	Temperature (°C)	Time (Secs)	Number of cycles
Preincubation	95	300	
Amplification	95	10	x45
	60	10	
	72	10	
Melting curve	95	5	
	65	60	
Cooling	40	30	

2.6 Statistical analysis

The Roche LightCycler 480 software was used for analysis of qPCR data. The output after conducting qPCR was the Cp (cross point) value, which is the number of cycles taken for the DNA to cross fluorescence threshold. All qPCR reactions were conducted in technical duplicates, except quantification of sequencing libraries using the Kapa quantification kit, which were conducted in triplicates (Section 2.7.3). Any technical replicates differing by more than 0.5 cycle were repeated. The software utilises the standard curve to calculate mean concentration of samples.

Every experimental condition for each gene in gene expression studies (Section 2.13), cell viability assays (Section 2.16), western blots (Section 2.14) and apoptosis/cell cycle analyses by FACS (fluorescence-activated cell sorting) (Section 2.17) was tested using at least three biological replicates, thus error bars represent standard error of the mean (SEM). Statistical significance between two conditions was determined by performing an unpaired t-test. A statistically significant p-value was considered as <0.05. Bars shown for qPCR reactions for genome-wide quality control checks display the range between technical duplicates.

2.7 DNase I-seq

2.7.1 Overview

The DNase I-seq method was used to identify and map regulatory elements in the VU40T(-), SCC040(-), VU147(+) and UDSCC2(+) HNSCC cell lines. The assay utilises the DNase I enzyme, which digests accessible regions of the chromatin more quickly than closed regions, therefore by digesting for a short amount of time, accessible regions are released (DNase I hypersensitive sites (DHSs)) **(Figure 2.2)**. The assay was performed as previously described [190].

Nuclei isolation

Open chromatin digested with 30-50 units
of DNase I for 3mins at 37°C

Small fragments (<300bp)
released

Size fractionation

Collection of 20 DNA fractions

Visualisation by gel electrophoresis

**Fractions 4-8 contain target
DNA fragments (100-200bp)**

qPCR

Library preparation

Genome-wide sequencing
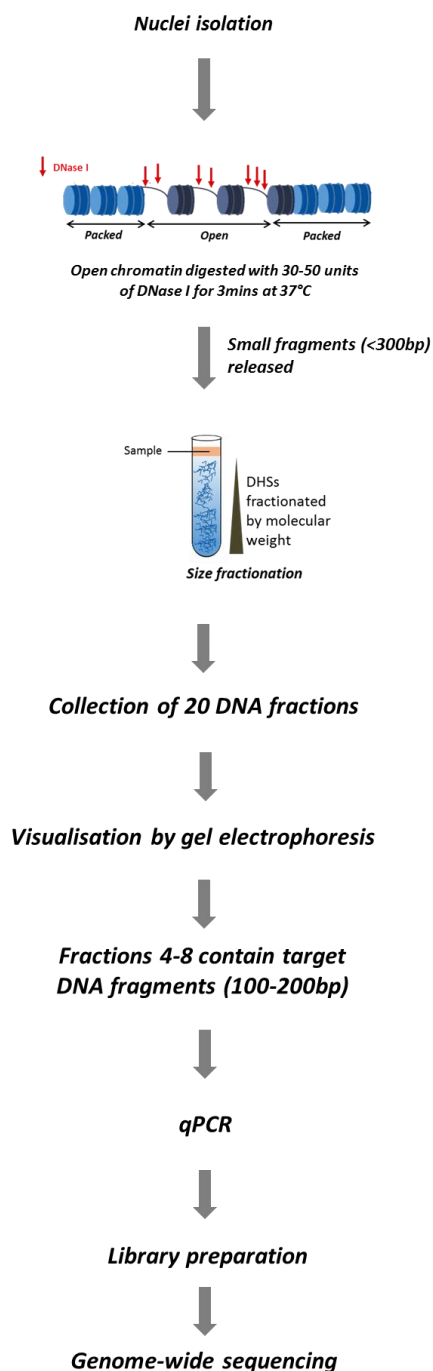
Figure 2.2. DNase I-seq flowchart.

Nuclei were isolated by treatment with a detergent (NP-40), at a concentration optimised for each cell line. The DNase I enzyme was used to preferentially digest open/unstable regions of the chromatin. Digestion was conducted for 3mins at 37ºC. A range of DNase I amounts between 30-50U were used, as the amount of DNase I can be cell type-specific. DNase I digestion releases small fragments of less than 300bp in size. DNA fragments were size-selected using a sucrose gradient fractionation method and 20 fractions were collected, which were visualised by gel electrophoresis. Due to limitations of the Illumina next-generation sequencing platform, DNA fragments between 100-200bp were desirable and were collected. The agarose gel revealed that fragments of this size separated into fractions 4-8. To verify collection of DNA, qPCR was conducted of genes known to be expressed in HNSCC and compared with a negative control region (No DHS). A DNA library was constructed, which was required to modify the DNA and facilitate compatibility with the sequencing platform, allowing genome-wide sequencing.

2.7.2 Nuclei isolation

Cell nuclei must first be isolated and intact nuclei with preserved chromatin structure are treated with DNase I. All buffers in this method were freshly prepared and filtered using the Corning 500ml vacuum filter/storage bottle system (Corning, UK). Before DNase I digestion, nuclei were isolated to allow DNase I access to the chromatin **(Figure 2.2)**. Cells were washed twice with pre-warmed (37ºC) PBS and harvested (Section 2.1.2). The cells were centrifuged at 387 *xg* at 4ºC for 5mins and washed twice with cold PBS supplemented with 1x complete EDTA-free protease inhibitor (PI) (Roche, UK). The supernatant was discarded and the suspension was resuspended in Buffer A (15mM Tris-HCl (pH 8.0), 100nM $NaCl_2$, 60mM KCl_2 1mM EDTA (pH 8.0), 0.5mM EGTA (egtazic acid) (pH 8.0), 100nM spermidine (Sigma-Aldrich, UK), 1x PI). The ratio of live:dead cells was calculated, as described in Section 2.1.3. Samples containing at least 90% live cells were carried forward. The concentration of cells was adjusted to $2x10^6$/ml in Buffer A. Next, an equal volume of Buffer A with the required concentration of NP-40 (Sigma-Aldrich, UK) was added and the cells were incubated for 10mins on ice. The concentration of NP-40 was adjusted for each cell line to give 95-100% nuclei. Four concentrations of NP-40 were used: 0.01-0.04% at increments of 0.01%. The optimal NP-40 concentration was determined by counting trypan stained nuclei (Section 2.1.3). After NP-40 treatment, the nuclei were centrifuged at 387 *xg* for 5mins at 4°C and the supernatant was discarded. The pellet was washed twice with PBS/1x PI solution and resuspended in 1ml of Buffer A. The nuclei were counted and adjusted to $2x10^7$/ml in Buffer A and centrifuged, as above.

2.7.3 DNase I digestion

The nuclei pellet was incubated at 37ºC for 1min. DNase I digestion buffer (150mM NaCl$_2$, 60mM CaCl$_2$, optimal DNase I amount (Roche, UK)) was prepared by adding 1ml of DNase I digestion buffer and 9ml of Buffer A and warming to 37°C for 10mins. The chromatin was digested by adding 1ml of DNase I digestion buffer per 1×10^7 of cells for 3mins at 37°C (**Figure 2.2**). The optimal amount of DNase I was established for each cell line by using three different amounts; 30, 40 or 50 units. This step was required as under-digestion of the sample with DNase I can lead to incomplete isolation of DHSs, while over-digestion can lead to the inclusion of non-specific and inaccessible sites. A negative DNase I control was also used to test for DNA degradation. The digestion was terminated by adding an equal volume of stop buffer (50mM Tris-HCl (pH 8.0), 150mM NaCl, 0.1% (v/v) SDS solution, 1mM EDTA (pH 8.0), 100nM spermidine (Roche, UK), RNase (500µg/ml) (Roche, UK)). The reaction was incubated overnight at 55°C. After the incubation, proteinase K (50ug/ml) (Roche, UK) was added and the sample was incubated at 55ºC for a further 5-7hrs. DNase I digestion was verified by gel electrophoresis on a 2% agarose gel (Section 2.3) (**Figure 2.2**).

2.7.4 Phenol/chloroform extraction

The DNA was purified by adding an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 (saturated with 10mM Tris (pH 8.0), 1mM EDTA) (Sigma-Aldrich, UK). The sample was mixed by inverting the tube until the solution appeared white and centrifuged at 2,755 *xg* at RT for 10mins. The aqueous phase containing the DNA was removed without disturbing the water/phenol interface and transferred to a fresh tube.

2.7.5 DHS fractionation

To conduct sucrose gradient DNA fractionation, a 9% (v/v) sucrose aqueous solution (Sigma-Aldrich, UK) was prepared and filtered **(Figure 2.2)**. Ten millilitres of the sucrose solution was added to a thin wall polypropylene centrifuge tube (Beckman Coulter, UK) and 2ml of the purified DNA was layered on the top. The sample was centrifuged using the Beckman Coulter Ultracentrifuge, SW40 Ti rotor at 25,000 rpm, RT for 24hrs. Nineteen fractions were collected using a syringe carefully positioned at the top of the solution. A volume of 500µl was taken for each fraction and were visualised by agarose electrophoresis to ensure the correct sized fractions were collected **(Figure 2.2)**.

2.7.6 DNA precipitation

Of the 19 fractions collected, those containing the 100-200bp DNA fragments were precipitated by adding 10µg/ml glycogen (Invitrogen), 10% (v/v) 3M sodium acetate (pH 7.0) (Sigma- Aldrich, UK) and 70% (v/v) isopropanol (Sigma-Aldrich, UK). The sample was incubated at -20ºC overnight to precipitate the DNA, centrifuged at 17,000 *xg* for 30mins and the supernatant was discarded. The pellet was washed with 70% ethanol, centrifuged, and was air-dried and resuspended in 25µl of nuclease-free water. The DNA was taken forward for quality control qPCR (next Section) and library preparation (Section 2.9).

2.7.7 DNase I hypersensitivity assay quality control qPCR

To confirm enrichment of DHSs, qPCR was conducted, as described in Section 2.5 **(Figure 2.2)**. QPCR primers were designed based on three positive control regions, one at the keratin 6A (*KRT6A*) promoter, which was selected as it is known to be

important in epithelial development and is constitutively expressed in epithelial cells [4, 6, 213, 214] **(Table 2.6)**. In addition, the region was used as a positive control in preliminary DNase I experiments with the UDSCC2(+) (Wiench, unpublished). Based on the initial experiment, two positive control primers were designed at two DHS peaks near the *IL-6* gene **(Table 2.6)**. The *IL-6* gene has been reported as being overexpressed in HNSCC [75]. In addition, a negative control region was selected near to the *KRT6A* gene, which lacks any DHSs based on ENCODE (ENCyclopaedia Of DNA Elements) data of 125 cell lines [164, 215]. To determine DHS enrichment by qPCR, a ratio was calculated from the amplification signals of the positive control regions against the negative control region. The required ratio was at least five (positive signal:negative signal), as described in Results, Section 3.6.

DNase I preparations that pass the quality control were taken for library preparation and sequenced as described in Section 2.9.

Table 2.6. DNase I quality control qPCR primer sequences.

Primer name	Forward (5' to 3')	Reverse (5' to 3')
Keratin 6A DHS	GAAGGTGAGCTTGCAGGTTG	GCTGGAAGGCAGGAGAATTT
No DHS	GTCAGGCTGGTCTCTCACTC	AAATACGGCCAGGCATGTTG
IL-6 Site 1	AGAAGCCACGCGGTGGCAAA	TCTGGCTCTCCCTGTGAGCGG
IL-6 Site 2	AGCGCTAGCCTCAATGACGAC	GCGGGTGGGGCTGATTGGAA

2.8 Chromatin immunoprecipitation (ChIP)

To investigate and identify genome-wide binding of transcription factors (TFs) the chromatin immunoprecipitation (ChIP)-seq method was used. This assay allows the identification of protein-DNA interactions by crosslinking and isolating protein-bound DNA fragments. The crosslinked complexes can then be de-crosslinked, the DNA can be sequenced and mapped to the human genome.

2.8.1 Chromatin collection

All buffers in this method were freshly prepared and filtered using the Corning 500ml vacuum filter/storage bottle system (Corning, UK). The ChIP assay was conducted as previously described [216]. The VU40T(-) and VU147(+) cells were cultured on 150mm dishes until 70% confluent. The medium was discarded, and cells were fixed using 10ml of pre-warmed (37ºC) and 1% (v/v) formaldehyde in PBS and incubated for 10mins at 37ºC **(Figure 2.3)**. The crosslinking reaction was inhibited by addition of 1ml of 2.5M glycine and then incubation for 5mins at RT, while rocking. The fixative was discarded and the cells were washed twice with ice-cold PBS. The cells were then scraped into a centrifuge tube and centrifuged for 4mins at 780 *xg* at 4ºC. The supernatant was discarded and the cell pellet was resuspended in 500μl SDS lysis buffer (1% (v/v) SDS, 10mM EDTA, 50mM Tris-HCL (pH 8.0), 1x PI) **(Figure 2.3)**. The cells were incubated on ice for 30mins and in order to shear the DNA by sonication the cell suspension was transferred to sonication tubes (Active Motif, UK).
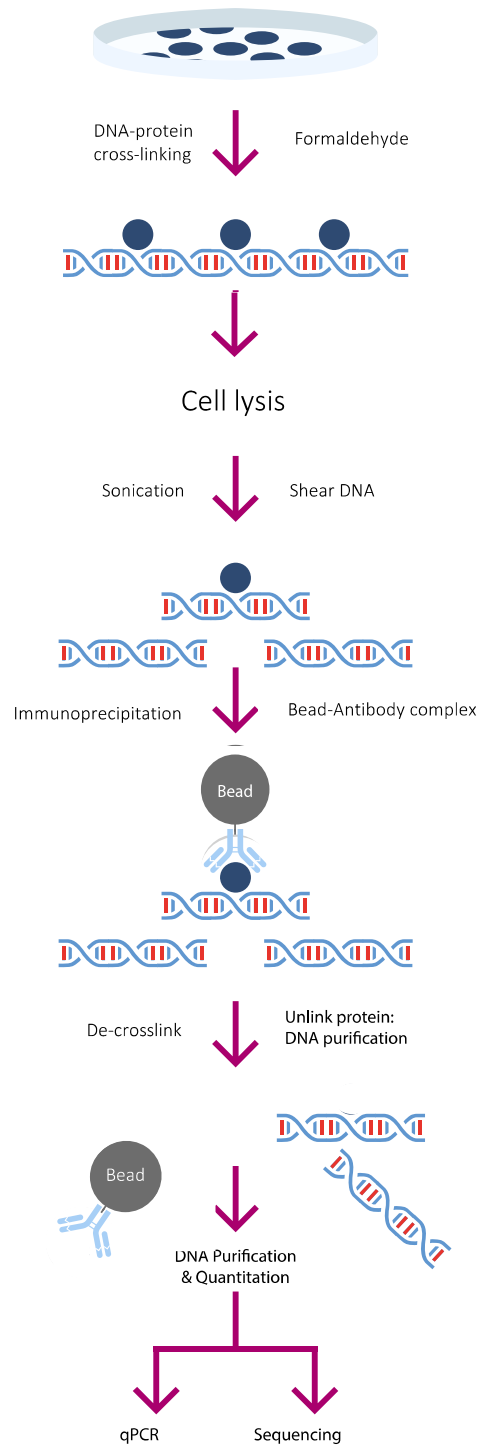
Figure 2.3. Chromatin immunoprecipitation (ChIP)-seq.

Cells were crosslinked using formaldehyde to preserve the DNA-protein complexes. The cells were lysed using an SDS buffer to isolate the chromatin. Chromatin was sheared by sonication in order to yield DNA fragments of ~200bp. Magnetic dynabeads were crosslinked to the antibody in preparation for the chromatin-bound protein of interest to bind the antibody-bound beads. Chromatin was also pre-cleared with the Dynabeads, to reduce background. The protein and DNA were de-crosslinked to isolate DNA fragments and qPCR or next-generation sequencing was conducted to interrogate DNA.

2.8.2 Sonication

Sonication was conducted using the EpiShear sonicator (Active Motif, UK) and the parameters used were 85% amplitude for 14 cycles (30secs of sonication and 30secs of cooling per cycle) **(Figure 2.3)**. After sonication, the sample was transferred to a centrifuge tube and centrifuged for 10mins at 13,000 xg, at $4°C$. A 5µl aliquot of the sonicated chromatin was taken to assemble a reaction for DNA quantification by Nanodrop Spectrophotometry (Section 2.4). The reaction for DNA quantification was de-crosslinked by adding 5µl of sample, 0.2M NaCl, 4mg/ml proteinase K (Roche, UK), 92% (v/v) SDS lysis buffer to a volume of 200µl and incubated at 65°C for 1hr. A second aliquot of sonicated chromatin was taken to visualise DNA fragment size by agarose gel electrophoresis (Section 2.3). The chromatin for DNA visualisation was de-crosslinked by adding 10µl of sample to 0.2M NaCl, 4mg/ml proteinase K (Roche, UK) and 21% (v/v) SDS lysis buffer to a volume of 20µl. The reaction was incubated for 1hr at 45ºC, followed by 4hrs at 65°C. The remaining sample was diluted five times in ChIP dilution buffer (0.01% (v/v) SDS, 1.1% (v/v) Triton-X 100, 1.2mM EDTA, 16.7mM Tris-HCl (pH 8.1), 167mM NaCl). Chromatin aliquots of 100µg were made and stored at $-80°C$.

2.8.3 Chromatin immunoprecipitation (ChIP)

ChIPs were performed using antibodies for p63, JUND, FOSL1 (FOS like 1, AP-1 transcription factor subunit) (FRA-1) and TEAD (TEA domain transcription factor)-4 TFs **(Table 2.7)**. The assay involved crosslinking magnetic Dynabeads with an antibody to form bead-antibody complexes **(Figure 2.3)**. Next, chromatin was added to the bead-antibody complexes, so that the antibody binds the chromatin

proteins. The complex was eluted and de-crosslinked, releasing the DNA fragment to which the protein was bound.

To crosslink the antibody to beads, 60µl of the magnetic beads, Dynabeads™ M-280 Sheep IgG anti-rabbit or anti-mouse (Thermo Fisher Scientific, UK) were added to centrifuge tubes. The beads were washed twice with 1ml of PBS for 5mins at 4ºC while rotating. To separate the beads from solution, a DynaMag™-2 Magnet magnetic rack was used. The tubes were placed on the rack for 1min until the solution appeared clear and the supernatant was removed. The tubes were removed from the rack and the corresponding species of antibody was added to the washed beads and pipetted several times to ensure the beads go into solution. A 'no antibody' control, containing PBS in place of antibody was also prepared. The beads and antibody with 1x PI were crosslinked for 6hrs. The beads were washed twice with low salt immune complex buffer (0.1% (v/v) SDS, 1% (v/v) Triton x-100, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 150mM NaCl) for 5mins at 4ºC while rotating.

Table 2.7. ChIP antibodies.

Antibody	Amount (µg per ChIP)	Species	Company (Catalogue No.)
TEAD4	10	Mouse	Abcam (ab58310)
P63	5	Mouse	Abcam (ab735)
JUND (329)	1	Rabbit	Santa Cruz (sc-74)
Fra-1 (FOSL1) (R-20)	1	Rabbit	Santa Cruz (sc-605)

Simultaneously, to reduce non-specific binding, chromatin was pre-cleared. To pre-clear, fresh beads were washed as described before. Next, 100μg of chromatin per ChIP was added to the washed beads and the suspension was incubated for 2hrs at 4ºC while rotating. The pre-cleared chromatin was separated from the beads as before. An aliquot of 10% of the chromatin sample volume was taken to serve as a background control for genomic deletions, insertions or amplifications (input DNA). The input DNA was processed at the elution step. The remaining pre-cleared chromatin was added to the antibody-bound beads and incubated overnight at 4ºC while rotating to ensure that the chromatin proteins bind to the antibody (**Figure 2.3**). The chromatin-bound beads were washed sequentially with 1ml of the following buffers for 15mins each: low salt immune complex buffer, high salt immune complex buffer (0.1% (v/v) SDS, 1% (v/v) Triton X-100, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 500mM NaCl), LiCl immune complex buffer (0.25M LiCl, 1% (v/v) NP-40, 1% (v/v) deoxycholate, 1mM EDTA, 10mM Tris-HCl (pH 8.1)) and 1x Tris-EDTA (TE) buffer (pH 8.0)).

The DNA from the ChIP samples and input were eluted and de-crosslinked from protein complexes by adding 400μl of elution buffer (10mM Tris-HCl (pH 8.0), 300mM NaCl, 55mM EDTA, 0.5% (v/v) SDS) and 50μg/1ml proteinase K (**Figure 2.3**). The beads were incubated 6-16hrs at 65°C, while shaking. The supernatant was removed and transferred to a centrifuge tube. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and thoroughly mixed by pipetting. A Phase Lock Light Gel (5PRIME, UK) was used to extract the aqueous phase of the solution. To do this, the Phase Lock tube was centrifuged at 12,000 *xg* for 30secs at RT to collect the gel at the bottom of the tube and the sample was transferred to the tube. To separate the aqueous phase from the organic phase, the

tube was centrifuged at 12,000 *xg* for 5mins at RT. The upper aqueous phase was transferred into a centrifuge tube. The DNA was precipitated as described in Section 2.7.6.

2.8.4 ChIP quality control qPCR

To determine ChIP enrichment, qPCR was conducted as described in Section 2.5. To identify potential binding sites for p63, TEAD4, JUND and FOSL1 TFs, the Cistrome Data Browser was used [217]. ChIP data for HNSCC cell lines used in this project were not available in the database, therefore data from similar cell types were used. Previously published data report overlap between DHSs and TF binding, therefore each TF site was verified as a DHS using our DNase I-seq data [188, 191, 218] **(Table 2.8)**. The selected binding sites are shown in Results, Chapter five, Section 5.6 and Appendix: Figures 3-7.

Table 2.8. ChIP qPCR primers.

Primer name	Forward (5' to 3')	Genomic location
p63 and TEAD4 ChIPs		
p63 site 1 (Fw)	ACCAGAAGCAATTATATGTGCGT	chr1:3,628,730-3,628,843
p63 site 1 (Rv)	GCTGCAATTTCGCCCTAAGT	
p63 site 2 (Fw)	TCTAAGGAGCAGTGGTAAGCC	chr15:42125252+42125365
p63 site 2 (Rv)	CACTCCCACCCAAAGCATTC	
TEAD4 site 1 (Fw)	ATCTCAGAATCCCAGTGCCC	chr12:47665103+47665245
TEAD4 site 1 (Rv)	ACTTCAGACAGAGACGAGGC	
TEAD4 site 2 (Fw)	ATACACCGCCACACTCACAG	chr12:52541325+52541450
TEAD4 site 2 (Rv)	GAATGTCTCACAGCACCACC	
JUND and FOSL1 ChIPs		
PC (Fw)	GCCTGGGATGACTAAGAGCT	chr11:66659803+66659903
PC Rv)	GGACTGTTGCTTTGCCCTTT	
LMNA (Fw)	CTGCCTCTGTACAATGCCAC	chr1:156093605+156093694
LMNA (Rv)	CGCAACTCCAGTAGAAGTGC	
ANXA2 (Fw)	CCCTACTGACTTGTTGCAGC	chr15:60685428+60685524
ANXA2 (Rv)	TTCTTCCTGCCGGTCTCTTT	
S100A10 (Fw)	GCTTCAAGGTCTGCTGTGAG	chr1:151961138+151961253
S100A10 (Rv)	CACTGCCATCTGAAACCCAT	

DNA sequences were downloaded using the UCSC browser. Primers were designed using Primer3 and were checked using the *in-silico* PCR design tool in UCSC. Primers were designed to produce a product between 90-150bp.

2.9 DNase I-seq and ChIP-seq library preparation

2.9.1 Library preparation overview

The DNA isolated during ChIP and DHS preparations must be made compatible for the Illumina next-generation sequencing platform. Therefore, the DNA was converted into sequencing libraries **(Figure 2.4)**. The synthesis of libraries involved three main reactions; the first was 'end-repair' in which the ends of fragmented DNA is repaired to produce blunt ends. DNA fragmentation can be caused by

DNase I digestion or shearing by sonication. The end-repair reaction required two enzymes, T4 polynucleotide kinase and T4 DNA polymerase. The first enzyme added phosphate groups to the 5' ends of damaged DNA fragments, while the second enzyme incorporated dNTPs (nucleoside triphosphates) and created blunt ends by removing any overhangs on the 5' and 3' ends of the DNA.

The second step was dA-tailing, which required the Klenow Large fragment in the presence of dAMPs (deoxyadenosine monophosphates) to produce dA-tails to the 3' end of DNA. This reaction prevented the formation of concatemers during the ligation step and facilitated adapter ligation **(Figure 2.4)**.

The third step is a ligation step to ligate adapters to either end of the DNA fragments. The adapters are short sequences of nucleotides and serve two main purposes; they allow the DNA fragments to bind to the sequencing platform and they contain an index sequence unique to each sample. This allowed the samples to be pooled and sequenced, while still being identified during data analysis. The libraries were then size-selected, followed by PCR amplification **(Figure 2.4)**.

NEB Next

- DNA end-repair
- AMPure XP bead clean-up
- dA-tailing
- Adapter ligation and USER enzyme digestion
- AMPure XP bead clean-up
- PCR Enrichment
- AMPure XP bead clean-up
- Size selection using AMPure beads
- AMPure XP bead clean-up

KAPA Hyper

- DNA end-repair and dA-tailing
- Adapter ligation
- AMPure XP bead clean-up and size-selection
- PCR Enrichment
- AMPure XP bead clean-up
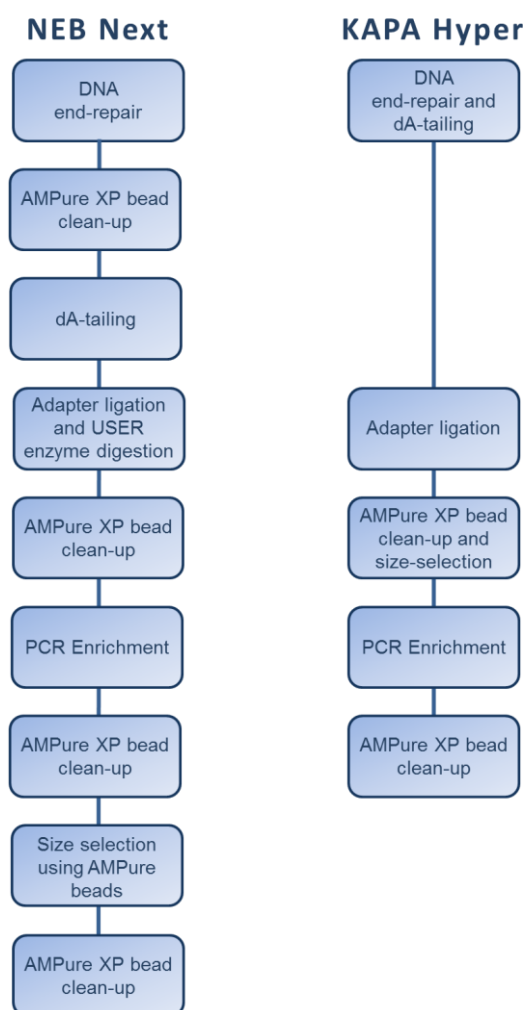
Figure 2.4. Library preparation workflow.
The NEB Next workflow: The **DNA end-repair** step, repairs the ends of fragmented DNA to produce blunt ends. The 5' and 3' overhangs are removed and 5' phosphate group added. The **AMPureXP bead clean-up** steps were required to remove salts and nucleotides which may interfere with downstream steps. The **dA-tailing step** was required to prevent the formation of concatemers during the ligation step and facilitated adapter ligation. The **adapter ligation** step was required to allow identification of each sample during sequencing analysis, as samples are pooled for sequencing and the adapter contains a unique index of 6 nucleotides. This step also involves digestion by the Uracil-Specific Excision Reagent (USER™) Enzyme. The USER enzyme cuts the adapter at a uracil nucleotide, creating a gap of one nucleotide, changing the conformation of the adapter so it is able to bind to the Illumina flowcell. **PCR enrichment** was required to increase library concentration. **Size selection** was conducted using the AMPure beads, where the ratio of beads:DNA could be adjusted to select for the desired DNA fragment size. **Kapa Hyper prep workflow:** This workflow combines the DNA end-repair and dA-tailing steps. The size-selection step is combined with a clean-up step. This is followed by PCR enrichment and the final clean-up step.

2.9.2 DNase I-seq library preparation

DNase I libraries were constructed using NEBNext ChIP-Seq Library Prep Master Mix Set and the Multiplex Oligos (Index Primers Set 1) (NEB, UK). All clean-up steps were conducted using Agencourt AMPure XP beads (Beckman Coulter, UK). Each use of the AMPure beads involved the following steps: the beads were vortexed to resuspend, added to the sample and pipetted several times to mix. The beads were incubated with the sample for 5mins at RT. A magnetic rack was used to separate the beads from solution. The supernatant was discarded, and the beads were washed twice for 30secs with 80% fresh ethanol and dried for 5mins before resuspension.

Library preparation was conducted by first performing the end repair reaction. The reaction was assembled by adding 10ng of DNA, 5μl of 10x NEBNext End Repair Reaction Buffer and 1μl of NEBNext End Repair Enzyme Mix (204 units/ml of T4 polynucleotide kinase and 61 units/ml of T4 DNA polymerase) **(Figure 2.4)**. The reaction volume was made up to 50μl with nuclease-free water and incubated at 20ºC for 30mins. To clean-up the DNA, 1.8x volumes of AMPure XP beads were added to the reaction and cleaned up, as described before. The beads were resuspended in 50μl of 0.1x TE buffer. The beads and DNA were separated as before and 44μl of DNA was taken forward.

The dA-tailing step was conducted by adding 44μl of DNA, 5μl of 10x NEBNext dA-Tailing Reaction Buffer and 1μl of the Klenow Fragment (3' to 5' exonuclease) **(Figure 2.4)**. The reaction was incubated at 37°C for 30mins and the DNA was cleaned-up as before. The DNA was eluted in 25μl 0.1x TE buffer.

To conduct adapter ligation, one of the twelve NEBNext adapters were ligated by combining 19μl of the DNA, 6μl of 5x Quick Ligation Reaction buffer, 1μl of 1.5μM NEBNext Adaptor and 4μl of 5x Quick T4 DNA Ligase **(Figure 2.4)**. The reaction was incubated for 15mins at RT. One microlitre of the adapter Uracil-Specific Excision Reagent (USER™) Enzyme Mix (1 unit/μl) was added to the reaction and incubated for a further 15mins at 37°C **(Figure 2.4)**. The DNA was cleaned up as before and eluted in 105μl of 0.1x TE buffer.

The DNA was size-selected using the appropriate volume of beads to obtain the optimum beads:DNA ratio **(Figure 2.4)**. The desired size for sequencing was 270bp (total library size), therefore 0.9x volumes of beads were used. The beads and DNA were incubated for 5mins at RT. A further 0.2x volumes of beads were added to clean-up the DNA as before. The DNA was eluted in 20μl of 0.1x TE buffer.

The library was amplified by PCR. The reaction was assembled by taking 20μl of adaptor-ligated DNA, 25μl of NEBNext High-Fidelity 2x PCR Master Mix, 2.5μM Universal PCR Primer, 2.5μM Index *N* Primer ('*N*' denotes the index number). The PCR amplification was conducted as shown in **Table 2.9**. A final clean-up step was conducted as described before and qPCR was conducted to verify DHS enrichment in libraries as described in Sections 2.5 and 2.7.7.

Table 2.9. PCR parameter for NEB Next library preparation.

Cycle step	Temperature (°C)	Time (secs)	Number of cycles
Initial denaturation	98	45	
Denaturation	98	15	
Annealing	60	30	x15
Extension	72	30	
Final extension	72	300	

2.9.3 ChIP-seq library preparation

ChIP libraries were prepared using the Kapa Hyper Prep kit (Roche, UK). The workflow for this kit offers the advantage of being less time consuming, as fewer clean-up steps are required and some of the steps are combined **(Figure 2.4)**. As described for DNase I library preparation, all clean-up steps were conducted using Agencourt AMPure XP beads (Beckman Coulter, UK). One change to the clean-up step was that the beads were incubated with the sample for 10mins at RT, instead of 5mins.

For each ChIP sample, 10ng of sample was made up to 50μl. The End Repair and A-Tailing buffer mix and enzyme mixes were pre-mixed (7μl and 3μl, respectively) and were added to the sample. The reaction was incubated at RT for 30mins and then at 65ºC for 30mins.

Adapter ligation was conducted at a ratio of 200:1 adapter:DNA and incubated for 15mins at RT. The DNA was cleaned-up by adding 0.8x volumes of AMPure beads. The supernatant was removed, and the beads were washed as before, with the added step of a 2mins incubation in 25μl elution buffer (10mM Tris-HCl). The supernatant was transferred to a fresh tube.

PCR was carried out by adding 25μl of the 2x Kapa HiFi HotStart ReadyMix and 5μl of 10x Kapa Library Amplification Primer Mix to the DNA and amplified as shown in **Table 2.10**. The sample was cleaned up by adding one volume of beads and the remaining clean-up steps were conducted as above.

To size-select DNA fragments of 150-350bp, 0.6x volumes of beads to DNA were added and incubated for 10mins. The beads were washed as above and eluted in

27μl of elution buffer. Aliquots of the sample were taken for quality control checks: 1.5μl of sample was taken for BioAnalyzer analysis (Section 2.11.2), 1μl was taken for Kapa quantification (Section 2.11.3) and the remainder was taken forward for Illumina next-generation sequencing (Section 2.12).

Table 2.10. PCR parameters for Kapa Hyper library preparation.

Cycle step	Temperature (°C)	Time (secs)	Number of cycles
Initial denaturation	98	45	
Denaturation	98	15	
Annealing	60	30	x10
Extension	72	30	
Final extension	72	60	

2.10 ATAC (Assay for Transposase-Accessible Chromatin)-sequencing

2.10.1 Overview

The Assay for Transposase Accessible Chromatin (ATAC)-seq is a method to map accessible chromatin and regulatory elements. The ATAC-seq method utilises a transposase enzyme called, Tn5, which probes and inserts sequencing adapters at accessible regions of the chromatin, combining the digestion and library preparation into one step [181]. The transposed DNA fragments were amplified by PCR. QPCR was conducted on the amplified transposed DNA to ensure sequencing libraries are minimally amplified, while still obtaining sufficient material to sequence. Library over-amplification may introduce GC bias. QPCR revealed how many more PCR amplification cycles are needed, therefore a second PCR step may have been required.

2.10.2 Nuclei isolation

A total of 50,000 cells were harvested from the OTK cells and were pelleted at 500 *xg* for 5mins, 4°C. The supernatant was discarded and the cell pellet was washed once with cold PBS by centrifuging at 500 *xg* for 10mins at 4°C. The cell pellet was resuspended in 50µl of cold lysis buffer (10mM Tris-HCl (pH 7.4), 10mM NaCl, 3mM MgCl2, 0.1% (v/v) IGEPAL CA-630). The lysate was centrifuged at 500 *xg* for 10mins at 4°C and the supernatant was discarded.

2.10.3 Transposition reaction and purification

The Nextera DNA Library Preparation Kit (Illumina, UK) was used to carry out the transposition reaction. The reaction was assembled by adding 25µl TD (2x reaction buffer), 2.5µl TDE1 (Nextera Tn5 Transposase) and 22.5µl nuclease-free water to a tube. The lysed cells were resuspended in the transposition reaction mix and incubated at 37°C for 30mins.

The transposed DNA was then purified using the MinElute PCR Purification kit (Qiagen, UK) as per manufacturer's instructions. All centrifugation steps were conducted at 17,900 *xg* for 1min at RT. First, to allow DNA to bind the spin column membrane, 5 volumes of Buffer PB were added to the reaction volume and transferred to the column. The column was centrifuged and the flow-through was discarded. Next, the column was washed with 750µl of Buffer PE. The column was centrifuged and the flow-through was discarded. The column was centrifuged again to remove any residual Buffer PE. The transposed DNA was eluted by adding 10µl of elution buffer and collected by centrifugation.

2.10.4 PCR amplification

The transposed DNA was amplified by PCR, followed by qPCR to determine the amount of DNA in the sample and establish the number of additional cycles required for the second PCR. In the first PCR, the number of amplification cycles were restricted to 5 cycles to limit GC bias and maintain DNA complexity (**Tables 2.11 and 2.12**). The qPCR reaction was assembled alongside a DNA negative control. The qPCR setup is outlined in **Tables 2.13 and 2.14**. To calculate the number of additional cycles required, the qPCR fluorescence signal was plotted against cycle number. The cycle number which corresponded to 25% of the maximum fluorescence intensity was the number of additional cycles needed.

Table 2.11. PCR reaction setup for ATAC libraries

Reagent	Volume (µl)
Transposed DNA	10
Nuclease-free water	10
25µM Custom Nextera PCR Primer 1	2.5
25µM Custom Nextera PCR Primer 2 (with barcode)	2.5
NEBNext High-Fidelity 2x PCR Master Mix	25

Table 2.12. PCR parameters for ATAC libraries.

Cycle Step	Temperature (ºC)	Time (secs)	Number of cycles
Extension	72	300	1
Initial denaturation	98	30	1
Denaturation	98	10	x5 + N
Annealing	63	30	
Final extension	72	60	

'N' denotes the number of additional cycles required after the first PCR step.

Table 2.13. qPCR reaction for ATAC libraries.

Reagent	Volume (µl)
PCR amplified DNA or water	5
Nuclease-free water	4.41
25µM Custom Nextera PCR Primer 1	0.25
25µM Custom Nextera PCR Primer 2	0.25
100x SYBR Green I	0.09
NEBNext High-Fidelity 2x PCR Master Mix	5µl

Table 2.14. qPCR parameters for ATAC libraries.

Cycle Step	Temperature (ºC)	Time (secs)	Number of cycles
Initial denaturation	98	30	1
Denaturation	98	10	x20
Annealing	63	30	
Final extension	72	60	

2.11 Quality control of sequencing libraries

2.11.1 Quality control qPCR

To verify enrichment of DNA in the DNase I hypersensitivity and ChIP assays sequencing libraries, qPCR was conducted using the parameters described in **Tables 2.4 and 2.5**. Primers used for DNase I and ChIP qPCRs are described in Sections 2.7.7 and 2.8.4, respectively.

2.11.2 Bioanalyzer

The Agilent 2100 Bioanalyzer instrument (Agilent Technologies, UK) was used to determine the amount of DNA in the libraries, as well as determine the library size. Libraries were diluted 50x and the DNA fragment size was verified by using the Agilent High Sensitivity DNA Assay protocol, as per the manufacturer's instructions (Agilent Technologies, UK). In order to be able to visualise the DNA,

a gel-dye mix was first prepared. The High Sensitivity dye concentrate (dye) and High Sensitivity DNA gel matrix reagent (gel) were equilibrated at RT for 30mins. Next, 15µl of the dye was added to the gel and were mixed by vortexing. A 'cleaning microfluidic chip' was loaded with 350µl of nuclease-free water and placed into the BioAnalyzer machine for 10secs to clean the electrodes. The cleaning Chip was removed, and the electrodes were left to dry until the samples were ready to be run. The samples were run using the microfluidic Agilent High Sensitivity DNA Chip (Chip). The Chip was prepared by loading 9µl of the gel-dye mix into the appropriate well, marked 'G'. The Chip was then loaded into the Priming Station, which applies pressure to remove any bubbles. The plunger of the Priming Station was pushed down for 60secs and then released. The remaining wells marked 'G' were loaded with 9µl of the gel-dye mix. Next, 5µl of the High Sensitivity DNA marker (marker) was loaded into the 'ladder' well, followed by 1µl of the High Sensitivity DNA ladder. Subsequently, 1µl of the sample was loaded into each of the sample wells or 1µl of marker into any 'unused' wells.

The Chip was then placed into an IKA MS2 Vortex Mixer (IKA Works) and vortexed for30 secs at 2,000 rpm before being loaded into the BioAnalyzer. Analysis was conducted using the 2100 Expert Software where an electropherogram was produced by plotting the fluorescence signal of the sample against the size in bp. The fluorescence signal is relative to the DNA ladder and so the amount of DNA can be determined.

2.11.3 Kapa Quantification

Libraries were quantified using the Kapa SYBR FAST LightCycler 480 qPCR kit, as per manufacturer's instructions. First, 1ml of 10x Illumina Primer Premix was

added to the 5ml 2x Kapa FAST qPCR Master Mix and was vortexed to mix. A serial dilution of each library was conducted, 1:1000, 1:2000 and 1:4000 to account for library concentrations which may be out of range of the six known standard concentrations. The libraries were diluted in library dilution buffer (10mM Tris-HCl (pH 8.0), 0.05% (v/v) Tween-20). A master mix was assembled of 12μl of Kapa SYBR FAST qPCR Master Mix and 4μl of water per sample. The master mix was added to the wells of a 96-well plate, followed by 4μl of diluted DNA or DNA standards 1-6. Samples were loaded in triplicates. The parameters for qPCR are shown in **Table 2.15**. For the analysis, the concentration in pM was calculated from the average of the three technical replicate library samples. The average pM was multiplied by the average bp size of DNA standards (452bp), divided by the average fragment size of the library in bp as indicated the BioAnalyzer (Section 2.11.2). The undiluted library concentration was then calculated by multiplying with the corresponding dilution factor.

Table 2.15. qPCR parameters for Kapa quantification of libraries.

Cycle Step	Temperature (ºC)	Time (Secs)	Number of cycles
Initial denaturation	95	300	
Denaturation	95	30	x35
Annealing/Extension	60	45	

2.12 Illumina sequencing and bioinformatics analysis

2.12.1 Illumina sequencing

Libraries were indexed according to the number of samples to be sequenced on each sequencing lane. Eight indexed libraries were pooled per lane and sequenced on the

HiSeq 2500 platform, using the V3 SBS chemistry (1x50 cycles) (Illumina, UK). A 10nM pool (20µl) of indexed of libraries was assembled.

2.12.2 Genome alignment

After library sequencing, a FASTQ file was obtained containing sequences of reads. Each read is assigned a quality score between 1 and 40. Using the HOMER (Hypergeometric Optimization of Motif EnRichment) software low quality reads, i.e. any read with a score of less than 30 that points to ambiguous nucleotide callings, were removed. Next, the Bowtie2 tool was used to align the reads to the hg19 version of the human genome [219]. The output was a BAM file, which contains the alignment data for the reads. The HOMER software was used to remove any duplicated and ENCODE blacklisted reads [220]. Blacklisted reads were first identified by the ENCODE consortium and are defined as reads which, when aligned to the genome, can be anomalous or display a high or low signal regardless of cell type and experimental technique from which they were derived [164]. Non-uniquely mapped reads were filtered out and peaks were called using the 'factor mode' in HOMER based on a 350bp fragment size [220, 221]. The BEDTools suite was used to calculate coverage maps, whose output includes reads mapped to chromosome number and the coverage depth (the number of reads for each nucleotide) [221]. When the coverage maps were created, a Bigwig file was also generated. This file was uploaded to the UCSC Genome Browser to allow for visualization of the cumulated reads [212]. Any analyses using HOMER and subsequent analyses were conducted by Dr Sam Clokie (West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital).

2.12.3 Transcription factor motif and pathway enrichment analyses

Potential TF binding sites within DHSs or ChIP peaks were identified through motif analysis using the 'findMotifsGenome' script within HOMER. Motifs were identified by first, controlling for background by taking at least 50,000 randomly selected background sequences from the DHS sequences (target sequences). Enriched motifs were identified by calculating the frequency of target sequences versus background sequences. The GC content of the target DHSs were also considered in order to avoid GC bias, which due to the presence of CpG islands (regions of dense C and G bases), can be introduced during the PCR step of library preparation [159, 222]. TF motifs were identified based on sequence similarity from a database of known TF motifs, derived from published ChIP-seq data.

Pathway enrichment analysis of genes of interest was conducted using the 'AnnotatePeaks' script within HOMER. This analysis assigns each DHS, enhancer cluster (EC), ATAC or ChIP peak to the closest transcription start site (TSS) of a gene to create a list of potentially regulated genes. The list of genes was run through a combination of two pathway overrepresentation databases: Kyoto Encyclopaedia of Genes and Genomes (KEGG) [223] and WikiPathways (WP) [224]. This analysis was also conducted using genes obtained from RNA-seq analyses.

2.12.4 Enhancer cluster identification

The 'superenhancer' script within HOMER was used to identify enhancer clusters (ECs) using 'called' DHS peaks. EC identification was based on the method first described by Whyte *et al.* [173]. DHS peaks within a distance of 12.5kb were

identified and concatenated. The EC signal was determined by a normalisation step, which involved normalisation of all reads minus normalised input reads. The putative enhancer signal was then sorted by score and normalised to the highest score, as well as to the number of putative enhancers. Any regions with a score above one were determined as an EC. However, it must be noted that in order to truly label these regions as enhancers, ChIP-seq assays for known enhancer marks and functional reporter assays must be conducted. Therefore, this method was used as a means of reinforcing data revealed by DNase I-seq, ChIP-seq and RNA-seq, rather than a method to identify novel regulators.

2.13 Gene expression analysis

2.13.1 RNA extraction

Cells were cultured in 6-well plates (Corning, UK) until 70% confluent and RNA was extracted using the RNeasy Mini kit (Qiagen, UK) as per manufacturer's instructions. The cells were washed twice in cold PBS. To lyse the cells and denature any RNases, 500μl of Buffer RLT containing 0.01% (v/v) β-mercaptoethanol was added. Next, 100% ethanol was added 1:1 to the sample, which was loaded onto the RNeasy Mini spin column and was centrifuged 13,000 *xg* for 30secs at RT, the flow-through was discarded. The RNA bound to the column membrane was washed twice with Buffer RPE and eluted with 20μl of nuclease-free water. The RNA was quantified on the Nanodrop 2000 Spectrophotometer (Section 2.4).

2.13.2 cDNA synthesis

CDNA synthesis was conducted using the iScript cDNA synthesis kit (Bio-Rad, UK). The reaction was set up by adding 4µl of 5x iScript reaction mix, 1µl of iScript reverse transcriptase to 1µg RNA and made up to 20µl with nuclease-free water. The reaction consisted of three steps, primer annealing, DNA polymerisation and enzyme deactivation. The reaction was incubated at 25ºC for 5mins (annealing), 42ºC for 30mins (polymerisation) and 85ºC for 5mins (enzyme deactivation).

2.13.3 qPCR

The cDNA samples were diluted 1:4 in nuclease-free water and qPCR was performed as described in Section 2.5. Primer sequences are shown in **Table 2.16**. The transcript amplification signal was normalised to the signal for the housekeeping genes *GAPDH* and *β-ACTIN*. Three biological replicates of RNA were used for cDNA synthesis and qPCR was conducted in technical duplicates.

Table 2.16. Gene expression primers

Gene	Forward (5'-3')	Reverse (5'-3')
ATF3	TTGCCATCCAGAACAAGCAC	CCTCGGCTTTTGTGATGGAC
BACH1	AACAGTGTCGTCAGAGTGGT	AGACGCTGCCAAAACTTCAG
β-ACTIN	AAAGACCTGTACGCCAACAC	GTCATACTCCTGCTTGCTGAT
FOS	GCTTCAACGCAGACTACGAG	AGTGACCGTGGGAATGAAGT
FOSL1 (FRA-1)	GACCTACCCTCAGTACAGCC	TCAGTTCCTTCCTCCGGTTC
FOSL2 (FRA-2)	GACAGGGTGTGAGTGAGACA	AGACGCCCTACTCAAGACAC
GAPDH	CCTGGCCAAGGTCATCCAT	AGGGGCCATCCACAGTCTT
JUN	CAACATGCTCAGGGAACAGG	CCCGACGGTCTCTCTTCAAA
JUNB	TGCACAAGATGAACCACGTG	GCTGAGGTTGGTGTAAACGG
JUND	AAATTCTCCGCCCCTTTCCT	AAAACAGAAAACCGGGCGAA
TP63	GTTTCGACGTGTCCTTCCAG	TCTGGATGGGGCATGTCTTT
TEAD1	TGCAAGGTTTGAGAATGGCC	ATGTTGTGCTCCGTCTTCAC
TEAD2	CATGACCCTCACCTGTTCCT	CCAGGACGCTGTTCATCATG
TEAD3	CTTCCTTGCCCCTTTGTGTC	AAACCCAGAGACCTACGCTC
TEAD4	AAGGATCTCTTCGAACGGGG	TGTTCTCGGGGCTCTCATAC

RNA sequences were downloaded from the UCSC browser. Primers were designed using Primer3 and were checked using the *in-silico* PCR design tool in UCSC. Where possible, primers were designed over an intron-exon junction to prevent amplification of contaminating DNA. Primers were designed to produce a product between 90-150bp.

2.13.4 RNA-seq

RNA was extracted, as described in Section 2.13.1. Three biological replicates were extracted and pooled in equal amounts. The libraries were prepared and sequenced by the Environmental Omics Facility, University of Birmingham. The sample pools were quantified using the Qubit Fluorometer (Thermo Fisher Scientific, UK). To assess RNA integrity, the RNA integrity number (RIN) was determined by the Agilent TapeStation instrument. All samples displayed a RIN of >9, indicating intact RNA [225]. Libraries for RNA-seq were made using the TruSeq Stranded mRNA Library Prep kit (Illumina, UK). The libraries were sequenced using the HiSeq 2500 instrument.

2.13.5 RNA-seq analysis

Within the Galaxy platform, the FASTQC file and the 'Trim sequences' tool were used assess and remove any reads with a Phred quality score of <30, as well as removing adapter sequences. The 'Hisat2' tool was used to align reads to the hg19 genome [226, 227]. The 'Stringtie' tool was used to assemble and quantify transcripts and each gene was assigned a FPKM value (fragments per kilobase exon per million mapped reads) [228]. The FPKM was normalised by the number of reads mapped and the gene length. The cut-off point for the FPKM value was set at 1 and genes with FPKM values above 1 were taken for pathway enrichment analysis using the Consensus Pathway database [229-231].

Principal components analyses were conducted using the 'PCAexplorer' script in the R software (Dr Sam Clokie, West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital).

2.14 Western blot analysis

2.14.1 Protein lysate collection

Cells were cultured in 100mm cell culture dishes until 70% confluent. The dishes were placed on ice and washed twice with ice-cold PBS. To lyse the cells, 500µl of RIPA lysis buffer (150mM NaCl, 1% (v/v) NP-40, 0.5% (v/v) sodium deoxycholate, 0.1% (v/v) SDS, 50mM Tris (pH 8.0)) was added to each dish and left on ice for 2mins. The lysate was scraped and transferred to a 1.5ml tube. The lysate was incubated for 45mins at 4ºC while rotating. The insoluble protein and cell debris were collected by centrifuging at 13,000 xg for 10mins and the RIPA soluble proteins was transferred to a tube.

2.14.2 Bradford protein assay

The Bradford assay was used to measure protein concentration. First, a standard curve was prepared using eight known bovine serum albumin (BSA) protein concentrations. To measure protein concentration, 200µl of 5x Bradford reagent (Bio-Rad, UK), 5µl sample or RIPA buffer (blank) and 795µl water were combined and added to a cuvette and measured on a nanophotometer (Geneflow).

2.14.3 Protein loading

To denature the protein, 1x Laemmli buffer (40% (v/v) glycerol, 0.4% (v/v) β-mercaptoethanol, 0.04% (w/v) bromophenol blue, 0.1M Tris-HCl (pH 6.8), 4% (v/v) SDS) was added to adjust protein concentration to 0.5µg/µl. The sample was heated to 95ºC for 10mins and centrifuged briefly to pool any liquid. Five

microlitres of Pageruler Prestained Protein ladder (Thermo Fisher Scientific) was loaded into each gel, followed by 10µg of each sample.

2.14.4 SDS-PAGE gel electrophoresis

Gels for sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) were prepared using the Mini Protean 3 Bio-Rad apparatus (Bio-Rad, UK). A 10% resolving gel (10% (v/v) 37.5:1 acylamide:bisacrylamide (Protogel, UK), 375mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS (sodium dodecyl sulphate) (Sigma-Aldrich, UK), 0.1% (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED) (Sigma-Aldrich, UK) and 0.1% (w/v) ammonium persulphate (APS) (Sigma-Aldrich, UK)) was made. The gel was covered with a layer of 100% isopropanol to achieve a uniform gel edge. Once the resolving gel had polymerised, the isopropanol was removed, and the remaining was soaked up with filter paper.

A stacking gel (4.5% (v/v) 37.5:1 acylamide:bisacrylamide (Protogel, UK), 375mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) TEMED and 0.1% (w/v) APS (ammonium persulphate)) was poured on top of the resolving gel. A comb was placed in the stacking gel and the gel was left to polymerise. The gels were transferred into MiniProtean 3 tanks (Bio-Rad, UK), containing 1x running buffer (25mM Tris-HCl (pH 7.5), 2mM glycine (Thermo Fisher Scientific) and 10% (w/v) SDS). The protein samples were separated by electrophoresis at 100V for 1hr.

2.14.5 Electrophoretic protein transfer

Two pieces of sponge, two pieces of Whatman chromatography paper (Sigma-Aldrich, UK) and one piece of Amersham Hybond P 0.45µm PVDF (polyvinylidene difluoride) membrane (GE Healthcare Life Sciences, UK) were cut

to gel size. The sponge and Whatman paper were soaked in transfer buffer (100mM Tris-HCl (pH 7.2), 200mM glycine (Sigma-Aldrich, UK), 0.05% (w/v) SDS, 20% (v/v) methanol). The membrane was soaked in 100% methanol, rinsed with distilled water and transferred to transfer buffer. To transfer, one sponge, Whatman paper, gel, membrane, Whatman paper and sponge were assembled in a Bio-Rad transfer cassette (Bio-Rad, UK). The 'sandwich' was flattened to ensure that any air bubbles were forced out. The cassette was put into a transfer tank (Bio-Rad, UK) and filled with transfer buffer. An ice pack was used to ensure the transfer buffer remained cool and the protein was transferred for 1hr at 200 V.

2.14.6 Protein visualisation

All washes were conducted in PBST (1x PBS with 0.1% (v/v) Tween-20) three times for 5mins. Non-specific binding was eliminated by incubating the membrane in 20% (w/v) non-fat milk (Marvel, Premier Foods) in PBST at RT for 1hr. When a membrane was probed with more than one antibody, whose proteins had different molecular weights, the membrane was cut with a scalpel before the blocking step. After blocking, the membrane was washed and incubated with a primary antibody in 5% (w/v) BSA or 5% (w/v) milk in PBST, either overnight at 4ºC or 1hr at RT **(Table 2.17)**. To assess for equal loading of protein, the membrane was also probed with the housekeeping protein lamin A/C **(Table 2.17)**. The membrane was washed and incubated with horse radish peroxidase (HRP)-conjugated secondary IgG antibody (NEB, UK) for 1hr at RT. If the molecular weight of a protein overlapped with that of lamin A/C, then the membrane was stripped. To strip, the membrane was incubated with stripping buffer (1.5% (w/v) glycine, 0.1% (w/v) SDS, 0.01%

(v/v) Tween-20 (pH 2.2)) for 15mins at RT and the buffer was replaced halfway through the incubation.

To detect the protein signal, the EZ-ECL (enhanced chemiluminescence) detection kit for HRP (Biological Industries, UK) was used. The detection reagents were mixed (1:1) and 1ml was applied to each membrane for 1min. Excess ECL reagent was removed by soaking up with a paper towel. The membrane was wrapped in plastic film and fixed into place in an autoradiography cassette. In a dark room, a Hyperfilm X-ray film (Amersham Biosciences, UK) was placed into the cassette for 1min and the exposure time was adjusted accordingly. The film was developed using a PROTEC ECOMAX automatic film processor (PROTEC GmbH & Co.).

Table 2.17. Antibodies for Western blotting.

Antibody	Species	Dilution	Incubation	Company (Catalogue No.)
Primary antibodies				
Lamin A/C	Rabbit	1: 10,000 in 5% BSA	1hr at RT	Santa Cruz (sc-20681)
EGFR	Rabbit	1: 1,000 in 5% milk	1hr at RT	Abcam (ab52894)
JUND (329)	Rabbit	1:200 in 5% milk	Overnight at 4°C	Santa Cruz (sc-74)
Fra-1 (FOSL1) (R-20)	Rabbit	1:200 in 5% milk	Overnight at 4°C	Santa Cruz (sc-605)
Secondary antibodies				
HRP-conjugated secondary IgG	Rabbit	1: 5,000 in 5% milk	1hr at RT	Cell Signalling (70745)

2.15 Preparation of drugs for cell culture

Verteporfin (VP) (Sigma-Aldrich, UK) was dissolved in 1ml of 100% DMSO (dimethyl sulfoxide) (Sigma-Aldrich, UK) as a stock concentration of 2mg/ml. The stock VP was aliquoted in 10μl aliquots, stored at -20°C and freeze-thawed once

before discarding. VP is light-sensitive and therefore was protected from light by wrapping in foil. Tyrphostin (AG1478) (Sigma-Aldrich, UK) was dissolved in 1ml of 100% ethanol to give a stock concentration of 1mg/ml. The stock AG1478 was stored at -20ºC in 20µl aliquots. Cetuximab was unavailable for purchase and therefore was requested from Merck, Germany. The drug was given as a ready-made stock solution of 5mg/ml. The complete composition of the drug buffer was not revealed. However, it was known that the buffer contained 0.01% Tween-80 (Personal communication with Dr. C. Stroh, Merck, Germany). The drug was stored in aliquots at 4ºC.

2.16 Cell viability assay

To assess changes in cell viability after drug treatment, the CellTiter-Blue assay (Promega, UK) was used. This kit is based on a reduction reaction of the blue non-fluorescent reagent, resazurin, to the pink fluorescent reagent resorufin by dehydrogenase enzymes. The reaction only takes place in the presence of dehydrogenase, which is produced by metabolically active cells.

VU40T(-) and VU147(+) cells were seeded at a density of 0.3×10^3 cells/well and 2.4×10^3/well, respectively in 96-well plates in a volume of 80µl. Six technical replicates were plated for each of the ten conditions. Two control conditions were used: (i) cells treated with vehicle and (ii) media in the absence of cells to control for background absorbance **(Table 2.18)**. The remaining conditions were treatments of eight different concentrations of verteporfin (VP) (Sigma-Aldrich, UK), Cetuximab (Merck, UK) or tyrphostin (AG1478) (Sigma-Aldrich, UK). The drugs and vehicles were made up in microcentrifuge tubes before 20µl of each was added

to the appropriate wells. The drugs were added to the cells 8hrs after seeding and the cells were treated for 96hrs. The media and drugs were replaced after 48hrs.

Following 96hrs of cell culture, 20µl of the CellTiter-Blue reagent (Promega, UK) was added to each well. The cells were cultured for approximately another 4hrs until the colour changed from blue to pink. The plate was read on a Synergy HT microplate reader (BioTek, UK). The absorbance for resazurin and resorufin were read at 600nm and 507nm, respectively. To determine changes in cell viability, firstly, the difference in absorbance was calculated by subtracting the absorbance of resorufin from the absorbance of resazurin. Secondly, to adjust for background, the absorbance signal for media only was subtracted. The final value was normalized to cells treated with vehicle only. At least three biological replicates were obtained for each condition. Statistical analysis was conducted as described in Section 2.6. The SigmaPlot software (Systat Software Inc, USA) was used to generate four-parameter logistic dose-response curves.

Table 2.18. Cell viability assay template.

	-Cells	+Cells								
		Vehicle	Drug (µM)							
Verteporfin (mM)	Media only	DMSO	720	360	180	90	45	22.5	11.25	5.6
Cetuximab (µM)	Media only	Ethanol	34.3	3.4	2.7	2	1.3	0.67	0.17	0.03
AG1478 (µM)	Media only	Ethanol	400	200	100	50	25	12.5	6.25	3.125

The 'media only' condition served as a control for background absorbance. The media, vehicle and cells condition served as a control for cell viability changes induced by the vehicle. VU40T(-) and VU147(+) cells were seeded 8hrs before drug treatment to allow the cells to adhere. Each treatment was for 96hrs and the drug and controls were replaced after 48hrs. Six technical replicates of each condition were plated. The experiment was repeated at least three times.

2.17 Fluorescence-Activated Cell Sorting (FACS)

Propidium iodide staining followed by fluorescence-activated cell sorting (FACS) was used to investigate changes in the cell cycle after drug treatment. Propidium iodide is a dye, which intercalates into DNA stoichiometrically. Once intercalated, the subsequent fluorescence levels of propidium iodide can be measured and are directly proportional to the amount of DNA [232]. Annexin V staining was used as a marker of apoptosis. The annexin V protein has high affinity for the phosphatidylserine receptor, which is exposed upon the induction of apoptosis [233].

2.17.1 Propidium iodide-staining

To prepare cells for propidium iodide staining and cell cycle analysis, the cells were harvested and counted as described in Sections 2.1.2 and 2.1.3, respectively. After counting, $1x10^6$ cells were fixed by adding 1ml of 70% cold ethanol. The ethanol was added drop-wise while vortexing to minimise cell clumping. The cells were incubated at 4ºC overnight and washed twice with cold PBS and centrifuged for 5mins at 850 xg, 4°C. The supernatant was discarded and 50µl of RNase (100µg/ml) and 200µl of propidium iodide (50 µg/ml) (Sigma-Aldrich, UK) were added to the cells and transferred to a 5ml polypropylene round bottom tube (BD Falcon, UK).

2.17.2 Annexin V-DNA double staining

To prepare the cells for annexin V-DNA double staining, the cells were harvested, counted as before and were prepared using the eBioscience Annexin V Apoptosis Detection FITC (fluorescein isothiocyanate) Kit, as per manufacturer's instructions

106

(Thermo Fisher Scientific, UK). The cells were washed twice with cold PBS for 5mins at 4ºC for 450 *xg* and resuspended at a concentration of 1×10^6 cells/ml in 1x Binding Buffer (10µM Hepes (pH 7.4), 140µM NaCl, 2.5mM $CaCl_2$). To detect annexin V binding, 100µl of the cell solution was transferred to a centrifuge tube and 5µl of annexin V conjugated with the fluorophore FITC (annexin V-FITC) was added to the cells. The cells were incubated at RT in the dark for 10-15mins. Next, the cells were washed in 2ml of 1x Binding Buffer and centrifuged for 5mins at 450 *xg*, RT. To stain for propidium iodide, the supernatant was discarded, and the cell pellet was resuspended in 200µl of Binding Buffer and 5µl of propidium iodide. The suspension was made up to 500µl with PBS. Two control conditions were included; cells were stained with only propidium iodide or only FITC to assess any bleed-through of signal. Post-acquisition analysis was conducted on the CyAn ADP flow cytometer (Beckman Coulter, UK).

2.17.3 FACS analysis

To gate out cell doublets and debris, two scatter plots were used, pulse width vs. forward scatter (FSC) and FSC vs. side scatter (SSC).

To determine DNA content and cell cycle distribution of cells, a histogram of propidium iodide emission vs. cell count was plotted. The appropriate peaks were gated to measure propidium iodide emission in the G1, S and G2/M phases. To measure levels of apoptosis, a scatter plot for propidium iodide vs. FITC was plotted and divided into quadrants based on subpopulations of necrotic cells (high propidium iodide and low FITC emission), apoptotic cells (low or high propidium iodide and high FITC emission) and viable cells (low PI and low FITC emission).

CHAPTER 3: CHARACTERISING THE *CIS*-REGULATORY NETWORK IN HNSCC

3.1 Introduction

Research efforts into understanding molecular changes underlying head and neck cancer (HNSCC) have been intensifying over the last few years. For example, The Cancer Genome Atlas (TCGA) consortium have highlighted somatic gene alterations in a panel of 279 HNSCC tumours, some of which reflect the clinical outcome of HNSCC patients [36]. For example, HNSCC patients harbouring mutations of the *TP53* gene and amplification of chromosome 11q13 have a less favourable clinical outcome than those without [36]. Furthermore, EGFR (epidermal growth factor receptor) overexpression and reduced overall patient survival particularly in HNSCC has been reported [36, 234]. However, despite a flurry of genome-wide HNSCC studies being published over the last few years [36, 66-71], the five-year survival rate of patients with HPV (human papillomavirus)-negative (HPV(-)) tumours currently stands at 25-40% and 70-80% for HPV-positive (HPV(+)) patients [101]. In addition, only a few new therapies are in clinical trials, and survival rate for patients who fail to respond to first-line therapy is less than one year [235]. The *cis*-regulatory network in HNSCC has not previously been investigated and filling this gap could help to pinpoint new therapeutic targets.

3.2 Aims

The first aim of this project was to identify and characterise regulatory elements in HNSCC cell lines. The second aim was to compare differences in regulatory regions in HPV(-) and HPV(+) HNSCC cells to highlight differences in *cis*-regulation underlying both tumour types.

To investigate the above aims, the DNase I-hypersensitivity assay followed by next generation sequencing (DNase I-seq) was conducted in four HNSCC cell lines; two HPV(-): VU40T(-) and SCC040(-) and two HPV(+): VU147(+) and UDSCC2(+). Transcription factor (TF) motif analysis and pathway enrichment analysis were applied to identify key transcriptional regulators. The findings were also compared to the genome-wide gene expression data obtained through RNA-seq.

3.3 DNase I hypersensitivity assay development: an overview

To isolate regulatory elements, the DNase I hypersensitivity assay was conducted in each of the above cell lines, followed by next-generation sequencing (DNase I-seq) (**Figure 3.1**). The assay is briefly described as follows and will be discussed in more detail later. Nuclei were isolated using a detergent (NP-40) to disrupt the cell membrane, while maintaining intact nuclei with preserved chromatin structure. The concentration of NP-40 was optimised for each cell line, as the concentration can be cell line-specific (**Figure 3.1: OP1**). DNase I digestion can also be cell-line specific and so the amount of DNase I enzyme was optimised (**Figure 3.1: OP2**). Since the enzyme digests open chromatin regions more readily than closed regions, by limiting the digestion time, only accessible regions were released, which are called DNase I hypersensitive sites (DHSs). DNase I over-digestion can result in

the collection of unspecific fragments, while under-digestion can lead to a poor yield of DHSs. Therefore, before conducting the experiment, the best experimental conditions were established for each HNSCC cell line by optimising the amount of DNase I and concentration of NP-40 used.

Once the optimisation steps were established, the efficiency of DNase I digestion was determined by agarose gel electrophoresis for each experiment (**Figure 3.1: QC1**). The digested chromatin was fractionated by size using ultra-centrifugation and the resulting fractions were visualised by agarose gel to ensure the DNA had been fractionated (**Figure 3.1: QC2**). Fractions enriched for DNA fragments between 100bp and 200bp were collected for downstream library preparation. QPCR (quantitative polymerase chain reaction) was also conducted as a quality control step to verify DHS enrichment by using DHS-specific primer sets (**Figure 3.1: QC3**). Library preparation was required to modify the DNA and make it compatible for the Illumina sequencing platform. Once libraries were prepared, they were amplified by PCR (polymerase chain reaction) and amplification was verified by qPCR (**Figure 3.1: QC4**). Library size was verified using a microfluidics gel-based system (Agilent Bioanalyzer system) (**Figure 3.1: QC5**) and libraries were quantified using a qPCR-based method (Kapa quantification) (**Figure 3.1: QC6**). Post sequencing genome alignment, peak calling and analyses is described from Section 3.9 onwards. The quality control (QC) measures were applied for every DHS test protocol and the actual experiment, as described below.
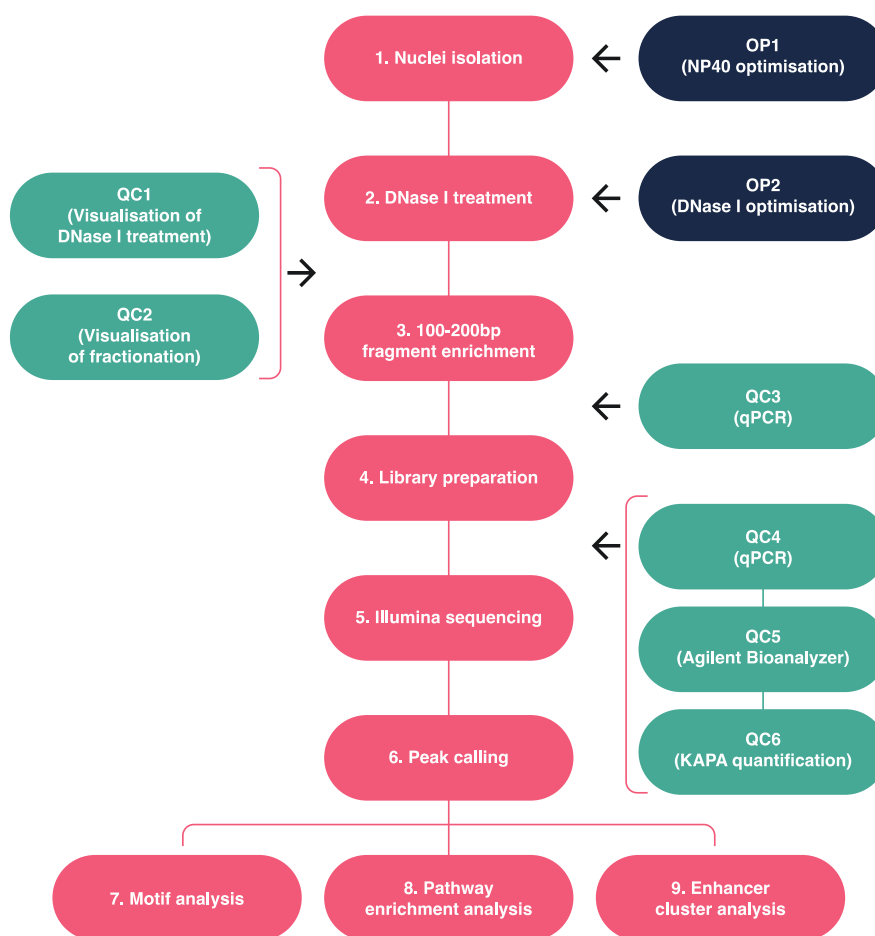
Figure 3.1. Overview of DNase I-seq assay optimisation (OP) and quality control (QC) steps.

Isolation of nuclei was optimised for each cell line by using a range of NP-40 concentrations **(OP1)**. The concentration of DNase I was optimised for each cell line to avoid incomplete digestion or over-digestion of the chromatin **(OP2)** and visualised on an agarose gel **(QC1)**. The fragments were separated by size using the sucrose gradient method and separation was visualised by agarose gel electrophoresis **(QC2)**. Fragments between 100-200bp were collected **(3)** and enrichment was assessed by qPCR using primers specific for known open chromatin regions and comparing to signal obtained from a 'no DHS' negative region **(QC3)**. Library preparation involved modifying the samples by adding adapters for Illumina HiSeq compatibility and indexed for identification during analysis **(4)**. Successful library preparation was assessed by performing qPCR on the library sample **(QC4)**. The total library size, which includes the insert and adapters, was measured using the Agilent Bioanalyzer system **(QC5)**. The concentration of library was measured using a qPCR-based assay (Kapa quantification kit) **(QC6)**. After sequencing, peaks were called using the HOMER software as described in Materials and Methods **(6)** and motif analysis **(7)**, pathway enrichment analysis **(8)** and enhancer cluster analysis **(9)** was conducted using the 'findMotifsGenome', 'AnnotatePeaks' and 'superenhancer' scripts in HOMER.

3.4 NP-40 and DNase I optimisation

To isolate the maximum number of nuclei without over-digesting the cells, the concentration of NP-40 was optimised in a small-scale test using $1x10^6$ cells per preparation. The optimisation step for NP-40 provided an indication of the concentration to use for the actual experiment. Four concentrations from 0.01% to 0.04% were tested at increments of 0.01%. The percentage of intact nuclei was assessed for each cell line by conducting trypan blue staining. In SCC040(-), VU40T(-) and VU147(+) cells the optimal NP-40 concentration used was established to be 0.02%, while in UDSCC2(+) cells it was 0.01%. The optimum NP-40 concentrations yielded between 51-60% nuclei.

To optimise the amount of DNase I, a small-scale test was conducted and a range of DNase I concentrations was tested for each cell line: 30U, 40U or 50U of DNase I per $1x10^6$ nuclei. To control for degraded chromatin, a sample was included where no DNase I enzyme was added. After digestion, the samples were run on a 2% agarose gel for visual inspection. The desirable level of DNase I digestion resulted in a limited release of smaller DNA fragments from the high molecular weight DNA band and was compared to the untreated negative control, as exemplified in **Figure 3.2**. Based on this optimisation, 50U of DNase I was selected for follow up experiments in all four HNSCC cell lines.
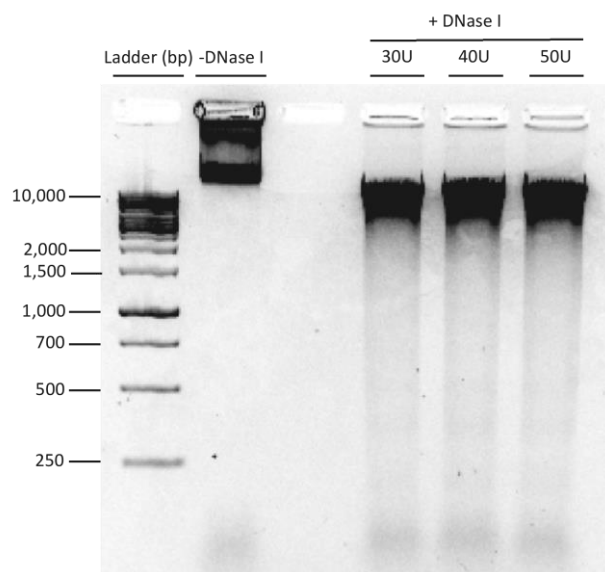
Figure 3.2. Agarose gel electrophoresis of DNase I digested DNA in SCC040(-) cells.
Nuclei from SCC040(-) cells were isolated and treated with 30U, 40U or 50U of DNase I
enzyme for 3 mins at 37°C. The samples were visualised on a 2% agarose gel. A 'no DNase
I' negative control was used to assess for endogenous chromatin digestion. The absence of
a DNA smear in the negative control indicates lack of endogenous chromatin digestion.

3.5 DNase I digestion and fractionation

After DNase I digestion, DNA was isolated from the chromatin and the fragments
were separated by size using a sucrose gradient **(Figure 3.1)**. A total of 19 fractions
containing the separated DNA fragments were collected for visualization on an
agarose gel **(Figure 3.3)**. Fragments between 100-200bp were enriched for library
preparation. It was observed that fractions 1-3 contained DNA fragment sizes less
than 100bp, while fractions 4-8 contained fragments largely between 100-200bp
(Figure 3.3). To verify this, fractions 4-8 were combined and visualized on the
Agilent Bioanalyzer system, as exemplified in **Figure 3.4**. The peak containing
fractions 4-8 displayed up to 300 fluorescence units. Bleed-through of DNA
fragments above the desired size of 200bp was observed. However, the majority of
the DNA were within the desired range and DNA fragments beyond this range were

filtered out during the size-selection step of library preparation, as described in

Materials and Methods, Section 2.9.
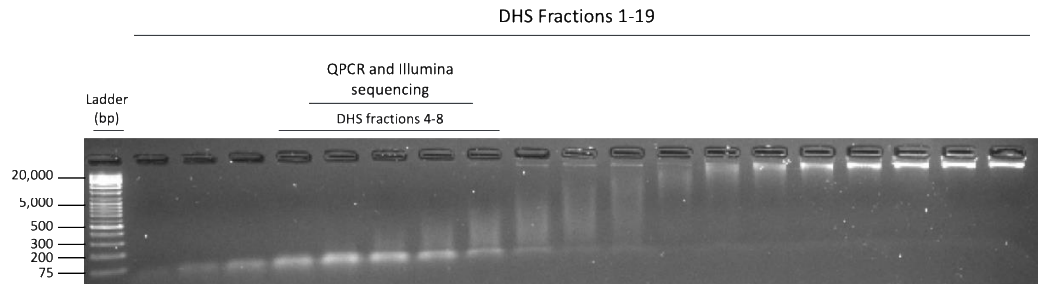
Figure 3.3. Agarose gel electrophoresis of fractionated DNase I fragments in SCC040(-) cells.
Generated DNA fragments were separated by ultra-centrifugation on a sucrose gradient. A total of 19 fractions were collected and visualised on an agarose gel to check for successful fractionation. Fractions containing DHSs of 100-200bp were merged (fractions 4-8).
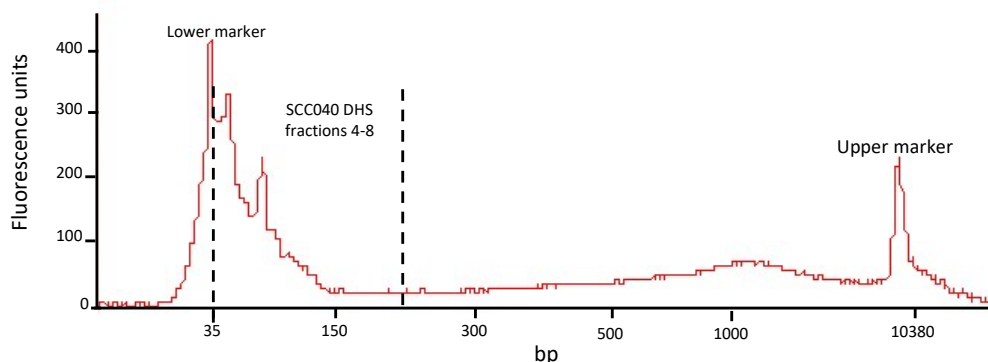
Figure 3.4. Agilent Bioanalyzer electropherogram for fractions 4-8 from SCC040(-) cells.
DNA fragments were fractionated by size and fractions 4-8 were collected. The lower marker peak is observed at 35bp and the upper marker peak at 10,380bp. The peak between 35bp and 200bp represents the combined 4-8 fractions (indicated by the dashed lines).

3.6 QPCR primer design

DHS enrichment in fractions 4-8 was assessed by conducting qPCR amplification of three positive control regions; keratin 6A (*KRT6A*) DHS region **(Figure 3.5)** and two DHS sites near the interleukin-6 (*IL-6*) gene (*IL-6* site 1 and site 2) **(Figure 3.6)**. The *KRT6A* gene was selected as it is known to be constitutively expressed in epithelial cells and was used as a positive control region in initial experiments for UDSCC2(+) (Unpublished, Wiench) [213, 214]. Upon sequencing and mapping the UDSCC2(+) DHS fragments, two DHS peaks close to the *IL-6* gene were observed **(Figure 3.6)**. The *IL-6* gene is often overexpressed in HNSCC and the IL-6 protein is a regulator of the HNSCC-related pathway, PI3K (phosphotidylinositol-3 kinase)/AKT (protein kinase B)/mTOR (mechanistic target of rapamycin) [75]. Therefore, it was hypothesised that amplification of *IL-6* may reflect DHS enrichment more closely in these cells than *KRT6A* amplification [75]. Two primer

sets were designed to the two DHS peaks close to *IL-6* (site 1 and site 2) **(Figure 3.6)**.

In addition to the positive control primers, a negative control region was selected near the *KRT6A* gene, which lacks any DHSs based on ENCODE (ENCyclopaedia Of DNA Elements) data of 125 cell lines [164] **(Figure 3.5)**. To determine DHS enrichment by qPCR, a ratio was calculated from the amplification signals of the positive control regions against the negative control region. Based on data obtained from sequenced samples, which displayed high background (not shown), a fold change of at least five (positive signal:negative signal) was set as the qPCR signal cut-off.
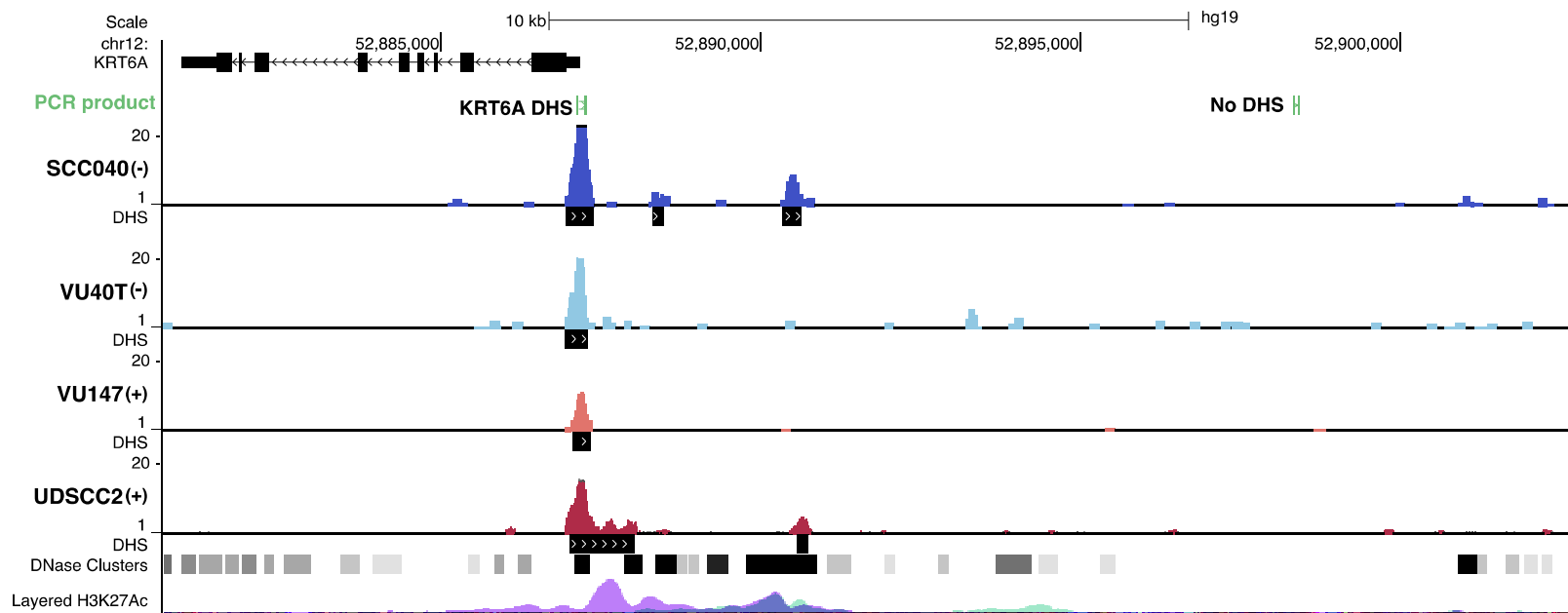
Figure 3.5. Primer locations for quality control of DNase I-digested samples at the *KRT6A* gene.

This figure displays the *KRT6A* DHS from subsequent DNase I-seq experiments conducted in HNSCC cell lines. Bigwig and BED files for SCC040(-) (dark blue), VU40T(-) (light blue), VU147(+) (orange) and UDSCC2(+) (red) were uploaded to the UCSC (University of California Santa Cruz) browser and visualised as coverage tracks. Identified (called) peaks (the cumulative reads for each location) for each cell line are depicted below the coverage tracks (depicted by the black bars). Primers were designed at a *KRT6A* DHS peak close to the 5' end (*KRT6A* DHS), which serves as a positive control signal for qPCR of epithelial cells. Primers were also designed, based on the ENCODE data, to a nearby region lacking DHS peaks ('no DHS') to assess for over-digestion and increased background signal.

117
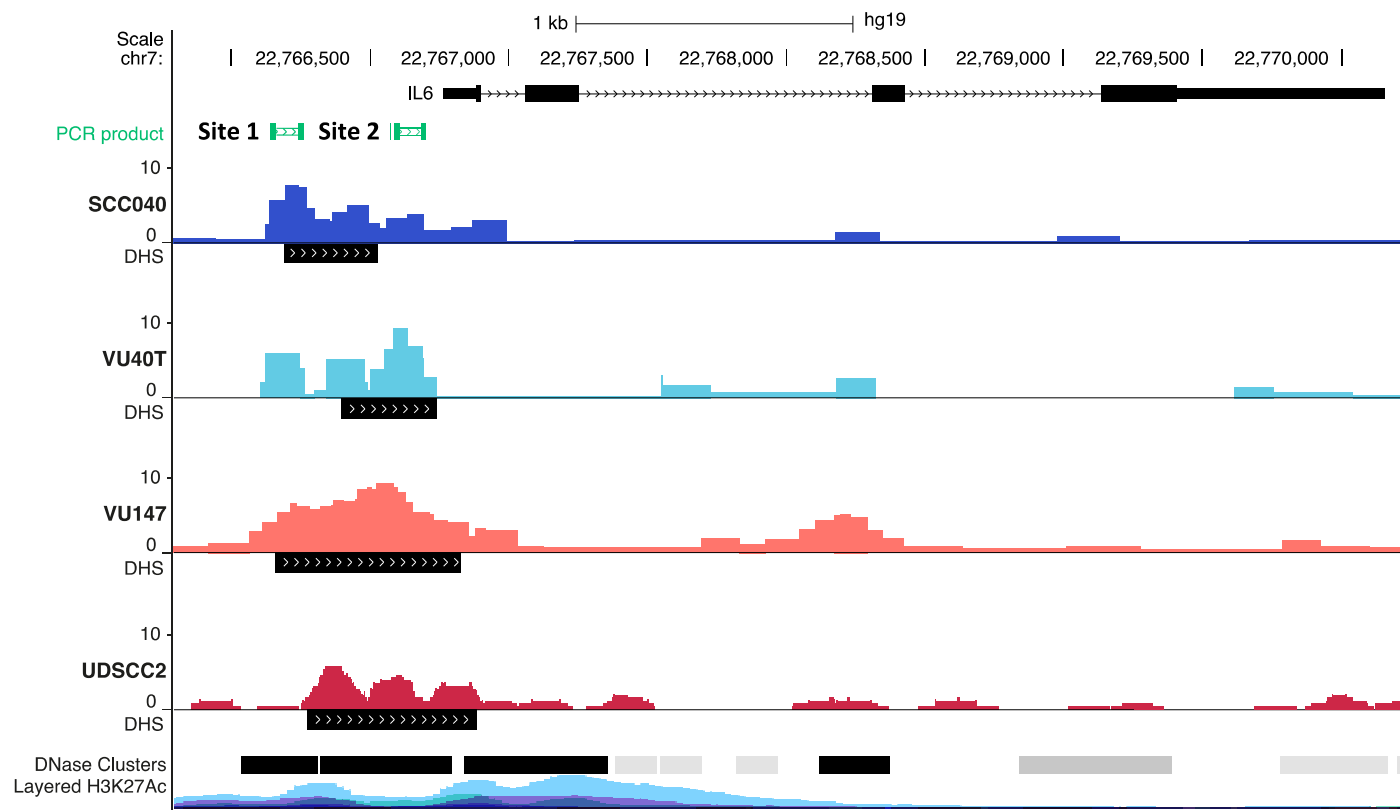
Figure 3.6. Primer locations for quality control of DNase I-digested samples near the *IL-6* gene.

The genomic features and DHS data are presented as in Figure 3.5. This figure displays the *IL-6* DHS from subsequent DNase I-seq experiments conducted in HNSCC cell lines. Positive control primers for qPCR were designed at two sites close to the *IL-6* gene (indicated by green bars, 'site 1', 'site 2').

3.7 Assessment of DHS enrichment by qPCR

All DNase I experiments yielded enough cells for two DNase I treatments. Therefore, for comparison, along with qPCR data from treatments with 50U of DNase I in all HNSCC cell lines, 30U DNase I in UDSCC2(+) cells and 40U DNase I in SCC040(-), VU40T(-) and VU147(+) cells are also shown **(Figure 3.7)**.

In SCC040(-), VU40T(-) and VU147(+) cells, DHS fold enrichment was assessed by comparing the amplification signals of *IL-6* sites 1 and 2 of 40U and 50U DHS samples, to that of the 'no DHS' region **(Figure 3.7)**. SCC040(-) DHSs amplified at *IL-6* site 1, displayed fold enrichment of 6.8- and 9.5-fold, for 40U and 50U preparations, respectively, while amplification of the *IL-6* site 2 revealed 8.3- and 17-fold enrichment, respectively **(Figure 3.7A)**. VU40T(-) DHSs amplified at *IL-6* site 1 displayed fold enrichment of 5.9-fold and 17.0-fold in 40U and 50U preparations, respectively, while amplification at the *IL-6* site 2 displayed 8.3- and 28.0-fold enrichment, respectively **(Figure 3.7B)**. VU147(+) DHSs amplified at *IL-6* site 1 displayed fold enrichment of 13.1-fold and 48.6-fold in 40U and 50U preparations, respectively, while amplification of *IL-6* site 2 displayed 15.5-fold and 74.8-fold enrichment, respectively **(Figure 3.7C)**. UDSCC2(+) DHS enrichment was assessed using the *KRT6A* DHS region. The DHSs were isolated with either 30U or 50U of DNase I **(Figure 3.7D)**. DHSs isolated with 30U of DNase I displayed fold enrichment of 28.8-fold, while 50U DHSs displayed 42.1-fold enrichment. Therefore, in all experiments, DHSs derived from 50U DNase I treatment displayed enrichment over background above the 5-fold cut-off and so were taken forward for library synthesis.
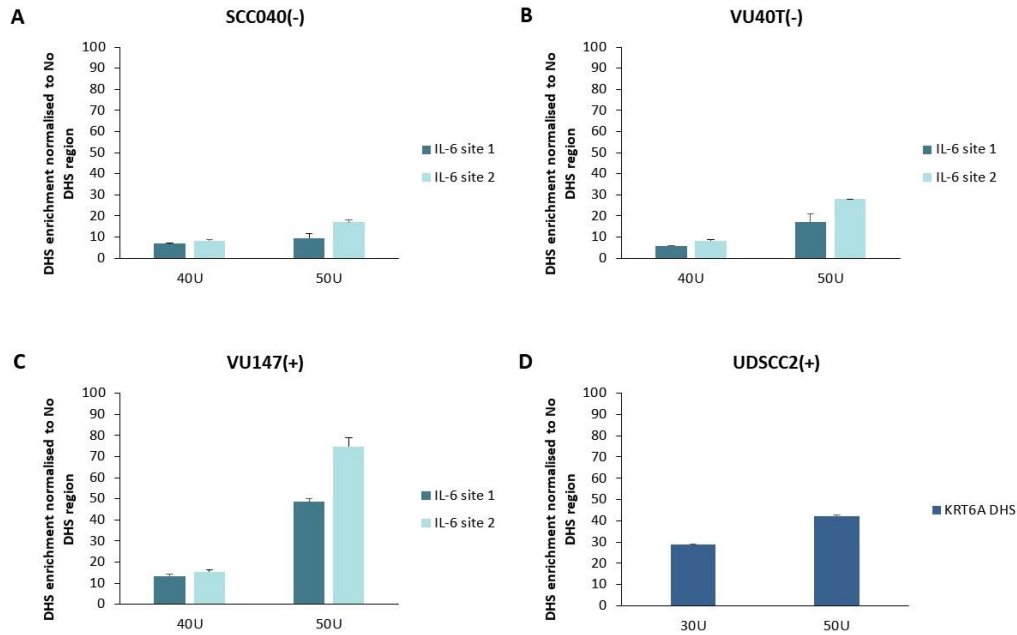
Figure 3.7. Size-selected DHS enrichment assessed by qPCR.
Nuclei were treated with 30U, 40U or 50U of DNase I and qPCR was conducted of sites 1 and 2 near the *IL-6* gene in (**A**) SCC040(-), (**B**) VU40T(-), (**C**) VU147(+) and of the *KRT6A* DHS region in (**D**) UDSCC2(+) DHSs. The signal was normalised to a negative control 'no DHS' region. The bars represent the range between technical duplicates.

3.8 DHS library preparation and quality control

After verification of DHS enrichment by qPCR, libraries were prepared, as described in Materials and Methods, Section 2.9. Library preparation involves an adapter ligation step, which modifies the DNA making it compatible with the sequencing platform. The adapter also contains a unique index sequence for each sample, allowing identification of the samples during analysis after pooling and sequencing. Once the DHS library was prepared, the DHS sequences (insert) were size-selected for ~270bp (~150bp DHS insert size + 120bp adapter size). PCR was conducted to amplify the library and amplification was determined by qPCR. The same primer sets were used as described in **Figures 3.5 and 3.6** (*KRT6A*, *IL-6* sites 1 and 2 and the 'no DHS' region) (**Figure 3.1: QC4**).

Amplification of the SCC040(-) DHSs at the *IL-6* sites 1 and 2 resulted in 134.8- and 434.8-fold enrichment, respectively **(Figure 3.8A)**. In the VU40T(-) library amplification of *IL-6* sites 1 and 2 displayed 281.3- and 824.6-fold enrichment, respectively **(Figure 3.8B)**. Amplification of the VU147(+) library displayed fold enrichment of 547.7- and 626.2-fold at *IL-6* sites 1 and 2, respectively **(Figure 3.8C)**. Amplification of the *KRT6A* region of the UDSCC2(+) library displayed a 15.8-fold enrichment **(Figure 3.8D)**. The low enrichment in the UDSCC2(+) library qPCR signal may be due to the size-selection step during library preparation, where DHSs out of the 100-200bp range are filtered out. Despite this, the signal after library preparation was above the 5-fold enrichment cut-off value.
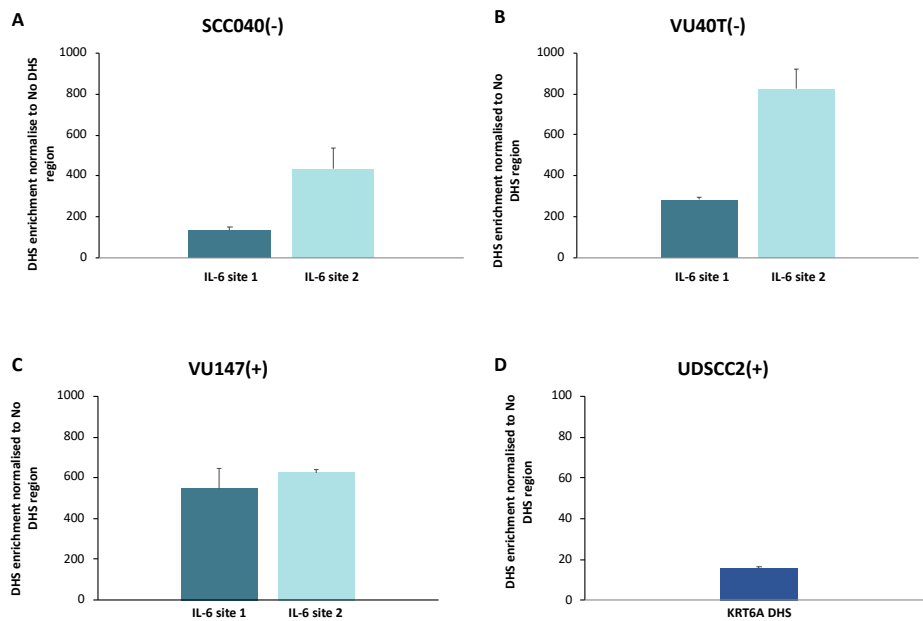
Figure 3.8. Assessment of DHS enrichment after library preparation by qPCR.
DHS enrichment of libraries was determined by amplification of DHSs near the *IL-6* gene (sites 1 and 2) in **(A)** SCC040(-), **(B)** VU40T(-), **(C)** VU147(+) or at the *KRT6A* gene in **(D)** UDSCC2(+). The qPCR signal obtained from DHSs was normalised using the 'no DHS' signal to calculate enrichment over background. The bars represent the range between technical duplicates.

After qPCR confirmation of library enrichment, library size was verified by using the Agilent Bioanalyzer system **(Figure 3.1: QC5 and Figure 3.9)**. Verification of the library size was required, as the desirable total size for Illumina HiSeq sequencing is 250-300bp.

The average library size for the SCC040(-) DHS library was 284bp **(Figure 3.9)**. A smaller peak can be observed with an average size of ~80bp, which represents primer dimers. However, the amount of DNA within this peak relative to the library DNA was not sufficient to significantly deplete reads from the library. Therefore, the SCC040(-) library was carried forward to the next step. The average library sizes for the VU40T(-), VU147(+) and UDSCC2(+) libraries were 275bp, 271bp and 461bp, respectively. The average library size for UDSCC2(+) DHSs was higher than the desired size range. However, the amount of DNA within the desired range was sufficient for sequencing and so the library was carried forward to the next step, as were the remaining libraries.

After verification of library size, the libraries were quantified using the qPCR-based Kapa quantification kit, as described in Materials and Methods, Section 2.11.3 **(Figure 3.1: QC6)**. After quantification, eight indexed libraries were pooled for sequencing; the amount of total library was 10nM and was verified by Kapa quantification. Quantification was required to achieve the correct cluster density on the sequencing platform. A cluster is formed when the DNA fragments undergo clonal amplification. Loading too much library could lead to over-clustering and poor accuracy in determining each read, while under-clustering could lead to too few reads being sequenced, resulting in poor resolution coverage maps.
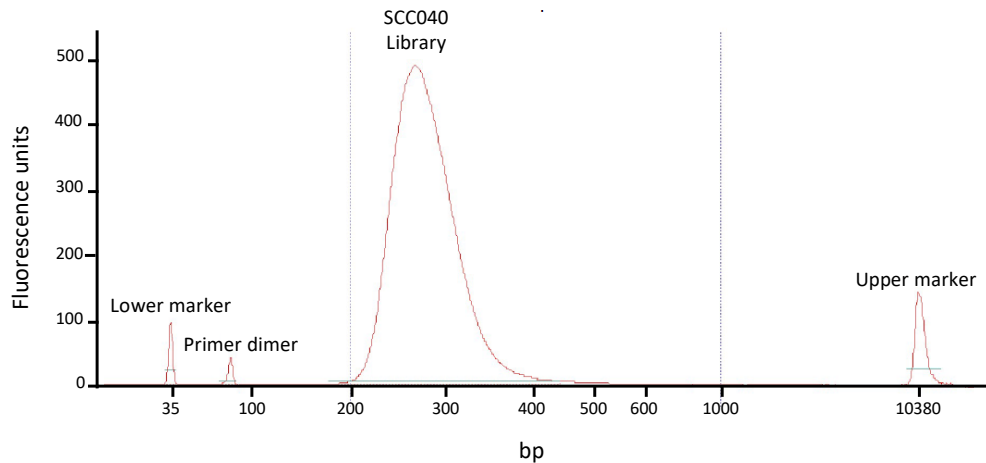
Figure 3.9. A screenshot of the electropherogram for the SCC040(-) library using the Agilent Bioanalyzer system.
The DNA marker can be observed on the electropherogram as two peaks; a lower marker at 35bp and an upper marker at 10,380bp. To determine the average library size, gates were set surrounding the library peak. The average size of the fragments in the SCC040(-) library was 284bp (indicated by the peak labelled 'SCC040 library'). A peak is visible at ~80bp, indicative of the presence of a small amount of primer dimers.

3.9 Mapping of DHSs in HNSCC cell lines

Between 20 and 40 million reads were uniquely mapped to the hg19 version of the human genome for each analysed cell line ('reads aligned' in **Table 3.1**). An example of the sequenced DHS data can be seen in **Figure 3.10** at the *ANXA1* (annexin 1) gene. The ANXA1 protein is often deregulated in cancers including HNSCC and is implicated in the HNSCC-related pathway, PI3K/AKT/mTOR [236, 237].

Peaks were identified (called) as DHS peaks using the HOMER (Hypergeometric Optimization of Motif EnRichment) software [220]. The input sample for the corresponding cell lines was used to improve peak calling specificity and reduce background. The peaks were called using the 'findpeaks' command and the 'variable peak' setting. The settings for peak width and minimum distance between peaks were based on default settings in HOMER and were adjusted by observing

changes in DHS-rich regions considered to be transcriptionally active in HNSCC such as, the keratin genes, *EGFR*, *TP63* and *ANXA* genes [4, 6, 10, 12, 36]. The peak width 'size' setting was set at 500 and the 'minimum distance' setting between peaks was 1000 (2x the peak size). The false discovery rate (FDR) was set at 0.01% for SCC040(-), VU40T(-) and VU147(+) cells and 0.05% for UDSCC2(+) cells. The 'false discovery rate' (FDR) setting allows control over false positive peak predictions. The FDR value negatively correlates with peak confidence and stringency, therefore, the lower the number the more stringently peaks will be called. In previously published data generated by the ENCODE consortium and others using DNase I-seq, the FDR was set between 0.1% to 1% [161, 220, 238]. Once the reads were mapped and the peaks were aligned to the genome for each of the four HNSCC cell lines (single cell line dataset), between 22,398 and 38,851 peaks were identified in the cell lines (**Table 3.1**).

Table 3.1. Number of mapped reads in four HNSCC cell lines.

Cell line	Number of reads	Number of reads aligned	Percentage of reads aligned	Number of peaks called
SCC040(-)	29,539,814	29,222,932	99.0%	30,137
VU147(+)	43,973,943	43,611,816	99.2%	38,815
VU40T(-)	21,154,428	20,812,496	98.4%	30,851
UDSCC2(+)	27,517,141	22,182,041	80.6%	22,398

The distribution of DHSs across the genome was investigated to determine whether it agreed with previously published data. The called peaks were annotated by genomic feature using the HOMER 'AnnotatePeaks' function. The peaks were annotated into the following categories: intergenic, intronic, promoter, exon, transcription termination site (TTS) and untranslated regions (UTRs) (**Figure 3.11A-D**). Agreeing with previously published data, the analysis revealed that the

majority of DHS peaks were located within intergenic and intronic regions; between 28.3% and 39.2%, respectively [161, 188, 238, 239]. Furthermore, between 10.5% and 18.5% of DHSs in promoter regions and between 8.7% and 15.1% in exons were observed [142-144, 239].

Using the HOMER 'mergepeaks' function, a 'merged-HNSCC' dataset (73,414 peaks) was generated in which peaks present in at least one cell line were considered. This enabled investigation of whether there was any overlap amongst DHSs within a heterogeneous HNSCC cell population of four cell lines with different HPV status and origin.
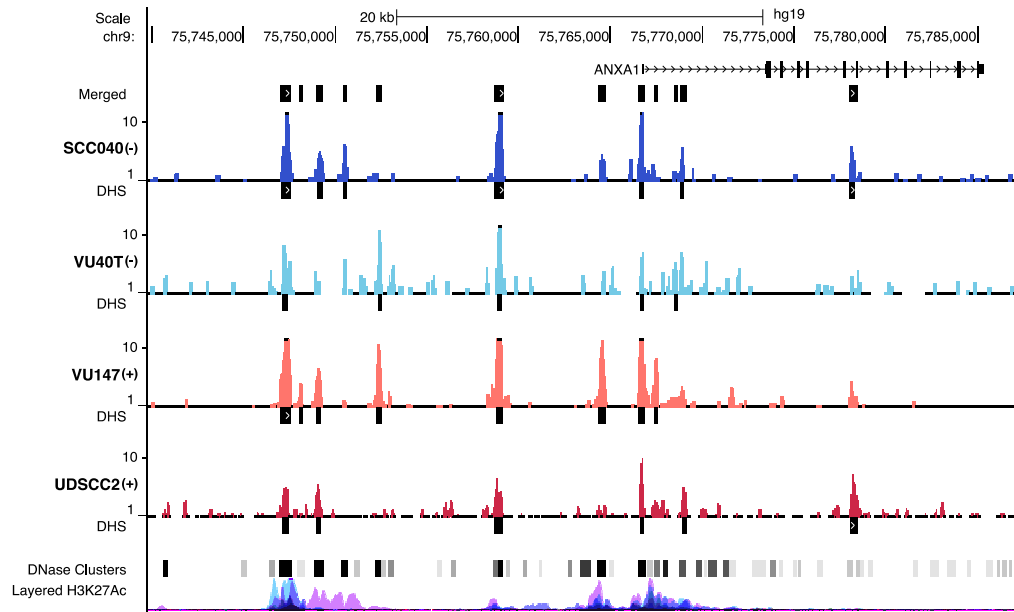
Figure 3.10. UCSC screenshot of DNase I-seq data at the *ANXA1* gene.
DHS reads were mapped to the hg19 genome and peaks were called using HOMER.
Coverage tracks for each cell line are displayed; SCC040(-) (dark blue), VU40T(-) (light
blue), VU147(+) (orange) and UDSCC2(+) (red). Called peaks for each cell line are
depicted below the coverage tracks (depicted by the black bars, labelled 'DHS'). The
'Merged' peak calling track corresponds to the merged DHS dataset, which was produced
by including DHS peaks present in at least one of the HNSCC cell lines. The 'DNase
Clusters' track displays previously published data; mapped DHS reads in a panel of 125
cell lines by the ENCODE consortium. The 'layered H3K27Ac' track displays previously
published data; the active histone mark, H3K27Ac in ENCODE cell lines. The pink peaks
in the H3K27Ac track represent the NHEK (normal human epidermal keratinocyte) cell
line, indicating that some of these DHSs are epithelial specific.

126
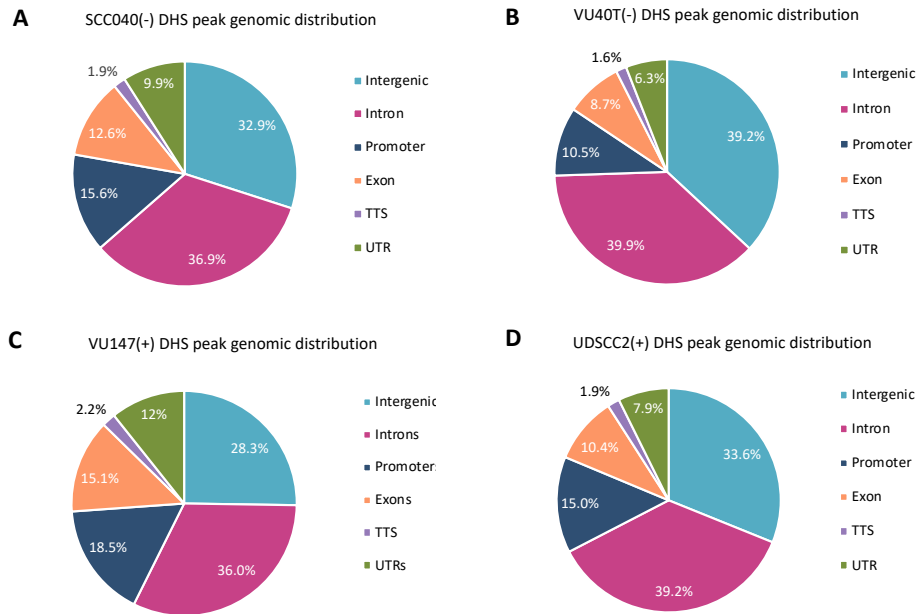
Figure 3.11. Genomic distribution of DHS peaks.
A total of **(A)** 30,137 DHS peaks were called in SCC040(-) cells, **(B)** 30,851 DHSs in VU40T(-) cells, **(C)** 38,815 DHSs in VU147(+) cells and **(D)** 22,398 DHS peaks in UDSCC2(+) cells. The locations of the DHSs were mapped using HOMER to the following genomic regions: intergenic regions, introns, promoters, exons, transcription termination site (TTS) and untranslated regions (UTR).

In total, 6.7% of DHS peaks in the 'merged-HNSCC' dataset were shared amongst all four cell lines and 13.1% were found in any three cell lines **(Figure 3.12A and B)**. Therefore, 24.4% were found in at least three of the investigated cell lines and are referred to as 'common peaks' **(Figure 3.12B and C)**. As much as 62.5% of 'merged-HNSCC' peaks were uniquely found in only one cell line. Next, the proportion of unique and common peaks for each cell line were investigated independently **(Figure 3.12C)**. Between 26% and 55% were unique in any given cell line and between 35% and 46% were common with at least two other cell lines. The highest proportion of unique peaks was observed for UDSCC2(+) cell line **(Figure 3.12C)**.

The high proportion of unique peaks observed in the 'merged-HNSCC' dataset, indicates high heterogeneity between the cell lines. However, it was observed that unique peaks were smaller than common peaks, which may reflect the stability of the DHS [188]. Therefore, the top 10% of peaks (7,340 peaks selected based on the peak score from the annotation file) were investigated. In this data set, 32.5% of peaks were shared between all cell lines, 65.8% of peaks were common (at least 3 out of 4) and only 19.2% were unique **(Figure 3.13A** and **3.13B)**. When each cell line was considered independently, there were more than 80% common peaks and 5-13% unique peaks **(Figure 3.13C)** as compared to the 'merged-HNSCC' dataset (35-46% and 26-55%, respectively) **(Figure 3.13C)**. This indicated that the most prominent regulatory network was shared between the HNSCC cell lines.
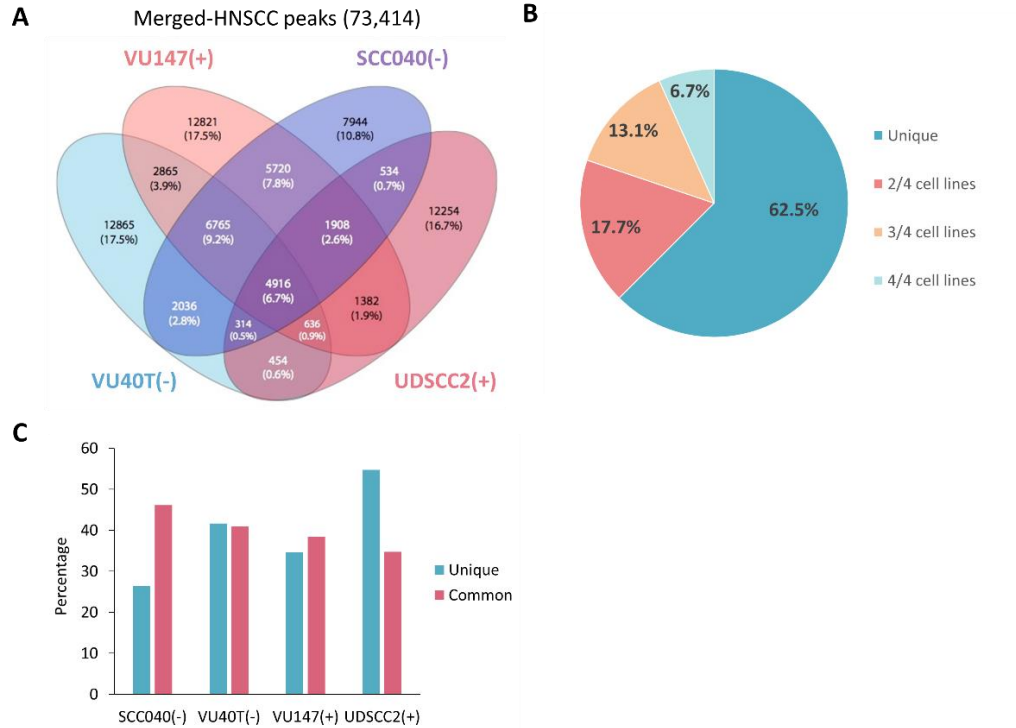
Figure 3.12. Cell line distribution of merged-HNSCC DHS peaks.
(A) A total of 73,414 DHS peaks were merged by considering any peaks which were present in at least one HNSCC cell line (SCC040(-), VU40T(-), VU147(+) and UDSCC2(+). **(B)** The DHS peaks were assigned to one of four categories: Unique to one cell line or present in 2/4, 3/4 or 4/4 cell lines. **(C)** The percentage of unique or common (present in at least 3/4 cell lines) was calculated for each cell line; 26.4%, 41.7%, 34.6% and 54.7% of DHSs were unique to SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively and 46.1%, 40.9%, 38.4% and 34.7% of DHSs were common in SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively.
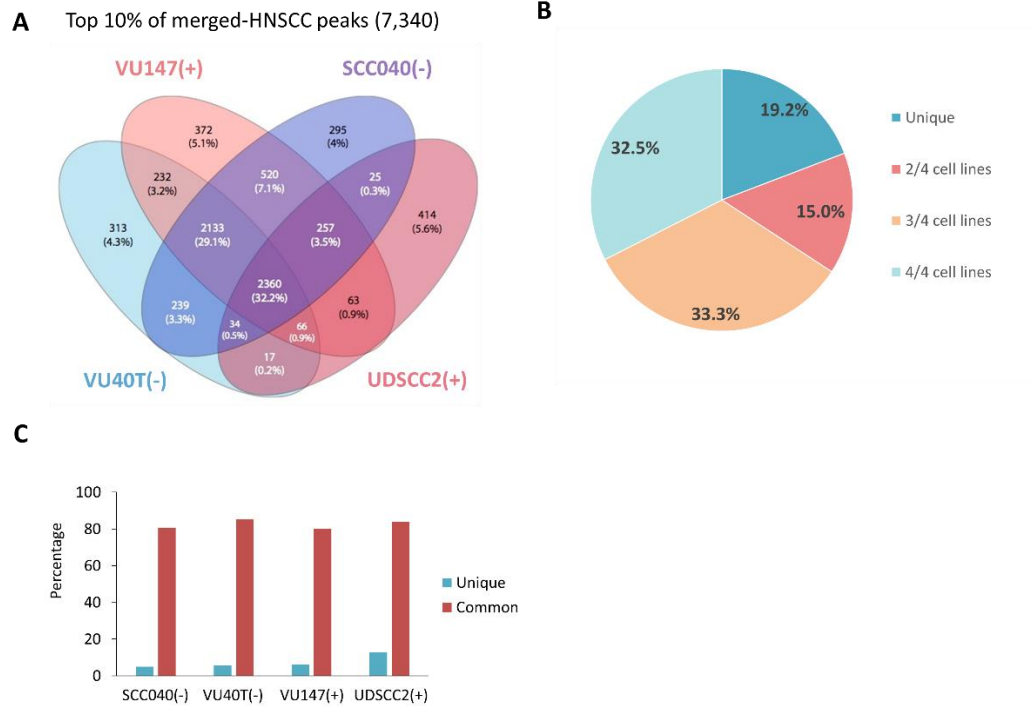
Figure 3.13. The cell line distribution of the top 10% of merged-HNSCC DHS peaks.
(A) 7,340 DHS peaks with highest peak scores were selected from the 'merged-HNSCC' as top 10% of merged DHS peaks **(B)** The DHS peaks were assigned to one of four categories: Unique to one cell line, present in any 2/4, 3/4 or 4/4 cell lines. **(C)** The percentage of unique or common (present in at least three out of four cell lines) was calculated for each cell line; 5.0%, 5.8%, 6.3% and 12.8% of DHSs were unique to SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively and 80.8%, 85.2%, 80.0% and 84.0% of DHSs were common in SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively.

3.10 Transcription factor motif analysis of DHS peaks

To determine which transcription factors (TFs) may be involved in HNSCC gene regulation, TF motif analysis was first performed for each HNSCC cell line using HOMER with the 'known motif' function. Motif analysis provides data on which sequences (motifs) the TFs preferentially bind to and predict motifs based on sequence similarity [220].

The number of TF motifs identified in DHSs of each cell line were: 177 TF motifs in VU40T(-), 177 in SCC040(-), 198 in VU147(+) and 180 in UDSCC2(+) cells. The analysis of the single cell line DHS dataset revealed a total of 217 different motifs across the four cell lines and over half (60.8%) of these were shared between all four. The top 20 motifs for each cell line are shown in **Appendix Tables 1-4**, which show high overlap between the lists. Due to similar TF motifs being identified between the cell lines, the top 20 TF motifs common between all four cell lines were taken for analysis **(Table 3.2)**. This list included the AP-1 (activator protein-1) motif (ATF3, BACH (BTB domain and CNC homolog)-1, -2, FOSL-1, -2, JUN, MAFK (MAF bZIP transcription factor K), NRF2/NF-E2 (nuclear factor erythroid 2–related factor 2)). The AP-1 TF functions as a dimer comprised of members from three TF families, the JUN family (c-JUN, JUNB, JUND), the FOS family (c-FOS, FOSB, FOSL1 (FOS like 1, AP-1 transcription factor subunit) (FRA-1), FOSL2 (FRA-2)) and the ATF (activating transcription factor) family (ATF1-3) [240]. Motifs for members of all three AP-1 families were identified in all HNSCC cell lines and the merged TF table **(Table 3.2** and **Appendix Table 1-4)**. Further motifs related to AP-1 were identified, including the BACH TF family, which were also observed in all HNSCC cell lines **(Table 3.2** and **Appendix Table**

1-4). In addition, the NRF2/NF-E2 and MAFK motifs were also present in the merged TF table **(Table 3.2)**. Other top 20 TF motifs included p53/p63 motifs, TEAD (TEA domain transcription factor), RUNX (runt-related transcription factor), KLF (Krüppel-like factor) family motifs, CTCF (CCCTC-binding factor) and BORIS (brother of regulator of imprinted sites, a paralogue TF of CTCF) TF motifs **(Table 3.2)**.

In addition to motifs common between the cell lines, cell line-specific TF motifs were also observed. In UDSCC2(+) DHSs, the LHX (LIM homeobox)-2, -3, ELK (ETS domain-containing protein)-4 and NF (nuclear factor)-1 TF motifs were observed. In VU40T(-) DHSs, the ERG (ETS-related gene) was observed **(Appendix: Tables 2 and 4)**.

The lists of the top 20 TF motifs, were compared by HPV status, which revealed the SP1 (specificity protein 1) motif to be enriched in only the HPV(+) cell lines, while the RUNX motif was observed only in the HPV(-) cell lines **(Appendix: Tables 2 and 4)**.

Table 3.2. Common top 20 statistically significant transcription factor motifs in four HNSCC cell lines.

Merged TF Motifs

Name	Motif logo
FOSL1 (Fra-1)	
ATF3	
BATF	
AP-1	
FOSL2 (Fra-2)	
JUN	
BACH1	
BACH2	
NF-E2	
NRF2	
p63	
p53	
TEAD4	
TEAD	
RUNX	
RUNX1	
BORIS	
CTCF	
KLF5	
MAFK	

Motif analysis was conducted using the 'findMotifsGenome' script within HOMER. Six TF motif families were identified: AP-1 (TGAnTCA) (dark blue), p53/p63/p73 (red), TEAD (cyan), RUNX (green), BORIS/CTCF (light blue) and KLF (white).

Due to the high sequence similarity between motifs, there can be indistinguishable overlap with other motifs, therefore 'de novo motif' analysis was also conducted. A total of 21, 23, 22 and 24 motifs were predicted in VU40T(-), SCC040(-), VU147(+) and UDSCC2(+), respectively. This analysis involved matching a target sequence to a known motif, while ranking other motifs by similarity to generate a list of *de novo* motifs. The top 10 motifs are shown for each cell line in **Figure 3.14**. It was found that all motifs identified earlier through the 'known motif' analysis (AP-1, p53/p63, RUNX, TEAD, CTCF, KLF and NRF motifs) are included **(Table 3.2)**. In addition, the *de novo* analysis identified the ETS (E twenty-six transcription factor), NFY (nuclear transcription factor Y), CRE (cAMP response element), GFY (general factor Y), YY1 (yin yang 1) and NF (nuclear factor)-1 motifs **(Figure 3.14)**.

The *de novo* motif analysis was also conducted to identify motifs which overlap with promoter DHSs and those which overlap with distal DHSs (>1kb from the promoter) **(Figure 3.15)**. Analysis of the two subsets (promoter and distal) revealed three groups of motifs: unique to promoter (ETS, NFY, CRE, NRF (nuclear respiratory factor)-1, GFY, YY1) unique to distal (FOXx/SRY, NF1, TEAD, p53/p63, RUNX, CTCF), and present in both (AP-1 and SP1/KLF) **(Table 3.3)**. While the SP1 and KLF motifs are separate TFs, they share a highly conserved DNA binding domain and are able to function in association with each other or separately [241]. The AP-1 motif is present in both elements, at distal DHSs AP-1 displays higher significance and appears in 3-times more targets compared to in promoters **(Figure 3.15)**.

134

Table 3.3. Motifs identified in promoter, distal DHSs or both.

Promoter only	Distal only	Both
ETS	FOXx/SRY	AP-1
NFY	NF1	SP1/KLF
CRE	TEAD	
NRF1	p53/p63	
GFY	RUNX	
YY1	CTCF	
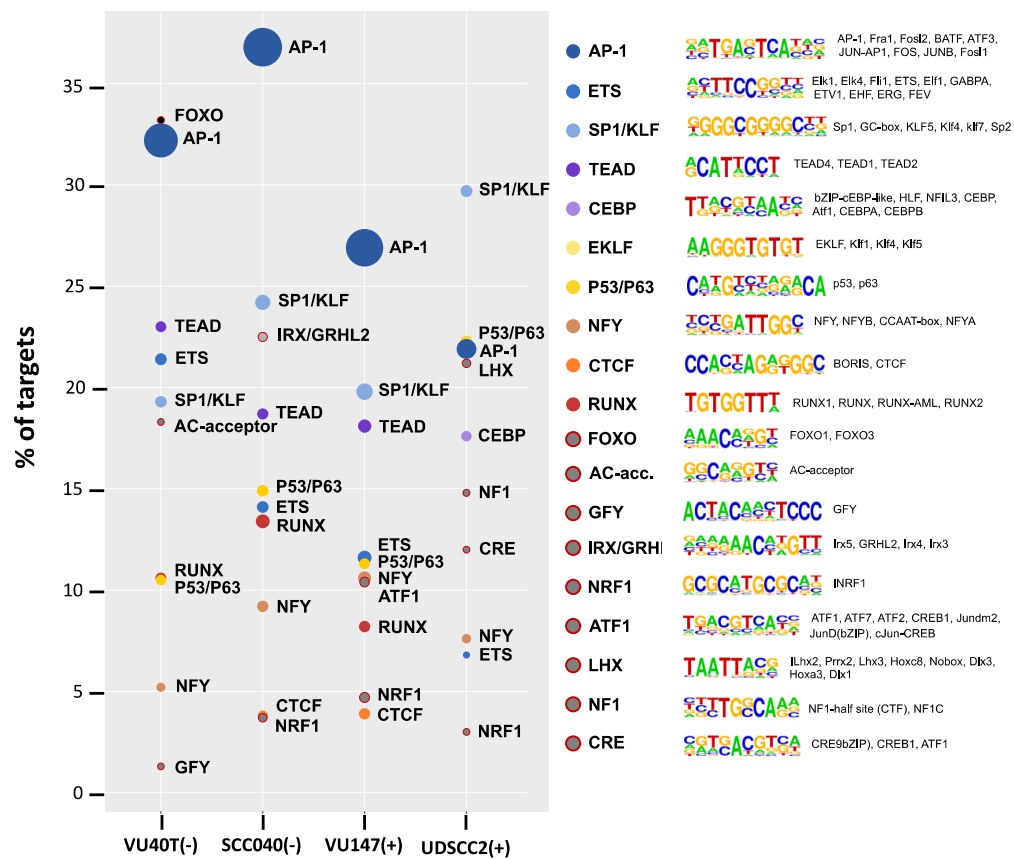
Figure 3.14. *De novo* **motif analysis of HNSCC cells.**
HOMER was used to predict *de novo* motifs using VU40T(-), SCC040(-), VU147(+) and UDSCC2(+) DHS peaks; a total of 21, 23, 22 and 24 motifs were identified, respectively. For each motif, the percentage of motifs in target sequences is displayed and the size of the circles represent p-value.
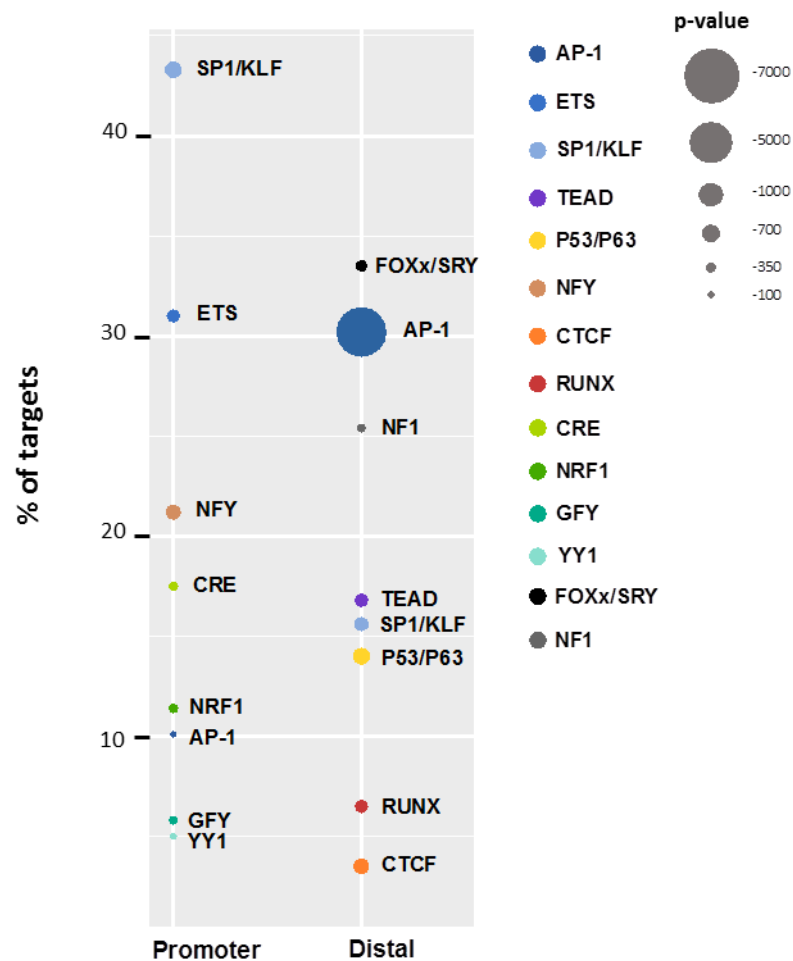
Figure 3.15. *De novo* **motifs identified in the merged-HNSCC DHS dataset in promoters or distal locations.**

De novo motif analysis was conducted in the merged-HNSCC DHS dataset using the 'findMotifsGenome' script within HOMER. The analysis was conducted in DHSs which were in proximity to the promoter TSS (<1kb) and in DHSs distal to the TSS (>1kb). The motifs are ranked by percentage of target sequences (DHSs) that contained them. The size of the circles represents p-value.

3.11 Pathway enrichment analysis of HNSCC DHSs.

To investigate which molecular processes the HNSCC DHSs may be involved in, pathway enrichment analysis was conducted in HOMER using the 'AnnotatePeaks' script. This analysis assigns each DHS to the closest TSS of a gene to create a list of potentially regulated genes. The list of genes was run through a combination of two pathway databases: Kyoto Encyclopaedia of Genes and Genomes (KEGG) and WikiPathways (WP) [223, 224]. The KEGG database highlights molecular processes and the WP database identifies cell signalling pathways. Between 283 and 608 signalling pathways were identified in the four HNSCC cell lines and between 10,709 and 13,705 genes were annotated in the pathways **(Table 3.4)**.

The top 20 pathways were investigated for each cell line **(Tables 3.5-3.8)**. The total number of annotated genes for the 20 pathways yielded between 1,882 and 2,654 genes **(Table 3.4)**. The most commonly annotated genes encoding TFs identified in motif analysis were the AP-1 genes; between 45-55% of pathways were annotated with an AP-1-related gene **(Tables 3.5-3.8)**. Other TFs highlighted in the merged TF motif analysis were less frequently annotated in the pathway enrichment analysis. For example, the *BACH1* gene was annotated only in the integrated breast cancer pathway in VU147(+) and UDSCC2(+) DHSs **(Tables 3.7 and 3.8)**. Of the *KLF* family, the *KLF11* gene was annotated in the VU40T(-) EGFR (epidermal growth factor receptor) signalling pathway, while *KLF2* was annotated in the Fox (Forkhead)-O signalling pathway **(Table 3.6)**. The *RUNX1* gene was annotated in the 'pathways in cancer' pathway in all four cell lines, while the *RUNX2* gene was annotated in the androgen receptor signalling pathway in UDSCC2(+) cells **(Table**

3.8). Genes of the *TEAD* TF family were annotated in the Hippo signalling pathway

in SCC040(-) and VU147(+) cells **(Tables 3.5 and 3.7)**.

Table 3.4. Summary of pathway and gene annotation.

Cell line	Number of pathways	Number of annotated genes	Number of annotated genes in top 20 pathways
SCC040(-)	288	12,794	2,229
VU40T(-)	283	13,705	2,580
VU147(+)	286	13,639	2,654
UDSCC2(+)	608	10,709	1,882

Table 3.5. The top 20 DHS pathways identified in SCC040(-) cells.

SCC040(-) DHS: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Pathways in cancer	7.03E-10	-21.08	327	257	75.6	AKT2-, -3, CCND1, CDK4, CDKN1A, CDKN2A, EGF, EGFR, FOS, IL6, JUN, MAPK-1, RB1, RUNX1, TGF-β3, WNT-2B, -3A, -4, -5B, -6, -7A, -7B, -8, -9A, -10, 16	KEGG
RB in Cancer	7.29E-08	-16.43	87	79	90.8	CCND1, CDKN1A, E2F1, RB1	WP
Integrated Breast Cancer Pathway	3.62E-07	-14.83	152	128	84.2	CCND1, EGFR, JUN, RB1	WP
Cell cycle	9.16E-07	-13.90	124	103	83.1	CCND1, CDKN1A, CDKN2A, E2F1, RB1	KEGG
AMPK signalling pathway	3.42E-06	-12.59	123	101	82.1	CCND1, MTOR	KEGG
EGFR1 Signalling Pathway	4.32E-06	-12.35	177	144	81.4	CTNND1, EGF, EGFR, FOS, JUN, JUND, MAPK14	WP
Endocytosis	4.39E-06	-12.34	203	158	77.8	EGFR, ERBB-3,-4	KEGG
ErbB signalling pathway	5.56E-06	-12.10	87	74	85.1	AREG, CDKN1A, EGF, EGFR, ERBB-2-4, EREG, JUN, MAPK-1, -8, -10, MTOR	KEGG
Proteoglycans in cancer	6.13E-06	-12.00	225	173	76.9	AKT-2, -3, CCND1, CDKN1A, EGFR, ERBB2, ERBB3, ERBB4, MAPK1, -11, -13, -14, WNT-2B, -3, -3A, -4, -5B, 7A, -7B, -8B, -9A, -10A, -16	KEGG
Viral carcinogenesis	1.99E-05	-10.83	206	158	76.7	ATF4, CCND1, CDKN1A, CDKN2A, JUN, MAPK1, RB1, YWHAE	KEGG
Spliceosome	2.38E-05	-10.65	130	104	80.0	-	KEGG
Insulin Signalling	3.86E-05	-10.16	163	131	80.04	FOS, JUN, MAPK-1, -4, -6 ,-8 ,-10, -13 , -14, MTOR	WP
DNA damage response (only ATM dependent)	5.98E-05	-9.72	91	77	84.6	CCND-1, -2, CDKN2A, ERBB2, FOSL1, JUN, MAPK-1, -8, -10, TP73, WNT-2B, -3, -3A, -4, -5B, -6, -7A, -7B, -10A, -11, -16	WP
α6-β4 Integrin Signalling Pathway	7.55E-05	-9.49	69	60	87.0	CDKN1A, EGFR, ERBB2, MTOR, TP73, YES1, YWHAE	WP
Prostate Cancer	9.93E-05	-9.22	109	90	82.6	ATF1, CCND1, E2F1, FOS, JUN, MAPK-8, -14, MTOR, RB1	WP
Adherens junction	1.14E-04	-9.08	73	61	83.6	EGFR, ERBB2, MAPK1, YES1	KEGG
p53 signalling pathway	1.61E-04	-8.73	68	57	83.8	CCND1, CDKN1A, CDKN2A, TP73	KEGG
Ribosome	1.72E-04	-8.67	132	103	78.0	-	KEGG
Small cell lung cancer	1.91E-04	-8.56	86	70	81.4	CCND1	KEGG
TGF-β Receptor Signalling Pathway	2.07E-04	-8.48	151	120	79.5	CCND1, CDKN1A, FOSB, JUN, JUNB, JUND, MAPK8, RB1, RUNX2, TP73	WP

Table 3.6. The top 20 DHS pathways identified in VU40T(-) cells.

VU40T(-) DHS: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Pathways in cancer	2.76E-07	-15.10	327	267	81.7	BIRC2,CDK6, CCND1, CDKN1A, CDKN2A, EGF, EGFR, ERBB2, FOS, IL-6, JUN, MAPK-1, -3, -8, -10, PIK3CA, RB1, RUNX1, TGF-β2, -β2, WNT-2, -2B, -3, -3A, -4, -5A,-6, -7A, -7B, -9A, -10A, -11, -16	KEGG
Endocytosis	3.43E-07	-14.88	203	172	84.7	EGF, EGFR, ERBB3, TGF-β2	KEGG
EGFR1 Signalling Pathway	3.58E-07	-14.84	177	155	87.6	EGF, EGFR, FOS, JUN, JUND, KLF11, MAPK-1,-3,-8,-14, PIK3CA	WP
Focal adhesion	2.13E-06	-13.06	207	173	83.6	EGF, EGFR, ERBB2, MAPK-1,-3,-8,-10, PIK3CA	KEGG
FoxO signalling pathway	4.92E-06	-12.22	132	114	86.4	CCND1, CDKN1A, MAPK-1,-9, -12, -14, PIK3CA, TGFB1	KEGG
Proteoglycans in cancer	8.35E-06	-11.69	225	185	82.2	CCND1, CDKN1A, EGFR, ERBB-2, -3, MAPK-1, -3, 11-14, PIK3CA, WNT-3, -3A, -4, -5A, -6, -7B, -9A, -10A, -16	KEGG
Adherens junction	1.91E-05	-10.87	73	66	90.4	CTNND1, EGFR, ERBB2,, MAPK-1, -3, YES1	KEGG
TGF-β Receptor Signalling Pathway	2.41E-05	-10.63	151	130	86.1	CCND1, CDKN1A, FOS, FOSB, JUN,JUNB, JUND, MAPK14, TGF-β2, YAP1A	WP
ErbB signalling pathway	2.58E-05	-10.57	87	77	88.5	AKT3, AREG, CDKN1A, EGFR, ERBB2, EREG, JUN, MAPK-1, -3, -8, -10, MTOR, PIK3CA, RAF1	KEGG
Axon guidance	3.95E-05	-10.14	127	108	85.0	CDK51, MAPK-1, -3	KEGG
DNA damage response (only ATM dependent)	6.76E-05	-9.60	91	81	89.0	AKT-2, -3, CCND1, CDKN1A, DVL1, FOSL1, JUN, MAPK1, MAPK10, MAPK8, MAPK9, PTEN, RBB2, TP73, WNT-2, -2B, -3, -3A, -5A, -6, -7A, -7B, -10A, -11, -16	WP
T cell receptor signalling pathway	1.17E-04	-9.06	104	89	85.6	AKT2, CD247, CDC42, CDK4, FOS, JUN, MAPK1, MAPK11, MAPK12, MAPK13, MAPK14, MAPK3, MAPK9, PIK3CA	KEGG
Wnt Signalling Pathway	1.61E-04	-8.73	60	55	91.7	CCND1, FOSL1, JUN, MAPK-9, 10, WNT-2, -2B, -3, -3A, -4, -5A, -5B, -6, -7A, -7B, -10A, -11, -16	WP
Integrated Pancreatic Cancer Pathway	1.97E-04	-8.53	194	161	83.0	CCND1, CDKN1A, CDKN2B, EGF, EGFR, ERBB2, JUN, JUNB, JUND, MAPK-1, -3, -4, -8, -9, -14, PIK3CA	WP
Hippo signalling pathway	2.03E-04	-8.50	153	126	82.4	AJUBA, AREG, BIRC2, CCND1, LATS2, MOB1A, MOB1B, NF2, TEAD1-4, TGF-β2, -β2, TP73, WNT-2, 2B, -3, -3A, -4, -5A, -5B, -6, -7A, -9A, -10A, YAP1, YWHAE	KEGG
Senescence and Autophagy	2.04E-04	-8.50	106	92	86.8	CDKN1A, JUN, MAPK-1, -14, MTOR	WP
Regulation of actin cytoskeleton	2.06E-04	-8.49	214	172	80.4	EGF, EGFR, MAPK3	KEGG
Prolactin Signalling Pathway	2.29E-04	-8.38	75	67	89.3	ERBB2, FOSB, JUN, MAPK-3, -6, -8, -9, -14, MTOR, PIK3CA	WP
Corticotropin-releasing hormone	2.54E-04	-8.28	90	79	87.8	FOS, FOSB, FOSL1, FOSL2, JUNB, JUND, KRT1, MAPK-1, -3, -8, -9, -14	WP
Circadian rhythm	2.61E-04	-8.25	30	29	96.7	–	KEGG

Table 3.7. The top 20 DHS pathways identified in VU147(+) cells.

VU147(+) DHS: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
EGFR1 Signalling Pathway	3.22E-12	-26.46	177	163	92.1	CTNND1, EGF, EGFR, FOS, JUN, JUND, MAPK-1, -8, -14, PIK3CA	WP
Interferon type I	5.56E-08	-16.70	49	49	100.0	MAPK14, MTOR	WP
Epstein-Barr virus infection	1.51E-07	-15.70	200	169	84.5	CDKN1A, JUN, MAPK10, MAPK11, MAPK13, MAPK14, PIK3CA, RB1, YWHAE	KEGG
Pathways in cancer	1.69E-07	-15.59	327	265	81.0	AKT3, BIRC2, CCND1, CDKN1A, CDKN2A, E2F1, EGF, EGFR, ERBB2, FOS, IL6, JUN, MAPK-8, -10, MTOR, PIK3CA, RB1, RUNX1, TGF-β1, -β3, WNT-2, -2B, -3, -3A, -4, -5B, -6, -7A, -7B, -9A, -10A, -11, -16	KEGG
RB in Cancer	3.42E-07	-14.89	87	81	93.1	CCND1, CDKN1A, E2F1	WP
Cell cycle	3.95E-07	-14.74	124	109	87.9	CCND1, CDKN1A, CDKN2A, TGF-β1, -β2, -β3, YWHAE	KEGG
TGF-β Receptor Signalling Pathway	4.50E-07	-14.61	151	133	88.1	ATF3, CCND1, CCND1, CDKN1A, FOS, FOSB, JUN, JUND, RUNX2, TGF-β1, -β2, TP73, YAP1	WP
Hippo signalling pathway	8.37E-07	-13.99	153	131	85.6	AJUBA, BIRC2, CCND1, CCND1, CTGF, LATS1, LATS2, MOB1A, TEAD1, TEAD4, TGF-β1, -β2, -β3, TP73, WNT-2B, -2B, -3, -3A, -4, -6, -7A, -7B, -8B, -9A, 10A, 11, 16	KEGG
Integrated Breast Cancer Pathway	1.08E-06	-13.73	152	133	87.5	BACH1, CCND1, E2F1, EGFR, JUN, MAPK1, MTOR, NF1, RB1	WP
Endocytosis	1.19E-06	-13.64	203	169	83.3	EGFR, ERBB-3, -4, TGF-β1	KEGG
HTLV-I infection	3.15E-06	-12.67	259	210	81.1	AKT-1, -3, ATF-1, -3, CCND1, CDKN1A, CDKN2A, E2F1, FOS, FOSL1, IL6, JUN, MAPK8, PIK3CA, RB1, TGF-β1, -β2, -β3, WNT-2, -2B, -3 -3A, -4, -5B, -6, 7B, -8B, -9A, -10A, -11, -16	KEGG
Viral carcinogenesis	3.44E-06	-12.58	206	170	82.5	CCNA1, CCND1, CDKN1A, CDKN2A, MAPK1, PIK3CA, YWHAE	KEGG
FoxO signalling pathway	4.92E-06	-12.22	132	113	85.6	CCND1, CDKN1A, CDKN2B, EGF, EGFR, IL6, KLF2, MAPK-1, -10, -11, -13, -14, PIK3CA, TGF-β1, -β2, -β3	KEGG
Insulin Signalling	6.06E-06	-12.01	163	140	85.9	FOS, JUN, MAPK-1, -4, -6, -8 -10, -11, 14, MTOR, PIK3CA	WP
Non-alcoholic fatty liver disease (NAFLD)	8.56E-06	-11.67	147	124	84.4	IL6, MAPK-8, -10, PIK3CA, TGF-β1	KEGG
Chronic myeloid leukaemia	9.35E-06	-11.58	73	66	90.4	AKT1-3, CCND1, CDKN1A, CDKN2A, E2F1, MAPK1, PIK3CA, RB1, RUNX1, TGF-β1, -β2, -β3	KEGG
Renal cell carcinoma	1.58E-05	-11.05	66	60	90.9	JUN, MAPK1, TGF-β1, PIK3CA	KEGG
DNA damage response (only ATM dependent)	3.64E-05	-10.22	91	81	89.0	AKT-1, -3, CCND1, CDKN1A, CDKN2A, ERBB2, FOSL1, MAPK-1, -8, -10, TP73, WNT-2, -2B, -3, -4, -5B, -6, -7A, -7B, -10A, -11, -16	WP
Integrated Pancreatic Cancer Pathway	3.84E-05	-10.17	194	162	83.5	AKT1, CCND1, CDKN1A, E2F1, EGF, EGFR, ERBB2, JUN, JUNB, JUND, MAPK-1, -4, -8, -14, PIK3CA, TGF-β1	WP
Huntington's disease	3.91E-05	-10.15	177	145	81.9	–	KEGG

Table 3.8. The top 20 DHS pathways identified in UDSCC2(+) cells.

UDSCC2(+) DHS: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Viral carcinogenesis	1.48E-10	-22.63	206	101	49.0	ATF2, ATF4, CDKN1A, JUN, MAPK3	KEGG
EGFR1 Signalling Pathway	7.60E-10	-21.00	177	93	52.5	ATF1, EGF, EGFR, FOS, JUN, MAPK-1, -3, -14	WP
Spliceosome	1.84E-06	-13.21	130	62	47.7	–	KEGG
p38 MAPK Signalling Pathway	2.01E-06	-13.12	34	24	70.6	ATF2, MAPK14	WP
Integrated Breast Cancer Pathway	2.24E-06	-13.01	152	74	48.7	ATF1, BACH1, EGFR, JUN, MTOR, MAPK1	WP
Androgen receptor signalling pathway	7.48E-06	-11.80	88	47	53.4	CDKN1A, JUN, RUNX2	WP
RB in Cancer	1.33E-05	-11.23	87	46	52.9	CDKN1A	WP
Chronic myeloid leukaemia	1.53E-05	-11.09	73	38	52.1	AKT2, CDKN1A	KEGG
Leptin signalling pathway	2.71E-05	-10.52	60	34	56.7	MAPK-1, -3, MTOR	WP
Pathways in cancer	2.94E-05	-10.44	327	126	38.5	CDKN1A, EGF, EGFR, FOS, JUN,MAPK-1, -10, MTOR, RUNX1, WNT5A	KEGG
Protein processing in endoplasmic reticulum	3.27E-05	-10.33	168	72	42.9	–	KEGG
TNF-α/NF-kB Signalling Pathway	3.72E-05	-10.20	194	86	44.3	–	WP
Insulin Signalling	5.00E-05	-9.90	163	74	45.4	FOS, JUN, MAPK-1, -3, -10, 14, MTOR	WP
Alcoholism	5.67E-05	-9.78	179	75	41.9	CALM1, FOSB, MAPK3	KEGG
TSH signalling pathway	5.73E-05	-9.77	64	35	54.7	FOS, JUN, MAPK-1, -3, -14, MTOR	WP
ErbB signalling pathway	5.83E-05	-9.75	87	42	48.3	AREG, CDKN1A, EGF, EGFR, JUN, MTOR, MAPK-1, -3, 10	KEGG
Cell cycle	9.05E-05	-9.31	124	55	44.4	CCNA1, CDKN1A	KEGG
Signalling of Hepatocyte Growth Factor Receptor	9.89E-05	-9.22	33	21	63.6	FOS, MAPK-1, -3	WP
mRNA processing	1.04E-04	-9.17	134	62	46.3	–	WP
p53 signalling pathway	1.22E-04	-9.01	68	34	50.0	CDKN1A, TP73	KEGG

A high degree of overlap between the pathways of the four cell lines was observed, which resulted in 48 different pathways being identified and were assigned to 17 broad functions: HNSCC-, cancer-related, cell adhesion, migration, proliferation, growth, signalling, senescence, CNS (central nervous system), immune function, cytoskeletal, development, DNA regulation, metabolism, protein/RNA processing, and viral **(Figure 3.16)**.

Based on cell line representation, the pathways could be grouped into six clusters. Half of the pathways were categorised into group (1), which contained pathways over-represented in all four cell lines **(Figure 3.16)**. Many group (1) pathways were HNSCC-related (5/24) or cancer-related (7/24). For example, the EGFR pathway (HNSCC-related) and pathways in cancer (cancer-related) showed high statistical significance in all four cell lines, while other HNSCC-related pathways included the TGF (transforming growth factor)-β signalling, p53 signalling and MAPK (mitogen activated protein kinase) pathways **(Figure 3.16)**. Pathways related to cell adhesion/migration and the cytoskeletal network, such as adherens junction, focal adhesion and regulation of actin cytoskeleton were also over-represented in group (1) **(Figure 3.16)**.

Group (2) included a subset of six pathways which were under-represented in UDSCC2(+) cells. The pathways included Hippo signalling and proteoglycans in cancer, both of which are involved in cell adhesion, migration and proliferation processes **(Figure 3.16)**.

Group (3) included eight pathways which were under-represented in VU40T(-) DHSs **(Figure 3.16)**. Half of these pathways were associated with RNA/protein processing. Furthermore, the viral carcinogenesis and Epstein Barr Virus (EBV)

pathways were observed in this group and were more over-represented in HPV(+) cell lines **(Figure 3.16)**.

Group (4) contained four pathways over-represented more in VU147(+) DHSs, one of which was the HTLV-1 (human T-lymphotropic infection) pathway **(Figure 3.16)**. Group (5) included the HNSCC-related WNT (wingless) signalling pathway and was over-represented in the VU40T(-) DHSs. Group (6) included three pathways which were the least represented in the HNSCC cell lines and included two pathways related to CNS and one related to metabolism **(Figure 3.16)**.

3.12 Enhancer cluster analysis of DHS peaks

The HNSCC regulatory landscape was searched for enhancer clusters (EC) (also known as 'super-enhancers') [173, 175]. These regions contain a higher density of TF binding sites than normal enhancers and so it was hypothesised that ECs may highlight key genes and processes involved in HNSCC.

EC (enhancer cluster) identification was based on the method first described by Whyte *et al*. and is described in Methods and Materials, Section 2.12.4. A total of 459 ECs in SCC040(-) cells, 477 ECs in VU40T(-), 380 ECs in VU147(+) and 558 ECs in UDSCC2(+) cells were identified. Examples of ECs are shown in **Figures 3.17 and 3.18** at the *EGFR* and *TP63* genes, respectively. EGFR has frequently been reported as being overexpressed in HNSCC and is associated with reduced patient survival [36, 67, 89, 90]. The *TP63* gene, which is normally involved in epithelial differentiation and proliferation is also deregulated in HNSCC and often exhibits amplification [36, 66-70, 72].

Pathways	Function	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)	log p-value
EGFR1 Signaling Pathway	HNSCC-related					-16.70
Pathways in cancer	Cancer-related					-14.89
ErbB signaling pathway	HNSCC-related					-12.01
Adherens junction	Cell adhesion					-10.15
Endocytosis	Cytoskeletal					-8.90
TGF-beta Receptor Signaling Pathway	HNSCC-related					-8.29
FoxO signaling pathway	Cell proliferation/death					-7.61
Integrated Pancreatic Cancer Pathway	Cancer-related					-7.05
Integrated Breast Cancer Pathway	Cancer-related					-5.40
Small cell lung cancer	Cancer-related					-4.47
Renal cell carcinoma	Cancer-related					-2.75
Chronic myeloid leukemia	Cancer-related					(1)
Insulin Signaling	Metabolism					
Cell cycle	Cell proliferation					
Focal adhesion	Cell adhesion					
Regulation of actin cytoskeleton	Cell adhesion/migration					
Alpha6-Beta4 Integrin Signaling Pathway	Cell adhesion/migration					
Leptin signaling pathway	Metabolism					
Signaling of Hepatocyte Growth Factor Receptor	Cell proliferation					
p38 MAPK Signaling Pathway	HNSCC-related					
Androgen receptor signaling pathway	Cell proliferation					
TSH signaling pathway	Cell proliferation					
p53 signaling pathway	HNSCC-related					
Prostate Cancer	Cancer-related					
DNA damage response (only ATM dependent)	DNA-related					
Proteoglycans in cancer	Cell adhesion/migration/proliferation					
Hippo signaling pathway	Cell adhesion/migration/proliferation					(2)
AMPK signaling pathway	Metabolism					
Prolactin Signaling Pathway	Metabolism					
Senescence and Autophagy	Cell proliferation					
Viral carcinogenesis	Viral					
RB in Cancer	HNSCC-related					
Epstein-Barr virus infection	Viral					
Spliceosome	RNA/protein processing					(3)
Ribosome	RNA/protein processing					
Protein processing in endoplasmic reticulum	RNA/protein processing					
TNF-alpha/NF-kB Signaling Pathway	Immune-related					
mRNA processing	RNA/protein processing					
Interferon type I	Immune-related					
HTLV-I infection	Viral					(4)
Non-alcoholic fatty liver disease (NAFLD)	Metabolism					
Huntington's disease	CNS					
Axon guidance	CNS					
T cell receptor signaling pathway	Immune-related					(5)
Wnt Signaling Pathway	HNSCC-related					
Circadian rhythm	CNS					
Corticotropin-releasing hormone	CNS					(6)
Alcoholism	Metabolism					

Figure 3.16. Pathway enrichment analysis based on HNSCC DHSs.
Analysis revealed 48 different pathways identified across the four HNSCC cell lines. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into six subsets based on cell representation: (**1**) over-represented in all (24/48; 50%), (**2**) under-represented in UDSCC2(+) (6/48; 13%), (**3**) under-represented in VU40T(-) (8/48; 17%), (**4**) over-represented in VU147(+) (4/48; 8%), (**5**) over-represented in VU40T(-) (3/48; 6%) and (**6**) least represented in HNSCC (3/48; 6%).
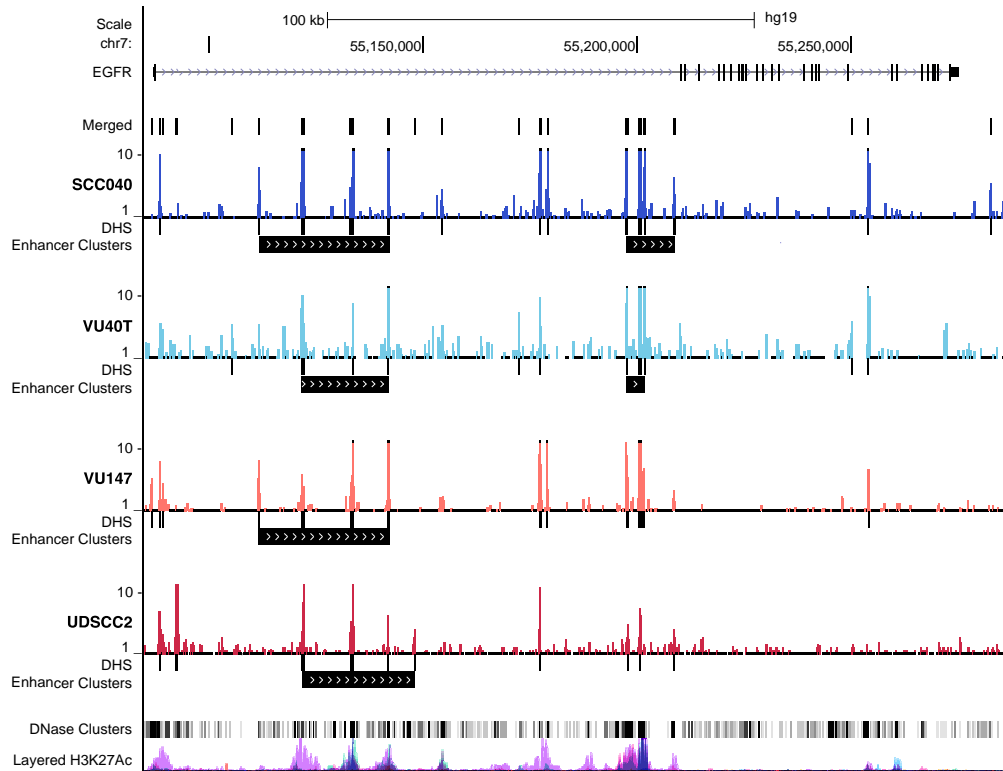
Figure 3.17. UCSC screenshot of enhancer cluster data at the *EGFR* gene.
HOMER was used to identify enhancer clusters (ECs) using 'called' DHS peaks. DHS peaks within a distance of 12.5kb were identified and concatenated. The enhancer cluster signal was normalised by taking EC reads minus normalised input reads. The signal was ranked by score and any regions with a score above one were determined an EC. A total of 459, 477, 380 and 558 ECs were identified in SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively. Coverage tracks for each cell line are displayed; SCC040(-) (dark blue), VU40T(-) (light blue), VU147(+) (orange) and UDSCC2(+) (red). Called EC peaks for each cell line are depicted below the coverage tracks (depicted by the black bars, labelled 'DHS'). The 'DNase Clusters' track contains mapped DHS reads in a panel of 125 cell lines by the ENCODE consortium. The 'layered H3K27Ac' track displays the active histone mark, H3K27Ac in ENCODE cell lines. The pink peaks in the H3K27Ac track represent the NHEK (normal human epidermal keratinocyte) cell line, indicating that some of these DHSs are epithelial specific.

146

Figure 3.18. UCSC screenshot of enhancer cluster data at the *TP63* gene.
HOMER was used to identify ECs using 'called' DHS peaks. DHS peaks within a distance of 12.5kb were identified and concatenated. The enhancer cluster signal was normalised by taking EC reads minus normalised input reads. The signal was ranked by score any regions with a score above one were determined an EC. A total of 459, 477, 380 and 558 ECs were identified in SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells, respectively. Coverage tracks for each cell line are displayed; SCC040(-) (dark blue), VU40T(-) (light blue), VU147(+) (orange) and UDSCC2(+) (red). Called EC peaks for each cell line are depicted below the coverage tracks (depicted by the black bars, labelled 'DHS'). The 'DNase Clusters' track contains mapped DHS reads in a panel of 125 cell lines by the ENCODE consortium. The 'layered H3K27Ac' track displays the active histone mark, H3K27Ac in ENCODE cell lines. The pink peaks in the H3K27Ac track represent the NHEK (normal human epidermal keratinocyte) cell line, indicating that some of these DHSs are epithelial specific.

147

The EC peaks identified in the HNSCC cell lines were assigned to the closest TSS to generate a list of genes associated with ECs. Genome-wide distribution of ECs was comparable in all four HNSCC cell lines and agreed with published data **(Appendix: Figure 1)** [44]. A total of 1,208 EC genes were identified across the four cell lines and the overlap of genes between the cell lines was investigated **(Figure 3.19)**. A quarter of the genes were specific to UDSCC2(+) DHSs (360/1208), while 121 (10%) annotated EC genes were VU147(+)-specific, 212 (18%) EC genes were VU40T(-)-specific and 177 (15%) EC genes were SCC040(-)-specific **(Figure 3.19A)**.

The overlap of genes was investigated further by calculating the percentage of genes present in all cell lines, in 3 out of 4 cell lines, in 2 out of 4 cell lines or in one cell line (cell line-specific) **(Figure 3.19B)**. It was found that the majority of genes were cell line-specific (72%), while 2.7% (33/1208) were found in all four cell lines, 7.2% (87/1208) were found in 3 out of 4 cell lines and 18% (217/1208) were found in 2 out of 4 cell lines **(Figure 3.19B)**. This pattern reflects the distribution of DHSs in the merged-HNSCC dataset **(Figure 3.12)** rather than the pattern in the top 10% of DHSs **(Figure 3.13)**, indicating that ECs do not simply reflect the size of the peaks.
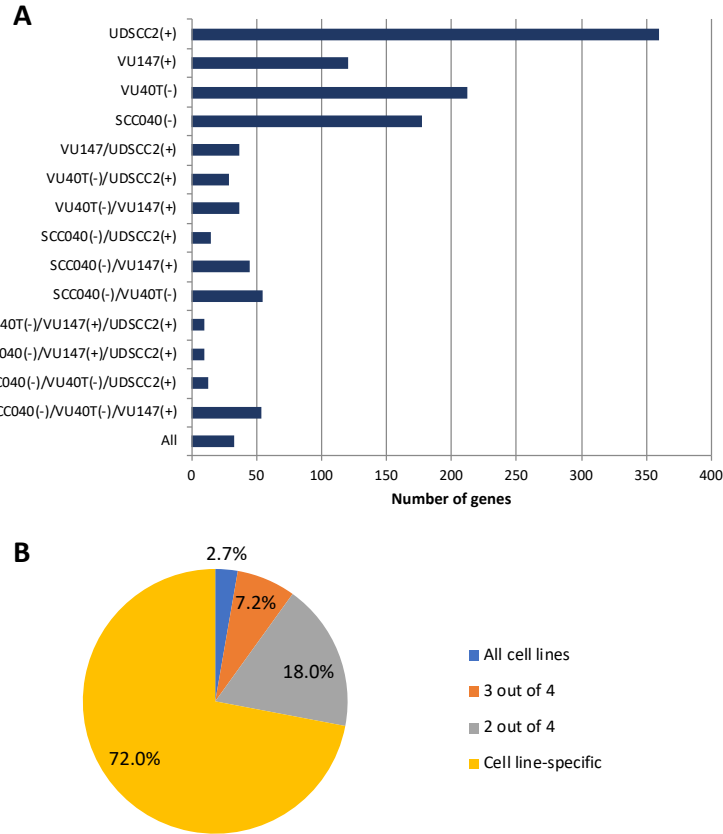
Figure 3.19. Genes annotated to HNSCC EC peaks.
EC peaks were assigned to the closest gene TSS using HOMER. **(A)** A total of 1,208 genes were annotated and the overlap between SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells was investigated. **(B)** Percentage overlap of genes was investigated by calculating the percentage of genes which were cell line-specific, present in all cell lines, in 3 out of 4 cell lines or 2 out of 4 cell lines.

The 33 genes found in all four cell lines were investigated further by conducting pathway enrichment analysis by using DAVID (Database for Annotation, Visualization and Integrated Discovery) [242] **(Figure 3.20)**. Despite the set being very small it clearly allowed for identification of the tissue as the 'head and neck neoplasia' using the Unigene database. Among the annotated genes in the 'head and neck neoplasia' pathway was the *WWTR1* (WW domain containing transcription regulator 1) gene which encodes the TAZ (tafazzin) protein, a positive effector of the Hippo pathway and *ANXA1*, which as mentioned earlier is often deregulated in cancers including in HNSCC [236, 243]. The CGAP (Cancer Genome Anatomy Project) database also revealed the 33 genes to be related to oesophageal and pharyngeal tumour disease **(Figure 3.20)**. Among the annotated genes in the oesophageal tumour disease pathway was the *CLDN1* gene which encodes claudin, a tight junction protein. The CLDN1 protein has been associated with invasion in HNSCC [244].
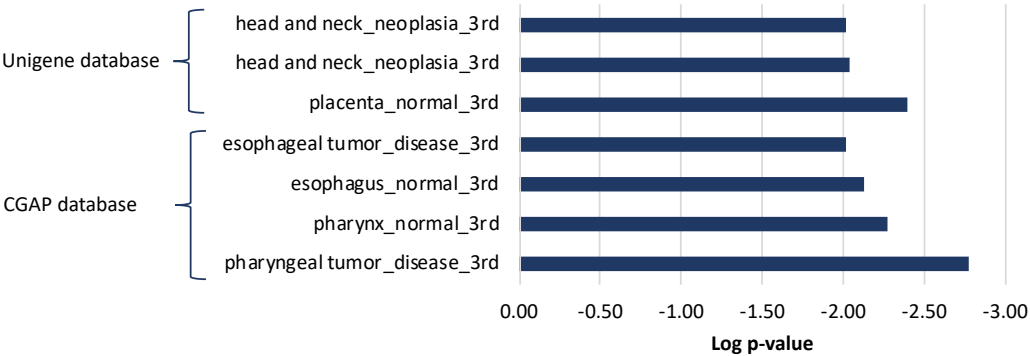
Figure 3.20. Pathway enrichment analysis of 33 genes common to all four HNSCC cell lines.
The DAVID tool was used to conduct pathway enrichment analysis. Two databases were used: Unigene and CGAP (Cancer Genome Atlas Project), which identified 3 and 4 pathways, respectively.

150

The overlap of TF genes was also investigated. A total of 23 TF genes were annotated in the HNSCC ECs, many of which were also identified by TF motif analysis, including *TP63*, AP-1 genes (*FOS*, *FOSB*, *FOSL1*), *ETS2*, SP1/KLF genes and *RUNX1* (**Table 3.9**). In addition, *LHX-2, LHX-3* and *DLX6* (distal-less homeobox 6) genes appear to be UDSCC2(+)-specific, reflecting TF motif analysis (**Tables 3.2 and 3.9**).

Table 3.9. TF genes identified during gene annotation of EC peaks.

TF family	EC gene	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)
p53	TP63	■	■	■	■
	TP73				■
AP-1	FOS				■
	FOSB			■	■
	FOSL1	■		■	
	ATF3			■	
ETS	ETS2		■		■
	ELF3	■	■		
	ETV3				■
	ETV5		■		
	ETV6		■		
	EHF			■	
SP1/KLF	SP1	■	■	■	
	KLF6		■		
	KLF7		■		
	KLF10		■		
	KLF13			■	
RUNX	RUNX1			■	
GRHL2	GRHL2	■	■		
LHX	LHX2				■
	LHX3				■
DLX	DLX3			■	
	DLX6				■

Pathway enrichment analysis was conducted using EC peaks for each HNSCC cell line; the top 20 most over-represented pathways were considered. A total of 1,179 genes were allocated to the ECs across the four cell lines (**Appendix: Tables 5-8**). Pathway enrichment analysis was conducted as described for Figure 3.16. A total of 49 different pathways were identified, grouped based on log p-value and categorised into six clusters (**Figure 3.21**).

Group (1) contained three pathways which were highly over-represented in all HNSCC cell lines: focal adhesion, α6β4 integrin signalling pathway and viral carcinogenesis (**Figure 3.21**).

Group (2) consisted of 16 pathways which are also highly involved in most of the cell lines but are less significant in UDSCC2(+) cells. This subset of pathways appear to commonly control processes such as cell adhesion/migration, as over-representation of the proteoglycans in cancer pathway, RAP1 (Rap1 GTPase) and Hippo signalling pathways are observed (**Figure 3.21**). RAP1 is a GTPase enzyme involved in the maintenance of cell-cell junctions [245].

Group (3) consisted of seven common pathways which included two HNSCC-related pathways, MAPK and TGF-β signalling pathways (**Figure 3.21**). The HTLV-1 viral pathway, cell cycle and FAS (fas cell surface death receptor) pathways were also over-represented, more so in HPV(+) cell lines.

Groups (4) and (5) contain 16 pathways which were over-represented in VU147(+) and in UDSCC2(+), respectively (**Figure 3.21**). Uniquely, these two subsets contained pathways associated with transcription and DNA regulation. In general, the differential regulation of UDSCC2(+) is more pronounced in EC analysis, with 28 out of 46 pathways being either more (group 5) or less (group 2) enriched in

UDSCC2(+) than in other cell lines. Interestingly, in group six, the SCC040(-) cell

line was most over-represented for the WNT pathway **(Figure 3.21)**.

Pathway enrichment analysis table

Pathway	Function	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)	Group
Focal adhesion	Cell adhesion					(1)
Alpha6-Beta4 Integrin Signaling Pathway	Cell adhesion/migration					(1)
Viral carcinogenesis	Viral					(1)
EGFR1 Signaling Pathway	HNSCC-related					(2)
Rap1 signaling pathway	Cell migration					(2)
Adherens junction	Cell adhesion					(2)
Leukocyte transendothelial migration	Immune function/cell migration					(2)
Proteoglycans in cancer	Cell adhesion/migration/proliferation					(2)
Pathways in cancer	Cancer-related					(2)
Regulation of actin cytoskeleton	Cytoskeletal					(2)
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Cell adhesion					(2)
Hippo signaling pathway	Cell adhesion/migration/proliferation					(2)
Tight junction	Cell adhesion					(2)
Epithelial cell signaling in Helicobacter pylori infection	Bacterial					(2)
Androgen receptor signaling pathway	Cell growth/proliferation					(2)
Bacterial invasion of epithelial cells	Bacterial					(2)
Integrated Breast Cancer Pathway	Cancer-related					(2)
Integrin-mediated cell adhesion	Cell adhesion					(2)
Pancreatic cancer	Cancer-related					(2)
Small cell lung cancer	Cancer-related					(2)
Apoptosis	Cell death					(2)
NF-kappa B signaling pathway	Immune function					(2)
Endocytosis	Cytoskeletal					(3)
MAPK signaling pathway	HNSCC-related					(3)
IL-3 Signaling Pathway	Immune function					(3)
FAS pathway and Stress induction of HSP regulation	Cell death					(3)
Cell cycle	Cell proliferation					(3)
HTLV-I infection	Viral					(3)
TGF-beta signaling pathway	HNSCC-related					(3)
Transcriptional misregulation in cancer	Transcription regulation					(4)
Ras signaling pathway	HNSCC-related					(4)
Signaling Pathways in Glioblastoma	CNS					(4)
Integrated Pancreatic Cancer Pathway	Cancer-related					(4)
Systemic lupus erythematosus	Immune function					(5)
DNA replication	DNA regulation					(5)
Alcoholism	Metabolism					(5)
Mismatch repair	DNA regulation					(5)
Nucleotide excision repair	DNA regulation					(5)
DNA polymerase alpha / primase complex	DNA regulation					(5)
RPA complex	DNA regulation					(5)
Estrogen signaling pathway	Cell growth/proliferation					(5)
IL-6 Signaling Pathway	Immune function					(5)
Endometrial cancer	Cancer-related					(5)
Signaling of Hepatocyte Growth Factor Receptor	Cell growth/proliferation					(5)
Hypertrophy Model	Protein processing					(5)
Wnt signaling pathway	HNSCC-related					(6)
IL-1 Signaling Pathway	Immune function					(6)
RANKL/RANK Signaling Pathway	Immune function					(6)
Pathogenic Escherichia coli infection	Bacterial					(6)

log -value legend: -14.13, -8.99, -5.09, -4.08, -2.37, -1.22, 0.00

Figure 3.21. Pathway enrichment analysis of enhancer clusters in four HNSCC cell lines.

Analysis revealed 49 different pathways across the four HNSCC cell lines. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into six groups by cell representation: **(1)** over-represented in all (3/49; 6%), **(2)** under-represented in UDSCC2(+) (19/49; 39%), **(3)** over-represented in HNSCC (7/49; 14%), **(4)** over-represented in VU147(+) (4/49; 8%), **(5)** over-represented in UDSCC2(+) (12/49; 24%), **(6)** over-represented in SCC040(-) or VU40T(-) (4/49; 8%).

In DHS and EC pathway enrichment analyses, 21 pathways were common to both datasets, including cell adhesion pathways (adherens junction, α6β4 integrin signalling pathway, focal adhesion, proteoglycans in cancer, regulation of actin cytoskeleton, Hippo), HNSCC-related (EGFR, MAPK, TGF-β and WNT), cancer-related pathways (integrated breast, pancreatic cancers, small cell lung cancer) and the viral carcinogenesis and HTLV-1 pathways (**Figure 3.16 and 3.21**).

3.13 RNA-seq analysis of HNSCC cell lines

Genome-wide gene expression was investigated by conducting RNA-seq analysis in the HNSCC cell lines. The Galaxy platform was used to analyse the data as described in Materials and Methods, Section 2.13.5. Within Galaxy, the Hisat2 tool was used to align reads to the hg19 genome [227]. Between 16 and 35 million reads were sequenced and between 92.6% and 97.5% of reads were aligned to the genome (**Table 3.10**).

Table 3.10. Summary of RNA-seq alignment data.

Cell line	No. of reads sequenced	% of reads aligned	No. of genes >1 FPKM
SCC040(-)	16,556,164	92.6%	8,071
VU40T(-)	34,941,803	97.5%	7,714
VU147(+)	30,411,068	96.4%	8,185
UDSCC2(+)	18,726,353	93.0%	8,807

The Stringtie tool within Galaxy was used to assemble and quantify transcripts [228]. Each gene was assigned an FPKM value (fragments per kilobase exon per million mapped reads). The FPKM is normalised using the number of mapped reads and the gene length. The cut-off point for the FPKM value was set at 1 and genes with FPKM values above 1 were considered as being expressed and taken for

further analysis [229, 230]. The total number of genes with an FPKM value above 1 were between 7,714 and 8,807 genes **(Table 3.10)**.

Clustering analysis was conducted to investigate variability between gene expression for SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) genes **(Figure 3.22)**. The greatest variability was observed between UDSCC2(+) cells and the remaining three cell lines, while the least variability was observed between the VU40T(-) and VU147(+) genes.

To investigate gene expression variability further, FPKM values were used to conduct principal components analysis (PCA) using the 'PCAexplorer' script within the R software **(Figure 3.23)**. PCA analysis generates components which are ranked by score. Score 1 represents the largest component, as most genes fit into this component (PC1), while scores 2 (PC2) and 3 (PC3) represent the second and third largest components, respectively **(Figure 3.23A)**. Over half of gene variability (54.3%) could be accounted for by PC1 **(Figure 3.23B)**. The analysis revealed that this was due to differences in gene expression between UDSCC2(+) cells and SCC040(-), VU40T(-) and VU147(+) cells **(Figure 3.23B)**. PC2 accounted for 23.7% of variability and this was found to be due to VU147(+) expression differences compared to VU40T(-), SCC040(-) and UDSCC2(+) cells **(Figure 3.23B)**. PC3 accounted for 22.1% of variability and this was due to differences between SCC040(-) gene expression and the remaining three cell lines **(Figure 3.23C)**. Hierarchical clustering of FPKM values for the four HNSCC cell lines further demonstrated the UDSCC2(+) genes to vary the most between the cell lines **(Figure 3.23D)**.
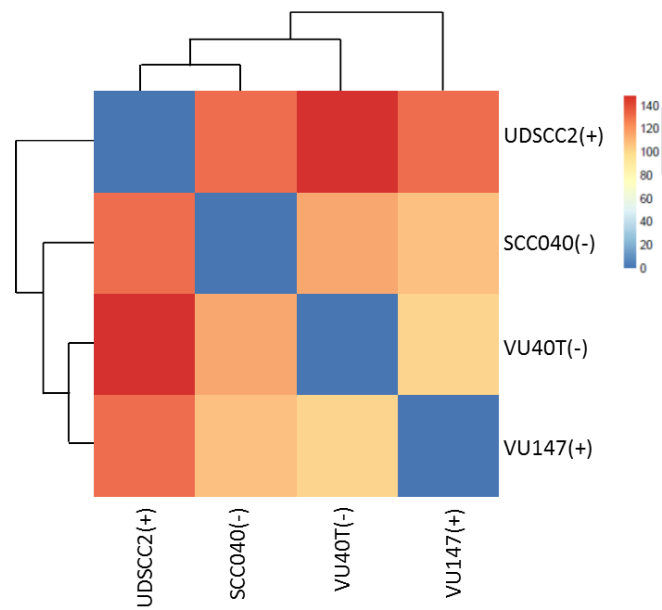
Figure 3.22. Clustering analysis of gene expression data.
RNA-seq data for SCC040(-), VU40T(-), VU147(+) and UDSCC2(+) cells yielded 8,071, 7,714, 8,185 and 8,807 genes, respectively with an FPKM (fragments per kilobase exon per million mapped reads) value above the cut-off value of 1. Clustering analysis was performed using the 'PCAexplorer' script within the R software. The scale represents variability from least gene expression variability to most.
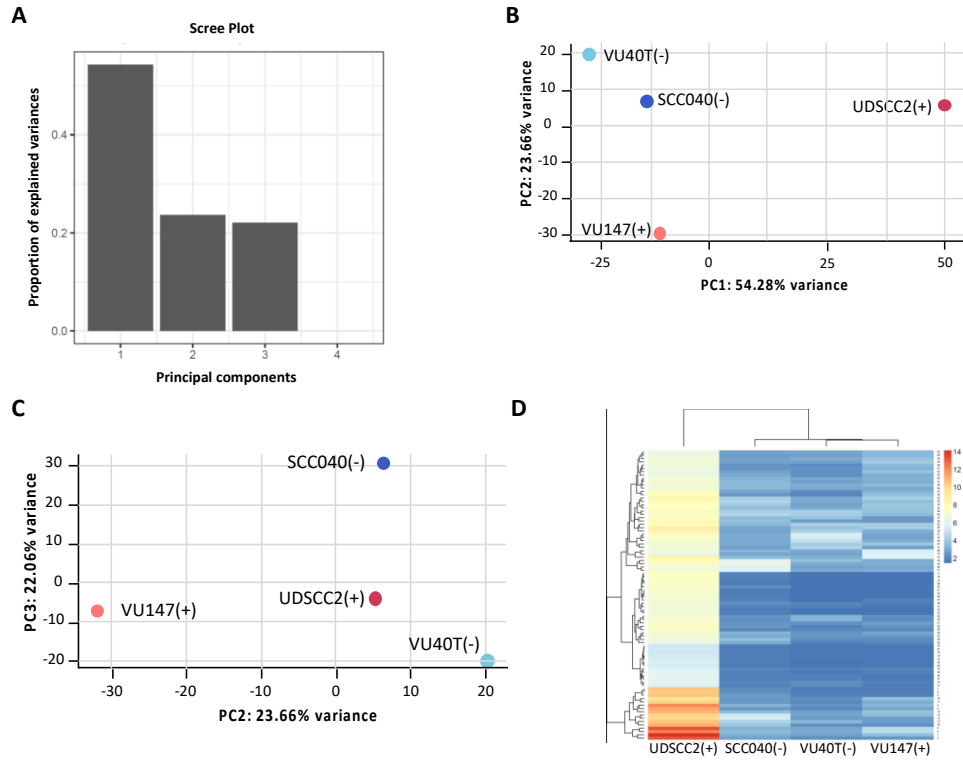
Figure 3.23. Principal components (PCA) and clustering analyses of gene expression data.
(**A**) Three principal components (PCs) were identified upon clustering of variabilities in gene expression: PC1, PC2 and PC3. (**B**) PC1 accounted for 54.28% of variability and this was due to differences between UDSCC2(+) gene expression and VU40T(-), SCC040(-) and VU147(+) expression. (**B and C**) PC2 accounted for 23.66% of variability and this was due to differences between VU147(+) gene expression and VU40T(-), SCC040(-) and VU147(+) expression. (**C**) PC3 accounted for 22.06% variability, which was due to expression differences between SCC040(-) and the remaining three cell lines. (**D**) Clustering analysis of FPKM values revealed the greatest variability of UDSCC2(+) gene expression compared to the remaining three cell lines. Analysis was conducted using the 'PCAexplorer' script within the R software.

To investigate which pathways the expressed genes in the HNSCC cell lines were associated with, pathway enrichment analysis was conducted using the ConsensusPath database [231]. The genes were sorted by FPKM values and the top 2000 genes were run through the database; the data for the KEGG and WP databases were extracted. The top 20 pathways were identified and grouped as described in **Figure 3.16**. A total of 35 pathways were identified, within which four groups were identified **(Figure 3.24)**.

Group (1) consisted of 11 pathways, which were highly over-represented in two or more cell lines. Over half of the pathways in this group (6/11) were associated with RNA/protein processing. In addition, the cell cycle pathway was also over-represented **(Figure 3.24)**.

Group (2) could be further split into two subsets: 2a and 2b. Subset 2a consisted of 11 pathways which were largely over-represented in UDSCC2(+) cells **(Figure 3.24)**. Subset 2b mainly consisted of pathways associated with microRNAs (miRNA) (small single-stranded, non-coding RNAs) and were under-represented in VU147(+) **(Figure 3.24)**.

Group (3) consisted of pathways which were well represented in the VU40T(-), VU147(+) and UDSCC2(+) cell lines, while being significantly over-represented in SCC040(-) cells **(Figure 3.24)**. This group contained pathways associated with cell adhesion, proliferation, migration, as well as the HNSCC-related pathway, TGF-β. This group therefore, reflects well the DHS pathway enrichment analysis. Striking is the lack of HNSCC-related signalling pathways such as the EGFR, ERBB, p53, Hippo and MAPK signalling pathways **(Figure 3.24)**.

Pathway	Function	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)		log p-value
mRNA Processing	Protein/RNA processing						7.5E-06
Translation Factors	Protein/RNA processing						1.9E-04
Cytoplasmic Ribosomal Proteins	Protein/RNA processing						4.1E-04
Ribosome	Protein/RNA processing						1.0E-03
Cell cycle	Cell proliferation						2.3E-03
RNA transport	Protein/RNA processing					(1)	7.4E-03
Protein processing in endoplasmic reticulum	Protein/RNA processing						
Androgen receptor signaling pathway	Cell growth/proliferation						
Electron Transport Chain	Metabolism						
Retinoblastoma (RB) in Cancer	HNSCC-related						
Dilated cardiomyopathy	Cell adhesion						
miR-targeted genes in leukocytes	micro RNA						
Spliceosome	Protein/RNA processing						
miR-targeted genes in epithelium	micro RNA					(2b)	
miR-targeted genes in muscle cell	micro RNA						
miR-targeted genes in lymphocytes	micro RNA						
miR-targeted genes in squamous cell	micro RNA						(2a)
Primary Focal Segmental Glomerulosclerosis FSGS	Renal degeneration						
Epstein-Barr virus infection	Viral						
Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)	Cell adhesion						
Pathogenic Escherichia coli infection	Bacterial						
Thyroid hormone signaling pathway	Cell growth/proliferation						
Tight junction	Cell adhesion						
Proteoglycans in cancer	Cell adhesion						
Regulation of actin cytoskeleton	Cytoskeletal						
Adherens junction	Cell adhesion					(3)	
TGF-beta Signaling Pathway	HNSCC-related						
Pathways in cancer	Cancer-related						
Estrogen signaling pathway	Cell growth/proliferation						
Focal adhesion	Cell adhesion						
ATF6-alpha activates chaperone genes	Protein/RNA processing						
Oocyte meiosis	Development						
Oxidative phosphorylation	Metabolism					(4)	
Parkinson's disease	CNS						
Progesterone-mediated oocyte maturation	Development						

Figure 3.24. Pathway enrichment analysis of genes expressed in HNSCC cells.
Genes were sorted by FPKM and the top 2,000 genes were taken for pathway enrichment analysis by using the ConsensusPath database; a total of 35 pathways were over-represented. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. Pathways were grouped into four groups by representation: **(1)** over-represented in all (11/35; 31%), **(2)** over-represented in UDSCC2(+) (11/35; 31%), **(2a)** under-represented in VU147(+) (6/35; 17%), **(2b)** over-represented in UDSCC2(+) (5/35; 14%), **(4)** under-represented in all (5/35; 14%).

3.14 Discussion

Genome-wide analysis of the HNSCC (head and neck squamous cell carcinoma) regulatory network was conducted by using the DNase I-seq assay to generate DNase I hypersensitive sites (DHSs). As DHSs are representative of open chromatin regions, they were used to gain insight into *cis*-regulation of HNSCC by analysing a heterogeneous population of four HNSCC cell lines, derived from the tongue, floor of the mouth and hypopharynx.

A large number (73,414) of potential regulatory elements were identified across the four cell lines with the majority of them being cell line-specific, therefore reflecting the heterogeneity of the cell lines used. However, these cell line-specific peaks were observed as being smaller than shared peaks. It has been suggested that small peaks may reflect the lack of stability of the DHS and it may be that cell line-specific DHSs may be present in a fraction of a given cell population [188]. Alternatively, it may be that cell-line specific peaks display dynamic chromatin accessibility, where DNA regulatory elements are transiently accessible for TF binding [180, 246, 247].

Despite the described heterogeneity, interrogation of the largest DHS peaks (top 10%) revealed that nearly 70% were shared between all four cell lines indicating that similar active regions underlie the core regulatory network in these HNSCC cell lines. Alternatively, the shared, largest DHSs could reflect key regulation of basic cell metabolism including replication and translation. TF motif and pathway enrichment analyses also show a prominent overlap between the cell lines. Therefore, despite large differences in the localization of regulatory elements, the pattern of signalling pathway activation remains generally the same.

Pathway enrichment analysis of HNSCC DHSs revealed two large groups of pathways: over-represented in all HNSCC cell lines and under-represented specifically in UDSCC2(+) cells. The UDSCC2(+) cell line had the highest proportion of unique peaks, which was further reflected in a differential TF motif profile, as well as in pathway enrichment analysis. This observation was further confirmed and enhanced by EC (enhancer cluster) and RNA-seq analyses. Therefore, UDSCC2(+) cells being hypopharyngeal and originating from reticulated epithelium may explain its distinct regulatory and genetic profile compared to the other HNSCC cell lines, which are of oral origin. To support this, previously published data highlight molecular differences in HNSCC of the oral cavity compared to the hypopharynx. For example, methylation of the *CDKN2A* (cyclin dependent kinase inhibitor 2A) promoter, which encodes for the cyclin/CDK inhibitor, p16, was found to be significantly higher in oral tumours compared to those of the hypopharynx [248]. Furthermore, loss-of-heterozygosity of the *TP53* gene was found to be significantly higher in hypopharyngeal tumours, compared to those derived from oral cancers [248].

TF motif analysis revealed that the HNSCC DHS peaks are dominated by the AP-1 motif, but also the presence of p53 (p63 and p73), TEAD (TEA domain transcription factor) (1-4), RUNX (runt-related transcription factor) (1-3), KLF (Krüppel-like factor) (1-17) and ETS (E twenty-six transcription factor) family motifs were detected. The ETS TF family consists of 13 subfamilies, which includes the ETS, ERG (ETS related gene), ELK, ELF (E74-like factor), ETV (ETS variant), EHV (equine herpesvirus) and EHF (ETS homologous factor) TF families [249-251]. Members of these TF families have been previously characterised in SCC: (i) JUN and FOSL1, of the AP-1 family, have previously been linked to

invasion and metastasis of SCC cells, including in HNSCC [117, 252-255]. In addition, constitutive activation of JUND has been linked to the induction of pro-inflammatory signals in HNSCC cells [240, 252, 255, 256]. (ii) The *TP53* gene is one of the earliest genes to undergo mutations in HPV(-) HNSCC tumours, and its degradation can be induced by E6 in HPV(+) tumours leading to cell cycle deregulation [29, 36, 66-71]. (iii) The p63 protein is essential in the organisation of the epithelium and is known to be overexpressed in HNSCC [12, 13, 16, 257-259]. (iv) Expression of the *KLF4* gene has been associated with poor prognosis in HNSCC patients and its overexpression in HNSCC cells has been shown to increase cell invasion and migration [260]. In addition, an EC has been identified, associated with the *KLF5* gene in HNSCC tumours, moreover focal amplifications of this EC have been linked to KLF5 protein overexpression and associated with SCC [179]. (v) Both *RUNX1* and *RUNX2* were annotated in DHS pathway enrichment analysis of all HNSCC cell lines or UDSCC2(+) only, respectively. In oral SCC tumours in mice, Runx1 expression has been seen to be pro-tumourigenic [261]. *RUNX2* has been reported to display abnormal methylation patterns in cancers, including HNSCC [262, 263]. RUNX3 TF overexpression has been shown to positively correlate with proliferation of HNSCC cell lines [264]. (vi) Evidence of gene alterations in the *ETS1*, *ETV3* and *ELF3* genes have been observed in HNSCC cell lines [265]. ETS-1/2 have been linked to activation of the RAS (resistance to audiogenic seizures)/MAPK pathway in epithelial cells (kidney and prostate cell lines) [266]. Therefore, TF motif analysis has revealed several TFs which are known to be involved in HNSCC development.

Until recently, the TEAD TFs had not been implicated in HNSCC. It has now been reported that TEAD, its activator YAP (yes-associated protein), and an inactive

form of p53 have a role in the regulation of a circular RNA transcript of the *PVT1* (pvt1 oncogene) gene (a long non-coding RNA), which is often upregulated in cancers, including HNSCC tumours [267].

In addition to common regulation, cell line specificity was also observed- LHX (LIM homeobox)/DLX in UDSCC2(+) and GRHL2 (grainyhead-like protein 2) in SCC040(-). Of the LHX TF family, *LHX-6* is a tumour suppressor gene, which undergoes hypermethylation in HNSCC tumours thereby silencing its expression [268]. This may be the case in UDSCC2(+) cells, as RNA-seq data generated from the UDSCC2(+) cells reveal no expression of *LHX-6* expression (data not shown). On the other hand, GRHL2 has been reported to activate the MAPK pathway and contribute to HNSCC development *in vivo* and is expressed in our SCC040(-) cells (data not shown) [269].

Furthermore, the TF motif analysis highlighted two types of regulatory factors: acting at promoters and at the distal regulatory elements (located >1kb from the TSS). The promoter-specific *de novo* motifs highlighted in this work have been found by previous groups to be signatures of TSSs and include YY1 (yin yang 1), ETS, NRF (nuclear respiratory factor)-1, NFY, CRE (cAMP response element) and GFY (golgi associated olfactory signaling regulator) [270, 271]. Distal DHS-unique motifs identified here included TEAD, RUNX and p53/p63 families. Further to this, there is evidence that TEAD and RUNX can act as DNA binding partners for the downstream Hippo proteins, YAP/TAZ and p63 may be a downstream target of EGFR [272, 273]. Therefore, this may reflect associations of TEAD, RUNX and p63 at distal regulatory elements and involvement in the Hippo and EGFR pathways [88, 92]. The AP-1 motif, although present in both groups also showed much stronger association with distal enhancers. Occupancy of TEAD4, JUND, FOSL1

and FOSL2 has previously been investigated in A459 (lung epithelial cell line). It was observed that nearly 90% of ChIP (chromatin immunoprecipitation)-seq peaks overlapped active enhancer regions and 12% of peaks with active promoters, agreeing with our data [274].

Pathway analysis of DHSs and ECs revealed many known HNSCC pathways (HNSCC-related). For example, EGFR, TGF (transforming growth factor)-β, MAPK, pRB (retinoblastoma protein) and p53 were over-represented in all four cell lines. EGFR is well known in HNSCC to be overexpressed in 90% of HNSCC tumours and is a therapeutic target to control aberrant EGFR expression [70, 91]. The TGF-β pathway is essential in epithelial homeostasis; disruption of the pathway in HNSCC mouse models by inhibition of Smad4 (mothers against decapentaplegic homolog 4), a key mediator of the pathway, can lead to the development of HNSCC tumours in mice [275]. The MAPK family are downstream effectors of the EGFR pathway and are involved in cell proliferation; *MAPK* gene levels have been reported to be significantly altered in SCC [66]. The p53 and pRB pathways are commonly disrupted in HNSCC either due to the action of the viral oncoproteins E6 and E7 in HPV (human papillomavirus)-(+) HNSCC tumours or, due to altering mutations in both HPV(+) and HPV(-) HNSCC tumours [20, 56-58].

Pathways associated with cancers other than HNSCC were also over-represented in DHS and EC pathway enrichment analysis (cancer-related pathways). For example, over-representation of the breast, pancreatic, renal, small cell lung, prostate cancers and chronic myeloid leukaemia were observed across the four cell lines. The presence of these pathways is perhaps expected, as many of the same genes are deregulated in different cancer types. For example, there is evidence that genes

belonging to the *MAPK*, *WNT*, AP-1 families and *EGFR* are associated with HNSCC but are also overexpressed in the cancers mentioned above [252, 276-278].

Analysis of pathways associated with DHSs and highlighted by genes marked by ECs, revealed pathways involved in cell adhesion and migration to be over-represented in the HNSCC cell lines. The reason for this may be that cell adhesion is strongly related to the identity of epithelial cells and underlies the structure of epithelium where cell-cell or cell-extracellular matrix contacts are of particular importance. However, in cancer, cell adhesion proteins become deregulated and such phenotypic changes aid cell invasion and metastasis [21, 116].

Pathway analysis of DHSs and ECs revealed an association with viral pathways. The HTLV-1 (human T-cell lymphotropic virus type 1) and EBV (Epstein Barr virus) pathways were over-represented more significantly in HPV(+) DHSs and ECs. The annotation of EBV and HTLV-1 could be explained by the fact that many of the same genes are disrupted in viral infections and therefore annotated. For example, upon EBV infection, the viral oncoprotein, latent membrane protein 1 (LMP1) has been shown to indirectly inhibit p53-mediated apoptosis, thereby disrupting p53 and its target genes associated with apoptosis, proliferation and cell survival [279]. This is also observed in HPV infected cells, where E6 can lead to p53 degradation [20, 57, 58]. Furthermore, HTLV-1 infection has been linked to AKT (protein kinase B) activation, which is part of the PI3K (phosphotidylinositol-3 kinase)/AKT/mTOR (mechanistic target of rapamycin) pathway. Aberrant AKT activation could deregulate some of the same pathway targets as in HPV infection as *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha)*, which is involved in the PI3K/AKT/mTOR pathway, is frequently amplified in both HPV(+) and HPV(-) tumours [36, 66-71, 252, 276-278, 280]. The presence

of other viral pathways in this data can also be explained by a study, which demonstrated that upon infection of any of four DNA viruses (HPV, EBV, adenovirus and polyomavirus), common pathways and proteins become deregulated, such as the NOTCH pathway [281]. In addition to viral pathways, the viral carcinogenesis pathway was also overrepresented in all four HNSCC cell lines in the EC pathway enrichment analysis. The viral carcinogenesis pathway describes the mechanism of carcinogenesis of HPV and other viruses (hepatitis B, hepatitis C, EBV, HTLV-I and Karposi's sarcoma-associated herpesvirus). As described before, the initial mechanisms of carcinogenesis may differ between the viruses, but the same downstream oncogenes may be involved, leading to overrepresentation of this pathway [281].

Pathway enrichment analysis based on DHS and EC data highlighted signalling pathways already linked to HNSCC or of potential importance (Hippo pathway), which were not found in analysis based on RNA-seq data. Instead, the latter highlighted the involvement of protein/RNA processes (over-representation of pathways such as translation factors, cytoplasmic ribosomal proteins, ribosome, protein processing in endoplasmic reticulum and RNA transport), as well as significant involvement of micro RNA (miRNA)-targeted processes in UDSCC2(+) cells. MiRNAs repress mRNA by blocking their translation or targeting the mRNA for degradation [134]. It has been reported by a number of studies that miRNA levels can be altered in the presence of HPV, which may explain the presence of miRNA pathways in the data here [134, 282-284].

The importance of ECs has previously been demonstrated by their association with cell identity [173, 175]. Reflecting this, our data revealed a very small number of genes (n=33) to be associated with ECs in all cell lines, clearly delineating the cells'

origin in HNSCC and highlighting the ECs' role in HNSCC development. As mentioned earlier, ECs have previously been associated with HNSCC cells, for example the *KLF5* EC [179]. In addition, our data highlighted the *EGFR* and *TP63* as ECs and the main pathways which appear to be regulated by EC genes in all cell lines, are associated with cell adhesion (focal adhesion and α6β4 integrin) and the viral carcinogenesis pathway. Interestingly, it was observed that many TFs, including those highlighted by TF motif analysis were regulated by ECs, including p63, p73, AP-1, KLF and RUNX. In addition, LHX/DLX ECs were revealed to be UDSCC2(+)-specific, reflecting the cell line specificity of UDSCC2(+) TF motifs. Therefore, EC analysis supports our data and agrees with previously published data. ECs may prove useful in pointing towards mechanisms of HNSCC.

Analysis of the regulatory network did not reveal clear differences between HPV(+) and HPV(-) cell lines. This may be due to the fact that UDSCC2(+) cells exhibit large differences compared to the remaining HNSCC cell lines. However, it must also be considered that only two cell lines from each subgroup were analysed and a larger sample size may reveal distinctions between HPV(+) and HPV(-) cells. Therefore, although the initial trigger for HNSCC development may differ (HPV oncoproteins in HPV(+) cells and mutations in both HPV(+) and HPV(-) cells), downstream *cis*-regulation and pathway signalling appears to be similar.

In conclusion, a combination of genome-wide analyses have been used to elucidate the regulatory and the transcriptome of HNSCC cells. The data have highlighted the majority of stable DHS peaks to overlap between the four HNSCC cell lines investigated, indicating a common core regulatory network shared by the cell lines. This was confirmed by TF motif analysis and pathway enrichment analysis, where many of the same TF motifs and pathways were identified in all four cell lines.

However, the UDSCC2(+) cells exhibited a subset of unique pathways and TF motifs which could be explained by their hypopharyngeal origin. DNase I-hypersensitivity assay identified HNSCC-related pathways and TF motifs, while RNA-seq revealed the involvement of miRNAs. In addition, our data has identified pathways known to be deregulated in HNSCC (EGFR, MAPK, WNT, TGF-β), as well as TFs associated with the development of HNSCC (AP-1, p63, KLF and RUNX). The Hippo pathway and its transcription effector, TEAD, has also been highlighted to be of importance in HNSCC cells.

Limitations of these results include the method of assigning the nearest gene to each DHS in order to conduct pathway enrichment analyses. While this method has been used previously to elucidate which processes distal regulatory elements are involved with [285, 286], the advent of chromatin conformation capture has highlighted the oversimplification of this method [287]. Chromatin conformation capture allows the spatial identification of long-range chromatin interactions between distal regulatory elements and promoters. For example, other groups by using chromatin conformation capture have found that 27% of distal elements can interact with its nearest gene, while 47% of distal elements can interact with the nearest active gene [142]. Therefore, by using chromatin conformation capture to investigate distal element-promoter interactions and conduct subsequent pathway analysis, identification of the biological processes involved in the cells could be improved.

A second limitation of the findings of this chapter is the small sample size of four cell lines. Investigations of this sample size was further limited by the use of a pharyngeal cell line (UDSCC2(+)), which exhibited distinct differences in its regulatory landscape, as discussed earlier. This limitation possibly hindered the

identification of differences between the *cis*-regulatory network of HPV(+) and HPV(-) cell lines. To overcome this limitation, the DNase I-hypersensitivity assay could be conducted in a larger sample size of oral SCC cell lines.

Finally, the results described in this chapter are incomplete without including data from normal, untransformed cells of similar origin; this has been approached in the next chapter (Chapter 4). Data in Chapters 3 and 4 also laid a foundation for follow-up analyses to investigate interactions between different regulatory networks identified (Chapter 5) and the Hippo and EGFR pathways (Chapter 6).

CHAPTER 4: CHARACTERISATION OF THE NON-TUMOURIGENIC ORAL KERATINOCYTE REGULATORY NETWORK

4.1 Introduction

In the previous chapter, *cis*-regulatory elements were identified by conducting genome-wide DNase I hypersensitive assay. TF motifs which were distal-specific and common to all HNSCC (head and neck squamous cell carcinoma) cell lines were identified and were indicative of TFs involved in the functioning of regulatory elements. These included AP-1, TEAD (TEA domain transcription factor), p53/p63/p73, RUNX (runt-related transcription factor) and CTCF (CCCTC-binding factor)/BORIS. In addition, pathway enrichment analysis revealed the EGFR (epidermal growth factor receptor) and Hippo pathways be involved in HNSCC.

4.2 Aims

In terms of selecting potential therapeutic targets, an understanding of how the HNSCC network differs from that of non-tumourigenic oral keratinocytes is required. One possibility of how the networks may differ is that certain TFs and signalling pathways are only activated in either cancer or in non-tumour cells. Alternatively, the same pathways could be recognized in both cell types but with different activities of their repressive and activating components. Based on the observations from Chapter 1, combined with published data, the latter option would

be more likely. Furthermore, cancer-related changes within a pathway could be due to replacement of specific TFs within a family, their overexpression due to genetic and non-genetic causes and/or enhancement of signalling pathway activation downstream of TF binding.

To test this hypothesis, the DNase I hypersensitivity assay was attempted in two non-tumour cell types: primary OTKs (oral tonsil keratinocytes) and OKF6 (hTERT immortalised oral keratinocytes). Several attempts were undertaken to achieve this goal, however, due to slow population doubling times of both cell types, a small-scale version of the DNase I assay was used where five million nuclei per DNase I preparation were used, rather than 20 million. The assay was carried out as described in Chapter 3 and tested initially using VU40T(-) cells. Unfortunately, this approach was not successful in the non-tumour cells. Therefore, an alternative method was explored called, ATAC (Assay for Transposase-Accessible Chromatin), followed by next-generation sequencing (ATAC-seq). ATAC-seq was performed in OTK, OKF6 and at this point a third primary non-tumour cell type was obtained, HOK (human oral keratinocyte) cells. ATAC-seq was unsuccessful in OKF6 and HOK cells, as will be discussed next. Therefore, in this chapter the small-scale DNase I-seq assay in VU40T(-), OKF6 and OTK cells is described, as well as, ATAC-seq data for OTK cells. In addition, RNA-seq data for HOK cells is shown, as it was unsuccessful in OTK and OKF6 cells.

4.3 Small-scale DNase I hypersensitivity assay test in VU40T(-) cells

The VU40T(-) cells were treated with 50U of DNase I based on the results described in Chapter 3, Section 3.4. An untreated sample served as the negative control and the samples were visualised on an agarose gel (**Figure 4.1**).
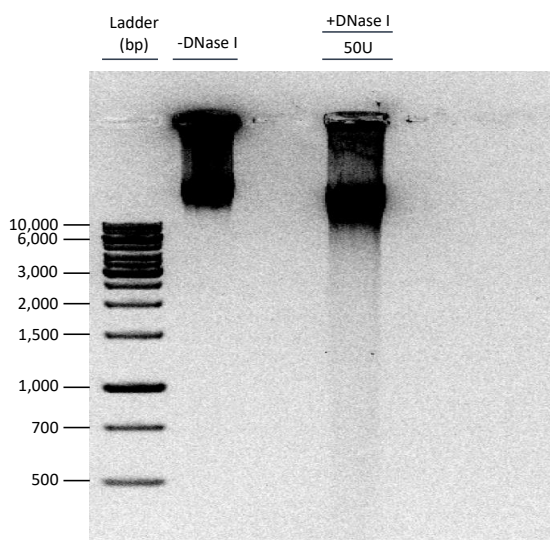
Figure 4.1. Agarose gel electrophoresis of DNase I digested chromatin.
Five million nuclei from VU40T(-) cells were isolated and treated with 50U of DNase I enzyme for 3 mins at 37°C. The sample was visualised on a 2% agarose gel. A 'No DNase I' negative control was used to assess for endogenous chromatin digestion. The absence of a DNA smear in the negative control indicates intact chromatin.

The release of small DNA fragments was observed from the upper DNA band of high molecular weight DNA which confirmed digestion. To enrich for the desired sized fragments of 100-200bp, the DNA fragments were separated by size using ultra-centrifugation; 13 fractions were collected and visualised by gel electrophoresis (**Figure 4.2A**). Upon visualisation, it was not clear whether fractionation was unsuccessful or whether fractionated fragments could not be

detected due to low DNA concentration. Therefore, to increase DNA concentration, fractions 1-3, 4-9 and 10-13 were combined based on DNA fragment sizes obtained in the large-scale assay. Combined fractions were visualised on an agarose gel **(Figure 4.2B)**. A strong band was observed in fractions 10-13 of 10kb; bands of the same size were also observed in fractions 1-3 and 4-9, though these were much weaker. A smear could also be observed between 50-200bp in each set of fractions, which may be the accumulation of smaller DNA fragments. Therefore, qPCR (quantitative polymerase chain reaction) was conducted to test for DHS (DNase I hypersensitive site) enrichment. The *KRT* (keratin)-*6A* DHS primers were used as a positive control and the 'no DHS' primers as the negative control **(Figure 4.2C)**. The fold-enrichments of fractions 1-4, 5-9 and 10-13 were 41.6-, 61.3- and 267.8-fold, respectively. This suggested that DHSs are enriched using the small-scale assay in VU40T(-) cells **(Figure 4.2C)**.
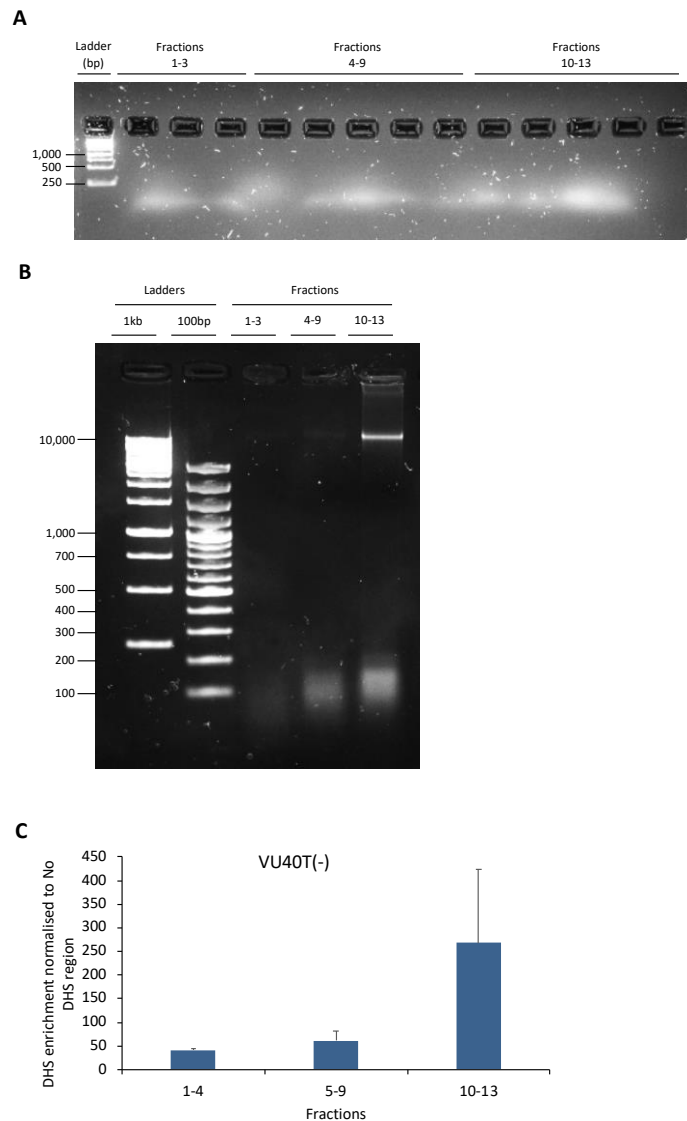
Figure 4.2. Fractionation and enrichment of VU40T(-) DNase I fragments.
(A) Generated DNA fragments were separated by centrifugation on a sucrose gradient and 13 fractions were collected and visualised on an agarose gel to check for fractionation. **(B)** Combined fractions 1-3, 4-9 and 10-13 were merged and visualised by agarose electrophoresis. **(C)** QPCR was conducted at the '*KRT6A* DHS' region. The signal was normalised to the 'no DHS' negative control region. The bars represent the range between technical duplicates.

4.4 Small-scale DNase I hypersensitivity assay in oral keratinocyte cells

Nuclei isolation was conducted in OTK, OKF6 and HOK cells. Due to the difficulty in obtaining a sufficient number of cells to isolate nuclei, only two NP-40 concentrations were tested, 0.01% and 0.02%. These two concentrations were chosen based on the lower range of NP-40 concentrations tested in HNSCC cells. It was found that the 0.02% NP-40 concentration yielded nuclei within the optimum range, as described in Chapter 3, Section 3.4.

A sufficient number of nuclei could not be isolated from the three cell types to optimise DNase I digestion. However, based on frequent observations of chromatin degradation in negative control samples of non-tumour cells (not shown), it was thought that non-tumour cells may display higher sensitivity to DNase I digestion. Therefore, based on the DNase I amount used in HNSCC cells, 30U and 40U were used in the non-tumour cells **(Figures 4.3A and B)**. DNase I digestion was not successful in HOK cells, as the nuclei preparation remained viscous, indicating intact DNA and so was not taken forward. In the OKF6 samples, the release of high molecular weight fragments in preparations digested with 30U or 40U of DNase I were observed, however no observable difference between the two digestions was detected **(Figure 4.3A)**. A very faint smear could be observed in the 'no DNase I' negative control, indicating some degradation of the chromatin **(Figure 4.3A)**. The OTK nuclei preparation treated with 40U of DNase I yielded the release of more DNA fragments compared to the 30U preparation. However, the 30U preparation appeared similar to the negative control sample, in which a faint smear could be observed, indicating chromatin degradation **(Figure 4.3B)**. In order to assess for

any DHS enrichment, the 40U OKF6 and OTK samples were taken forward for qPCR.

In order to confirm the enrichment of DHSs, qPCR was conducted on each fraction set using the *KRT6A* DHS primers and its signal was normalised to the signal obtained from amplification of the 'no DHS' region. In OKF6 cells, fold change over the negative control signal for 40U DNase I preparation was 0.4-fold for fractions 1-3 and 4-9, and 0.7-fold for fractions 10-13 (**Figure 4.4A**). In OTK cells, the 40U nuclei preparation displayed fold change of 1.6-, 2.5- and 0.7-fold (**Figure 4.4B**). Therefore, in OKF6 cells, there was no DHS enrichment over background and in OTK cells, an enrichment of maximum 2.5-fold was observed. This result indicated that the small-scale assay could not be used in OKF6 and OTK cells to isolate DHSs.
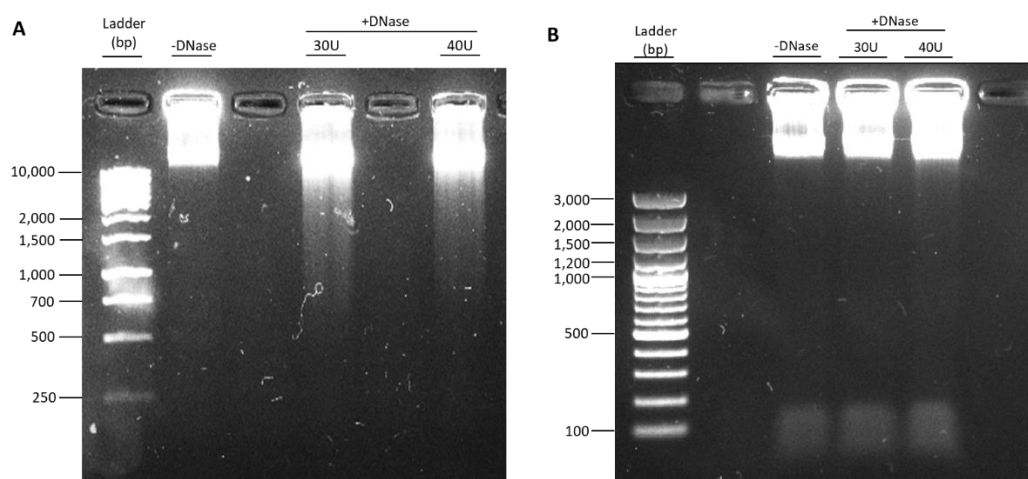
Figure 4.3. Visualisation of DNase I digested chromatin by agarose gel electrophoresis.
Nuclei were isolated from (**A**) OKF6 cells and (**B**) OTK cells and treated with 30U or 40U of DNase I enzyme for 3 mins at 37°C. Samples were visualised on a 2% agarose gel. A 'no DNase I' negative control was used to assess for endogenous chromatin digestion.
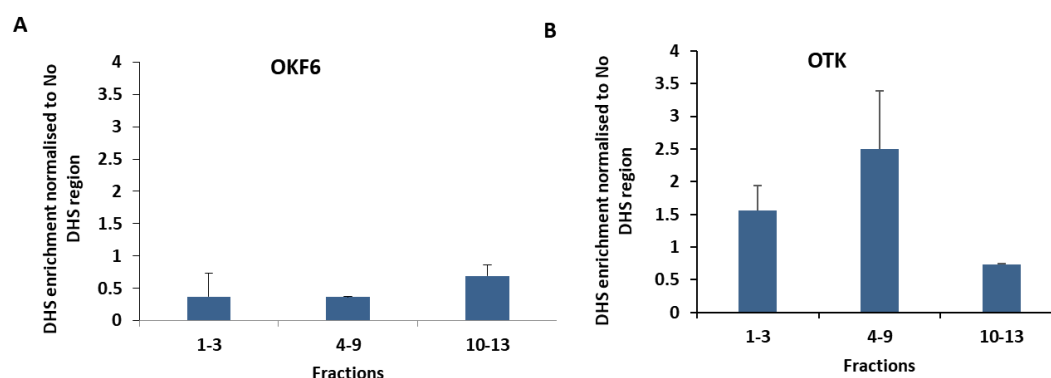
Figure 4.4. Enrichment of DHSs in non-tumour keratinocytes assessed by qPCR.
Fractions between 1-3, 4-9 or 10-13 were merged and collected from **(A)** OKF6 and **(B)**
OTK cells. QPCR was conducted at the *KRT6A* DHS region. The signal was normalised to
a negative control 'no DHS' region. The bars represent the range between technical
duplicates.

4.5 ATAC-seq sample preparation

The ATAC (Assay for Transposase Accessible Chromatin)-seq method allows for

genome-wide mapping of accessible chromatin regions with the advantage of

requiring fewer cells (25,000-50,000) compared to DNase I-seq [181]. This

technique involves a digestion with a transposase enzyme, which is flanked by

adapter sequences. When the enzyme cuts at open regions of the genome, adapters

are also inserted, thereby combining the chromatin digestion and library preparation

steps. ATAC-seq was used to identify open chromatin regions in OTK cells and

HOK cells. Only the data for OTK cells is shown here, as sequencing of HOK and

OKF6 cells was unsuccessful.

The transposition reaction was conducted in OTK cells, followed by PCR

(polymerase chain reaction) to amplify the library. After five cycles of PCR, qPCR

was carried out to quantify the library and calculate additional PCR cycles needed,

as recommended by Buenrostro *et al.* [181]. To calculate the number of additional

PCR cycles required, the SYBR fluorescence values were plotted against cycle number **(Figure 4.5)**. The number of cycles corresponding to 25% of the maximum fluorescence signal indicated the number of additional cycles required. The OTK library required eight additional PCR cycles, which was within the range of 6-11 cycles recommended for maintaining library complexity [181] **(Figure 4.5)**. The size of the final OTK library was assessed using the Bioanalyzer system. DNA fragments were observed between 100bp to 1kb; the desirable fragment size for sequencing is less than 1kb and so the library was of sufficient quality to be sequenced **(Figure 4.6)** [181].
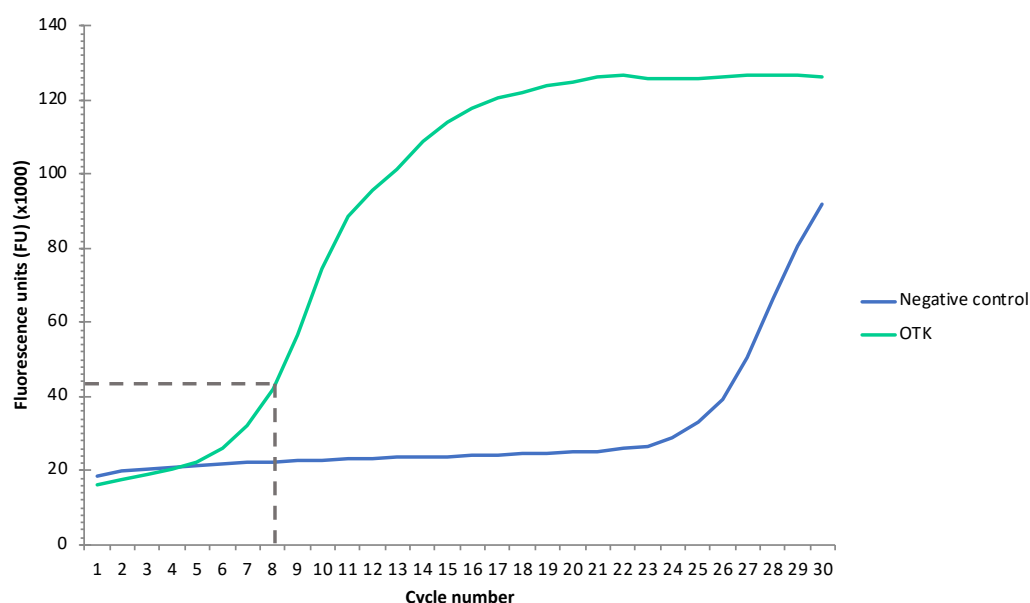
Figure 4.5. Enrichment of ATAC library by PCR assessed by qPCR.
QPCR was carried out using primers specific to the library adapter sequences. The SYBR fluorescence signal values were plotted against cycle number. To calculate the number of additional PCR cycles required: the baseline fluorescence value was taken away from the fluorescence saturation value; 25% of this value was calculated and the baseline value was added on (43,591 FU*).
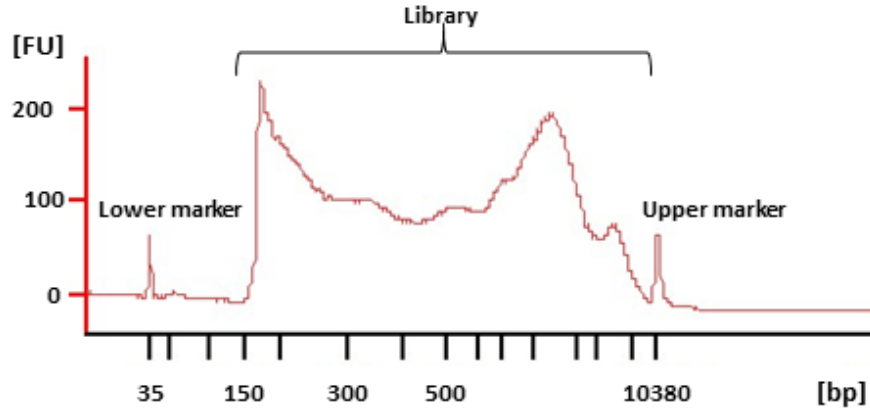
Figure 4.6. Electropherogram of OTK ATAC library using the Agilent Bioanalyzer system.
The DNA marker can be observed on the electropherogram as two peaks; a lower marker at 35bp and an upper marker at 10,380bp. To determine the average library size, gates were set surrounding the library peak. The ATAC-library size was between 100bp and 1kb.

4.6 Mapping of ATAC-seq peaks

The ATAC library was sequenced as described in Materials and Methods, Section 2.12. A total number of 37,953,387 reads were obtained and 47.3% of reads were aligned. The HOMER (Hypergeometric Optimization of Motif EnRichment) software was used to identify OTK ATAC-seq peaks by using the 'findpeaks' command and the 'variable peak' setting. The following parameters were applied: fragment length of 350bp, minimum distance of 400bp and FDR (false discovery rate) of 0.01%. A total of 60,329 peaks were called. An example of the sequenced ATAC-seq data at the *TP63* gene can be seen in **Figure 4.7**.

The called ATAC peaks were annotated by genomic feature using the HOMER 'AnnotatePeaks' function. The peaks were annotated into the following categories, intergenic, intron, promoter, exon, TTS and UTR (untranslated region) **(Figure 4.8)**. ATAC regions were distributed similarly between intergenic and intronic

regions (36.7% and 39% of regions, respectively) while 16.5% of regions overlapped with promoters, agreeing with the distribution of DHSs in HNSCC cells (**Figure 4.8 and Figure 3.11**). This pattern of distribution agrees with published literature [288, 289].
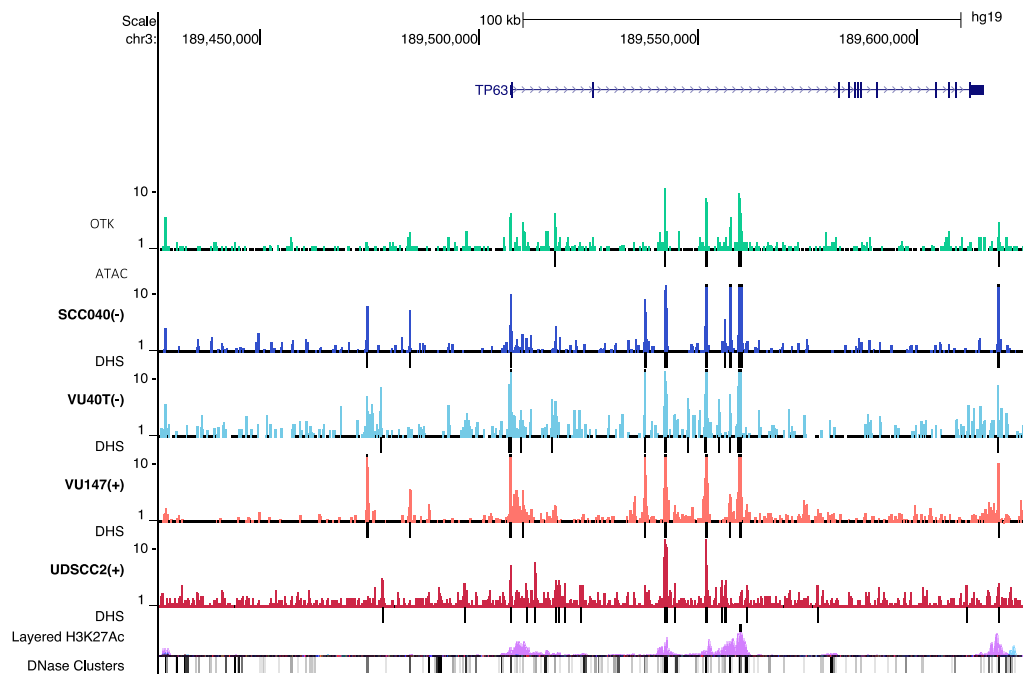
Figure 4.7. UCSC screenshot of ATAC-seq peaks at the *TP63* gene.
The ATAC-seq reads were aligned to the hg19 genome and the Bigwig file was uploaded to the UCSC (University of California Santa Cruz) genome browser, along with the four HNSCC DHS files. The ATAC-seq coverage track is displayed in green. OTK ATAC-seq peaks were called using HOMER and a total of 60,329 peaks were called (called peaks are depicted by black bars below the coverage track). The 'layered H3K27Ac' track displays previously published data; the active histone mark, H3K27Ac in a panel of 125 ENCODE (ENCyclopaedia Of DNA Elements) cell lines. The pink peaks in the H3K27Ac track represent the NHEK (normal human epidermal keratinocyte) cell line, indicating that some of these DHSs are epithelial specific. The 'DNase Clusters' track displays previously published data; mapped DHS reads in ENCODE cell lines.

181
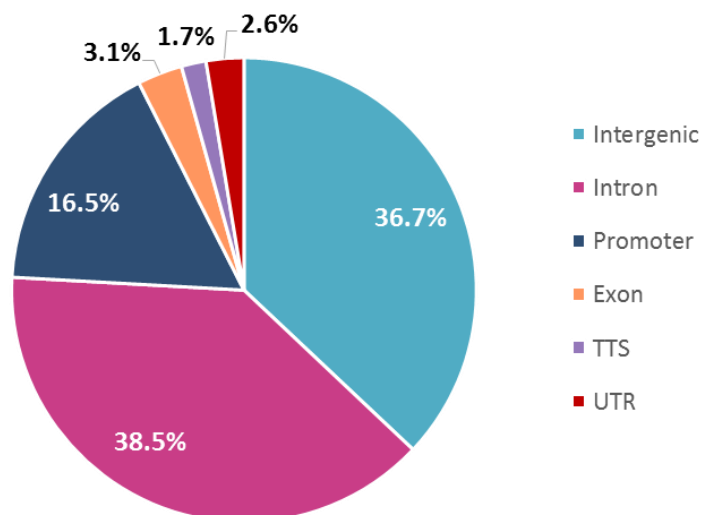
Figure 4.8. Genomic distribution of OTK ATAC-seq peaks.
A total of 60,329 OTK ATAC-seq peaks were annotated using the 'AnnotatePeaks' function in HOMER to the following regions: intergenic regions, introns, promoters, exons, transcription termination sites (TTS) and untranslated regions (UTR).

4.7 Peak location overlap of OTK and HNSCC cell line peaks

The locations of called peaks were used to investigate the overlap between open regions in non-tumour cells and HNSCC cells. The 'PCAexplorer' script within the R software was used to construct a heatmap based on peak location (**Figure 4.9A**). All HNSCC DHS peaks and all OTK ATAC peaks (merged HNSCC-OTK dataset) were considered. The largest variation between peak location was between the OTK and HNSCC cells, while SCC040(-) and VU147(+) cells were the most similar (**Figure 4.9A**).

Differences in peak locations of HNSCC and non-tumour cells was investigated further by principal components analysis (PCA) (**Figure 4.9B**). PCA analysis revealed two main components (PC1 and PC2). PC1 separated UDSCC2(+) peaks

182

from the HNSCC cell lines and OTK cells, while PC2 separated the HNSCC DHS

peaks and OTK ATAC peaks (**Figure 4.9B**). Therefore, peak location variability is

observed between the non-tumour and HNSCC cells.
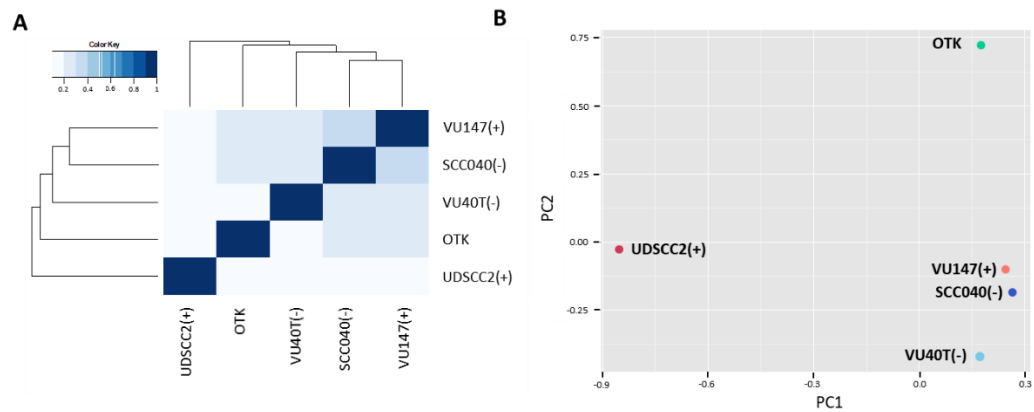
Figure 4.9. Differences in peak locations of ATAC-seq and DNase I-seq.
The 'PCAexplorer' script in the R software was used to analyse variability between ATAC
peak and HNSCC DHS peak locations. (**A**) A variation heatmap based on OTK and
HNSCC cell line peak location was constructed. Clusters were assigned a score between
0.1 and 1 based on variability (a score of 1 indicates no variability). (**B**) Principal
components plot based on peak location revealed two principal components; PC1 which
separated UDSCC2(+) and PC2 which separated OTK peaks.

4.8 Transcription factor motif analysis in OTK ATAC open regions

HOMER was used to conduct known TF motif analysis in order to determine which transcription factors (TFs) may be involved in the OTK cells. A total of 185 known TF motifs were identified and the top 20 motifs are shown in **Figure 4.10A**. Five groups of motifs were identified for the following TFs: AP-1, CTCF/BORIS, p53, KLF (Krüppel-like factor)/SP1 (specificity protein 1) and ETS (E twenty-six transcription factor) **(Figure 4.10A)**. The motif which appeared the most frequently in the list of top 20 motifs (11/20) was the AP-1 TF motif. In addition, the percentage of 'target sequences with motif' was investigated and it was found that between 3% and 36% of sequences harboured the AP-1 motif in OTK cells **(Figure 4.10A)**. The motifs for CTCF and its paralogue BORIS were found in, 8.2% and 12% of target sequences, respectively **(Figure 4.10A)**. Members of the p53 family which includes p63 and p73 were also enriched in OTK cells. The p63 and p53 motifs were present in 4.9% and 12.8% of target sequences with motif, respectively **(Figure 4.10A)**. Motifs for the KLF/SP1 TFs were also enriched. Motifs for KLF-4 and -5 were found in 38.7% and 14.1% of target sequences, respectively, while the SP-1 motif was enriched in 15% of sequences **(Figure 4.10A)**. The FLI1 (friend leukemia integration 1) and ELK1 (ETS domain-containing protein 1) TFs which are members of the ETS TF family were found in 28.7% and 18% of target sequences, respectively **(Figure 4.10A)**.

The OTK top 20 TF list and the HNSCC-merged motif list were compared **(Figure 4.10B)**. It was found that the majority of motifs were shared between the two datasets (16/20; 80%) and were enriched in OTK cells and at least one HNSCC cell

line. Only four TFs were specific to OTK cells, which included SP1, KLF4, FLI1 and ELK **(Figure 4.10B)**.

In addition to 'known motif' analysis, *de novo* motif analysis was conducted to predict similar motifs. The analysis was conducted in OTK cells and compared to the four HNSCC cell lines; a total of 32 motifs were predicted in OTK ATAC regions and the top 10 TF motifs were analysed (AP-1, ETS, KLF/SP-1, TEAD, p53/p63, NFY, RUNX, CTCF, CEBP (CCAAT/enhancer-binding protein) **(Figure 4.11A)**. Agreeing with known motif analysis, the AP-1 TF motif displayed the highest enrichment in target sequences in OTK, SCC040(-) and VU147(+) cells, in addition, AP-1 was the most statistically significant motif in these cell lines **(Figure 4.11A)**.

The overlap of *de novo* motifs was compared between OTK and HNSCC cells to highlight HNSCC-specific, OTK-specific and motifs shared between non-tumour and HNSCC cells **(Figure 4.11B)**. One OTK-specific TF was identified, EKLF (erythroid kruppel-like factor or KLF1), while enrichment of the CEBP (CCAAT/enhancer-binding protein) motif was shared between OTK and one other HNSCC cell line (UDSCC2(+)) **(Figure 4.11B)**. Six motifs were shared between all cell types which included AP-1, ETS, KLF/SP1, p53/p63 and NFY **(Figure 4.11B)**. The TEAD and RUNX motifs were shared between all cell types except UDSCC2(+) cells, while CTCF was enriched in OTK, SCC040(-) and VU147(+) cells **(Figure 4.11B)**. Eight TF motifs were HNSCC-specific and included FOX (Forkhead)-(*x*), AC-acceptor, GFY (golgi associated olfactory signaling regulator), IRX (Iroquois Homeobox Protein)/GRHL (grainyhead-like protein), NRF (nuclear respiratory factor)-1, LHX (LIM homeobox), NF (nuclear factor)-1 and CRE (cAMP response element) **(Figure 4.11B)**.

Name	log P-value	% Target Sequences with Motif	% Background Sequences with Motif	Fold-change	Motif logo
FOSL1	-1.65E+04	33.00	6.29	19.06	
ATF3	-1.61E+04	35.30	7.54	21.36	
BATF	-1.55E+04	34.40	7.40	21.51	
AP1	-1.45E+04	36.17	8.79	24.30	
FOSL2	-1.32E+04	24.18	3.95	16.34	
JUN	-1.14E+04	19.40	2.84	14.64	
BACH2	-5.56E+03	12.54	2.47	19.70	
CTCF	-2.38E+03	8.22	2.33	28.35	
P63	-2.30E+03	12.76	5.02	39.34	
P53	-1.93E+03	4.94	1.07	21.66	
BORIS	-1.89E+03	12.02	5.07	42.18	
NF-E2	-1.71E+03	3.69	0.67	18.16	
MAFK	-1.67E+03	8.05	2.87	35.65	
BACH1	-1.65E+03	3.37	0.57	16.91	
SP1	-1.45E+03	15.15	7.99	52.74	
NRF2	-1.42E+03	3.00	0.53	17.67	
KLF5	-1.27E+03	38.67	28.35	73.31	
KLF4	-1.25E+03	14.07	7.60	54.02	
FLI1	-1.18E+03	28.74	19.77	68.79	
ELK1	-9.94E+02	18.00	11.31	62.83	

Figure 4.10. The top 20 TF motifs predicted in OTK ATAC-seq peaks.
(**A**) HOMER known motif prediction was used to predict motifs in the OTK ATAC-peaks. A total of 185 motifs were predicted; the top 20 are shown in the table. (**B**) The top 20 OTK motifs were compared with those identified in the HNSCC-merged DHS dataset. In (**B**) green boxes indicate motifs predicted in both OTK and HNSCC peaks and white boxes indicate motifs predicted only in OTK peaks.
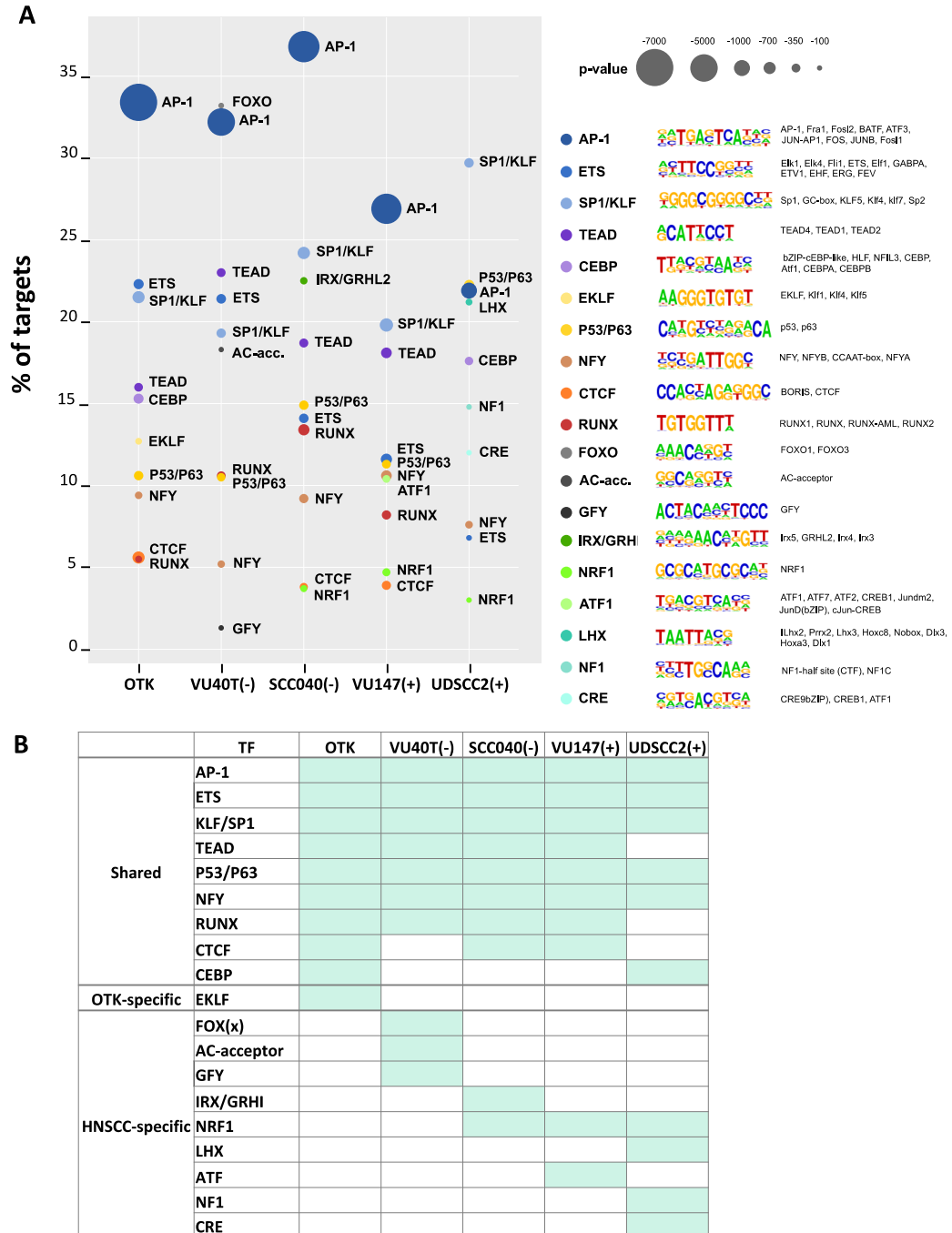
Figure 4.11. Top 10 *de novo* motifs in OTK cells and HNSCC cell lines.
(A) HOMER was used to predict *de novo* motifs using OTK ATAC peaks and HNSCC DHS peaks. A total of 32, 21, 23, 22 and 24 motifs were predicted for OTK, VU40T(-), SCC040(-), VU147(+) and UDSCC2(+) cells, respectively. For each motif, the percentage of motifs in target sequences is displayed and the size of the circles represent p-value. **(B)** The overlap of TF motifs between OTK and HNSCC was investigated. A total of 19 motifs were divided into three categories: shared (10/19; 53%), OTK-specific (1/19; 5%) and HNSCC-specific (8/19; 42%).

4.9 Pathway enrichment analysis of OTK ATAC open regions

HOMER was used to conduct pathway enrichment analysis by assigning each open chromatin region to the nearest gene TSS. A total of 277 pathways were identified and the top 20 pathways were investigated **(Appendix: Table 9)**. The top 20 pathways were combined with those top 20 identified in the four HNSCC cell lines **(Figure 3.16)** which yielded 55 different pathways **(Figure 4.12)**. As before, the pathways were ranked by p-value and assigned to the same 17 broad functions **(Figure 4.12)**.

Pathways	Function	OTK	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)	log-value	
Pathways in cancer	Cancer-related						-24.84	
EGFR1 Signaling Pathway	HNSCC-related						-19.81	
TGF-beta Signaling Pathway	HNSCC-related						-15.67	
ErbB signaling pathway	HNSCC-related						-13.14	
Focal adhesion	Cell adhesion						-7.73	
Integrated Pancreatic Cancer Pathway	Cancer-related						-5.26	
Endocytosis	Cytoskeletal						-0.51	
Alpha6-Beta4 Integrin Signaling Pathway	Cell adhesion/migration							
FoxO signaling pathway	Cell proliferation							
Chronic myeloid leukemia	Cancer-related							
Cell cycle	Cell proliferation							
Small cell lung cancer	Cancer-related							(1)
p53 signaling pathway	HNSCC-related							
Renal cell carcinoma	Cancer-related							
Regulation of actin cytoskeleton	Cytoskeletal							
Prostate Cancer	Cancer-related							
Insulin Signaling	Metabolism							
Integrated Breast Cancer Pathway	Cancer-related							
Adherens junction	Cell adhesion							
MAPK Signaling Pathway	HNSCC-related							
Leptin signaling pathway	Metabolism							
Signaling of Hepatocyte Growth Factor Receptor	Cell growth/proliferation							
DNA damage response (only ATM dependent)	DNA regulation							
Senescence and Autophagy	Senescence							(2)
Prolactin Signaling Pathway	Metabolism/immune function/development							
AMPK signaling pathway	Metabolism							
Hippo signaling pathway	Cell adhesion/migration/proliferation							
Fc gamma R-mediated phagocytosis	Immune function							
Rap1 signaling pathway	Cell migration							
Phosphatidylinositol signaling system	Cell signalling							
Inositol phosphate metabolism	Metabolism							
Inflammatory mediator regulation of TRP channels	CNS/Immune function							
Wnt Signaling Pathway	HNSCC-related							
Proteoglycans in cancer	Cell adhesion/migration/proliferation							(3a)
Huntington's disease	CNS							
Non-alcoholic fatty liver disease (NAFLD)	Metabolism							
Interferon type I	Immune function							
Circadian rhythm	CNS							
Axon guidance	CNS							
Ubiquitin mediated proteolysis	Protein/RNA processing							
Dopaminergic synapse	CNS							
Protein processing in endoplasmic reticulum	Protein/RNA processing							(3b)
Thyroid signalling hormone pathway	Development/cell growth							
TNF-alpha/NF-kB Signaling Pathway	Immune function							
RB in Cancer	HNSCC-related							
Androgen receptor signaling pathway	Cell growth/proliferation							(4)
mRNA processing	Protein/RNA processing							
Corticotropin-releasing hormone	CNS							
Viral carcinogenesis	Viral							
Epstein-Barr virus infection	Viral							
HTLV-I infection	Viral							
Spliceosome	Protein/RNA processing							(5)
Ribosome	Protein/RNA processing							
T cell receptor signaling pathway	Immune function							
Alcoholism	Metabolism							

Note: the column (3a) and (3b) groups are both within group (3).

Figure 4.12. Pathway enrichment analysis of genes assigned to HNSCC DHSs and ATAC peaks in OTK cells.

Analysis revealed 55 enriched pathways in OTK cells compared with four HNSCC cell lines. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into five groups by cell line representation: **(1)** over-represented in all (22/55; 40%), **(2)** under-represented in UDSCC2(+) (4/55; 7%), **(3)** over-represented in OTK (17/55; 31%), **(3a)** under-represented in UDSCC2(+) (15/55; 27%), **(3b)** over-represented in UDSCC2(+) (2/55; 4%), **(4)** over-represented in UDSCC2(+) (4/55; 7%), **(5)** over-represented in HNSCC only (7/55; 13%).

189

When grouped by cell type representation, five groups of pathways were identified. Group (1) contained pathways over-represented equally in OTK and HNSCC cells and contained the most pathways (22/55) (**Figure 4.12**). Many of the pathways were HNSCC-related (5/22; EGFR, ERBB, TGF (transforming growth factor)-β, p53 and MAPK signalling pathways), cancer-related (7/22) or cell adhesion pathways (3/22). Among the most statistically significant pathways in this group were the EGFR, ERBB and 'pathways in cancer' (**Figure 4.12**). Group (2) contained four pathways all of which were under-represented in UDSCC2(+) cells and were largely over-represented in the remaining cell types. The majority of pathways in this group (3/4) were involved in metabolism (**Figure 4.12**). Group (3) contained 17 pathways under-represented in HNSCC cells, while being strongly over-represented in OTK cells. This group could be divided into two subsets: 3a and 3b. Subset 3a included 15 pathways which were strongly under-represented in UDSCC2(+) cells, while subset 3b included two pathways which were over-represented in UDSCC2(+) (protein processing in endoplasmic reticulum and thyroid signalling hormone pathways) (**Figure 4.12**). Pathways in subset 3a included the five pathways related to CNS: inflammatory mediator regulation of TRP (transient receptor potential) channels, Huntington's disease, circadian rhythm, axon guidance and dopaminergic synapse. In addition, some of the most over-represented pathways in OTK cells was the Hippo pathway, RAP1 (Rap1 GTPase) and phosphatidylinositol signalling pathways. Interestingly, the HNSCC-related pathway WNT (wingless) was also over-represented in OTK cells (**Figure 4.12**). Group (4) consisted of five pathways which were under-represented in OTK cells and over-represented in UDSCC2(+) cells and in at least one other HNSCC cell line. Pathways in this group included the pRB (retinoblastoma protein) in

cancer, androgen receptor signalling and mRNA processing pathways **(Figure 4.12)**. Group (5) consisted of seven pathways which were not represented in OTK cells at all and strongly over-represented in HNSCC cells. Three pathways in this group were associated with viral infection (viral carcinogenesis, Epstein-Barr virus infection, HTLV (Human T-cell leukemia virus)-I pathways) **(Figure 4.12)**. There were no pathways which were strictly OTK-specific and absent in HNSCC cells.

In summary, the majority of HNSCC-related, cancer-related and cell adhesion pathways appear to be involved in OTK cell function to the same extent as in HNSCC cells. Having said this, a number of pathways were clearly more involved in OTK cells than in HNSCC cell lines as shown in Group (3a), which included pathways such as Hippo, CNS (central nervous system)-related pathways, phosphatidylinositol signalling system and WNT pathways **(Figure 4.12)**.

4.10 Enhancer cluster analysis of OTK ATAC open regions

To further investigate ATAC open regions in OTK cells, the HOMER software was used to conduct enhancer cluster (EC) analysis, which revealed 592 ECs in OTK ATAC peaks. Pathway enrichment analysis was conducted on the ATAC EC-associated genes and to highlight differences, the top 20 pathways were combined with the top 20 HNSCC EC (enhancer cluster) pathways to give 59 pathways **(Appendix: Tables 5-8 and 10, Figure 4.13)**. Pathways were grouped by cell representation which revealed six groups. Group (1) consisted of three pathways (focal adhesion, α6β4 integrin and Fc gamma R-mediated phagocytosis pathways) which were strongly over-represented in all cell types **(Figure 4.13)**. Group (2)

consisted of six pathways which were over-represented more significantly in OTK cells than in HNSCC. This group included the phosphatidylinositol signalling system, inositol phosphate, dopaminergic synapse and two pathways related to protein/RNA processing (protein processing in endoplasmic reticulum and ubiquitin mediated proteolysis) **(Figure 4.13)**. Group (3) consisted of 13 pathways which were largely under-represented in UDSCC2(+) cells and over-represented in the remaining HNSCC cell lines and OTK cells **(Figure 4.13)**. Nearly one third of the pathways are associated with cell adhesion. Notably, the pathways in cancer and Hippo signalling pathways were significantly over-represented in OTK cells and the EGFR pathway was significantly over-represented in the OTK, SCC040(-) and VU147(+) cells **(Figure 4.13)**. Group (4) consisted of 16 pathways which were consistently over-represented in OTK cells and over-represented in at least one HNSCC cell line, though largely under-represented in SCC040(-) cells **(Figure 4.13)**. Group (4) pathways can be summarised as HNSCC-related (RAS, WNT, ERBB and MAPK signalling pathways), CNS associated (Huntington's, signalling pathways in glioblastoma and inflammatory mediator regulation of TRP channels) and immune function-related (IL (interleukin)-3, IL-6 and RANKL/RANK (tumour necrosis factor receptor superfamily member 11a) signalling) **(Figure 4.13)**. Group (5) consisted of seven pathways which were consistently under-represented in OTK cells and were largely under-represented in HNSCC cells. This group included two pathways associated with DNA regulation (DNA replication and mismatch repair) and the TGF-β pathway **(Figure 4.13)**. Group (6) consisted of 14 pathways which were HNSCC-specific. This group contained pathways associated with DNA regulation (nucleotide excision repair, DNA polymerase α/primase complex and RPA (replication protein A) complex) and immune function (NF-κB signalling,

systemic lupus erythematosus and leukocyte transendothelial migration) **(Figure 4.13)**. In addition, the transcriptional misregulation pathway, and two viral pathways (viral carcinogenesis and HTLV-I infection) were also present in this group **(Figure 4.13)**.

Pathway	Function	Group	log p-value
Focal adhesion	Cell adhesion	(1)	-24.14
Fc gamma R-mediated phagocytosis	Immune function	(1)	-15.91
Alpha6-Beta4 Integrin Signaling Pathway	Cell adhesion/migration	(1)	-12.05
Thyroid signaling hormone pathway	Development/cell growth	(2)	-11.04
Phosphatidylinositol signaling system	Cell signalling	(2)	-9.69
Protein processing in endoplasmic reticulum	Protein/RNA processing	(2)	-7.24
Dopaminergic synapse	CNS	(2)	-2.60
Inositol phosphate metabolism	Metabolism	(2)	-0.76
Ubiquitin mediated proteolysis	Protein/RNA processing	(2)	
Proteoglycans in cancer	Cell adhesion/migration/proliferation	(3)	
Apoptosis	Cell death	(3)	
Pathways in cancer	Cancer-related	(3)	
Hippo signaling pathway	Cell adhesion/migration/proliferation	(3)	
EGFR1 Signaling Pathway	HNSCC-related	(3)	
Adherens junction	Cell adhesion	(3)	
Pancreatic cancer	Cancer-related	(3)	
Androgen receptor signaling pathway	Cell growth/proliferation	(3)	
Integrated Pancreatic Cancer Pathway	Cancer-related	(3)	
Integrin-mediated cell adhesion	Cell adhesion	(3)	
Integrated Breast Cancer Pathway	Cancer-related	(3)	
Regulation of actin cytoskeleton	Cytoskeletal	(3)	
Rap1 signaling pathway	Cell migration	(3)	
Huntington's disease	CNS	(4)	
Ras signaling pathway	HNSCC-related	(4)	
Small cell lung cancer	Cancer-related	(4)	
Wnt signaling pathway	HNSCC-related	(4)	
Cell cycle	Cell proliferation	(4)	
ErbB signaling pathway	HNSCC-related	(4)	
IL-3 Signaling Pathway	Immune function	(4)	
Endocytosis	Cytoskeletal	(4)	
MAPK signaling pathway	HNSCC-related	(4)	
FAS pathway and Stress induction of HSP regulation	Cell death	(4)	
Hypertrophy Model	Protein processing	(4)	
IL-6 Signaling Pathway	Immune function	(4)	
Inflammatory mediator regulation of TRP channels	CNS/Immune function	(4)	
RANKL/RANK Signaling Pathway	Immune function	(4)	
Signaling of Hepatocyte Growth Factor Receptor	Cell growth/proliferation	(4)	
Signaling Pathways in Glioblastoma	CNS	(4)	
TGF-beta signaling pathway	HNSCC-related	(5)	
DNA replication	DNA regulation	(5)	
IL-1 Signaling Pathway	Immune function	(5)	
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Cell adhesion	(5)	
Pathogenic Escherichia coli infection	Bacterial	(5)	
Mismatch repair	DNA regulation	(5)	
Estrogen signaling pathway	Cell growth/proliferation	(5)	
Viral carcinogenesis	Viral	(6)	
Epithelial cell signaling in Helicobacter pylori infection	Bacterial	(6)	
NF-kappa B signaling pathway	Immune function	(6)	
Bacterial invasion of epithelial cells	Bacterial	(6)	
Leukocyte transendothelial migration	Immune function/cell migration	(6)	
Nucleotide excision repair	DNA regulation	(6)	
DNA polymerase alpha / primase complex	DNA regulation	(6)	
Systemic lupus erythematosus	Immune function	(6)	
RPA complex	DNA regulation	(6)	
Alcoholism	Metabolism	(6)	
Endometrial cancer	Cancer-related	(6)	
Tight junction	Cell adhesion	(6)	
Transcriptional misregulation in cancer	Transcription regulation	(6)	
HTLV-I infection	Viral	(6)	

Cell-line columns: OTK, SCC040(-), VU40T(-), VU147(+), UDSCC2(+).

Figure 4.13. Pathway enrichment analysis of genes assigned to enhancer clusters in OTK cells and HNSCC cell lines.

Analysis revealed 59 different pathways across the OTK and four HNSCC cell lines. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into 6 groups by cell representation: (**1**) over-represented in all (3/59; 5%), (**2**) more over-represented in OTK (6/59; 10%), (**3**) under-represented in UDSCC2(+) (13/59; 22%), (**4**) consistently over-represented in OTK (16/59; 12%), (**5**) more over-represented in HNSCC (7/59; 12%), (**6**) exclusively over-represented in HNSCC (14/59; 24%).

194

To investigate similarities between EC-related genes in OTK and HNSCC cells, the overlap between the two datasets was analysed **(Figure 4.14)**. A total of 592 OTK EC genes were annotated and 72% of these genes were OTK-specific (427/592), while 28% (165/592) of genes were shared with HNSCC cells **(Figure 4.14)**.

The ConsensusPath database was used to conduct pathway enrichment analysis on each set of genes: OTK-specific, HNSCC-specific and shared between OTK and HNSCC (shared), which revealed 16, 80 and 43 pathways, respectively. All 16 pathways for OTK-specific genes and the top 20 pathways for the remaining two gene datasets were taken for further analysis, which gave a total of 47 different pathways **(Figure 4.15)**. The pathways were ranked by p-value and grouped by cell representation, which formed six groups and were categorised into 19 broad functions: bacterial, bone development, cancer-related, cell adhesion, death, differentiation, growth, proliferation, migration, signalling, cytoskeletal, HNSCC-related, immune function, kidney degeneration, metabolism, micro RNA (miRNA), senescence, transcription and viral **(Figure 4.15)**.
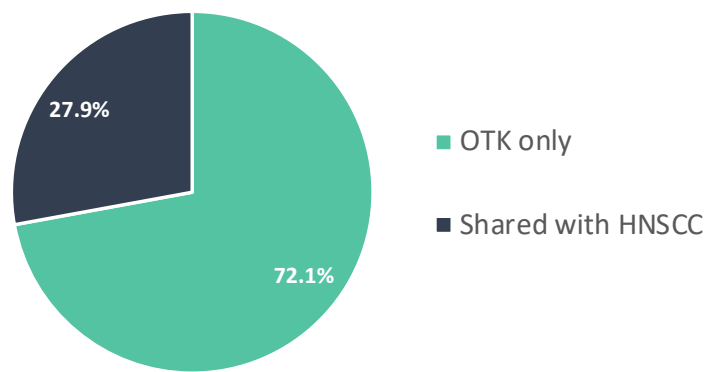
Figure 4.14. Percentage overlap of EC genes between OTK and HNSCC cells.
HOMER was used to assign 592 genes to each OTK EC. The EC genes were compared with those assigned to HNSCC ECs. The percentage of OTK-specific EC genes and those shared with HNSCC EC genes was calculated.

Pathway	Function	OTK-specific	HNSCC-specific	OTK/HNSCC	Group	log p-value
VEGFA-VEGFR2 Signaling Pathway	Cell growth	■	■	■	**(1)**	2.4E-06
EMT transition in Colorectal Cancer	Cancer-related	■	■	■		8.5E-06
Focal Adhesion	Cell adhesion	■	■	■		1.6E-05
Bacterial invasion of epithelial cells	Bacterial	■	■			3.4E-05
EGF-EGFR Signaling Pathway	Cell growth	■	■		**(2)**	5.0E-05
Androgen receptor signaling pathway	Cell growth	■	■			2.0E-04
Alpha 6 Beta 4 signaling pathway	Cell adhesion	■	■			9.4E-04
Rap1 signaling pathway	Cell adhesion/migration	■	■			2.2E-03
Insulin Signaling	Metabolism	■	■			3.1E-03
Nuclear Receptors Meta-Pathway	Transcription		■	■		
Primary Focal Segmental Glomerulosclerosis FSGS	Kidney degeneration		■	■		
Small cell lung cancer	Cancer-related		■	■		
TGF-beta Signaling Pathway	HNSCC-related		■	■	**(3)**	
miR-targeted genes in epithelium	miRNA		■	■		
miR-targeted genes in muscle cell	miRNA		■	■		
Regulation of actin cytoskeleton	Cytoskeletal		■	■		
JAK-STAT	Cell proliferation/differentiation/migration/death		■	■		
EPO Receptor Signaling	Cell growth/proliferation	■				
Notch Signaling Pathway	HNSCC-related	■				
RORA activates gene expression	Transcription	■				
Kit receptor signaling pathway	Cell growth/proliferation	■			**(4)**	
NR1D1 (REV-ERBA) represses gene expression	Transcription	■				
Vitamin D Receptor Pathway	Transcription	■				
HIF1A and PPARG regulation of glycolysis	Metabolism	■				
Cellular senescence	Senescence		■			
Fc gamma R-mediated phagocytosis	Immune function		■			
Hair Follicle Development	Cell growth		■			
Hippo signaling pathway	Cell adhesion/migration/proliferation		■			
Leukocyte transendothelial migration	Cell migration		■			
miR-targeted genes in lymphocytes	miRNA		■		**(5)**	
Osteoclast differentiation	Bone development		■			
Pathways in cancer	Cancer-related		■			
Proteoglycans in cancer	Cell adhesion		■			
RANKL-RANK (Receptor activator of NFKB (ligand)) Signaling Pathway	Immune function		■			
Viral carcinogenesis	Viral		■			
Adherens junction	Cell adhesion			■		
Arrhythmogenic Right Ventricular Cardiomyopathy	Cell adhesion			■		
Cellular response to heat stress	Cell growth/death			■	**(6a)**	
Glucocorticoid Receptor Pathway	Cell signalling			■		
Integrin-mediated Cell Adhesion	Cell adhesion			■		**(6)**
Adipogenesis	Metabolism			■		
Endometrial cancer	Cancer-related			■		
Gastric cancer	Cancer-related			■		
Hfe effect on hepcidin production	Metabolism			■	**(6b)**	
IL-7 Signaling Pathway	Immune function			■		
miRNA Regulation of DNA Damage Response	miRNA			■		
Regulation of Apoptosis by Parathyroid Hormone-related Protein	Cell death			■		

Figure 4.15. Pathway enrichment analysis of enhancer cluster genes.

HOMER was used to assign genes to enhancer clusters. The overlap of EC genes between HNSCC and OTK was investigated: 427 genes were OTK-specific, 1,044 genes were HNSCC-specific and 165 genes were shared between OTK and HNSCC cells. Pathway enrichment analysis was conducted using the ConsensusPath tool on each gene set. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped in six groups: **(1)** over-represented in all (3/47; 6%), **(2)** over-represented in OTK- and HNSCC-specific genes (6/47; 13%), **(3)** HNSCC-specific and shared (8/47; 17%), **(4)** OTK-specific (7/47; 15%), **(5)** HNSCC-specific (11/47; 23%), **(6)** shared (12/47; 26%), **(6a)** highly over-represented in shared (5/47; 11%), **(6b)** over-represented in shared (7/47; 15%).

Group (1) consisted of three pathways which were over-represented in all three datasets (VEGF (vascular endothelial growth factor) signalling pathway, EMT (epithelial to mesenchymal) transition in colorectal cancer and focal adhesion pathways) (**Figure 4.15**). Group (2) consisted of six pathways which were over-represented in OTK-specific and HNSCC-specific genes. These pathways included two cell growth associated pathways (EGFR and androgen receptor signalling), which were over-represented more significantly in HNSCC-specific genes. A pathway involved in bacterial infection (bacterial invasion of epithelial cells pathway) was significantly over-represented in both the datasets (**Figure 4.15**). The remaining three pathways ($\alpha4\beta6$, RAP1 and insulin signalling) were also over-represented in both gene sets. Group (3) consisted of eight pathways, which were over-represented in HNSCC-specific and shared gene sets: the HNSCC- and cancer-related pathways (TGF-β and small cell lung cancer, respectively) were more highly over-represented in HNSCC-specific pathways, as were the miRNA-related pathways (miR-targeted genes in epithelium and muscle cell) (**Figure 4.15**). Group (4) consisted of seven OTK gene-specific pathways, which were largely related to cell growth/proliferation and transcription. The transcription-associated pathways included the 'RORA (retinoic acid related orphan receptor A) activates gene expression' pathway, the 'NR1D1 (REV-ERBA) represses gene expression' (NR1D1-nuclear receptor subfamily 1 group D member 1) pathway and the vitamin D receptor pathway (**Figure 4.15**). The cell growth/proliferation pathways included the EPO (erythropoietin) receptor signalling pathway and the kit receptor signalling pathway. Notably, this group also included the NOTCH (neurogenic locus notch homolog) pathway as being over-represented specifically in OTK cells (**Figure 4.15**). Group (5) consisted of 11 pathways which were strongly HNSCC-specific

(**Figure 4.15**). Distinctly from OTK-specific pathways, group (5) included pathways related to immune function (FC gamma R-mediated phagocytosis pathway), senescence (cellular senescence pathway), cell adhesion/migration/proliferation (Hippo pathway), cell migration (leukocyte transendothelial migration pathway) and the viral carcinogenesis pathway (**Figure 4.15**). Group (6) consisted of 12 pathways which were over-represented in the 'shared' gene set. Based on statistical significance, this group could be divided into two subsets: 6a and 6b (**Figure 4.15**). Subset 6a included five pathways which displayed higher statistical significance than those in subset 6b and included three pathways involved in cell adhesion (adherens junction, arrhythmogenic right ventricular cardiomyopathy and integrin-mediated cell adhesion pathways) (**Figure 4.15**). The remaining pathways in subset 6a included the cellular response to heat stress and glucocorticoid receptor pathways. Subset 6b included two pathways related to metabolism: 'HFE effect on hepcidin production pathway' (HFE-homeostatic iron regulator), which is involved in iron metabolism and the adipogenesis pathway (**Figure 4.15**). Interestingly, no HNSCC-related pathways were over-represented in this group, but the cancer-related endometrial and gastric pathways were over-represented in subset 6b (**Figure 4.15**).

The EC genes were further analysed to investigate genes encoding for TFs in OTK cells and compare with the TF genes annotated in HNSCC cells. A total of 25 genes were identified that could be divided into nine families: p53, AP-1, ETS, SP1, KLF, RUNX, GRHL, LHX and DLX (distal-less homeobox) (**Table 4.1**). The p53 gene family includes *TP53*, *TP63* and *TP73*. While *TP53* was not annotated as an EC gene, *TP63* was HNSCC-specific and *TP73* was annotated in UDSCC2(+) and OTK cells (**Table 4.1**). Other HNSCC-specific EC genes included the AP-1 genes

(*FOSL1* (FOS like 1, AP-1 transcription factor subunit) and *ATF3*), ETS genes (*ELF3*, *ETV-3* (ETS variant 3), *-5*, *-6* and *EHF* (ETS homologous factor), *SP1*, KLF genes (*KLF-7*, *-13*), *RUNX1*, *GRHL2*, *LHX-2*, *-3* and *DLX-3*, *-6* (**Table 4.1**). OTK-specific EC genes included *ETS1* and *KLF9* (**Table 4.1**). EC genes shared between OTK and HNSCC cells included, *FOS*, *FOSB*, *ETS2* and *KLF-6*, *-10*.

Table 4.1. TF genes assigned to enhancer clusters in HNSCC and OTK cells.

TF family	EC gene	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)	OTK
p53	TP63	■	■	■		
	TP73				■	■
AP-1	FOS				■	■
	FOSB			■	■	■
	FOSL1	■				
	ATF3			■	■	
ETS	ETS2		■	■		■
	ELF3	■			■	
	ETV3				■	
	ETS1					■
	ETV5		■			
	ETV6		■			
	EHF			■		
SP1/KLF	SP1	■	■			
	KLF6		■			■
	KLF7		■			
	KLF9					■
	KLF10		■			■
	KLF13			■		
RUNX	RUNX1			■		
GRHL2	GRHL2	■				
LHX	LHX2				■	
	LHX3				■	
DLX	DLX3			■	■	
	DLX6				■	

4.11 Genome-wide gene expression analysis (RNA-seq) in non-tumour keratinocyte cells

Genome-wide RNA-seq was conducted to elucidate differences between HNSCC gene expression and non-tumourigenic primary cells to put the ATAC-seq data into context. RNA-seq was unsuccessful in OTK cells, however data was obtained for another primary non-tumour oral cell type- HOK (human oral keratinocyte). HOK RNA-seq analysis was conducted as described for HNSCC cell lines (Chapter 3, Section 3.12). The Galaxy platform was used to analyse the data as described in Materials and Methods, Section 2.13.5. Within Galaxy, the Hisat2 tool was used to align reads to the hg19 genome. A total of 26,281,754 reads were sequenced and 92.3% of reads were aligned to the genome. Only genes with an FPKM (fragments per kilobase exon per million mapped reads) value above 1 were considered for downstream analysis; a total of 12,228 genes were analysed.

The 'PCAexplorer' script in the R software was used to investigate differences in HNSCC gene expression compared to that in HOK cells. Gene FPKM values were used to conduct the PCA analysis (**Figure 4.16**). Four principal components were identified (PC1-4); over 40% of variability could be accounted for by PC1 (**Figure 4.16A**). It was found that PC1 accounted for differences in HNSCC gene expression compared to the HOK cells (**Figure 4.16B**). In addition, hierarchical clustering of FPKM values for the five cell types further highlighted gene expression differences in HOK cells compared to HNSCC (**Figure 4.16C**).
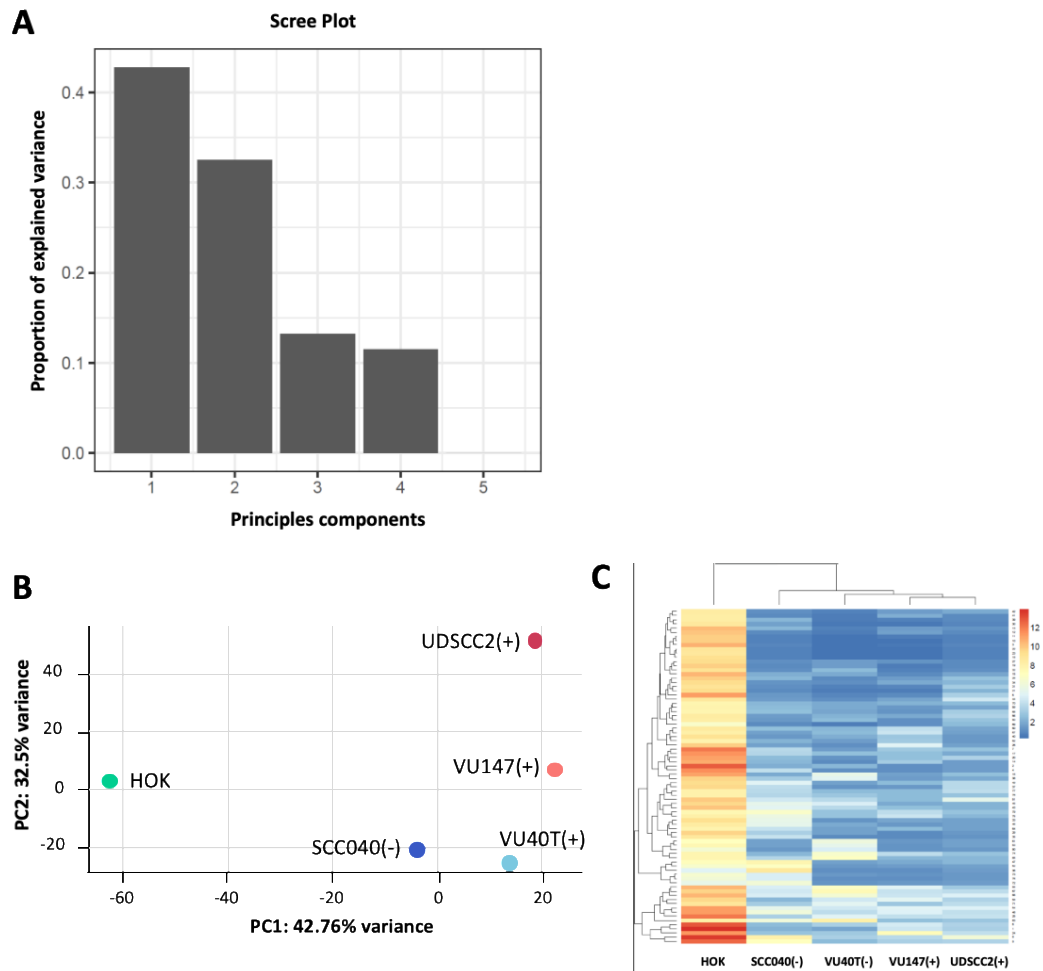
Figure 4.16. Principal components (PCA) and clustering analyses of gene expression data.

(A) Four principal components (PCs) were identified upon clustering of gene FPKM values: PC1 (42.76%), PC2 (32.5%), PC3 (13.2%) and PC4 (11.5%). **(B)** PC1 accounted for 42.76% of variability due to differences between HNSCC gene expression and HOK expression. PC2 accounted for 32.5% of variability due to differences between UDSCC2(+) gene expression compared to the remaining HNSCC cell lines and HOK cells. (PCA plots for PC3 and PC4 are shown in Appendix: Figure 2). **(C)** Clustering analysis of FPKM values revealed greatest variability to be between HOK gene expression and HNSCC expression. Analysis was conducted using the 'PCAexplorer' script within the R software.

To further investigate differences in gene expression, pathway enrichment analysis was conducted on the top 2,000 genes ranked by FPKM values for the HOK cells and four HNSCC cell lines. The analysis was conducted using the ConsensusPath database and data for KEGG (Kyoto Encyclopedia of Genes and Genomes) and Wikipathways were extracted. As before, the pathways were ranked by p-value and grouped by cell representation, which resulted in two main groups (**Figure 4.17**). Group (1) consisted of 26 pathways which were significantly over-represented in HOK cells. This group could be further divided into three subsets: 1a, 1b and 1c (**Figure 4.17**). Subset 1a contained four pathways which were significantly over-represented in HOK and nearly all HNSCC cells (**Figure 4.17**). Pathways in this subset exclusively contained pathways related to protein/RNA processing, including mRNA processing, translation factors, cytoplasmic ribosomal proteins and ribosome pathways (**Figure 4.17**). Subset 1b contained 19 pathways which were more significantly over-represented in HOK cells compared to HNSCC cells and were largely associated with miRNA (miR-targeted genes in leukocytes, muscle cell, lymphocytes and epithelium pathways), protein/RNA processing (RNA transport, protein processing in endoplasmic reticulum and spliceosome pathways) and CNS (Parkinson's, Huntington's and Alzheimer disease). Subset 1c consisted of four HOK-specific pathways which included pathways associated with miRNA in squamous cells and proteasome degradation and two pathways were associated with metabolism (non-alcoholic fatty liver disease and thermogenesis pathways) (**Figure 4.17**). Group (2) contained 14 pathways which were less represented in the five cell types. This group could be divided into two subsets: 2a and 2b. Subset 2a consisted of five pathways which showed comparable statistical significance in HOK cells and in at least one HNSCC cell line (**Figure 4.17**).

203

Pathways in this subset were associated with cell proliferation (cell cycle and androgen receptor signalling pathway) and adhesion (adherens and proteoglycans in cancer). In addition, the HNSCC-related pathway TGF-β was present in this group, however, the pathway was not enriched in VU40T(-) cells **(Figure 4.17)**. Subset 2b consisted of nine pathways which were more over-represented in HNSCC than in HOK cells. This subset contained cell adhesion related pathways (tight junction, ARVC (arrhythmogenic right ventricular cardiomyopathy), dilated cardiomyopathy), pRB in cancer pathway, pathways in cancer and three pathways related to development (thyroid signalling hormone, oocyte meiosis and progesterone-mediated oocyte maturation pathways) **(Figure 4.17)**.

Pathway	Function	HOK	SCC040(-)	VU40T(-)	VU147(+)	UDSCC2(+)	Group	Group
mRNA Processing	Protein/RNA processing						(1a)	(1)
Translation Factors	Protein/RNA processing						(1a)	(1)
Cytoplasmic Ribosomal Proteins	Protein/RNA processing						(1a)	(1)
Ribosome	Protein/RNA processing						(1a)	(1)
RNA transport	Protein/RNA processing						(1b)	(1)
Protein processing in endoplasmic reticulum	Protein/RNA processing						(1b)	(1)
miR-targeted genes in epithelium	miRNA						(1b)	(1)
Focal adhesion	Cell adhesion						(1b)	(1)
Regulation of actin cytoskeleton	Cytoskeletal						(1b)	(1)
Estrogen signaling pathway	Cell growth/proliferation						(1b)	(1)
Spliceosome	Protein/RNA processing						(1b)	(1)
VEGFA-VEGFR2 Signaling Pathway	Cell growth/proliferation						(1b)	(1)
Epstein-Barr virus infection	Viral						(1b)	(1)
Electron Transport Chain	Metabolism						(1b)	(1)
Pathogenic Escherichia coli infection	Bacterial						(1b)	(1)
Primary Focal Segmental Glomerulosclerosis FSGS	Renal degeneration						(1b)	(1)
Parkinson's disease	CNS						(1b)	(1)
miR-targeted genes in leukocytes	miRNA						(1b)	(1)
miR-targeted genes in muscle cell	miRNA						(1b)	(1)
miR-targeted genes in lymphocytes	miRNA						(1b)	(1)
Alzheimer	CNS						(1b)	(1)
Huntington's disease	CNS						(1b)	(1)
Oxidative phosphorylation	Metabolism						(1b)	(1)
Non-alcoholic fatty liver disease (NAFLD)	Metabolism						(1c)	(1)
Thermogenesis	Metabolism						(1c)	(1)
miR-targeted genes in squamous cell	miRNA						(1c)	(1)
Proteasome Degradation	Protein/RNA processing						(1c)	(1)
Cell cycle	Cell proliferation						(2a)	(2)
Androgen receptor signaling pathway	Cell growth/proliferation						(2a)	(2)
Adherens junction	Cell adhesion						(2a)	(2)
Proteoglycans in cancer	Cell adhesion						(2a)	(2)
TGF-beta Signaling Pathway	HNSCC-related						(2a)	(2)
Thyroid signaling hormone pathway	Development/cell growth						(2b)	(2)
Tight junction	Cell adhesion						(2b)	(2)
ATF6-alpha activates chaperone genes	Protein/RNA processing						(2b)	(2)
Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)	Cell adhesion						(2b)	(2)
Dilated cardiomyopathy	Cell adhesion						(2b)	(2)
Oocyte meiosis	Development						(2b)	(2)
Retinoblastoma (RB) in Cancer	HNSCC-related						(2b)	(2)
Pathways in cancer	Cancer-related						(2b)	(2)
Progesterone-mediated oocyte maturation	Development						(2b)	(2)

log -value legend: 5.6E-05, 1.2E-04, 6.0E-04, 2.0E-03, 3.0E-03, 7.2E-03

Figure 4.17. Gene expression analysis of HOK and HNSCC cells by RNA-seq.
Genome-wide RNA-seq was conducted in non-tumourigenic primary HOK cells and in four HNSCC cell lines. The genes were ranked by FPKM (fragments per kilobase exon per million mapped reads) value and the top 2000 genes were taken for pathway enrichment analysis by using the Consensus DB tool. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped by cell representation into two groups and five subgroups: (1) significantly over-represented in HOK (27/41; 66%), (1a) over-represented in all (4/41; 10%), (1b) over-represented in HOK (19/41; 12%), (2) under-represented in all (14/41; 34%) (2a) over-represented in all (5/41; 12%), (2b) over-represented in HNSCC (9/41; 22%).

4.12 Discussion

To investigate regulatory elements in non-tumourigenic oral keratinocytes, ATAC (Assay for Transposase-Accessible Chromatin)-seq was conducted as an alternative to the DNase I-hypersensitivity assay. The genomic distribution of ATAC-seq peaks displays high similarity to HNSCC (head and neck squamous cell carcinoma) DHS (DNase I hypersensitive site) peaks, suggesting that locations of open regions in both techniques are comparable as has been suggested previously [161, 188, 238, 239].

The transcription factor (TF) motif analyses of ATAC peaks revealed the EKLF (erythroid Krüppel-like factor) (KLF1) and KLF4 motifs to be OTK (oral tonsil keratinocyte)-specific. In addition, the OTK-specific EC-regulated TF genes included *KLF9*. The 17-member KLF TF family are involved in many different processes. For example, erythropoiesis (KLF-1, -2, -3 and -11), adipogenesis (KLF-2, -3 and -7), cardiac homeostasis (KLF-5, -6, -10, -13, -15 and -16) and epithelial homeostasis (KLF-4, -5, -9 and -13) [290].

KLF4 is known to be a negative cell cycle regulator and in line with this, a decrease in KLF4 protein expression in oral SCC samples has been reported [260, 291]. KLF4 normally binds the cyclin-dependent kinase inhibitor gene (*CDKN1*) and the cyclin B1 gene (*CCNB1*) to regulate the cell cycle [13, 16, 292]. In addition, upon DNA damage KLF4 can block the *CCNE1* gene which encodes cyclin E, thereby activating cell cycle checkpoints and preventing cell cycle progression [293]. Downregulation of KLF9 has been observed in oesophageal SCC compared to the surrounding normal tissue and KLF9 plays a role in the inhibition of growth, migration and metastasis of oesophageal SCC cells [294].

Therefore, the absence of the KLF4 and KLF9 motifs from the top 20 HNSCC motifs and their presence in the OTK cells may indicate suppression of KLF4- and KLF9-mediated cell cycle processes in these HNSCC cell lines. This decrease was also reflected by our RNA-seq data when comparing *KLF4* and *KLF9* expression in HOK (human oral keratinocyte) cells and HNSCC cells (data not shown).

KLF1 is central to erythropoiesis and plays a role in gene regulation, haem biosynthesis and cell cycle regulation [295]. Despite the role of KLF1 being largely restricted to erythroid tissues, recently, increased KLF1 levels in cervical cancer cells have been reported and has been linked to cell proliferation, metastasis and invasion via interaction with the PI3K (phosphotidylinositol-3 kinase)/AKT (protein kinase B) pathway [296, 297]. This evidence indicates KLF1 as having a pro-tumourigenic role in epithelial cells. Our RNA-seq data however demonstrates higher *KLF1* expression in HOK cells compared to HNSCC cells. However, in our HNSCC cell lines, the KLF1 motif and transcript levels are not detected and so it may be that *KLF1* overexpression does not contribute to HNSCC in these cells. This difference could be attributed to the fact that HOK cells are of foetal origin. Having said this, even during development, KLF1 expression remains restricted to erythrocytes [85].

Pathway overrepresentation analysis revealed pathways previously linked to HNSCC to be shared between OTK and HNSCC cells, for example, the TGF (transforming growth factor)-β and EGFR pathways. EGFR is frequently reported as being overexpressed in HNSCC cells, but it is also essential for non-tumour keratinocyte function in processes such as, proliferation, migration, adhesion and survival [26, 81]. Indeed, blocking of EGFR in non-tumour keratinocyte cells can result in cell death [298]. TGF-β is a cytokine, which in non-tumour cells can inhibit

proliferation and induce cell differentiation, while in cancer cells, including in HNSCC, TGF-β can be pro-tumourigenic [118, 299]. Therefore, the presence of similar pathways in both OTK and HNSCC cells could be explained by the fact that these pathways are needed for the normal maintenance of cells. Moreover, as in the case of TGF-β, some genes and their proteins can exert tumour suppressor affects in some environments and tumourigenic in others.

Other pathways in our data common to both non-tumour and HNSCC cells were associated with micro RNAs (miRNAs). MiRNAs are small RNA molecules which block mRNA translation or target mRNA for degradation [134]. It has been reported by a number of studies that miRNA levels can be altered in the presence of HPV (human papillomavirus) [134, 282-284, 300]. On the other hand, in non-tumourigenic oral cells, miRNAs are required for normal cell function [301].

Analysis of EC (enhancer cluster) genes revealed annotated Hippo pathway genes to be HNSCC-specific. In addition, analysis of OTK ATAC peaks revealed the Hippo pathway to be highly enriched. The Hippo pathway is important in regulating cell growth through cell adhesion molecule and cell-cell signalling [92, 302, 303]. However, in cancer, loss of cell-cell signalling leads to aberrant cell growth [304, 305]. Therefore, it may be that signal inputs to the Hippo pathway such as mechanical cell stress function in non-tumour cells, while in cancer cells this may be lost. Further investigation of EC annotated genes of the Hippo pathway revealed *SMAD3* (mothers against decapentaplegic homolog 3) and the TGF-β receptor 1 gene (*TGFBR1*). SMAD3 is part of the SMAD TF family and is activated by the TGF-β receptor [88]. Cross-talk between the Hippo and TGF-β pathways and their role in EMT (epithelial to mesenchymal transition) has previously been reported. It

has been shown that one of the Hippo transcriptional regulators, TAZ (tafazzin) can form a heteromeric complex with SMAD2/3/4 and upon TGF-β activation the complex can activate TGF-β target genes including those involved in EMT [88, 306]. Therefore, it may be that in non-tumour cells upstream components of the Hippo pathway are active, which function to restrict cell growth, while in cancer cells, downstream transcriptional regulators are active resulting in aberrant cell growth.

EC pathways specific to HNSCC cells included the viral carcinogenesis pathway. Annotated genes in this pathway included genes which have previously been reported to be deregulated in HNSCC (*PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) and *JUN*) [36, 66-70, 72, 117, 252, 255]. The overlap in pathways affected in both HPV(-) and HPV(+) cell lines is evident in this data, as described in Chapter 1 **(Table 1.1 and Figure 1.8)**. However, further investigations may uncover novel genes and pathways which are deregulated in both HPV(-) and HPV(+) cells, providing an opportunity for the development of new therapies.

The NOTCH pathway was found to be over-represented in the OTK-specific EC dataset. This is in line with published data where *NOTCH1* inactivating mutations are frequently observed in HNSCC [36, 66-71]. Another study has reported that Notch1 overexpression leads to cell cycle arrest and differentiation by induction of p21 expression [22]. Therefore, this data indicates that *NOTCH1* silencing may have a role in these HNSCC cell lines as a tumour suppressor gene.

Limitations of this chapter include the use of ATAC-seq, as the method yields fewer mapped peaks compared to DNase I-seq. The inclusion of mitochondrial DNA

contributes to this limitation despite the method being improved in recent years, as discussed in Section 1.11.3. Additionally, it may be due to the inclusion of feeder layer cells, which are mouse embryonic fibroblast cells required for the growth of human keratinocytes. While mitochondrial DNA and mouse DNA can be computationally filtered out after sequencing, during sequencing, such DNA can deplete the number of reads from the nuclear DNA [195]. This may be one of the reasons OKF6 and HOK ATAC-seq did not work, as upon sequencing OKF6 and HOK ATAC samples displayed very high background and so peaks could not be called.

A third limitation of this chapter is that the RNA-seq data for non-tumourigenic cells was derived from HOK cells, while the ATAC-seq data was derived from OTK cells; the RNA-seq experiment was unsuccessful for the OTK and OKF6 RNA samples. The reason for this is likely to be a technical issue, as the quality of RNA indicated by the RNA integrity number was sufficient. Due to time constrictions and limited amount of RNA, it was not possible to repeat the sequencing.

Making comparisons between HOK RNA with OTK ATAC open regions must be conducted with caution, as the HOK cells are of foetal origin, while OTK cells are derived from an adult. Data derived from the same cell type would overcome this limitation. However, while the origin of HOK and OTK cells differ, encouragingly, the pathways identified in pathway enrichment analysis showed some overlap with those in the HNSCC cells, and also overlap with the each other (i.e. HOK RNA-seq pathways and OTK ATAC-seq pathways). Therefore, useful data can be derived from the HOK cells, but it must be interpreted with caution. A further limitation is that the RNA-seq and ATAC-seq data was derived from one cell type each.

Therefore, to strengthen this data, the sample size of non-tumour cells can be increased, and the sequencing assays can be conducted in the same cell types.

In summary, comparison of the HNSCC cell lines' regulatory network and that of non-tumourigenic oral cells has revealed some differences, for example, the presence of the NOTCH pathway in non-tumour cells and the viral carcinogenesis pathway in HNSCC cells. However, overall, large overlap is also evident between the cell types involving the EGFR and Hippo pathways, as well as the KLF and AP-1 TF families.

CHAPTER 5: INVESTIGATION OF p63, TEAD4 AND AP-1 TRANSCRIPTION FACTOR BINDING IN HNSCC CELL LINES

5.1 Introduction

The previous two results chapters aimed to map open chromatin regions in four HNSCC (head and neck squamous cell carcinoma) cell lines and in non-tumourigenic primary oral keratinocytes. The open chromatin regions were interrogated using known and *de novo* transcription factor (TF) motif analyses in order to investigate which TFs may be involved in these HNSCC cells. Analysis of known motifs in HNSCC indicated the involvement of five TF families: AP-1, p53/p63, KLF (Krüppel-like factor), RUNX (runt-related transcription factor) and TEAD (TEA domain transcription factor). This finding was supported by *de novo* motif analysis in both HNSCC and non-tumour cells. Furthermore, the motifs appeared to be unique to distal DNase I hypersensitive sites (DHSs), apart from the AP-1 motif which was also found in promoter DHSs, though stronger association at distal regions was observed **(Figure 3.15)**. Therefore, these TF families were investigated further.

The AP-1 motif was one of the most enriched TF motifs in the merged-HNSCC dataset and in the OTK (oral tonsil keratinocyte) cells. The AP-1 TF is a dimer consisting of two main families: JUN (JUN, JUNB, JUND) and FOS (FOS, FOSL1 (FOS like 1, AP-1 transcription factor subunit) and FOSL2). Homodimers can form of JUN proteins, while JUN and FOS proteins can form heterodimers [307].

Members of the JUN and FOS families have been associated with many cancers, including HNSCC [253, 255, 308].

The p63 motif was also enriched in our HNSCC and OTK cells and was uniquely present in distal HNSCC cell line DHSs. The p63 protein is a key epithelial regulator and is essential in epithelial organisation [12]. However, in HNSCC cells, *TP63* is often over-amplified and has pro-survival effects [257, 309]. Indeed, in our data we observed the presence of an EC (enhancer cluster) at the *TP63* gene, specific to HNSCC cell lines, which aligned with a higher level of *TP63* expression observed in the RNA-seq data. The *TP63* gene is rarely deleted unlike its family member p53, which often undergoes inactivating mutations in cancer [36, 66-71]. Furthermore, a study in which p63 ChIP was conducted in non-tumourigenic human foreskin keratinocyte (HFK) cells revealed that p63 binding was associated with processes such as senescence, while in laryngeal cells, p63 was associated with positive growth factor signalling [138]. Therefore, p63 could serve as an alternative therapeutic target to p53.

The RUNX1-3 TF family motif was enriched in the HNSCC cells, OTK cells and uniquely at distal HNSCC DHSs. In addition, the *RUNX1* and *RUNX2* genes were annotated in the HNSCC DHSs. Both RUNX1 and RUNX3 have been implicated in HNSCC [255, 261, 264, 310] and *RUNX2* has been reported to display abnormal methylation patterns in HNSCC [262, 263].

The motif for the KLF TF family is enriched in our HNSCC cell lines. The KLF family has previously been implicated in cancer; *KLF4* has been associated with poor patient prognosis in HNSCC, while in oral SCC cell lines, KLF5 overexpression has been observed through a decrease in a suppressive micro RNA

(miRNA) which regulates *KLF5* expression [310, 311]. In addition, whole-genome sequencing data derived from HNSCC tumours have revealed recurrent focal amplifications of enhancer clusters (ECs) near the *KLF5* gene, which can lead to KLF5 protein overexpression [179].

The TEAD motif was enriched in all four HNSCC cell lines investigated here. TEADs are a family of four TFs and are the downstream TF effectors of the Hippo pathway, as discussed next [312].

5.2 Hippo pathway

The potential involvement of the Hippo pathway in HNSCC was a novel finding at the time of the project's development. The Hippo pathway was over-represented in pathway enrichment analysis of HNSCC and OTK cells. The pathway is involved in organ growth and regulation of cell-cell contact inhibition by influencing transcription of genes involved in cell proliferation and growth [88, 92] **(Figure 5.1)**. The pathway involves a cascade of kinase proteins: MST (macrophage stimulating)-1 and -2 and LATS (Large Tumour Suppressor kinase)-1 and -2 [88, 92]. When active, MST1/2 form a complex and phosphorylate a scaffold protein, SAV1 (Salvador homolog 1). MST1/2 then phosphorylate LATS1/2 and its scaffold protein MOB1A/1B **(Figure 5.1)**. LATS1/2 in turn phosphorylate and promote cytoplasmic stabilisation of the downstream transcriptional effectors, TAZ (tafazzin) and YAP (yes-associated protein) [88, 92]. Upon receiving signals such as mechanical stress, in their phosphorylated state YAP/TAZ are restricted to the cytoplasm, have no transcriptional effect and are eventually targeted for degradation **(Figure 5.1)**. In environments of low cell density, YAP/TAZ are unphosphorylated and free to translocate to the nucleus [88] **(Figure 5.1)**. To induce

214

transcription, YAP and TAZ must bind to a TF such as TEAD which are a family of four highly homologous proteins (TEAD1-4) and share a highly conserved DNA binding domain [312]. The TEAD TF family are the main mediators of YAP/TAZ; studies designed to disrupt YAP/TAZ/TEAD interactions have revealed a decreased response to EMT (epithelial to mesenchymal transition) and anchorage-independent growth [313]. In addition to TEAD, the KLF4, RUNX-1, -3 and AP-1 (JUN) TFs can also serve as binding partners for the Hippo transcriptional regulators, YAP/TAZ [272].
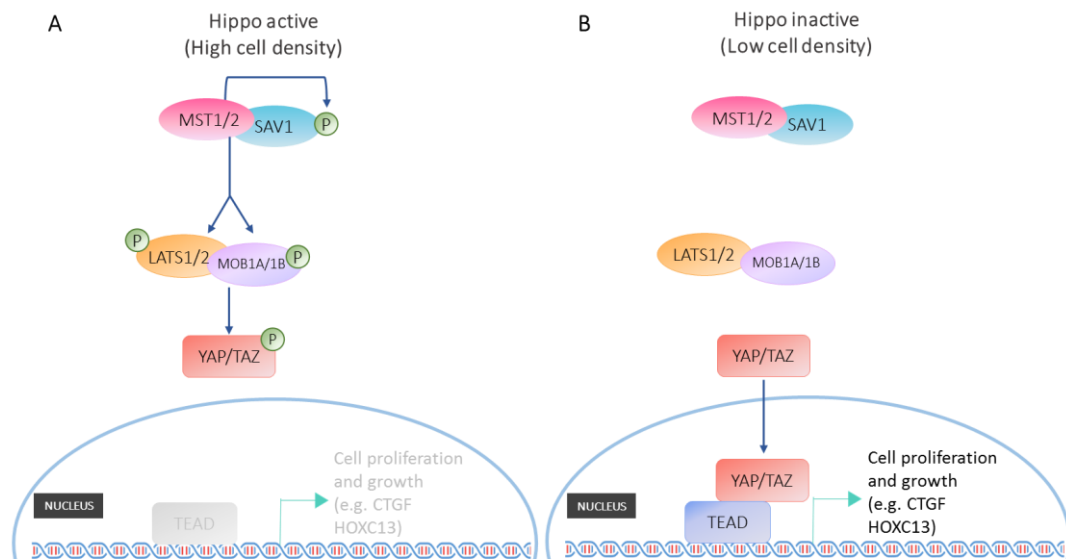
Figure 5.1. The Hippo signalling pathway.
(A) Upon activation of the Hippo pathway, the MST1/2 kinase proteins form a complex with the scaffold protein SAV1 and phosphorylate it. The MST1/2 kinases then phosphorylate the LATS1/2 kinases and its scaffold proteins MOB1A/1B. LATS1/2 then phosphorylate YAP/TAZ, stabilising them in the cytoplasm and inhibiting transcription. **(B)** In environments of low cell density, the Hippo pathway is switched off leading to dephosphorylation of the core kinases (MST1/2 and LATS1/2), the scaffold proteins (SAV1 and MOB1A/1B) and of YAP/TAZ. YAP/TAZ are then free to translocate to the nucleus and activate gene transcription of Hippo target genes such as *CTGF*, which is involved in processes such as cell proliferation and growth.

5.3 The role of the Hippo pathway in HNSCC

As briefly addressed in the main introduction (Section 1.5), the role of the Hippo pathway in recent years has become more prominent in HNSCC, particularly the role of YAP (recently reviewed by Segrelles *et al.*, 2018) [314]. For example, a study in which a total of 221 HNSCC tumours from the TCGA and University Hospital of A Coruña cohorts were interrogated, found *PIK3CA* gene overexpression to correlate with YAP protein inactivation and expression of Hippo-target genes [315]. Furthermore, the gene for the cadherin protein FAT1, which is involved in cell-cell signalling via the Hippo pathway, has been found to undergo frequent inactivation mutations and deletions **(Table 1.1)** [36, 66, 68, 70, 71, 316]. The loss of FAT1 has been found to result in inactivation of YAP, correlating with disease progression [93]. In addition, a study conducted in 35 HNSCC tumours found positive correlation between YAP expression and tumour metastasis, moreover, YAP expression co-localised at the tumours' invasive margin [87]. Therefore, the role of the Hippo pathway in HNSCC has previously been established, however its role at the chromatin level is lacking and will be addressed in this chapter.

5.4 Aims

In this chapter, the aim was to investigate the role of the AP-1, p53/p63, KLF, RUNX and TEAD TFs and the EGFR (epidermal growth factor receptor) and Hippo pathways by conducting *in silico* analysis of a panel of primary HNSCC tumours obtained from The Cancer Genome Atlas (TCGA) database [36]. This was followed by gene expression analysis by qPCR (quantitative polymerase chain reaction) in

our cell lines. Based on this analysis, the role of p63-AP1-TEAD in the HNSCC transcriptional network was further investigated by conducting ChIP-seq (chromatin immunoprecipitation-sequencing) to investigate their genome-wide binding patterns and possible interactions.

5.5 *In silico* analysis of a panel of 496 HNSCC tumours obtained from The Cancer Genome Atlas (TCGA) database

Before investigating the TF network in the HNSCC cell lines, analysis of a panel of 496 HNSCC tumours was conducted to establish whether the same TF network could be of importance in primary HNSCC tumours. The tumour panel was obtained from The Cancer Genome Atlas (TCGA) and data analysis was conducted using the cBioPortal tool [36, 317]. The analysis involved information on HPV (human papillomavirus) status, gene alterations and RNA expression. The genes considered were of the Hippo signalling pathway (*YAP1, WWTR1* (WW domain containing transcription regulator 1) (encodes TAZ) and the *TEAD* TF family), the AP-1 TF (*FOS*, *FOSL1*, *FOSL2*, *JUN*, *JUNB* and *JUND*), *KLF4/5*, *RUNX1-3*, *TP63* and *EGFR*. It was observed that the majority of gene alterations were amplifications and upregulation of RNA transcript levels **(Figure 5.2)**. Of the TF genes, the *TP63* gene was the most frequently altered in the HNSCC tumour panel (36%), followed by *FOSL1* (12%) and *TEAD4* (10%) **(Figure 5.2)**. Both *KLF-4* and *-5* were altered in 8% of tumours, while *RUNX1-3* were altered in 7%, 5% and 4% of tumours, respectively. Of pathway genes, the *EGFR* gene was altered in 21% of tumours, *YAP1* and *WWTR1* was altered in 17% and 22% of tumours, respectively **(Figure**

5.2). Therefore, alterations in the genes for p63, AP-1, EGFR and the Hippo pathways appear to be associated with HNSCC tumours and predominantly involve gene amplification and RNA upregulation, suggesting that the investigated TFs are involved in pro-tumourigenic processes and pathways.

Figure 5.2. Gene alterations in the TCGA provisional HNSCC tumour cohort of 496 tumours.
The cBioPortal tool was used to visually represent the percentage of tumours with one or more gene alteration in any given gene. Rows represent genes and columns represent samples. Gene alterations include amplification (red), deletion (blue), missense mutation (green) or truncating mutation (black).

The incidence of co-occurrence and mutual exclusivity between gene alterations in the same tumour was investigated. Statistically significant co-occurrence between *TP63/WWTR1*, *FOS/FOSB*, *YAP1/JUND*, *TP63/KLF4* were identified (**Table 5.1**). No genes investigated displayed statistically significant mutual exclusivity (**Table 5.1**). Therefore, an association appears to be present in the TCGA tumour panel between genes for Hippo and AP-1, as well as *TP63* and *KLF4*.

Table 5.1. *In silico* **analysis of co-occurrent gene alterations performed using the TCGA, 'provisional HNSCC' tumour cohort of 496 tumours.**

Gene A	Gene B	Neither	A Not B	B Not A	Both	Log odds ratio	Adjusted p-Value
TP63	WWTR1	290	96	29	81	2.133	<0.001
FOS	FOSB	453	23	9	11	>3	<0.001
YAP1	JUND	407	72	7	10	2.089	0.013
TP63	KLF4	306	152	13	25	1.354	0.016

The cBioPortal tool was used for analysis. Log odds ratio (OR) was calculated for gene A and gene B to reveal the likelihood of alterations in the two genes being mutually exclusive or co-occurent. OR = (A*D)/(B*C); where A = number of cases altered in both genes; B = number of cases altered in Gene 1 but not Gene 2; C = number of cases altered in Gene 2 but not Gene 1; and D = number of cases altered in either gene. Alterations in each gene pair is then assigned to a category; mutually exclusive, tendency to co-occur or no association [318].

Using the MEXPRESS tool, which allows visualisation of TCGA clinical data, correlation between HPV status and gene expression of the AP-1, *TEAD*, *KLF*, *RUNX* gene families, as well as *TP63*, *YAP1* and *EGFR* was investigated [319]. An example plot for the *TEAD4* gene is shown in **Figure 5.3**. *TEAD4* expression displayed statistically significant negative correlation with HPV status (P=0.009) (**Figure 5.3**). Analysis of other genes revealed a statistically significant positive correlation between HPV status and expression of *JUND*, *TEAD2* and *RUNX3* (**Table 5.2**). Negative correlation was found between expression of *TEAD1*,

TEAD4, *FOSL1*, *FOSL2*, *RUNX2*, *TP63*, *EGFR*, *YAP1*, *WWTR1* and HPV status **(Table 5.2)**. No statistically significant correlation was observed between HPV status and *FOS*, *FOSB*, *JUN*, *JUNB*, *KLF4*, *KLF5*, *RUNX* and *TEAD3* **(Table 5.2)**. Therefore, this data strengthens the association of the genes for AP-1, p63, EGFR and Hippo with HNSCC and their affects could be dependent on HPV status.

Table 5.2. Correlation between gene expression and HPV presence using the TCGA HNSCC data.

Gene	P-value	Correlation between gene expression and HPV presence
JUND	1.54E-02	Positive
TEAD2	1.52E-02	Positive
RUNX3	1.45E-02	Positive
TEAD1	1.41E-02	Negative
EGFR	3.41E-06	Negative
FOSL1	7.42E-08	Negative
RUNX2	4.4E-03	Negative
TEAD4	9.03E-03	Negative
TP63	1.73E-04	Negative
YAP1	1.94E-09	Negative
FOSL2	2.98E-04	Negative
WWTR1	2.44E-06	Negative
FOS	0.469	None (not significant)
FOSB	0.747	None (not significant)
JUN	0.494	Negative (not significant)
JUNB	0.768	Positive (not significant)
KLF4	0.812	Positive (not significant)
KLF5	0.405	Negative (not significant)
RUNX1	0.141	Positive (not significant)
TEAD3	0.575	Positive (not significant)
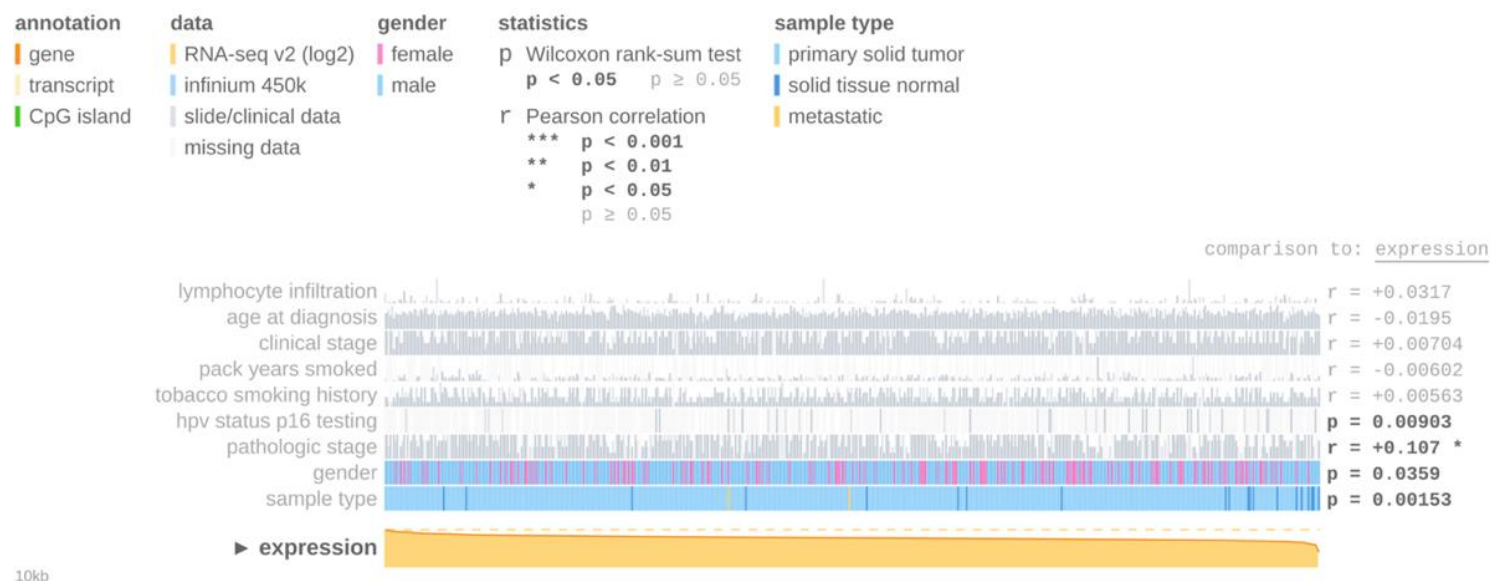
Figure 5.3. Visualisation of clinical data using the TCGA cohort of HNSCC tumours and its correlation with *TEAD4* gene expression.
The 'MEXPRESS' tool was used to visualise data for the *TEAD4* gene. The samples were ordered by *TEAD4* expression which displays the statistical significance of the relationship between gene expression and eight other parameters: lymphocyte infiltration, age at diagnosis, clinical stage, pack years smoked, tobacco smoking history, HPV status p16 testing, pathologic stage, gender, sample type. However, only the correlation between 'HPV status p16 testing' and gene expression were analysed here. HPV status was determined by the presence of p16 by immunohistochemistry. Samples showing a positive signal for p16 are indicated by the dark grey bars in the 'hpv status p16 testing' track. Though the MEXPRESS tool only displays p16 staining for the tumours, the TCGA also conducted RNA-seq to determine HPV status; if >1000 mapped reads aligned to the *E6* and *E7* genes, then the tumour was classed as HPV(+) [36].

5.6 QPCR analysis of transcription factor genes

To investigate the role of the genes highlighted above, qPCR was conducted to determine gene expression in our four HNSCC cell lines (SCC040(-), VU40T(-), VU147(+) and UDSCC2(+)). Gene expression in the normal, immortalised OKF6 cell line was also assessed as a control. Gene expression levels were normalised to the *GAPDH* housekeeping gene.

Expression levels of members of the AP-1 TF family (*FOS, FOSB, FOSL1, FOSL2, JUN, JUNB* and *JUND*) were investigated (**Figure 5.4A-G**). The *FOS* gene was expressed at very low levels in OKF6, VU40T(-), VU147(+) and UDSCC2(+) cells, while high expression was observed only in SCC040(-) cells (**Figure 5.4A**). On the other hand, *FOSB* was expressed at very low levels in SCC040(-), however *FOSB* was expressed at consistent levels in OKF6, VU40T(-) and UDSCC2(+) cells, while in VU147(+) cells *FOSB* was highly expressed (**Figure 5.4B**). Expression of *FOSL2*, *JUN*, *JUNB*, *JUND* was highest in SCC040(-) cells, while *FOSL1* was highly expressed in VU40T(-) and VU147(+) (**Figure 5.4C-G**). OKF6 cells display uniform expression across the AP-1 genes apart from high expression of *JUND* and no expression of *FOS* (**Figure 5.4A-G**).

The expression levels of *ATF3*, a member of the ATF (activating transcription factor) family, which can dimerise with the JUN family, and expression levels of *BACH* (BTB domain and CNC homolog)-*1*, a TF harbouring the AP-1 motif were investigated [320]. The SCC040(-) cells expressed *ATF3* most highly, while lowest expression was observed in VU40T(-) and UDSCC2(+) cells (**Figure 5.4H**).
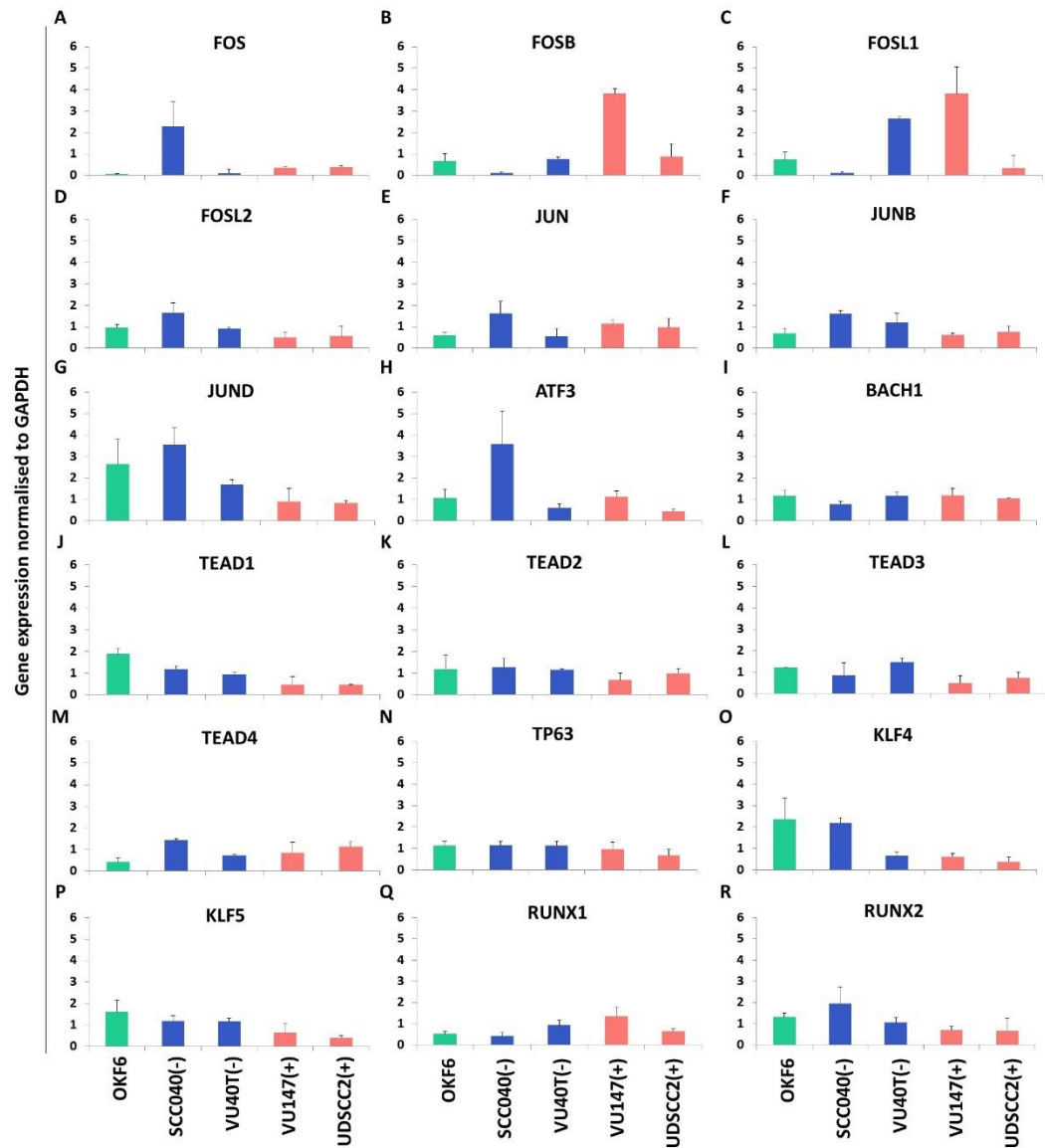
Figure 5.4 TF gene expression levels in HNSCC and OKF6 cell lines were assessed by qPCR.

QPCR was conducted to determine the expression levels in OKF6 (green bars) and HNSCC (blue bars: HPV(-) cells, red bars: HPV(+) cells) cells of the **(A)** *FOS*, **(B)** *FOSB*, **(C)** *FOSL1*, **(D)** *FOSL2*, **(E)** *JUN*, **(F)** *JUNB*, **(G)** *JUND*, **(H)** *ATF3*, **(I)** *BACH1*, **(J)** *TEAD1*, **(K)** *TEAD2*, **(L)** *TEAD3*, **(M)** *TEAD4*, **(N)** *TP63*, **(O)** *KLF4*, **(P)** *KLF5*, **(Q)** *RUNX1* and **(R)** *RUNX2*. Expression levels were normalised to *GAPDH*. The error bars represent SEM of biological triplicates.

Expression levels of the TEAD TF gene family were investigated (**Figure 5.4J-M**). *TEAD1* expression was highest in OKF6 cells and lowest in VU147(+) and UDSCC2(+) cell lines (**Figure 5.4J**). *TEAD2* expression was relatively consistent between all five cell lines, while *TEAD3* expression was lowest in SCC040(-), VU147(+) and UDSCC2(+) cells (**Figure 5.4K and L**). *TEAD4* was expressed highest in the HNSCC cell lines (**Figure 5.4M**).

Expression levels of *TP63* was also investigated and levels were consistent in four cell lines apart from UDSCC2(+) cells where levels showed a small decrease (**Figure 5.4N**).

Due to the large gene family of the KLF TFs, only genes most well studied in HNSCC were selected for investigation, which included *KLF4* and *KLF5* [179, 310, 311] (**Figure 5.4O and P**). Expression levels of *KLF4* was highest in OKF6 and SCC040(-) cells, while *KLF5* expression was lowest in both VU147(+) and UDSCC2(+) cells.

Gene expression of the RUNX family members, *RUNX1* and *RUNX2* were investigated (**Figure 5.4Q and R**). *RUNX1* expression was highest in VU40T(-) and VU147(+) cells, while *RUNX2* expression was highest in SCC040(-) cells and lowest in VU147(+) and UDSCC2(+) cells.

In summary, *in silico* TCGA analysis revealed the p63, AP-1, EGFR and the Hippo pathways to have strong association with HNSCC. Of these genes, the *TP63* gene was the most altered gene in the TCGA HNSCC tumour cohort and its expression negatively correlated with the HPV status of the tumours. In our data, *TP63* was expressed in all four HNSCC cell lines investigated. Of the AP-1 genes, *FOSL1* was the most altered AP-1 gene in the TCGA data and negatively correlated with

tumour HPV status. Furthermore, in our cell lines, *FOSL1* was one of the most highly expressed genes. The TCGA HNSCC tumours displayed *JUND* alterations least frequently, however *JUND* alterations were significantly co-occurrent with *YAP1* alterations and displayed significant positive correlation with HPV status. In addition, the *JUND* gene was one of the most highly expressed AP-1 genes in our HNSCC cell lines. Of the TEAD genes, *TEAD4* was the most highly expressed in the TCGA data analysis and was most highly expressed in all four HNSCC cell lines compared to OKF6 cells. Therefore, for the next step of the investigation, genome-wide TF binding of FOSL1, JUND, p63 and TEAD4 were investigated by ChIP-seq.

5.7 ChIP-seq of TEAD4, p63, FOSL1 and JUND

ChIP-seq was conducted in VU40T(-) and VU147(+) cell lines, as they displayed high gene expression for all four TFs. ChIP was performed in three independent biological replicates, as described in Chapter 2, Section 2.8. Before the ChIP samples were sequenced, successful immunoprecipitation and enrichment for binding sites were tested.

5.7.1 ChIP quality control

QPCR was conducted to determine successful ChIP using primers designed at putative TF binding sites. These sites were selected based on ChIP-seq data obtained from the Cistrome database [217]. ChIP-seq data for the four HNSCC cell lines used here were not available, therefore data from similar cell types were used to find TF binding sites and design qPCR primers. The following datasets were used: p63 ChIP-seq data from an oesophageal SCC cell line (KYSE-70) [79],

TEAD4 ChIP-seq data from a lung carcinoma cell line (A549) [321] and ChIP-seq data for JUND and FOSL1 from a cervical cancer cell line (HeLa-S3) [322]. Using a list of top putative binding sites, a binding site was selected and verified as a DHS site in our data for the VU40T(-) and VU147(+) cell lines, indicating possible TF binding. Binding sites near the *PC* (pyruvate carboxylase), *ANXA2* (annexin 2), *LNMA* (lamin A/C), and *S100A10* (S100 calcium binding protein A10) genes were selected as sites to assess FOSL1 and JUND ChIP enrichment **(Figure 5.5 and Appendix: Figures 3-5)**, two sites for p63 (site 1 near to the *TP73* and site 2 near to the *JMJD7* (jumonji domain containing 7) gene) **(Appendix: Figure 6 and Figure 5.6)** and two sites for TEAD4 ChIPs (sites 1 and 2) were also selected **(Appendix: Figure 7 and Figure 5.7)**.
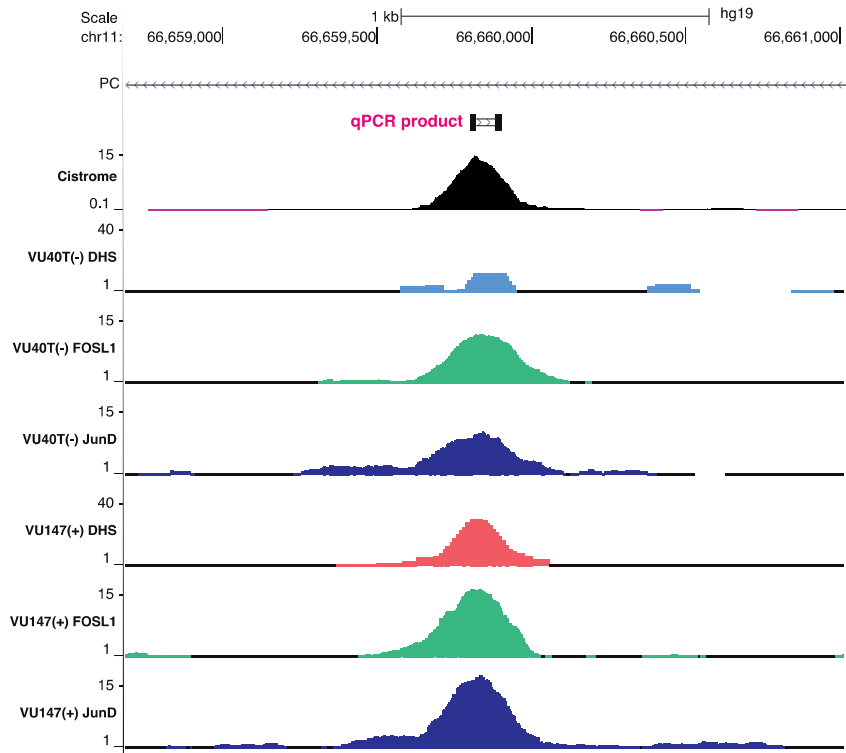
Figure 5.5. UCSC (University of California Santa Cruz) browser screenshot for FOSL1 and JUND ChIPs at the *PC* gene.

QPCR primers were selected near the *PC* gene to test for potential FOSL1 and JUND binding. The qPCR product (mapped on to the coverage tracks) overlaps with subsequent FOSL1 and JUND TF ChIP peaks in the VU40T(-) and VU147(+) cell lines (green and blue tracks, respectively); the data were obtained from the Cistrome database and is shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines.
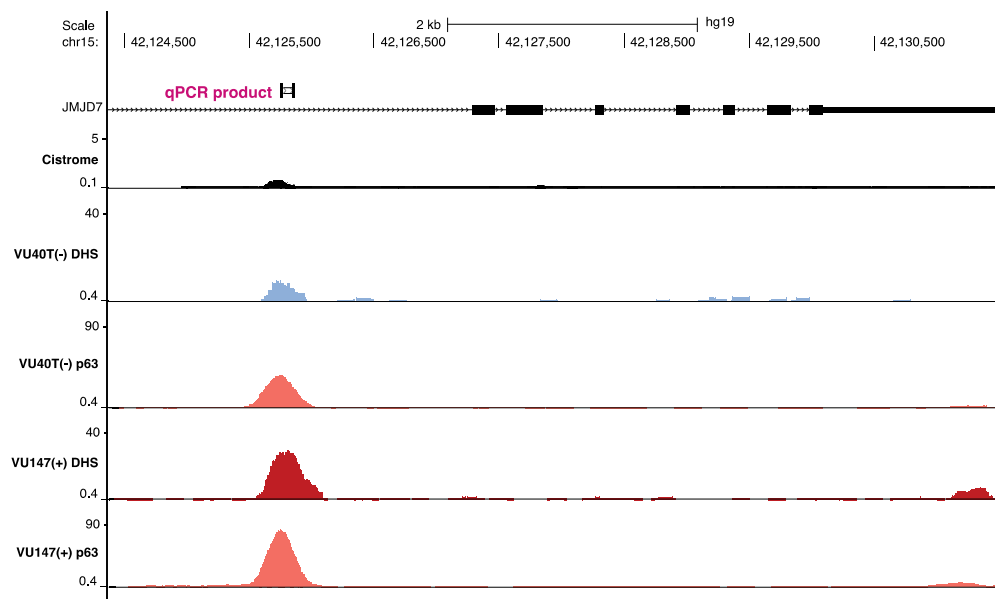
Figure 5.6. UCSC browser screenshot for p63 ChIP at the *JMJD7* gene (site 2).
QPCR primers were selected near the *JMJD7* gene to test for potential p63 binding. The qPCR product (mapped on to the coverage tracks) overlaps with subsequent p63 TF ChIP peaks in the VU40T(-) and VU147(+) cell lines (orange tracks); the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines.
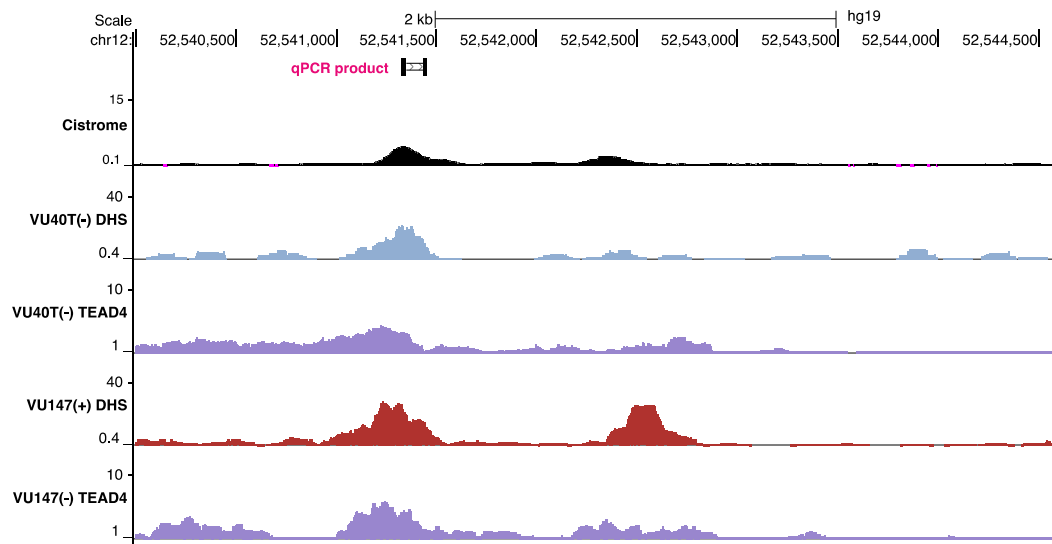
Figure 5.7. UCSC browser screenshot for TEAD4 ChIP (site 2).
QPCR primers were selected ~2.2kb upstream of the *KRT* (keratin)-*80* TSS to test for potential TEAD4 binding. The qPCR product (mapped on to the coverage tracks) overlaps with subsequent TEAD4 TF ChIP peaks in the VU40T(-) and VU147(+) cell lines (purple tracks); the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines. The remaining tracks show the outcome of subsequent ChIP-seq experiments in VU40T(-) and VU147(+) cells indicating binding at TEAD4 (purple tracks).

Each AP-1 primer set was first tested in one biological replicate of VU40T(-) FOSL1 and JUND ChIPs **(Figure 5.8A)**. To assess the ChIP enrichment at the putative regions the 'no DHS' primer was used as a negative control. The data were normalized against the signal from an input sample and shown as '% of input'. The signal was compared to the 'no antibody' control sample **(Figure 5.8A)**. ChIP enrichment was observed at all locations when compared to the 'no antibody' signal and as expected, enrichment was much lower at the 'no DHS' region. For the FOSL1 ChIP the observed fold change for the positive control primers above the 'no antibody' sample was between 10- and 60-fold and between 3- and 7-fold in the 'no DHS' region. For the JUND ChIP the observed fold change over the 'no antibody' sample was between 8- and 50-fold and between 6- and 10-fold over the 'no DHS' region for JUND ChIP. **(Figure 5.8A)**. As enrichment was the highest for the PC region, this primer set was used to assess the ChIP efficiency in three independent ChIP experiments using VU40T(-) and VU147(+) cells **(Figure 5.8B and C)**.
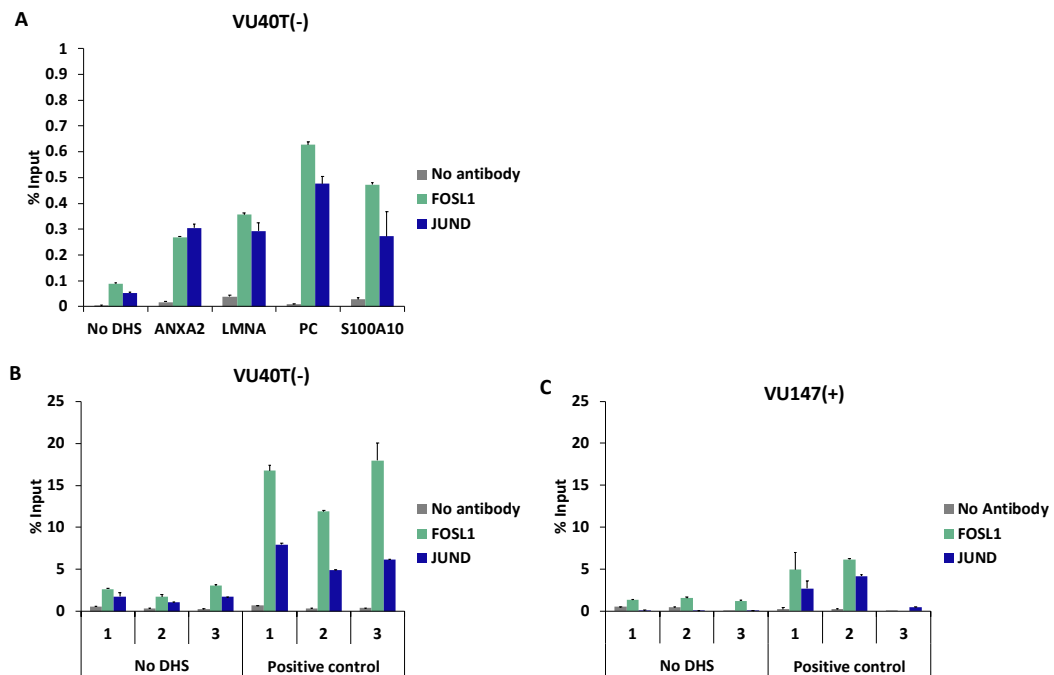
Figure 5.8. VU40T(-) AP-1 ChIP assessed by qPCR.

(A) Four AP-1 positive control qPCR primers were tested in VU40T(-) FOSL1 and JUND ChIPs. A negative control primer designed to a 'no DHS' region was also used. The positive and negative qPCR signals were assessed in the 'no antibody' sample to control for background and in the input sample to calculate the ChIP percentage input. The PC primer was used to assess the three biological replicates in **(B)** VU40T(-) and **(C)** VU147(+) ChIP samples. The bars represent the range between sample duplicates.

For each biological replicate of FOSL1 ChIP in VU40T(-) cells, binding displayed 24.0-, 29.8- and 45.0-fold change at the PC site over the 'no antibody' signal and 15.5-, 14.2-, 17.2-fold change over the 'no DHS' region (**Figure 5.8B**). For each replicate of JUND binding at the PC site over the 'no antibody' sample, the signal showed 11.3-, 12.3- and 15.5-fold change and 4.4-, 4.5- and 3.7-fold change over the 'no DHS' region (**Figure 5.8B**). In VU147(+) cells, FOSL1 enrichment over the 'no antibody' sample was 24.5-, 20.7- and 60.0-fold and 28.6-, 25.8- and 0.5-fold over the 'no DHS' region in the three replicates (**Figure 5.8B**). JUND binding fold enrichment over the 'no antibody' sample was 13.5-, 14.0- and 50-fold and 39-, 140- and 10-fold change over the 'no DHS' signal for each replicate (**Figure 5.8C**). Based on data obtained from sequenced samples which displayed high background (not shown), a fold change of at least three over the 'no antibody' signal was set as the qPCR signal cut-off. Therefore, all three VU40T(-) ChIP biological replicates for FOSL1 or JUND were pooled and sequenced. For VU147(+) cells, only biological replicates one and two were pooled and taken for sequencing, as the ChIPs for replicate three did not appear to work.

Primers were also tested for p63 and TEAD4 ChIPs. Primers for p63 sites 1 and 2 were designed near to the *TP73* and *JMJD7* genes, respectively (**Appendix: Figure 6 and Figure 5.6**). Primers for TEAD4 sites 1 and 2 were designed ~35kb downstream of the *PCED1B* (PC-esterase domain containing 1B) TSS and ~2.2kb upstream of the KRT80 TSS, respectively (**Appendix: Figure 7 and 5.7**).

Each p63 primer set was tested in one biological replicate of p63 or TEAD4 ChIP from VU40T(-) cells. p63 ChIP fold change over 'no antibody' was 1.6- and 14.8-fold for sites 1 and site 2, respectively, while fold change over the 'no DHS' region

was 2.9- and 4.8-fold, respectively **(Figure 5.9A)**. Therefore, the p63 site 2 primer set was selected to assess enrichment of other p63 ChIP samples.

VU40T(-) TEAD4 ChIP biological replicate 1 was also tested using the TEAD4 primer sets **(Figure 5.9B)**. TEAD4 ChIP fold change over 'no antibody' was 1.2- and 5.9-fold for TEAD4 sites 1 and 2, respectively, while fold change over the 'no DHS' region was 0.5- and 2.7-fold, respectively **(Figure 5.9B)**. Therefore, the TEAD4 site 2 primer set was selected to assess enrichment of other TEAD4 ChIP samples.

The remaining two p63 ChIP biological replicates for VU40T(-) cells were assessed by using the p63 site 2 primer set **(Figure 5.9C)**. p63 ChIP fold change over the 'no antibody' samples was 23.0- and 6.8-fold for p63 ChIP replicates 2 and 3, respectively, while over 'no DHS' enrichment was 7.0- and 2.8-fold, respectively **(Figure 5.9C)**.

TEAD4 ChIP enrichment of VU40T(-) biological replicates two and three was assessed using the TEAD4 site 2 primer set **(Figure 5.9D).** TEAD4 fold change over 'no antibody' was 8.7- and 2.8-fold, while enrichment over the 'no DHS' region was 0.9- and 0.6-fold, for replicates 2 and 3, respectively **(Figure 5.9D)**.
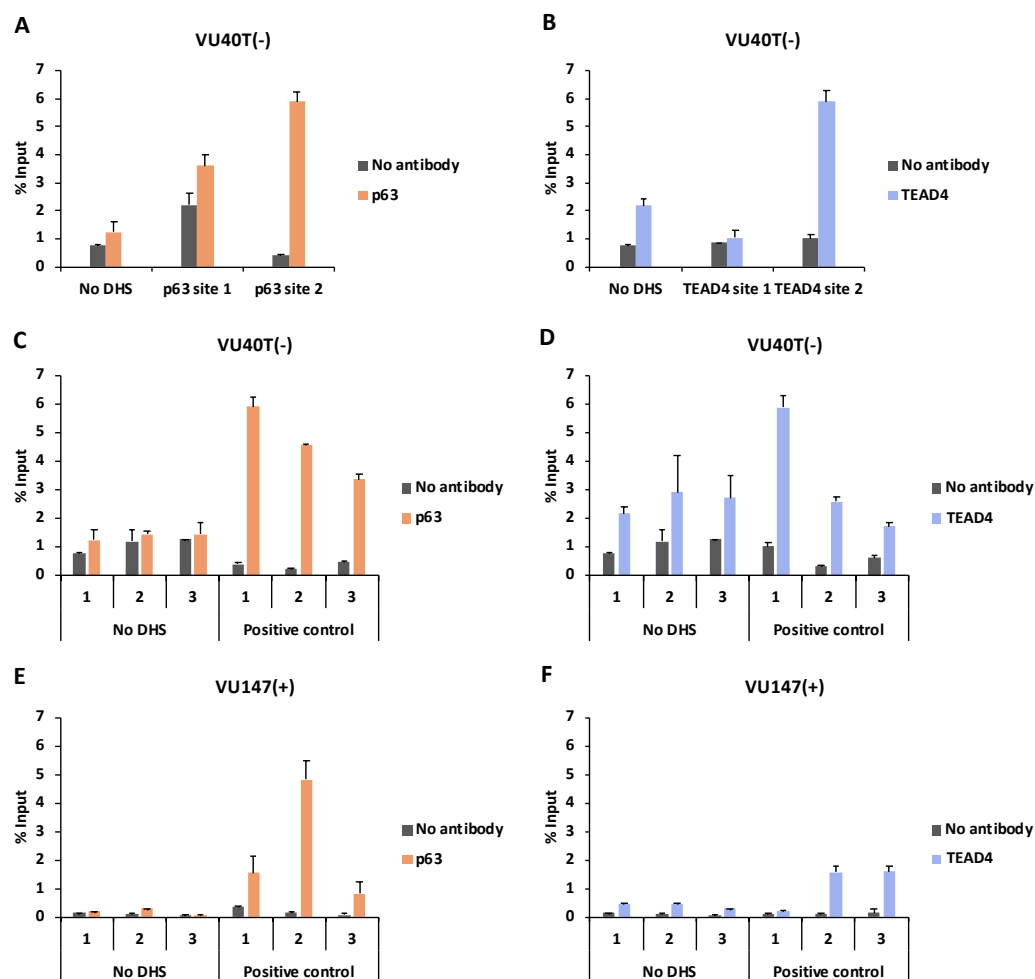
Figure 5.9. QPCR was conducted to assess p63 and TEAD4 ChIP enrichment in VU40T(-) and VU147(+) cells.

Two primer sets were designed to two putative binding sites in (**A**) p63 (p63 site 1 and site 2) and (**B**) TEAD4 (TEAD4 site 1 and site 2). The 'no DHS' primer was used as a negative control. A 'no antibody' sample was also included to control for background. ChIP was conducted in three biological replicates in VU40T(-) cells for (**C**) p63 and (**D**) TEAD4 and three biological replicates in VU147(+) cells for (**E**) p63 and (**F**) TEAD4. The bars represent the range between technical duplicates.

Therefore, for sequencing of VU40T(-) p63 ChIPs, all three biological replicates were quantified and combined in equal concentrations. For VU40T(-) TEAD4 ChIP samples, only biological replicate one showed enrichment over the negative control region, therefore this experiment was taken for sequencing.

In VU147(+) cells, p63 ChIP samples were assessed by qPCR using the p63 site 2 primer set (**Figure 5.9E**). p63 fold change over 'no antibody' was 4.4-, 48 and 8.0-fold, respectively, while fold change over 'no DHS' was 8.0-, 16.0-, and 8.0-fold for replicates 1-3, respectively (**Figure 5.9E**). VU147(+) TEAD4 ChIP was assessed using the TEAD4 site 2 primer. TEAD4 ChIP fold change over 'no antibody was 2.0-, 16.0- and 16.0-fold for replicates 1-3, respectively. TEAD4 ChIP fold change over 'no DHS' was 0.5-, 3.5- and 5.4-fold for replicates 1-3, respectively (**Figure 5.9F**). Therefore, all three VU147(+) p63 ChIP biological replicates were pooled and taken for sequencing, while TEAD4 replicates 2 and 3 were pooled and sequenced.

5.7.2 Mapping of ChIP-seq data

The sequencing libraries were synthesised for the p63, TEAD4, FOSL1 and JUND ChIP samples described above and sequencing was conducted as described in Methods and Materials, Sections 2.9.3 and 2.12. In VU40T(-) cells, between 10,137 and 75,970 peaks were found across the four TF ChIPs and between 7,704 and 49,309 peaks in VU147(+) cells (**Table 5.3 and Figure 5.10**). p63 shows the highest number of binding sites in both cell lines, which could be either due to the widespread transcriptional regulation (as expected from cells originating from squamous epithelium) or due to high affinity and specificity of the p63 antibody.

Table 5.3. Number of mapped ChIP-seq peaks in VU40T(-) and VU147(+) cell lines.

ChIP	VU40T(-)	VU147(+)
p63	75,970	49,309
TEAD4	10,137	11,927
FOSL1	39,824	7,446
JUND	27,537	7,704
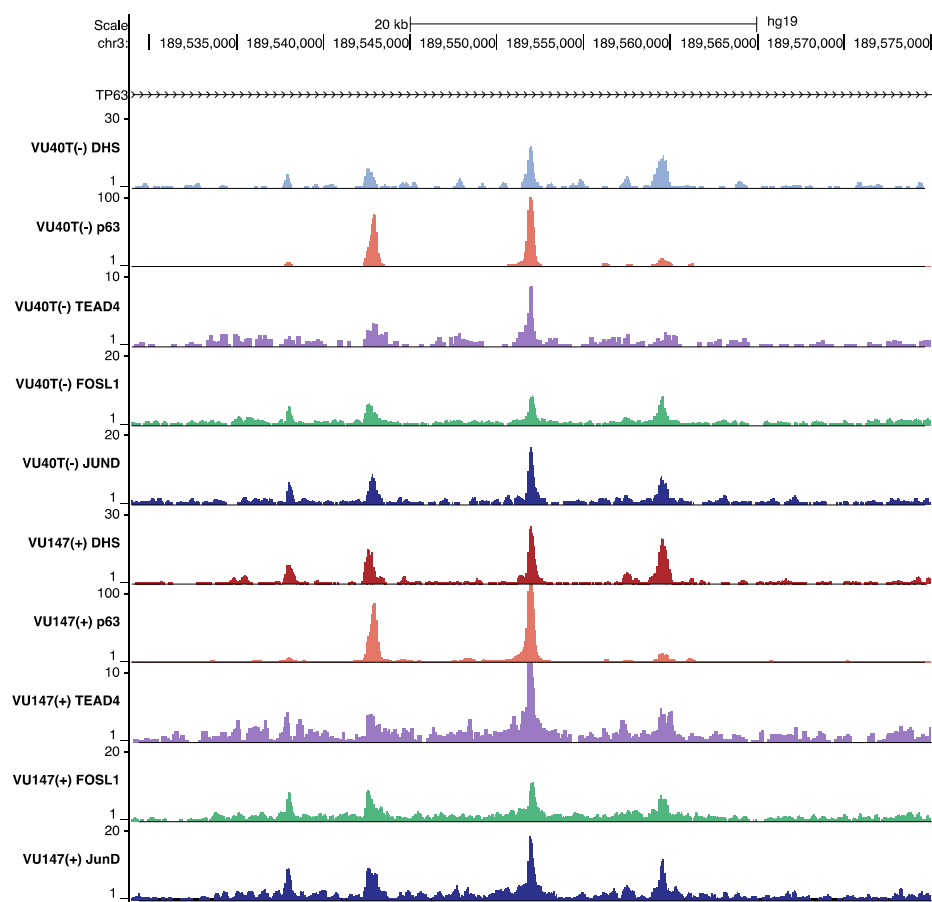
Figure 5.10. UCSC screenshot of VU40T(-) and VU147(+) p63, TEAD4, FOSL1 and JUND ChIP-seq peaks at the *TP63* gene.

The ChIP-seq reads for p63, TEAD4, FOSL1 and JUND ChIPs conducted in VU40T(-) cells and VU147(+) cells were aligned to the hg19 genome and the Bigwig files were uploaded to the UCSC genome browser, along with the files for VU40T(-) DHS (light blue track) and VU147(+) (red track). The ChIP-seq peaks were called using HOMER for p63 (orange track), TEAD4 (purple track), FOSL1 (green track) and JUND (blue track), respectively.

5.7.3 ChIP-seq pathway enrichment analysis

HOMER (Hypergeometric Optimization of Motif EnRichment) was used to conduct pathway enrichment analysis for each TF ChIP-seq experiment in VU40T(-) and VU147(+) cells, in which the closest gene was assigned to each peak, as before (Chapter 1, Section 3.10 and Chapter 2, Section 4.9). The top 20 pathways for each TF ChIP were taken for further analysis.

In VU40T(-) cells a total of 35 pathways were over-represented and could be grouped into nine clusters based on TF enrichment **(Figure 5.11)**. The largest of which was group (1) and consisted of 11 pathways over-represented in all four TFs. Six of these pathways were associated with cell adhesion and migration, including proteoglycans in cancer, Hippo, leukocyte transendothelial migration, adherens junction, RAP1 (Rap1 GTPase) signalling, myometrial relaxation and contraction pathways. The EGFR pathway was also highly over-represented in this group **(Figure 5.11)**. Interestingly, the Hippo pathway appears highly over-represented in all four TFs, although the highest association is observed in the TEAD4 ChIP dataset. Group (2) consisted of four pathways under-represented in TEAD4 and included the HNSCC-related pathways WNT (wingless) and RAS (resistance to audiogenic seizures) signalling **(Figure 5.11)**. Group (3) consisted of four pathways under-represented in p63 including the PI3K (phosphotidylinositol-3 kinase)-AKT (protein kinase B) signalling, α6β4 integrin signalling and tight junction pathways. Other HNSCC-related pathways were over-represented in groups (4), (5) and (7), which display pathways over-represented in p63/JUND, TEAD4/JUND and p63, respectively **(Figure 5.11)**. These pathways included ERBB signalling, TGF (transforming growth factor)-β and MAPK. Cancer-related pathways were also

over-represented in group (5) which included renal cell carcinoma and small cell

lung cancer pathways (**Figure 5.11**).

Figure 5.11. Pathway enrichment analysis of p63, TEAD4, FOSL1 and JUND ChIP-seq in VU40T(-) cells.
Pathway enrichment analysis was conducted, and the top 20 pathways were taken from each ChIP-seq experiment. The analysis revealed a total of 35 different pathways. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into 9 groups by TF: (**1**) over-represented in all (11/35; 31%), (**2**) under-represented in TEAD4 (4/35; 11%), (**3**) under-represented in p63 (4/35; 11%), (**4**) under-represented in TEAD4/FOSL1 (6/35; 6%), (**5**) under-represented in p63/FOSL1 (3/35; 7%), (**6**) under-represented in FOSL1 (1/35; 3%), (**7**) over-represented in p63 (3/35; 9%), (**8**) over-represented in FOSL1 (1/35; 3%), (**9**) under-represented in all (2/35; 6%).

In VU147(+) cells, a total of 36 pathways were over-represented and could be grouped into seven clusters **(Figure 5.12)**. One of the largest groups was group (1), which included pathways over-represented in all ChIP datasets **(Figure 5.12)**. Several pathways in group (1) overlapped with those in VU40T(-) group (1) (pathways in cancer, proteoglycans in cancer, Hippo, EGFR, endocytosis, adherens junction and RAP1 signalling) **(Figure 5.12)**. Group (2) included four pathways under-represented in p63, three of which were associated with cell adhesion (focal adhesion, ECM (extracellular matrix)-receptor interaction and ARVC pathways). Group (3) included six pathways under-represented in TEAD4. Interestingly, cell adhesion-related pathways were not present in this group. Group (4) was the second largest group and included nine pathways over-represented in p63/JUND, relating largely to immune function. Groups (5) and (6) included pathways less represented in the ChIP datasets, while group (7) included pathways under-represented in FOSL1 ChIP **(Figure 5.12)**.

Figure 5.12. Pathway enrichment analysis of p63, TEAD4, FOSL1 and JUND ChIP-seq in VU147(+) cells.

Pathway enrichment analysis was conducted, and the top 20 pathways were taken from each ChIP-seq experiment. The analysis revealed a total of 36 different pathways. The pathways were ranked by log p-value; green represents the least statistically significant pathways, red represents most significant and blank boxes indicate no enrichment. The pathways were grouped into 7 groups by TF: **(1)** over-represented in all (11/36; 31%), **(2)** under-represented in p63 (4/36; 11%), **(3)** under-represented in TEAD4 (6/36; 17%), **(4)** under-represented in TEAD4/FOSL1 (9/36; 25%), **(5)** over-represented in FOSL1 (1/36; 3%), **(6)** over-represented in TEAD4 (1/36; 3%), **(7)** under-represented in FOSL1 (4/36; 11%).

243

ChIP-seq pathway enrichment analysis for the four TFs revealed that over half of the VU40T(-) pathways (51%; 18/35) overlapped with 47% (17/36) of VU147(+) pathways **(Figure 5.11 and 5.12)**. In both cell lines, all four TFs are strongly involved in cancer-related pathways (pathways in cancer and proteoglycans in cancer), cell adhesion, migration or proliferation-related pathways (Hippo, adherens junction and RAP1) and the EGFR pathway. In addition, in both cell lines the p63 TF was most involved in cell adhesion/migration/proliferation-related pathways (RAP1, adherens and regulation of actin cytoskeleton), as well as in HNSCC-related pathways RAS/MAPK and ERBB. On the other hand, p63 was less involved in other HNSCC-related pathways, namely PI3K/AKT. In both cell lines, p63 was also less involved in some of the cell-adhesion related pathways, though these were distinct in the two cell lines; tight junction and $\alpha6\beta4$ integrin signalling pathways in VU40T(-) cells and focal adhesion and ECM-receptor interaction pathways in VU147(+) cells **(Figure 5.11 and 5.12)**.

Regarding TEAD4 ChIP pathway enrichment, in both cell lines the TEAD4 TF was largely involved in pathways associated with cell adhesion, migration and proliferation, which included the Hippo, focal adhesion and adherens junction pathways **(Figure 5.11 and 5.12)**. TEAD4 was less involved in WNT, RAS/MAPK and regulation of actin cytoskeleton in both cell lines. In addition, TEAD4 was less involved in pathways associated with metabolism such as, insulin signalling, glycerolipid metabolism, vitamin A and carotenoid metabolism in VU40T(-) cells and insulin signalling and prolactin signalling in VU147(+) cells **(Figure 5.11 and 5.12)**.

The AP-1 factors were associated in both cell lines, largely with cell adhesion pathways (Hippo, adherens junction, RAP1, $\alpha6\beta4$ integrin signalling pathways), as

well as the EGFR/PI3K/AKT pathway. Furthermore, the AP-1 factors were less involved in MAPK in both cell lines, while FOSL1 was less involved in ERBB, TGF-β in both cell lines and WNT in VU147(+) cells. Interestingly, in both cell lines JUND is more associated than FOSL1 with the Hippo and EGFR signalling pathways (**Figure 5.11 and 5.12**).

Furthermore, many of the same pathways overlapped with the DHS HNSCC pathways. In VU40T(-) cells, 49% of ChIP pathways overlapped with DHS pathways (**Figure 5.11 and 3.16**). In VU147(+) cells, 43% of pathways overlapped with DHS pathways (**Figure 5.11 and 3.16**).

5.7.4 Transcription factor motif analysis of p63, TEAD4 and AP-1 ChIP-seq data

De novo motif analysis was carried out using VU40T(-) ChIP-seq datasets for p63, TEAD4, FOSL1 and JUND for which between 22 and 23 *de novo* motifs were predicted. The top five TF motifs are shown for each dataset (**Figure 5.13**). In the p63 ChIP-seq dataset, the most highly enriched motifs were p63 and AP-1, HMBOX (homeobox)-HES, TEAD and RUNX (**Figure 5.13**). In the FOSL1 and JUND ChIP-seq datasets the most enriched motif was AP-1; other motifs in the FOSL1 dataset were CEBP (CCAAT Enhancer Binding Protein)/HLF (hepatic leukemic factor), ETS (E twenty-six transcription factor), RUNX and TEAD, while in JUND ChIP data the motifs were ETS, GRHL (grainyhead-like protein), TEAD and RUNX (**Figure 5.13**). In the TEAD4 ChIP-seq motif dataset, the AP-1 motif was the most enriched, followed by TEAD, p53/p63, EGR (early growth response) and RUNX (**Figure 5.13**).

Over half (56.9%) of the p63 ChIP-seq peaks were found to be p63-specific and did not show overlap with AP-1 or TEAD4 peaks. *De novo* motif analysis was conducted on the p63-specific peaks to reveal the p63 motif to be the most enriched, followed by TEAD, ERG (ETS related gene), AP-1 and KLF5 (**Figure 5.13**).

In summary, the ChIPs for the four TFs indicate high enrichments for respective TF motifs as well as common enrichment for AP-1, TEAD and RUNX motifs. However, they also show specificity: HMBOX/HES (hairy and enhancer of split) for p63, CEBP/HLF for FOSL1, GRHL for JUND and EGR for TEAD4 (**Figure 5.13**). In addition, AP-1 TFs associate more with the ETS family motif, while p53/p63 is not present in the top five motifs. The presence of the TEAD motif in

246

the p63-specific peaks, but lack of peak overlap may be due to the poorer quality of

TEAD4 immunoprecipitation.
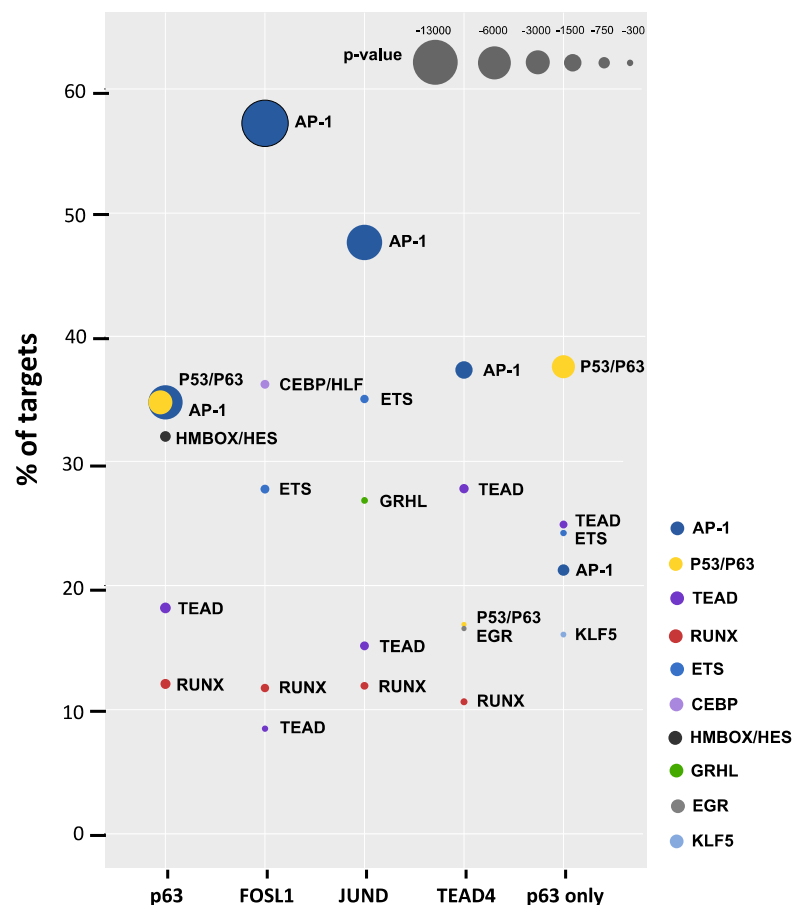
Figure 5.13. TF motif analysis of VU40T(-) ChIP-seq peaks.
ChIP-seq was conducted in p63, FOSL1 and JUND to obtain three sets of peaks representing TF binding events. A fourth dataset was formed by analysing p63 peaks (p63 only) which did not overlap with any of the remaining ChIP-seq datasets for FOSL1, JUND and TEAD4. *De novo* motif analysis was conducted in each dataset using the 'findMotifsGenome' script within HOMER. A total of 22, 22, 22, 23 and 24 motifs were predicted for p63, FOSL1, JUND, TEAD4 and p63 only ChIP-seq peaks, respectively. The top 5 motifs were ranked by percentage of target sequences that contained them. The size of the circles represents p-value.

5.7.5 Analysis of transcription factor binding overlap

To determine whether the TF motif co-occurrence reflects physical co-binding between the TFs, overlap between ChIP peaks was investigated in both cell lines. This was achieved by using HOMER to create a merged-ChIP peaks dataset which, similarly to the merged-DHS dataset, indicates regions of overlap. In VU40T(-) and VU147(+) cells it was found that the majority of JUND peaks overlap with p63 peaks (58% and 62%, respectively), while only 21% and 10% of p63 peaks overlap with JUND, respectively **(Figure 5.14A and B)**. This pattern was also observed in the FOSL1/p63 overlap data where in VU40T(-) cells 53% of FOSL1 peaks overlapped with 28% of p63 peaks, while in VU147(+) cells, 56% of FOSL1 peaks overlapped with 9% of p63 peaks **(Figure 5.14A and B)**. Between both of the AP-1 factors, in VU40T(-) cells, 36% of FOSL1 peaks overlapped with 51% of JUND peaks and in VU147(+) cells, 50% of FOSL1 peaks overlapped with 49% of JUND peaks **(Figure 5.14A and B)**. Next, the overlap between TEAD4 and p63 was investigated. In VU40T(-) cells, 2% of p63 peaks overlapped with 18% of TEAD4 peaks, while in VU147(+) cells 68% of TEAD4 peaks overlapped with 17% of p63 peaks **(Figure 5.14A and B)**. Finally, the overlap between TEAD4 and AP-1 was investigated in VU40T(-) cells and it was found that 8% of TEAD4 peaks overlap with 2% of FOSL1 peaks, while 7% of TEAD4 peaks overlap with 3% of JUND peaks **(Figure 5.14A and B)**.
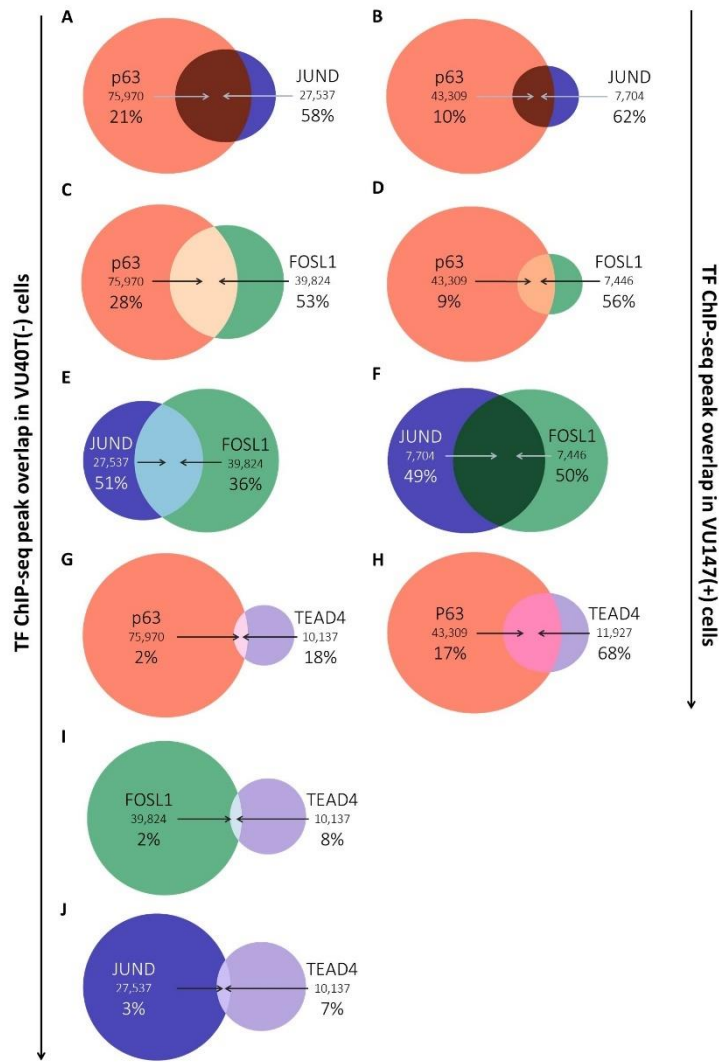
Figure 5.14. Overlap between p63, JUND, FOSL1 and TEAD4 ChIP-seq peaks in VU40T(-) and VU147(+) cells.

P63 and JUND peak location was investigated to find percentage overlap between TF binding in **(A)** VU40T(-) cells: 75,970 p63 ChIP-seq peaks and 27,537 JUND peaks were mapped; 21% of p63 peaks overlapped with 58% of JUND peaks and in **(B)** VU147(+) cells: 43,309 p63 peaks and 7,704 JUND peaks were mapped; 10% of p63 peaks overlapped with 62% of JUND peaks. p63 and FOSL1 peak location was investigated in **(C)** VU40T(-) cells: 75,970 p63 peaks and 39,824 FOSL1 peaks were mapped; 28% of p63 peaks overlapped with 53% of FOSL1 peaks and in **(D)** VU147(+) cells: 43,309 p63 peaks and 7,446 FOSL1 peaks were mapped; 9% of p63 peaks overlapped with 56% of FOSL1 peaks. JUND and FOSL1 peak location was investigated in **(E)** VU40T(-) cells: 27,537 JUND peaks and 39,824 FOSL1 peaks were mapped; 51% of JUND peaks overlapped with 36% of FOSL1 peaks and in **(F)** VU147(+) cells: 7,704 JUND peaks and 7,446 FOSL1 peaks were mapped; 49% of JUND peaks overlapped with 50% of FOSL1 peaks. TEAD4 and p63 peak location was investigated in **(G)** VU40T(-) cells: 75,970 p63 peaks and 10,137 TEAD4 peaks were mapped; 2% of p63 peaks overlapped with 18% of TEAD4 peaks and in **(H)** VU147(+) cells: 49,309 p63 peaks and 11,927 TEAD4 peaks were mapped; 17% of p63 peaks overlapped with 68% of TEAD4 peaks. In VU40T(-) cells, **(I)** FOSL1 peak overlap with TEAD4 peaks was investigated: 39,824 FOSL1 and 10,137 TEAD4 peaks were mapped; 2% of FOSL1 peaks overlapped with 8% of TEAD4 peaks and **(J)** 27,537 JUND peaks and 10,137 TEAD4 peaks were mapped: 3% of JUND peaks overlapped with 7% of TEAD4 peaks.

249

The distribution of ChIP peaks across the four TFs was analysed in VU40T(-) cells to investigate the frequency of binding events **(Figure 5.15)**. Due to similar involvement of the FOSL1 and JUND TFs, the peaks for FOSL1 and JUND were combined and called AP-1. Analysis of peak distribution revealed that nearly half (47.6%) of peaks were bound by p63 only **(Figure 5.15)**. AP-1 only bound peaks were 15.1%, while 26.2% of peaks were bound by p63 and AP-1. It was also observed that a small percentage of peaks bound either only TEAD4, TEAD4/AP-1 or TEAD4/p63 (1.0%, 0.3% and 0.7%, respectively). However, when all four factors are bound, the percentage of peaks increases to 9.1% **(Figure 5.15)**.

Interestingly, the frequency of TEAD4 binding increases the most in the presence of both p63 and both AP-1 factors. Therefore, the localisation of TEAD4 binding with FOSL1, JUND and p63 was investigated in VU40T(-) cells by analysing peak location proximity; p63, FOSL1 and JUND binding was investigated 3kb upstream and downstream of each TEAD4 peak. The analysis revealed that indeed the majority of TEAD4 binding directly overlaps with p63, FOSL1 or JUND **(Figure 5.16A-C)**.
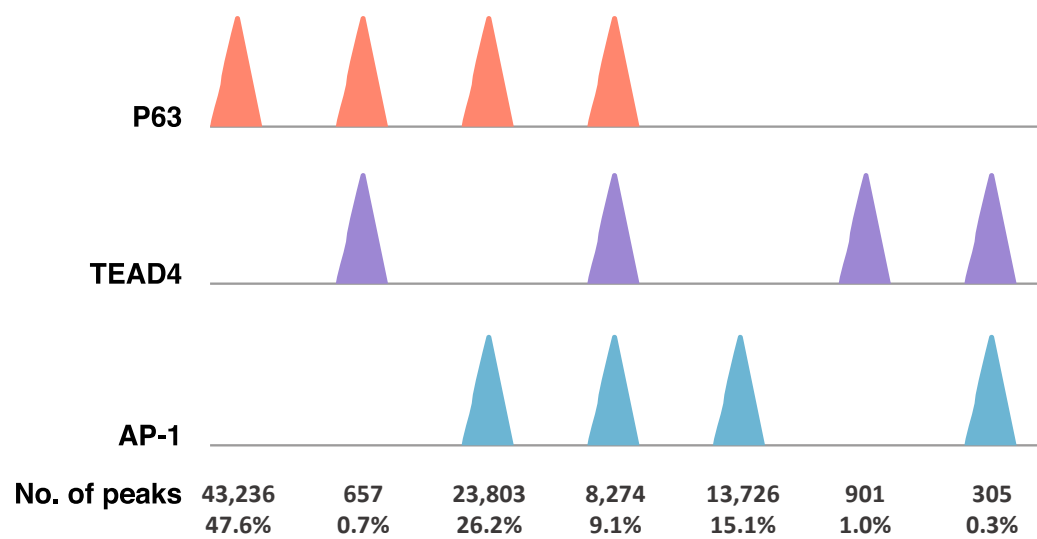
Figure 5.15. ChIP peak distribution of p63, TEAD4 and AP-1 peaks in VU40T(-) cells. The 'mergepeaks' function in HOMER was used to investigate peak location to analyse the distribution of ChIP peaks in p63, TEAD4 and AP-1 (JUND, FOSL1). A merged-ChIP peaks dataset was created to count distinct binding events, as well as co-binding. A total of 90,902 peaks were identified for p63 (43,236 peaks), TEAD4 (901 peaks) and AP-1 (FOSL1 and JUND) (13,726 peaks).
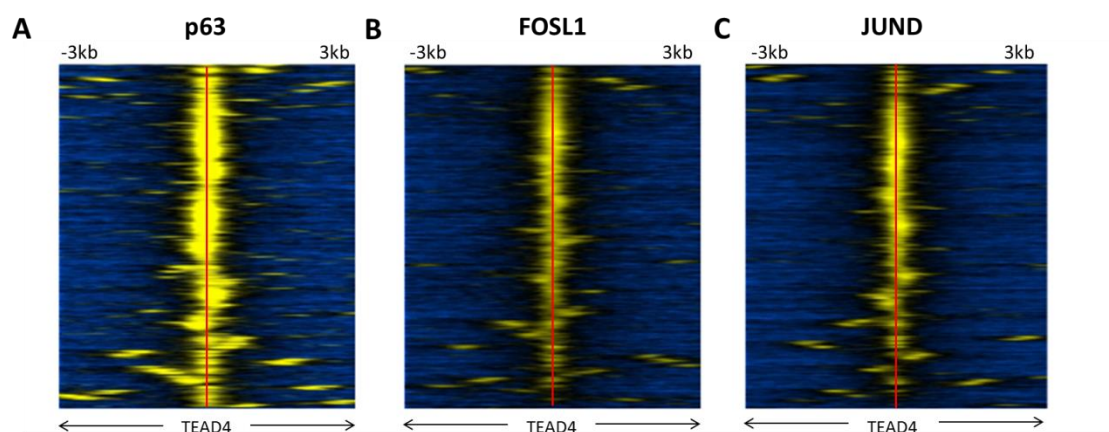
Figure 5.16. Visual representation of co-localisation of TEAD4 binding in VU40T(-) peaks with p63, FOSL1 and JUND peaks.

By using the 'AnnotatePeaks' tool in HOMER, TEAD4 peaks (centre of peaks visualized by red vertical line) were annotated 3kb upstream and downstream with **(A)** p63, **(B)**, FOSL1 and **(C)** JUND reads (indicated by yellow). The resulting matrix file was clustered using the GeneCluster 3.0 program and the following settings: 'log transform data', 'center genes', 'Cluster', 'Correlation (centered)' and 'Centroid linkage'. Heatmap images were created using the TreeView program (Dr Sam Clokie).

5.8 Discussion

In silico analysis of genetic alterations in the TCGA (The Cancer Genome Atlas) panel of HNSCC (head and neck squamous cell carcinoma) tumours revealed the *TP63* gene to be one of the most frequently altered genes, agreeing with previously published data [66-70, 72]. Further to this, our data revealed that nearly 50% of the HNSCC cell line regulatory network identified by chromatin immunoprecipitation-sequencing (ChIP-seq) is regulated by p63. Amplification of the p63 gene has been linked to HNSCC tumours by several studies [12, 13, 16, 257, 258]. p63 overexpression can lead to inhibition of epithelial differentiation and uncontrolled self-renewal of keratinocyte cells, which could lead to cancers such as HNSCC [16]. Therefore, deregulation of p63 may be one of the mechanisms by which the HNSCC cell lines investigated here function.

Upon the analysis of the p63, FOSL1 and JUN VU40T(-) and VU147(+) ChIP-seq datasets, more peaks were called in the VU40T(-) ChIP samples compared to the VU147(+) samples. A reason for this could simply be that the TFs bind more frequently in VU40T(-) cells than in VU147(+) cells. This could suggest differential expression; however, our gene expression data indicates that these TFs are expressed at similar levels. In addition, the number of DNase I-hypersensitive sites were similar in both cell lines, indicating that the same number of open sites are available for TF binding to occur. Technical differences may be a more likely explanation for the discrepancy in the number of ChIP-seq peaks. For example, inefficiency of the formaldehyde cross-linking step can lead to fewer protein-DNA complexes being preserved. As a result, these regions are sequenced less frequently within the cell population, leading to fewer peaks being called. To overcome this,

the sequencing depth could be increased to improve the mapping of such sites. A further reason may be higher background in the VU147(+) input sample compared to the VU40T(-) input, which could lead to the identification of fewer peaks in the ChIP samples upon normalisation to input. It could be argued that perhaps VU147(+) cells exhibit greater genomic instability compared to VU40T(-) cells, leading to the discrepancy in ChIP-seq peak numbers. However, all ChIP samples are normalised against a corresponding input sample to control for genome amplifications and deletions. In addition, a false discovery rate is applied to allow control over false positive peak predictions. Therefore, without conducting further experiments, this method of analysis may not be suitable for comparison of specific TF binding sites between different ChIP samples and cell lines. However, in order to investigate genome-wide processes by, for example, pathway enrichment analysis, this method has been useful in highlighting processes relevant to HNSCC, as similar pathways were identified in both cell types, as discussed next.

Our initial hypothesis of AP-1 and TEADs being involved in HNSCC cell line gene regulation was also confirmed by our ChIP-seq data. Gene expression analysis by qPCR revealed expression levels of *FOSL1* and *JUND* to be among the highest. In addition, *FOSL1* was the most amplified AP-1 gene in the TCGA cohort. In HNSCC cell lines, FOSL1 and the JUN family of proteins have previously been found to promote expression of the IL (interleukin)-8 protein, which can promote tumour growth, invasion and metastasis [255, 323]. To support these findings, our AP-1 ChIP-seq data reveal the involvement of pathways associated with cell growth, migration and invasion (Hippo, adherens junction, RAP1 signalling and α6β4 integrin signalling pathways). Mechanisms by which AP-1 may induce pro-tumourigenic effects such as invasion and cell growth have been reported

previously. In mouse models it has been shown that a reduction in Jun and Fosl1 protein expression through knockdown of the histone demethylase enzyme Kdm4a can inhibit cell invasion and tumour growth [252]. In addition, JUNB knockout in HNSCC cell lines has been shown to reduce cell invasion and migration [253]. Further to this, published studies have reported FOSL1 protein overexpression in HNSCC cells to be associated with processes such as cell adhesion, proliferation and migration through activation of the PI3K (phosphotidylinositol-3 kinase)/AKT pathway [254, 256]. Our data supports this finding, as AP-1 ChIP-seq revealed strong over-representation of the PI3K/AKT pathway.

In summary, our AP-1 ChIP-seq and gene expression data in HNSCC cell lines, as well as analysis of genetic alterations in the TCGA HNSCC tumours, agree on the involvement of AP-1 in HNSCC. This is also supported by previously published data which demonstrates the role of AP-1 in tumour growth, invasion, migration and metastasis.

One of the mechanisms by which the Hippo pathway has been recognized in HNSCC is by associations between YAP (yes-associated protein) expression and HNSCC tumours [87, 324, 325]. For example, YAP protein expression has been shown to correlate well with HNSCC tumour grade and metastasis [87]. Furthermore, overexpression of the *YAP1* gene has been reported to positively correlate with poor prognosis in HNSCC patients [324]. The analysis of the TCGA cohort showed that the *YAP1* gene was altered in 17% of tumours and even more frequent upregulation was observed for *WWTR1* (22%), which encodes the TAZ (tafazzin) protein. The functions of YAP and TAZ have been investigated in kidney epithelial cells. Read outs such as transcription of Hippo target genes, cell size and migration were measured and it was found that YAP is largely able to compensate

for the loss of TAZ [326]. Interestingly, in HNSCC, the alterations in YAP and TAZ appear to be largely mutually exclusive, and at least 39% of HNSCC tumours are characterized by an alteration in one of the genes. Therefore, it may be that YAP could be a more effective therapeutic target than TAZ. In addition to genetic alterations of *YAP1* and *WWTR1*, the *TEAD4* gene was altered in 10% of the TCGA cohort and upregulation of the *TEAD4* transcript was observed in our HNSCC cell lines compared to the non-tumourigenic OKF6 cells. This agrees with published data in which upregulation of both *TEAD4* RNA and protein levels have been observed in HNSCC cell lines compared to primary human normal oral keratinocyte (HNOK) cells [316]. Furthermore, in the same study, TEAD4 protein overexpression was also observed in HNSCC tissues compared to surrounding normal tissue [316]. Therefore, genetic amplifications occur to a large degree in the Hippo pathway, deregulating expression of Hippo target genes and affect transcriptional regulators (YAP and TAZ) more than TFs themselves (TEADs).

As well as investigations into YAP and TEAD separately, their association with each other has also been established by previous studies. For example, using TCGA data, it has been shown that both *TEAD4* RNA levels and YAP protein, correlate well with HNSCC tumour onset and levels increase with higher tumour grade and stage [327]. Furthermore, in HNSCC cell lines it was shown that TEAD knockdown can block activation of YAP target genes, thereby slowing cell growth [316, 328]. Another study reported YAP/TAZ target genes to be involved in HNSCC tumour progression, among which were, *TEAD1*, *TEAD4*, *ETS1* and *JUN* [327]. Therefore, in recent years, the involvement of the Hippo pathway at the transcriptional level in HNSCC has been demonstrated, agreeing with the observations in our data.

To support the role of YAP and TEAD in tumour progression, our TEAD4 ChIP-seq pathway enrichment analysis revealed that TEAD4 forms a regulatory network involved in the Hippo, focal adhesion and adherens junction pathways. Disruption of cell-cell and cell-ECM (extracellular matrix) junctions have been demonstrated to be a hallmark of cell invasion, metastasis and tumour progression [329]. Therefore, overall, our ChIP-seq data strongly support the involvement of p63, AP-1 and TEAD in HNSCC and in enhancing HNSCC-related signalling pathways. The analysis also shows large degree of overlap and therefore potential co-regulation within p63/AP-1/TEAD regulatory network.

Our ChIP-seq data have demonstrated significant AP-1 and p63 co-binding, in addition, the data reveals the AP-1 motif to be enriched in p63 ChIP-seq motif analysis. This has also been observed by Saladi *et al.*, 2017, where p63 ChIP-seq data conducted in a laryngeal SCC cell line revealed the FOSL1 motif to be amongst the top enriched motifs [138]. This finding is supported by Si *et* al., 2016, in a different laryngeal SCC cell line, TF motif analysis of p63 ChIP-seq data revealed high enrichment of the AP-1 motif [330]. Moreover, Saladi *et* al., 2017 demonstrated by luciferase reporter assays the induction of the AP-1 response element by ΔNp63α and upregulation of the *FOSL1* gene by microarrays [330]. Furthermore, this study also used three publicly available microarray datasets of a total 125 HNSCC tumour samples and conducted *in silico* analysis. A statistically significant correlation was found between *TP63* and *JUNB*/*FOSL1* gene expression [330].

ChIP-seq data also revealed a particularly strong overlap of TEAD4 peaks with p63 peaks and revealed enrichment of TEAD and p63 motifs in both datasets, indicating association between the TFs. Published data investigating the role of TEAD and

p63 is lacking, however it has been demonstrated that TEAD can bind to the *ΔNp63* promoter to induce its repression [331]. Associations between p63 and YAP have also been reported in HNSCC. For example, the ΔNp63 protein can directly bind the *YAP1* promoter to inhibit its expression [138, 139]. Furthermore, a study conducted in a laryngeal SCC cell line reports that p63 knockdown can reduce YAP-dependent oncosphere formation [138]. While on the other hand, it has been demonstrated that YAP can promote the degradation of p63 in HNSCC cells [87, 332, 333]. In addition, YAP/TEAD interactions have been commonly associated with EMT (epithelial to mesenchymal transition), cell transformation and cell growth, which are processes highlighted by our pathway enrichment analyses [272, 334, 335]. Therefore, published data demonstrates that p63, YAP and TEAD can influence each other's expression, moreover the proteins can work together to promote HNSCC tumour progression. It may be that these effects are induced by YAP/TEAD interactions, however it must be considered that TEAD is not the only transcription partner for YAP, as KLF (Krüppel-like factor)-4, RUNX (runt-related transcription factor) and p73 can also interact with YAP [272].

Furthermore, our data revealed overlap between TEAD4 and AP-1 binding, in addition, motif analysis of AP-1 and TEAD ChIP-seq data, showed enrichment of both AP-1 and TEAD motifs in both ChIP experiments. This finding is supported by investigations conducted by other groups. Lui *et al.*, 2016 conducted motif analysis of JUND and FOSL1 ChIP-seq peaks in cell lines derived from four different cancers: lung adenocarcinoma, colorectal carcinoma, endometrial adenocarcinoma and neuroblastoma. The analysis revealed the TEAD motif to be among the top most enriched motifs in the JUND and FOSL1 peaks [274]. Further to this, co-regulation between YAP and JUND has previously been assessed using

reporter assays. It was found in a colorectal carcinoma cell line that YAP/TAZ knockdown, but not JUND knockdown resulted in decreased TEAD4 expression, while expression levels of Hippo target genes *ANKRD1* (Ankyrin repeat domain containing) and *DOCK9* (dedicator of cytokinesis 9) significantly decreased after TEAD and JUND inhibition [274]. Another study demonstrated by reporter assays, the association of TEAD and AP-1 with the transcriptional regulation of the YAP target genes, *CTGF* and *ANKRD1* [272]. This confirmed co-regulation between TEAD and AP-1. There is also evidence highlighting a central role for TEAD in YAP/TEAD/AP-1-mediated gene regulation. For example, using breast cancer cells, transfection of YAP and JUN/FOS dimers together can increase mammosphere size, while AP-1 alone has no effect. This phenotype was dependent on TEAD, as mutated TEAD yielded no change in phenotype [272]. Therefore, TEAD seems to be central to transcriptional and phenotypic change mediated by YAP/TEAD/AP-1. Altogether our data and published evidence indicates co-regulation between AP-1 and TEAD4 in HNSCC cell lines and other cancers.

In addition to evidence of co-regulation between TEAD and AP-1, TEAD/AP-1 co-binding at regulatory elements has been observed. A study showed that in a lung adenocarcinoma cell line nearly 90% of TEAD4/AP-1 (FOSL1 and FOSL2) peaks overlapped with active enhancer marks (H3K4me1 and H3K27ac) [274]. This finding supports our data which demonstrated TEAD4 and AP-1 motif enrichment in distal DHSs (DNase I hypersensitive sites) of the HNSCC cell lines. Further strengthening this finding, co-regulation between TEAD and AP-1 TFs has also been reported in breast carcinoma cells at active enhancers [272]. JUN ChIP-seq experiments revealed that the vast majority of JUN peaks overlapped with YAP/TAZ/TEAD4 peaks and most were found at active enhancers associated with

the regulation of cell proliferation [272]. Altogether, the data suggests co-regulation and perhaps interaction between TEAD and AP-1 at regulatory elements to drive transcription of Hippo target genes. In line with this, our pathway enrichment analysis revealed a subset of pathways involving the TEAD4, FOSL1 and JUND TFs, which were largely related to cell adhesion pathways, including the Hippo pathway [88, 272, 274].

In summary, our data and that of others have highlighted AP-1, p63 and TEAD as being important epithelial transcriptional regulators, and in the context of HNSCC are involved in processes such as tumour growth, invasion, migration and metastasis. However, published evidence report associations between two of the three factors investigated and data reporting associations between all three TFs is lacking. In the case of TEAD4, our data indicates a stronger association between the three TFs, than between one or two, as TEAD4 binding events increase by nine-fold when all three TFs bind together compared to one or two TFs binding in the case of TEAD4 (**Figure 5.17**).

Although the three TF groups appear to be key for HNSCC tumour and/or cell line regulation, there are possibly other factors contributing to this network. Analysis of p63-specific ChIP-seq peaks demonstrated enrichment of the KLF5 motif. Associations between p63 and KLF5 have previously been identified and found to induce tumour-promoting effects. For example, KLF5 overexpression has been reported by EC (enhancer cluster) activation in HNSCC cell lines, leading to activation of cancer-related genes, including *TP63* [179]. Agreeing with this, our RNA-seq data reveals that *KLF5* was the most highly expressed KLF gene in all four HNSCC cell lines, on the other hand, *KLF5* was overexpressed more in the RNA-seq data for the non-tumour HOK (human oral keratinocyte) cells (data not

shown). However, in mice studies of epidermal keratinocyte cells, Klf5 overexpression studies have shown to decrease levels of *TAp63* and *ΔNp63* leading to abnormal epithelial stratification and hypoproliferation of keratinocytes [336]. This contradiction in findings may be due to the study being conducted during early development, during which Klf5 levels are restricted and cell maintenance can be taken over by other members of the Klf family [337].

It appears that TEAD4 is an important part of the described regulatory network, as evidenced by visual representation of p63, FOSL1 and JUND co-localisation with TEAD4 (**Figure 5.16**), as well as TCGA data highlighting that nearly 70% of HNSCC tumours contain an alteration in the gene of a Hippo transcriptional regulator (*YAP*, *WWTR1*, *TEAD1-4*) (**Figure 5.2**). Therefore, the Hippo pathway may present a therapeutic target in HNSCC, which is explored in the next chapter.

A limitation of this chapter is the ChIP-seq analysis of one oral SCC cell line (VU40T(-)), namely the TF motif analysis and the investigation of the AP1-p63-TEAD4 network. Further work could include the analysis of the VU147(+) ChIP-seq datasets for p63, FOSL1, JUND and TEAD4 ChIPs. In addition, the use of HNSCC tumour samples or, as mentioned before, increasing the number of oral SCC cell lines investigated would strengthen the data. Furthermore, determining the samples' HPV status could allow for the identification of differences in the AP1-p63-TEAD4 TF network in the context of HPV infection.
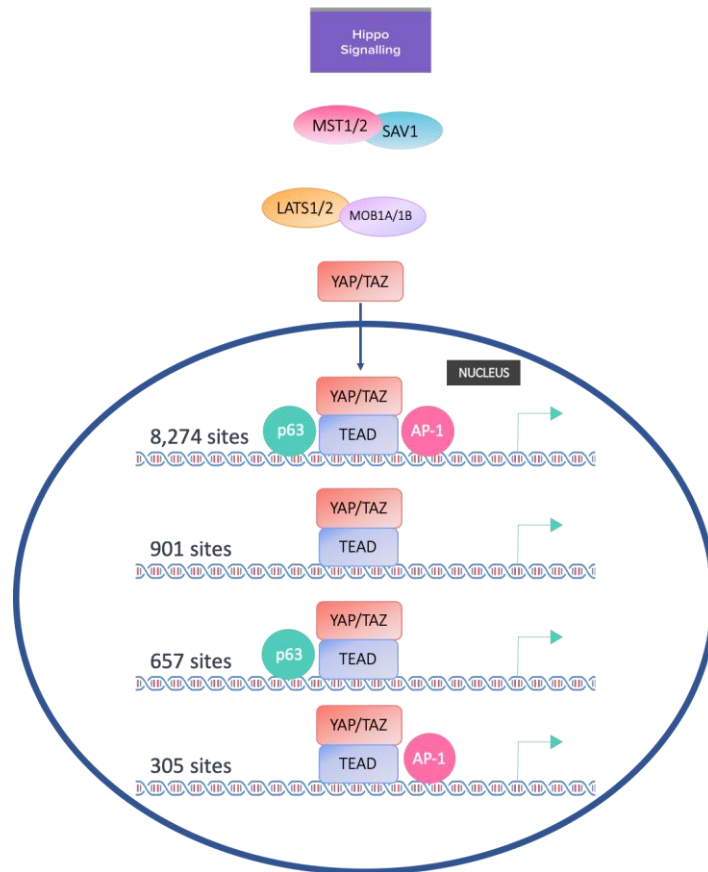
Figure 5.17. Summary of p63, TEAD4 and AP-1 (FOSL1/JUND) binding events.
In environments of low cell density, the Hippo pathway is switched off, leading to dephosphorylation of the core kinases (MST1/2 and LATS1/2), the scaffold proteins (SAV1 and MOB1A/1B) and YAP/TAZ [88, 92]. YAP/TAZ are then free to translocate to the nucleus and activate gene transcription of Hippo target genes such as *CTGF*, which is involved in processes such as cell proliferation and growth [88, 92]. Our data reveals that the total number of ChIP-seq binding events involving TEAD4/FOSL1/JUND (AP-1) were 305, p63/TEAD4 were 657, TEAD4 only were 901 and p63/TEAD4/AP-1 were 8,274. Our ChIP-seq data indicates a stronger association between three TFs, than between one or two, as evidenced by the increase of TEAD4 binding events by nine-fold in the presence of AP-1 and p63. Therefore, it may be that p63, TEAD4, FOSL1 and JUND can regulate Hippo target genes together, perhaps by recruitment of TEAD4 to the chromatin by p63 and/or AP-1.

CHAPTER 6: INVESTIGATION OF THE HIPPO PATHWAY AS A POTENTIAL THERAPEUTIC TARGET IN HNSCC

6.1 Introduction

The previous chapters have highlighted the p63, AP (activator protein)-1 and TEAD (TEA domain transcription factor)-4 transcription factors (TFs) to be important in the function of our HNSCC (head and neck squamous cell carcinoma) cell lines. In addition, the TEAD4 TF has been observed to be important to the HNSCC regulatory network. Furthermore, the discussions of the previous two results chapters have described published data on the role of the Hippo pathway in cell growth and proliferation in HNSCC cells [88, 92]. In line with this, a potential therapy for HNSCC may be the pharmacological disruption of the Hippo pathway, as discussed next.

6.1.1 Therapeutic targets of the Hippo pathway

In cancer cells, several compounds have previously been identified to activate the Hippo pathway, promoting phosphorylation of its components and inhibiting transcription of Hippo target genes. This resulted in decreased cell growth and proliferation [92, 302, 303]. The compounds can be categorised into three main groups based on where in the pathway they cause their effect: (1) activators of Hippo kinases (MST (macrophage stimulating)-1/2 and LATS (large tumour suppressor protein kinase)-1/2), (2) activators of transcription mediator proteins

YAP/TAZ (tafazzin), and (3) inhibitors of YAP/TAZ/TEAD interactions (**Figure 6.1A**).
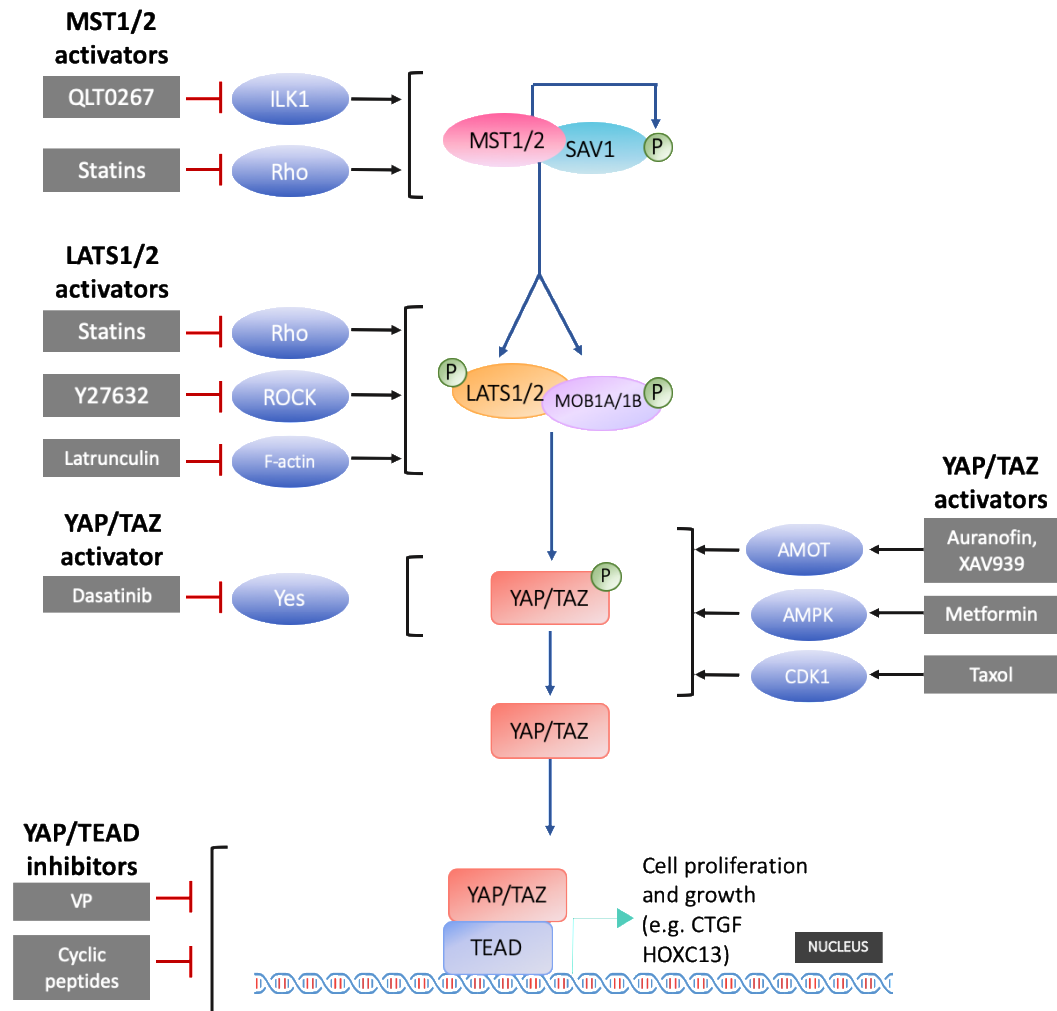
Figure 6.1. Druggable targets of the Hippo pathway.
Activators of the Hippo pathway can be classed into four main groups: (i) MST1/2 activators promote MST1/2 phosphorylation: **QLT0267** inhibits ILK1 and **statins** inhibit Rho GTPases, (ii) LATS1/2 activators promote LATS1/2 phosphorylation: **statins** inhibit Rho GTPases, **Y27632** inhibits ROCK and **latrunculin** inhibits F-actin and, (iii) YAP/TAZ activators promote YAP/TAZ phosphorylation: **dasatinib** inhibits the YES kinase protein, **auranofin** and **XAV939** promote AMOT proteins, **metformin** activates AMPK and **taxol** activates CDK1.
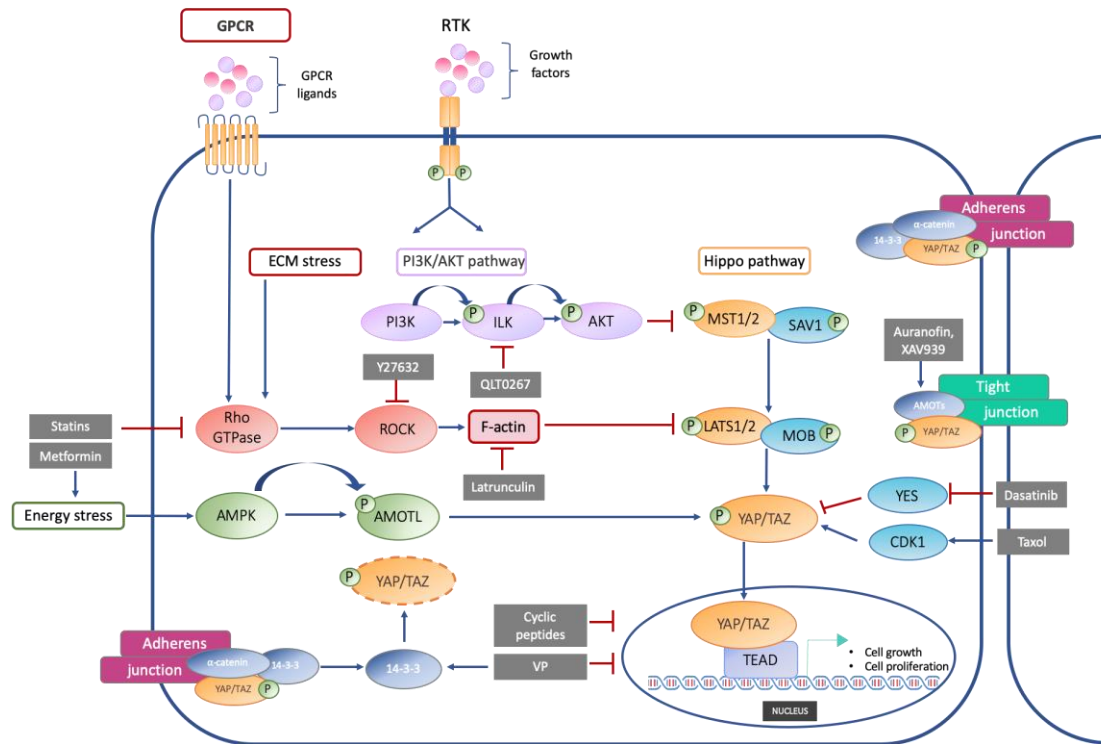
Figure 6.2. Druggable targets of the Hippo pathway and their mechanisms.
Activation of GPCRs (G-protein coupled receptors) leads to activation of Rho GTPases and in turn ROCK. ROCK has a role in F-actin assembly and F-actin monomers can inhibit LATS1/2 phosphorylation allowing YAP/TAZ nuclear localisation. Binding of growth factors to receptor tyrosine kinases can activate the PI3K/ILK/AKT (protein kinase B) pathway. PI3K can phosphorylate ILK, which then phosphorylates AKT. AKT then phosphorylates a site on MST1/2 which can block its kinase activity. Energy stress activates the AMPK kinase, which phosphorylates AMOTL, which in turn phosphorylates YAP/TAZ. AMOT proteins can bind to and sequester phosphorylated YAP/TAZ at tight junctions. The YES protein kinase phosphorylates YAP/TAZ promoting nuclear localisation to initiate gene transcription. On the other hand, CDK1 phosphorylates YAP/TAZ to promote its cytoplasmic retention.

Many of the compounds are able to disrupt the Hippo pathway through interference of the cytoskeletal network. For example, an activator of MST1/2 called QLT2067, functions by inhibiting ILK1 (integrin-linked kinase 1) [303] **(Figure 6.1)**. ILK1 is part of the PI3K (phosphotidylinositol-3 kinase)/ILK/AKT axis where PI3K can phosphorylate ILK1, which in turn phosphorylates AKT leading to phosphorylation of a MST1/2 site, inhibiting its kinase activity [338-340] **(Figure 6.2)**. A second type of MST1/2 inhibitor exploits mechanical ECM (extracellular matrix) stress as being a trigger of the Hippo pathway. Rho (rhodopsin) GTPases and ROCK (Rho-associated coiled-coil containing protein kinase) are required for assembly and disassembly of actin filaments during ECM stress [341] **(Figure 6.2)**. To achieve this, Rho GTPase and ROCK can inactivate LATS1/2 and MST1/2 stimulated by an upstream cholesterol synthesis pathway [342]. Drugs such as statins can disrupt this process by inhibiting Rho GTPases leading to activation of MST1/2 and LATS1/2, while a compound called Y27632 can inhibit ROCK, which promotes LATS1/2 phosphorylation [92, 302, 303] **(Figure 6.1 and 6.2)**. A third drug called latrunculin can bind and inhibit polymerisation of F-actin monomers which normally inactivate LATS1/2 [92, 302, 303] **(Figure 6.1 and 6.2)**.

Compounds which target YAP/TAZ exert their effect by promoting YAP/TAZ phosphorylation and cytoplasmic retention. Therapeutic activators of YAP include dasatinib which can inhibit the YES kinase protein **(Figure 6.1 and 6.2)**. YES phosphorylates a YAP tyrosine site, which promotes YAP nuclear localisation [302, 303, 342, 343]. A second mechanism by which YAP/TAZ can be inhibited is by stabilisation of a family of tight junction scaffold proteins called AMOT (angiomotin) (AMOT, AMOTL1 and AMOTL2). AMOT proteins can sequester YAP/TAZ to tight junctions and prevent YAP/TAZ nuclear localisation and

activate LATS1/2 phosphorylation [303] **(Figure 6.2)**. Two compounds (Auranofin and XAV939) can promote the stabilisation of AMOT proteins thereby preventing translocation of YAP/TAZ to the nucleus [303, 344] **(Figure 6.1)**. Metformin treatment can result in energy stress, leading to activation of AMP kinase (AMPK) which in turn phosphorylates AMOTL, promoting YAP/TAZ cytoplasmic retention [303] **(Figure 6.1 and 6.2)**. Taxol can activate cyclin dependent kinase 1 (CDK1), which phosphorylates YAP, promoting its cytoplasmic retention, while phosphorylation of TAZ promotes its degradation [303, 345] **(Figure 6.1 and 6.2)**.

The third group of compounds function to disrupt YAP/TAZ/TEAD interactions. Verteporfin (VP) is an FDA (food and drug administration) approved photosensitizer used in the treatment of macular degeneration, but it has also been found that VP can disrupt the binding of YAP to TEAD [346-350]. VP has been shown to reduce nuclear YAP levels and increase levels of 14-3-3, which sequesters YAP/TAZ to adherens junctions and targets it for degradation [303, 342, 343, 351] **(Figure 6.1 and 6.2)**. In addition, it has been shown that VP can directly bind to YAP to disrupt its interaction with TEAD [348, 349, 352]. Cyclic peptides have also been designed to specifically inhibit YAP/TEAD interaction. Zhou *et al*. report a YAP-like peptide, which binds to TEAD blocking YAP/TEAD interaction [353] **(Figure 6.1 and 6.2)**.

6.1.2 Cross-talk between the Hippo and EGFR pathways

In HNSCC, EGFR (epidermal growth factor receptor) overexpression is observed in up to 90% of HNSCC tumours [36, 70, 91]. However, despite EGFR overexpression being observed in the majority of HNSCC tumours, only 10-20%

of patients respond to treatment with cetuximab, a monoclonal antibody, which binds and inhibits the EGFR ligand-binding domain [109, 354].

Investigations of our cell lines and *in silico* analysis of TCGA (The Cancer Genome Atlas) primary HNSCC tumours have highlighted both the Hippo pathway and EGFR to be significantly overexpressed. Furthermore, as the previous chapters have demonstrated, the TEAD4 TF is closely involved in other elements of the regulatory network, including those regulating EGFR response. Therefore, it was hypothesised that investigations into Hippo/EGFR pathway cross-talk may offer an approach to attenuate cetuximab-resistance. To support this hypothesis, a study has described a feedback loop between the EGFR and Hippo pathways in cervical SCC [355]. The study reported that constitutively active or overexpressed YAP led to overexpression of the EGFR ligands, *AREG* (amphiregulin), *EGFR* and *TGF* (transforming growth factor)-α (tumour growth factor α) and treatment with TGF-α or AREG resulted in activation of EGFR/PI3K/MAPK and inactivation of the Hippo pathway through EGFR phosphorylation and YAP dephosphorylation, respectively [355]. This resulted in increased migration and proliferation of cervical cancer cells, suggesting a positive feedback loop between the EGFR and Hippo pathways [355]. Additionally, co-treatment of cervical SCC cells with AG1478 (EGFR tyrosine kinase inhibitor) and VP resulted in inhibition of colony formation and reduction in levels of AREG [355]. Subsequent to the initiation of this PhD project, cross-talk between the Hippo and EGFR pathways has also been observed by other groups in oesophageal, lung and kidney carcinomas [356-358].

6.1.3 Aims

Our data and published data suggest an interaction between the Hippo and EGFR pathways. As HNSCC patients often display considerable cetuximab resistance, the primary aim of this chapter was to investigate the effect of Hippo inhibition in HNSCC cells. The secondary aim was to observe whether disruption of both the Hippo and EGFR pathways could attenuate cetuximab resistance.

6.2 The effect of Verteporfin (VP) on VU40T(-) and VU147(+) HNSCC cell lines

The first experiment was designed to examine whether disruption of the Hippo pathway can affect HNSCC cell line viability. For these experiments, verteporfin (VP) was used, as it offers the advantage of disrupting the Hippo pathway at the transcription level.

6.2.1 The effect of VP treatment on HNSCC cell line viability

A cell viability assay was conducted in VU40T(-) cells to find the most effective VP treatment conditions. The cells were treated with vehicle and a serial dilution of eight concentrations of VP between 0.72µM and 0.00563µM (**Figure 6.3**). The absorbance levels, which are proportional to the metabolic activity of the cells, were measured after 24, 48, 72 or 96hrs of VP treatment and the values were normalized to the vehicle control (**Figure 6.3**).
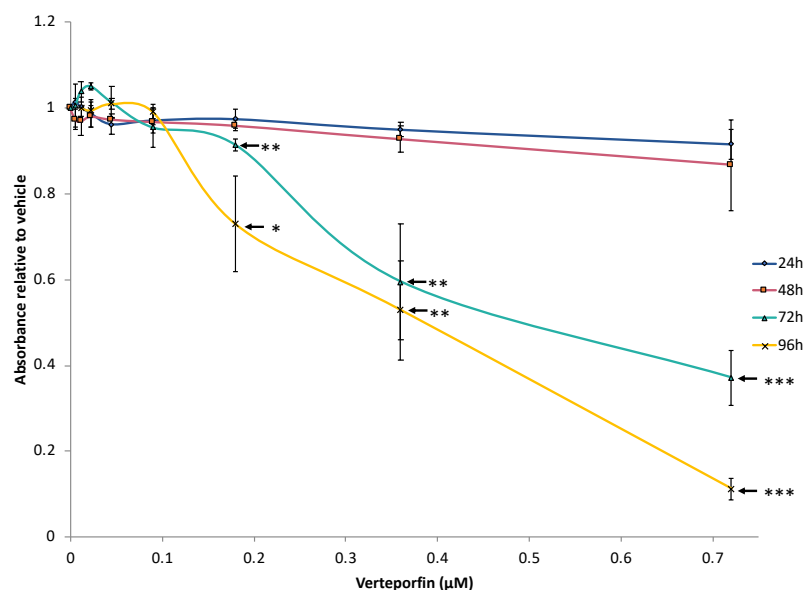
Figure 6.3. VP treatment reduces VU40T(-) cell viability in a dose-dependent manner. VU40T(-) cells were treated with vehicle (DMSO), 0.72, 0.36, 0.18, 0.1, 0.05, 0.023, 0.011 or 0.00563µM of VP. Absorbance levels were measured every 24hrs for 96hrs. A significant reduction in cell viability was observed at 72hrs and 96hrs post-treatment. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test (*P<0.05, **P<0.005, ***P<0.0005).

After 24 and 48hrs of VP treatment there was little observable effect on cell viability even at the highest concentration. It could be that the drug was exerting a cytostatic effect, however, this cannot be tested using the cell viability assay alone. A significant reduction in viability was observed after 72hrs and 96hrs for the three highest VP concentrations (0.72µM and 0.36µM) **(Figure 6.3)**. As the largest observable decrease in cell viability was after 96hrs, in all further experiments the cells were treated with VP for 96hrs.

In order to find the 20% inhibitory concentration (EC20) of VP in VU40T(-) and VU147(+) cells, they were treated for 96hrs with the six highest VP concentrations described in Figure 6.2. A dose-response curve was plotted to establish the EC20 and EC50 for VU40T(-) and VU147(+) cells, as described in Materials and

Methods, Section 2.16 (**Figure 6.4A and 6.4B**). The EC20 was established at

0.38µM and 0.08µM in VU40T(-) and VU147(+) cells, respectively. The EC50 was

established at 0.81µM and 0.15µM, respectively. Therefore, VU147(+) cells

demonstrated higher sensitivity to VP treatment compared to VU40T(-) cells.
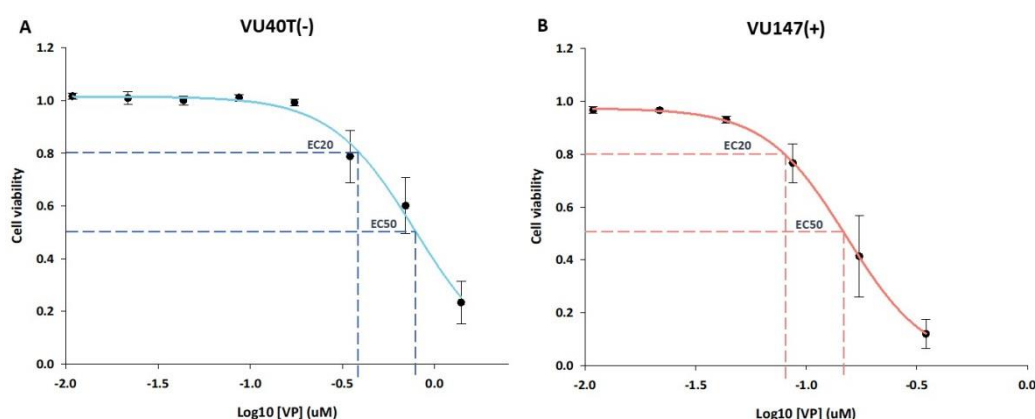
Figure 6.4. Cell viability dose-response curves for cell viability after VP treatment.
SigmaPlot was used to plot sigmoidal dose-response curves based on cell viability in cells
treated for 96hrs with vehicle (DMSO), 0.72, 0.36, 0.18, 0.1, 0.05 or 0.023µM VP in (**A**)
VU40T(-) cells and (**B**) VU147(+) cells. EC50 and EC20 were calculated by determining
the concentrations which reduce viable cells by 50% or 20%, respectively. The EC20 was
established at 0.38µM and 0.08µM, the EC50 was established at 0.81µM and 0.15µM for
VU40T(-) and VU147(+) cells, respectively. The error bars represent SEM of biological
triplicates.

6.2.2 The effect of VP treatment on apoptosis and cell cycle distribution in VU40T(-) cells

To establish the effect of VP treatment on apoptosis and the cell cycle profile of VU40T(-) cells, two sets of FACS (fluorescence-activated cell sorting) analyses were conducted. VU40T(-) cells were treated with EC20 VP for 96hrs. The proportion of VU40T(-) cells in each cell cycle phase was determined by propidium iodide staining **(Figure 6.5A)**. No change in the percentage of cells in S phase was found after VP treatment (15.6% vs. 14.9% in control and treated, respectively). However, there was a small but significant decrease in cells in G2/M phase after VP treatment; from 20% to 18%, while cells in G1 phase increased after VP treatment from 59.9% to 63.6% **(Figure 6.5A)**. Therefore, it could be that VP treatment in VU40T(-) cells can alter the cell cycle distribution. However, the observed change was very small and further experiments to support this data would need to be conducted.

In the second FACS experiment the effect of VP treatment on apoptosis was tested. VU40T(-) cells were stained with propidium iodide and FITC (fluorescein)-conjugated annexin V **(Figure 6.6A and B)**. A reduction in viable cells after VP treatment was observed, from 87.1% to 77.6%, however this reduction was not statistically significant **(Figure 6.6C)**. The percentage of cells in early apoptosis before and after treatment was similar (0.5% and 0.3%, respectively) **(Figure 6.6C)**. There was a statistically significant increase in the percentage of cells in late apoptosis after VP treatment from 8.8% to 12.7% **(Figure 6.6C)**. An increase in necrotic cells was also observed after VP treatment from 3.6% to 9.3% **(Figure 6.6C)**. Therefore, the data suggests that VP may be able to promote apoptosis in

VU40T(-) cells. However, the observed changes are small and further experiments

should be conducted to confirm this, as will be discussed later.
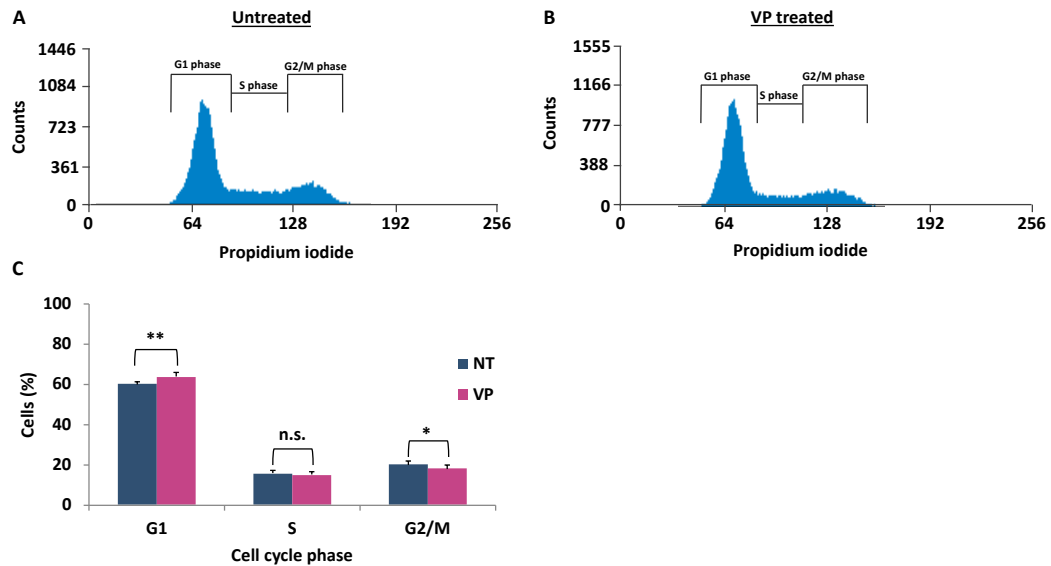
Figure 6.5. VP treatment alters the cell cycle distribution of VU40T(-) cells.
VU40T(-) cells were treated with EC20 VP (0.38μM) or vehicle for 96hrs. Cell cycle was
assessed by propidium iodide staining after **(A)** vehicle or **(B)** VP treatment. The proportion
of cells in G1, S or G2/M phase was measured. **(C)** A significant increase in the proportion
of cells in G1 phase and decrease in the proportion of cells in G2/M was observed after VP
treatment. The error bars represent SEM of 7 biological replicates. P-value was determined
by conducting an unpaired t-test (*P<0.05, **P<0.005, 'n.s' denotes 'not significant').
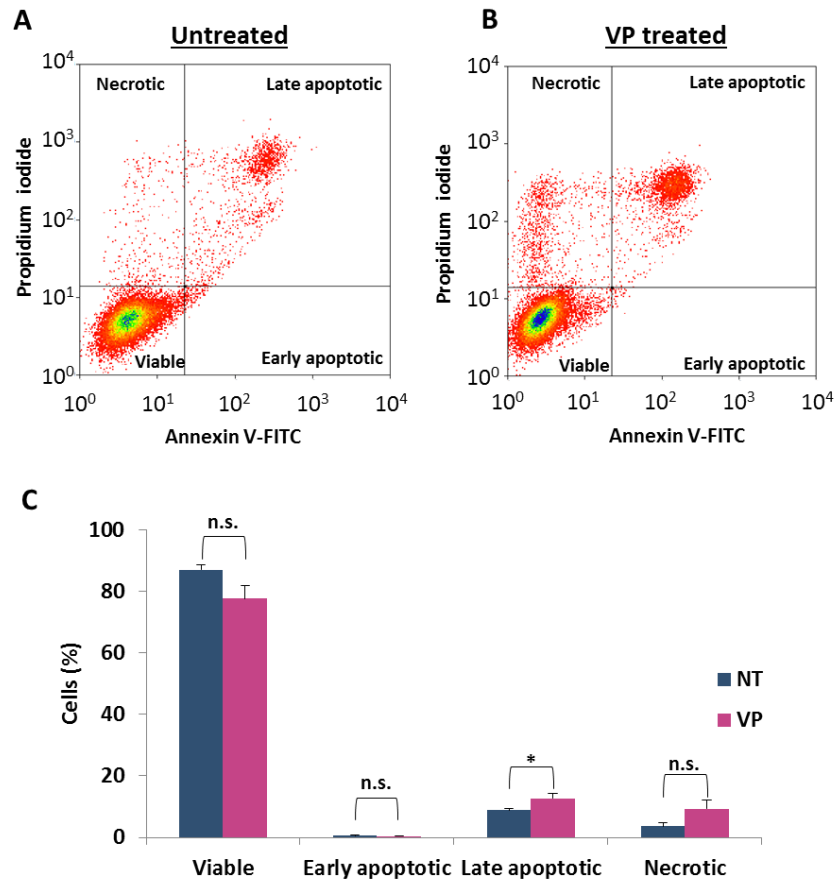
Figure 6.6. VP treatment promotes apoptosis in VU40T(-) cells.
VU40T(-) cells were treated with EC20 VP (0.38μM) or vehicle for 96hrs. Cells were double stained with annexin-FITC V and propidium iodide to measure the proportion of viable, early apoptotic, late apoptotic or necrotic cells after **(A)** vehicle or **(B)** VP treated cells. **(C)** A significant increase in cells in late apoptosis was observed. The error bars represent SEM of biological triplicates. P-value was determined by conducting an unpaired t-test (*P<0.05, 'n.s' denotes 'not significant').

6.2.3 The effect of VP treatment on TEAD4 chromatin binding in the VU40T(-) cell line

To determine whether VP treatment can disrupt TEAD binding, TEAD4 ChIP (chromatin immunoprecipitation)-qPCR (quantitative polymerase chain reaction) was conducted in VU40T(-) cells after VP treatment **(Figure 6.7)**. VU40T(-) cells were treated for 96hrs with either EC20 VP or vehicle. For each condition three biological replicates of chromatin were collected and in addition to conducting the TEAD4 ChIP, an input sample was collected and a 'no antibody' ChIP control were included. QPCR was carried out on each sample to observe changes in TEAD4 binding after VP treatment. To do this, three TEAD4 binding sites were selected based on three criteria. The first criterion was to use TEAD4 chromatin immunoprecipitation-sequencing (ChIP-seq) peak scores (which are indicative of protein binding) from VU40T(-) ChIP-seq data from the previous chapter. A cut-off score was set to identify a subset of genes where TEAD4 is known to bind. The second criterion was RNA-seq data obtained from VU40T(-) vehicle and VP treated cells (discussed in Section 6.3). Output for differential gene expression analysis was used to obtain the values for log2 fold change between vehicle and VP treated cells. The third criterion was the base mean value (mean number of counts for each gene), which was taken into account to ensure the gene had robust basal expression levels. The *YWHA* (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein)-*E*, KCTD (potassium channel tetramerization domain containing)-2 and *ADAM* (eukaryotic translation initiation factor 3 subunit J)-*9* genes were selected as putative TEAD4 binding sites and qPCR was conducted at these regions **(Figure 6.7)**. The samples were normalised to either VP input or vehicle input and to 'no

antibody' samples. Though not significant, a decrease in TEAD4 binding was observed at the *YWHAE*, *KCTD2* and *ADAM* genes (1.6-, 2- and 20-fold decrease, respectively).
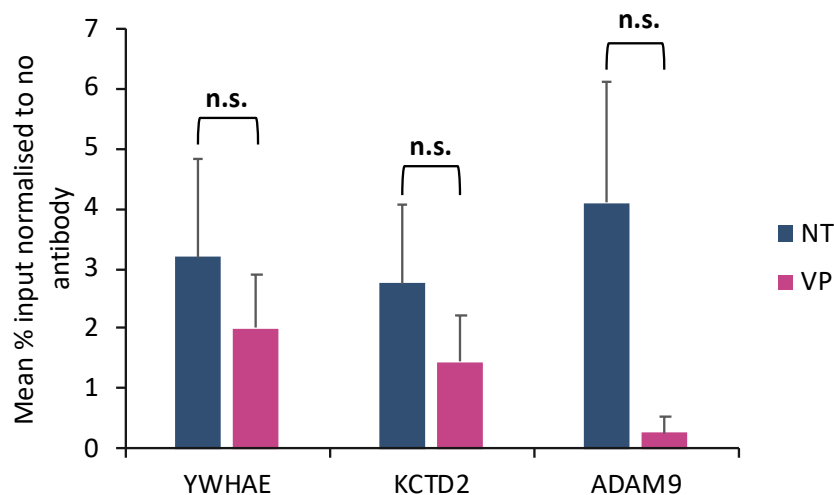
Figure 6.7. TEAD4 ChIP-qPCR analysis in VU40T(-) cells.
VU40T(-) cells were treated with either EC20 VP or vehicle DMSO for 96hrs. TEAD4 ChIP was conducted followed by qPCR at three putative TEAD4 binding sites at the *YWHAE*, *KCTD2* and *ADAM9* genes. ChIP signal was normalised to input to obtain percentage input and this value was normalised to the 'no antibody' signal. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test (n.s denotes 'not significant').

6.2.4 The effect of VP treatment on AP-1 and p63 transcription factor gene expression in VU40T(-) and VU147(+) cell lines

To investigate any changes to the p63/AP-1/TEAD regulatory network as a result of disruption of TEAD4 binding, gene expression levels of the aforementioned TFs were investigated by qPCR after VP treatment. Expression levels of the AP-1 genes were first investigated. In VU40T(-) cells, expression levels of *FOSB* (FosB proto-oncogene, AP-1 transcription factor subunit), *JUNB* (JUNB proto-oncogene, AP-1

transcription factor subunit) and *JUND* (JUND proto-oncogene, AP-1 transcription factor subunit) remained stable, while levels of *FOSL1* (FOS like 1, AP-1 transcription factor subunit) and *JUN* increased and a small decrease in *JUNB* levels were observed **(Figure 6.8A)**. *FOS* (FOS proto-oncogene, AP-1 transcription factor subunit) was expressed at very low levels in VU40T(-) cells. In VU147(+) cells, *JUNB* expression decreased 2.4-fold **(Figure 6.8A)**.

Expression levels of *TP63* were investigated in VU40T(-) cells and an increase of 2.5-fold was observed, while no change in VU147(+) cells was observed **(Figure 6.8B)**. The investigated alterations after VP treatment, however, were not significant.
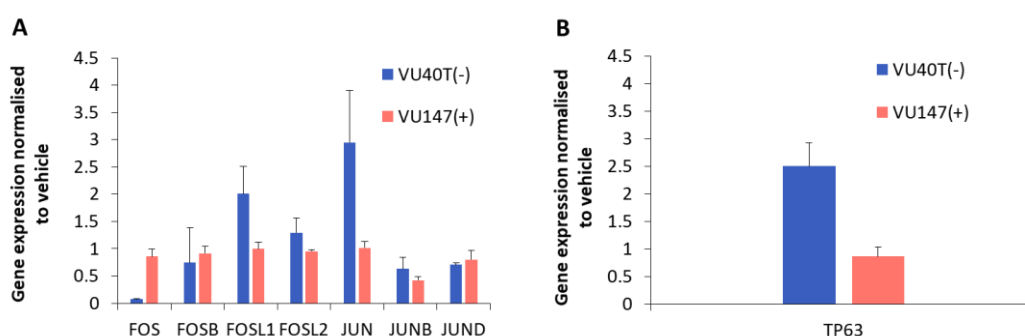
Figure 6.8. QPCR gene expression analysis of genes of the AP-1 and p63 TF genes in VU40T(-) and VU147(+) cells after VP treatment.
VU40T(-) and VU147(+) cells were treated with EC20 VP. Gene expression changes were investigated in **(A)** AP-1 TF genes (*FOS*, *FOSL1*, *FOSL2*, *JUN*, *JUNB*, *JUND*) and **(B)** *TP63*. The error bars represent the SEM between biological triplicates. Statistical significance was calculated using an unpaired t-test. Gene expression changes were not significant between untreated and VP treated samples.

6.3 RNA-seq analysis of VP treated VU40T(-) cells

To explore genome-wide changes in gene expression after VP treatment, RNA-seq was conducted in VU40T(-) cells. RNA was extracted from three biological replicate samples of EC20 VP (0.38μM) treated VU40T(-) cells and of vehicle treated samples. RNA-seq data was analysed as described in Materials and Methods, Section 2.13.5. A total of 7,713 genes were identified in the vehicle sample and 8,038 genes in the VP treated sample. Within the Galaxy platform, the Stringtie file was input into the DESeq2 tool to conduct differential gene expression analysis. Genes whose log2 fold change was >1 or <-1 were taken for pathway enrichment analysis. Pathway enrichment analysis was conducted using the ConsensusPath database and a total of 64 pathways were annotated with the upregulated genes and 83 pathways were annotated with the downregulated genes.

The top 20 pathways for each dataset are presented here (**Figure 6.9**). Among the upregulated pathways was the apoptosis pathway, indicating promotion of anti-tumourigenic effects (**Figure 6.9A**). However, the most prominent pathways were cell adhesion pathways (focal adhesion, integrin mediated cell adhesion, focal adhesion-PI3K/AKT/mTOR (mechanistic target of rapamycin) signalling and NOTCH (neurogenic locus notch homolog) pathways). Downregulated pathways included RNA and protein processing pathways such as, mRNA processing, spliceosome, RNA transport and RNA degradation (**Figure 6.9B**). In addition, HNSCC-related pathways were downregulated such as pRB (retinoblastoma protein) in cancer, EGF-EGFR, MAPK and TGF-β pathways (**Figure 6.9B**). Interestingly, the Hippo pathway was not overrepresented in either dataset.
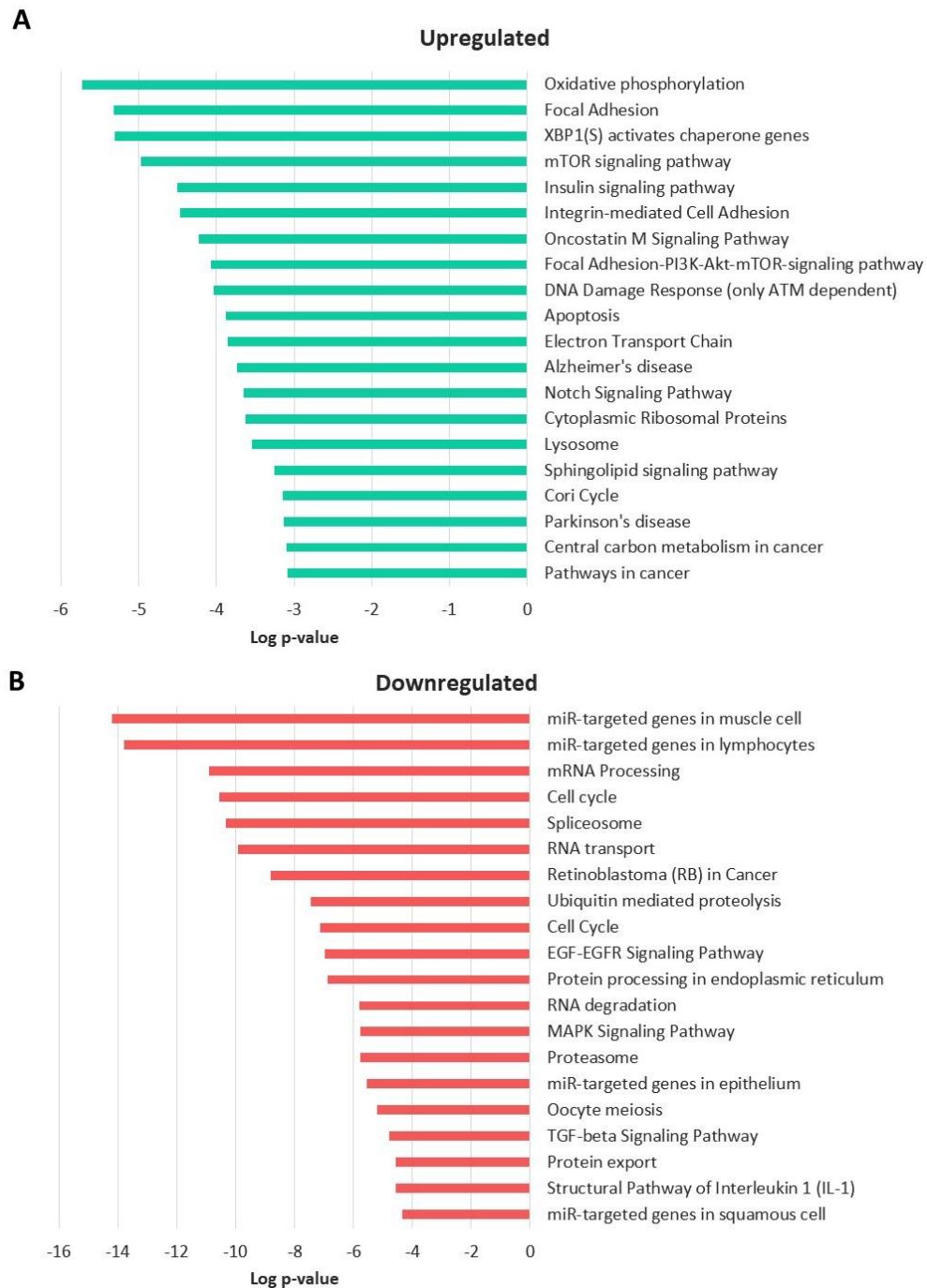
Figure 6.9. Pathway enrichment analysis of RNA-seq data from VP treated VU40T(-) cells

VU40T(-) cells were treated with EC20 VP or vehicle for 96hrs and RNA-seq analysis was conducted. Genes with an FPKM (fragments per kilobase exon per million mapped reads) above 1 were taken forward for pathway enrichment analysis (7,713 genes in the vehicle sample were identified and 8,038 genes in the VP sample were identified). Using the Deseq2 tool, differentially expressed genes were identified: 1,324 genes were upregulated after VP treatment and 1,259 genes were downregulated. Pathway enrichment analysis was conducted using the ConsensusPath database. **(A)** A total of 64 pathways were annotated with the upregulated genes and **(B)** 83 pathways were annotated with the downregulated genes. The top 20 pathways for each dataset are presented here.

6.4 Establishing a VP response in non-tumourigenic primary keratinocyte cells

6.4.1 The effect of VP treatment on non-tumourigenic cell viability

The TEAD4 ChIP-qPCR data demonstrated that VP treatment may disrupt TEAD4 binding, while RNA-seq data revealed that this could result in the downregulation of HNSCC-related processes. Therefore, the next aim was to investigate what effect disruption of TEAD4 binding would have in non-tumour oral cells.

The effect of VP was tested in primary non-tumour cells (human oral keratinocytes (HOKs)) to allow for comparison of VP response between non-tumour cells with HNSCC cells. The same range of VP concentrations previously tested in the HNSCC cells were also used in the HOK cells. The effect on HOK cells was comparable to that of VU147; the EC50 was 0.1μM compared to 0.08μM, respectively **(Figure 6.10)**. VP had the least effect on the viability of the VU40T(-) cells. Therefore, HOK cells are similarly sensitive to VP compared to VU147(+) cells and are more sensitive than VU40T(-) cells.
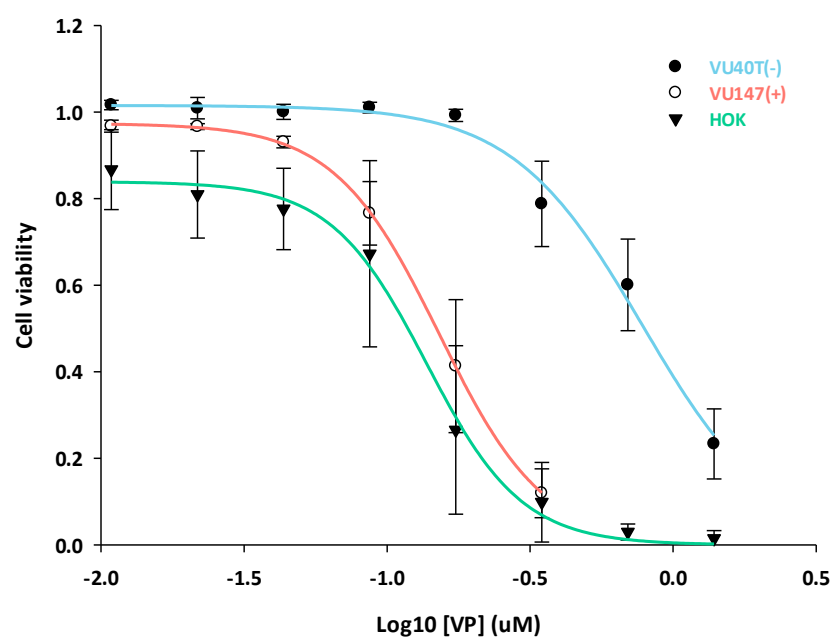
Figure 6.10. VP dose-response curve in HNSCC and non-tumour cells.
VU40T(-), VU147(+) and HOK cells were treated with 0.72, 0.36, 0.18, 0.1, 0.05, 0.023, 0.011 or 0.00563μM of VP. Absorbance levels were measured every 24hrs for 96hrs and normalised to vehicle. SigmaPlot was used to plot sigmoidal dose-response curves. The error bars represent SEM of biological triplicates.

6.5 Investigation of cross-talk between the EGFR and Hippo pathways

6.5.1 Basal EGFR protein levels in HNSCC and non-tumour cells

Due to previously published evidence of Hippo/EGFR pathway cross-talk, the next aim was to observe changes in genes associated with the EGFR pathway after VP treatment [355-358]. First, basal levels of the EGFR protein were examined in the VU40T(-), VU147(+) and in non-tumour HOK cells, and EGFR levels were normalised to lamin A/C housekeeping protein. Levels of EGFR were higher in both the HNSCC cell lines, compared to the HOK cells, agreeing with previously published data [36, 70, 91] **(Figure 6.11A and B)**. Therefore, HNSCC cells express higher EGFR protein levels than do non-tumour cells, indicating activation of the EGFR signalling pathway.

As discussed earlier, a feedback loop between YAP, EGFR and EGFR-like ligands (AREG and EREG) has previously been reported [355]. In order to determine whether a decrease in TEAD4 binding after VP treatment could induce changes in *AREG*, *EREG* and *EGFR* transcript levels in HNSCC cell lines, VU40T(-) and VU147(+) cells were treated with VP and transcript levels were investigated **(Figure 6.12)**. In both VU40T(-) and VU147(+) cell lines, after VP treatment, a small decrease in transcript levels of all three ligands was observed **(Figure 6.12)**. However, this change was not statistically significant and further experiments would be required to conclude the observed reduction in transcript levels of EGF-like ligands after VP treatment.
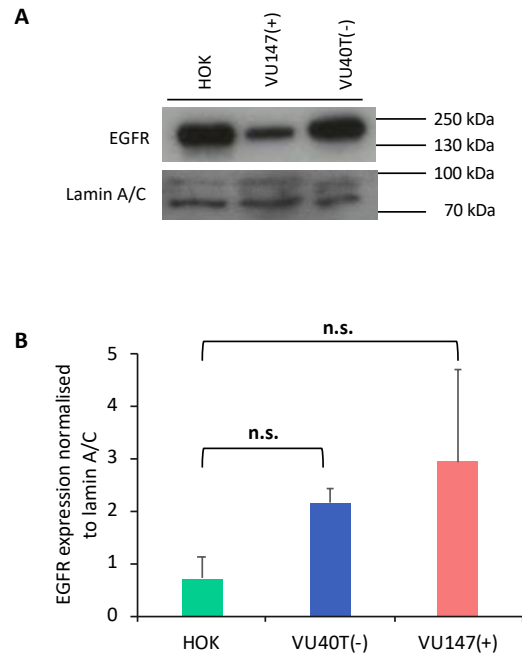
Figure 6.11. EGFR protein expression by western blot in HNSCC cell lines and non-tumour primary cells.
Western blot to show EGFR protein expression in HOK, VU40T(-) and VU147(+) cells. Lamin A/C was used as a loading control. **(B)** HOK, VU40T(-) and VU147(+) EGFR expression levels were normalised to levels of lamin A/C. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test; n.s denotes 'not significant'.
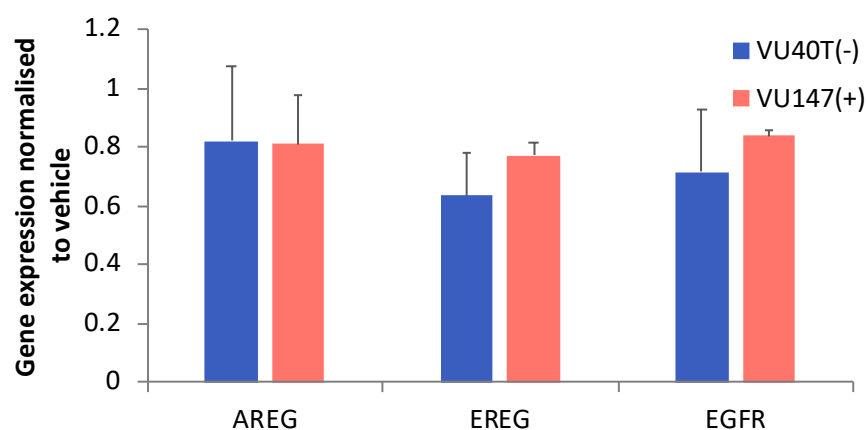
Figure 6.12. QPCR gene expression analysis of genes of the EGFR and EGF-like ligands in VU40T(-) and VU147(+) cells after VP treatment.
VU40T(-) and VU147(+) cells were treated with EC20 VP (0.38µM). Gene expression changes were investigated in *AREG*, *EREG* and *EGFR*. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test. Gene expression changes were not significant between untreated and VP treated samples.

6.5.2 The effect of EGFR inhibition on HNSCC cells

The next aim was to disrupt the EGFR pathway with the aim of reducing viability of HNSCC cells. To achieve this, two known EGFR inhibitors were used, a monoclonal antibody, cetuximab, and tyrphostin (AG1478), which is a tyrosine kinase inhibitor that blocks the EGFR receptor at its intracellular domain [354, 359].

The first experiment was to treat the VU40T(-) and VU147(+) cells with cetuximab and investigate the effect on cell viability. Previously published studies conducted on HNSCC cell lines have used between 5-240nM of cetuximab [360-362]. Therefore, to elicit a response, the cells were treated with 10pM to 10μM for 96hrs at increments of one order of magnitude. Despite the high concentrations of cetuximab used, no effect on cell viability was observed in either cell line (**Figure 6.13A and B**).

Due to previous evidence of interaction between the EGFR and Hippo pathways, it was hypothesised that cetuximab may have an effect when used as a co-treatment with VP [355]. Therefore, VU40T(-) and VU147(+) cells were treated with the same range of cetuximab concentrations as before (**Figure 6.13**) and co-treated with EC20 or EC50 VP, as determined by the dose-response curve (**Figure 6.4**). As expected, in both HNSCC cell lines treatment with EC20 or EC50 VP resulted in a 20% and 50-60% reduction in cell viability, respectively (**Figure 6.14A and B**). However, in both cell lines after co-treatment with cetuximab, cell viability remained at the same level as with VP only treatment (**Figure 6.14**). Therefore, in both VU40T(-) and VU147(+) cells, an effect of VP with cetuximab treatment was not observed.
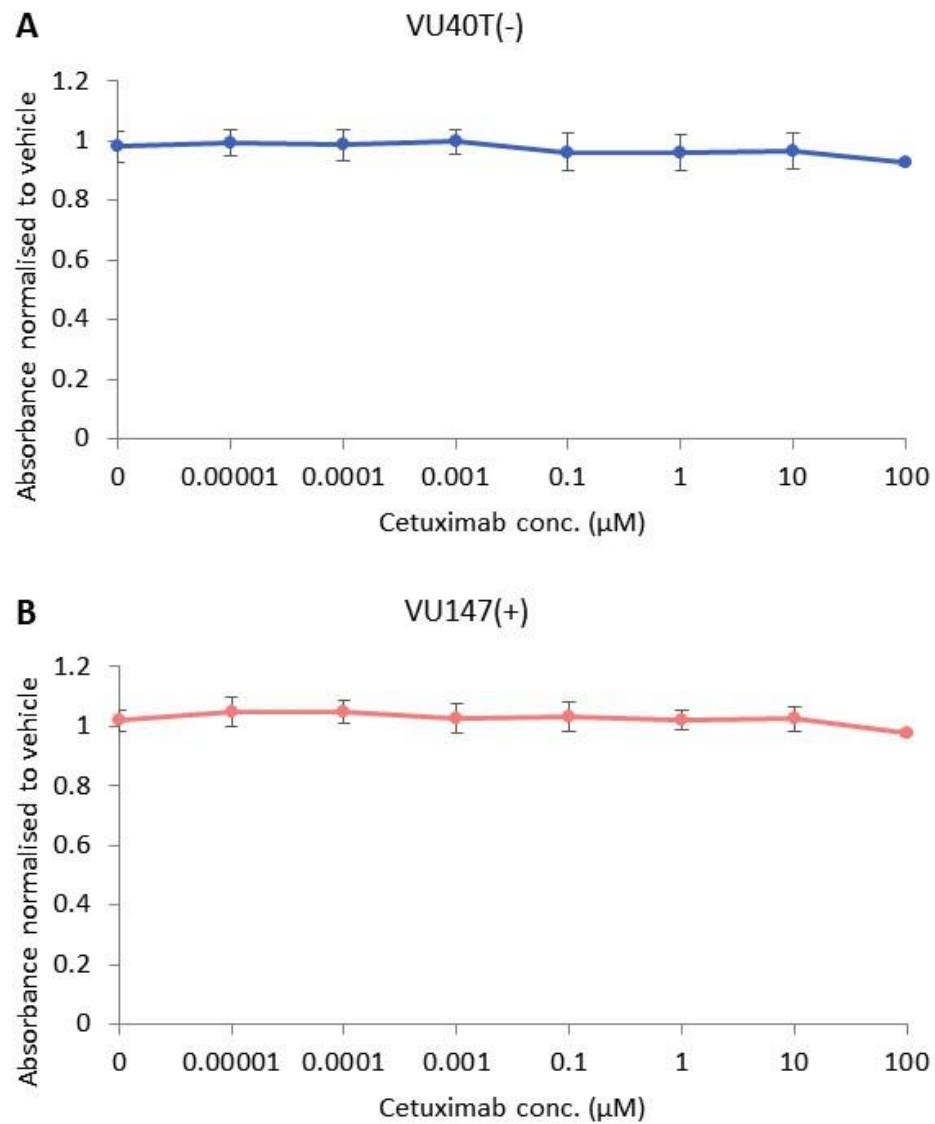
Figure 6.13. Cell viability assay in HNSCC cells after cetuximab treatment.
(A) VU40T(-) and **(B)** VU147(+) cells were treated with seven concentrations of cetuximab
between 10pM-100μM at increments of one order of magnitude for 96hrs. The error bars
represent SEM of biological triplicates.
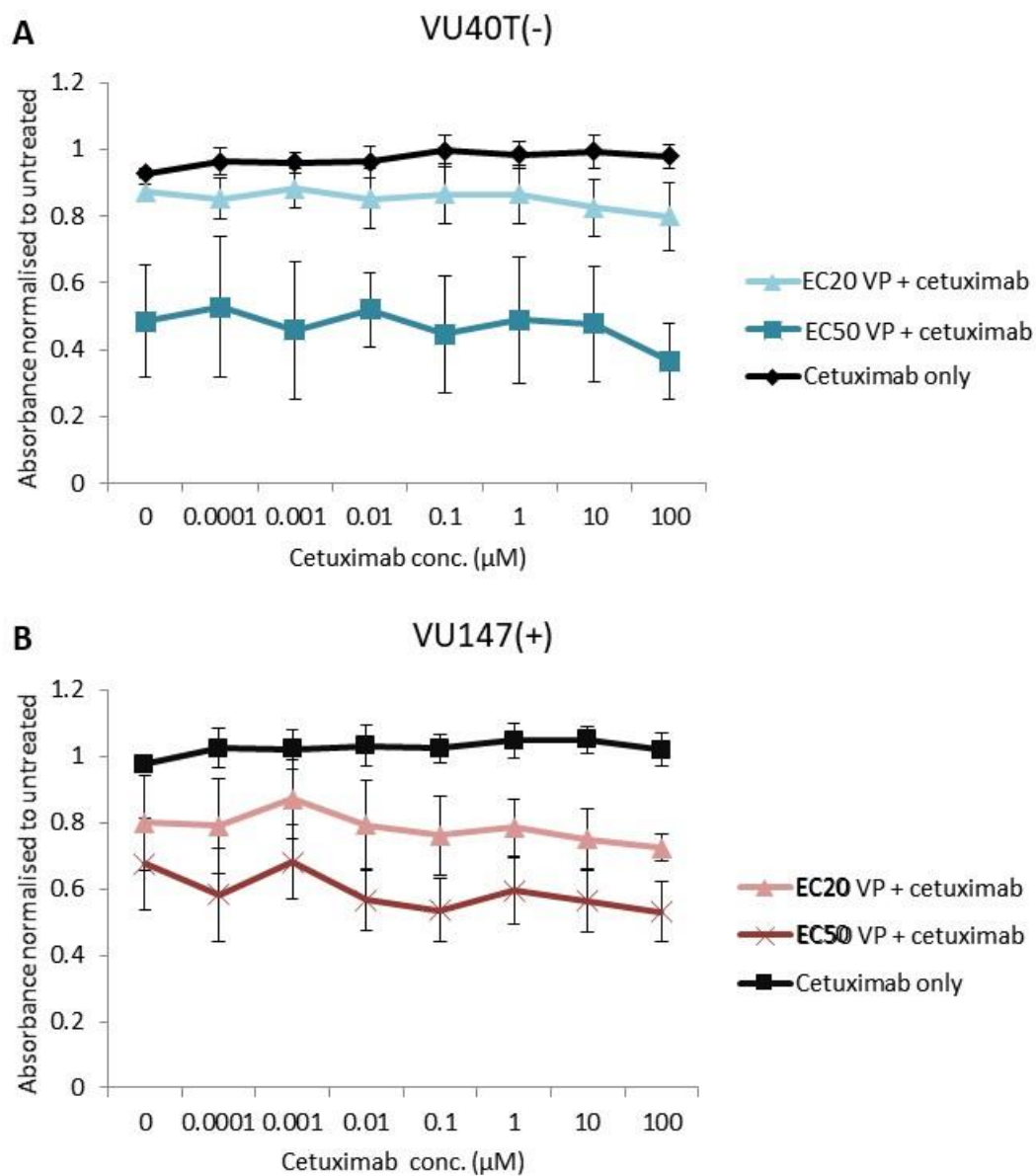
Figure 6.14. Cell viability assay in HNSCC cells with cetuximab and VP co-treatment.
(**A**) VU40T(-) and (**B**) VU147(+) cells were treated for 96hrs with EC20 VP only (0.38μM and 0.08μM, respectively) or co-treated with 0.0001, 0.001, 0.01, 0.1, 1, 10 or 100μM of cetuximab. The error bars represent SEM of biological triplicates.

Since cetuximab did not influence cell viability in the HNSCC cell lines, the VU40T(-) cells were treated with an alternative EGFR inhibitor, AG1478. A time-course for a cell viability assay was conducted to determine the most effective timepoint. Cells were treated for 24, 48, 72 or 96hrs with AG1478; a range of seven concentrations from 0.0625µM to 4µM at 2-fold increments were used. A dose-dependent response was observed at each timepoint, however as there was no significant difference between 72 and 96hrs of treatment, 96hrs of treatment was used for further experiments **(Figure 6.15A)**. A dose-response curve was plotted to find the EC20 of the cells after 96hrs of treatment, which was 0.58µM (**Figure 6.15B**).

To investigate the effect of AG1478 and VP co-treatment in VU40T(-) cells, the cells were treated with either EC20 VP (0.38µM), EC20 AG1478 (0.58µM) or both **(Figure 6.16)**. A 20% reduction in cell viability was observed after EC20 VP only treatment and a 10% reduction with EC20 AG1478. A 30% reduction in cell viability was observed after co-treatment **(Figure 6.16)**. Therefore, upon single treatments, the expected reduction in cell viability was observed. However, upon co-treatment of VP and AG1478, while a larger decrease in cell viability was observed compared to single treatments, these results were not statistically significant.
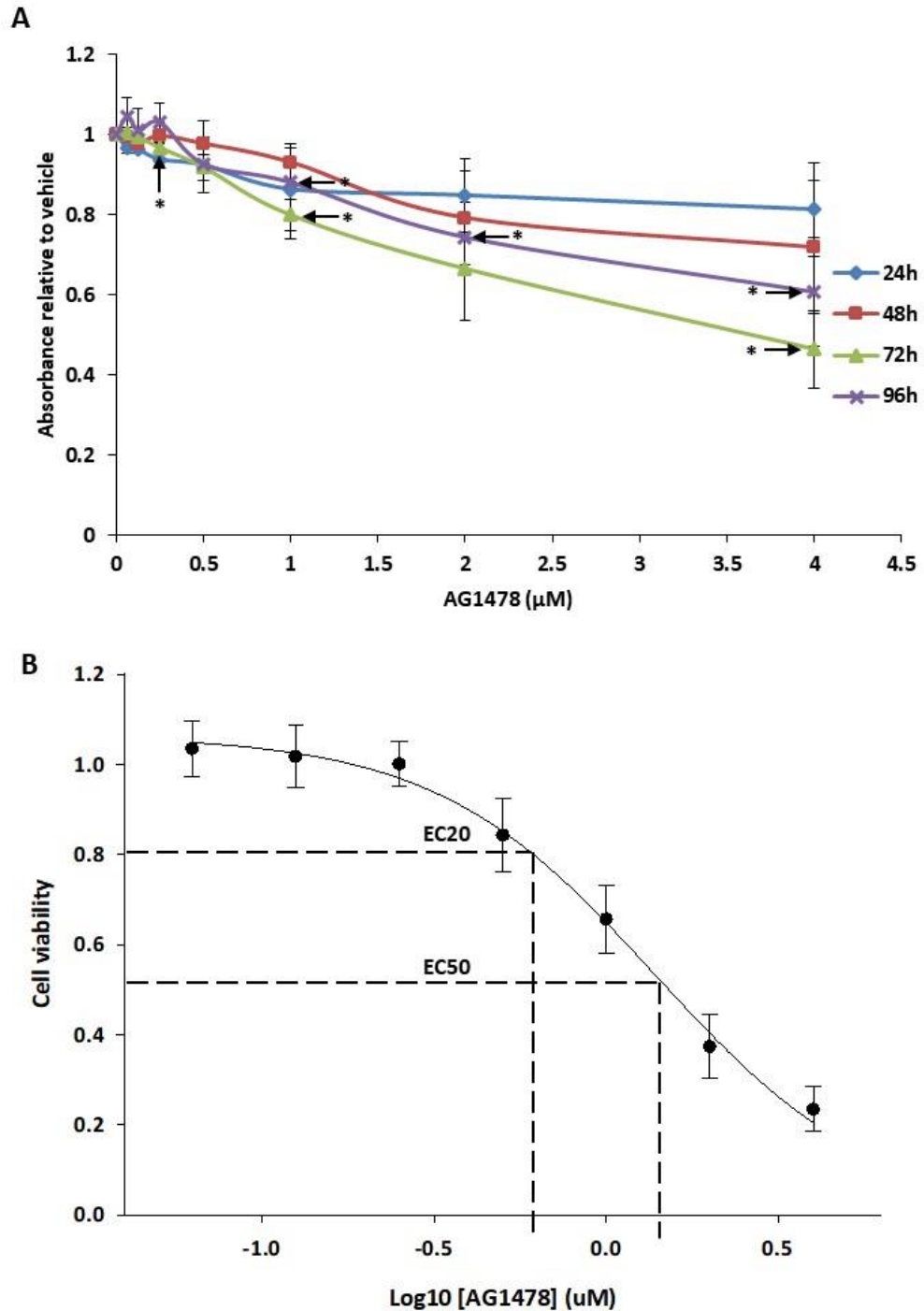
Figure 6.15. Cell viability assay in VU40T(-) cells treated with AG1478.
(A) To find the AG1478 EC20 and the most effective timepoint, VU40T(-) cells were treated with vehicle (ethanol) or 0.0625, 0.125, 0.25, 0.5, 1, 2 or 0.4μM of AG1478 for 24, 48, 72 or 96hrs. Absorbance levels were measured every 24hrs for 96hrs and normalised to vehicle. **(B)** A 96hrs sigmoidal dose-response curve was plotted using SigmaPlot to find the EC20 (0.58μM) and EC50 (1.39μM), which were calculated by determining the concentrations which reduce viable cells by 20% or 50%, respectively. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test (*P<0.05, **P<0.005, ***P<0.0005).
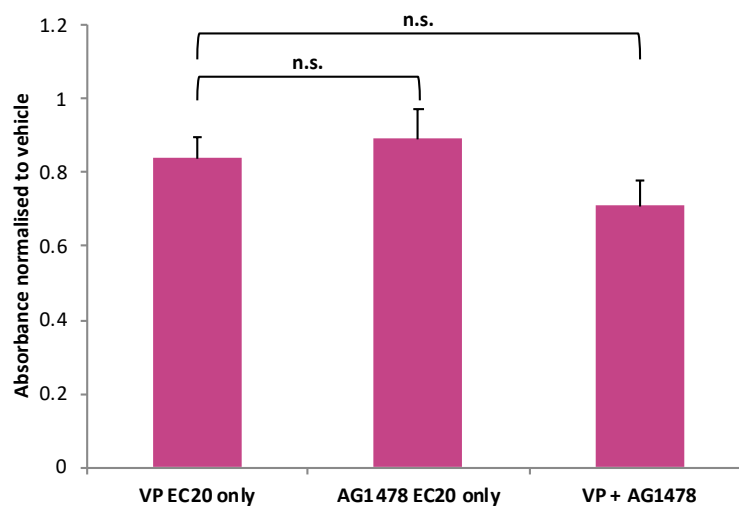
Figure 6.16. Co-treatment of VU40T(-) cells with EC20 VP and AG1478.
VU40T(-) cells were treated for 96hrs with vehicle, EC20 VP, EC20 AG1478 or both. Cell viability was measured by normalising to vehicle. The error bars represent SEM of biological triplicates. Statistical significance was calculated using an unpaired t-test (n.s. denotes 'not significant').

6.6 Discussion

The data described in the previous results chapters have highlighted the Hippo pathway as being one of the key pathways in the *cis*-regulation of the HNSCC (head and neck squamous cell carcinoma) cell lines. Despite the Hippo pathway being extensively investigated by many groups, in HNSCC, the focus has largely been on YAP (yes-associated protein) and its correlation with tumour grade and cell proliferation [87, 93, 314, 363]. However, based on our evidence, it was hypothesised that the investigation of the Hippo downstream transcription effector, TEAD4 could hold value as a therapeutic target. Published investigations of the TEAD transcription factors (TFs), while not extensive in HNSCC, have been reported to correlate with tumour grade and progression [316]. Therefore, due to the association of Hippo with HNSCC tumour progression, verteporfin (VP) was used to disrupt YAP/TEAD interaction with the aim of disrupting the Hippo pathway and its downstream targets.

Our data demonstrated that VP treatment can reduce cell viability of both VU40T(-) and VU147(+) cells. This effect was accompanied by a small increase in cells in G1 phase, a small decrease in cells in G2/M phase and a small increase in cells appearing to undergo apoptosis. Due to the small changes, taken together, these data indicate either slowing of cell proliferation, cell cycle arrest, senescence or cell death. Therefore, it cannot be determined with these experiments alone, the mechanism of VP in reducing cell viability. The reason for the small changes could be due to the concentration of VP used (EC20) where only 20% of cells are affected by this dose. Therefore, a higher concentration such as EC50 could be used to exert a larger measurable effect. In addition, supporting data would be useful such as

trypan blue staining and counting the cells before and after VP treatment. This would provide data on whether the cells are viable and whether they are in fact undergoing proliferation. However, to support the above findings, our RNA-seq pathway enrichment analysis of VP treated VU40T(-) cells revealed that genes involved in the apoptosis pathway and cell cycle pathway are upregulated and downregulated, respectively. These findings while not previously investigated in HNSCC, are in line with published data derived from other cancers. For example, VP treatment can promote cell cycle arrest in pancreatic, endometrial and retinoblastoma cancer cells [349, 364, 365], reduce viability of ovarian cancer cells, reduce cell growth in bladder cancer cell lines and increase apoptosis in pancreatic cancer cells [364, 366, 367]. Therefore, it can be concluded that VP results in reduced cell viability in the HNSCC cells here. In addition, our preliminary data and published data of others suggest that this effect could be through the induction of apoptosis and alteration of cell cycle distribution. However, in the cell lines investigated here, further experiments to support this are required.

TEAD4 ChIP-qPCR was conducted at three of the most differentially expressed genes after VP treatment. A reduction in TEAD4 chromatin binding was observed, however, this result was not statistically significant. A reason for this could be biological variation between ChIP experiments, introduced by the dynamic nature of TF binding. For example, there is evidence that TFs can bind for durations of between 0.5-12 seconds [247]. This could, therefore, introduce variability between cell populations. In addition, TEAD4 ChIP-seq variability can be augmented by Hippo pathway stimuli, such as cell growth and density, therefore TEAD chromatin binding may vary depending on the degree of cell-cell contact. To reduce variability, cell density could be controlled more strictly by conducting cell counts

to ensure similar final cell densities before carrying out ChIP. To further strengthen this data and investigate whether the effect of VP is specific to the TEAD TF, negative control regions could be tested by qPCR, i.e. genes where TEAD4 binds but gene expression does not change after VP treatment. Therefore, the data for the effect of VP on TEAD4 chromatin binding, while not statistically significant, which could be attributed to biological variation, a trend towards reduced TEAD4 binding after VP treatment can be observed. To support this, while previously TEAD binding after VP treatment has not directly been investigated, it has been shown in breast cancer cells by luciferase reporter assays that VP can reduce TEAD activity [368].

The three genes selected to investigate differential binding were *YWHAE* (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein E), *KCTD* (potassium channel tetramerization domain containing)-*2* and *ADAM9* (eukaryotic translation initiation factor 3 subunit J 9). The *YWHAE* gene encodes 14-3-3ε, which is one of seven members of the 14-3-3 protein family [351, 369]. One of the functions of the 14-3-3 proteins is sequestering YAP/TAZ (tafazzin) to adherens junctions when the Hippo pathway is active and downstream transcription is inactive **(Figure 6.1)**. VP treatment has been seen to increase levels of 14-3-3σ (encoded by *YWHAZ*) and reduce nuclear YAP levels [351]. Furthermore, published evidence reports that VP can directly bind YAP [346-350]. However, our RNA-seq data, revealed a decrease in *YWHAE* transcript level after VP treatment, as well as of other members of the 14-3-3 gene family (*YWHAB*, *YWHAH* and *YWHAZ*). Published findings suggest that VP may bind to YAP in the cytoplasm, preventing its transit to the nucleus and resulting in a decrease in nuclear YAP levels, while also increasing 14-3-3 activity. To investigate whether VP influences

YAP localisation, immunofluorescence for the YAP protein could be conducted. Alternatively, in the cell lines here, it may be that VP can also induce YAP degradation or reduce *YAP1* gene transcription by other mechanisms, leading to a reduction in 14-3-3 transcription. To determine this, investigation of YAP protein localisation and 14-3-3 protein levels would be required. To support this hypothesis, an alternative YAP-independent mechanism for VP has been reported by Dasari *et al*. in an endometrial cancer cell line. The cells were transfected with siRNA for *YAP1* and VP treatment still resulted in a cytotoxic effect. On the other hand, in an endometrial cancer organoid model, nuclear YAP levels were downregulated after exposure to VP resulting in cytotoxicity, which has also been observed in other cell types, as discussed earlier [303, 342, 343, 351]. Therefore, it may be that VP can influence the activity of the Hippo pathway independently of YAP, and so 14-3-3 activity is not required to maintain low YAP levels.

The protein encoded by the second differentially expressed gene, KCTD2 is an adaptor protein for the cullin3 E3 ubiquitin ligase complex, which targets pRB (retinoblastoma protein) for degradation [20, 370]. The pRB protein is often disrupted in HNSCC, leading to aberrant cell cycle progression [56, 62]. Interestingly, RNA-seq analysis revealed that after VP treatment *KCTD2* levels decrease. On the other hand, RNA-seq pathway enrichment analysis revealed the pRB pathway to be downregulated by VP treatment. A reason for this discrepancy may be that some genes annotated in the pRB pathway can be overexpressed and contribute to tumour progression. For example, the gene which encodes the E2F3 (E2 factor 3) protein, a TF for the pRB pathway, was annotated in our data. E2F3 has been shown to be overexpressed in renal cell carcinoma and involved in

inducing transcription of genes associated with cell proliferation [371]. Therefore, it may be that VP can, to some extent, restore cell cycle control.

The *ADAM9* (metalloproteinase disintegrin cysteine-rich protein 9) gene has previously been observed to be overexpressed in many cancers including in oral SCC cell lines. It has been demonstrated that ADAM9 can induce anti-apoptotic effects mediated by the EGFR (epidermal derived growth factor)/AKT (protein kinase B) pathway [372]. Encouragingly, our RNA-seq data revealed that VP reduces *ADAM9* transcript levels [372].

The primary non-tumour HOK (human oral keratinocyte) cells displayed reduced viability with a similar EC50 value to VU147(+) cells, but more sensitivity than VU40T(-) cells. To investigate these differences, VP treatment of VU147(+) and HOK cells followed by RNA-seq and ChIP-qPCR experiments could be conducted to compare with the data derived from the VU40T(-) cells to determine differences in the mechanism of action of VP.

One of the mechanisms by which VP could induce a reduction in cell viability could be by inhibition of TEAD-mediated transcription. For example, Takeuchi *et al*. have shown that knockdown of TEAD4 in oral SCC cell lines can promote cell cycle arrest, which agrees with observations of our data [316]. Furthermore, it has been shown in epithelial breast cancer cells, that disruption of TAZ/TEAD activity can abolish cell proliferation [373]. Therefore, based on our and published evidence, it may be that disruption of TEAD binding can reduce cell viability through transcription inhibition of TEAD target genes. However, as mentioned earlier further experiments must be conducted to conclude that VP results in disruption of TEAD4 binding in VU40T(-) cells.

RNA-seq data may also offer insights into mechanisms by which VP can exert its affects. Upon receiving signals of energy and mechanical stress, the main function of the Hippo pathway is regulation of cell growth and proliferation by mediating the cytoskeletal network and cell adhesion molecules [313, 314, 374] **(Figure 6.1)**. Cell adhesion molecules can be tumour suppressive as they can restrict aberrant cell growth and cell invasion [304, 329]. In cancer however, normal Hippo signalling is lost, leading to aberrant cell-cell signalling, proliferation and growth [304, 305]. Our RNA-seq data after VP treatment revealed upregulation of predominantly cell adhesion related pathways including the NOTCH pathway, which has been demonstrated to be tumour suppressive [21, 22] and frequently inactivated in HNSCC [36, 66-71]. While the Hippo pathway was not annotated in the top 20 pathways of either dataset (upregulated or downregulated genes), the pathway was annotated at the 28th position in the upregulated dataset, suggesting that it is active, and its downstream effects are inhibited. Therefore, one mechanism of action of VP may be to restore signalling between the Hippo pathway and cell adhesion molecules.

Pathways downregulated as a result of VP treatment revealed many to be related to RNA and protein processing. Aberrant splicing and upregulation of RNA/protein processing has been reported previously in HNSCC cells and in other cancer types, leading to activation of genes associated with cell survival, such as EGFR [375-377]. Our data revealed HNSCC-related pathways such as, EGFR, MAPK and TGF-β signalling to be downregulated after VP treatment, indicating that VP could reduce HNSCC cell viability by diminishing the effect of pathways involved in HNSCC cell survival [56, 66, 74, 77, 78, 275].

In the previous chapter, it was highlighted that TEAD binding overlaps with other TFs and could contribute to cross-talk between signalling pathways. Here we investigated whether disrupting TEAD binding affects gene expression of the AP-1 and p63 TFs. The observed changes in AP-1 and *TP63* gene expression after VP treatment were not statistically significant. As mentioned earlier, the non-significant result could be due to treatment with the EC20 dose and it could be that EC50 treatment may yield a significant change. In VU40T(-) cells, levels of *TP63* increased, while in VU147(+) cells levels decreased. There is published evidence that TEAD can bind to the *ΔNp63* promoter to induce its repression and so it could be that VP is able abolish this effect by disrupting TEAD binding [331]. On the other hand, in HNSCC cells, YAP has a role in p63 degradation, while VP can promote YAP degradation [87, 303, 332, 333, 342, 343, 351]. Therefore, the contrasting effect of VP treatment in VU40T(-) and VU147(+) cells on *TP63* expression may be due to the complex associations between YAP/TEAD/p63. However, further investigations would be required to elucidate the mechanisms underlying this, including repeating the experiment with EC50 VP doses to investigate changes in gene expression and the effect on TEAD4 binding. In addition, interrogation of p63 and YAP protein levels could strengthen this data.

Next, it was investigated whether VP treatment would affect transcription of AP-1 genes, which would indicate alterations in AP-1 activity. Although not statistically significant, in VU40T(-) cells, *FOSL1* expression levels increased after VP treatment, while *JUNB* levels in VU147(+) cells decreased. JUNB knockdown has been reported to reduce metastasis and invasion in HNSCC cell lines [253]. It has also been reported that knockdown of FOSL1 can result in slowed HNSCC cell line proliferation, migration and invasion [256]. However, the function of AP-1 dimers

can be complicated as their function can differ depending on their binding partner [307, 378]. Therefore, these reported effects may be dependent on dimer composition. While VP treatment results in changes in AP-1 gene expression, which could be in part due to disruption of TEAD, AP-1 and p63, binding would have to be assessed by ChIP-seq to determine whether AP-1 binding has in fact been altered at genes involved in tumour promotion. In addition, as mentioned before, further experiments are required to determine whether VP and TEAD4 activity has influence on AP-1 and p63 gene expression.

To explore the EGFR pathway, EGFR protein expression was investigated in VU40T(-) and VU147(+) cells and compared with expression in normal HOK cells. Although both HNSCC cell lines showed higher EGFR protein expression over normal oral cells, the result was not conclusive. Therefore, this experiment would have to be repeated. Further investigation of the EGFR pathway was conducted by treating the VU40T(-) and VU147(+) cells with cetuximab. No detectable change in cell viability was observed. To test whether cetuximab does have an effect, a cell line which is sensitive to cetuximab could be used as a positive control. On the other hand, treatment with the tyrosine kinase inhibitor, AG1478 resulted in a reduction in cell viability in a dose-dependent manner, which has also been reported by other groups [379-381]. A reason for a reduction in cell viability with AG1478 and not with cetuximab may be that in HNSCC cell lines, cetuximab-resistance has been attributed to changes in levels of specific ERBB proteins, including ERBB-2 and -3, which treatment with tyrosine kinase inhibitors can overcome [113, 382]. For example, AG1478 has been reported to target and block ERBB2 activity [382]. Therefore, it may be that AG1478 can target components of the EGFR/PI3K

(phosphotidylinositol-3 kinase) pathway which contribute to cetuximab-resistance, rendering cetuximab ineffective.

Gene expression levels of the EGF-like ligands, *AREG*, *EREG* and of *EGFR* showed a small decrease after VP treatment. This is in agreement with a previous study in which overexpression of YAP resulted in increased transcript and protein levels of AREG and EGFR [355]. In addition, the same study demonstrated that treatment with AREG resulted in dephosphorylation of the core kinases of the Hippo pathway (MST (macrophage stimulating)-1/2 and LATS (large tumour suppressor protein kinase)-1/2) and of YAP, resulting in increased EGFR and AREG expression [355]. Therefore, it may be that this feedback loop is also active in HNSCC cell lines. However, changes in gene expression of EGF-like ligands was not statistically significant, which may be due to the use of a low concentration of VP. To strengthen this finding, cells could be treated with a higher concentration of VP, alternatively protein levels of the EGF-like ligands after EC20 VP treatment could be assessed. In addition, treatment with protein EGF-like ligands before and after VP treatment, followed by investigation of the Hippo kinases and YAP could be assessed.

One of the aims of this project was to investigate whether EGFR and Hippo pathway inhibition, by VP and cetuximab co-treatment, would have a synergistic effect on HNSCC cell viability. However, such an effect was not observed, moreover the reduction in cell viability was not statistically significant. However, the expected 20% reduction in cell viability was observed, as the cells were treated with EC20 doses of both drugs. The reason for not observing a synergistic effect may be that receptor tyrosine kinases and the Hippo pathway contribute to HNSCC cell viability but do so independently. Alternatively, it may be that the PI3K/AKT pathway is

activated independently of receptor tyrosine kinases where AG1478 exerts its effect [383].

In conclusion, we observe that VP treatment can reduce cell viability of HNSCC cell lines in a dose-dependent manner. VP may exert this effect by altering the cell cycle distribution of these cells and inducing apoptosis, as described in published literature [349, 364-367]. However, the observed effects are too small to determine this result conclusively. Having said this, the potential effect of VP on the cell cycle and induction of apoptosis is encouraging, as it supported by RNA-seq data where upregulation of pathways such as apoptosis and NOTCH and downregulation of pathways such as the cell cycle is observed. In addition, VP appears to upregulate cell adhesion pathways, which may have anti-tumourigenic effects in HNSCC cells, while downregulating HNSCC-related signalling pathways (EGFR, MAPK, TGF-β). Although VP appears to exert tumour suppressive effects in both HNSCC cell lines, ultimately leading to a decrease in cell viability, non-tumourigenic oral keratinocytes appear to be similarly affected, thereby limiting the clinical potential of VP.

Limitations of the findings in this chapter include the lack of statistically significant results with the use of VP, which may have been obtained with use of a higher VP dose, as supported by published data [349, 364-367]. Therefore, while the work in this chapter is preliminary and is thus far encouraging, further experiments are required to determine the mechanism of action of VP and its potential as a therapy.

CHAPTER 7: GENERAL DISCUSSION AND FUTURE WORK

The aim of this study was to investigate the HNSCC (head and neck squamous cell carcinoma) cell line *cis*-regulatory network and compare the findings with non-tumourigenic cells to reveal how the network may differ. The investigation was conducted with the ultimate aim of using the regulatory networks to identify therapeutic targets.

To address the first aim of elucidating the HNSCC cell line *cis*-regulatory network, investigations of open chromatin regions led to the identification of over 70,000 regulatory elements, primarily driven by the p63-AP1-TEAD transcription factor (TF) network. The network appears to be highly overlapping, suggesting cross-talk of signalling pathways on a transcriptional level. This addresses a gap in the literature concerning p63 and its association with TEAD (TEA domain transcription factor) and AP-1. ChIP (chromatin immunoprecipitation)-seq data in this study have demonstrated close overlap of TEAD4 binding with p63 and AP-1. Moreover, our data demonstrated that binding of all three TFs occur more frequently than binding of TEAD4 alone. This indicates that there could be co-regulation between the three factors. Cell transfection experiments using short interfering RNAs (siRNA) or short hairpin loop RNAs (shRNA) could be conducted to investigate whether TF binding is dependent on another TF within the network. There is evidence from colorectal carcinoma cells that while AP-1 and TEAD4 can co-regulate transcription, physical interaction may occur and is mediated by the SRC (proto-oncogene, non-receptor tyrosine kinase) family of transcription co-factors

[274]. To determine this in HNSCC cells, co-immunoprecipitation assays could be conducted to examine whether any combination of p63, AP-1 and TEAD4 TFs physically interact.

Apart from the p63-AP1-TEAD network, other potentially contributing TFs have been identified which are worth further investigation, such as the KLF (Krüppel-like factor) TF family. In particular, KLF5 which has been associated with an EC in HNSCC cell lines and tumours and has been associated with activation of genes involved in SCC cell identity [179]. Furthermore, the transcriptional network in the HNSCC cell lines and non-tumour cells appears to be largely similar, with some notable exceptions such as KLF4 and KLF9. Both these TFs may be more active in the regulatory network of non-tumour cells and are known to play anti-proliferative and pro-apoptotic roles [293, 294].

The TEAD and p63 motifs were found to be specific to distal sites, while the AP-1 motif was more prominent in distal sites as compared to promoter elements. To support this, binding of TEAD4, JUND and FOSL1 (FOS-like 1, AP-1 transcription factor subunit) has previously been shown to largely overlap with active enhancer regions [274]. To verify the binding of the TFs at enhancer elements, H3K27ac and H3K4me1 chromatin immunoprecipitation (ChIP-seq) experiments could be conducted as these histone modifications form part of the signature for active enhancers [129, 162, 170]. The resulting data can be mapped to the genome to investigate the overlap of active enhancer histone marks with the TF ChIP-seq data. This can be followed by enhancer/promoter reporter assays to establish whether the promoter can in fact be driven by the identified putative enhancer. The latter would be particularly important as the identified network is based mostly on distal regulatory elements and current analyses of enhancer-gene assignments are

oversimplified, as discussed in Chapter 3. An alternative method such as chromatin conformation capture could identify more accurately distal element-promoter interactions and improve the identification of the biological processes involved.

Interestingly, the investigation of open chromatin regions in HNSCC cells lines and non-tumour cells revealed that they were dominated by the AP-1 motif. Experiments to explore the role of AP-1 could include knockdown or knockout of AP-1 subunits. However, due to the varied composition of AP-1 (consisting of a member of the FOS family (FOS, FOSB, FOSL1 and FOSL2) and a member of the JUN family (JUN, JUNB, JUND)), the TF can be heterogeneous in its function. For example, JUNB knockdown has been reported to reduce HNSCC cell line metastasis and invasion, but not cell proliferation [253]. On the other hand, knockdown of FOSL1 resulted in slowed HNSCC cell line proliferation, migration and invasion [256]. Another study conducted in epidermal SCC has shown that JUN can promote transformation of keratinocytes, while JUNB inhibits transformation and cell growth [384]. Moreover, it has been reported that the AP-1 dimer composition can alter based on HNSCC tumour progression [385]. This was reflected in the data here where the cell lines are derived from various tumour stages and AP-1 subunit transcript levels differed not only between cell lines, but also between subunits. In addition, our FOSL1 and JUND ChIP-seq pathway analysis revealed distinct categories of pathways, which were more over-represented either in FOSL1 or JUND ChIP. Therefore, despite AP-1 forming a main part of the HNSCC cell line regulatory network, interrogation of the role of AP-1 could be difficult due to its highly heterogenic function.

The interrogation of the HNSCC cell line *cis*-regulatory network further highlighted enhancer clusters (ECs) regulatory regions, which have been reported to control

genes of particular importance in maintaining cell identity [173, 175]. By comparing HNSCC cell line and OTK (oral tonsil keratinocyte) pathway enrichment analyses of ECs, it was found that the EGFR (epidermal growth factor receptor), TGF (transforming growth factor)-β, MAPK and WNT (wingless) pathways appear in both datasets. This indicates that these pathways are required for oral epithelial cell identity, but also have the potential to become deregulated in cancer, which has been reported previously [29, 36, 66, 102, 103, 275, 386]. Our data also revealed that over-represented pathways in HNSCC and non-tumour cell EC datasets, as well as in AP-1, p63 and TEAD4 TF ChIP-seq data, included not only HNSCC-associated pathways (EGFR and WNT), but also the Hippo pathway, as well as pathways associated with cell-cell and cell-ECM (extracellular matrix) regulation (adherens junction, focal adhesion and regulation of actin cytoskeleton).

The finding of differential ECs between HNSCC and non-tumour cells was evident upon identification of HNSCC-specific and OTK-specific EC genes. For example, Hippo pathway was HNSCC-specific, while the NOTCH (neurogenic locus notch homolog) pathway was OTK-specific. As already discussed, Hippo signalling and TEAD activity appear to be important in the regulation of HNSCC cells, while NOTCH is tumour suppressive in HNSCC [23]. A reason for the differential EC identification between non-tumour and HNSCC cells could be gene overexpression. For example, mutations in a *KLF5* EC have been reported to not only lead to KLF5 protein overexpression, but also *TP63* activation in HNSCC cell lines [179]. p63 overexpression can lead to inhibition of epithelial differentiation and uncontrolled self-renewal of keratinocyte cells [16]. Therefore, mutations within regulatory elements for HNSCC-promoting genes could be one of the mechanisms for the presence of ECs. A strategy to determine whether this is the case would be to

compare the DHS and ChIP-seq data with copy number alterations and mutational analyses of the cell lines. Analysis of ECs reveals them to be important in both HNSCC and non-tumour oral cells forming the basis of epithelial regulation but can be deregulated in cancer. In addition, based on TF motif and ChIP-seq analysis, the regulation of these elements is likely to be conducted, at least in part, by the p63, TEAD4 and AP-1 TFs, expression of which could be driven by genetic mutations and amplifications.

One of the aims of this project was to use the HNSCC cell line regulatory network to identify therapeutic targets. Verteporfin (VP) had previously been shown to disrupt YAP (yes-associated protein)/TEAD binding, however, TEAD-DNA binding had not previously been investigated [346-350]. Data in this project has demonstrated that VP treatment can result in decreased cell viability in the VU40T(-) and VU147(+) cell lines, agreeing with published data [349, 364-367]. While it is not conclusive from the data here, VP may exert its effect through induction of apoptosis and altering the cell cycle distribution. These findings however, are preliminary and further studies will be required to draw conclusions as to the mechanism of action of VP and its effect on TEAD4 binding. Furthermore, there are some limitations of VP which may hinder its use in a clinical setting. Firstly, the sensitivity of HOK (human oral keratinocyte) cells to VP is similar to that of VU147(+) cells and more sensitive than VU40T(-) cells. Before VP can be used clinically to target the Hippo pathway, the mechanism underlying the difference in sensitivity especially between HOK and HNSCC cells needs to be elucidated. A second limitation is that VP has been reported to be poorly soluble in plasma, as well as having a low plasma half-life leading to low blood concentrations [387]. Thirdly, VP is photosensitive and there is evidence that exposure to ambient light

while in cell culture can lead to the formation of protein aggregates, which may interfere with cell growth [388]. Therefore, an alternative therapy could be cyclic peptides which, similarly to VP, directly inhibit TEAD activity but are not light sensitive and are more selective in their targets [353, 389]. Alternatively, shRNA could be used to inhibit the YAP binding domain of TEAD, or its DNA-binding domain, as has been proposed previously [390, 391].

Another aim of this project was to investigate differences underlying the *cis*-regulation of the HPV (human papillomavirus)-(+) and HPV(-) cell lines. However, no clear difference in *cis*-regulation was observed, possibly due to the small number of cell lines. To address this, the number of HPV(+) and HPV(-) models would need to be increased to investigate this further.

In conclusion, our data confirms published findings in HNSCC cells, but also contributes novel findings such as the relationship between AP-1, TEAD4 and p63. The *cis*-regulatory network underlying both HNSCC cell lines and non-tumourigenic oral keratinocyte cells involves AP-1, TEAD4 and p63, which influence key pathways involved in normal epithelial function, such as EGFR, WNT, and Hippo. These pathways in turn influence key processes including cell-cell adhesion and cell-ECM signalling. However, these pathways can become deregulated and promote HNSCC.

The effect of VP on the Hippo pathway, particularly TEAD activity cannot be concluded by the results described in this thesis alone. Further experiments must be conducted to strengthen the mechanism of action of VP in these cells. However, preliminary data indicate that VP can influence processes involved in reducing cell viability and can alter gene expression of genes involved in apoptosis, as well as

affecting other HNSCC-related pathways. Therefore, the Hippo pathway has been demonstrated as being a potential therapeutic target and VP may be a useful tool in investigations of the pathway, but further experiments must be conducted to determine the effect of VP and it may not be suitable in a clinical setting.

APPENDIX

Table 1. Top 20 statistically significant transcription factor motifs in SCC040(-) cells. Motif analysis was conducted using the 'findMotifsGenome' script within HOMER. A total of 177 TF motifs were found.

SCC040(-) TF Motifs

Name	log P-value	% of Target Sequences with Motif	% of Background Sequences with Motif	Fold-change	Motif logo
FOSL1 (FRA-1)	-1.63E+04	30.60	4.34	7.05	
ATF3	-1.55E+04	32.32	5.30	6.10	
BATF	-1.50E+04	31.83	5.33	5.97	
FOSL2 (FRA-2)	-1.47E+04	22.75	2.32	9.81	
AP-1	-1.42E+04	32.82	6.09	5.39	
JUN	-1.33E+04	18.80	1.61	11.68	
BACH2	-6.11E+03	11.05	1.34	8.25	
P63	-3.63E+03	11.15	2.52	4.42	
P53	-2.73E+03	5.00	0.60	8.33	
NF-E2	-2.27E+03	3.68	0.37	9.95	
BACH1	-2.22E+03	3.47	0.33	10.52	
CTCF	-2.20E+03	4.70	0.70	6.71	
NRF2	-1.93E+03	3.06	0.29	10.55	
RUNX	-1.85E+03	13.07	5.36	2.44	
KLF5	-1.68E+03	21.45	11.65	1.84	
RUNX1	-1.67E+03	16.71	8.17	2.05	
MAFK	-1.58E+03	5.91	1.56	3.79	
RUNX2	-1.49E+03	14.12	6.69	2.11	
RUNX-AML	-1.41E+03	12.42	5.67	2.19	
BORIS	-1.41E+03	5.17	1.34	3.86	

Table 2. Top 20 statistically significant transcription factor motifs in VU40T(-) cells. Motif analysis was conducted using the 'findMotifsGenome' script within HOMER. A total of 177 motifs were found.

VU40T(-) TF Motifs

Name	log P-value	% of Target Sequences with Motif	% of Background Sequences with Motif	Fold-change	Motif logo
FOSL1 (FRA-1)	-1.31E+04	37.59	4.58	8.21	
ATF3	-1.29E+04	39.94	5.53	7.22	
BATF	-1.26E+04	39.16	5.45	7.19	
AP-1	-1.23E+04	40.55	6.13	6.62	
FOSL2 (FRA-2)	-1.06E+04	26.67	2.46	10.84	
JUN	-9.34E+03	21.65	1.69	12.81	
BACH2	-4.33E+03	12.54	1.36	9.22	
P63	-1.98E+03	11.38	2.76	4.12	
NF-E2	-1.56E+03	4.03	0.35	11.51	
MAFK	-1.54E+03	6.83	1.26	5.42	
P53	-1.43E+03	5.21	0.76	6.86	
BACH1	-1.43E+03	3.86	0.36	10.72	
NRF2	-1.26E+03	3.32	0.30	11.07	
KLF5	-1.14E+03	21.48	10.98	1.96	
RUNX	-9.12E+02	12.92	5.70	2.27	
TEAD4	-8.68E+02	14.94	7.23	2.07	
RUNX1	-8.52E+02	16.44	8.37	1.96	
ERG	-8.15E+02	23.01	13.57	1.70	
RUNX-AML	-8.06E+02	12.45	5.71	2.18	
FLI1	-7.85E+02	17.12	9.14	1.87	

Table 3. Top 20 statistically significant transcription factor motifs in VU147(+) cells.
Motif analysis was conducted using the 'findMotifsGenome' script within HOMER. A total of 198 TF motifs were found.

VU147(+) TF Motifs

Name	log P-value	% of Target Sequences with Motif	% of Background Sequences with Motif	Fold-change	Motif logo
FOSL1 (FRA-1)	-1.95E+04	32.31	4.12	7.84	
ATF3	-1.89E+04	34.05	4.95	6.88	
BATF	-1.85E+04	33.59	4.91	6.84	
AP-1	-1.77E+04	34.47	5.58	6.18	
FOSL2 (FRA-2)	-1.68E+04	23.67	2.24	10.57	
JUN	-1.54E+04	19.81	1.56	12.70	
BACH2	-6.24E+03	10.83	1.33	8.14	
P63	-3.44E+03	10.77	2.58	4.17	
P53	-2.76E+03	5.04	0.65	7.75	
CTCF	-2.72E+03	5.24	0.73	7.18	
NF-E2	-2.36E+03	3.68	0.37	9.95	
BACH1	-2.21E+03	3.50	0.36	9.72	
KLF5	-1.89E+03	24.65	13.83	1.78	
TEAD4	-1.82E+03	15.04	6.86	2.19	
BORIS	-1.79E+03	6.24	1.61	3.88	
NRF2	-1.79E+03	3.04	0.35	8.69	
MAFK	-1.63E+03	5.74	1.50	3.83	
TEAD	-1.63E+03	12.96	5.78	2.24	
SP1	-1.56E+03	9.64	3.73	2.58	
FLI1	-1.45E+03	17.27	9.20	1.88	

Table 4. Top 20 statistically significant transcription factor motifs in UDSCC2(+) cells.

Motif analysis was conducted using the 'findMotifsGenome' script within HOMER. A total of 180 TF motifs were found.

UDSCC2(+) TF Motifs

Name	log P-value	% of Target Sequences with Motif	% of Background Sequences with Motif	Fold-change	Motif logo
FOSL1 (FRA-1)	-1.93E+03	20.99	3.80	5.52	
ATF3	-1.89E+03	22.26	4.42	5.04	
BATF	-1.87E+03	22.25	4.48	4.97	
AP-1	-1.78E+03	23.14	5.10	4.54	
FOSL2 (FRA-2)	-1.72E+03	14.81	1.95	7.59	
JUN	-1.47E+03	11.90	1.40	8.50	
P63	-1.16E+03	13.34	2.49	5.36	
NFY	-7.80E+02	18.12	6.26	2.89	
KLF5	-7.07E+02	28.77	13.90	2.07	
P53	-6.66E+02	5.27	0.59	8.93	
BACH2	-6.36E+02	6.86	1.16	5.91	
SP1	-5.50E+02	12.30	4.06	3.03	
LHX2	-4.07E+02	20.12	10.26	1.96	
LHX3	-3.67E+02	27.23	16.25	1.68	
NF1	-3.50E+02	8.35	2.84	2.94	
KLF4	-3.39E+02	9.80	3.76	2.61	
BACH1	-2.91E+02	2.47	0.31	7.97	
NKX6.1	-2.78E+02	36.27	25.38	1.43	
ELK4	-2.68E+02	11.09	5.13	2.16	
ELK1	1.00E-111	2.35	0.31	7.58	

311
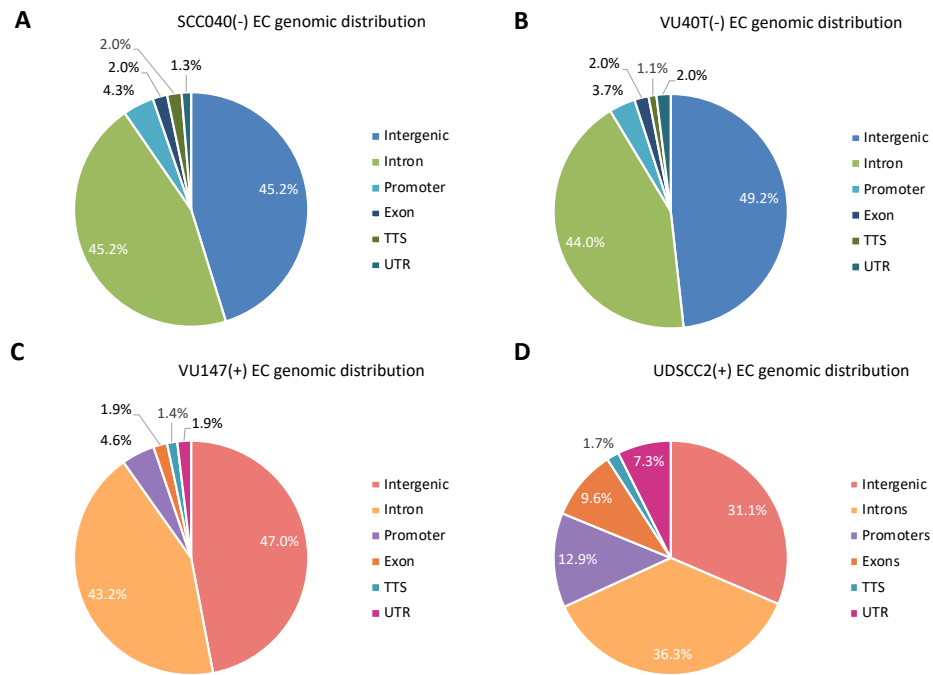
Figure 1. Genome-wide distribution of enhancer cluster (EC) peaks.
The locations of the ECs were mapped to intergenic, intronic, promoter, untranslated regions (UTR), transcription termination site (TTS), exon. This was conducted by HOMER, Genome Ontology function. The majority of ECs were located in intergenic or intronic regions.

Table 5. Top 20 Enhancer cluster pathways found in SCC040 cells.
A total of 459 ECs were found and by using the 'AnnotatePeaks' script within HOMER pathway, overrepresentation analysis was conducted to identify 352 pathways.

SCC040(-) Enhancer Clusters: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
EGFR1 Signaling Pathway	4.85E-05	-9.934	177	15	8.47	CAV2, CTNND1, EGFR, ELF3, ERRFI1, MAP3K14, PLCG2, PLEC, PTK2B, RAC1, SMAD3, SOCS3, SP1, TGIF1, VAV2	WP
Adherens junction	8.73E-05	-9.346	73	8	10.96	ACTN4, BAIAP2, CDH1, CTNND1, EGFR, IQGAP1, RAC1, SMAD3	KEGG
Proteoglycans in cancer	1.33E-04	-8.925	225	14	6.22	ANK3, CAV2, CD44, COL21A1, CTTN, EGFR, IQGAP1, MIR21, PLAU, PLCG2, RAC1, SDC1, VAV2, WNT7A	KEGG
Wnt Signaling Pathway NetPath	1.82E-04	-8.612	115	11	9.57	BTRC, CDH1, CTNNBIP1, DKK1, DLG1, JUP, LRP5, RAC1, SMAD3, TFAP2A, WNT7A	WP
Alpha6-Beta4 Integrin Signaling Pathway	3.91E-04	-7.848	69	8	11.59	DST, EGFR, ITGA6, LAMA3, PLEC, RAC1, SFN, SMAD3	WP
Leukocyte transendothelial migration	5.16E-04	-7.57	118	9	7.63	ACTN4, CLDN1, CTNND1, F11R, PLCG2, PTK2B, RAC1, RAPGEF4, VAV2	KEGG
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	6.09E-04	-7.404	74	7	9.46	ACTN4, CDH2, DSG2, ITGA6, ITGB6, JUP, LMNA	KEGG
Viral carcinogenesis	7.40E-04	-7.209	206	12	5.83	ACTN4, CDK6, DLG1, EGR2, GTF2A1, HDAC11, HIST1H2BM, HIST1H2BO, HIST1H4E, PKM, RAC1, TBPL2	KEGG
Regulation of actin cytoskeleton	1.03E-03	-6.874	214	12	5.61	ACTN4, BAIAP2, BDKRB2, EGFR, IQGAP1, ITGA6, ITGB6, MYH9, MYLK, RAC1, SSH3, VAV2	KEGG
Hippo signaling pathway	3.21E-03	-5.741	153	9	5.88	BTRC, CDH1, DLG1, FRMD6, ID1, SMAD3, SMAD7, WNT7A, WWTR1	KEGG
HTLV-I infection	5.07E-03	-5.284	259	12	4.63	BCL2L1, DLG1, EGR2, FOSL1, MAP3K14, NRP1, POLD4, SLC2A1, SMAD3, TBPL2, TNFRSF1A, WNT7A	KEGG
Pathways in cancer	5.09E-03	-5.281	327	14	4.28	BCL2L1, CDH1, CDK6, EGFR, FADD, GSTP1, ITGA6, JUP, LAMA3, PLCG2, RAC1, SLC2A1, SMAD3, WNT7A	KEGG
Wnt signaling pathway	6.15E-03	-5.092	139	8	5.76	BTRC, CTNNBIP1, DKK1, FOSL1, LRP5, RAC1, SMAD3, WNT7A	KEGG
Integrated Breast Cancer Pathway	6.18E-03	-5.087	152	10	6.58	ANXA1, CDH1, EGFR, FADD, GADD45A, NCOA3, RAC1, SMAD7, SP1, ZMYND8	WP
Apoptosis	6.94E-03	-4.971	86	6	6.98	BCL2L1, CAPN2, FADD, IRAK2, MAP3K14, TNFRSF1A	KEGG
Focal adhesion	7.77E-03	-4.858	207	10	4.83	ACTN4, CAPN2, CAV2, EGFR, ITGA6, ITGB6, LAMA3, MYLK, RAC1, VAV2	KEGG
NF-kappa B signaling pathway	9.08E-03	-4.701	91	6	6.59	BCL10, BCL2L1, MAP3K14, PLAU, PLCG2, TNFRSF1A	KEGG
Pancreatic cancer	9.61E-03	-4.645	66	5	7.58	BCL2L1, CDK6, EGFR, RAC1, SMAD3	KEGG
Epithelial cell signaling in Helicobacter pylori infection	1.09E-02	-4.522	68	5	7.35	EGFR, F11R, MAP3K14, PLCG2, RAC1	KEGG
TGF Beta Signaling Pathway	1.40E-02	-4.268	55	5	9.09	FST, ITGB6, SMAD3, SMAD7, TGIF1	WP

Table 6. Top 20 Enhancer cluster pathways found in VU40T(-) cells.
A total of 477 ECs were found and by using the 'AnnotatePeaks' script within HOMER pathway, overrepresentation analysis was conducted to identify 346 pathways.

VU40T(-) Enhancer Clusters: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Leukocyte transendothelial migration	7.28E-07	-14.13	118	13	11.02	ACTN4, ARHGAP35, CLDN1, CLDN16, CTNNB1, CTNND1, PTK2, PTK2B, RAC1, RAPGEF3, RAPGEF4, RASSF5, VAV2	KEGG
Adherens junction	1.60E-05	-11.04	73	9	12.33	ACTN4, CDH1, CTNNB1, CTNND1, EGFR, PTPN1, PVRL1, RAC1, TJP1	KEGG
Alpha6-Beta4 Integrin Signaling Pathway	8.04E-05	-9.429	69	9	13.04	DST, EGFR, ITGB4, PLEC, PTK2, RAC1, SFN, YWHAQ, YWHAZ	WP
Rap1 signaling pathway	9.98E-05	-9.212	211	14	6.64	CDH1, CTNNB1, CTNND1, EGFR, EPHA2, ID1, LPAR3, PARD6B, RAC1, RAPGEF3, RAPGEF4, RASSF5, TIAM1, VAV2	KEGG
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	1.25E-04	-8.989	74	8	10.81	ACTN4, CDH2, CTNNB1, ITGA2, ITGA3, ITGB4, ITGB6, JUP	KEGG
Hippo signaling pathway	2.87E-04	-8.158	153	11	7.19	BTRC, CDH1, CTNNB1, DLG1, FRMD6, ID1, PARD6B, STK3, WWTR1, YWHAQ, YWHAZ	KEGG
Focal adhesion	3.07E-04	-8.087	207	13	6.28	ACTN4, ARHGAP35, CAPN2, CAV1, CTNNB1, EGFR, ITGA2, ITGA3, ITGB4, ITGB6, PTK2, RAC1, VAV2	KEGG
Androgen receptor signaling pathway	5.27E-04	-7.548	88	9	10.23	CAV1, CTNNB1, EGFR, ETV5, FHL2, NCOA3, PTK2, RAC1, SP1	WP
RANKL/RANK Signaling Pathway	5.35E-04	-7.533	54	7	12.96	FHL2, IKBKB, LYN, PTK2, RAC1, TNFRSF11B, TRAF6	WP
MAPK signaling pathway	7.34E-04	-7.217	256	14	5.47	DUSP5, DUSP6, DUSP7, EGFR, HSPA1A, HSPB1, IKBKB, IL1A, MAP3K13, MECOM, RAC1, RPS6KA5, STK3, TRAF6	KEGG
Arrhythmogenic right ventricular cardiomyopathy	7.47E-04	-7.199	74	8	10.81	ACTN4, CDH2, CTNNB1, ITGA2, ITGA3, ITGB4, ITGB6, JUP	WP
Integrin-mediated cell adhesion	1.25E-03	-6.688	99	9	9.09	CAPN2, CAV1, ITGA2, ITGA3, ITGB4, ITGB6, PTK2, RAC1, VAV2	WP
MAPK signaling pathway	1.31E-03	-6.64	163	12	7.36	DUSP5, DUSP6, DUSP7, EGFR, HSPA1A, HSPB1, IKBKB, IL1A, MAP3K13, RAC1, STK3, TRAF6	WP
EGFR1 Signaling Pathway	2.64E-03	-5.937	177	12	6.78	CAV1, CTNND1, EGFR, ELF3, EPS8, ERRFI1, PLEC, PTK2B, RAC1, RPS6KA5, SP1, VAV2	WP
Viral carcinogenesis	3.30E-03	-5.714	206	11	5.34	ACTN4,DLG1,HIST1H2BB,HIST1H2BG,HIST1H4H,LYN,PKM,RAC1,TBPL2,YWHAQ,YWHAZ	KEGG
IL-1 Signaling Pathway	3.51E-03	-5.652	12	3	25.00	IL1A, IL1RAP, TRAF6	WP
Regulation of actin cytoskeleton	4.41E-03	-5.424	214	11	5.14	ACTN4, ARHGAP35, EGFR, ITGA2, ITGA3, ITGB4, ITGB6, PTK2, RAC1, TIAM1, VAV2	KEGG
Bacterial invasion of epithelial cells	4.54E-03	-5.395	76	6	7.89	CAV1, CDH1, CTNNB1, PTK2, RAC1, SEPT9	KEGG
Pathogenic Escherichia coli infection	5.20E-03	-5.259	55	5	9.09	CDH1, CLDN1, CTNNB1, YWHAQ, YWHAZ	KEGG
Tight junction	6.16E-03	-5.09	134	8	5.97	ACTN4, AMOTL1, CLDN1, CLDN16, CTNNB1, PARD6B, TJP1, TJP2	KEGG

Table 7. Top 20 Enhancer cluster pathways found in VU147(+) cells.

A total of 380 ECs were found and by using the 'AnnotatePeaks' script within HOMER pathway overrepresentation analysis was conducted to identify 396 pathways.

VU147(+) Enhancer Clusters: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
EGFR1 Signaling Pathway	7.32E-06	-11.82	177	19	10.73446	CAV2, CBLC, CTNND1, EGFR, ELF3, ERRFI1, MAP2K3, MAP3K14, MAPK7, MTA2, MYC, PLD1, PLEC, PTK2B, RAC1, RALGDS, SP1, TGIF1, VAV2	WP
Rap1 signaling pathway	5.57E-05	-9.796	211	17	8.056872	ACTB, CDH1, CTNND1, EFNA5, EGFR, EPHA2, GNAI2, ID1, MAP2K3, MET, PARD6B, RAC1, RALGDS, RAPGEF1, SIPA1L3, VAV2, VEGFB	KEGG
Focal adhesion	1.51E-04	-8.796	207	16	7.729469	ACTB, ACTN4, CAPN2, CAV2, EGFR, FLNA, ITGA2, ITGB6, LAMA5, LAMB3, MET, PTK2, RAC1, RAPGEF1, VAV2, VEGFB	KEGG
Endocytosis	3.98E-04	-7.829	203	15	7.389163	AP2S1, ARF6, ARRB2, CAV2, CBLC, CLTB, EGFR, EPN3, HSPA1A, MET, NEDD4, PARD6B, PIP5KL1, PLD1, SMURF1	KEGG
Integrin-mediated cell adhesion	5.08E-04	-7.585	99	11	11.11111	CAPN2, CAPNS1, CAV2, ITGA2, ITGB6, MAP2K3, MAPK7, PTK2, RAC1, RAPGEF1, VAV2	WP
Alpha6-Beta4 Integrin Signaling Pathway	5.10E-04	-7.581	69	9	13.04348	DST, EGFR, LAMA5, LAMB3, MET, PLEC, PTK2, RAC1, SFN	WP
Small cell lung cancer	5.13E-04	-7.575	86	9	10.46512	BCL2L1, CDKN1B, CDKN2B, ITGA2, LAMA5, LAMB3, MYC, PTK2, RXRA	KEGG
Pathways in cancer	5.44E-04	-7.516	327	20	6.116208	BCL2L1, CBLC, CDH1, CDKN1B, CDKN2B, EGFR, FADD, ITGA2, JUP, LAMA5, LAMB3, MET, MYC, PLD1, PTK2, RAC1, RALGDS, RASSF1, RXRA, VEGFB	KEGG
Adherens junction	7.72E-04	-7.167	73	8	10.9589	ACTB, ACTN4, CDH1, CTNND1, EGFR, MET, RAC1, TJP1	KEGG
Bacterial invasion of epithelial cells	1.01E-03	-6.898	76	8	10.52632	ACTB, CAV2, CBLC, CDH1, CLTB, MET, PTK2, RAC1	KEGG
Transcriptional misregulation in cancer	1.14E-03	-6.775	179	13	7.26257	BCL2L1, CDKN1B, CDKN2C, DOT1L, HIST1H3H, JUP, MET, MYC, PLAU, PTK2, RXRA, SP1, SPINT1	KEGG
Leukocyte transendothelial migration	1.35E-03	-6.611	118	10	8.474576	ACTB, ACTN4, CLDN16, CLDN9, CTNND1, GNAI2, PTK2, PTK2B, RAC1, VAV2	KEGG
Focal Adhesion	1.44E-03	-6.542	184	15	8.152174	ACTB, CAV2, EGFR, FLNA, ITGA2, ITGB6, LAMA5, LAMB3, MAP2K3, MAPK7, MET, PTK2, RAC1, RAPGEF1, VEGFB	WP
MAPK signaling pathway	1.58E-03	-6.45	256	16	6.25	ARRB2, EGFR, FLNA, GADD45B, HSPA1A, HSPB1, IL1A, MAP2K3, MAP3K11, MAP3K14, MAP3K8, MAPK7, MKNK2, MYC, RAC1, TNFRSF1A	KEGG
Integrated Pancreatic Cancer Pathway	2.44E-03	-6.015	194	15	7.731959	ANXA1, BBC3, CDKN1B, CDKN2B, EGFR, FADD, FTL, GPRC5A, MAPK7, MYC, NOXA1, RAC1, RXRA, SP1, TNFRSF1A	WP
Proteoglycans in cancer	3.21E-03	-5.74	225	14	6.222222	ACTB, CAV2, CBLC, EGFR, FLNA, ITGA2, MET, MIR21, MYC, PLAU, PLAUR, PTK2, RAC1, VAV2	KEGG
MAPK signaling pathway	3.63E-03	-5.62	163	13	7.97546	ARRB2, EGFR, FLNA, HSPA1A, HSPA5, HSPB1, IL1A, MAP3K11, MAP3K14, MAP3K8, MAPK7, MYC, RAC1	WP
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	3.84E-03	-5.563	74	7	9.459459	ACTB, ACTN4, ITGA2, ITGB6, JUP, LMNA, SGCA	KEGG
Signaling Pathways in Glioblastoma	5.74E-03	-5.161	80	8	10	CDKN1B, CDKN2B, CDKN2C, EGFR, ERRFI1, MAP2K3, MDM4, MET	WP
Ras signaling pathway	8.55E-03	-4.762	226	13	5.752212	ARF6, BCL2L1, EFNA5, EGFR, EPHA2, ETS2, MET, PLD1, RAC1, RALGDS, RASSF1, RIN1, VEGFB	KEGG

315

Table 8. Top 20 Enhancer cluster (EC) pathways found in UDSCC2(+) cells.
A total of 558 ECs were found and by using the 'AnnotatePeaks' script within HOMER pathway overrepresentation analysis was conducted to identify 209 pathways.

UDSCC2(+) Enhancer Clusters: Pathway overrepresentation

Term	Enrichment	logP	# of genes in germ	# target genes in term	% of targets in term	Gene symbols	Database
Systemic lupus erythematosus	4.01E-05	-10.12	134	6	4.48	HIST1H4I,HIST1H2AE,HIST1H2BI,HIST1H3H,HIST1H4B,HIST1H2BO	KEGG
Alcoholism	2.01E-04	-8.514	179	6	3.35	HIST1H4I,HIST1H2AE,HIST1H2BI,HIST1H3H,HIST1H4B,HIST1H2BO	KEGG
DNA replication	6.91E-04	-7.277	36	3	8.33	RPA4,PCNA,PRIM2	KEGG
Endocytosis	2.86E-03	-5.857	203	5	2.46	PSD3,AP2M1,HSPA1A,CAV2,EGFR	KEGG
Focal adhesion	3.11E-03	-5.773	207	5	2.42	GSK3B,RAPGEF1,MYLPF,CAV2,EGFR	KEGG
IL-3 Signaling Pathway	3.73E-03	-5.592	104	4	3.85	GSK3B,ID1,RAPGEF1,HSPB1	WP
Mismatch repair	5.55E-03	-5.193	23	2	8.70	RPA4,PCNA	KEGG
Hypertrophy Model	6.75E-03	-4.998	20	2	10.00	ATF3,IFRD1	WP
HTLV-I infection	8.04E-03	-4.824	259	5	1.93	GSK3B,ATF3,PCNA,CDC27,FOS	KEGG
Estrogen signaling pathway	1.26E-02	-4.37	100	3	3.00	FOS,HSPA1A,EGFR	KEGG
Viral carcinogenesis	1.75E-02	-4.044	206	4	1.94	HIST1H4I,HIST1H2BI,HIST1H4B,HIST1H2BO	KEGG
MAPK signaling pathway	1.78E-02	-4.028	163	4	2.45	FOS,HSPB1,HSPA1A,EGFR	WP
Signaling of Hepatocyte Growth Factor Receptor	1.78E-02	-4.027	33	2	6.06	RAPGEF1,FOS	WP
RPA complex	1.95E-02	-3.937	4	1	25.00	RPA4	KEGG
DNA polymerase alpha / primase complex	1.95E-02	-3.937	4	1	25.00	PRIM2	KEGG
Nucleotide excision repair	2.21E-02	-3.814	47	2	4.26	RPA4,PCNA	KEGG
Cell cycle	2.24E-02	-3.799	124	3	2.42	GSK3B,PCNA,CDC27	KEGG
FAS pathway and Stress induction of HSP regulation	2.33E-02	-3.76	38	2	5.26	HSPB1,LMNA	WP
IL-6 Signaling Pathway	2.33E-02	-3.758	99	3	3.03	GSK3B,FOS,HSPB1	WP
Focal Adhesion	2.65E-02	-3.629	184	4	2.17	GSK3B,RAPGEF1,CAV2,EGFR	WP

Table 9. Top 20 statistically significant transcription factor motifs in OTK cells.
A total of 185 motifs were identified. Pathway enrichment analysis was conducted in HOMER using the 'AnnotatePeaks' script and a total of 277 pathways were identified.

OTK ATAC-seq: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Pathways in cancer	1.63E-11	-24.84	327	297	90.8	AKT-1, -2, -3, ERBB2, MAPK-1, -8, -9, -10, PIK3CA, RUNX1, TGFB1, TGFB2, TGFB3, WNT-2, 2B, 3, 3A, 4, 5A, 5B, 6, 7A, 7B, 8B, 9A, 9B, 10A, 11, 16	KEGG
Hippo signaling pathway	3.28E-11	-24.14	153	147	96.1	AJUBA, AREG, CCND1, CTGF, CTNNA1-3, LATS1-2, TEAD1-4, TGFB1-3, TP73, WNT-1, -2, -3, -3A, -4, -5A, -5B, -6, -7, -7B, -8B, -9A, -9B, -10A, -11, -16, WWTR1, YAP1, YWHAE	KEGG
EGFR1 Signaling Pathway	2.50E-09	-19.81	177	169	95.5	ATF1, CTNND1, EGF, EGFR, FOS, JUN, JUND, KLF11, KRT-7, -8, -17, -18, MAPK-1, -8, -14, PIK3CA	WP
Proteoglycans in cancer	1.23E-07	-15.91	225	203	90.2	CCND1, CDKN1A, EGFR, ERBB2, ERBB3, ERBB4, MAPK-1, -11, -12, -13, -14, PIK3CA, TGFB1, TGFB2, WNT-2, -2B, -3, -3A, -4, -5A, -5B, -6, -7, -7B, -8B, -9A, -9B, -10A, -11, -16	KEGG
Fc gamma R-mediated phagocytosis	1.33E-07	-15.83	91	88	96.7	MAPK1, PIK3CA	KEGG
Rap1 signaling pathway	1.57E-07	-15.67	211	191	90.5	EGF, EGFR, MAPK-1, -11-14, PIK3CA	KEGG
TGF-beta Receptor Signaling Pathway	2.58E-07	-15.17	151	143	94.7	ATF-2, -3, CCND1, CDKN1A, FOS, JUN, JUNB, JUND, MAPK14, MAPK8, RUNX2, TGFB1, TGFB2, TGFB3, YAP1	WP
Inflammatory mediator regulation of TRP channels	9.27E-07	-13.89	99	94	94.9	MAPK8-14, PIK3CA	KEGG
Phosphatidylinositol signaling system	1.35E-06	-13.51	81	78	96.3	PIK3CA	KEGG
Thyroid hormone signaling pathway	1.51E-06	-13.40	119	111	93.3	CCND1, MAPK1, NOTCH1-3, PIK3CA, WNT4	KEGG
Inositol phosphate metabolism	1.97E-06	-13.14	61	60	98.4	PIK3CA	KEGG
Protein processing in endoplasmic reticulum	3.30E-06	-12.62	168	152	90.5	ATF-4, -6, MAPK8-10	KEGG
Huntington's disease	5.85E-06	-12.05	177	159	89.8		KEGG
Ubiquitin mediated proteolysis	8.52E-06	-11.67	137	125	91.2	WWP1	KEGG
Dopaminergic synapse	9.27E-06	-11.59	130	119	91.5	ATF-2, -4, MAPK8-14	KEGG
Integrated Pancreatic Cancer Pathway	1.61E-05	-11.0353	194	177	91.2	CCDKN1A, E2F1, EGF, EGFR, ERBB2, JUN, JUNB, JUND, MAPK-1, -4, -8, -9, -14, TGFB1	WP
VEGF signaling pathway	1.90E-05	-10.8724	61	59	96.7	MAPK-1, -11, -14, PIK3CA	KEGG
Interferon type I	1.97E-05	-10.8347	49	49	100	MAPK14	WP
p53 signaling pathway	2.58E-05	-10.567	68	65	95.6	CDKN1A, CDKN2A	KEGG
Alpha6-Beta4 Integrin Signaling Pathway	3.76E-05	-10.1877	69	67	97.1	PIK3CA, TP73	WP

Table 10. Top 20 Enhancer cluster (EC) pathways found in OTK cells.
A total of 592 ECs were found and by using the 'AnnotatePeaks' script within HOMER
pathway overrepresentation analysis was conducted to identify 274 pathways.

OTK ATAC-seq Enhancer Clusters: Pathway overrepresentation

Term	Enrichment	logP	# of genes in term	# target genes in term	% of targets in term	Gene symbols	Database
Pathways in cancer	1.63E-11	-24.84	327	297	90.8	CCND1, CDKN-1A, -2A, E2F1, EGF, EGFR, ERBB2, JUN, MAPK-1, -8-10, PIK3CA, RUNX1, TGFB1-3, WNT-3, -3A, -4, -5A, -5B, -6, -7A, 7B, -8B, -9A, -9B, -10A, -11, 16	KEGG
Hippo signaling pathway	3.28E-11	-24.14	153	147	96.1	ATF1, EGF, EGFR, JUN, JUND, KLF11, KRT17, KRT18, KRT7, KRT8, MAPK1, MAPK14, MAPK8, PIK3CA	KEGG
EGFR1 Signaling Pathway	2.50E-09	-19.81	177	169	95.5	ATF1, EGF, EGFR, FOS, JUN, JUND, KLF11, KRT17, KRT18, KRT7, KRT8 MAPK-1, -8, -14, PIK3CA	WP
Proteoglycans in cancer	1.23E-07	-15.91	225	203	90.2	CDKN-1A, EGFR, ERBB2-4, MAPK-1, -11-14, TGFB1-3, WNT-2, -2B, -3, -3A, -4, -5A, 5B, -6, -7A, -7B, -8B, -9A, -9B, 1-A, 11, 16	KEGG
Fc gamma R-mediated phagocytosis	1.33E-07	-15.83	91	88	96.7	PIK3CA	KEGG
Rap1 signaling pathway	1.57E-07	-15.67	211	191	90.5	EGF, EGFR, LAT, MAPK-1, -11-14, PIK3CA	KEGG
TGF-beta Receptor Signaling Pathway	2.58E-07	-15.17	151	143	94.7	ATF-2, -3, CDKN1A, FOS, FOSB, JUN, JUNB, JUND, MAPK-8, -14, SP1, TGFB1-3, TP73, YAP1	WP
Inflammatory mediator regulation of TRP channels	9.27E-07	-13.89	99	94	94.9	MAPK8-14,PIK3CA	KEGG
Phosphatidylinositol signaling system	1.35E-06	-13.51	81	78	96.3	PIK3CA	KEGG
Thyroid hormone signaling pathway	1.51E-06	-13.40	119	111	93.3	MAPK1, NOTCH1, PIK3CA, WNT4	KEGG
Inositol phosphate metabolism	1.97E-06	-13.14	61	60	98.4	PIK3CA	KEGG
Protein processing in endoplasmic reticulum	3.30E-06	-12.62	168	152	90.5	ATF-4, -6, MAPK8-10	KEGG
Huntington's disease	5.85E-06	-12.05	177	159	89.8	-	KEGG
Ubiquitin mediated proteolysis	8.52E-06	-11.67	137	125	91.2	WWP1	KEGG
Dopaminergic synapse	9.27E-06	-11.59	130	119	91.5	MAPK8-14	KEGG
Integrated Pancreatic Cancer Pathway	1.61E-05	-11.04	194	177	91.2	CCND1, CDKN1A, E2F1, JUN, JUNB, JUND, KRAS, MAPK-1, -4, -8, -9, -14	WP
VEGF signaling pathway	1.90E-05	-10.87	61	59	96.7	MAPK-1, 11-14	KEGG
Interferon type I	1.97E-05	-10.83	49	49	100.0	MAPK14	WP
p53 signaling pathway	2.58E-05	-10.57	68	65	95.6	CCND1, CDKN2A,TP73	KEGG
Alpha6-Beta4 Integrin Signaling Pathway	3.76E-05	-10.19	69	67	97.1	EGFR, PIK3CA	WP
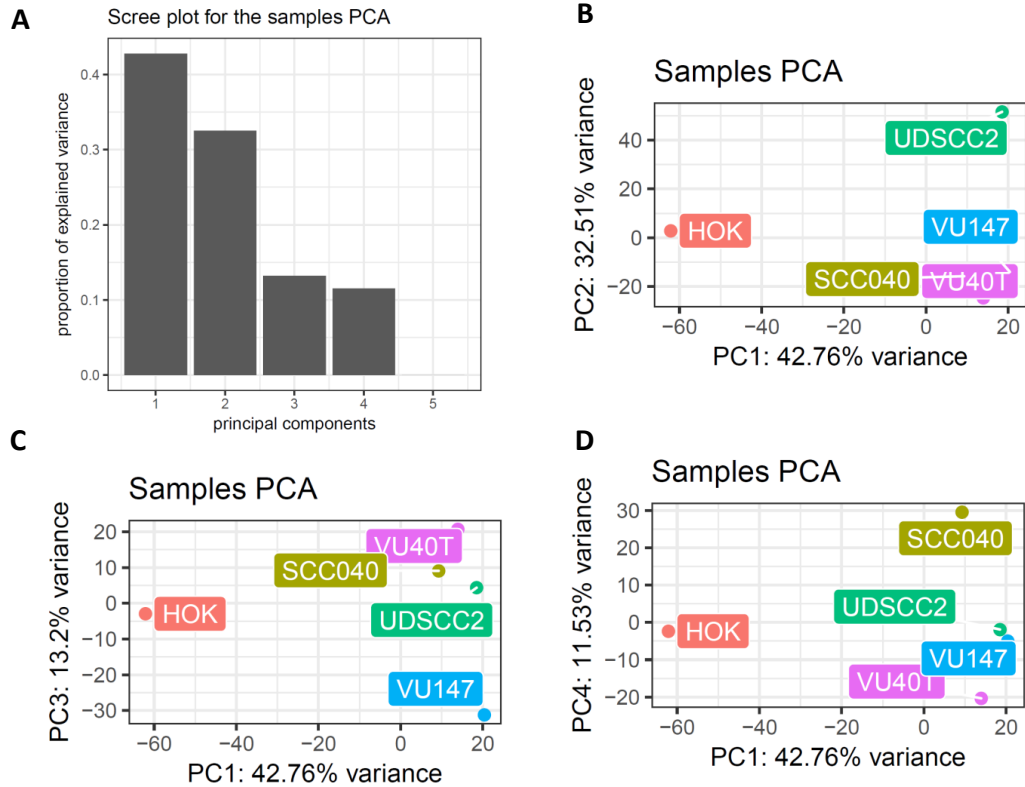
Figure 2. Principal components analysis (PCA) of OTK and HNSCC cells.

Analysis was conducted using the 'PCAexplorer' script within the R software. **(A)** Four principal components (PCs) were identified upon clustering of gene FPKM values: PC1 (42.76%), PC2 (32.51%), PC3 (13.2%) and PC4 (11.53%). **(B)** PC1 accounted for 42.76% of variability due to differences in HNSCC gene expression compared to HOK cells. PC2 accounted for 32.5% of variability due to differences in UDSCC2(+) gene expression compared to the remaining HNSCC cell lines and HOK cells. **(C)** PC3 accounted for 13.2% of variability due to differences in VU147(+) cells compared to the remaining HNSCC cells and HOK cells. **(D)** PC4 accounted for 11.53% of variability due to differences in SCC040(-) cells compared to the remaining cell types.
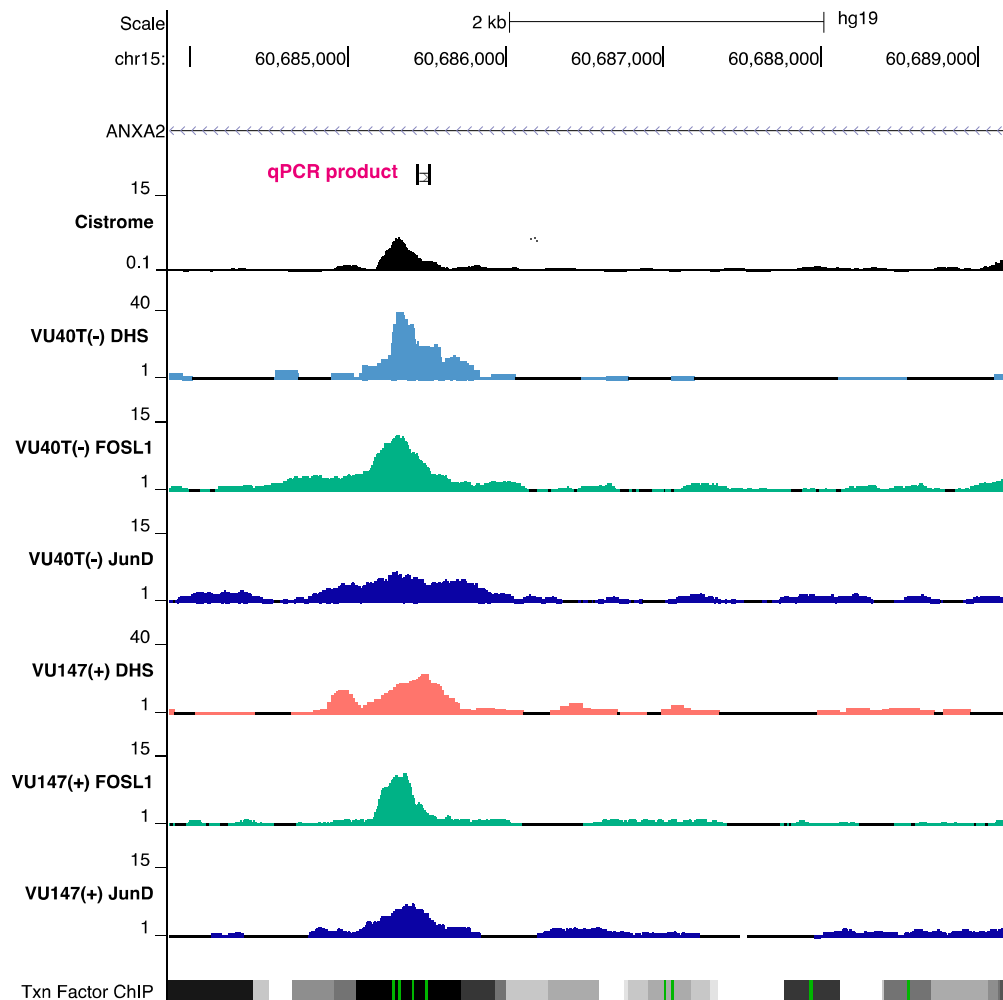
Figure 3. UCSC browser screenshot for FOSL1 and JUND ChIPs at the *ANXA2* gene. qPCR primers were selected near the *ANXA2* gene to test for potential FOSL1 and JUND binding. The qPCR product (mapped on to the coverage tracks) overlaps with FOSL1 and JUND TF ChIP peaks in the VU40T(-) and VU147(+) cell lines; the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (red track) cell lines. The 'Txn Factor ChIP' track shows regions of TF binding in ENCODE cell lines.
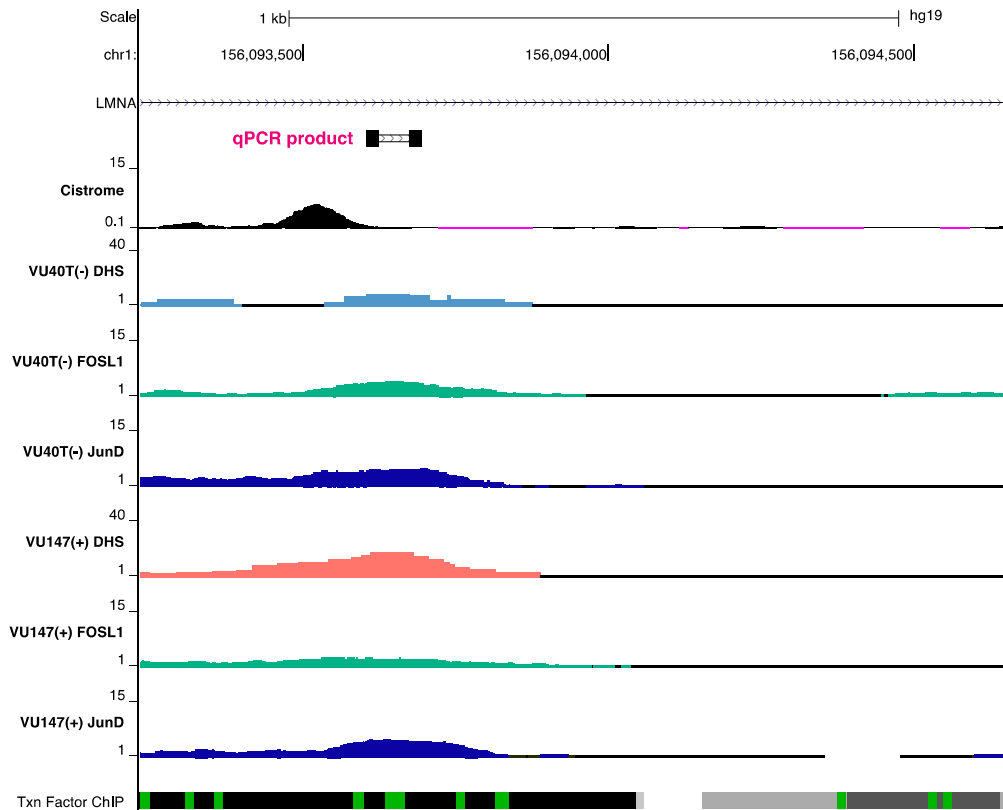
Figure 4. UCSC browser screenshot for FOSL1 and JUND ChIPs at the *LMNA* gene. QPCR primers were selected near the *LMNA* gene to test for potential FOSL1 and JUND binding. The qPCR product (mapped on to the coverage tracks) overlaps with FOSL1 and JUND TF ChIP peaks in the VU40T(-) and VU147(+) cell lines; the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines. The 'Txn Factor ChIP' track shows regions of TF binding in ENCODE cell lines.
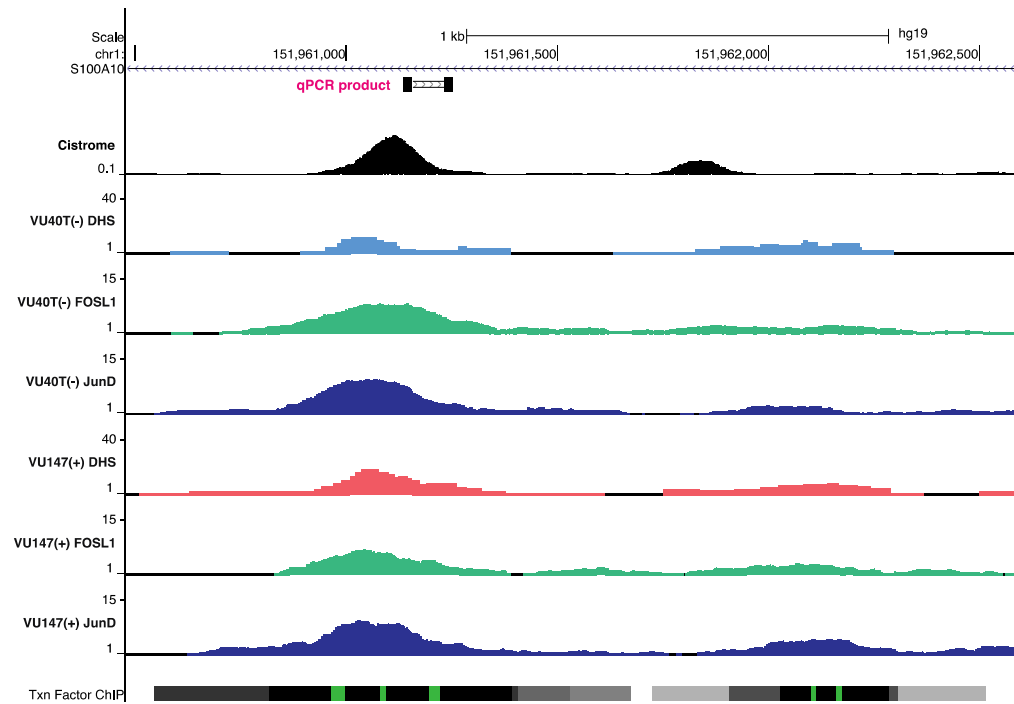
Figure 5. UCSC browser screenshot for FOSL1 and JUND ChIPs at the *S100A10* gene.

QPCR primers were selected near the *S100A10* gene to test for potential FOSL1 and JUND binding. The qPCR product (mapped on to the coverage tracks) overlaps with FOSL1 and JUND TF ChIP peaks in the VU40T(-) and VU147(+) cell lines; the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines. The 'Txn Factor ChIP' track shows regions of TF binding in ENCODE cell lines.
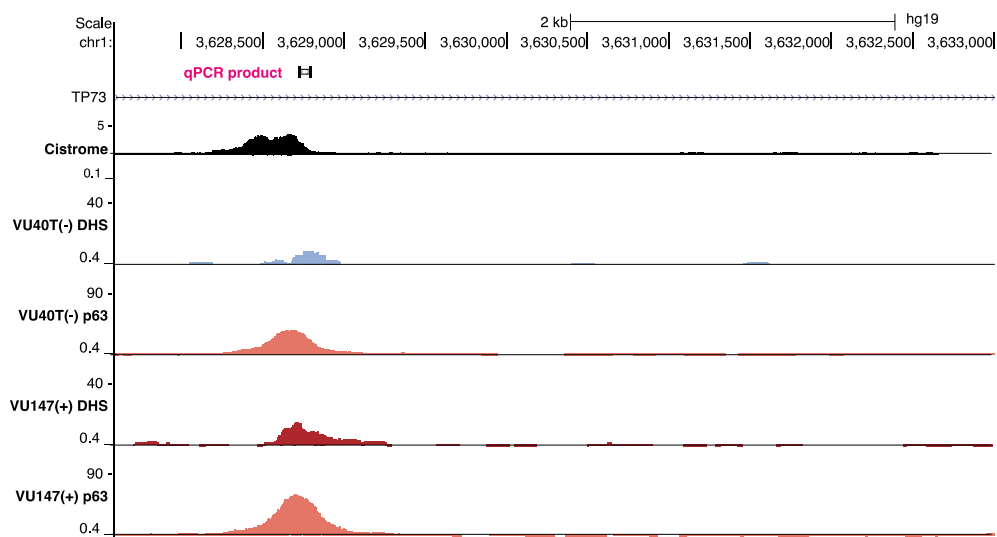
Figure 6. UCSC browser screenshot for p63 ChIP at the *TP73* gene (site 1).
QPCR primers were selected near the *TP73* gene to test for potential p63 binding. The qPCR product (mapped on to the coverage tracks) overlaps with p63 TF ChIP peaks in the VU40T(-) and VU147(+) cell lines; the data were obtained from the Cistrome database and are shown as a black track. The site also overlaps with a DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines.
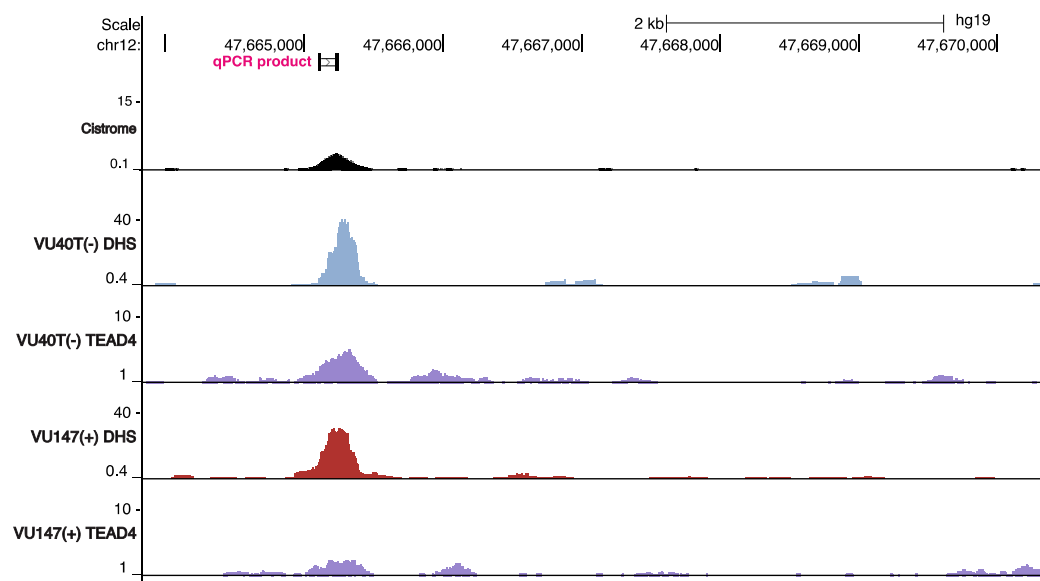
Figure 7. UCSC browser screenshot for TEAD4 ChIP (site 1).
QPCR primers were selected ~35kb downstream of the *PCED1B* TSS to test for potential
TEAD4 binding. The qPCR product (mapped on to the coverage tracks) overlaps with
TEAD4 TF ChIP peaks in the VU40T(-) and VU147(+) cell lines; the data were obtained
from the Cistrome database and are shown as a black track. The site also overlaps with a
DHS in both VU40T(-) (blue track) and VU147(+) (pink track) cell lines.

REFERENCES

1. Squier, C.A. and M.J. Kremer, *Biology of oral mucosa and esophagus.* J Natl Cancer Inst Monogr, 2001(29): p. 7-15.
2. Larjava, H., L. Koivisto, L. Hakkinen, and J. Heino, *Epithelial integrins with special reference to oral epithelia.* J Dent Res, 2011. **90**(12): p. 1367-76.
3. Roberts S., *Papillomaviruses*, in *eLS*. 2015.
4. Winning, T.A. and G.C. Townsend, *Oral mucosal embryology and histology.* Clin Dermatol, 2000. **18**(5): p. 499-511.
5. Moll, R., M. Divo, and L. Langbein, *The human keratins: biology and pathology.* Histochem Cell Biol, 2008. **129**(6): p. 705-33.
6. Wong, P. and P.A. Coulombe, *Loss of keratin 6 (K6) proteins reveals a function for intermediate filaments during wound repair.* J Cell Biol, 2003. **163**(2): p. 327-37.
7. Lesch, C.A., C.A. Squier, A. Cruchley, D.M. Williams, and P. Speight, *The permeability of human oral mucosa and skin to water.* J Dent Res, 1989. **68**(9): p. 1345-9.
8. Lodish, H.F., *Molecular cell biology*. 7th ed. 2013, New York: W.H. Freeman and Co. xxxiii, 1154, 58 p.
9. Desouza, M., P.W. Gunning, and J.R. Stehn, *The actin cytoskeleton as a sensor and mediator of apoptosis.* Bioarchitecture, 2012. **2**(3): p. 75-87.
10. Yang, A., R. Schweitzer, D. Sun, M. Kaghad, N. Walker, R.T. Bronson, C. Tabin, A. Sharpe, D. Caput, C. Crum, and F. McKeon, *p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development.* Nature, 1999. **398**(6729): p. 714-8.
11. Mills, A.A., B. Zheng, X.J. Wang, H. Vogel, D.R. Roop, and A. Bradley, *p63 is a p53 homologue required for limb and epidermal morphogenesis.* Nature, 1999. **398**(6729): p. 708-13.
12. Yoh, K. and R. Prywes, *Pathway Regulation of p63, a Director of Epithelial Cell Fate.* Front Endocrinol (Lausanne), 2015. **6**: p. 51.
13. Koster, M.I. and D.R. Roop, *Mechanisms regulating epithelial stratification.* Annu Rev Cell Dev Biol, 2007. **23**: p. 93-113.
14. Koster, M.I., B. Marinari, A.S. Payne, P.N. Kantaputra, A. Costanzo, and D.R. Roop, *DeltaNp63 knockdown mice: A mouse model for AEC syndrome.* Am J Med Genet A, 2009. **149A**(9): p. 1942-7.
15. Romano, R.A., K. Smalley, C. Magraw, V.A. Serna, T. Kurita, S. Raghavan, and S. Sinha, *Delta Np63 knockout mice reveal its indispensable role as a master regulator of epithelial development and differentiation.* Development, 2012. **139**(4): p. 772-782.
16. Nguyen, B.C., et al., *Cross-regulation between Notch and p63 in keratinocyte commitment to differentiation.* Genes Dev, 2006. **20**(8): p. 1028-42.
17. Truong, A.B., M. Kretz, T.W. Ridky, R. Kimmel, and P.A. Khavari, *p63 regulates proliferation and differentiation of developmentally mature keratinocytes.* Genes Dev, 2006. **20**(22): p. 3185-97.
18. Suh, Y., I. Amelio, T. Guerrero Urbano, and M. Tavassoli, *Clinical update on cancer: molecular oncology of head and neck cancer.* Cell Death Dis, 2014. **5**: p. e1018.
19. Andersson, E.R., R. Sandberg, and U. Lendahl, *Notch signaling: simplicity in design, versatility in function.* Development, 2011. **138**(17): p. 3593-612.
20. Rampias, T., C. Sasaki, and A. Psyrri, *Molecular mechanisms of HPV induced carcinogenesis in head and neck.* Oral Oncol, 2014. **50**(5): p. 356-63.

21.   Dotto, G.P., *Notch tumor suppressor function.* Oncogene, 2008. **27**: p. 5115–5123.

22.   Rangarajan, A., C. Talora, R. Okuyama, M. Nicolas, C. Mammucari, H. Oh, J.C. Aster, S. Krishna, D. Metzger, P. Chambon, L. Miele, M. Aguet, F. Radtke, and G.P. Dotto, *Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation.* EMBO J, 2001. **20**(13): p. 3427-36.

23.   Beck, T.N. and E.A. Golemis, *Genomic insights into head and neck cancer.* Cancers of the Head & Neck, 2016. **1**(1): p. 1.

24.   Noguti, J., D.E.M. CF, T.A. Hossaka, M. Franco, C.T. Oshima, R.A. Dedivitis, and D.A. Ribeiro, *The role of canonical WNT signaling pathway in oral carcinogenesis: a comprehensive review.* Anticancer Res, 2012. **32**(3): p. 873-8.

25.   Liu, F. and S.E. Millar, *Wnt/beta-catenin signaling in oral tissue development and disease.* J Dent Res, 2010. **89**(4): p. 318-30.

26.   Cassell, A. and J.R. Grandis, *Investigational EGFR-targeted therapy in head and neck squamous cell carcinoma.* Expert Opinion on Investigational Drugs, 2010. **19**(6): p. 709-722.

27.   Zieske, J.D. and M. Wasson, *Regional variation in distribution of EGF receptor in developing and adult corneal epithelium.* J Cell Sci, 1993. **106 (Pt 1)**: p. 145-52.

28.   Patel, A.N., J.M. Mehnert, and S. Kim, *Treatment of recurrent metastatic head and neck cancer: focus on cetuximab.* Clin Med Insights Ear Nose Throat, 2012. **5**: p. 1-16.

29.   Leemans, C.R., P.J.F. Snijders, and R.H. Brakenhoff, *The molecular landscape of head and neck cancer.* Nat Rev Cancer, 2018. **18**(5): p. 269-282.

30.   Coca-Pelaz, A., R.P. Takes, K. Hutcheson, N.F. Saba, M. Haigentz, Jr., C.R. Bradford, R. de Bree, P. Strojan, V.J. Lund, W.M. Mendenhall, I.J. Nixon, M. Quer, A. Rinaldo, and A. Ferlito, *Head and Neck Cancer: A Review of the Impact of Treatment Delay on Outcome.* Adv Ther, 2018. **35**(2): p. 153-160.

31.   Hashibe, M., P. Brennan, S. Benhamou, X. Castellsague, C. Chen, M.P. Curado, L. Dal Maso, A.W. Daudt, E. Fabianova, and L. Fernandez, *Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium.* J Natl Cancer Inst, 2007. **99**: p. 777-89.

32.   Dayyani, F., C.J. Etzel, M. Liu, C.H. Ho, S.M. Lippman, and A.S. Tsao, *Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC).* Head Neck Oncol, 2010. **2**: p. 15.

33.   Mourad, M., T. Jetmore, A.A. Jategaonkar, S. Moubayed, E. Moshier, and M.L. Urken, *Epidemiological Trends of Head and Neck Cancer in the United States: A SEER Population Study.* J Oral Maxillofac Surg, 2017. **75**(12): p. 2562-2572.

34.   Noguti, J., C.F. De Moura, G.P. De Jesus, V.H. Da Silva, T.A. Hossaka, C.T. Oshima, and D.A. Ribeiro, *Metastasis from oral cancer: an overview.* Cancer Genomics Proteomics, 2012. **9**(5): p. 329-35.

35.   National Cancer Institute. *Head and Neck Cancer Regions*. 2012 [cited 2018 12 Sep].

36.   The Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-582.

37.   Chaturvedi, A.K., W.F. Anderson, J. Lortet-Tieulent, M.P. Curado, J. Ferlay, S. Franceschi, P.S. Rosenberg, F. Bray, and M.L. Gillison, *Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers.* J Clin Oncol, 2013. **31**: p. 4550-9.

38.   D'Souza, G. and A. Dempsey, *The role of HPV in head and neck cancer and review of the HPV vaccine.* Prev Med, 2011. **53 Suppl 1**: p. S5-S11.

39. Gillison, M.L., A.K. Chaturvedi, W.F. Anderson, and C. Fakhry, *Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma.* J Clin Oncol, 2015. **33**: p. 3235-42.

40. Wagner, S., S.J. Sharma, N. Wuerdemann, J. Knuth, H. Reder, C. Wittekindt, and J.P. Klussmann, *Human Papillomavirus-Related Head and Neck Cancer.* Oncol Res Treat, 2017. **40**(6): p. 334-340.

41. Graham, S.V., *Keratinocyte Differentiation-Dependent Human Papillomavirus Gene Regulation.* Viruses, 2017. **9**(9): p. E245.

42. Harden, M.E. and K. Munger, *Human papillomavirus molecular biology.* Mutat Res Rev Mutat Res, 2017. **772**: p. 3-12.

43. Chung, C.H., A. Bagheri, and G. D'Souza, *Epidemiology of oral human papillomavirus infection.* Oral Oncol, 2014. **50**: p. 364-9.

44. Jiang, Y., et al., *Co-activation of super-enhancer-driven CCAT1 by TP63 and SOX2 promotes squamous cancer progression.* Nat Commun, 2018. **9**(1): p. 3619.

45. Venuti, A., F. Paolini, L. Nasir, A. Corteggio, S. Roperto, M.S. Campo, and G. Borzacchiello, *Papillomavirus E5: the smallest oncoprotein with many functions.* Mol Cancer, 2011. **10**: p. 140.

46. Um, S.H., N. Mundi, J. Yoo, D.A. Palma, K. Fung, D. MacNeil, B. Wehrli, J.S. Mymryk, J.W. Barrett, and A.C. Nichols, *Variable expression of the forgotten oncogene E5 in HPV-positive oropharyngeal cancer.* J Clin Virol, 2014. **61**(1): p. 94-100.

47. Kivi, N., D. Greco, P. Auvinen, and E. Auvinen, *Genes involved in cell adhesion, cell motility and mitogenic signaling are altered due to HPV 16 E5 protein expression.* Oncogene, 2008. **27**(18): p. 2532-41.

48. Brand, T.M., et al., *Cross-talk Signaling between HER3 and HPV16 E6 and E7 Mediates Resistance to PI3K Inhibitors in Head and Neck Cancer.* Cancer Res, 2018. **78**(9): p. 2383-2395.

49. Gloss, B. and H.U. Bernard, *The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element.* J Virol, 1990. **64**(11): p. 5577-84.

50. Oh, S.T., S. Kyo, and L.A. Laimins, *Telomerase activation by human papillomavirus type 16 E6 protein: induction of human telomerase reverse transcriptase expression through Myc and GC-rich Sp1 binding sites.* J Virol, 2001. **75**(12): p. 5559-66.

51. Liu, Y., Y. Pan, W. Gao, Y. Ke, and Z. Lu, *Whole-Genome Analysis of Human Papillomavirus Types 16, 18, and 58 Isolated from Cervical Precancer and Cancer Samples in Chinese Women.* Sci Rep, 2017. **7**(1): p. 263.

52. Burd, E.M., *Human papillomavirus and cervical cancer.* Clin Microbiol Rev, 2003. **16**(1): p. 1-17.

53. Fakhry, C., G. D'Souza, E. Sugar, K. Weber, E. Goshu, H. Minkoff, R. Wright, E. Seaberg, and M. Gillison, *Relationship between prevalent oral and cervical human papillomavirus infections in human immunodeficiency virus-positive and -negative women.* J Clin Microbiol, 2006. **44**(12): p. 4479-85.

54. Chung, C.H. and M.L. Gillison, *Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications.* Clin Cancer Res, 2009. **15**.

55. Doorbar, J., *The E4 protein; structure, function and patterns of expression.* Virology, 2013. **445**(1-2): p. 80-98.

56. Andl, T., T. Kahn, A. Pfuhl, T. Nicola, R. Erber, C. Conradt, W. Klein, M. Helbig, A. Dietz, H. Weidauer, and F.X. Bosch, *Etiological involvement of oncogenic human papillomavirus in tonsillar squamous cell carcinomas lacking retinoblastoma cell cycle control.* Cancer Res, 1998. **58**(1): p. 5-13.

57.	Wilkie, M.D., A.S. Lau, N. Vlatkovic, T.M. Jones, and M.T. Boyd, *Metabolic signature of squamous cell carcinoma of the head and neck: Consequences of TP53 mutation and therapeutic perspectives.* Oral Oncol, 2018. **83**: p. 1-10.

58.	Lechner, M., et al., *Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV-tumors.* Genome Med, 2013. **5**(5): p. 49.

59.	Lue, R.A., S.M. Marfatia, D. Branton, and A.H. Chishti, *Cloning and characterization of hdlg: the human homologue of the Drosophila discs large tumor suppressor binds to protein 4.1.* Proc Natl Acad Sci U S A, 1994. **91**(21): p. 9818-22.

60.	Woods, D.F., C. Hough, D. Peel, G. Callaini, and P.J. Bryant, *Dlg protein is required for junction structure, cell polarity, and proliferation control in Drosophila epithelia.* Journal of Cell Biology, 1996. **134**(6): p. 1469-1482.

61.	Mantovani, F., P. Massimi, and L. Banks, *Proteasome-mediated regulation of the hDlg tumour suppressor protein.* J Cell Sci, 2001. **114**(Pt 23): p. 4285-92.

62.	Huh, K., X. Zhou, H. Hayakawa, J.Y. Cho, T.A. Libermann, J. Jin, J.W. Harper, and K. Munger, *Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor.* J Virol, 2007. **81**(18): p. 9737-47.

63.	Zhang, C.Y., W. Bao, and L.H. Wang, *Downregulation of p16(ink4a) inhibits cell proliferation and induces G1 cell cycle arrest in cervical cancer cells.* Int J Mol Med, 2014. **33**(6): p. 1577-85.

64.	McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1-2): p. 57-79.

65.	Parfenov, M., et al., *Characterization of HPV and host genome interactions in primary head and neck cancers.* Proc Natl Acad Sci U S A, 2014. **111**(43): p. 15544-9.

66.	Campbell, J.D., et al., *Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas.* Cell Rep, 2018. **23**(1): p. 194-212 e6.

67.	Seiwert, T.Y., et al., *Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas.* Clin Cancer Res, 2015. **21**(3): p. 632-41.

68.	Stransky, N., et al., *The mutational landscape of head and neck squamous cell carcinoma.* Science, 2011. **333**(6046): p. 1157-60.

69.	Agrawal, N., et al., *Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.* Science, 2011. **333**(6046): p. 1154-7.

70.	Pickering, C.R., et al., *Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers.* Cancer Discov, 2013. **3**(7): p. 770-81.

71.	Pickering, C.R., J.X. Zhang, D.M. Neskey, M. Zhao, S.A. Jasser, J.P. Wang, A. Ward, C.J. Tsai, M.V.O. Alves, and J.H. Zhou, *Squamous cell carcinoma of the oral tongue in young Non-smokers is genomically similar to tumors in older smokers.* Clin Cancer Res, 2014. **20**: p. 3842-8.

72.	Melino, G., *p63 is a suppressor of tumorigenesis and metastasis interacting with mutant p53.* Cell Death Differ, 2011. **18**: p. 1487-99.

73.	Gen, Y., K. Yasui, Y. Zen, K. Zen, O. Dohi, M. Endo, K. Tsuji, N. Wakabayashi, Y. Itoh, Y. Naito, M. Taniwaki, Y. Nakanuma, T. Okanoue, and T. Yoshikawa, *SOX2 identified as a target gene for the amplification at 3q26 that is frequently detected in esophageal squamous cell carcinoma.* Cancer Genet Cytogenet, 2010. **202**(2): p. 82-93.

74.	Pedrero, J.M., D.G. Carracedo, C.M. Pinto, A.H. Zapatero, J.P. Rodrigo, C.S. Nieto, and M.V. Gonzalez, *Frequent genetic and biochemical alterations of the PI 3-*

K/AKT/PTEN pathway in head and neck squamous cell carcinoma. Int J Cancer, 2005. **114**(2): p. 242-8.

75.   Choudhary, M.M., T.J. France, T.N. Teknos, and P. Kumar, *Interleukin-6 role in head and neck squamous cell carcinoma progression.* World J Otorhinolaryngol Head Neck Surg, 2016. **2**(2): p. 90-97.

76.   Surviladze, Z., R.T. Sterk, S.A. DeHaro, and M.A. Ozbun, *Cellular Entry of Human Papillomavirus Type 16 Involves Activation of the Phosphatidylinositol 3-Kinase/Akt/mTOR Pathway and Inhibition of Autophagy.* Journal of Virology, 2013. **87**(5): p. 2508-2517.

77.   Pim, D., P. Massimi, S.M. Dilworth, and L. Banks, *Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A.* Oncogene, 2005. **24**(53): p. 7830-8.

78.   Menges, C.W., L.A. Baglia, R. Lapoint, and D.J. McCance, *Human papillomavirus type 16 E7 up-regulates AKT activity through the retinoblastoma protein.* Cancer Res, 2006. **66**(11): p. 5555-9.

79.   Watanabe, H., et al., *SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas.* J Clin Invest, 2014. **124**(4): p. 1636-45.

80.   Kolev, V., A. Mandinova, J. Guinea-Viniegra, B. Hu, K. Lefort, C. Lambertini, V. Neel, R. Dummer, E.F. Wagner, and G.P. Dotto, *EGFR signalling as a negative regulator of Notch1 gene transcription and function in proliferating keratinocytes and cancer.* Nat Cell Biol, 2008. **10**(8): p. 902-11.

81.   Leemans, C.R., B.J.M. Braakhuis, and R.H. Brakenhoff, *The molecular biology of head and neck cancer.* Nat Rev Cancer, 2011. **11**(1): p. 9-22.

82.   Rothenberger, N.J. and L.P. Stabile, *Hepatocyte Growth Factor/c-Met Signaling in Head and Neck Cancer and Implications for Treatment.* Cancers (Basel), 2017. **9**(4): p. E39.

83.   McLaughlin-Drubin, M.E., D. Park, and K. Munger, *Tumor suppressor p16INK4A is necessary for survival of cervical carcinoma cell lines.* Proc Natl Acad Sci U S A, 2013. **110**(40): p. 16175-80.

84.   Smeets, S.J., B.J.M. Braakhuis, S. Abbas, P.J.F. Snijders, B. Ylstra, M.A. van de Wiel, G.A. Meijer, C.R. Leemans, and R.H. Brakenhoff, *Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus.* Oncogene, 2006. **25**(17): p. 2558-64.

85.   Song, Y., et al., *Identification of genomic alterations in oesophageal squamous cell cancer.* Nature, 2014. **509**(7498): p. 91-5.

86.   Chien, H.T., S.D. Cheng, W.Y. Chuang, C.T. Liao, H.M. Wang, and S.F. Huang, *Clinical Implications of FADD Gene Amplification and Protein Overexpression in Taiwanese Oral Cavity Squamous Cell Carcinomas.* PLoS One, 2016. **11**(10): p. e0164870.

87.   Ge, L., M. Smail, W. Meng, Y. Shyr, F. Ye, K.H. Fan, X. Li, H.M. Zhou, and N.A. Bhowmick, *Yes-associated protein expression in head and neck squamous cell carcinoma nodal metastasis.* PLoS One, 2011. **6**(11): p. e27529.

88.   Mauviel, A., F. Nallet-Staub, and X. Varelas, *Integrating developmental signals: a Hippo in the (path)way.* Oncogene, 2012. **31**(14): p. 1743-1756.

89.   Burtness, B., J.E. Bauman, and T. Galloway, *Novel targets in HPV-negative head and neck cancer: overcoming resistance to EGFR inhibition.* Lancet Oncol, 2013. **14**: p. e302-9.

90.   Sanchez-Vega, F., et al., *Oncogenic Signaling Pathways in The Cancer Genome Atlas.* Cell, 2018. **173**(2): p. 321-337 e10.

91.   Grandis, J.R. and D.J. Tweardy, *TGF-alpha and EGFR in head and neck cancer.* J Cell Biochem Suppl, 1993. **17F**: p. 188-91.

92. Guo, L. and L. Teng, *YAP/TAZ for cancer therapy: opportunities and challenges (review).* Int J Oncol, 2015. **46**(4): p. 1444-52.

93. Martin, D., et al., *Assembly and activation of the Hippo signalome by FAT1 tumor suppressor.* Nat Commun, 2018. **9**(1): p. 2372.

94. Ang, K.K., J. Harris, R. Wheeler, R. Weber, D.I. Rosenthal, P.F. Nguyen-Tan, W.H. Westra, C.H. Chung, R.C. Jordan, and C. Lu, *Human papillomavirus and survival of patients with oropharyngeal cancer.* New Engl J Med, 2010. **363**: p. 24-35.

95. Ndiaye, C., M. Mena, L. Alemany, M. Arbyn, X. Castellsagué, L. Laporte, F.X. Bosch, S. de Sanjosé, and H. Trottier, *HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis.* Lancet Oncol, 2014. **15**(12): p. 1319-31.

96. Cattani, P., A. Siddu, S. D'Onghia, S. Marchetti, R. Santangelo, V.G. Vellone, G.F. Zannoni, and G. Fadda, *RNA (E6 and E7) assays versus DNA (E6 and E7) assays for risk evaluation for women infected with human papillomavirus.* J Clin Microbiol, 2009. **47**(7): p. 2136-41.

97. Badaracco, G., C. Rizzo, B. Mafera, B. Pichi, D. Giannarelli, S.S. Rahimi, M.G. Vigili, and A. Venuti, *Molecular analyses and prognostic relevance of HPV in head and neck tumours.* Oncol Rep, 2007. **17**(4): p. 931-9.

98. Maxwell, J.H., J.R. Grandis, and R.L. Ferris, *HPV-Associated Head and Neck Cancer: Unique Features of Epidemiology and Clinical Management.* Annu Rev Med, 2016. **67**: p. 91-101.

99. Ferris, R.L., *Immunology and immunotherapy of head and neck cancer.* J Clin Oncol, 2015. **33**: p. 3293-304.

100. Huang, S.H., W. Xu, J. Waldron, L. Siu, X. Shen, L. Tong, J. Ringash, A. Bayley, J. Kim, and A. Hope, *Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas.* J Clin Oncol, 2015. **33**: p. 836-45.

101. Kozakiewicz, P. and L. Grzybowska-Szatkowska, *Application of molecular targeted therapies in the treatment of head and neck squamous cell carcinoma.* Oncol Lett, 2018. **15**(5): p. 7497-7505.

102. Kalyankrishna, S. and J.R. Grandis, *Epidermal growth factor receptor biology in head and neck cancer.* J Clin Oncol, 2006. **24**(17): p. 2666-72.

103. Grandis, J.R. and D.J. Tweardy, *Elevated levels of transforming growth factor alpha and epidermal growth factor receptor messenger RNA are early markers of carcinogenesis in head and neck cancer.* Cancer Res, 1993. **53**(15): p. 3579-84.

104. Blasco, M.A., P.F. Svider, S.N. Raza, J.R. Jacobs, A.J. Folbe, P. Saraf, J.A. Eloy, S. Baredes, and A.M. Fribley, *Systemic therapy for head and neck squamous cell carcinoma: Historical perspectives and recent breakthroughs.* Laryngoscope, 2017: p. 2565-2569.

105. Bedi, A., X. Chang, K. Noonan, V. Pham, R. Bedi, E.J. Fertig, M. Considine, J.A. Califano, I. Borrello, C.H. Chung, D. Sidransky, and R. Ravi, *Inhibition of TGF-beta enhances the in vivo antitumor efficacy of EGF receptor-targeted therapy.* Mol Cancer Ther, 2012. **11**(11): p. 2429-39.

106. Ravi, R., E.J. Fuchs, A. Jain, V. Pham, K. Yoshimura, T. Prouser, S. Jalla, X. Zhou, E. Garrett-Mayer, S.H. Kaufmann, R.D. Schulick, D.M. Pardoll, and A. Bedi, *Resistance of cancers to immunologic cytotoxicity and adoptive immunotherapy via X-linked inhibitor of apoptosis protein expression and coexisting defects in mitochondrial death signaling.* Cancer Res, 2006. **66**(3): p. 1730-9.

107. Johns, T.G., R.B. Luwor, C. Murone, F. Walker, J. Weinstock, A.A. Vitali, R.M. Perera, A.A. Jungbluth, E. Stockert, L.J. Old, E.C. Nice, A.W. Burgess, and A.M. Scott, *Antitumor efficacy of cytotoxic drugs and the monoclonal antibody 806 is*

enhanced by the EGF receptor inhibitor AG1478. Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15871-6.

108. Tebbutt, N., M.W. Pedersen, and T.G. Johns, *Targeting the ERBB family in cancer: couples therapy.* Nat Rev Cancer, 2013. **13**(9): p. 663-73.

109. Vermorken, J.B., J. Trigo, R. Hitt, P. Koralewski, E. Diaz-Rubio, F. Rolland, R. Knecht, N. Amellal, A. Schueler, and J. Baselga, *Open-label, uncontrolled, multicenter phase II study to evaluate the efficacy and toxicity of cetuximab as a single agent in patients with recurrent and/or metastatic squamous cell carcinoma of the head and neck who failed to respond to platinum-based therapy.* J Clin Oncol, 2007. **25**(16): p. 2171-7.

110. Van Damme, N., P. Deron, N. Van Roy, P. Demetter, A. Bols, J. Van Dorpe, F. Baert, J.L. Van Laethem, F. Speleman, P. Pauwels, and M. Peeters, *Epidermal growth factor receptor and K-RAS status in two cohorts of squamous cell carcinomas.* BMC Cancer, 2010. **10**: p. 189.

111. Licitra, L., R. Mesia, F. Rivera, E. Remenar, R. Hitt, J. Erfan, S. Rottey, A. Kawecki, D. Zabolotnyy, M. Benasso, S. Storkel, S. Senger, C. Stroh, and J.B. Vermorken, *Evaluation of EGFR gene copy number as a predictive biomarker for the efficacy of cetuximab in combination with chemotherapy in the first-line treatment of recurrent and/or metastatic squamous cell carcinoma of the head and neck: EXTREME study.* Ann Oncol, 2011. **22**(5): p. 1078-87.

112. Grandis, J.R., Q. Zeng, S.D. Drenning, and D.J. Tweardy, *Normalization of EGFR mRNA levels following restoration of wild-type p53 in a head and neck squamous cell carcinoma cell line.* Int J Oncol, 1998. **13**(2): p. 375-8.

113. Yonesaka, K., et al., *Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab.* Sci Transl Med, 2011. **3**(99): p. 99ra86.

114. Barnea, I., S. Haif, R. Keshet, V. Karaush, S. Lev-Ari, A. Khafif, A. Shtabsky, Y. Yarden, A. Vexler, and R. Ben Yosef, *Targeting ErbB-1 and ErbB-4 in irradiated head and neck cancer: results of in vitro and in vivo studies.* Head Neck, 2013. **35**(3): p. 399-407.

115. Harrington, K., et al., *Randomised Phase II study of oral lapatinib combined with chemoradiotherapy in patients with advanced squamous cell carcinoma of the head and neck: rationale for future randomised trials in human papilloma virus-negative disease.* Eur J Cancer, 2013. **49**(7): p. 1609-18.

116. Frederick, B.A., B.A. Helfrich, C.D. Coldren, D. Zheng, D. Chan, P.A. Bunn, Jr., and D. Raben, *Epithelial to mesenchymal transition predicts gefitinib resistance in cell lines of head and neck squamous cell carcinoma and non-small cell lung carcinoma.* Mol Cancer Ther, 2007. **6**(6): p. 1683-91.

117. Boeckx, C., L. Blockx, K.O. de Beeck, R. Limame, G.V. Camp, M. Peeters, J.B. Vermorken, P. Specenier, A. Wouters, M. Baay, and F. Lardon, *Establishment and characterization of cetuximab resistant head and neck squamous cell carcinoma cell lines: focus on the contribution of the AP-1 transcription factor.* Am J Cancer Res, 2015. **5**(6): p. 1921-38.

118. White, R.A., S.P. Malkoski, and X.J. Wang, *TGF[beta] signaling in head and neck squamous cell carcinoma.* Oncogene, 2010. **29**(40): p. 5437-5446.

119. Mehra, R., et al., *Efficacy and safety of pembrolizumab in recurrent/metastatic head and neck squamous cell carcinoma: pooled analyses after long-term follow-up in KEYNOTE-012.* Br J Cancer, 2018. **119**(2): p. 153-159.

120. Alberts, B., *Molecular biology of the cell.* 4th ed. 2002, New York: Garland Science. xxxiv, 1548 p.

121. Quina, A.S., M. Buschbeck, and L. Di Croce, *Chromatin structure and epigenetics.* Biochem Pharmacol, 2006. **72**(11): p. 1563-9.

122. Tremethick, D.J., *Higher-order structures of chromatin: the elusive 30 nm fiber.* Cell, 2007. **128**(4): p. 651-4.

123. Figueiredo, L.M., G.A. Cross, and C.J. Janzen, *Epigenetic regulation in African trypanosomes: a new kid on the block.* Nat Rev Microbiol, 2009. **7**(7): p. 504-13.

124. Djebali, S., et al., *Landscape of transcription in human cells.* Nature, 2012. **489**(7414): p. 101-8.

125. Clamp, M., B. Fry, M. Kamal, X. Xie, J. Cuff, M.F. Lin, M. Kellis, K. Lindblad-Toh, and E.S. Lander, *Distinguishing protein-coding and noncoding genes in the human genome.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19428-33.

126. Kellis, M., et al., *Defining functional DNA elements in the human genome.* Proc Natl Acad Sci U S A, 2014. **111**(17): p. 6131-8.

127. Yilmaz, A. and E. Grotewold, *Components and mechanisms of regulation of gene expression.* Methods Mol Biol, 2010. **674**: p. 23-32.

128. Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 29-59.

129. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.

130. Dowen, J.M., Z.P. Fan, D. Hnisz, G. Ren, B.J. Abraham, L.N. Zhang, A.S. Weintraub, J. Schujiers, T.I. Lee, K. Zhao, and R.A. Young, *Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes.* Cell, 2014. **159**(2): p. 374-387.

131. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature, 2012. **485**(7398): p. 381-5.

132. Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren, *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.

133. Padeken, J., P. Zeller, and S.M. Gasser, *Repeat DNA in genome organization and stability.* Curr Opin Genet Dev, 2015. **31**: p. 12-9.

134. Wald, A.I., E.E. Hoskins, S.I. Wells, R.L. Ferris, and S.A. Khan, *Human papillomavirus alters microRNA profiles in squamous cell carcinoma of the head and neck (SCCHN) cell lines.* Head & neck, 2011. **33**(4): p. 504-512.

135. Ong, C.T. and V.G. Corces, *Enhancer function: new insights into the regulation of tissue-specific gene expression.* Nat Rev Genet, 2011. **12**(4): p. 283-93.

136. Vaquerizas, J.M., S.K. Kummerfeld, S.A. Teichmann, and N.M. Luscombe, *A census of human transcription factors: function, expression and evolution.* Nat Rev Genet, 2009. **10**(4): p. 252-63.

137. Lambert, S.A., A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, and M.T. Weirauch, *The Human Transcription Factors.* Cell, 2018. **172**(4): p. 650-665.

138. Saladi, S.V., K. Ross, M. Karaayvaz, P.R. Tata, H. Mou, J. Rajagopal, S. Ramaswamy, and L.W. Ellisen, *ACTL6A Is Co-Amplified with p63 in Squamous Cell Carcinoma to Drive YAP Activation, Regenerative Proliferation, and Poor Prognosis.* Cancer Cell, 2017. **31**(1): p. 35-49.

139. Li, Y., F. Kong, Q. Shao, R. Wang, E. Hu, J. Liu, C. Jin, D. He, and X. Xiao, *YAP Expression and Activity Are Suppressed by S100A7 via p65/NFkappaB-mediated Repression of DeltaNp63.* Mol Cancer Res, 2017. **15**(12): p. 1752-1763.

140. Novo, C.L., B.M. Javierre, J. Cairns, A. Segonds-Pichon, S.W. Wingett, P. Freire-Pritchett, M. Furlan-Magaril, S. Schoenfelder, P. Fraser, and P.J. Rugg-Gunn, *Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition.* Cell Rep, 2018. **22**(10): p. 2615-2627.

141.   Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.

142.   Sanyal, A., B.R. Lajoie, G. Jain, and J. Dekker, *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.

143.   Abbasi, A.A., Z. Paparidis, S. Malik, D.K. Goode, H. Callaway, G. Elgar, and K.H. Grzeschik, *Human GLI3 intragenic conserved non-coding sequences are tissue-specific enhancers.* PLoS One, 2007. **2**(4): p. e366.

144.   Birnbaum, R.Y., et al., *Coding exons function as tissue-specific enhancers of nearby genes.* Genome Res, 2012. **22**(6): p. 1059-68.

145.   Lomvardas, S., G. Barnea, D.J. Pisapia, M. Mendelsohn, J. Kirkland, and R. Axel, *Interchromosomal interactions and olfactory receptor choice.* Cell, 2006. **126**(2): p. 403-13.

146.   de Kok, Y.J., et al., *Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4.* Hum Mol Genet, 1996. **5**(9): p. 1229-35.

147.   Visel, A., M.J. Blow, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, and L.A. Pennacchio, *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.

148.   Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.

149.   Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol Cell, 2013. **49**(5): p. 825-37.

150.   Zhang, Y. and D. Reinberg, *Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails.* Genes Dev, 2001. **15**(18): p. 2343-60.

151.   Rippe, K., P.H. von Hippel, and J. Langowski, *Action at a distance: DNA-looping and initiation of transcription.* Trends Biochem Sci, 1995. **20**(12): p. 500-6.

152.   Mora, A., G.K. Sandve, O.S. Gabrielsen, and R. Eskeland, *In the loop: promoter-enhancer interactions and bioinformatics.* Brief Bioinform, 2016. **17**(6): p. 980-995.

153.   He, B., C. Chen, L. Teng, and K. Tan, *Global view of enhancer-promoter interactome in human cells.* Proc Natl Acad Sci U S A, 2014. **111**(21): p. E2191-9.

154.   Zuin, J., J.R. Dixon, M.I. van der Reijden, Z. Ye, P. Kolovos, R.W. Brouwer, M.P. van de Corput, H.J. van de Werken, T.A. Knoch, I.W.F. van, F.G. Grosveld, B. Ren, and K.S. Wendt, *Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.* Proc Natl Acad Sci U S A, 2014. **111**(3): p. 996-1001.

155.   Castilho, R.M., C.H. Squarize, and L.O. Almeida, *Epigenetic Modifications and Head and Neck Cancer: Implications for Tumor Progression and Resistance to Therapy.* Int J Mol Sci, 2017. **18**(7): p. E1506.

156.   Zhao, Y. and B.A. Garcia, *Comprehensive Catalog of Currently Documented Histone Modifications.* Cold Spring Harb Perspect Biol, 2015. **7**(9): p. a025064.

157.   Hong, L., G.P. Schroth, H.R. Matthews, P. Yau, and E.M. Bradbury, *Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA.* J Biol Chem, 1993. **268**(1): p. 305-14.

158.   Bird, A.P., *DNA methylation and the frequency of CpG in animal DNA.* Nucleic Acids Res, 1980. **8**(7): p. 1499-504.

159.   Long, M.D., D.J. Smiraglia, and M.J. Campbell, *The Genomic Impact of DNA CpG Methylation on Gene Expression; Relationships in Prostate Cancer.* Biomolecules, 2017. **7**(1): p. E15.

160. Wajed, S.A., P.W. Laird, and T.R. DeMeester, *DNA methylation: an alternative pathway to cancer.* Ann Surg, 2001. **234**(1): p. 10-20.

161. Consortium, E.P., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

162. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.

163. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.

164. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

165. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.

166. Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, and R. Jaenisch, *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

167. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

168. Rada-Iglesias, A., R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, and J. Wysocka, *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

169. Hu, Z. and W.W. Tee, *Enhancers and chromatin structures: regulatory hubs in gene expression and diseases.* Biosci Rep, 2017. **37**(2): p. BSR20160183.

170. Inoue, F. and N. Ahituv, *Decoding enhancers using massively parallel reporter assays.* Genomics, 2015. **106**(3): p. 159-164.

171. Placek, B.J., L.N. Harrison, B.M. Villers, and L.M. Gloss, *The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform.* Protein Sci, 2005. **14**(2): p. 514-22.

172. Pham, T.H., C. Benner, M. Lichtinger, L. Schwarzfischer, Y. Hu, R. Andreesen, W. Chen, and M. Rehli, *Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states.* Blood, 2012. **119**(24): p. e161-71.

173. Whyte, W.A., D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young, *Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.* Cell, 2013. **153**(2): p. 307-319.

174. Lovén, J., H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, and R.A. Young, *Selective inhibition of tumor oncogenes by disruption of super-enhancers.* Cell, 2013. **153**(2): p. 320-34.

175. Hnisz, D., B.J. Abraham, T.I. Lee, A. Lau, V. Saint-Andre, A.A. Sigova, H.A. Hoke, and R.A. Young, *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.

176. Pasquali, L., et al., *Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants.* Nat Genet, 2014. **46**(2): p. 136-143.

177. Parker, S.C., et al., *Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants.* Proc Natl Acad Sci U S A, 2013. **110**(44): p. 17921-6.

178. Chen, X., J.X. Loo, X. Shi, W. Xiong, Y. Guo, H. Ke, M. Yang, Y. Jiang, S. Xia, M. Zhao, S. Zhong, C. He, L. Fu, and F. Li, *E6 Protein Expressed by High-Risk HPV Activates*

Super-Enhancers of the EGFR and c-MET Oncogenes by Destabilizing the Histone Demethylase KDM5C. Cancer Res, 2018. **78**(6): p. 1418-1430.

179.   Zhang, X., et al., *Somatic Superenhancer Duplications and Hotspot Mutations Lead to Oncogenic Activation of the KLF5 Transcription Factor.* Cancer Discov, 2018. **8**(1): p. 108-125.

180.   Wiench, M., T.B. Miranda, and G.L. Hager, *Control of nuclear receptor function by local chromatin structure.* FEBS J, 2011. **278**(13): p. 2211-30.

181.   Buenrostro, J.D., B. Wu, H.Y. Chang, and W.J. Greenleaf, *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-9.

182.   John, S., P.J. Sabo, R.E. Thurman, M.H. Sung, S.C. Biddie, T.A. Johnson, G.L. Hager, and J.A. Stamatoyannopoulos, *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-8.

183.   Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.

184.   Gross, D.S. and W.T. Garrard, *Nuclease hypersensitive sites in chromatin.* Annu Rev Biochem, 1988. **57**: p. 159-97.

185.   Furey, T.S., *ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions.* Nat Rev Genet, 2012. **13**(12): p. 840-52.

186.   Barba, M., H. Czosnek, and A. Hadidi, *Historical perspective, development and applications of next-generation sequencing in plant virology.* Viruses, 2014. **6**(1): p. 106-36.

187.   Schwartz, S., R. Oren, and G. Ast, *Detection and removal of biases in the analysis of next-generation sequencing reads.* PLoS One, 2011. **6**(1): p. e16685.

188.   Natarajan, A., G.G. Yardimci, N.C. Sheffield, G.E. Crawford, and U. Ohler, *Predicting cell-type-specific gene expression from regions of open chromatin.* Genome Res, 2012. **22**(9): p. 1711-22.

189.   Song, L., et al., *Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity.* Genome Res, 2011. **21**(10): p. 1757-67.

190.   John, S., P.J. Sabo, T.K. Canfield, K. Lee, S. Vong, M. Weaver, H. Wang, J. Vierstra, A.P. Reynolds, R.E. Thurman, and J.A. Stamatoyannopoulos, *Genome-scale mapping of DNase I hypersensitivity.* Curr Protoc Mol Biol, 2013. **Chapter 27**: p. Unit 21 27.

191.   Tsompana, M. and M.J. Buck, *Chromatin accessibility: a window into the genome.* Epigenetics Chromatin, 2014. **7**(1): p. 33.

192.   Giresi, P.G., J. Kim, R.M. McDaniell, V.R. Iyer, and J.D. Lieb, *FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin.* Genome Res, 2007. **17**(6): p. 877-85.

193.   Gaulton, K.J., et al., *A map of open chromatin in human pancreatic islets.* Nat Genet, 2010. **42**(3): p. 255-9.

194.   Simon, J.M., P.G. Giresi, I.J. Davis, and J.D. Lieb, *Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA.* Nat Protoc, 2012. **7**(2): p. 256-67.

195.   Buenrostro, J.D., P.G. Giresi, L.C. Zaba, H.Y. Chang, and W.J. Greenleaf, *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.

196.   Corces, M.R., et al., *An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues.* Nat Methods, 2017. **14**(10): p. 959-962.

197.   Solomon, M.J., P.L. Larsen, and A. Varshavsky, *Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene.* Cell, 1988. **53**(6): p. 937-47.

198. Orlando, V., *Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation.* Trends Biochem Sci, 2000. **25**(3): p. 99-104.

199. O'Neill, L.P. and B.M. Turner, *Immunoprecipitation of native chromatin: NChIP.* Methods, 2003. **31**(1): p. 76-82.

200. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nat Rev Genet, 2009. **10**(10): p. 669-80.

201. Lindberg, K. and J.G. Rheinwald, *Three distinct keratinocyte subtypes identified in human oral epithelium by their patterns of keratin expression in culture and in xenografts.* Differentiation, 1990. **45**(3): p. 230-41.

202. Rheinwald, J.G., W.C. Hahn, M.R. Ramsey, J.Y. Wu, Z. Guo, H. Tsao, M. De Luca, C. Catricalà, and K.M. O'Toole, *A Two-Stage, p16(INK4A)- and p53-Dependent Keratinocyte Senescence Mechanism That Limits Replicative Potential Independent of Telomere Status.* Molecular and Cellular Biology, 2002. **22**(14): p. 5157-5172.

203. Dickson, M.A., W.C. Hahn, Y. Ino, V. Ronfard, J.Y. Wu, R.A. Weinberg, D.N. Louis, F.P. Li, and J.G. Rheinwald, *Human keratinocytes that express hTERT and also bypass a p16(INK4a)-enforced mechanism that limits life span become immortal yet retain normal growth and differentiation characteristics.* Mol Cell Biol, 2000. **20**(4): p. 1436-47.

204. Hermsen, M.A., H. Joenje, F. Arwert, M.J. Welters, B.J. Braakhuis, M. Bagnay, A. Westerveld, and R. Slater, *Centromeric breakage as a major cause of cytogenetic abnormalities in oral squamous cell carcinoma.* Genes Chromosomes Cancer, 1996. **15**(1): p. 1-9.

205. White, J.S., J.L. Weissfeld, C.C.R. Ragin, K.M. Rossie, C.L. Martin, M. Shuster, C.S. Ishwad, J.C. Law, E.N. Myers, J.T. Johnson, and S.M. Gollin, *The influence of clinical and demographic risk factors on the establishment of head and neck squamous cell carcinoma cell lines.* Oral Oncol, 2007. **43**(7): p. 701-12.

206. Martin, C.L., S.C. Reshmi, T. Ried, W. Gottberg, J.W. Wilson, J.K. Reddy, P. Khanna, J.T. Johnson, E.N. Myers, and S.M. Gollin, *Chromosomal imbalances in oral squamous cell carcinoma: examination of 31 cell lines and review of the literature.* Oral Oncol, 2008. **44**(4): p. 369-82.

207. Steenbergen, R.D., M.A. Hermsen, J.M. Walboomers, H. Joenje, F. Arwert, C.J. Meijer, and P.J. Snijders, *Integrated human papillomavirus type 16 and loss of heterozygosity at 11q22 and 18q21 in an oral carcinoma and its derivative cell line.* Cancer Res, 1995. **55**(22): p. 5465-71.

208. Martin, D., M.C. Abba, A.A. Molinolo, L. Vitale-Cross, Z. Wang, M. Zaida, N.C. Delic, Y. Samuels, J.G. Lyons, and J.S. Gutkind, *The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies.* Oncotarget, 2014. **5**(19): p. 8906-23.

209. Drexler, H.G. and C.C. Uphoff, *Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention.* Cytotechnology, 2002. **39**(2): p. 75-90.

210. Desjardins, P. and D. Conklin, *NanoDrop microvolume quantitation of nucleic acids.* J Vis Exp, 2010(45).

211. Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen, *Primer3--new capabilities and interfaces.* Nucleic Acids Res, 2012. **40**(15): p. e115.

212. Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

213. Chan, J.K.L., D. Yuen, P.H. Too, Y. Sun, B. Willard, D. Man, and C. Tam, *Keratin 6a reorganization for ubiquitin-proteasomal processing is a direct antimicrobial response.* J Cell Biol, 2018. **217**(2): p. 731-744.

214. Lloyd, C., Q.C. Yu, J. Cheng, K. Turksen, L. Degenstein, E. Hutton, and E. Fuchs, *The basal keratin network of stratified squamous epithelia: defining K15 function in the absence of K14.* J Cell Biol, 1995. **129**(5): p. 1329-44.

215. Rosenbloom, K.R., et al., *ENCODE data in the UCSC Genome Browser: year 5 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D56-63.

216. Wiench, M., S. John, S. Baek, T.A. Johnson, M.H. Sung, T. Escobar, C.A. Simmons, K.H. Pearce, S.C. Biddie, P.J. Sabo, R.E. Thurman, J.A. Stamatoyannopoulos, and G.L. Hager, *DNA methylation status predicts cell type-specific enhancer activity.* EMBO J, 2011. **30**(15): p. 3028-39.

217. Liu, T., et al., *Cistrome: an integrative platform for transcriptional regulation studies.* Genome Biol, 2011. **12**(8): p. R83.

218. Shu, W., H. Chen, X. Bo, and S. Wang, *Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains.* Nucleic Acids Res, 2011. **39**(17): p. 7428-43.

219. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

220. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass, *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.

221. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-842.

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

223. Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-D361.

224. Slenter, D.N., et al., *WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research.* Nucleic Acids Res, 2018. **46**(D1): p. D661-D667.

225. Schroeder, A., O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, and T. Ragg, *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Mol Biol, 2006. **7**: p. 3.

226. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.* Nucleic Acids Res, 2018. **46**(W1): p. W537-W544.

227. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements.* Nat Methods, 2015. **12**(4): p. 357-60.

228. Pertea, M., D. Kim, G.M. Pertea, J.T. Leek, and S.L. Salzberg, *Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.* Nat Protoc, 2016. **11**(9): p. 1650-67.

229. Trakhtenberg, E.F., N. Pho, K.M. Holton, T.W. Chittenden, J.L. Goldberg, and L. Dong, *Cell types differ in global coordination of splicing and proportion of highly expressed genes.* Sci Rep, 2016. **6**: p. 32249.

230. Brooks, M.J., H.K. Rajasimha, J.E. Roger, and A. Swaroop, *Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes.* Mol Vis, 2011. **17**: p. 3034-54.

231. Kamburov, A., U. Stelzl, H. Lehrach, and R. Herwig, *The ConsensusPathDB interaction database: 2013 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D793-800.

232. Ormerod, M.G., M.K. Collins, G. Rodriguez-Tarduchy, and D. Robertson, *Apoptosis in interleukin-3-dependent haemopoietic cells. Quantification by two flow cytometric methods.* J Immunol Methods, 1992. **153**(1-2): p. 57-65.

233. Schutte, B., R. Nuydens, H. Geerts, and F. Ramaekers, *Annexin V binding assay as a tool to measure apoptosis in differentiated neuronal cells.* J Neurosci Methods, 1998. **86**(1): p. 63-9.

234. Nicholson, R.I., J.M. Gee, and M.E. Harper, *EGFR and cancer prognosis.* Eur J Cancer, 2001. **37 Suppl 4**: p. S9-15.

235. Ribeiro, I.P., F. Caramelo, L. Esteves, J. Menoita, F. Marques, L. Barroso, J. Migueis, J.B. Melo, and I.M. Carreira, *Genomic predictive model for recurrence and metastasis development in head and neck squamous cell carcinoma patients.* Sci Rep, 2017. **7**(1): p. 13897.

236. Garcia Pedrero, J.M., M.P. Fernandez, R.O. Morgan, A. Herrero Zapatero, M.V. Gonzalez, C. Suarez Nieto, and J.P. Rodrigo, *Annexin A1 down-regulation in head and neck cancer is associated with epithelial differentiation status.* Am J Pathol, 2004. **164**(1): p. 73-9.

237. Alvarez-Teijeiro, S., S.T. Menendez, M.A. Villaronga, E. Pena-Alonso, J.P. Rodrigo, R.O. Morgan, R. Granda-Diaz, C. Salom, M.P. Fernandez, and J.M. Garcia-Pedrero, *Annexin A1 down-regulation in head and neck squamous cell carcinoma is mediated via transcriptional control with direct involvement of miR-196a/b.* Sci Rep, 2017. **7**(1): p. 6790.

238. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.

239. Wang, Y.-M., P. Zhou, L.-Y. Wang, Z.-H. Li, Y.-N. Zhang, and Y.-X. Zhang, *Correlation between DNase I hypersensitive site distribution and gene expression in HeLa S3 cells.* PLoS One, 2012. **7**(8): p. e42414.

240. Ozanne, B.W., L. McGarry, H.J. Spence, I. Johnston, J. Winnie, L. Meagher, and G. Stapleton, *Transcriptional regulation of cell invasion: AP-1 regulation of a multigenic invasion programme.* Eur J Cancer, 2000. **36**(13 Spec No): p. 1640-8.

241. Presnell, J.S., C.E. Schnitzler, and W.E. Browne, *KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes.* Genome Biol Evol, 2015. **7**(8): p. 2289-309.

242. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.

243. Zhang, L., S. Yang, D.O. Wennmann, Y. Chen, J. Kremerskothen, and J. Dong, *KIBRA: In the brain and beyond.* Cell Signal, 2014. **26**(7): p. 1392-9.

244. Jimenez, L., S.K. Jayakar, T.J. Ow, and J.E. Segall, *Mechanisms of Invasion in Head and Neck Cancer.* Arch Pathol Lab Med, 2015. **139**(11): p. 1334-48.

245. Kooistra, M.R., N. Dube, and J.L. Bos, *Rap1: a key regulator in cell-cell junction formation.* J Cell Sci, 2007. **120**(Pt 1): p. 17-22.

246. Stavreva, D.A., A. Coulon, S. Baek, M.H. Sung, S. John, L. Stixova, M. Tesikova, O. Hakim, T. Miranda, M. Hawkins, J.A. Stamatoyannopoulos, C.C. Chow, and G.L. Hager, *Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing.* Genome Res, 2015. **25**(6): p. 845-57.

247. Swift, J. and G.M. Coruzzi, *A matter of time - How transient transcription factor interactions create dynamic gene regulatory networks.* Biochim Biophys Acta Gene Regul Mech, 2017. **1860**(1): p. 75-83.

248. Shiga, K., T. Ogawa, K. Katagiri, F. Yoshida, M. Tateda, K. Matsuura, and T. Kobayashi, *Differences between oral cancer and cancers of the pharynx and larynx on a molecular level.* Oncol Lett, 2012. **3**(1): p. 238-243.

249. Shepherd, T. and J.A. Hassell, *Role of Ets transcription factors in mammary gland development and oncogenesis.* J Mammary Gland Biol Neoplasia, 2001. **6**(1): p. 129-40.

250. Cooper, C.D., J.A. Newman, H. Aitkenhead, C.K. Allerston, and O. Gileadi, *Structures of the Ets Protein DNA-binding Domains of Transcription Factors Etv1, Etv4, Etv5, and Fev: DETERMINANTS OF DNA BINDING AND REDOX REGULATION BY DISULFIDE BOND FORMATION.* J Biol Chem, 2015. **290**(22): p. 13692-709.

251. Fossum, S.L., M.J. Mutolo, A. Tugores, S. Ghosh, S.H. Randell, L.C. Jones, S.H. Leir, and A. Harris, *Ets homologous factor (EHF) has critical roles in epithelial dysfunction in airway disease.* J Biol Chem, 2017. **292**(26): p. 10938-10949.

252. Ding, X., H. Pan, J. Li, Q. Zhong, X. Chen, S.M. Dry, and C.Y. Wang, *Epigenetic activation of AP1 promotes squamous cell carcinoma metastasis.* Sci Signal, 2013. **6**(273): p. ra28 1-13, S0-15.

253. Hyakusoku, H., D. Sano, H. Takahashi, T. Hatano, Y. Isono, S. Shimada, Y. Ito, J.N. Myers, and N. Oridate, *JunB promotes cell invasion, migration and distant metastasis of head and neck squamous cell carcinoma.* J Exp Clin Cancer Res, 2016. **35**: p. 6.

254. Mangone, F.R., M.M. Brentani, S. Nonogaki, M.D. Begnami, A.H. Campos, F. Walder, M.B. Carvalho, F.A. Soares, H. Torloni, L.P. Kowalski, and M.H. Federico, *Overexpression of Fos-related antigen-1 in head and neck squamous cell carcinoma.* Int J Exp Pathol, 2005. **86**(4): p. 205-12.

255. Ondrey, F.G., G. Dong, J. Sunwoo, Z. Chen, J.S. Wolf, C.V. Crowl-Bancroft, N. Mukaida, and C. Van Waes, *Constitutive activation of transcription factors NF-(kappa)B, AP-1, and NF-IL6 in human head and neck squamous cell carcinoma cell lines that express pro-inflammatory and pro-angiogenic cytokines.* Mol Carcinog, 1999. **26**(2): p. 119-29.

256. Zhang, X., J. Wu, S. Luo, T. Lechler, and J.Y. Zhang, *FRA1 promotes squamous cell carcinoma growth and metastasis through distinct AKT and c-Jun dependent mechanisms.* Oncotarget, 2016. **7**(23): p. 34371-83.

257. Sniezek, J.C., K.E. Matheny, M.D. Westfall, and J.A. Pietenpol, *Dominant negative p63 isoform expression in head and neck squamous cell carcinoma.* Laryngoscope, 2004. **114**(12): p. 2063-72.

258. Weber, A., U. Bellmann, F. Bootz, C. Wittekind, and A. Tannapfel, *Expression of p53 and its homologues in primary and recurrent squamous cell carcinomas of the head and neck.* Int J Cancer, 2002. **99**(1): p. 22-8.

259. Rocco, J.W., C.-O. Leong, N. Kuperwasser, M.P. DeYoung, and L.W. Ellisen, *p63 mediates survival in squamous cell carcinoma by suppression of p73-dependent apoptosis.* Cancer Cell, 2006. **9**(1): p. 45-56.

260. Tai, S.K., M.H. Yang, S.Y. Chang, Y.C. Chang, W.Y. Li, T.L. Tsai, Y.F. Wang, P.Y. Chu, and S.L. Hsieh, *Persistent Kruppel-like factor 4 expression predicts progression and poor prognosis of head and neck squamous cell carcinoma.* Cancer Sci, 2011. **102**(4): p. 895-902.

261. Scheitz, C.J., T.S. Lee, D.J. McDermitt, and T. Tumbar, *Defining a tissue stem cell-driven Runx1/Stat3 signalling axis in epithelial cancer.* EMBO J, 2012. **31**(21): p. 4124-39.

262. Lu, Y., B. Zabihula, W. Yibulayin, and X. Liu, *Methylation and expression of RECK, P53 and RUNX genes in patients with esophageal cancer.* Oncol Lett, 2017. **14**(5): p. 5293-5298.

263. Sartor, M.A., D.C. Dolinoy, T.R. Jones, J.A. Colacino, M.E. Prince, T.E. Carey, and L.S. Rozek, *Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis.* Epigenetics, 2011. **6**(6): p. 777-87.

264. Tsunematsu, T., Y. Kudo, S. Iizuka, I. Ogawa, T. Fujita, H. Kurihara, Y. Abiko, and T. Takata, *RUNX3 has an oncogenic role in head and neck cancer.* PLoS One, 2009. **4**(6): p. e5892.

265. Singchat, W., E. Hitakomate, B. Rerkarmnuaychoke, A. Suntronpong, B. Fu, W. Bodhisuwan, S. Peyachoknagul, F. Yang, S. Koontongkaew, and K. Srikulnath, *Genomic Alteration in Head and Neck Squamous Cell Carcinoma (HNSCC) Cell Lines Inferred from Karyotyping, Molecular Cytogenetics, and Array Comparative Genomic Hybridization.* PLoS One, 2016. **11**(8): p. e0160901.

266. Plotnik, J.P. and P.C. Hollenhorst, *Interaction with ZMYND11 mediates opposing roles of Ras-responsive transcription factors ETS1 and ETS2.* Nucleic Acids Res, 2017. **45**(8): p. 4452-4462.

267. Verduci, L., M. Ferraiuolo, A. Sacconi, F. Ganci, J. Vitale, T. Colombo, P. Paci, S. Strano, G. Macino, N. Rajewsky, and G. Blandino, *The oncogenic role of circPVT1 in head and neck squamous cell carcinoma is mediated through the mutant p53/YAP/TEAD transcription-competent complex.* Genome Biol, 2017. **18**(1): p. 237.

268. Wang, X., C. He, and X. Hu, *LIM homeobox transcription factors, a novel subfamily which plays an important role in cancer (review).* Oncol Rep, 2014. **31**(5): p. 1975-85.

269. Chen, W., K.L. Kang, A. Alshaikh, S. Varma, Y.L. Lin, K.H. Shin, R. Kim, C.Y. Wang, N.H. Park, K. Walentin, K.M. Schmidt-Ott, and M.K. Kang, *Grainyhead-like 2 (GRHL2) knockout abolishes oral cancer development through reciprocal regulation of the MAP kinase and TGF-beta signaling pathways.* Oncogenesis, 2018. **7**(5): p. 38.

270. Benner, C., S. Konovalov, C. Mackintosh, K.R. Hutt, R. Stunnenberg, and I. Garcia-Bassets, *Decoding a signature-based model of transcription cofactor recruitment dictated by cardinal cis-regulatory elements in proximal promoter regions.* PLoS Genet, 2013. **9**(11): p. e1003906.

271. Xie, X., J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, and M. Kellis, *Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals.* Nature, 2005. **434**(7031): p. 338-45.

272. Zanconato, F., M. Forcato, G. Battilana, L. Azzolin, E. Quaranta, B. Bodega, A. Rosato, S. Bicciato, M. Cordenonsi, and S. Piccolo, *Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth.* Nat Cell Biol, 2015. **17**(9): p. 1218-1227.

273. Matheny, K.E., C.E. Barbieri, J.C. Sniezek, C.L. Arteaga, and J.A. Pietenpol, *Inhibition of epidermal growth factor receptor signaling decreases p63 expression in head and neck squamous carcinoma cells.* Laryngoscope, 2003. **113**(6): p. 936-9.

274. Liu, X., H. Li, M. Rajurkar, Q. Li, J.L. Cotton, J. Ou, L.J. Zhu, H.L. Goel, A.M. Mercurio, J.S. Park, R.J. Davis, and J. Mao, *Tead and AP1 Coordinate Transcription and Motility.* Cell Rep, 2016. **14**(5): p. 1169-1180.

275. Bornstein, S., R. White, S. Malkoski, M. Oka, G. Han, T. Cleaver, D. Reh, P. Andersen, N. Gross, S. Olson, C. Deng, S.L. Lu, and X.J. Wang, *Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation.* J Clin Invest, 2009. **119**(11): p. 3408-19.

276. Burotto, M., V.L. Chiou, J.M. Lee, and E.C. Kohn, *The MAPK pathway across different malignancies: a new perspective.* Cancer, 2014. **120**(22): p. 3446-56.

277. Chong, C.R. and P.A. Janne, *The quest to overcome resistance to EGFR-targeted therapies in cancer.* Nat Med, 2013. **19**.

278. Zhan, T., N. Rindtorff, and M. Boutros, *Wnt signaling in cancer.* Oncogene, 2017. **36**(11): p. 1461-1473.

279. Martin, K.A., L.N. Lupey, and I. Tempera, *Epstein-Barr Virus Oncoprotein LMP1 Mediates Epigenetic Changes in Host Gene Expression through PARP1.* J Virol, 2016. **90**(19): p. 8520-30.

280. Cherian, M.A., H.H. Baydoun, J. Al-Saleem, N. Shkriabai, M. Kvaratskhelia, P. Green, and L. Ratner, *Akt Pathway Activation by Human T-cell Leukemia Virus Type 1 Tax Oncoprotein.* J Biol Chem, 2015. **290**(43): p. 26270-81.

281. Rozenblatt-Rosen, O., et al., *Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins.* Nature, 2012. **487**(7408): p. 491-5.

282. Harden, M.E. and K. Munger, *Human papillomavirus 16 E6 and E7 oncoprotein expression alters microRNA expression in extracellular vesicles.* Virology, 2017. **508**: p. 63-69.

283. Gao, D., Y. Zhang, M. Zhu, S. Liu, and X. Wang, *miRNA Expression Profiles of HPV-Infected Patients with Cervical Cancer in the Uyghur Population in China.* PLoS One, 2016. **11**(10): p. e0164701.

284. Lajer, C.B., E. Garnaes, L. Friis-Hansen, B. Norrild, M.H. Therkildsen, M. Glud, M. Rossing, H. Lajer, D. Svane, L. Skotte, L. Specht, C. Buchwald, and F.C. Nielsen, *The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer.* Br J Cancer, 2012. **106**(9): p. 1526-34.

285. Welch, R.P., C. Lee, P.M. Imbriano, S. Patil, T.E. Weymouth, R.A. Smith, L.J. Scott, and M.A. Sartor, *ChIP-Enrich: gene set enrichment testing for ChIP-seq data.* Nucleic Acids Res, 2014. **42**(13): p. e105.

286. Crawford, G.E., et al., *Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS).* Genome Res, 2006. **16**(1): p. 123-31.

287. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.

288. Thomas, S.Y., G.S. Whitehead, M. Takaku, J.M. Ward, X. Xu, K. Nakano, M.R. Lyons-Cohen, H. Nakano, K.M. Gowdy, P.A. Wade, and D.N. Cook, *MyD88-dependent dendritic and epithelial cell crosstalk orchestrates immune responses to allergens.* Mucosal Immunol, 2018. **11**(3): p. 796-810.

289. Zacharias, W.J., D.B. Frank, J.A. Zepp, M.P. Morley, F.A. Alkhaleel, J. Kong, S. Zhou, E. Cantu, and E.E. Morrisey, *Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor.* Nature, 2018. **555**(7695): p. 251-255.

290. McConnell, B.B. and V.W. Yang, *Mammalian Kruppel-like factors in health and diseases.* Physiol Rev, 2010. **90**(4): p. 1337-81.

291. Li, W., M. Liu, Y. Su, X. Zhou, Y. Liu, and X. Zhang, *The Janus-faced roles of Kruppel-like factor 4 in oral squamous cell carcinoma cells.* Oncotarget, 2015. **6**(42): p. 44480-94.

292. Zhang, W., D.E. Geiman, J.M. Shields, D.T. Dang, C.S. Mahatan, K.H. Kaestner, J.R. Biggs, A.S. Kraft, and V.W. Yang, *The gut-enriched Kruppel-like factor (Kruppel-like factor 4) mediates the transactivating effect of p53 on the p21WAF1/Cip1 promoter.* J Biol Chem, 2000. **275**(24): p. 18391-8.

293. Yoon, H.S., A.M. Ghaleb, M.O. Nandan, I.M. Hisamuddin, W.B. Dalton, and V.W. Yang, *Kruppel-like factor 4 prevents centrosome amplification following gamma-irradiation-induced DNA damage.* Oncogene, 2005. **24**(25): p. 4017-25.

294. Qiao, F., F. Yao, L. Chen, C. Lu, Y. Ni, W. Fang, and H. Jin, *Kruppel-like factor 9 was down-regulated in esophageal squamous cell carcinoma and negatively regulated beta-catenin/TCF signaling.* Mol Carcinog, 2016. **55**(3): p. 280-91.

295. Yien, Y.Y. and J.J. Bieker, *EKLF/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination.* Mol Cell Biol, 2013. **33**(1): p. 4-13.

296. Tallack, M.R. and A.C. Perkins, *KLF1 directly coordinates almost all aspects of terminal erythroid differentiation.* IUBMB Life, 2010. **62**(12): p. 886-90.

297. Zhu, B., Q. Liu, Q. Han, B. Zeng, J. Chen, and Q. Xiao, *Downregulation of Kruppellike factor 1 inhibits the metastasis and invasion of cervical cancer cells.* Mol Med Rep, 2018. **18**(4): p. 3932-3940.

298. Jost, M., T.M. Huggett, C. Kari, L.H. Boise, and U. Rodeck, *Epidermal growth factor receptor-dependent control of keratinocyte survival and Bcl-xL expression through a MEK-dependent pathway.* J Biol Chem, 2001. **276**(9): p. 6320-6.

299. Seoane, J. and R.R. Gomis, *TGF-beta Family Signaling in Tumor Suppression and Cancer Progression.* Cold Spring Harb Perspect Biol, 2017. **9**(12): p. a022277.

300. Lubov, J., M. Maschietto, I. Ibrahim, A. Mlynarek, M. Hier, L.P. Kowalski, M.A. Alaoui-Jamali, and S.D. da Silva, *Meta-analysis of microRNAs expression in head and neck cancer: uncovering association with outcome and mechanisms.* Oncotarget, 2017. **8**(33): p. 55511-55524.

301. Liu, J., K.M. Drescher, and X.M. Chen, *MicroRNAs and Epithelial Immunity.* Int Rev Immunol, 2009. **28**(3-4): p. 139-54.

302. Johnson, R. and G. Halder, *The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment.* Nat Rev Drug Discov, 2014. **13**(1): p. 63-79.

303. Wu, L. and X. Yang, *Targeting the Hippo Pathway for Breast Cancer Therapy.* Cancers (Basel), 2018. **10**(11): p. E422.

304. Moh, M.C. and S. Shen, *The roles of cell adhesion molecules in tumor suppression and cell migration: a new paradox.* Cell Adh Migr, 2009. **3**(4): p. 334-6.

305. Sun, S. and K.D. Irvine, *Cellular Organization and Cytoskeletal Regulation of the Hippo Signaling Network.* Trends Cell Biol, 2016. **26**(9): p. 694-704.

306. Thiery, J.P., H. Acloque, R.Y. Huang, and M.A. Nieto, *Epithelial-mesenchymal transitions in development and disease.* Cell, 2009. **139**(5): p. 871-90.

307. Hess, J., P. Angel, and M. Schorpp-Kistner, *AP-1 subunits: quarrel and harmony among siblings.* J Cell Sci, 2004. **117**(Pt 25): p. 5965-73.

308. Xu, H., X. Jin, Y. Yuan, P. Deng, L. Jiang, X. Zeng, X.S. Li, Z.Y. Wang, and Q.M. Chen, *Prognostic value from integrative analysis of transcription factors c-Jun and Fra-1 in oral squamous cell carcinoma: a multicenter cohort study.* Sci Rep, 2017. **7**(1): p. 7522.

309. Rocco, J.W., C.O. Leong, N. Kuperwasser, M.P. DeYoung, and L.W. Ellisen, *p63 mediates survival in squamous cell carcinoma by suppression of p73-dependent apoptosis.* Cancer Cell, 2006. **9**(1): p. 45-56.

310. Limame, R., K. Op de Beeck, F. Lardon, O. De Wever, and P. Pauwels, *Kruppel-like factors in cancer progression: three fingers on the steering wheel.* Oncotarget, 2014. **5**(1): p. 29-48.

311. Shi, W., J. Yang, S. Li, X. Shan, X. Liu, H. Hua, C. Zhao, Z. Feng, Z. Cai, L. Zhang, and D. Zhou, *Potential involvement of miR-375 in the premalignant progression of oral squamous cell carcinoma mediated via transcription factor KLF5.* Oncotarget, 2015. **6**(37): p. 40172-85.

312. Kaneko, K.J. and M.L. DePamphilis, *Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors.* Developmental Genetics, 1998. **22**(1): p. 43-55.

313. Pan, D., *The hippo signaling pathway in development and cancer.* Dev Cell, 2010. **19**(4): p. 491-505.

314. Segrelles, C., J.M. Paramio, and C. Lorz, *The transcriptional co-activator YAP: A new player in head and neck cancer.* Oral Oncology, 2018. **86**: p. 25-32.

315. Garcia-Escudero, R., C. Segrelles, M. Duenas, M. Pombo, C. Ballestin, M. Alonso-Riano, P. Nenclares, R. Alvarez-Rodriguez, G. Sanchez-Aniceto, A. Ruiz-Alonso, J.L. Lopez-Cedrun, J.M. Paramio, and C. Lorz, *Overexpression of PIK3CA in head and neck squamous cell carcinoma is associated with poor outcome and activation of the YAP pathway.* Oral Oncol, 2018. **79**: p. 55-63.

316. Takeuchi, S., A. Kasamatsu, M. Yamatoji, D. Nakashima, Y. Endo-Sakamoto, N. Koide, T. Takahara, T. Shimizu, M. Iyoda, K. Ogawara, M. Shiiba, H. Tanzawa, and K. Uzawa, *TEAD4-YAP interaction regulates tumoral growth by controlling cell-cycle arrest at the G1 phase.* Biochem Biophys Res Commun, 2017. **486**(2): p. 385-390.

317. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

318. Gao, J., B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, and N. Schultz, *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.

319. Koch, A., T. De Meyer, J. Jeschke, and W. Van Criekinge, *MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data.* BMC Genomics, 2015. **16**: p. 636.

320. Thompson, M.R., D. Xu, and B.R. Williams, *ATF3 transcription factor and its emerging roles in immunity and cancer.* J Mol Med (Berl), 2009. **87**(11): p. 1053-60.

321. Gertz, J., D. Savic, K.E. Varley, E.C. Partridge, A. Safi, P. Jain, G.M. Cooper, T.E. Reddy, G.E. Crawford, and R.M. Myers, *Distinct properties of cell-type-specific and shared transcription factor binding sites.* Mol Cell, 2013. **52**(1): p. 25-36.

322. Pope, B.D., et al., *Topologically associating domains are stable units of replication-timing regulation.* Nature, 2014. **515**(7527): p. 402-5.

323. Chan, L.P., L.F. Wang, F.Y. Chiang, K.W. Lee, P.L. Kuo, and C.H. Liang, *IL-8 promotes HNSCC progression on CXCR1/2-meidated NOD1/RIP2 signaling pathway.* Oncotarget, 2016. **7**(38): p. 61820-61831.

324. Eun, Y.G., D. Lee, Y.C. Lee, B.H. Sohn, E.H. Kim, S.Y. Yim, K.H. Kwon, and J.S. Lee, *Clinical significance of YAP1 activation in head and neck squamous cell carcinoma.* Oncotarget, 2017. **8**(67): p. 111130-111143.

325. Jerhammar, F., A.C. Johansson, R. Ceder, J. Welander, A. Jansson, R.C. Grafstrom, P. Soderkvist, and K. Roberg, *YAP1 is a potential biomarker for cetuximab resistance in head and neck cancer.* Oral Oncology, 2014. **50**(9): p. 832-839.

326. Plouffe, S.W., K.C. Lin, J.L. Moore, 3rd, F.E. Tan, S. Ma, Z. Ye, Y. Qiu, B. Ren, and K.L. Guan, *The Hippo pathway effector proteins YAP and TAZ have both distinct and overlapping functions in the cell.* J Biol Chem, 2018. **293**(28): p. 11230-11240.

327. Hiemer, S.E., L. Zhang, V.K. Kartha, T.S. Packer, M. Almershed, V. Noonan, M. Kukuruzinska, M.V. Bais, S. Monti, and X. Varelas, *A YAP/TAZ-Regulated Molecular Signature Is Associated with Oral Squamous Cell Carcinoma.* Mol Cancer Res, 2015. **13**(6): p. 957-68.

328. Muramatsu, T., I. Imoto, T. Matsui, K. Kozaki, S. Haruki, M. Sudol, Y. Shimada, H. Tsuda, T. Kawano, and J. Inazawa, *YAP is a candidate oncogene for esophageal squamous cell carcinoma.* Carcinogenesis, 2011. **32**(3): p. 389-98.

329. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. **119**(6): p. 1420-8.

330. Si, H., et al., *TNF-alpha modulates genome-wide redistribution of DeltaNp63alpha/TAp73 and NF-kappaB cREL interactive binding on TP53 and AP-1 motifs to promote an oncogenic gene program in squamous cancer.* Oncogene, 2016. **35**(44): p. 5781-5794.

331. Valencia-Sama, I., Y. Zhao, D. Lai, H.J. Janse van Rensburg, Y. Hao, and X. Yang, *Hippo Component TAZ Functions as a Co-repressor and Negatively Regulates DeltaNp63 Transcription through TEA Domain (TEAD) Transcription Factor.* J Biol Chem, 2015. **290**(27): p. 16906-17.

332. Chatterjee, A., T. Sen, X. Chang, and D. Sidransky, *Yes-associated protein 1 regulates the stability of DeltaNp63alpha.* Cell Cycle, 2010. **9**(1): p. 162-7.

333. Ehsanian, R., M. Brown, H. Lu, X.P. Yang, A. Pattatheyil, B. Yan, P. Duggal, R. Chuang, J. Doondeea, S. Feller, M. Sudol, Z. Chen, and C. Van Waes, *YAP dysregulation by phosphorylation or DeltaNp63-mediated gene repression promotes proliferation, survival and migration in head and neck cancer subsets.* Oncogene, 2010. **29**(46): p. 6160-71.

334. Zhao, B., X. Ye, J. Yu, L. Li, W. Li, S. Li, J. Yu, J.D. Lin, C.Y. Wang, A.M. Chinnaiyan, Z.C. Lai, and K.L. Guan, *TEAD mediates YAP-dependent gene induction and growth control.* Genes Dev, 2008. **22**(14): p. 1962-71.

335. Lamar, J.M., P. Stern, H. Liu, J.W. Schindler, Z.G. Jiang, and R.O. Hynes, *The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain.* Proc Natl Acad Sci U S A, 2012. **109**(37): p. E2441-50.

336. Sur, I., B. Rozell, V. Jaks, A. Bergstrom, and R. Toftgard, *Epidermal and craniofacial defects in mice overexpressing Klf5 in the basal layer of the epidermis.* J Cell Sci, 2006. **119**(Pt 17): p. 3593-601.

337. Lin, S.C., M.A. Wani, J.A. Whitsett, and J.M. Wells, *Klf5 regulates lineage formation in the pre-implantation mouse embryo.* Development, 2010. **137**(23): p. 3953-63.

338. Nho, R.S., H. Xia, J. Kahm, J. Kleidon, D. Diebold, and C.A. Henke, *Role of integrin-linked kinase in regulating phosphorylation of Akt and fibroblast survival in type I collagen matrices through a beta1 integrin viability signaling pathway.* J Biol Chem, 2005. **280**(28): p. 26630-9.

339. Lehtinen, M.K. and A. Bonni, *Demystifying MST family kinases in cell death.* Curr Mol Med, 2008. **8**(4): p. 313-8.

340. Serrano, I., P.C. McDonald, F. Lock, W.J. Muller, and S. Dedhar, *Inactivation of the Hippo tumour suppressor pathway by integrin-linked kinase.* Nat Commun, 2013. **4**: p. 2976.

341. Sit, S.T. and E. Manser, *Rho GTPases and their role in organizing the actin cytoskeleton.* J Cell Sci, 2011. **124**(Pt 5): p. 679-83.

342. Juan, W.C. and W. Hong, *Targeting the Hippo Signaling Pathway for Tissue Regeneration and Cancer Therapy.* Genes (Basel), 2016. **7**(9): p. E55.

343. Nakatani, K., T. Maehama, M. Nishio, H. Goto, W. Kato, H. Omori, Y. Miyachi, H. Togashi, Y. Shimono, and A. Suzuki, *Targeting the Hippo signalling pathway for cancer treatment.* J Biochem, 2017. **161**(3): p. 237-244.

344. Zhao, B., L. Li, Q. Lu, L.H. Wang, C.Y. Liu, Q. Lei, and K.L. Guan, *Angiomotin is a novel Hippo pathway component that inhibits YAP oncoprotein.* Genes Dev, 2011. **25**(1): p. 51-63.

345. Johnson, D.E., *Targeting Cell Survival Pathways to Enhance Response to Chemotherapy*. 1st ed. Cancer sensitizing agents for chemotherapy volume 3. 2018, UK: Academic Press. 308.

346. Henney, J.E., *From the Food and Drug Administration.* JAMA, 2000. **283**(21): p. 2779.

347. Sen Sharma, S. and S.S. Majumdar, *Transcriptional co-activator YAP regulates cAMP signaling in Sertoli cells.* Mol Cell Endocrinol, 2017. **450**: p. 64-73.

348. Liu-Chittenden, Y., B. Huang, J.S. Shim, Q. Chen, S.J. Lee, R.A. Anders, J.O. Liu, and D. Pan, *Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP.* Genes Dev, 2012. **26**(12): p. 1300-5.

349. Dasari, V.R., V. Mazack, W. Feng, J. Nash, D.J. Carey, and R. Gogoi, *Verteporfin exhibits YAP-independent anti-proliferative and cytotoxic effects in endometrial cancer cells.* Oncotarget, 2017. **8**(17): p. 28628-28640.

350. Zhang, W.Q., Y.Y. Dai, P.C. Hsu, H. Wang, L. Cheng, Y.L. Yang, Y.C. Wang, Z.D. Xu, S. Liu, G. Chan, B. Hu, H. Li, D.M. Jablons, and L. You, *Targeting YAP in malignant pleural mesothelioma.* J Cell Mol Med, 2017: p. 2663-2676.

351. Wang, C., X. Zhu, W. Feng, Y. Yu, K. Jeong, W. Guo, Y. Lu, and G.B. Mills, *Verteporfin inhibits YAP function through up-regulating 14-3-3sigma sequestering YAP in the cytoplasm.* Am J Cancer Res, 2016. **6**(1): p. 27-37.

352. Kandoussi, I., W. Lakhlili, J. Taoufik, and A. Ibrahimi, *Docking analysis of verteporfin with YAP WW domain.* Bioinformation, 2017. **13**(7): p. 237-240.

353. Zhou, Z., T. Hu, Z. Xu, Z. Lin, Z. Zhang, T. Feng, L. Zhu, Y. Rong, H. Shen, J.M. Luk, X. Zhang, and N. Qin, *Targeting Hippo pathway by specific interruption of YAP-TEAD interaction using cyclic YAP-like peptides.* FASEB J, 2015. **29**(2): p. 724-32.

354. Boeckx, C., K. Op de Beeck, A. Wouters, V. Deschoolmeester, R. Limame, K. Zwaenepoel, P. Specenier, P. Pauwels, J.B. Vermorken, M. Peeters, G. Van Camp, M. Baay, and F. Lardon, *Overcoming cetuximab resistance in HNSCC: the role of AURKB and DUSP proteins.* Cancer Lett, 2014. **354**(2): p. 365-77.

355. He, C., D. Mao, G. Hua, X. Lv, X. Chen, P.C. Angeletti, J. Dong, S.W. Remmenga, K.J. Rodabaugh, J. Zhou, P.F. Lambert, P. Yang, J.S. Davis, and C. Wang, *The Hippo/YAP pathway interacts with EGFR signaling and HPV oncoproteins to regulate cervical cancer progression.* EMBO Mol Med, 2015. **7**(11): p. 1426-49.

356. Song, S., S. Honjo, J. Jin, S.-S. Chang, A.W. Scott, Q. Chen, N. Kalhor, A.M. Correa, W.L. Hofstetter, C.T. Albarracin, T.-T. Wu, R.L. Johnson, M.-C. Hung, and J.A. Ajani, *The Hippo Coactivator YAP1 Mediates EGFR Overexpression and Confers Chemoresistance in Esophageal Cancer.* Clin Cancer Res, 2015. **21**(11): p. 2580-90.

357. Hong, S.A., S.H. Jang, M.H. Oh, S.J. Kim, J.H. Kang, and S.H. Hong, *Overexpression of YAP1 in EGFR mutant lung adenocarcinoma prior to tyrosine kinase inhibitor therapy is associated with poor survival.* Pathol Res Pract, 2018: p. 335-342.

358. Xia, H., X. Dai, H. Yu, S. Zhou, Z. Fan, G. Wei, Q. Tang, Q. Gong, and F. Bi, *EGFR-PI3K-PDK1 pathway regulates YAP signaling in hepatocellular carcinoma: the mechanism and its implications in targeted therapy.* Cell Death Dis, 2018. **9**(3): p. 269.

359. Jerhammar, F., A.C. Johansson, R. Ceder, J. Welander, A. Jansson, R.C. Grafstrom, P. Soderkvist, and K. Roberg, *YAP1 is a potential biomarker for cetuximab resistance in head and neck cancer.* Oral Oncol, 2014. **50**(9): p. 832-9.

360. Rebucci, M., P. Peixoto, A. Dewitte, N. Wattez, M.A. De Nuncques, N. Rezvoy, C. Vautravers-Dewas, M.P. Buisine, E. Guerin, J.P. Peyrat, E. Lartigau, and A. Lansiaux, *Mechanisms underlying resistance to cetuximab in the HNSCC cell line: role of AKT inhibition in bypassing this resistance.* Int J Oncol, 2011. **38**(1): p. 189-200.

361. Mandic, R., C.J. Rodgarkia-Dara, L. Zhu, B.J. Folz, M. Bette, E. Weihe, A. Neubauer, and J.A. Werner, *Treatment of HNSCC cell lines with the EGFR-specific inhibitor cetuximab (Erbitux) results in paradox phosphorylation of tyrosine 1173 in the receptor.* FEBS Lett, 2006. **580**(20): p. 4793-800.

362. Jedlinski, A., S. Garvin, A.C. Johansson, P.H. Edqvist, F. Ponten, and K. Roberg, *Cetuximab sensitivity of head and neck squamous cell carcinoma xenografts is associated with treatment-induced reduction in EGFR, pEGFR, and pSrc.* J Oral Pathol Med, 2017. **46**(9): p. 717-724.

363. Garcia-Escudero, R., C. Segrelles, M. Duenas, M. Pombo, C. Ballestin, M. Alonso-Riano, P. Nenclares, R. Alvarez-Rodriguez, G. Sanchez-Aniceto, A. Ruiz-Alonso, J.L. Lopez-Cedrun, J.M. Paramio, and C. Lorz, *Overexpression of PIK3CA in head and neck squamous cell carcinoma is associated with poor outcome and activation of the YAP pathway.* Oral Oncology, 2018. **79**: p. 55-63.

364. Wei, H., F. Wang, Y. Wang, T. Li, P. Xiu, J. Zhong, X. Sun, and J. Li, *Verteporfin suppresses cell survival, angiogenesis and vasculogenic mimicry of pancreatic ductal adenocarcinoma via disrupting the YAP-TEAD complex.* Cancer Sci, 2017. **108**(3): p. 478-487.

365. Brodowska, K., A. Al-Moujahed, A. Marmalidou, M. Meyer Zu Horste, J. Cichy, J.W. Miller, E. Gragoudas, and D.G. Vavvas, *The clinically used photosensitizer Verteporfin (VP) inhibits YAP-TEAD and human retinoblastoma cell growth in vitro without light activation.* Exp Eye Res, 2014. **124**: p. 67-73.

366. Feng, J., J. Gou, J. Jia, T. Yi, T. Cui, and Z. Li, *Verteporfin, a suppressor of YAP-TEAD complex, presents promising antitumor properties on ovarian cancer.* Onco Targets Ther, 2016. **9**: p. 5371-81.

367. Dong, L., F. Lin, W. Wu, Y. Liu, and W. Huang, *Verteporfin inhibits YAP-induced bladder cancer cell growth and invasion via Hippo signaling pathway.* Int J Med Sci, 2018. **15**(6): p. 645-652.

368. Gibault, F., F. Bailly, M. Corvaisier, M. Coevoet, G. Huet, P. Melnyk, and P. Cotelle, *Molecular Features of the YAP Inhibitor Verteporfin: Synthesis of Hexasubstituted Dipyrrins as Potential Inhibitors of YAP/TAZ, the Downstream Effectors of the Hippo Pathway.* ChemMedChem, 2017. **12**(12): p. 954-961.

369. Leal, M.F., H.F. Ribeiro, J.A. Rey, G.R. Pinto, M.C. Smith, C.A. Moreira-Nunes, P.P. Assumpcao, L.M. Lamarao, D.Q. Calcagno, R.C. Montenegro, and R.R. Burbano, *YWHAE silencing induces cell proliferation, invasion and migration through the up-regulation of CDC25B and MYC in gastric cancer cells: new insights about YWHAE role in the tumor development and metastasis process.* Oncotarget, 2016. **7**(51): p. 85393-85410.

370. Kim, E.J., S.H. Kim, X. Jin, X. Jin, and H. Kim, *KCTD2, an adaptor of Cullin3 E3 ubiquitin ligase, suppresses gliomagenesis by destabilizing c-Myc.* Cell Death Differ, 2017. **24**(4): p. 649-659.

371. Gao, Y., H. Li, X. Ma, Y. Fan, D. Ni, Y. Zhang, Q. Huang, K. Liu, X. Li, L. Wang, Y. Yao, Q. Ai, and X. Zhang, *E2F3 upregulation promotes tumor malignancy through the transcriptional activation of HIF-2alpha in clear cell renal cell carcinoma.* Oncotarget, 2017. **8**(33): p. 54021-54036.

372. Tanasubsinn, P., W.P.P. Aung, S. Pata, W. Laopajon, A. Makeudom, T. Sastraruji, W. Kasinrerk, and S. Krisanaprakornkit, *Overexpression of ADAM9 in oral squamous cell carcinoma.* Oncol Lett, 2018. **15**(1): p. 495-502.

373. Zhang, H., C.Y. Liu, Z.Y. Zha, B. Zhao, J. Yao, S. Zhao, Y. Xiong, Q.Y. Lei, and K.L. Guan, *TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition.* J Biol Chem, 2009. **284**(20): p. 13355-62.

374. Meng, Z., T. Moroishi, and K.L. Guan, *Mechanisms of Hippo pathway regulation.* Genes Dev, 2016. **30**(1): p. 1-17.

375. Geles, K.G., W. Zhong, S.K. O'Brien, M. Baxter, C. Loreth, D. Pallares, and M. Damelin, *Upregulation of RNA Processing Factors in Poorly Differentiated Lung Cancer Cells.* Transl Oncol, 2016. **9**(2): p. 89-98.

376. Radhakrishnan, A., et al., *Dysregulation of splicing proteins in head and neck squamous cell carcinoma.* Cancer Biol Ther, 2016. **17**(2): p. 219-29.

377. Sveen, A., S. Kilpinen, A. Ruusulehto, R.A. Lothe, and R.I. Skotheim, *Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes.* Oncogene, 2016. **35**(19): p. 2413-27.

378. Eferl, R. and E.F. Wagner, *AP-1: a double-edged sword in tumorigenesis.* Nat Rev Cancer, 2003. **3**(11): p. 859-68.

379. Sheu, J.J., et al., *Functional genomic analysis identified epidermal growth factor receptor activation as the most common genetic event in oral squamous cell carcinoma.* Cancer Res, 2009. **69**(6): p. 2568-76.

380. Pickhard, A.C., J. Margraf, A. Knopf, T. Stark, G. Piontek, C. Beck, A.L. Boulesteix, E.Q. Scherer, S. Pigorsch, J. Schlegel, W. Arnold, and R. Reiter, *Inhibition of radiation induced migration of human head and neck squamous cell carcinoma cells by blocking of EGF receptor pathways.* BMC Cancer, 2011. **11**: p. 388.

381. Baba, Y., T. Maeda, A. Suzuki, S. Takada, M. Fujii, and Y. Kato, *Deguelin Potentiates Apoptotic Activity of an EGFR Tyrosine Kinase Inhibitor (AG1478) in PIK3CA-Mutated Head and Neck Squamous Cell Carcinoma.* Int J Mol Sci, 2017. **18**(2): p. 262.

382. Yamaoka, T., F. Yan, H. Cao, S.S. Hobbs, R.S. Dise, W. Tong, and D.B. Polk, *Transactivation of EGF receptor and ErbB2 protects intestinal epithelial cells from TNF-induced apoptosis.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11772-7.

383. Du, L., J. Shen, A. Weems, and S.L. Lu, *Role of phosphatidylinositol-3-kinase pathway in head and neck squamous cell carcinoma.* J Oncol, 2012. **2012**: p. 450179.

384. Jin, J.Y., H. Ke, R.P. Hall, and J.Y. Zhang, *c-Jun promotes whereas JunB inhibits epidermal neoplasia.* J Invest Dermatol, 2011. **131**(5): p. 1149-58.

385. Mishra, A., A.C. Bharti, D. Saluja, and B.C. Das, *Transactivation and expression patterns of Jun and Fos/AP-1 super-family proteins in human oral cancer.* Int J Cancer, 2010. **126**(4): p. 819-29.

386. Alamoud, K.A. and M.A. Kukuruzinska, *Emerging Insights into Wnt/beta-catenin Signaling in Head and Neck Cancer.* J Dent Res, 2018. **97**(6): p. 665-673.

387. Morishita, T., F. Hayakawa, K. Sugimoto, M. Iwase, H. Yamamoto, D. Hirano, Y. Kojima, N. Imoto, T. Naoe, and H. Kiyoi, *The photosensitizer verteporfin has light-independent anti-leukemic activity for Ph-positive acute lymphoblastic leukemia and synergistically works with dasatinib.* Oncotarget, 2016. **7**(35): p. 56241-56252.

388. Konstantinou, E.K., S. Notomi, C. Kosmidou, K. Brodowska, A. Al-Moujahed, F. Nicolaou, P. Tsoka, E. Gragoudas, J.W. Miller, L.H. Young, and D.G. Vavvas, *Verteporfin-induced formation of protein cross-linked oligomers and high molecular weight complexes is mediated by light and leads to cell toxicity.* Sci Rep, 2017. **7**: p. 46581.

389. Joo, S.H., *Cyclic peptides as therapeutic agents and biochemical tools.* Biomol Ther (Seoul), 2012. **20**(1): p. 19-26.

390. Pobbati, A.V., X. Han, A.W. Hung, S. Weiguang, N. Huda, G.Y. Chen, C. Kang, C.S. Chia, X. Luo, W. Hong, and A. Poulsen, *Targeting the Central Pocket in Human Transcription Factor TEAD as a Potential Cancer Therapeutic Strategy.* Structure, 2015. **23**(11): p. 2076-86.

391. Shi, Z., F. He, M. Chen, L. Hua, W. Wang, S. Jiao, and Z. Zhou, *DNA-binding mechanism of the Hippo pathway transcription factor TEAD4.* Oncogene, 2017. **36**(30): p. 4362-4369.