Oral Squamous Cell Carcinomas and their Genetic Variants in Association with Extra Capsular Spread

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Priyanka Bhattacharya

August 2017

Declaration

I declare that the work in this dissertation was carried out in accordance with the regulations of the University of Liverpool. The work is original, except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. The author in conjunction with her supervisor, Professor David Ross Sibson, conducted all aspects of experimental design and planning for this study. The author performed all experiments included in this thesis entirely independently, except running the DNA fragment libraries from cell culture on SOLiD platform and initial data analysis of whole genome sequencing and Ion data analysis using Perl script for which assistance was used. The thesis has not been presented to any other University for examination either in the United Kingdom or overseas.

This thesis is dedicated to three pillars of my life whose love, support and strength have got me this far.

My mother
My father
And
My husband

Abstract

Oral Squamous Cell Carcinomas and their Genetic Variants in Association with Extra Capsular Spread

Priyanka Bhattacharya

Oral squamous cell carcinoma (OSCC) is the sixth most common cancer worldwide with an incidence of ~600,000 cases per year and younger age groups, particularly between 40-69 years, are increasingly becoming affected. OSCC has an overall five-year survival of only ~30% when metastatic disease is present. The most adverse prognostic clinical predictor for OSCC is extracapsular spread (ECS) from lymph node metastasis. The molecular determinants of ECS are undetermined and their characterization could significantly assist patient management, identifying those having increased risk and suggesting new possibilities for therapy. The aim of this study was to identify genetic variants associated with ECS. Previous work using aCGH had found an unexpectedly low frequency of TP53 alteration in OSCCs from a local cohort enriched for cases with ECS. SNP genotyping was therefore used to further investigate allelic imbalance of TP53 and for comparison CCND1 in primary tumour samples from the series (n=46; 19 node negative [N-], 7 [N+ECS-], 20 [N+ECS-]). CCND1 gain and TP53 loss were found to be associated with nodal status but not ECS. Five OSCC primary tumour cell lines (2 N- & 3 N+ECS+) were then screened by whole genome sequencing for genetic variants that may be associated with ECS. Variants were found in a broad range of genes but poor coverage (< 5 fold mean coverage) from some regions of the genome, in particular NOTCH1, suggested that there may have been technical limitations. However, pathway analysis implicated the NOTCH pathway may have had significance, consistent with literature reports. Additional next generation sequencing assays targeting NOTCH1 were therefore developed and used to analyze the gene in these cell lines. High coverage (150-200 fold mean coverage) of NOTCH1 sequences were achieved and in one of the ECS+ve cell lines (Liv7K), a rare, potentially deleterious nonsynonymous polymorphism (rs61751543) was found. Further screening for NOTCH1 and TP53 genetic variants associated with ECS was therefore performed for primary tumour samples (n=50; 21[N-], 11 [N+ECS-], 18 [N+ECS-]). Sanger resequencing was used for independent confirmation of candidate variants identified. Poorly covered samples (n=10) were excluded. 7/40 (17.5%) samples showed NOTCH1 variants, with 6/14 (43%) ECS+ samples confirmed to have NOTCH1 variants, compared to 1/26 (3.8%) ECS-, Fisher's exact test p = 0.0044. None of these samples displayed variants in TP53. Using this approach, comparing the frequency of the variants to node status, we produced evidence for a unique subset of ECS positive OSCC cases characterized by having disruption of NOTCH1 and absence of TP53 alteration. The NOTCH1 variants were clustered within the region for the extracellular domain (ECD), suggesting that intercellular communication and response could be disrupted. Further functional studies using an ECS +ve primary OSCC cell line as well as invasive and non-invasive controls, showed invasive cell lines demonstrated decreased cell growth and migration in response to Notch inhibition by DAPT (25 µM) treatment, with western blotting revealing absence of Notch1 and 3 after week three and Notch4 by week four only in Liv7K. These findings suggested the importance of further functional studies to assess the effect of Notch inhibition on markers of aggressive phenotype. Taken together our observations overall pave the way to future studies exploring the downstream effects of NOTCH1 ECD mutations and their potential role in OSCC.

Acknowledgements

Firstly, I would like to thank my supervisor Professor David Ross Sibson who has supported me throughout my PhD with his constant guidance and encouragement. I am also thankful to him for originally giving me the opportunity to join his group as a research assistant and from there with his constant guidance I have developed my research project into the present PhD. I wish one day I can perfect the power of thinking that he has taught me throughout my PhD.

I am also thankful to Dr. Bryony Lloyd who initially gave me a foundation stone for my PhD by way of her data to use as a starting point for my project and also for guiding me in the early days of establishing the next generation sequencing techniques that I have used. I thank all the previous lab members especially Dr. Mike Davies, Dr. Fiona Gibbs, Dr. Elizabeth Shaw and Dr. Stephen Bennett, Dr. John Stanbury, Mrs. Lynn Frodsham for their advice and support at various times during my PhD.

Thanks to Mr. Jag Dhanda and Consultant surgeon, Mr. Richard Shaw, and pathologist, Dr. Janet Risk, for kindly providing me primary oral cancer cell lines and all the clinical information required for the samples used in the study.

Special thanks to Dr. Chris Mundy who has given me constant encouragement and advice throughout my PhD. His help and time made the long days while doing experiments more tolerable, especially during some of the most challenging periods of my PhD when he was always there to turn to for support. I would also like to thank my Father and Mother-in-law, Drs P.S & R Mukhopadhyay, for their support during my PhD. Their help at a crucial time when I was a new mother and writing this thesis and preparing for my viva is something that I am truly grateful for.

Finally, I would like to thank Clatterbridge Cancer Research who funded my research. Without their support none of this would have been possible.

Table of Contents

List of Figures

List of Tables

Abbreviations

S.I (Systeme International d'Unites) abbreviations for units are used in this thesis. Other abbreviations are listed below.

ANK = Ankyrin Repeats

APH1 = Anterior Pharynx-defective 1

APP = Amyloid Precursor Protein

Aβ = Amyloid Beta

BAM = Binary Sequence Alignment

BLAST = Basic Local Alignment Search Tool

BSA = Bovine Serum Albumin

CAN = Candidate Cancer Genes

CCND1 = Cyclin D1

CDK = Cyclin Dependent Kinases

cDNA = Complementary DNA

CGH = Comparative Genomic Hybridisation

CRD = Cysteine Rich Domain

cRNA = Complementary RNA

CT = Computed Tomography

Ct = Threshold Cycle

CTF = C-terminal Fragment

DAPT = N-[N-(3,5-difluorophenacetyl)- L-alanyl]-S-phenylglycine t-butyl

DMSO = Dimethyl Sulfoxide

ddNTP = Dideooxynucleotides

DNA = Deoxyribonucleic Acid

dNTPs = Deoxynucleotidephosphates

dsDNA = Double Stranded Deoxyribonucleic Acid

EB = Elution Buffer

EBV = Epstein Barr Virus

ECD = Extracellular Domain

ECS = Extracapsular Spread

EGFR = Epidermal Growth Factor Receptor

EMT = Epithelial Mesenchymal Transition

ePCR = Emulsion PCR

FA = Fanconi Anaemia

FBS = Fetal Bovine Serum

FFPE = Formalin Fixed Paraffin Embedded

FISH = Fluorescence In-situ Hybridisation

FRET = Fluorescence Resonance Energy Transfer

HD = Homozygous Deletion

HES = Hairy Enhancer of Split Family

Hg = Human Genome

HNSCC = Head and Neck Squamous Cell Carcinoma

HPLC = High-Performance Liquid Chromatography

HPV = Human Papilloma Virus

HRC1 = Human Random Control Panel 1

IGV = Integrative Genomics Viewer

ISP = Ion Sphere Particles

LFS = Li-Fraumeni Syndrome

LOH = Loss of Heterozygosity

MAPK = Mitogen-Activated Protein Kinase Pathway

MGB = Minor Groove Binder

MRI = Magnetic Resonance Imaging

mRNA = Messenger RNA

NCBI = National Center for Biotechnology Information

NCSTN = Nicastrin

NFQ = Nonfluorescent Quencher

NGS = Next Generation Sequencing

NICD = Notch Intracellular Domain

nm = Nanometers

NRR = Negative Regulatory Region

NTF = N-Terminal Fragment

OD = Optical Density

OSCC = Oral Squamous Cell Carcinoma

PBS = Phosphate Buffered Saline

PCR = Polymerase Chain Reaction

PEN2 = Presenilin Enhancer 2

PET = Positron Emission Tomography

PPi = Pyrophosphate

PSEN1 = Presenilin 1

RAM = RBP-Jκ-Associated Module Domain

Rb = Retinoblastoma Protein

RNA = Ribonucleic Acid

SCC = Squamous Cell Carcinoma

SCNA = Somatic Copy Number Alterations

SFF = Standard Flowgram Format

SNPs = Single Nucleotide Polymorphisms

STAT = Signal Transducers and Activators of Transcription

TAD = Transcription Activation Domain

TAE buffer = Tris-acetate-EDTA buffer

Tm = Melting Temperature

TMAP = Torrent Mapping Alignment Program

TP53 = Tumour protein 53

TSG = Tumour suppressor gene

WFA = Work flow analysis.

WGS = Whole genome sequencing

Chapter 1 : Introduction

1.1 The clinical biology of OSCC

1.1.1 Epidemiology of OSCC

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide with an incidence of ~600,000 cases per year [1] and oral squamous cell carcinoma (OSCC) which represents 90% of oral cavity tumours [2] is the most common subtype. In 2011, there were 6,767 new cases of oral cancer diagnosed in the UK (Cancer Research UK, 'CancerStats') with a male-female ratio of 3:1 [3]. OSCC refers specifically to squamous cell carcinomas arising in the oral cavity or oropharynx (Figure 1.) [4]. In Western countries, ~50% of all cases of OSCC affect the tongue (20% - 40%) and the floor of the mouth (15% - 20%) [5]. The gingivae, palate, retromolar area and the buccal and labial mucosa are less frequently affected [6]. The five year overall survival (OS) is ~50% [7] but only ~30% when metastatic disease is present [8]. Although the incidence of OSCC increases with age, a recent trend of concern is an increasing occurrence in younger age groups, particularly in individuals between aged 40-69 years [3]. Risk factors include alcohol and tobacco use, see 1.1.2.1 below. Given the decreasing tobacco use in western countries, a worldwide increase in alcohol consumption or Human Papilloma Virus (HPV) infection are possible reasons for the increase [9] and the current epidemiological data indicates an increasing demand for research to improve the diagnosis and management of this challenging disease [10].

1.1.2.1 Lifestyle risk factors for OSCC

The chief risk factors for oral carcinogenesis are genetics, tobacco or alcohol use and infections [9]. Smoking is linked to an estimated 65% of oral cancer cases in the UK [3]. Betel quid, smokeless tobacco, ionising radiation and certain occupational exposures are also associated with oral cancer but are less prevalent in the UK [9]. Alcohol has been demonstrated to enhance the permeability of the oral epithelium, acts as a solvent for tobacco carcinogens, induces basal-cell proliferation, and generates free radicals and acetaldehyde, which have the capacity to cause DNA damage [11].

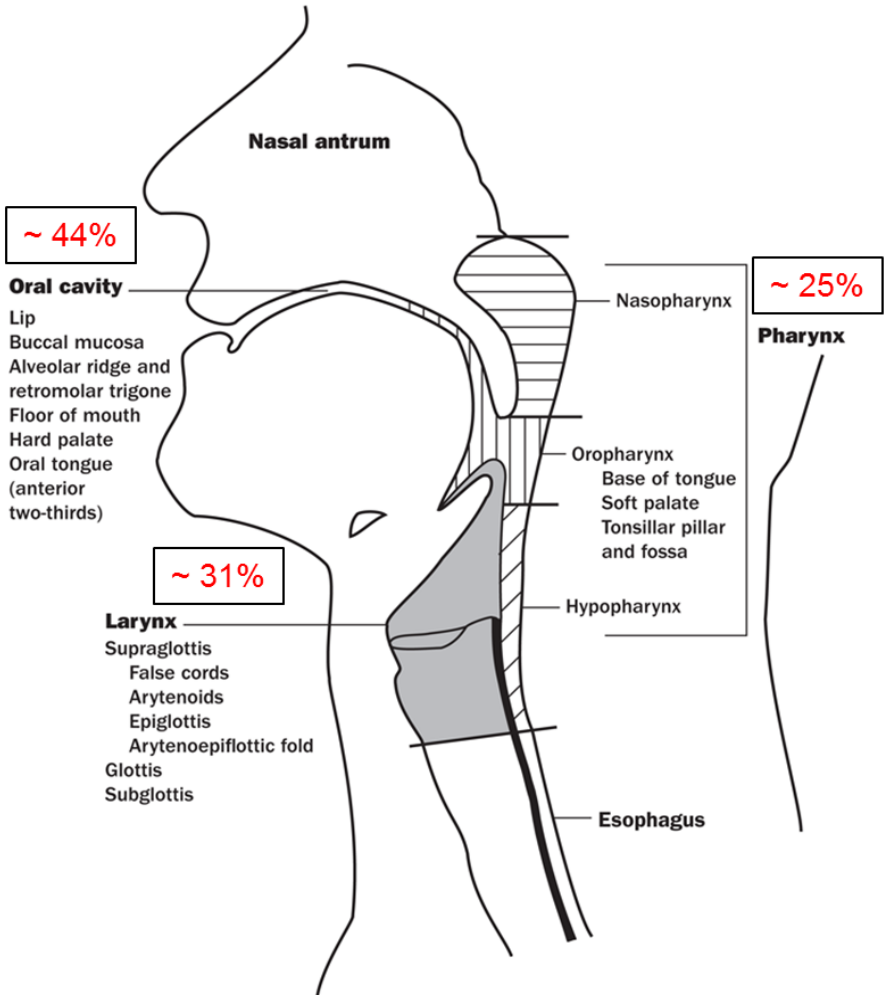

Figure 1. Proportions of HNSCC at various head and neck anatomical sites relevant to OSCC. OSCCs are found within the oral cavity (the lips, buccal mucosa, alveolar ridge, retromolar trigone, floor of mouth, hard palate and oral tongue) and oropharynx (base of tongue, soft palate, tonsillar pillar and fossa). OSCC constitutes the most common type of HNSCC.

(Adapted from Head and Neck Tumours May 01, 2014 | JA Ridge, R Mehra, MN Lango, S Feigenberg http://www.cancernetwork .com/ cancer-management/head-and-neck-tumours [Accessed September, 2015])

Alcohol-associated malnutrition and immune suppression may further promote carcinogenesis and it may act synergistically with the products of tobacco metabolism in the pathogenesis of oral squamous cell carcinoma with smoking increasing acetaldehyde exposure time and levels [12].

1.1.2.2 Human Papilloma Virus in OSCC

In 1983 *Syrjanen et al.* first reported human papilloma viruses (HPV) to be associated with squamous cell lesions at various sites of the body, including the oral cavity [13]. In females, HPV infections account for more than 50% of infection-linked cancers worldwide, although in males it is only 5% [14]. Estimates of HPV prevalence in OSCC vary according to factors that include geographic population differences, specimen type, sample preparation methods, and type of HPV detection method [15]. Overall, there is evidence that infection with high-risk human papilloma viruses (HPV), particularly HPV-16, increases the risk of OSCC especially in the oral tongue and oropharyngeal areas [16]. The impact of HPV infection may depend on site of infection. HPV expression was associated with a favorable outcome in oropharyngeal cancer (OPC) compared to oral cavity cancer (OCC) in a single center population-based study [17]. However, interpretation of significance is complex, because extracapsular spread from lymph nodes (ECS), a poor prognostic indicator in OSCC [18] (see section 1.3) and, was found to have no association on disease specific survival in OPC cases, regardless of HPV status but was found to be associated with a worse prognosis in OCC cases [19]. Some evidence suggests that HPV is associated with cancer in non-smokers and non-drinkers, however the degree of oral HPV infection may increase with tobacco and/or alcohol use with the reason for this remaining unclear [16]. Overall studies suggest that HPV positivity is associated with a better prognosis in HNSCC [20, 21].

1.1.2.3 Genetic risk factors for OSCC

Genetic conditions associated with an increased risk of developing OSCC include Li-Fraumeni Syndrome, Plummer-Vinson Syndrome, Fanconi Anemia (FA), dyskeratosis

congenita, xeroderma pigmentosum and discoid lupus erythematosus [22]. FA is a rare autosomal recessive disorder, first described in 1927, can involve the genes BRCA1, BRCA2, FANCD2, and FANCG, which are also frequently affected in sporadic OSCC [22]. This suggests that the process leading to early occurrence of oral cancer in FA patients follows a similar pathway of carcinogenesis to non-FA sporadic cancer patients [23].

The clinical behavior of OSCC varies greatly from patient to patient, site to site and sometimes even within individual sub sites [22]. Treatment approaches and outcomes may vary significantly among the different sub-sites [24].

1.2 The pathology of OSCC

Pathological stage based on the Tumour (T) Node (N) Metastasis (M) (TNM) staging method, is widely used for head and neck cancer. Key pathological factors that are correlated with poor prognosis in OSCC are summarised in Table 1, [25]; larger tumours and presence of metastasis are strongly associated with poor prognosis [26].

The presence and nature of cervical nodal metastasis has long been recognized as having strong prognostic significance for OSCC, with increased burden having a strong correlation with distant metastasis and overall survival [27]. Factors of key relevance; are metastatic status (nodal metastasis present versus absent); laterality of positive nodes, number of positive nodes; percentage of positive nodes (ratio of positive nodes: total number of assessed nodes); number of positive anatomical levels; highest positive anatomical level; size of largest metastatic deposit; presence and extent of spread outside the nodal capsule (ECS); soft tissue tumour deposits and pN stage [28]. Primary tumour histopathology, poorly differentiated tumours [29] and those displaying evidence of perineural [30] or perivascular invasion [31] are all also associated with poor prognosis predicting cervical metastasis, locoregional recurrence and distant metastasis. Prognostic prediction achieved by histopathology and clinical considerations

alone is unsatisfactory with a relatively high level of disease recurrence even following first-line treatment [8].

Table 1. Summary of key pathological characteristics correlated with poor prognosis in OSCC. (EGFR=Epidermal Growth Factor, MVD=Microvessel Density)

Factor	Parameter	Poor Prognostic Characteristic	References
Primary Tumour Gross Characteristics	*Tumour Dimensions*	*> 2cm*	*[32]*
	Tumour Thickness	*> 5mm*	*[33]*
	Total Tumour Volume	*> 6mm^3*	*[34]*
	Margin Status	*Positive*	*[34]*
	Tumour Site	*Floor of mouth, tongue, soft palate, retromolar area, alveolus*	*[35]*
Histopathology	*Malignancy Grading*	*>10 point score*	*[29]*
	Pattern of Invasion	*Small clusters of scattered, dispersed cells*	*[29]*
	Perineural Invasion	*Present*	*[30]*
	Perivascular Invasion	*Present*	*[31]*
Cervical Lymph Node Metastasis	*Number of Metastatic Lymph Nodes*	*>2 lymph node +ve*	*[27]*
	Node Location	*Outside lower neck involvement*	*[27]*
	Node Size	*>2cm*	*[27]*
	Extracapsular Spread	*Macroscopic > Microscopic*	*[18]*
Molecular Factors	*p53*	*Mutation positive*	*[36]*
	Angiogenesis Related Factors	*MVD elevated*	
	Cyclin D1	*Positive in primary*	
	EGFR and TGF	*Positive in primary*	
	Human Papilloma Virus	*Absence in primary*	*[37]*

1.3 Extracapsular Spread (ECS)

Extranodal extension termed as extracapsular spread (ECS) is defined as invasive cancer showing at least penetration of the nodal capsule or extension through the nodal capsule into the peri-nodal adipose tissue [38] (Figure 2.). ECS as a pathological entity has been extensively characterized over the last few decades. It has been found to be a

poor prognostic characteristic having significance in treatment planning for a number of different cancers; gastric [39], oesophageal [40], breast [41], cervical [42] as well as HNSCC [43-46].

The presence of ECS was originally documented in most studies as either present or absent, however, recent studies suggest that the extent of ECS, similar to growth of the primary growth, may also be a key factor for prognostication [18, 47]. The simplest and most widely adopted method of characterizing ECS (and the method adopted locally) is classification of ECS as macroscopic, tumour extension out of lymph nodes evident on dissection, or microscopic, tumour extension only evident on histological assessment. As one would intuitively expect, macroscopic ECS has been demonstrated to carry a worse prognostic outcome [18], albeit site dependent [19]. The best means for describing the extent of ECS remains controversial, with gross ECS clearly being more straightforward to recognize than microscopic ECS. The potential importance of this was highlighted by one study proposing a grading system for ECS [47], (summarized in Table 2.). They found that only ECS grade 4 (soft tissue metastasis, tumour mass without residual nodal tissue or architecture) was associated with poorer outcomes, but not independently of T-stage and other variables suggesting standardisation in ECS reporting is required. Nevertheless, the presence and degree of ECS is an established prognostic indicator for HNSCC.

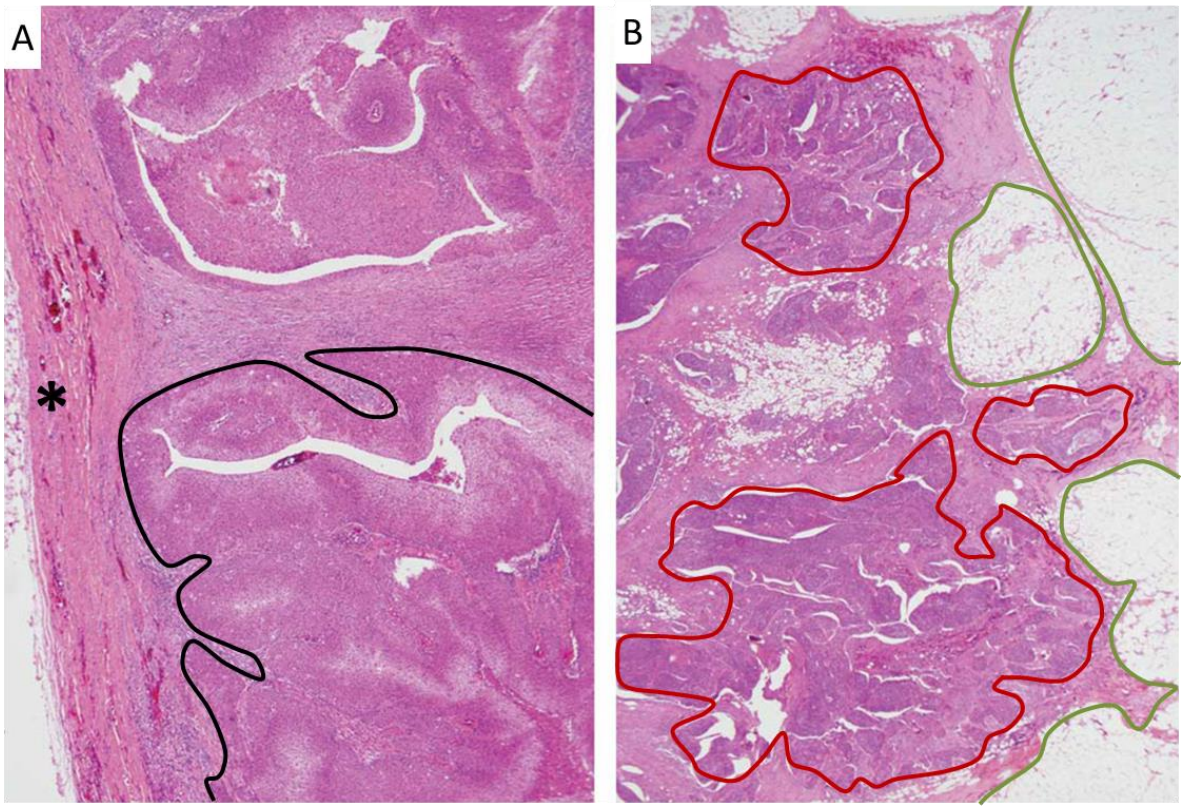

Figure 2. Representative Histopathology of ECS. Haemotoxylin and eosin stained sections of lymph nodes from a neck dissection specimen of a patient with metastatic OSCC showing **A**. a section through a lymph node harboring metastatic deposits of OSCC (one outlined in black), abutting but not breaching the nodal capsule (designated by *) and thus considered as an ECS-ve node and **B.** section through a node with fronds of metastatic OSCC (outlined in red) breaching the capsule and invading into the peri-nodal adipose tissue (outlined in green) considered as ECS+ve.

Table 2. Comparison of two different methods of ECS classification. Although Microscopic ECS in the *Woolgar et al.* classification broadly equates to grades 2 & 3 in the *Lewis et al.* classification and Macroscopic ECS broadly equates to grade 4 there may be considerable overlap in practice, with the extremes of the spectrum more easily definable. *(Compiled from Lewis et al., Mod Pathol. (2011) Nov; 24(11): 1413–1420 & Woolgar et al., Eur Arch Otorhinolaryngol (2013) 270:1581–1592)*

ECS Schematic Illustration					
***Lewis et al.* ECS Grade**	**0**	**1**	**2**	**3**	**4**
Description	Tumour confined to the substance of the lymph node (surrounded by lymphoid tissue)	Tumour reaching the capsule of the lymph node (no intervening lymphoid tissue) and with thickening of the overlying capsule	Tumour in perinodal tissue but extending ≤1mm beyond the lymph node capsule	Tumour in perinodal tissue and extending >1mm beyond the lymph node capsule	Soft tissue metastasis. Tumour mass without residual nodal tissue or architecture (no germinal centers or subcapsular sinus)
***Woolgar et al.* ECS description**		**Early MicroECS**	**MicroECS**		**MacroECS**
Description		Tumour confined to the lymph node, but desmoplasia alone extends to the immediate pericapsular area	Tumour breaching the nodal capsule / invading perinodal fat, which is only detected on histological assessment		Tumour spreading beyond node capsule evident on macroscopic assessment / dissection

1.3.1 Significance of ECS for OSCC

ECS is a reliable and significant poor prognostic indicator for OSCC [18]. The association of extracapsular extension with poorer outcomes in head and neck squamous cell carcinoma was first recognized in the 1970's [48] and numerous subsequent studies further confirmed this association [45, 49]. Studies show that ECS from regional lymph nodes is a feature of the group of patients with the worst prognosis [50]. The presence of lymph node metastasis alone, is associated with decreased survival of ~50% [51, 52]. Surprisingly, even in the light of a wealth of data supporting

this prognostic significance, ECS is not considered as part of routine TNM classification and must be documented separately on histopathological reporting [53]. The relative importance of extracapsular spread (ECS) *versus* the number of positive nodes remains controversial, with studies finding that ECS, not the number of positive nodes, was an independent predictor of poor survival [52].

A research group based in Liverpool, UK, published a cohort of 561 OSCC cases, where Cox regression modeling revealed that the single most predictive marker for death was extra-capsular spread from cervical lymph nodes ($p<0.001$) [18]. In the ECS group, overall survival at 5 years was about 24%. The study showed that ECS associated with a doubling of the incidence of local recurrence and distant metastases, but a tripling of regional relapse, even when surgical margins were generous and interestingly relapse was more frequently at the primary site. This suggests that the aggressive phenotype is expressed not only in the metastatic cells but also within the primary tumour. Overall, the study showed that ECS was associated with a doubling of the incidence of local recurrence and distant metastasis and that recurrence occurred sooner in ECS than in non-ECS cases (206 vs 334 days, $p = .04$) [18].

Early-to-moderate-stage OSCC (stages I-III) is most often treated surgically and radiotherapy [54, 55]. In advanced (stage IV) disease, multidisciplinary non-surgical approaches are being used to improve disease control, prolong survival, and maintain an acceptable quality of life for patients [56, 57]. The accurate prediction of ECS, prior to surgery, could help to better guide management. ECS is frequently present not only in patients with advanced nodal metastasis, but is also detected in a small number of patients with early-nodal disease and with a clinically negative neck, making ECS prediction more difficult [58]. Imaging techniques such as positron emission tomography (PET), computed tomography (CT) and magnetic resonance imaging (MRI) may be currently used for this purpose but are often not sufficiently sensitive for pre-treatment staging of metastasis and ECS [59].

Poor prognostic molecular factors are increasingly being defined in OSCC. For instance, a recent multivariate analysis revealed patients with TP53 mutations were 2.4 times more likely to have loco-regional recurrence evidenced by positive cervical lymph nodes being present [36]. Use of such factors could further refine the usefulness of histopathological classifications.

1.4 Aggressive molecular markers of OSCC

Molecular biology advances are providing a wealth of information on genetic abnormalities associated with cancer and often aggressive clinical characteristics can be predicted on their basis [60]. Adapting the National Cancer Institute definition of a biomarker, a biomarker of aggressive phenotype in OSCC would be a biological molecule found in blood, other body fluids, or tissues associated with OSCC carrying a poor prognosis [61]. Genetic biomarkers in OSCC could either be inherited or somatic and identified as mutations in DNA derived from OSCC tumour tissue [62]. Genomic instability is an enabling characteristic that is associated with the acquisition of hallmark capabilities of oncogenesis. Beroukhim et al., assessed copy number alterations across 3131 cancer specimens from 26 different cancer types. 158 regions identified were significantly amplified in different cancer types and these SCNAs were found to be associated with several cancers [63]. Gene rearrangements also have been shown to have the ability to trigger a subset of mutational events that result in inactivation of tumour suppressor genes [64-66]. The development of aggressive markers of disease first requires recognition of an oncogenic genotype, but later stage events are expected to drive aggressive behavior.

1.4.1 The development of an oncogenic phenotype

Current understanding of cancer indicates that a single genetic change is rarely responsible for developing malignant tumours, but sequential alterations in several genes leads to cancer [67]. This concept was first developed by Fearon and Vogelstein, who demonstrated that colorectal cancer cells specific types of genetic alterations occur in order of the development of morphological changes from normal to

benign tumour (adenoma) through to a malignant tumour or carcinoma [68]. It has been suggested that 3-10 genetic events are involved in common adult malignancies in humans and indeed mathematical modeling based on age specific tumours suggests that 3-7 independent hits are required for tumour formation [69]. The multistep process of carcinogenesis reflects a series of genetic alterations that are responsible for the transformation to a malignant cell and this concept can be used to better understand the specific genetic events responsible for the development of OSCC.

1.4.2 Late genetic events in OSCC development

Loss of heterozygosity (LOH) of chromosomal region 17p and mutation of the TP53 gene are two key genetic alterations that occur in the later stage of progression from dysplasia to invasive squamous carcinoma that may accelerate the rate of genetic alterations in oral carcinogenesis. Amplification of 11q13 has been found in 40% of cases of oral squamous dysplasia cases [70]. In general, loss of chromosomal material at 9p, 3p, and 17p is observed in relatively high proportions of dysplastic lesions whereas losses at 13q and 18q are observed more frequently in carcinomas than in dysplasia and they are also associated with later stages of carcinogenesis [71].

1.4.2.1 TP53 and aggressive OSCC phenotypes

The prevalence of mutations in HNSCC varies between 30% -70% depending on the type of tumour material, and heterogeneity of tumour sites examined. Most of the mutations are found in the conserved regions of TP53 exons 5–8 [72-76]. However, some studies have demonstrated that 20% of all TP53 gene mutations occur outside exons 5–8 in HNSCC. The pattern of mutations outside and inside exons 5–8 also differed. Within exons 5–8 region, the gene alterations were mainly missense mutations, however outside exon 5–8 region, the majority of changes were deletions, insertions, or nonsense changes [74].

TP53 has 11 exons, the first is noncoding, and tp53 consists of 393 amino acids divided into 4 main regions [77]. The protein contains a critical DNA-binding domain central region, a C-terminal domain tetramerisation region, and finally the N-terminal domain

harboring a transactivation region [78]. Disruptive TP53 mutations that lead to loss of function preferentially affect the DNA-binding domain of the protein (exons 5 to 8) [79]. Agrawal et al. found greater than 63% of TP53 mutations to be missense, with the remainder predicted to be inactivating (16% insertion or deletion [indel], 16% nonsense, 8% splice site mutations). Similarly, Stransky et al. reported 50% of TP53 mutations as missense, the remainder of which were various predicted inactivating mutations. Some studies have demonstrated an association of TP53 mutations with aggressive HNSCC behaviour. A multicentre trial analysing 420 patients using a TP53 hybridisation chip approach, analysing all known TP53 mutations, showed a decrease in survival by more than 1.5-fold in HNSCC patients in the presence of TP53 disrupting mutations [80]. The mutation status of TP53 has also been associated with poor response to cisplatin and fluorouracil in patients with HNSCC [75]. A prospective trial analysing 106 patients showed TP53 mutation status to be an independent predictor of response to fluorouracil and cisplatin, with a nonresponse risk 2.7-fold higher than patients with wild-type TP53 [75]. TP53 mutations are strongly associated with loco regional recurrence following primary radiotherapy [81] and in some studies predict loco regional failure in patients receiving radiotherapy as adjuvant therapy following surgical resection [82].

With specific relation to poor OSCC outcome, ultra-deep (>1000×) sequencing of 345 samples detected that TP53 was the most frequently mutated gene [83] with other studies reporting up to 50% of OSCC cases carrying a TP53 mutation [84]. A Dutch study revealed 14/31 OSCC samples tested harboured TP53 mutations with five missense mutations and nine resulting in a truncated protein. [85]. In a series of Taiwanese patients, TP53 mutations were observed in 32.91% of cases and were significantly correlated with an advanced form of OSCC leading to poor survival [86], whereas in a Mexican cohort of 112 OSCC cases the mutation rate was observed at 47.3% with a greater frequency of mutation observed with worsening grade [87]. These results are further supported by the frequent detection of TP53 mutation in human OSCC cell lines [88] as well as experiments using *in vitro* RNA interference. Irimie et al. reported that by inhibiting the expression of mutant p53 they observed a decreased proliferation and induction of apoptosis/autophagy in SSC-4 cells. Furthermore, they

observed decreased migration and an accompanying decrease in the expression of VEGF [89]. Evidence regarding TP53 variants in association with ECS is limited.

1.4.3 Molecular Characterization of ECS

Researchers have started to molecularly characterize ECS with a recent study profiling the DNA methylome in a cohort of OSCC cases local to Liverpool. Differential gene methylation levels distinguished OSCC tumour samples from control matched cases with putative methylation signatures for ECS and recurrence identified. The data also supported the concept of concordant methylation or CpG island methylator phenotype (CIMP) in OSCC with an association between 'CIMP-high' and worse prognosis [90]. Another study on 127 OSCC patients showed that EGFR copy number aberrations significantly associated with ECS [58]. Summarizing our present understanding a genetic progression model of multistep oral carcinogenesis is summarized in the Figure 3. Although some understanding exists already with regard to the multistep process of OSCC development, genetic mutations related to aggressive phenotype/ECS are not yet fully understood. Evidently, the observations mentioned above indicate that a better understanding of the genetic mechanisms of ECS and pretreatment molecular classification would be of great help to allow more effective, targeted and potentially personalized therapeutic approaches.

1.4.4 The role of cellular diversity

The invasive growth and metastasis ability of cancer cells also depends upon the crosstalk between cancer cells and cells of the neoplastic stroma. The stroma of the primary tumour provides circulating cancer cells arising from metastatic lesions with a favourable site for reseeding and colonization [91]**.** Traditionally, the cancer cells within tumours have been portrayed as reasonably homogeneous cell populations, which can initiate tumours and drive tumour progression as a consequence of the oncogenic and tumour suppressor mutations that define cancer as a genetic disease. In recent years, intratumour heterogeneity and the existence of cancer stem cells (CSCs) has added further complexity [92]. Many human tumours are histopathologically diverse, which

reflects clonal heterogeneity. Different regions are demarcated by various degrees of differentiation, proliferation, vascularity, inflammation and/or invasiveness. Once primary tumours have formed, the CSCs have the capability to self-renew and form the great bulk of many tumours [93, 94]. The discovery of CSCs in tumours indicates that a single and genetically homogeneous population of cells within a tumour may nevertheless be phenotypically heterogeneous due to the presence of cells in distinct states of differentiation. With the recent major advances in DNA (and RNA) sequencing technology, this is supported by in-depth sequence analysis of tumour cell genomes microdissected from different sectors of the same tumour, which has revealed a striking intratumoural genetic heterogeneity [95].

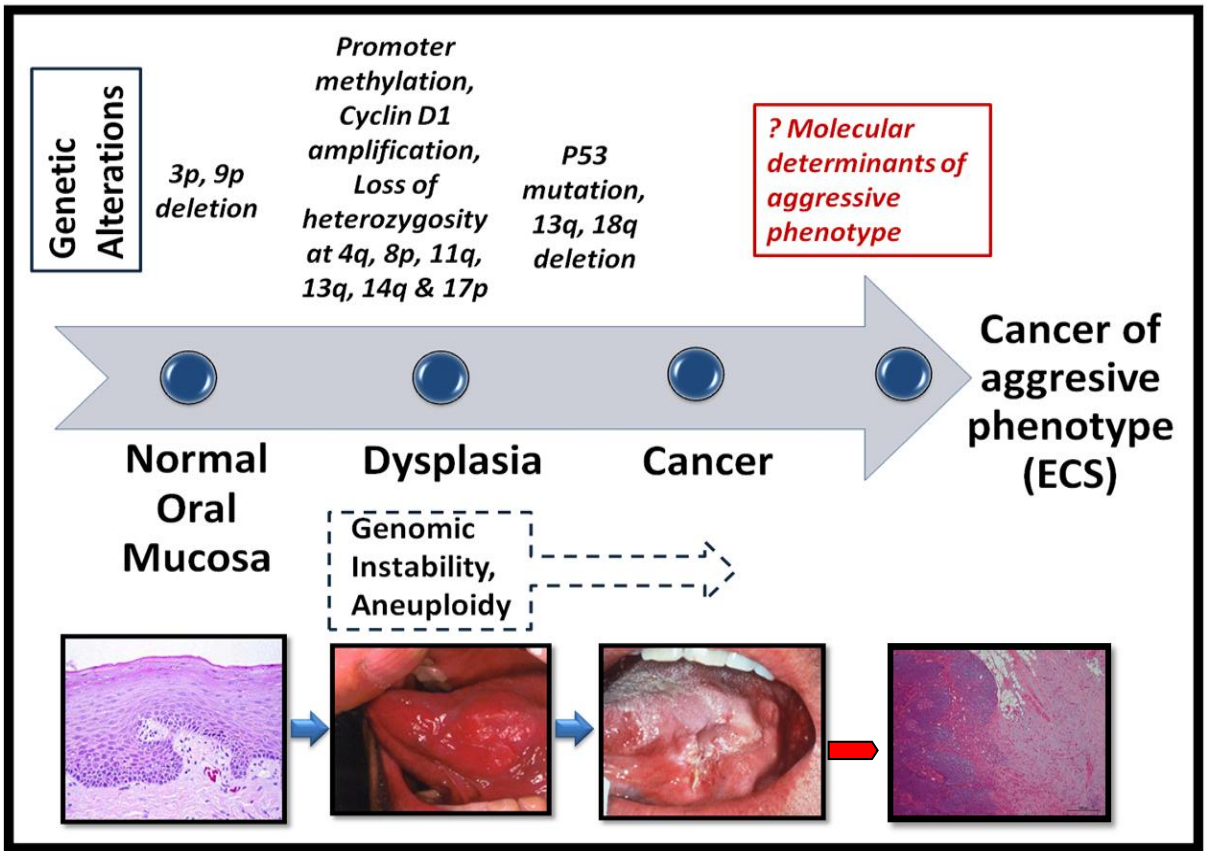

Figure 3. Genetic progression model of multistep oral carcinogenesis. Transformation of normal epithelium by multiple genetic alterations leads to dyplasia and subsequently invasive carcinoma. The key genetic alterations leading to genes influencing ECS are yet to be fully determined (highlighted by the red boxes). The final image is a haematoxylin and eosin stained cervical lymph node with the arrow indicating a capsular breach required for the identification of ECS.*(Modified from J dent Res; 2008; 87;14)*

1.5 Genomics applied to OSCC

Combining the observations from these studies with previous genomic and functional analyses of HNSCC, it is evident that HPV positive and HPV negative HNSCC have distinct mutation profiles with key differences, summarized in Table 3. Lack of TP53 mutations in HPV positive HNSCCs and the loss of CDKN2A/B in HPV negative HNSCC are the greatest differentiators [96]. There are a relatively small number of oncogenes (PIK3CA was the most frequently mutated oncogene among HPV positive patients) affected by activating mutations supporting the role of tumour suppressor pathways including p53, Rb/INK4/ARF, and Notch in disease pathogenesis and progression [96]. These pathways have roles in key processes like cellular proliferation, squamous epithelial differentiation, cell survival, and invasion/metastasis, all of which are central to the pathogenesis of HNSCC. Very few large-scale sequencing studies investigating mutational profiles of OSCC cases have been performed. One study published recently using 345 formalin-fixed paraffin-embedded (FFPE) samples obtained from pN+ OSCC patients with advanced stage (mainly stage III/IV) for sequencing of the mutational hotspots of 45 cancer-related genes, identified 1269 non-synonymous mutations in 276 OSCC samples. In their analysis-TP53, PIK3CA, CDKN2A, HRAS and BRAF were the most frequently mutated genes [83]. However, sub-group analyses examining profiles specifically for oral squamous cell carcinoma cases with or without ECS was not performed.

OSCC genetic profile variation may reflect site-specific impact of various casual agents and differences in clinical presentation. For example, tobacco smoke carcinogens, increase the prevalence and spectrum of TP53 mutation [97]. Carcinomas in patients who have never smoked harbour fewer p53 mutations, typically arise from the tongue, more commonly involve women and affect very young or old patients [98]. An important causative agent for oropharynx carcinomas is oncogenic human papillomavirus (HPV), with more than 50% harbouring integrated HPV DNA [99]. Mechanistically, the E6 and E7 viral oncoproteins bind and inactivate TP53 and retinoblastoma gene products respectively [100]. These HPV positive oropharyngeal tumours compose a distinct pathological entity with its own clinical spectrum and basaloid morphology [99],

illustrating the emerging role of genetic characterization as a potential means of determining prognosis and influencing management.

Table 3. Comparison of common genomic pathway alterations and specific genomic alterations by functional category in HPV-negative versus HPV-positive HNSCC.
(Adapted from Hammerman et al. Cancer Discov. 2015 March; 5(3): 239–244.)

Genetic Pathways and Alterations	HPV−	HPV+
RTK Amplification	>20% with ERBB family, FGFR, INS1R	Rare
RTK mutations/fusions	Rare	FGFR2/3 mutations in >10%, FGFR3-TACC3 fusions
H/K/NRAS, NF1	5–10%, HRAS may be most common	5–10%, NF1 loss may be more common
PIK3CA amplification/mutation	Common ~30%	Very common >50%
TP53	Genomic loss in nearly all cases	HPV-driven loss
Cell cycle deregulation	Loss of CDKN2A/RB1 by multiple mechanisms in nearly all cases, CCND1/CDK4/CDK6 amplification common (30%)	HPV-driven CDKN2A loss, E2F1 amplification (20%)
Oxidative Stress Regulation	Common activation of NRF2/KEAP1/CUL3 (25%)	Rare
Differentiation	Common loss of NOTCH1/FAT1/AJUBA, TP63 gain	NOTCH1 loss less common, TP63 gain more common
Immune evasion	Uncommon HLA mutations, <10%	HLA, B2M mutations & TRAF3 loss

HPV-negative HNSCCs are molecularly diverse, as suggested by expression profiling that demonstrates diverse biologic subclasses within HPV-negative disease. These studies demonstrate “atypical” HNSCCs constituting 20% of HPV-negative cases and the vast majority of HPV-positive HNSCCs and demonstrate lack of EGFR amplification

and/or overexpression [101, 102]. An intriguing mutational pattern directly relevant to this study was identified by The Cancer Genome Atlas Cancer Network in a subset of HPV-negative OSCCs having few to no copy number alterations and statistically significant enrichment for HRAS, CASP8 and PIK3CA mutations and lack of TP53 mutation [21].

1.5.1 The impact of genomics on treatment strategies

To date, treatment approaches for HNSCC have been dictated by the anatomic site of the primary tumour with oral cavity cancers treated primarily with surgical resection and pharyngeal and laryngeal tumours with chemoradiation [5]. Although targeted drugs like epidermal growth factor receptor (EGFR)-specific antibody, cetuximab, combined with radiotherapy are in use, the survival rate has not markedly improved in recent decades because of frequent locoregional recurrences, distant metastases and secondary tumours. Cetuximab remains the only targeted agent approved for HNSCC in present clinical use [96]. The development of targeted therapeutics for HNSCCs has relatively slow, due in part to the lack of therapeutic biomarkers and incomplete understanding of their genomic landscape. Furthermore, the wide-scale sequencing studies outlined previously have revealed predominant changes in tumour suppressor genes without numerous oncogene targets.

If other potential targets for HNSCC are considered, FGFRs, the PI3K/AKT Pathway, Cyclin Dependent Kinases and STAT3 are candidates. FGFR1 amplifications are found in HPV-negative patients at a rate of approximately 10% and appear to be enriched in non-oropharynx tumours. FGFR2 and FGFR3 mutations and FGFR3-TACC3 fusions are of particular interest because they are associated with significant responses to FGFR TKIs in pre-clinical models and in early phase clinical settings, including a case report of a major response to pazopanib in a patient with a FGFR2 mutated tongue cancer [103]. PIK3CA is commonly amplified and/or mutated in patients with HNSCCs (37% of cases in TCGA), and PIK3CA alterations are enriched in HPV-positive patients. More than 50% of patients with HNSCC have activating somatic alterations in the PI3K/AKT/mTOR pathway, stimulating initiatives to develop targeted, small molecule

inhibitors. Up to 32% of HNSCCs display an amplification or mutation of CCDN1, CDK4 or CDK6 with the majority of these alternations found in HPV-negative patients [21]. CDK4/6 inhibitors may therefore be effective in-patient cohorts with high rates of CCND1 amplification as evidenced by ongoing clinical trials in other cancers and these agents are well-tolerated as both single agents and in combination with other therapies [96]. Finally, HNSCCs frequently display hyperactivation of STAT3 via a variety of mechanisms and STAT3 pathway inhibitors are currently being explored in both pre-clinical models and in early phase clinical trials.

In summary, the limited information about molecular carcinogenesis of HNSCC as well as the genetic and biological heterogeneity of the disease has hampered the development of new therapeutic strategies. Further understanding of the specific molecular carcinogenesis of OSCC, harboring some distinct genomic characteristics when compared to other HNSCCs, will hopefully will lead to novel therapies having improved outcomes for every individual patient [104], especially those with the aggressive phenotype of ECS.

1.6 Summary and study objective

Recent significant broad next generation sequencing studies as well as numerous more targeted specific genetic studies suggest that OSCC requiring mutations in a number of key genes to develop is no different to other cancers. However, there appear to be differences in the spectrum of genes that are involved. Although the key molecular determinants of OSCC are becoming more defined our understanding of the molecular drivers of aggressive disease, particularly as evidenced by ECS, remains limited.

1.7 Overall hypothesis

Significant genetic changes in key genes or gene families drive extracapsular spread of OSCC.

1.8 Overall Aim

The overall aim of this study is to elucidate the molecular determinants of OSCC associated with ECS.

Chapter 2 : Materials and Methods

2.1 Reagents

All molecular biology kits and chemicals used in this study are listed along with the manufacturer and catalogue number in Appendix A.

2.2 Clinical series

OSCC HPV-ve primary tumour samples (n=50) ranging from T1 to T4 stage, with no nodal involvement, positive nodes or positive nodes plus ECS were selected from Liverpool head and neck tissue bank. The clinical features of the OSCC samples used in this study are listed in the table 4.

2.3 ECACC Human Random Control (HRC) DNA

Control DNA used in this study was obtained from Human Random Control Panel 1 (HRC1) from Sigma-Aldrich. The purified HRC DNA consists of DNA from 96 random individuals with a concentration of 100 ng/µl where the DNA is dissolved in 10mM Tris (pH8) and 1mM EDTA.

2.4 Macromolecule purification from tissues

DNA, RNA and protein were all rapidly isolated from pre-weighed, fresh frozen tissues, each sample for all three macromolecules using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, catalogue no: 80004). Purification was performed according to the manufacturer's protocol.

Table 4. OSCC case series clinical data. (TNM staging, tumour site and survival data) FOM = Floor of Mouth

Patient Number	Pathological Stage	T stage	ECS	Nodal status	Site	Death	Date of death
3483	pT2N1M0	2	No	pN+	ant pillar fauces	no	N/A
3402	pT4N0M0	4	No	pN0	FOM	no	N/A
3395	pT2N0Mx	2	No	pN0	FOM	no	N/A

3385	pT4N0	4	No	pN0	retromolar	yes	18/10/2009
3233	pT2N2cM0	2	ECS	pN+	FOM	yes	08/04/2008
3465	pT2N2bM0	2	No	pN+	tongue	no	N/A
3455	pT3N1M0	3	ECS	pN+	tongue	yes	29/09/2009
3499	pT4N2cM0	4	ECS	pN+	lower alveolus	yes	01/04/2010
3495	pT2N1Mx	2	No	pN+	tongue	no	N/A
3481	pT2N1Mx	2	No	pN+	soft palate	no	N/A
3484	pT3N2bM0	3	ECS	pN+	tongue	yes	02/12/2009
3270	pT2N2bM0	2	ECS	pN+	retromolar	yes	08/06/2009
3286	pT2N2b	2	ECS	pN+	FOM	yes	21/07/2008
3230	pT2N1Mx	2	ECS	pN+	tongue	yes	06/07/2006
3241	pT2N2bMx	2	ECS	pN+	FOM	no	N/A
3248	pT2N2bM0	2	ECS	pN+	FOM	no	N/A
3167	T2N2Mx	2	ECS	pN+	FOM	yes	24/10/2005
3366	pT2N2bMx	2	ECS	pN+	tongue	yes	12/10/2008
3345	pT2N2cMx	2	ECS	pN+	tongue	yes	31/12/2007
3240	pT3N0M0	3	No	pN0	max sinus	yes	16/12/2006
3313	pT4aN0Mx	4a	No	pN0	FOM	no	N/A
3261	pT4aN0M0	4a	No	pN0	mandible	no	N/A
3238	pT4N0M0	4	No	pN0	mandible	no	N/A
3318	pT3N0M0	3	No	pN0	tongue	no	N/A
3291	pT2N0M0	2	No	pN0	base of tongue	no	N/A
3258	pT2N0M0	2	No	pN0	FOM	yes	29/06/2007
3383	pT1N0Mx	1	No	pN0	supraglottis	no	N/A
3251	pT4N1M0	4	ECS	pN+	tongue	yes	21/07/2006
3272	pT1N0Mx	1	No	pN0	FOM	no	N/A
3276	pT2N0Mx	2	No	pN0	FOM	no	N/A
3288	pT1N0Mx	1	No	pN0	tongue	no	N/A

Patient Number	Pathological Stage	T stage	ECS	Nodal status	Site	Death	Date of death
3315	pT2N0Mx	2	No	pN0	FOM	no	N/A
3236	pT2N0Mx	2	No	pN0	mandible	no	N/A
3242	pT4aN0Mx	4a	No	pN0	lower alveolus	yes	24/12/2007
3340	pT2N0M0	2	No	pN0	tongue	no	N/A
3274	pT2N1Mx	2	No	pN+	tongue	No	N/A
3254	pT2N2cMx	2	No	pN+	FOM	Yes	22/10/2007
3277	pT2N2bM0	2	No	pN+	ant pillar fauces	No	N/A
3289	pT2N1Mx	2	ECS	pN+	FOM	Yes	15/01/2010
3237	pT4aN2aMx	4a	ECS	pN+	mandible	Yes	19/10/2007
3444	pT2N2cMx	2	ECS	pN+	tongue	No	N/A
3300	pT1N1M0	1	No	pN+	tongue	No	N/A
3302	pT2N2bM0	2	No	pN+	FOM	No	N/A
3533	pT2N2bM0	2	ECS	pN+	retromolar	No	N/A
3301	pT4N0Mx	4	No	pN0	maxillary alveolus	No	N/A
3463	pT2N1Mx	2	ECS	pN+	tongue/FOM	No	N/A
3354	pT2N1Mx	2	No	pN+	tongue	No	N/A
3461	pT4N2cM0	4	ECS	pN+	FOM	No	N/A
3464	pT2N2bMx	2	ECS	pN+	tongue	No	N/A
3480	pT3N2bM0	3	ECS	pN+	tongue	No	N/A

2.5 DNA analysis

2.5.1 SNP Genotyping assays

Quantitative PCR (q-PCR) using minor groove binders (MGB) Probes (TaqMan® SNP Genotyping Assays ABI, USA) were used for Quantitative PCR (Q-PCR/q-PCR) to make allele calls [105]. SNPs from the National Centre for Biotechnology Information

(NCBI) dbSNP database were selected on the basis of high level of heterozygosity in the European population and are listed in Table 5.

Table 5. SNP Genotyping assays with their SNP-ID.

Gene	SNP ID	Level of heterozygocity in European population	Assay number
TP53	RS1614984	40%	C_8727721_10
TP53	RS1642785	30%	C_2880090_10
TP53	RS2909430	22%	C_26603915_10
TP53	RS12951053	16%	C_2880088_10
TP53	RS1042522	30%	C_2403545_10
Cyclin D1	RS3212891	43%	C_25800715_10
Cyclin D1	RS649392	43%	C_744724_10
Cyclin D1	RS678653	43%	C_744721_10

Reactions were performed in an iCycler thermal cycler (Bio-Rad Laboratories) and analysed using the iCycler iQ real time PCR detection system and software (version3.0). Samples were diluted to 10ng/µl and 2µl used with 0.5µL probe mix + 4µl RT-PCR mix (Taqman assay universal PCR mastermix Amplitaq (R) Gold DNA polymerase (Biorad, catalogue number 4304437) + 3.5µL DNase-free, sterile-filtered water. Thermal cycling was performed at 95 ^{0}C for 3 min for heat activation of the enzyme, followed by 50 cycles of 92 ^{0}C for 15 sec for denaturation and 60 ^{0}C for 1 min 30 sec for annealing /extension.

The ΔΔCt method (input allele1 amount / allele2 amount against those for output allele1 amount / allele2 amount) was used for allele calls.

2.5.2 Size estimation of DNA

Agarose Gel Electrophoresis:

1.2% or 0.8% agarose gels according to the expected size of the DNA to be analysed were made by combining 1.2g or 0.8g agarose (Fisher scientific, Catalogue number

BMA 50002) and 100 ml of 1X TAE buffer. The gel was stained with SYBRsafe (Invitrogen, Catalogue no. S33102) at a final concentration of 500ng/ml. Samples were loaded with 1 volume of loading dye (containing 3 marker dyes, bromophenol blue, xylene cyanol, and orange G) and 5 volumes of each sample and electrophoresis performed at 100 volts for 30-60 minutes compared to size markers comprising1kb (NEB-catalogue no: N3232S) and 100 base pair (bp) (NEB-Catalogue no: N3231S) ladders. Gels were imaged using Typhoon 9400 (Amersham Biosciences) and analysed by Image Quant (Molecular Dynamics).

Agilent Bioanalyser Analysis:

A DNA 1000 kit (Agilent, catalogue no:50671504) was used for size estimation of DNA according to the manufacturer's instructions. A DNA chip was placed on the chip priming station where 9.0 µl of the gel-dye mix were added at the bottom of the well-marked chip and primed the chip with gel-dye mix. 5 µl of DNA marker was added into each well and then 1 µl of DNA ladder and samples loaded into the wells. 1 µl of deionized water was added to unused wells. The chip was vortexed for 60 seconds at 2400 rpm and then the chip was inserted into the Agilent 2100 Bioanalyzer. Agilent software was then used for fragment size and library quantification visualized as an electrophoregram.

2.6 OSCC cell culture

2.6.1 Primary OSCC cell culture

Primary OSCC cells were cultured using 10ml medium [90% Dulbecco's Modified Eagle's Medium (Sigma Aldrich, Catalogue no. D5796) + 10% Fetal bovine serum (FBS) [pre-warmed to 37°C] per 75cm² flask. The cells were maintained at 37°C with 5% CO2. The medium was changed twice a week.

When sufficient growth occurred (~ 80% confluent), then the cells were released from plates by trypsinisation (TrypLE Express,GIBCO, Catalogue no 12406-013). Trypsinised cells were harvested by centrifugation at 1200 rpm for 5 minutes. After removal of the

supernatant, cells were washed with 6 ml of DMEM medium was added for replating and divided between two 75cm² flask. A further 15 ml each of DMEM medium was added.

For storage, cell freezing medium (Sigma- catalogue no. C6295) was added, the cells were snap frozen in liquid nitrogen and then and kept at -80°C.

2.6.2 Secondary OSCC cell culture

Secondary cell lines used for this study were: BHY from DSMZ (Braunschweig, Germany, Catalogue no: ACC 404), H357 from UK, (Catalogue no: 06092004). These cells were cultured in 25cm² flasks using the same medium and conditions as primary cell lines.

2.6.3 Cell line authentication

Mycoplasma testing and single tandem repeat (STR) profiling were carried out every month to ensure absence of cross-contamination in all cell lines used in the study by another fellow lab member who was responsible for maintaining the cell lines. STR profiling was performed using the GenePrint 10 system (Promega Catalogue no. B9510) and 10 ng of DNA from each cell line used for amplification. After mixing the amplified product with internal lane standard 600 allele ladder, analysis was performed by 3130 Genetic Analyzer (Life Technologies) using GeneMapper ID-X software. Mycoplasma testing was carried out using e-MycoTM plus Mycoplasma PCR Detection Kit (Intron Biotechnology Catalogue no. 17341).

2.6.4 Gamma secretase inhibition in OSCC cell lines

DAPT (N-[N-(3,5-difluorophenacetyl)- L-alanyl]-S-phenylglycine t-butyl ester) (Sigma-Aldrich- Catalogue no: D5942) was used as a gamma secretase inhibitor to investigate the effect of the Notch pathway inhibition in different OSCC cell lines especially in relation to their invasive properties. 2.7 mg DAPT was dissolved in 250 µl DMSO dimethyl sulfoxide (DMSO, Sigma Catalogue no. D2650) and then added to 50 ml of media to give a 125 µM Stock solution. Half of the 25cm² flasks were then treated with 1ml of the DAPT media with 4 ml DMEM normal media to make a final concentration of 25 µM DAPT and half of the 25cm² flasks were treated with 5 ml standard DMEM media.

2.7 DNA Sequencing analysis

2.7.1 Primer design

Primers were designed Primer3 Input (version 0.4.0), and were purchased from Eurogentec (Southampton, UK). The primer sequences and the sizes of the amplified products are shown in the Table 6, to 8.

Table 6. Forward (F) and reverse (R) primer sequences for long range PCR to cover all the exon regions for TP53 and NOTCH1 genes.

Gene	Exon	Sequence	PCR product length
TP53	TP53 _EX_1F	GTTTCCATGTACTGAAAGCAA	690 base pair
	TP53 _EX_1R	AAATACACGGAGCCGAGA	
	TP53_EX_2F	CTTCTCTGCAGGCCCAGGT	4290 base pair
	TP53 _EX_9R	TGTCTTTGAGGCATCACTGC	
	TP53 _EX_10F	CTCCCCCTCCTCTGTTGCTG	3500 base pair
	TP53 _EX11_R	GGAGGCTGTCAGTGGGGAAC	
NOTCH1	NOTCH1_EX1_F	CCAGCCGGGGAAGAGAGG	2002 base pair
	NOTCH1_EX2_R	CCCTCGACAAAGCAACAGGT	
	NOTCH1_EX3_F	ACTAACTGCCCTGGCACATC	1330 base pair

	NOTCH1_EX4_R	ACACGCAGTCTGGGGAACT	
	NOTCH1_EX5_F	AGCTGCCGGAACCTCTTAAA	697 base pair
	NOTCH1_EX5_R	CCTCAATGCCAGGACACC	
	NOTCH1_EX6_F	TTAGGGGACAGGGAGCTCAG	3640 base pair
	NOTCH1_EX12_R	CCTCTGAGCACAGTGCAGTC	
	NOTCH1_EX13_F	GCCCTTAGGGAGCATGTGTA	4980 base pair
	NOTCH1_EX18_R	GCACACCTCTCTGTGCAGTC	
	NOTCH1_EX19_F	CCCTAGGGTTGAGCAGAAGG	3908 base pair
	NOTCH1_EX25_R	CAAGTTCAGGTCCTCCCTCA	
	NOTCH1_EX26_F	AGGCCAGCATGCAGTTCTAA	3550 base pair
	NOTCH1_EX30_R	TAGGGGAGAGAGGCAGGTG	
	NOTCH1_EX31_F	CAGACTGAGCACCCGTCTCT	4850 base pair
	NOTCH1_EX34_R	GAAGGCTTGGGAAAGGAAGC	

Table 7. Forward (F) and reverse (R) primer sequences for re-targeting regions in NOTCH1 gene.

Gene	Exon	Sequence	PCR product length
NOTCH1	EX1_NTCH1_F	CTGAGCCTCACTAGTGCCTC	388 base pair
	EX1_NTCH1_R	CCAAAGTTTCCAAAGGGCGC	
	EX2_NTCH1_F	CGGGTGAGACTGACCTCTCT	220 base pair
	EX2_NTCH1_R	GGCCATCTCCAGAAGACAAT	
	EX3_NTCH1_F	AGCTGTGGCGGGGCCTTC	244 base pair
	EX3_NTCH1_R	GGCAGCGGCACTTGTACT	
	EX4A_NTCH1_F	ACCCTCTTGTCCCCTTGTCT	222 base pair
	EX4A_NTCH1_R	GTAGGAGCCGACCTCGTTGT	
	EX4B_NTCH1_F	CACTGCCCACCCAGCTTC	238 base pair
	EX4B_NTCH1_R	GCACACTCGTGGGTGACG	
	EX26A_NTCH1_F	AGCCCCCTGTACGACCAGTA	240 base pair
	EX26A_NTCH1_R	CACGCTTGAAGACCACGTT	
	EX26B_NTCH1_F	GTGGTGGTGGTGCTGATG	238 base pair
	EX26B_NTCH1_R	GTCCATGGGGTCCAGCTC	
	EX27_NTCH1_F	CCAGCCCCTCTCTGATTGT	215 base pair
	EX27_NTCH1_R	CTGCAGGCAGAGCCTGTT	

EX28_NTCH1_F	CGCAGGTGAGACCGTGGA	260 base pair
EX28_NTCH1_R	ACCGGGGACCCAGAAGCAG	
EX29_NTCH1_F	CTCACTTCTCGACCCCTCAC	194 base pair
EX29_NTCH1_R	AGGGAGTGAGCAGAGCCTGT	
EX30_NTCH1_F	GGCTCTGCTCACTCCCTCTA	206 base pair
EX30_NTCH1_R	GCTGCTGGCACCCTTACC	

Table 8A. Primer sequences for Sanger affirmation of TP53 variants.

Oligo name	Chemistry	Mod5'	Mod3'	Sequence	Bases
TP53_ EX4F	DNA	None	None	GACCTGGTCCTCTGACTGCT	20
TP53 _EX4R	DNA	None	None	ATGAGACTTCAATGCCTGGC	20
TP53_EX5F	DNA	None	None	TTCCAGTTGCTTTATCTGTTCA	22
TP53_EX6R	DNA	None	None	TCAGTGGCCCTCCGGGTGAG	20
TP53 _EX7F	DNA	None	None	CCTCATCTTGGGCCTGTGTT	20
TP53_EX7R	DNA	None	None	TCTTTTCTCTGGCTTTGGGA	20
TP53 _EX8F	DNA	None	None	CACCTTTCCTTGCCTCTTTCC	21
TP53_EX8R	DNA	None	None	AGCTACAACCAGGAGCCATT	20

Table 8B. Primer sequences for Sanger affirmation of NOTCH1 gene variants.

Oligo name	Chemistry	Mod5'	Mod3'	Sequence	Bases
NTCH1_EX5F	DNA	None	None	ACCTCAGGGAAGAGGCTGAC	20
NTCH1_EX5R	DNA	None	None	TGCCTGTGAGTGCAGTTTAG	20
NTCH1_EX8F	DNA	None	None	GTGGTGTGCAGTGAGGTGTT	20
NTCH1_EX8R	DNA	None	None	AGCCTCGACTCGGTTTCC	18
NTCH1 _EX14F	DNA	None	None	CCATAGGGCATTGCAGACC	19
NTCH1 _EX14R	DNA	None	None	GTCTCCAGCTCCCCAGACTC	20
NTCH1_EX18F	DNA	None	None	ACTCACCCTTCCGTCCTCTC	20
NTCH1_EX18R	DNA	None	None	CCTTCCACGGCCTCACTC	18

NTCH1_EX20F	DNA	None	None	ACCCAGCTGACCCCAATCT	19
NTCH1_EX20R	DNA	None	None	GCCCACAGAGGACCTTGAT	19
NTCH1_EX21F	DNA	None	None	CACCTGTCTACCACCCCATC	20
NTCH1_EX21R	DNA	None	None	GGCCACAACCCTTACCCTAG	20
NTCH1_EX23F	DNA	None	None	GGCCACTGACGAAACCTG	18
NTCH1_EX23R	DNA	None	None	GAGAGGCTCACCCTGCTG	18
NTCH1_EX25F	DNA	None	None	ACCGTCCTGTCTTCCCTCTC	20
NTCH1 _EX25R	DNA	None	None	CAAGTTCAGGTCCTCCCTCA	20
NTCH1_EX31F	DNA	None	None	CACCCGTCTCTGCCTCTG	18
NTCH1_EX31R	DNA	None	None	CGTAAGCCTGGCCACTGC	18
NTCH1_EX32F	DNA	None	None	CCTGAGCCTCTCCCTGTTG	19
NTCH1_EX32R	DNA	None	None	AACCCAGTCCCACCCGTC	18
NTCH1_EX33F	DNA	None	None	TCTCCAACCTACCCCATCTG	20
NTCH1_EX33R	DNA	None	None	TTTGGCCCTCACTTCTCTGT	20
NTCH1_EX34AF	DNA	None	None	GCTTCCTCTGGTGATGGAAC	20
NTCH1_EX34AR	DNA	None	None	GTGGGCCAGTCTCAAAGG	18
NTCH1_EX34BF	DNA	None	None	AACCACCTGCCTGGGATG	18
NTCH1_EX34BR	DNA	None	None	GCGCCGTTTACTTGAAGG	18

2.7.2 Polymerase chain reaction (PCR)

Water was used as a negative control and ECACC control panel DNA as a positive control.

2.7.2.1 Long range PCR

Long range PCRs were performed at 93°C for 3minutes for enzyme activation, followed by 35 cycles of 93°C for 15 seconds for denaturation, 62°C for 30 seconds for annealing, and 68°C for 1 minute/kb for extension using the Qiagen long range PCR protocol with 20 ng of DNA (<=2000 base pair products) or 50 ng (for >=2000 base pair

length products) in 50 µl reaction volumes. Each reaction contained 5 µl of 10x Long Range PCR buffer (Qiagen, catalogue number 206402), 2.5 µl of dNTP mix (final concentration: 500µM of each), 2 µl of each forward and reverse primer (from 10 µM primer stock concentration) and 2 units Long Range PCR enzyme mix (Qiagen, catalogue number 206402).

2.7.2.2 Short range PCR

Short range PCR (~50-1000 base pair targets), were performed at 95 °C / 5 mins for enzyme activation, followed by 35 cycles of 94 °C denaturation for 30 sec, 63°C annealing for 30sec, and 72°C extension for 1 min followed by 72°C for 10 mins in 20 ul reactions containing 20 ng of target DNA, 2 µl of 10x PCR buffer (Qiagen), 2 µl of dNTP mix (2 mM), 2 µl of each primer (forward and reverse primer from 10 µM concentration), 0.8 µl of Mgcl2 from 25m M concentration and 0.2 µl Taq DNA Polymerase (5 units/µl, Qiagen, catalogue no. 201203). Tumour DNA, along with random control DNA (positive control) was used for PCR amplification.

2.7.3 Purification of PCR products

QIAquick PCR Purification Kit was used (Qaigen, catalogue no: 28104) for purification of PCR products according to the manufacturer's protocol.

2.7.4 Library preparation for next generation sequencing

2.7.4.1 DNA Shearing

DNA fragments, mean target fragment size and range of 165 bp and 150 to 180 bp, respectively were achieved by sonicating input DNA by using a Covaris S2 System (Applied Biosystems). 1 µg of DNA diluted in 1X Low TE Buffer were used in a Covaris

micro tube was in a total volume of 100 µL for Cycles at 5 °C, with an upper limit of 30 °C, in Frequency sweeping mode with Water Quality Testing Function, Off and a Duty Cycle of 10%, Intensity 5, Cycles/burst 100, and Time 60 seconds.

2.7.4.2 Size estimation of sheared DNA by bioanalyzer

A DNA 1000 kit (Agilent, catalogue no:50671504) was used for size estimation of sheared DNA in an Agilent Bioanalyser according to the manufacturer's instructions as described in section 2.5.2.

2.7.4.3 DNA end repair

The ends of sheared DNA (1 µg in a total volume of 100 µl) were polished and 5' phosphorylated using enzymes supplied with the kit. (Applied Biosystems, Catalogue no. 4443713). 4 µl of End Polishing Enzyme 1 (10U/ µl) and 16 µl of End Polishing Enzyme 2 (5U / µl), 20 ul 5x end polishing buffer and 8 µl of 10m M dNTP mix in a total volume of 200 µl at room temperature for 30 minutes.

2.7.4.4 Adaptor ligation

5 base unique bar-coded adaptors were designed according to the sequences from Ion Xpress™ Plus gDNA and Amplicon Library Preparation user guide (Publication Part Number 4471989 Rev. C) and purchased from Eurogentec (Southampton, UK). The adaptor sequences are tabulated in table 9A and B. Ligation of adaptors to purified sheared DNA was performed using 10 µl of 10X Ligase Buffer, 1.5 µl of DNA Ligase enzyme, 22.2 µl of adaptors from 20 µM stock concentration (20 pmolar excess over the amount of sheared DNA ends) in a total volume of 100 µl at room temperature overnight.

Table 9. Forward and Reverse P1 and A adaptor sequences. [A] The table summarizes the forward and reverse P1 adaptor sequences. [**B]** The forward and reverse sequences of barcoded A adaptors used in this study.

9A.

Oligo name	Sequence
P1F	CCA-CTA-CGC-CTC-CGC-TTT-CCT-CTC-TAT-G)GG-CAG-TCG-GTG-AT
P1R	ATC-ACC-GAC-TGC-CCA-TAG-AGA-GGA-AAG-CGG-AGG-CGT-AGT-GGT-T

9B.

Oligo name	Chemistry	Mod5'	Mod3'	Sequence	Bases	Purification	%GC
A_ITbc_8F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGA-CTT-GA	38	HPLC-RP	52.6
A_ITbc_8R	DNA	N/A	N/A	TCA-AGT-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	50
A_ITbc_9F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGG-ATC-AG	38	HPLC-RP	55.3
A_ITbc_9R	DNA	N/A	N/A	CTG-ATC-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	52.5
A_ITbc_11F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGG-GCT-AC	38	HPLC-RP	57.9
A_ITbc_11R	DNA	N/A	N/A	GTA-GCC-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	55
A_ITbc_12F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGC-TTG-TA	38	HPLC-RP	52.6
A_ITbc_12R	DNA	N/A	N/A	TAC-AAG-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	50
A_ITbc_13F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGG-CAC-TA	38	HPLC-RP	55.3
A_ITbc_13R	DNA	N/A	N/A	TAG-TGC-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	52.5
A_ITbc_16F	DNA	N/A	N/A	TTC-CAT-CTC-ATC-CCT-GCG-TGT-CTC-CGA-CTC-AGA-CCA-GT	38	HPLC-RP	55.3
A_ITbc_16R	DNA	N/A	N/A	ACT-GGT-CTG-AGT-CGG-AGA-CAC-GCA-GGG-ATG-AGA-TGG-AAT-T	40	HPLC-RP	52.5

2.7.4.5 Purification of size fractionated DNA

The adaptor ligated DNA was electrophoresed in an E-Gel (2% agarose, Invitrogen catalogue no GG01002) for size selection and DNA in the appropriate size range collected using a Gilson pipette.

2.7.4.6 Nick translation

Nick translation was performed using Hot star Taq plus DNA polymerase enzyme (Qiagen, catalogue no. 203605) using 72°C for 30 minutes according to the Short Range PCR reaction conditions 2.7.2.2.

2.7.4.7 Emulsion PCR

5 µl of nick translated diluted library (40pM final concentration) was added to 200 µl of OneTouch Reagent Mix and 100 µl of OneTouch Enzyme Mix and 595 µl nuclease free water in a 1.5-ml LoBind Tube, to make an amplification solution of 900 µl at room temperature. 100 µl Ion sphere particles were added and the solution vortexed at maximum speed for 5 seconds before microcentrifugation for 2 seconds. The Ion OneTouch Template (Life Technologies, Catalogue no. 4468660*)* containing template preparation reagents was used for emulsion PCR according to the manufacturer's instructions.

2.7.4.8 Bead capture

After ePCR, templated beads carrying adaptored library fragment were enriched by annealing to the P2 adaptor using the bead Enricher (LifeTechnologies) according to the manufacturer's instructions.

2.7.5 Next Generation Sequence analysis

2.7.5.1 SOLiD sequencing (set up and parameters)

The quality and quantity of the sequencing beads was determined by a 6 hour work flow analysis (WFA) run with a small fraction of the templated beads before actual sequencing. 600-700 million optimal beads were sequenced on a full SOLiD 4 slide in the order reverse 35 bp read and forward 50 bp read for 12 days on a SOLiD sequencer (ABI), according to the SOLiD 4 user guide.

2.7.5.2 Ion Torrent^{TM} sequencing (setup and parameters)

The Ion PGM 200 Sequencing Kit (Life Technologies, Catalogue no. 447400) was used for the Ion torrent sequencer. Templated beads were centrifuged for 2 minutes at 15,500 × g after transferring to a new 0.2-mL PCR tube and adding 100 µl of annealing buffer. Supernatant was removed to leave leaving slightly less than 15 µl. 12 µl of sequencing primer was added and the pellet disrupted by pipetting. The sequencing primer was annealed at 95ºC for 2 minutes and then 37ºC for 2 minutes. 1 µl of Sequencing Polymerase enzyme was added and the reaction incubated at room temperature for 5 minutes. Enriched beads were loaded onto the Ion 316 Chip according to manufacturer's instruction (Ion Sequencing Kit Quick Reference v2.0, Publication Part Number 4469867 Rev.C).

2.7.6 Sanger sequencing

2.7.6.1 Sample preparation for Sanger sequencing

Dye terminator sequencing reactions were performed using DYEnamic ET dye terminator kit (GE healthcare). Forward and reverse primers (Table 8) were used separately by setting up two different reactions for each sample for forward and reverse sequencing. In each reaction- 4 µl of PCR amplified DNA, 1.5 µl of water, 0.5 µl primer (5 µM) and 4 µl DYEnamic ET reagent mix were added. The sample was incubated in a

thermal cycler for 25 cycles and the program was 95°C for 20 seconds, 50°C for 15 seconds, 60°C for 60 seconds. AutoSeq96 plates (Amersham Pharmacia Biotech) were used for removal of unincorporated dye labelled terminators and sample preparation for Sanger sequencing following manufacturer's instructions.

2.7.6.2 Analysis of Sanger sequencing reactions

A MegaBACE 1000 detection system (Amersham Biosciences) was used for analysing sequencing reactions by capillary electrophoresis.

2.8. Bioinformatic analysis

2.8.1 Primary sequence analysis by Ion Torrent suite 2.2 software

Primary sequence analysis was performed using the Torrent Suite 2.2. Both SFF (Standard Flowgram Format) and FASTQ files were generated. Raw acquisition data was converted into base calls and reads, Data was normalized by signal correction. After calculation of an overall quality metric, low-quality bases were removed from the output by either filtering out entire reads or trimming low-quality 3' ends. Adapter trimming was also performed.

2.8.1.1 Data alignment against reference genome

Reads were aligned using TMAP aligner (The Torrent Mapping Alignment Program for Ion Torrent) to reference sequences and alignments converted to BAM (Binary Sequence Alignment) files. Genome Reference Consortium Human Build 37 patch release 13 was used as the reference genome.

2.8.1.2 Coverage analysis

Coverage Analysis was performed to provide a detailed breakdown of the level of sequence coverage produced for each targeted genomic region.

2.8.1.3 Variant detection

Torrent Variant Detection Algorithms were used for initial variant calling according to the Bayesian method for germ-line SNP detection when coverage was low to medium (between 10-50) or the frequentist approach for positions with higher coverage (>60) that can be used to detect low frequency variants. Variants were reported with a p value.

2.8.1.4 Genome 1000 database to identify known/unknown variants

All the variants generated from samples were then checked as to whether they are present in the 1000 genomes database or they were novel variants using the 1000 genomes project specific version of the Ensembl Browser found at http://browser.1000genomes.org.

2.8.2 Bioinformatic analysis of Ion data using Active Perl

Secondary sequence data analysis was performed using a custom designed Active Perl script for identifying repeat sequences by RepeatMasker (downloaded from http://www. repeatmasker. org), Blast analysis for genome alignment using a local copy of NCBI Blast 2.2.28+ (Basic local alignment search tool, download from- ftp://ftp.ncbi.nlm.nih.gov/blast/ executables /blast+/2.2.28/). The Perl installation was ActivePerl (www. activestate. com). Reads were extracted from source (. sff) files downloaded from the Ion Torrent server. Two 1 megabase regions were extracted from the GRCh37 assembly of the human genome (http://www.ncbi.nlm.nih.

gov/projects/genome/assembly/grc/ human /): base 138000000 to base 138999999 of chr9 to encompass NOTCH1 and base 7000000 to base 7999999 of chr17 to encompass TP53. Coverage data of fifty bases from either side of the amplicon that were taken into consideration to ensure sufficient amplicon edge coverage. After making sure there is an adequate coverage for each gene, a base-by-base comparison of subject (Blast hit) and query (corresponding region of source read) was made for each hit. Differences were catalogued by type (query insertion, query deletion, actual base change). Known polymorphisms were checked using the UCSC-SNP137 (University of California Santa Cruz Genome Centre:http:// hgdownloadsoe.ucscedu /goldenPath/ hg19/ snp137Mask/) source sequence. This sequence comprises the same assembly of the human genome used for Blast (GRCh37).

2.8.3 Bioinformatics analysis of SOLiD data

2.8.3.1 Alignment of reads

The BWA aligner (Burrows Wheeler Aligner) was used for matching reads to the reference genome (hg19) and generate alignments in the SAM (Sequence Alignment/Map) format for compatibility with Variant calling and other downstream analyses implemented in SAM tools. The software was downloaded from: http://sourceforge.net/projects/bio-bwa/files/latest/download?source=files.The processing was performed following user instructions.

2.8.3.2 Variant detection

The Cake software was used for single nucleotide variants detection. Putative raw variant calls made by the software algorithm were filtered to identify common variation (variants in 1000 genome project dbSNP) for identification of germ line variants and read position filtering (removal of SNPs in read positions where there is a drop in base call quality) [106]. The software was downloaded from: http://cakesomatic.sourceforge.net/ and applied according to the user manual.

For detecting insertions and deletions the Pindel algorithm was used which can detect breakpoints of large deletions (1bp-10kb), medium sized insertions (1-20bp), inversions, tandem duplications. Pindel only considers perfect matching. [107]. Pindel was downloaded from: http://gmt.genome.wustl.edu/pindel/current/ and used according to the user instructions.

2.8.3.3 VEP Analysis

The effect of the sequence variants on the 19,135 protein coding genes in the RefSeq database [108] was annotated using the Variant Effect Predictor (VEP) [109], which predicts the consequence of each sequence variant on all neighbouring RefSeq genes based on a set of 35 consequence terms defined by the Sequence Ontology (SO) [110], such as loss of function including stop gained/lost variants, frameshift indels, donor/acceptor splice and initiator codon variants, missense variants, inframe indels and splice region variants, synonymous variants, 3′ and 5′ UTR variants, deep intronic and inter-genic variants.

2.8.3.4 Reactome Pathway Analysis

RefSeq variant data generated from SOLiD sequencing was entered into the Reactome browser (http://www.reactome.org) to provide a list of significantly down or upregulated pathways that would result from any genetic variants detected in the data.

2.8.4 Sanger sequencing data analysis

Sequence analyzer (Amersham Biosciences) and Sequencer (version4.7, Gene Codes Corporation) were used to analyse the sequence data. BLAST was used to perform searches using the consensus nucleotide sequence generated by the Sequencer to confirm the sequence of the PCR products for respective genes.

2.9 Statistical analysis

All the statistical analysis was performed using student's t-test or Fisher's exact test (SPSS version 16.0).

2.10 Cell functional assays

2.10.1 Cell viability assay

~3000 OSCC cells were seeded per well of a 96-well Nunclon plate. Serial dilutions of DAPT in media were carried out and two concentration ranges were prepared: 50µM to 0.001µM and 50µM to 0.05µM. Triplicates for each DAPT concentration and two controls (0.5% DMSO; no DMSO or DAPT) were set up. For both concentration ranges experiments were performed for 24 hours. EZ4U Cell Proliferation and Cytotoxicity Assay (Biomedica Gruppe, catalogue number BI-5000) was used according to manufacturer's instructions. Here, 20 µl of dye solution was added to 200 µl sample and incubated at 37°C for 3 hours. After incubation, absorbance was measured using Modulus Microplate Reader (Turner Biosystems) at 450 nm with 620 nm as a reference. The reference absorbance at 620 nm was used to correct for nonspecific background values.

2.10.2 Cell growth rates determination by Crystal violet assay

~50,000 OSCC cells were seeded in each 25cm² of flask. Three flasks were treated with DAPT (25µM concentration, Sigma-Aldrich- Catalogue no: D5942), while another three flasks were left untreated. Subsequently the cells were fixed for 5 minutes with 10% formalin and stained for 30 minutes with 0.05% Crystal Violet in distilled water. After that the cells were washed 2x with tap water and the flasks drained by inversion for a couple of minutes. Then, 1 ml of methanol was added straight to 25cm² of flasks to solubilize the dye and 200 µl of the liquid from each flask was then added to 96 well plates for reading the plates at 560nm OD. The flasks were also photographed for colony counts.

For DAPT inhibition the cells were treated with DAPT inhibitor (25 μM) for four consecutive weeks. The proliferation assay procedure was performed every week in triplicates (DAPT+/- treated flasks). The number of viable cells was assessed using a spectrophotometer measuring absorbance at 560 nm. Background measurements were deducted from all recordings. A control measurement of crystal violet dye without any cells was measured on each run (mean absorbance 2.5 on all occasions).

2.10.3 Cell migration assay

6 well tissue culture plates were collagen-coated by incubation in 0.2 mg/ml of collagen type I solution (Sigma) for 2 h at 37 °C before rinsing with phosphate buffered saline (PBS, Invitrogen). Each well was seeded with cells to a final density of 100,000 cells per well were maintained at 37 °C and 5% CO2 for 24 h to permit cell adhesion and the formation of a confluent monolayer. Confluent monolayers were then scoured with a sterile 200 microlitre pipette tip to leave a scratch of approximately 0.4–0.5 mm in width. Three wells for each cell line were left untreated, while three were DAPT-treated (25μM). After an hour, the plate was set up for phase contrast live imaging. A picture was taken for a particular position in each well approximately every 1 hour. The experiment was run for 60 hours. The digitized images were then analyzed using ImageJ software (https://imagej.en.softonic.com) to measure %coverage of surface over time with the experiment repeated in triplicate.

2.11 Protein analysis

2.11.1 Protein Quantification by Bradford assay

Bradford ULTRA kit (Novexin, catalogue number BFU05L) was utilised for protein quantification according to manufacturer's instructions. A dilution series of protein standards was made using the BSA (Bovine serum albumin protein, stock concentration 2mg/ml) with the dilution range falling within the protein range: 0.1 mg/ml – 1.5 mg/ml. Then 20 μl of the samples, standards and a blank (no protein) was mixed with 300 μl

BradfordUltra reagent in triplicates and absorbance were read at 595 nm. Absorbance was measured using Modulus Microplate Reader, with unknown sample protein concentrations determined by comparison to the reference standards.

2.11.2 Western blotting to investigate protein expression

For Western blot (WB) analysis, 50 µg of cell protein lysates were separated on 10% precast polyacrylamide gel (Bio-Rad - Catalogue no. 456-1033EDU) and then electrically blotted onto PVDF membrane (Bio-Rad) using Trans-Blot® Turbo™ Transfer System (Bio-Rad). Membranes were probed for Notch1-4 antibodies. For Notch1 western blotting, membranes were probed with rabbit monoclonal antibody (Abcam, Catalogue no: ab52627), for Notch2 rabbit polyclonal antibody (Abcam, catalogue no ab137665), for Notch3 rabbit polyclonal antibody (Abcam, Catalogue no: ab23426) and finally for Notch4 mouse monoclonal antibody (Abcam, catalogue no:ab134831) were used. All the primary antibodies were used in 1:2000 dilutions. Membranes were then washed with washing buffer and probed with secondary antibody. For Notch1-3, horseradish peroxidase conjugated rabbit secondary antibody (Amersham, GE Healthcare- Catalogue no. NA934) was used in 1:10,000 dilution and for Notch4, horseradish peroxidase conjugated mouse secondary antibody (GE Healthcare, Catalogue no. NA931) was used in 1:20,000 dilution. For amyloid beta precursor protein western blot, anti-amyloid beta precursor protein antibody (Abcam: ab32136) was used. β-actin was used for loading control where the membranes were probed with rabbit polyclonal antibody to human β-actin (Abcam, Catalogue no: ab8227) diluted in 1:2000 and with horseradish peroxidase conjugated secondary rabbit antibody diluted in 1:10,000. The membranes were washed again using washing buffer. The bands were visualized using enhanced chemiluminescence (GE Healthcare, Catalogue number RPN2108) on X-ray film. The size of protein bands was estimated by running protein ladder (Thermo Fisher Scientific; Catalogue number: 26619) parallel where broad molecular weight range (10-250 kDa) was covered. All the western blots for Notch1- 4 protein were performed in triplicate.

Chapter 3 : Allelic imbalance at TP53 and CCND1 loci in relation to histopathological subtypes of OSCC

3.1 Introduction

German pathologist David von Hansemann reported aberrant mitosis as a feature of cancer over 125 years ago in 1890 [111]. By 1914 Theodor Boveri had developed this into the hypothesis that chromosomal abnormalities were responsible for the transition from normal to malignant proliferation and that the changes were genetic in nature [112]. A hundred years of progress, culminating in the use of Next Generation Sequencing has led to a detailed description of somatic changes that have arisen in a wide variety of epithelial and non-epithelial malignancies.

3.1.1 Methods for detecting Chromosomal alterations and their significance

The development of methods for cytogenetic analysis was a major breakthrough in the study of chromosomal abnormalities in cancer with the discovery in patients with chronic myelogenous leukaemia (CML) by Nowell and Hungerford (1960) of the Philadelphia chromosome, a region specific, reciprocal translocation between chromosome 9 and chromosome 22 [113]. Combining karyotyping with giemsa banding and fluorescence-in-situ-hybridisation (FISH) allowed further discoveries to be made with increased resolution. Kallioniemi and colleagues, in 1992 introduced comparative genomic hybridisation (CGH) for detecting DNA amplifications and deletions at higher resolution [114]. Ried et al. described in an early study utilising CGH, recurrent alterations in sporadic colorectal cancers (CRC) in which genomic gains affecting chromosomes 7, 8q, 13, and 20q and genomic losses of chromosomes 4, 8p, 17p, and 18q were often observed [115]. Increased resolution power became possible through the advent of Array based CGH (a-CGH), pioneered by Solinas-Toldo et al. in 1997 and Pinkel et al. in 1998, where DNAs from separate genomic regions were spotted out onto an ordered array for probing [116, 117].

The utility of a-CGH was significantly aided by the Human Genome Project allowing the creation of a library of cloned DNA fragments throughout the human genome. These could then be used as the target region probes on DNA microarrays. The initial approaches used arrays with elements from spotting DNA obtained directly from large insert genomic clones such as BACs [118]. Such strategies were then largely

superseded by oligonucleotide microarrays that were less laborious and provided higher genomic resolution. However, other genetic studies have shown that tumour genomes have a wide variety of copy-number phenotypes, indicating different types of genetic instability. For example, colorectal cancer has different levels and types of genomic aberrations, attributable to differences in mismatch-repair competence [119]. The larger arm-level somatic copy number aberrations (SCNAs) detected by CGH methods, occur 30 times more frequently than focal SCNAs when adjusted for size [63]. However, these broad regions of gain or loss can include hundreds of genes, several of which are likely to be potential 'driver' cancer genes [120]. As a result, narrowing the list of potential 'drivers' in large SCNAs has proved difficult [63]. Although large alterations have long been associated with poor outcomes, underlying mechanisms and gene-specific biomarkers remain elusive [121, 122].

Recent advances in genomics aided by next generation sequencing used with large patient cohorts has provided us with a means of assessing the identity of potential drivers within large SCNAs. Next generation sequencing has also provided a means of detecting structural changes accompanied by deep sequencing for the detection of variants. In tumours of squamous cell origin, such as head and neck or bladder squamous cell carcinoma, the gain of chromosome 3q was found to be the most prevalent alteration. Other genomic imbalances commonly observed in these tumour subtypes include gains located at chromosome 1, 7, and 20, with losses located at 4, 11, 16, 17, and 19 [21, 123].

3.1.2 Allelic Imbalance in OSCC

Patterns of DNA copy number aberration (CNA) including loss of heterozygosity (LOH) found in OSCC have been considered to be characteristic for different points in the natural history of the disease [124]. PCR for detecting RFLPs and microsatellites was initially used for detecting copy number change [125] but thereafter the majority of studies used CGH and array-CGH. Oligonucleotide microarray based CGH (a-CGH) has been used to score allelic imbalance across the OSCC genome. Publications for

key investigations are listed in Table 10. with the methods used and the regions found to be affected. Huang et al. analysed DNA from primary tumour tissue (n=34 OSCC samples of 75 total HNSCC samples) with matching normal tissue controls, and observed gains in regions 3q,5p and 8q with loss of 3p for all HNSCC and additional gains in 1p and 11q12-13 specific for OSCC [126]. Smeets et al. also used OSCC primary tumour tissue samples (n=37) but reference genomic DNA was from a random panel of blood donors as controls [127]. Gains and losses, at chromosomes 3q, 8q and 11q13 and losses of 3p observed were similar but additional losses of 5q, 11q23 and 18q were found. Oga et al. assessed copy number variation in OSCC primary tumour samples (n=17) covering a range of different stages (stage 1-4). Although gains (3q, 8q, Xp and Xq) and losses (3p, 4p, 8p, 9p, 13q and 17p) were found in a number of regions, the total number of gains and losses was greater in stage III/IV samples (mean 8.11+/-4.91) compared to stage I/II (mean 1.11 +/-1.49), with 3q gains found more significantly in stage III & IV tumours (6/8) v stage I/II (1/9) [128].

Table 10. Summary of key publications reporting the use of CGH to investigate allelic imbalance in OSCC primary tumour samples.

Tumour materials	Control samples	Method /s	Observation	References
OSCC primary tumour tissue (n=17)	Healthy volunteers	Comparative genomic hybridization (CGH)	Gains in regions 3q, 8q, Xp and Xq. Losses detected in 3p, 4p, 8p, 9p, 13q and 17p. Total number of gains significantly greater in stage III/IV than stage I/II. 3q gain in stage III & IV tumours (6/8) v stage I/II (1/9).	[128]
OSCC primary tumour tissue (n=34, of 75 HNSCC samples)	Matching normal tissues	Comparative genomic hybridization (CGH)	Gains in regions 3q,5p and 8q with loss of 3p and 17p for all HNSCC. Additional gains in 1p and 11q12-13 specific for OSCC	[126]
OSCC primary tumour tissue (n=47)	Matching normal tissues	Array- comparative genomic hybridization (CGH)	Gains in regions 8q, 9q and 11q, most frequent losses of 3p, 4q and 17p. Deletion of 3p and gain of 11q more prevalent with lymph node metastasis.	[129]
OSCC primary tumour tissue (n=37)	Reference genomic DNA from random panel of blood donors	Array- comparative genomic hybridization (CGH)	Gains at chromosomes 3q, 7q, 8q and 11q13 and losses at chromosomes 3p, 5q, 11q23 and 18q.	[127]
OSCC primary tumour tissue (n=54)	Marginal normal tissues	Array- comparative genomic hybridization (CGH)	Gain of regions including 3q, 5p, 8q, 9q, 11q, and 20q. Loss of 3p, 8p, 9p, 17p and 18q regions	[130]
OSCC metastatic primary tumours with paired lymph node metastases (n=25)	Non-metastatic primary tumours	Array-comparative genomic hybridization (CGH)	Gain of regions including 3q, 9q,11q13, 14q, 16p, 18p and 20q. Loss of 3p, 4q, 8p, 9p,10p, 13q, 17p, 18q and 21q regions. Gains of 7p, 8q and 17q more common with lymph node metastasis	[131]
OSCC primary tumour tissue (n=205)		Array- comparative genomic hybridization (CGH) coupled with whole genome sequencing	Losses of 3p and 8p, and gains of 3q, 5p and 8q chromosomal regions with activating mutations of PIK3CA or HRAS, coupled with inactivating mutations of CASP8, NOTCH1 and TP53	[21]
OSCC primary tumour tissue enriched for ECS cases (n=65, 22 ECS)	Normal controls (n=6)	Array- comparative genomic hybridization (CGH)	Gain of regions 3q, 5p, 7p, 8q, 9q, 11q, 14q, 17q, 20p, 22q. Loss of 3p and 8p regions.	(B. Lloyd unpublis-hed data)

The evidence gained from similar studies has provided some insight into the association of allelic imbalance observed with prognosis. Shu-Chun et al. found in their cohort of OSCC tumour samples that deletion of 3p (4/23 N0 v 12/24 N+) and gain of 11q (2/23 N0 v 9/24 N+) were more prevalent in those samples with lymph node metastasis [129]. Yoshioka et al., reported gains of 7p (2/10 N0 v 10/15 N+), 8q (5/10 N0 v 14/15 N+) and 17q (2/10 N0 v 11/15 N+) that were more common with lymph node metastasis in a cohort of OSCC samples [131]. As previously discussed (section 1.2), lymph node metastasis in OSCC is associated with an aggressive phenotype and is a key poor prognostic indicator. However, OSCC patients with lymph node metastasis do not have a uniformly poor prognosis [103, 132] with other independent risk factors including lymph node metastasis harbouring ECS being a further predictor of poor prognosis (section 1.3). Conversely, a powerful a-CGH study combined with whole genome next generation sequencing on a large OSCC cohort (n=205) by the Cancer Genome Atlas consortium reported that the absence of allelic imbalance in OSCC may be associated with good outcomes. A subgroup of oral cavity tumours was identified with favourable clinical outcomes displaying infrequent copy number alterations with activating mutations of PIK3CA or HRAS, coupled with inactivating mutations of CASP8, NOTCH1 and TP53 [21]. These data built on a previous study related to the Cancer Genome Atlas network that used a-CGH in a more limited number of OSCC cases (n=37) and classified samples on the basis of the genetic pattern detected. They concluded that three groups were present, one characterized by low level of chromosomal aberration (average 2.8 chromosomal breakpoints, n = 8), another by medium level (average 30.0 chromosomal breakpoints, n = 26) and one with high level (average 52.8 chromosomal breakpoints, n = 5) chromosomal aberrations. This classification was significantly ($p = 0.003$) associated with survival. In contrast, there was no significant association with pathological tumour stage or lymph node status [133].

Together, these studies suggest that both the level of chromosomal aberration and the specific sites of any changes are of relevance. A greater number of chromosomal aberrations are clearly more likely to give significant allelic changes. Lymph node metastasis and overall survival have been considered but no studies specifically

considered ECS. Detailed analysis of TCGA data by Roy et al. linked chromosomal aberration to clinical outcome [134]. Specific interrogation of their data with regards to outcomes for HNSCC (n=522) reveals that 3q and 8q regions most frequently have gains and 3p and 8p regions most frequently have losses. However, gain of 7p is most significantly associated with poor survival, suggesting a complex picture. To genetically characterise the basis for aggressive phenotype in OSCC, further we conducted an investigation of allelic imbalance associated with different histopathological subtypes of OSCC, specifically focussing on ECS, the most adverse clinical prognosticator.

3.1.3 Investigation of allelic imbalance in association with ECS: a previous pilot study

Previously our laboratory had used aCGH to examine copy number changes in relation to ECS. The series was enriched for cases having associated ECS (20 ECS+ and 26 ECS- cases, section 2.2). We predicted that particular molecular pathways would be altered according to the presence of ECS and this would be reflected in relative enrichment of chromosomal gains or losses in the regions where the corresponding alleles were located. DNA from fresh frozen primary tumour samples (n=46) had been analysed using Nimblegen arrays having 720,000 probes at an average density of 4,260 probes/μm^2. DNA from each sample was compared to a uniform mixture of DNA from 6 random, unaffected individuals. The CGH data obtained by this approach revealed a number of regions showing amplifications or deletions that had had also been found to be affected in other epithelial malignancies (B. Lloyd unpublished data). Significant copy number alterations included chromosomal arm amplifications of regions 3q, 5p, 7p, 8q, 14q, 20p, 22q and chromosomal arm deletions of regions 3p, 8p. This is summarised in the ideogram of Figure 4. Considering this data in the context of the established literature, discussed previously in section 3.1.2 and summarised in table 10, there was overlap but also some differences. Affected chromosomal regions in common were 3q, 5p, 8q, 9q and 11q for gains (≥3 studies) and for losses (≥3 studies) were 3p, 8p,9p. Gains of 14q, 20p, 22q were unique to the literature.

No gains or losses were specifically associated with ECS and therefore genes within the segments that showed significant copy number change within our data were subjected to hypergeometric analysis in comparison to canonical pathways to determine which processes may be enriched. Pathways for multiple processes were implicated, including cytoskeletal remodeling: TGF and WNT pathway ($p = 7.8 \times 10^{-8}$), keratin filaments ($p = 1.4 \times 10^{-7}$), transcription: receptor mediated HIF regulation ($p = 4.1 \times 10^{-7}$), immune response: IL9 signaling pathway ($p = 1.1 \times 10^{-6}$) and chemotaxis ($p = 3.8 \times 10^{-6}$), Notch signaling and TNFs/NF-kB/IAP apoptosis for both T4 and T2 tumours ($p = 9.7 \times 10^{-5}$ and $p = 1.0 \times 10^{-3}$, respectively). More detailed segmentation analysis to narrow down the minimal overlapping regions of loss/gain between samples in commonly affected regions found complex patterns. The 11q13.2-11q13.3 region was found to have aberrations in a large proportion of tumour samples (21/46) with predominant gains (20/46) particularly in ECS cases (11/20) (Figure 5.). This 0.514 Mb region contains 5 key genes: MYEOV, CCND1, FGF19, FGF4 & FGF3, all playing a key role in cell proliferation and differentiation. In contrast, some studies of OSCC have reported significant loss of 17p13 region (harboring the TP53 gene) [135, 136], loss of 17p13 was very infrequent in our cohort (1/46 demonstrating a deletion). Significant differences between ECS and Non-ECS cases were not found by the study by pathway analysis.

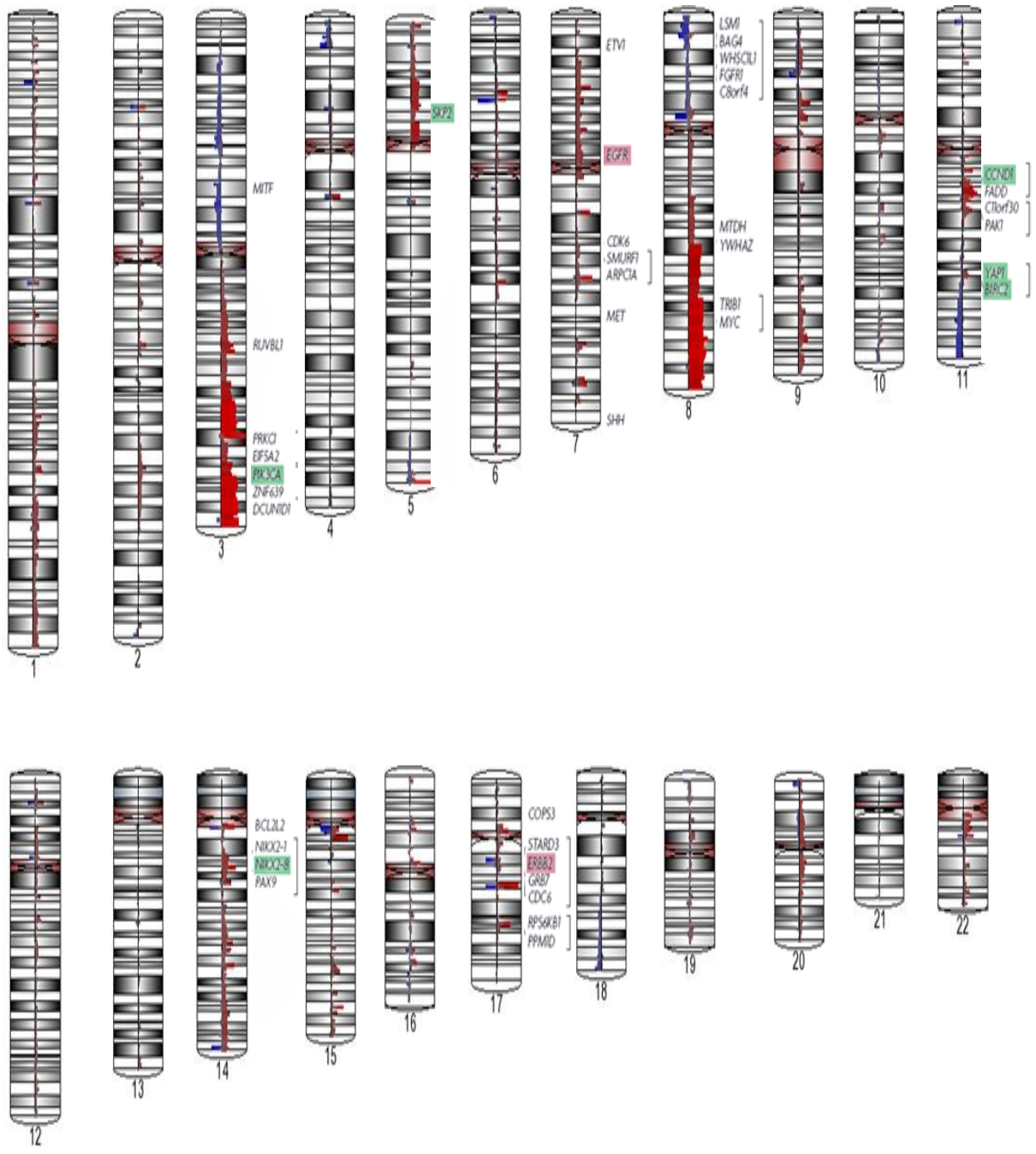

Figure 4. Summary ideogram of allelic imbalances found by CGH in our series of OSCC cases. Significantly amplified regions are highlighted in green and significant deleted regions in pink, with the locations of known cancer related genes listed at their appropriate locations next to the ideogram. (B. Lloyd, unpublished)

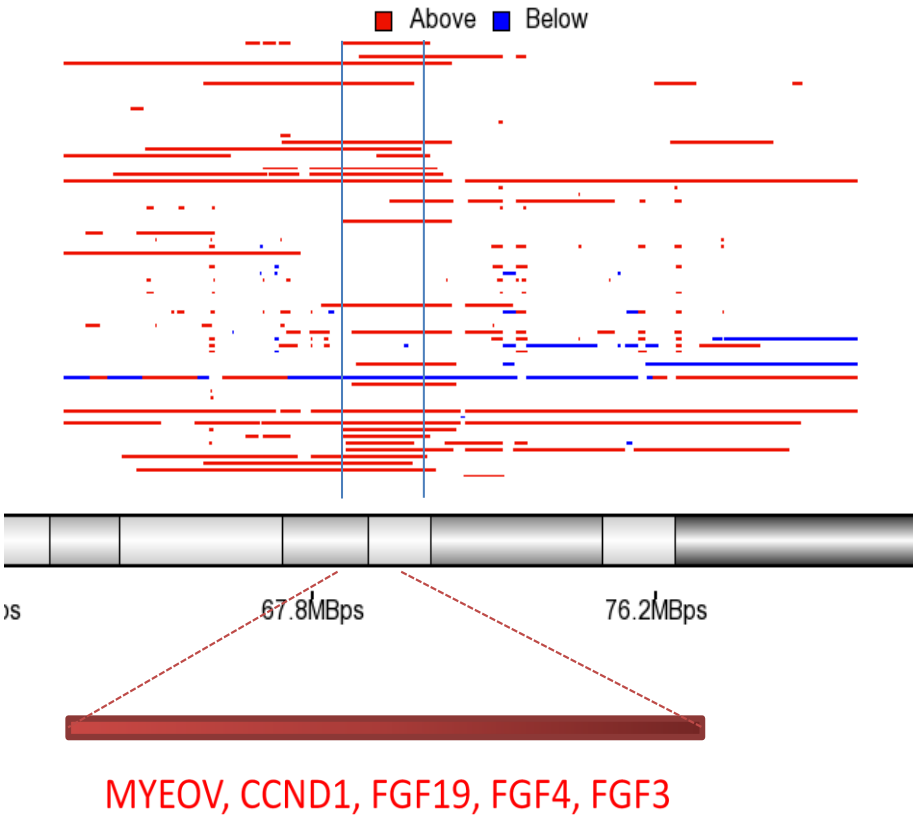

Figure 5. Minimal regions of allelic imbalance found by CGH in our series of OSCC cases shown for 11q13.2 – 11q13.3. Red lines signify regions of average gain and blue signify regions of average loss of genomic DNA. [Data courtesy of Dr. Bryony Lloyd]

3.1.4 Significance of chromosomal regions 11q13 and 17p13 for OSCC

CCND1 and TP53 are involved with key biological aspects of cancer development, regulation of the cell cycle and apoptosis respectively. TP53 and CCND1 both show high levels of allelic imbalance in HPV-negative HNSCC [16].

11q13, the location of the CCND1 gene is frequently amplified in different cancers [137]. In the case of OSCC, allelic imbalance is observed for 11q region and high-level amplification (>4 copy gains) of CCND1 has been found in up to 22% of cases [136]. Over-expression rates of CCND1 was also found for 25-70% of patients with oral

cancers [138]. Cyclin D1 protein overexpression was significantly associated with lymph node metastasis (p = 0.002). 11q13 and in particular CCND1 may therefore be significant drivers of oral cancers.

TP53 is a well-known and understood tumour suppressor gene, very commonly mutated in human cancers (~ 50%); including 25-69% of HPV negative oral cancers [102, 104] Loss of the TP53 gene in terms of allelic imbalance, dominates the genomic landscape of OSCC in common with many other tumour types [135, 136]. p53 protein downregulation occurs in OSCC cases and is associated with poor patient outcome [139]. No studies have been conducted so far comparing associations of these two genes (CCND1 and TP53) with ECS.

No regional gain or loss was found to associate with ECS in our a-CGH analysis. Similarly, very little is known of any molecular signatures that are associated with ECS. However, although our cohort has expected frequencies of 11q13 gain, there appears to be a significant deficit of TP53 loss compared to literature reports. Five studies had already found 11q gains in between 25 and 70 % of cases, with two studies reporting 17p loss in between 50-75% of cases (see Table 10.) whereas we found 43% and 4% respectively in our a-CGH study. A possible explanation is that our samples may be genetically different from other cohorts, with a low rate of TP53 change. There may be technical explanations, for example admixture of tumour cells and normal cells leading to reduced sensitivity of detecting allelic loss [140]. However, pathology reports (not shown), suggested that tumour cells comprised average ~70% of our samples (n=46) and therefore sensitivity should not have been compromised in this way. Gains of more than 1 copy were observed for 11q13 in our series. We were concerned that our approach was underestimating 17p loss and only detecting higher levels of 11q gain. Array CGH also leaves open the possibility of balanced allelic loss and gain [141]. I therefore decided to examine the validity of our results specifically for candidate regions 11q13 and 17p by using SNP analysis for detecting actual allelic imbalance, in the expectation that this would be more sensitive than aCGH.

Multiple approaches have been used to detect allelic imbalance using end-point readings of PCR [142], but they may not be accurate in quantifying the ratio between

the alleles. The accuracy of SNP analysis can be further improved by quantitative PCR (q-PCR), since two or more possible alleles can be labelled with a different flourescent dye each, allowing allele calls to be made according to the ratio of possible fluorescent probes present. q-PCR also has a greater dynamic range than aCGH, further improving the possibility of detecting allelic imbalance. Chen et al. demonstrated using this method that a Pearson Correlation Coefficient (R^2) of 0.9798 and 0.9868 when compared to a standard curve of fixed ratios with the ability to quantify allelic imbalance ranging from 2:1 – 8:1 [143].

Detection of a clinical association is dependent on study power. We estimated that a study performed on TP53 assuming a significance level of 0.05 and a power of 80% with an expected frequency of 25-69% for allelic imbalance, requires only 12-52 samples to demonstrate such a difference at this power. Similarly, for CCND1 the power calculation is the same, if similar assumptions were made with a potential frequency of 25-70% considered. As our available sample set was 46 cases and other gains and losses were too infrequent to be useful, CCND1 and TP53 were considered to be the most appropriate candidates for testing in our genetically distinct sample set.

3.2 Hypothesis and Aim

Hypothesis: ***Increased copy number change in CCND1 or TP53 is associated with ECS, an aggressive phenotype of OSCC.***

Aim: To quantify and compare CCND1 and TP53 gene allelic imbalance in OSCC samples of defined histopathological subtypes using single nucleotide polymorphisms (SNPs) via a q-PCR method.

3.3 Results

3.3.1.1 Establishment and Verification of functioning SNP PCR

Several SNPs were selected to allow for non-informative individual for TP53 (resides in 17p) and CCND1 (resides in 11q) from the National Centre for Biotechnology Information (NCBI) dbSNP database on the basis of a high level of heterozygosity in the European population. Predesigned SNP QPCR assays were selected (see section 2.5.1, Table 5). Standard PCRs were performed (section 2.5.1) and the products analysed by agarose gel electrophoresis (reference section 2.5.2) to identify the conditions giving the clearest expected bands and therefore greatest specificity.

Company specified pre-optimised conditions were applied for one TP53 (C-26603915) probe targeted against an intronic TP53 SNP (ID: rs2909430, location: Chr.17: 7578645 on NCBI Build 37) and were then repeated for all other probes of interest. The C-26603915 assay was performed using DNA from 6 randomly selected samples from the human control DNA panel (section 2.3) and 2 negative DNA controls. The presence of expected ~120bp bands support successful PCR amplification.

3.3.1.2 Preliminary classification of homozygote and heterozygote samples

Next, 96 samples from the human control DNA panel were screened for candidate homozygote and heterozygote samples that could be used to test for the sensitivity and specificity of allelic discrimination. The samples were subject to q-PCR using optimal conditions for enzyme activation, template denaturation and annealing/extension (see section 2.5.1). CFX manager^{TM} software was used to monitor the q-PCR reaction and the baseline and detection threshold were both set for sensitive significant amplification detection above background. Duplicate runs were performed for each probe.

Figure 6. demonstrates RFU values (the increase of fluorescence level) and Ct (PCR threshold cycle) values obtained using the C-26603915 TP53 assay. A successfully amplified product (ΔCt<50) for both FAM dye (used to detect allele1) & HEX dye (used

to detect allele2) probes indicated a heterozygote, whereas an amplified product for FAM but not HEX (ΔCt=50), or vice versa, indicated a homozygote for either Allele 1 or 2 respectively. A negative control was also used with an expected ΔCt of 50 recorded for both FAM and HEX in the absence of DNA. These criteria were applied using the data analysis software allowing the cut offs (blue and red lines) to be drawn, segregating the samples into these distinct categories. On this basis, the graph shows 35 samples were classified as heterozygote, with the remainder either homozygote for allele 1 (33) or 2 (28). These data confirmed the ability of the method to discriminate between the different alleles of a SNP for both genes, allowing for the more detailed assessment of sensitivity utilising the homozygote detected samples.

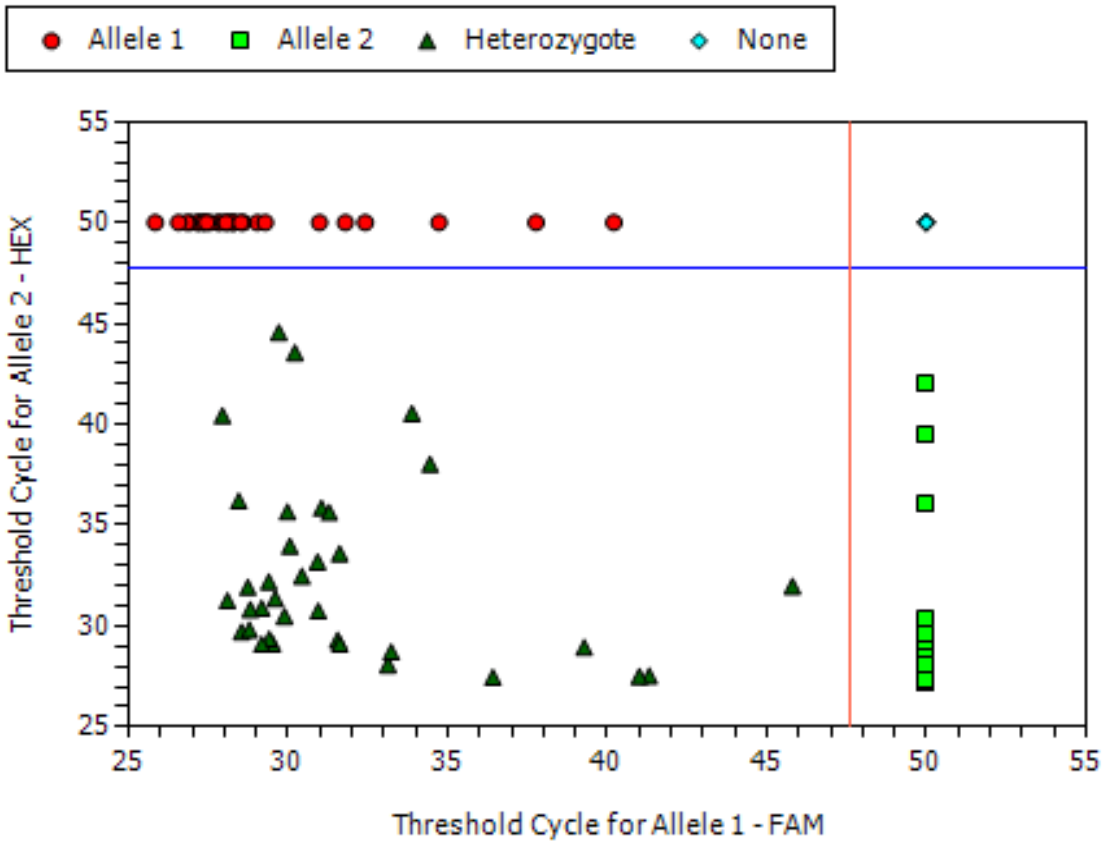

Figure 6. Threshold cycle analysis for maximum fluorescence intensity using the C-26603915 TP53 assay with human random control DNA samples.

3.3.1.3 Specificity of detection of minor alleles

Having identified a number of homozygote and heterozygote samples as outlined in the previous section, 2 samples of each homozygote were mixed in different fixed proportions and amplified by the C-26603915 TP53 probe to determine minor allele detection specificity. The q-PCR was performed in triplicates as previously described (see section 2.5.1). Table11. demonstrates the RFU and Ct values for both SNP probes generated from the first set of experiments. The lower the Ct value, the higher percentage of the respective allele was deemed present. Firstly, the mean Ct values for allele 1 (Table 11. pink rows), revealed discernible differences between the different samples tested with a mean Ct1 for 100% allele 1 of 23, 90% gave 26 and 50% gave 28. The mean Ct values for allele 2 (Table 11. white rows) were similar. The negative control with no DNA displayed ΔCt=50 for both fluorescence recordings confirming no or little DNA. These observations suggested that the method was sufficiently sensitive enough to detect a 10% difference in allele amount.

Table 11. Fluorescence level in relation to Threshold cycle for human control DNA samples mixed in predetermined proportions.

% allele 1	% allele 2	Mean Fluorescence level (RFU)	Mean PCR threshold cycle (Ct)	Call
100	0	5005	23	Allele 1
0	100	5052	23	Allele 2
90	10	4904	26	Heterozygote
50	50	4548	28	Heterozygote
50	50	4488	28	Heterozygote
10	90	4917	26	Heterozygote
0	0	3	50	None
0	0	1	50	None

3.3.1.4. Assessment of ability to quantify allelic imbalance

We then went on to vary the overall DNA amount in a further experiment to model a situation of allelic imbalance that potentially we could encounter on assessment of tumour samples. The method used the previously identified human control samples homozygous for allele 1 and 2 (see section 3.3.1.3) and q-PCR performed with the same optimised conditions (section 2.5.1). Stock test samples were prepared with a total DNA amount of 200 ng and mixture of alleles in a variety of proportions of allele 1: allele 2 (200:0, 190:10, 180:20, 160:40, 100:100, 40:160, 20:180, 10:190, 0:200). 5 subsequent 1 in 5 dilutions were made of these stock test samples, allowing the overall DNA amount to be varied from 200ng - 0.064 ng. All samples were analysed in triplicate and then used to plot a standard curve. For the TP53 (C_26603915) probe there was a correlation observed for both allele 1 and 2 with coefficient of determination values (R^2) of 0.62 and 0.67 respectively (Fig 7. A & B). A possible reason for R^2 being <1 was due to variability across samples with low values of DNA (particularly <10ng) for both allele 1 and 2 being overestimated by this method potentially as a result of non-specific probe binding. The amount of over-estimation was different for different probes, with the TP53 probe illustrated here generally over-estimating at around the same level.

A. B.

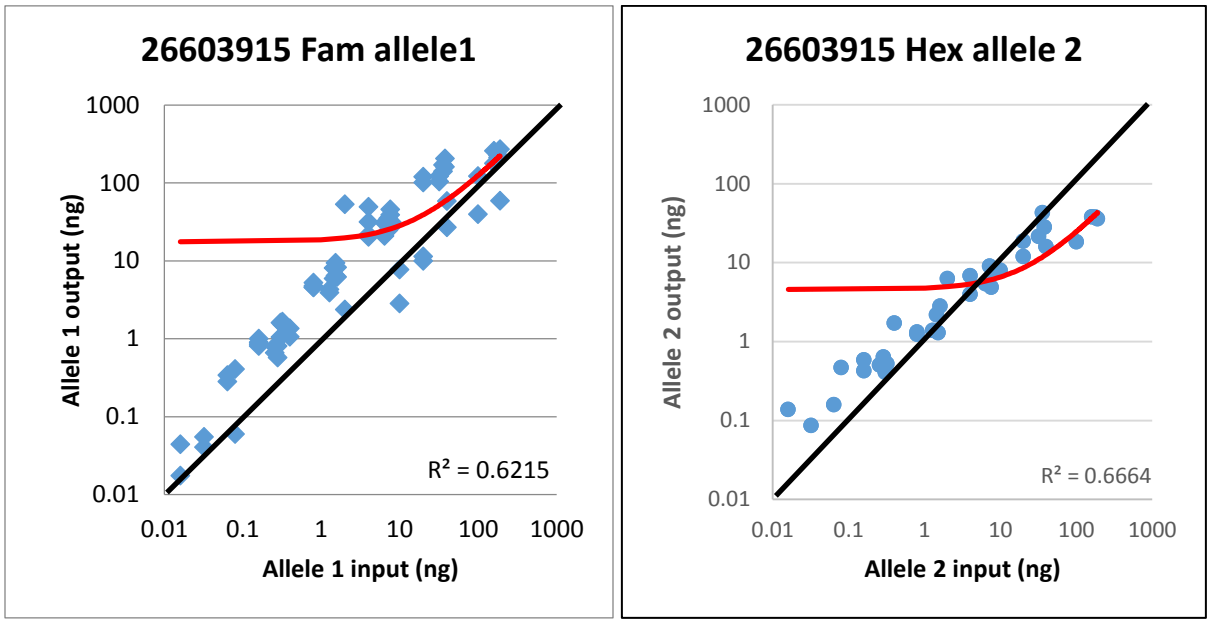

Figure 7. Allelic amount and allelic imbalance compared to standard (antilog) curves from control DNA. A., B. Scatter plots of input amount of DNA v output amount (ng) recorded following q-PCR for allele 1 (A.) and allele 2 (B.) Red line = line of best-fit, black line = expected standard curve, corresponding R^2 values are demonstrated on graphs.

Having established that the method lacked sensitivity to accurately quantify differences, it was necessary to further clarify the level of allelic imbalance at which a call of amplification or deletion could be confidently made. To clarify this level of sensitivity further so that a limit of detection (LOD) with regards ΔΔCt could be identified, above or below which samples could reliably be considered to have either a TP53 or CCND1 deletion or amplification, a further experiment was performed. qPCR was performed similarly to previously (see section 2.5.1) using the C_26603915 probe (TP53) with the same human control DNA homozygote allele 1 and 2 samples as used previously. A DNA amount of 20ng/µl was used initially with both alleles then mixed with less than 4 fold allelic imbalance differences (20v80, 25v75, 30v70, 70v30, 75v25, 80v20). All samples were then further diluted 1 in 10 to give a 2ng/µl total DNA concentration. Each

sample was assayed with eight replicates with the ΔΔCt detected for each sample recorded.

Figure 8. demonstrates the ΔΔCt values from this experiment using the C_26603915 TP53 probe. samples mixed as 70v30,75v25 and 80v20 with expected ΔΔCt values of 2,3 and 4 respectively, returned mean ΔΔCt values of 4.18, 6.56 and 6, suggesting over-estimation of quantification, but they were all statistically different from the reference control on student's t-test ($p<0.05$). Samples mixed in the opposite manner as 30v70, 25v75 and 20v80 with expected ΔΔCt values of 0.5, 0.33 and 0.25 had mean ΔΔCt values of 0.61, 0.55 and 0.49 respectively. Only the 20v80 (expected ΔΔCt = 0.25) in the 1 in 10 dilution group sample demonstrated statistical significance compared to the reference control ($p<0.05$) on student's t-test. The assay limit of detection (LOD) was determined on the basis of the samples that remained statistically different from reference (1:1) with the lowest detected imbalance above control for amplification and the highest detected imbalance below the reference for deletion (see arrows on Figure. 8 for TP53). For the TP53 SNP assay this meant cut offs when average values of ΔΔCt detected are >2 or <0.25. Hence, by applying these thresholds to the interpretation of subsequent tumour data we could conclude with greater confidence whether allelic imbalance was significantly present or not.

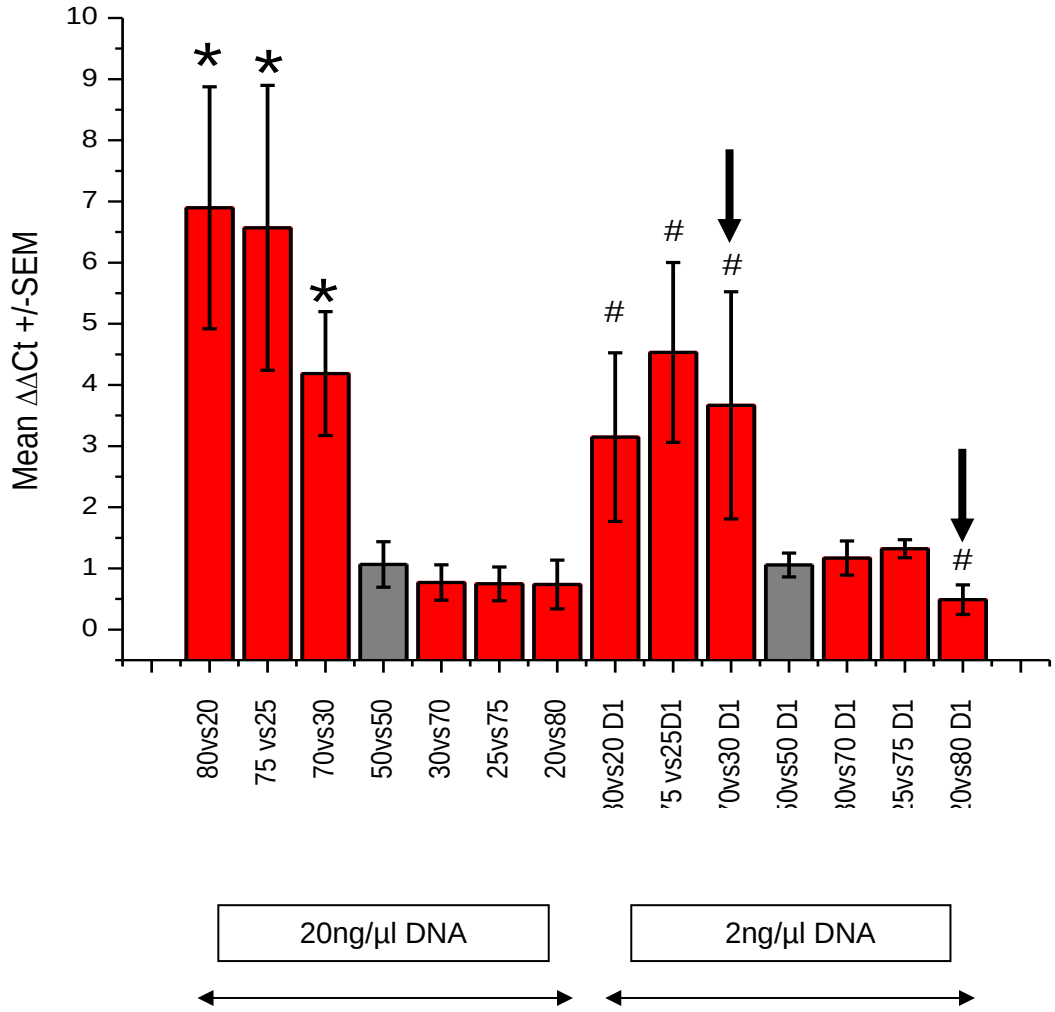

Figure 8. Mean fold difference for fixed levels of allelic imbalance using the C_26603915 TP53 probe. All readings had eight replicates with the findings shown for two different DNA amounts (Original dilution=20ng/µl total DNA concentration, Dilution1 (D1) =2ng/µl total DNA conc.) *=p<0.05 on student's t-test compared to original dilution 50v50 control, #=p<0.05 on student's t-test compared to dilution 1 50v50 control. Arrows mark limits of detection of the assay.

3.3.1.5 Statistical analysis of allelic discrimination

The optimisation experiments described in sections (3.3.1.2 – 3.3.1.5) were repeated for all TP53 and CCND1 SNP probes (Table 5). As the first part of statistical analysis to select the best TP53 and CCND1 probes to use on subsequent tumour samples, we specifically assessed how sensitive the q-PCR method was to detect heterozygote samples for a particular probe, by performing a statistical analysis based on the findings from all probes once the experiment described in 3.3.1.3 was replicated with them. The triplicate Ct values for both alleles and all probes at a 10% difference sensitivity level (100% v 90%) were compared by unpaired t-tests. Table 12. summarises the p values, confirming a statistical significance ($p<0.05$) in all comparisons except C_8727721, C_2880090 assay for TP53 and C_744724 for CCND1. The TP53 probe C_26603915, previously used as an illustrative example, did show significant difference at 10% sensitivity for minor allelic discrimination (allele 1 $p=0.010$ and allele 2 $p=0.024$) and is highlighted in red. Probes that did not show a significant difference were immediately excluded from consideration for use on OSCC tumour samples.

The remaining probes that were identified as sensitive enough for 10% minor allelic difference detection, C_26603915, C_2880088, C_2403545 for TP53 and C_744721, C_25800715 CCND1 probes, had their data produced from the 3.3.1.4 set of experiments analysed for final probe selection. Table 13. Shows a summary of the parameters for allelic discrimination by q-PCR for all these five probes. C_26603915 and C_744721 probes were the most suitable for tumour sample analysis based on their greatest sensitivity for allelic discrimination as well as least levels of over/under estimation of $\Delta\Delta$Ct.

We can conclude that the q-PCR was not reliable for accurate consistent quantification of smaller differences, however, the method can detect allelic imbalances when an LOD Ct threshold was applied (≥ 2 ≤ 0.25 for TP53 & ≥ 3 ≤ 0.5 for CCND1).

Table 12. TP53 and CCND1 probe ability to detect a 10% minor allelic difference. Student's t-test results for all TP53 and CCND1 probes comparing Ct values recorded by q-PCR between 100% v 90% (i.e. 10% difference) allele groups are displayed.

Gene	SNP ID	Heterozygocity	Assay number	Allele	p value
TP53	RS1614984	40%	C_8727721	1	0.065
				2	0.072
TP53	RS1642785	30%	C_2880090	1	0.051
				2	0.076
TP53	RS2909430	22%	C_26603915	1	0.010
				2	0.024
TP53	RS12951053	16%	C_2880088	1	0.034
				2	0.028
TP53	RS1042522	30%	C_2403545	1	0.044
				2	0.050
Cyclin D1	RS3212891	43%	C_25800715	1	0.012
				2	0.054
Cyclin D1	RS649392	43%	C_744724	1	0.089
				2	0.050
Cyclin D1	RS678653	43%	C_744721	1	0.005
				2	0.008

***Summary of optimisation experiments*:**

Table 13. Summary of key q-PCR selection parameters for TP53 and CCND1 allelic discrimination for the 5 assays selected for use with tumour samples. The key parameters from experiments described in 3.3.1.1 (Establishment and Verification of functioning SNP PCR), 3.3.1.2 (Preliminary classification of homozygote and heterozygote samples), 3.3.1.3 (Specificity of detection of minor alleles) and 3.3.1.4 (Assessment of ability to quantify allelic imbalance) using 3 TP53 SNP probes and 2 CCND1 probes are described above. Selected probes demonstrating the best levels of sensitivity that were used further for allelic imbalance analysis in OSCC samples are shown in red.

	C_26603915 TP53 probe	C_2880088 TP53 probe	C_2403545 TP53 probe	C_744721 CCND1 probe	C_25800715 CCND1 probe
100 v 90 p value Allele 1	0.010	0.034	0.044	0.005	0.012
100 v 90 p value Allele 2	0.024	0.028	0.050	0.008	0.054
R^2 Allele 1	0.62	0.61	0.60	0.96	0.67
R^2 Allele 2	0.67	0.58	0.62	0.58	0.49
R^2 ΔΔCt	0.64	0.60	0.61	0.51	0.50
Output ΔΔCt when input = 2	4.18	4.30	4.21	1.69	1.53
Limit of Detection	≥2 , ≤0.25	≥2 , ≤0.25	≥2 , ≤0.25	≥3 , ≤0.5	≥3 , ≤0.5

3.3.2 Allelic imbalance of TP53 and CCND1 loci in OSCC Tumour samples

The specific aim of this part of the study was to test the hypotheses that increased copy number change in CCND1 or TP53 gene were associated with ECS, an aggressive phenotype of OSCC. Optimisation experiments had determined that allelic imbalances could be detected by the method with confidence limits applied. By using an ECS enriched sample set of exactly the same samples (n=46) that had previously been investigated by CGH, direct comparisons between results obtained by the two assay systems could be made.

q-PCR assay was set up as described (section 2.5.1) and only heterozygote samples would be informative for the analysis. For TP53, 22/46 samples were heterozygote, whereas for Cyclin D1 37/46 samples were heterozygote, and the homozygote samples were excluded from analysis. All mean $\Delta\Delta$Ct values recorded for the tumour samples assessed by q-PCR using both CCND1 and TP53 probes are shown in Figures 9A &B. To complete the analysis a student's t- test comparing the sample mean $\Delta\Delta$Ct to the reference was performed to assess for significant allelic imbalance. For TP53 8/22 samples overall were found to have allelic imbalance with 6 samples (SP, SY, SR, S9, SV, SU) found to have deletions and 2 samples (S11, S8) showed amplification. For CCND1, 15/37 samples in total displayed allelic imbalance with 4 samples (S1, S2, S4, S18) showing deletion and 11 samples (SW, S16, S15, S21, SY, SX, SI, SR, S19, SF, S11) showing amplification. Overall the rates of allelic imbalance for TP53 and CCND1 were 36.4% and 40.5% respectively, both ~10-20% below levels of allelic imbalance reported in the literature for the regions harbouring these genes. These initial observations then allowed for further assessments to be performed with regards correlation of the findings with the previous CGH data as well as the histopathological subtype of the sample.

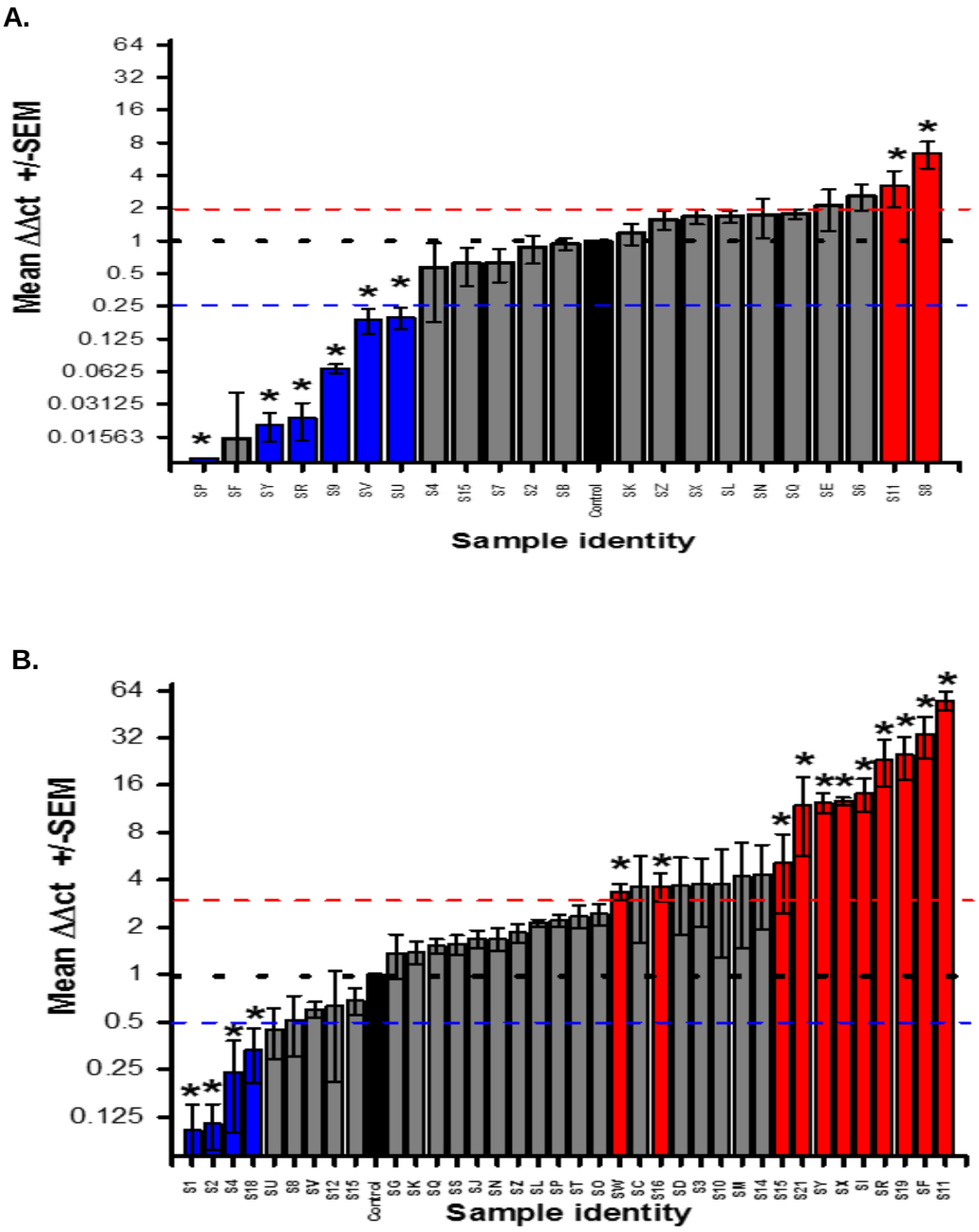

Figure 9. [A]. TP53 allelic imbalance in OSCC tumour samples. [B]. CCND1 allelic imbalance in OSCC tumour samples. Bar chart representing mean ΔΔCt value +/- standard error of the mean (5 repeats) for all heterozygote samples. Limits of detection determined from previous optimisation experiments were ΔΔCt >2 (dashed red line) & *p<0.05 v control for amplification, ΔΔCt <0.25 (dashed blue line) & *p<0.05 v control for deletion. Control ΔΔct = 1. ΔΔct >3 & *p<0.05 v control. ΔΔct <0.5 & *p<0.05 v control (student's t-test)

3.3.3 Comparison of TP53 and CCND1 SNP genotyping data and a-CGH data in OSCC samples

To validate our earlier a-CGH observations a direct comparison was required with regards to concordance of allelic imbalance detected by both methods and identification of any significant differences between both sets of findings. For TP53, of the samples identified as heterozygote, 2/22 showed amplification & 6/22 showed deletion, compared to 1 amplification and 1 deletion identified by the previous a-CGH study. Both these levels of TP53 change were significantly lower than that reported in the literature. For Cyclin D1 of the heterozygote samples 11/37 showed amplification and 4/37 deletion, compared to 20 amplified and 1 deletion in the a-CGH study. If sample-specific data was compared across both techniques, overall agreement with the a-CGH data was 63.6% for TP53 and 75.7% for CyclinD1 with a free marginal kappa statistic of 0.47 & 0.53 respectively. Table 14. summarises the contrasting findings from both techniques.

The difference in concordance of both methods and the fewer CCND1 amplifications observed by q-PCR along with greater TP53 deletions may be explained by the respective technical advantages / disadvantages of both techniques. As only heterozygote samples could be considered for the q-PCR, sample size was smaller and earlier experimental optimisation had revealed a tendency of the TP53 probe towards over-estimating allelic imbalance as well as a tendency of the CCND1 probe used towards under-estimation (as previously discussed in section 3.3.1.4). On the other hand, lack of probe hybridisation in the a-CGH method can lead to predominant under-estimation of both amplifications and deletions and is particularly less sensitive than PCR based methods in detecting deletions [115]. Although low levels of TP53 allelic imbalance were found by both methods, the two sets of findings did not show complete concordance and hence the SNP genotyping data produced here could not fully validate the earlier a-CGH observations.

Table 14. Summary of CCND1 and TP53 allelic imbalance detected by SNP assay and a-CGH.

	CCND1	TP53
Total sample number	46	46
No. of heterozygote samples	37	22
No. amplified heterozygote samples	11	2
No. amplified in CGH data	20	1
No. deleted heterozygote samples	4	6
No. deleted CGH data	1	1
No. of samples specifically concordant with a-CGH data (In only heterozygote samples)	28	14
% agreement between both techniques	75.7%	63.6%
% disagreement between both techniques	24.3%	36.4%
Free-marginal kappa (range -1 to 1)	0.53	0.47

3.3.4 Correlation of allelic imbalance with histopathological subtype

We went on to assess the proportion of samples showing significant allelic imbalance from node positive and ECS positive patients to look for any association with pathological stage. The a-CGH study had previously found that TP53 and CCND1 allelic imbalance were associated with advanced nodal status but not ECS (see section 3.1.3). The number of ECS+ve samples showing allelic imbalance was firstly identified. The total number of heterozygote samples for TP53 were 22 where only 2 samples showed amplification and 6 samples showed deletion. Out of these 6 samples, 4 samples were

node positive ECS-ve samples and only 2 ECS+ve samples showed TP53 deletion. The majority of heterozygote tumour samples showed no allelic changes for TP53 gene. On the other hand, total numbers of heterozygote samples for CyclinD1 were 37, where 11 samples showed amplification and only 4 samples showed deletion. The majority of the samples showing amplification were ECS+ve samples (8/11, 73%). Table 15. Illustrates the precise number of heterozygote tumour samples according to pathological data correlated with the q-PCR values. As there were no other previously reported studies of ECS and allelic imbalance we did not have reference levels that this could be compared to, however it was in keeping with our previous a-CGH study (see section 3.1.3). To specifically assess whether these findings had any significant association with ECS, statistical analysis was performed using Fisher's exact test, a statistical test performed for contingency tables and appropriate for small sample sizes. ECS positive samples were not found to be statistically significantly correlated with TP53 deletion or CCND1 amplification, whereas, TP53 deletion and CCND1 amplification were found to be correlated with nodal status in this data set (TP53 deletion $p = 0.040$ and CCND1 amplification $p = 0.050$)

In conclusion, both SNP analysis and the previous a-CGH study have similar findings when correlated with clinical data, with our study finding only an association of TP53 deletion and CCND1 amplification with advanced nodal status but not ECS. The fact that both methods come to similar conclusions with regards to association with one marker of aggressive phenotype (nodal status). The findings indicate **our cohort may harbour a novel, genetically distinct subset of cases** but did not identify a potential driver of ECS. As a result a further detailed molecular characterisation of ECS, possibly through mutational analysis, was suggested.

Table 15. Overall number of OSCC samples demonstrating TP53 or CCND1 allelic imbalance detected by q-PCR in relation to Nodal status and ECS.

		Total numbers				p value (Fisher's exact test)	
		N-ve	N+ve ECS-ve	N+ve ECS+	All samples	Nodal status	ECS
Number of samples (%)		19	7	20	46		
TP53	Total heterozygote	8	7	7	22		
	Amplification	0 (0)	1 (14)	1 (14)	2 (9)		
	Deletion	**0 (0)**	**4 (57)**	**2 (29)**	6 (27)	0.040	0.651
	Unchanged	8 (100)	2 (29)	4 (57)	14 (64)		
CCND1	Total heterozygote	14	5	18	37		
	Amplification	**1**	**2**	**8**	11	0.050	0.060
	Deletion	2	1	1	4		
	Unchanged	11	2	9	22		

3.4 Discussion

3.4.1 The identification of allelic imbalance by q-PCR SNP analysis

SNP analysis by q-PCR is a technique that has some distinct advantages in that it does not require labour-intensive DNA processing. It can label probes with two or more different fluorophores allowing for allelic qualitative assay in a single tube reaction and it also enables the detection of allelic imbalance in any double-stranded DNA sequence. Our SNP assay optimization experiments showed in a detailed and robust manner that the q-PCR method employed, could detect allelic imbalance when average values of $\Delta\Delta$Ct detected by the technique had certain limits of detection thresholds applied (≥3 for CCND1 or ≥2 for TP53 amplification or ≤0.5 for CCND1 and ≤0.25 for TP53 deletion). However, the method was not sufficiently sensitive to allow confident quantitative analysis.

Although the method has been reported previously in the literature as being able to sensitively quantify allelic imbalance [144], the poor performance in quantifying actual fold differences in amplification and deletion observed here is most likely due to artefact produced from nonspecific probe binding as also previously described [145]. Thus, variability may arise due to different test amplification properties and reference loci, as well as the differential yield sensitivity of each amplification reaction to the precise conditions [146]. The sensitivity of the assay could also have varied because of the amplification of normal cell DNA mixed with the tumour sample. When analysing heterogeneous tumour tissue, unjustified imbalance calling may occur. As the SNP resolution is relatively low, an inconclusive result may be obtained when a substantial number of the employed SNPs are homozygous and as a result non-informative. However, this issue could be overcome by increasing the number of SNPs to be analysed.

Although the development of genome-wide CNV screening has been used and applied to assess allelic imbalance in OSCC samples, including array-CGH and SNP arrays (see Table 10), However, q-PCR is a frequent computationally identified loci validation method [147]. q-PCR has been used in large-scale CNV analysis in detecting disease

associations in OSCC [148]. Thus, findings previously from the literature in combination with our own optimisation and understanding of the limitations of the q-PCR method still justified the further use of the method for the assessment of CCND1 and TP53 allelic imbalance in relation to ECS.

3.4.2 Allelic imbalance assessment for CCND1 and TP53 in association with ECS

The overall rate of allelic imbalance by the SNP assay was found to be 41% for CCND1 and 36% for TP53 in heterozygote samples. This compares to previous reports of the over-expression rate of CCND1 being 25-70% in patients with oral cancers [138] but for TP53 around 50% [149]. One explanation for this along with the overall lack of association observed with ECS in this study could be explained by the reduced sample size from exclusion of homozygote samples. Indeed, only 22 samples showed heterozygosity for TP53 gene whereas 37 samples showed heterozygosity for CCND1 gene. This has previously been reported as a drawback leading to inconclusive findings in other cancer studies [150].

The SNP assay in both instances found greater numbers of deletions than a-CGH (CCND1: 4 in SNP assay /1 in a-CGH, and for TP53: 6 in SNP assay /1 in a-CGH) that contribute to the slight difference in overall findings in between techniques. Indeed previous studies have shown that a-CGH has the disadvantage of being less sensitive in detecting deletions than PCR based methods [141]. Specifically, in the case of TP53, optimization experiments for the TP53 probe used in the quantitative PCR had already revealed slight over-estimation, which might also explain the difference in the number of samples deleted in the q-PCR data compared to the a-CGH data. Along with this, the sensitivity of both analyses can be hampered by admixture of tumour with normal stromal cells as well as the consideration that tumours can be composed of sub clones affecting analysis [151]. The sensitivity of a-CGH depends on visual interpretation and although digital image analysis can be applied to improve sensitivity, large copy differences are still more readily detected than smaller ones [115]. The difference observed in concordance between both techniques (free marginal kappa statistic of 0.47

for TP53 and 0.53 for CyclinD1) could also be explained by the q-PCR method suffering from nonspecific probe binding that may lead to both over and under estimation.

In total in our analysis 28.6% node positive ECS+ve samples showed deletion for TP53 gene and 44.4% ECS+ve samples showed amplification for CCND1 gene. After Fisher's exact test was applied, p values correlated with TP53 deletion and CCND1 amplification with ECS positive samples did not reach any statistical significance. However, TP53 deletion and CCND1 amplification were correlated with nodal status (TP53 deletion p=0.040, CCND1 amplification p=0.050). Future work increasing the total number of samples in different groups especially with regard to ECS cases (20 cases in this series) needs to be performed before stronger conclusions can be drawn especially in relation to the aggressive phenotype of OSCC, ECS. Furthermore, due to the drawbacks of both the a-CGH and SNP assay approaches, a more detailed quantitative assessment with re-validation of the findings obtained by this part of the study potentially using Multiplex ligation-dependent probe amplification (MLPA) techniques. MLPA is a multiplex method that can detect abnormal copy numbers [152]. Using MLPA for copy number detection offers some advantages over CGH and FISH techniques; because it can identify the frequent, single gene aberrations, it can be used on purified DNA and compared to array CGH, MLPA is a cost effective and technically uncomplicated method [115].

3.4.3 Future directions arising from allelic imbalance analysis for ECS biomarker discovery

There are several other issues that arise from this part of the study. Firstly, copy number changes may not necessarily be reflected in expression so further studies clarifying ccnd1 and p53 protein levels may prove useful in interpreting the data. Nevertheless, copy number alterations have been published in the literature to be associated with a variety of different cancers including head and neck cancer. In a high profile study of 279 HNSCC cases, 39 regions of recurrent copy number loss and 23 regions of recurrent copy number gain were identified [21]. After the start of this study there have been papers demonstrating that Cyclin D1 amplification [135, 153] and TP53 deletion [135] are related to OSCC cases. Amplification of CCND1 gene has been

reported in 25-70% of oral cancers premalignant lesions suggesting that cyclin D1 gene amplification may be an early event during oral tumourigenesis [70]. Another study on 164 OSCC cases showed that Cyclin D1 allelic imbalance is an independent predictor for worse survival in OSCC in HPV negative cases and a promising biomarker for predicting lymph node metastasis in patients with clinically Stage I–II OSCC. [154].

The results from our samples found CCND1 amplification and TP53 deletion to be correlated with nodal status. This is consistent with another study where frequent gain was observed in CCND1, exclusively in OSCC patients with cervical lymph node metastasis [130]. Overall, the contradictory results could be because of micro metastases present in the samples, which are known to be potentially missed by pathologists examining a dissected specimen [18]. Our study did not find many changes in TP53 in our subset of OSCC cases. This finding contrasts with other genome wide, mutational and genome copy profiling data reported for all head & neck cancers, which predominantly report TP53 as a commonly mutated gene (around 50%) [102]. Loss of heterozygosity (LOH) was also observed for the TP53 gene in oral cancer series [155].

A potential explanation for this could be provided by the Cancer Genome Atlas project that assembled the genomic data of 3,299 tumour samples of 12 tumour type and a closer inspection of the distribution of data showed a striking inverse relationship between copy number alterations and somatic mutations placed at the extremes of genomic instability. The data showed that the tumours had either a large number of copy number alterations or a large number of somatic mutations, never both. They refer to this trend as the cancer genome hyperbola [137]. Keeping that trend in mind, lack of TP53 allelic imbalance leads to a possibility of finding mutations in our OSCC series. Although many studies have focused on allelic imbalance with relation to HNSCC and in particular OSCC, none of these studies have specifically examined any correlation with extra capsular spread of OSCC. This suggested a mutational analysis of ECS in OSCC, as a further progression of the present study, would be worthwhile with the overall aim of better molecular characterisation of this aggressive phenotype in OSCC.

Chapter 4 : Screening for genomic variants in primary OSCC cell lines

4.1 Introduction

Results in Chapter 3 suggested that although TP53 loss and CCND1 were associated with nodal status but not ECS, the low frequency of TP53 change may be associated with a previously unidentified subtype of OSCC. As a result, we initially decided to perform a mutational screen on new primary cell lines, of known histopathology, obtained through local collection.

4.1.1 Large scale genome sequencing in cancer

Greenman and co-workers undertook comprehensive Sanger Sequencing of 518 protein kinase-encoding genes in 210 different cancers including lung, breast, colorectal, gastric, testis, renal, ovarian, glioma, melanoma and acute lymphoblastic leukaemia for somatic mutation detection in coding exons and splice junctions, reporting more than 1,000 somatic mutations overall [156]. Most somatic mutations were deemed potential 'passengers', not contributing to oncogenesis, whereas 158 predicted 'driver' mutations were identified on the basis of a statistical model comparing the observed to-expected ratio of synonymous (no amino-acid change) mutations with that of nonsynonymous (altered amino acid) mutations. Most kinase mutations were 'single hits' [156]. Sjoblom and co-workers used a different strategy, but reached similar conclusions. Analysis of 13,023 genes in 11 colorectal and 11 breast cancers revealed that individual tumours gather an average of ~90 mutant genes with only a subset of these genes contributing to the neoplastic process. They identified 1,307 validated nucleotide changes in 1,149 genes, of which 189 were significantly mutated. A few overlapping driver mutations were identified in the kinase genes analyzed in these two studies. [157]. Further whole exome sequencing of a series of glioma, breast, colorectal and pancreatic cancers [158-160] had also suggested distinct 'driver' and 'passenger' mutations co-exist in cancers, with driver mutations being key to altering cellular pathways associated with the acquisition of aggressive phenotype. Such studies have become increasingly comprehensive as Next Generation methods of sequencing have been invoked [161-164].

4.1.2 Next generation sequencing applied to HNSCC

Like the cancer studies outlined in 4.1.1, whole exome sequencing has also been conducted for HNSCC. Approximately 100 HPV positive and negative HNSCC specimens (sites include oral cavity, oropharynx, hypopharynx, larynx, and sinonasal cavity) were analysed independently by two groups to understand the pathology of HNSCC [101, 102]. One study employed whole exome sequencing analysis (150-fold mean sequence coverage of targeted exonic regions) from 74 tumour-normal pairs and revealed that frequent mutation occurs in previously known TP53, CDKN2A, PTEN, PIK3CA, and HRAS genes. Along with this at least 30% of cases showed mutations in NOTCH1, IRF6, and TP63 genes that regulate squamous differentiation implicating its deregulation as a major driver of HNSCC carcinogenesis. This analysis identified only 39 genes with high statistical significance. However, most mutated genes did not reach any statistical significance which suggests these could be passenger events. The most commonly mutated gene in HNSCC was TP53, whereas around 11% of the HNSCC tumours showed NOTCH1 point mutations. In addition, non-synonymous point mutations in NOTCH2 or NOTCH3 were also found in 11% of the samples. Recurrent mutations were observed in less well characterized genes like SYNE1 and SYNE2 (20% and 8% of HNSCC samples respectively) along with two apoptosis related genes CASP8 (8%) and DDX3X (4%) [102].

Myers et al. had broadly similar findings employing whole exome sequencing analysis to study 32 primary tumours and identified 911 candidate somatic mutations in 725 genes. They found recurrent mutations in RIMBP2, PIK3AP1, SI, NRXN2, EPHA7, NRXN3, RASA1, HRAS, RXFP3, PIK3CA, TP53, NOTCH1, CDKN2A and FBXW7. They further analyzed the sequences of these genes in additional HNSCC cases where somatic mutations in TP53, NOTCH1, CDKN2A, PIK3CA, FBXW7 and HRAS were identified in 47%, 15%, 9%, 6%, 5% and 4% of patients, respectively. The remaining genes were not observed to be mutated in more than one of the additional samples analyzed. They also found TP53 and NOTCH1 were the most frequently mutated genes in HNSCC [101]. If both studies are compared, there was a five-fold difference in the average

number of mutations reported per tumour although the two studies analyzed etiologically similar cases using related sequencing platforms. One explanation for this diversity could be the heterogenous nature of HNSCC or site-specific differences, risk-factor exposure related differences or indeed differences in aggressive phenotype amongst the samples used with limited consideration of detailed histopathological (such as ECS) and long term clinical correlates.

Another study highlighted new insights into HNSCC from studying 279 primary tumours from HNSCC patients [21]. Whole-exome sequencing identified somatically mutated genes which are mainly located in regions of Copy number alterations. 37,061 non-synonymous somatic variants were detected with 11 statistically enriched genes. TP53 mutation among HPV negative samples was found at a higher rate (86%) than the previously reported papers. Among inactivating mutations, four genes were found predominantly in HPV negative tumours. Two of them are CDKN2A and TP53 associated with cell cycle and survival and the other two were FAT1 and AJUBA, linked to Wnt/b-catenin signalling. Apart from this, a frequently mutated novel gene, the nuclear receptor binding SET domain protein 1 (NSD1), was identified in 33 HNSCCs. Significant inactivating mutations were found in genes linked to squamous differentiation including in NOTCH1 (19%), consistent with other reports. Other heavily mutated genes were NOTCH2 (9%), and NOTCH3 (5%), and the TP63 target gene ZNF750 (4%). This study identified additional mutations in TRAF3, RB1 and NFE2L genes. Of the known oncogenes, only PIK3CA achieved statistical significance. These recent studies predict that the inactivation of AJUBA, as well as FAT1 and NOTCH1, may converge to uncheck Wnt/b-catenin signalling, implicated in deregulation of cell polarity and differentiation, then leading to accelerated progression in HNSCC [21].

These studies used fresh, primary tumour tissues. A complimentary follow-up study used DNA samples from 252 formalin fixed paraffin-embedded (FFPE) HNSCC tumours and gave similar results. In both HPV positive and negative HNSCCs, mutations were more commonly identified in tumour suppressor genes compared to oncogenes. The

most common altered pathway by mutation was PIK3CA (30%) followed by PTEN (15%), AKT1 (5%), RICTOR (4%), mTOR (2%), AKT2 (2%), and PIK3R1 (2%) [80].

The studies differed in a number of ways. Stransky et al. performed whole exome-sequencing of 74 tumour/normal pairs, with 150-fold mean coverage, while Agrawal et al. only sequenced 32 tumour/ normal pairs, with a mean coverage of 77-fold and followed up with targeted follow-up sequencing of frequently mutated genes in an additional 88 patients. For HPV-negative tumours, Stransky et al. found a mutation rate of 4.83 per Mb while Agrawal et al found a mutation rate of 20.6 per Mb. The large discrepancy between the two groups may have arisen from differences in bioinformatics techniques for mutation calling or differences in coverage of sequencing. Since these initial studies, two other groups have determined mutations specifically in OSCC [136, 165].

The main observations from these studies are summarized in Table 16. Two of the studies required additional sequencing because of mean coverage falling below 100. Each study had slightly different patient cohorts and characteristics and different rates of significantly mutated genes was observed between studies. Use of formal statistical testing also varied. Nevertheless, selecting the most frequently altered genes based on high mutational frequency or relation to aggressive phenotype allows useful comparisons.TP53, CDKN2A, CASP8, FAT1, NOTCH1, HRAS, PIK3CA, MLL2 and FBXW7 were most frequently altered. All studies found TP53 to be the most mutated gene, the first two studies found NOTCH1 to be the second most mutated but for the more recent studies this was FAT1. FAT1 was not found to be mutated at all by Agrawal et al. Pickering et al. had the lowest rate of HPV positivity and found the lowest rate of NOTCH1 mutation at 9%. The India ICGC cohort had the highest proportion of smokers in their cohort and found significantly greater rates of CASP8 mutation compared to the other studies. The comparison suggests that while TP53 may play a consistent central role in the pathogenesis of OSCC the other genes that may contribute to and drive disease progression with aggressive phenotype may be quite heterogeneous in nature and dependent on cohort-specific environmental as well as sequencing method factors.

Table 16. Summary of whole exome sequencing studies in HNSCC/OSCC. Sample characteristics and mutations observed by four next generation sequencing studies of HNSCC and OSCC. OC= oral cavity, OP= oropharynx, H= hypopharynx, L= larynx, SC= Sino nasal cavity.

	Stransky et al.	Agrawal et al.	Pickering et al.	India ICGC	TCGA
Total number of samples	74	32	40	50	279
Sequencing Platform	Illumina	Illumina & SOLiD	Illumina & SOLiD	Illumina	Illumina
Disease sites	OC, OP, H, L, SC	OC, OP, L, H	OC	OC	OC, OP, L
Smokers	89%	75%	78.6%	96%	79%
HPV positivity	14%	25%	2.5%	26%	13%
Mean coverage	150	77	119	38	95
Whole exome	72	32	40	50	279
Whole genome	2	0	0	0	29
Additional tumours	NA	88	NA	60	NA
Significantly mutated genes	39	6	53	10	19
TP53	62%	47%	66%	62%	72%
CDKN2A	12%	9.2%	0%	2%	22%
CASP8	8%	3%	10%	34%	9%
FAT1	12%	0%	28%	40%	23%
NOTCH1	14%	15%	9%	16%	19%
HRAS	5%	4%	9%	12%	4%
PIK3CA	8%	6%	11%	6%	21%
MLL2	11%	0%	9.5%	10%	0%
FBXW7	5.4%	5%	4.7%	2%	5%

Although these studies give valuable genomic insights into OSCC, there is no study reported so far screening for mutations in OSCC that may be related to ECS. In addition, they were performed using either fresh OSCC primary tumour samples or FFPE fixed samples, with no possibility of follow up studies in matching OSCC cell lines.

Few studies have characterised OSCC cell lines by focusing on candidate genes. Soichi et al. identified mutations in three out of four OSCC cell lines for APC and Axin genes, both genes regulating the Wnt pathway and implicating its role in OSCC [166]. Kozaki et al. identified PIK3CA missense mutations in 21.4% (3/14) of OSCC primary cell lines derived from patients at different stages (I- IV). The highest frequency of PIK3CA mutations was observed in stage IV of the disease with a statistically significant association between the frequency of mutation and stage ($P = 0.042$, Fisher's exact test) suggesting a role for the PI3K-AKT signaling pathway in OSCC progression [167]. The same pathways were implicated by the Exome sequencing studies in OSCC tumour samples as previously discussed supporting the value of using primary OSCC cell lines as being representative for understanding the pathogenesis of OSCC and significance regarding ECS.

4.1.3 Primary cell lines for candidate gene selection in OSCC

In our lab, as a parallel project, five primary keratinocyte cell lines had been cultured from fresh oral cancer tissue samples having known histopathology. As a preliminary screen for genetic variants that may be associated with aggressive behaviour, low pass whole genome SOLiD sequencing was performed on these cell lines. The cell lines used and their corresponding nodal status are listed in Table 17. Cell lines from node positive patients but without ECS were not available at this time, reducing the possible types of comparison.

Table 17. Primary OSCC cell lines used for Low Pass WGS.

Cell line	Phenotype
Liv4K	Node negative
Liv22K	Node negative
Liv7K	Node positive ECS+
Liv12K	Node positive ECS+
Liv26K	Node positive ECS+

4.2 Hypothesis and aim

Hypothesis: ***Specific genetic alterations are associated with ECS in OSCC.***

Aim: Compare primary OSCC cell lines by whole genome sequencing utilising the SOLiD platform to identify gene variants that may have a relation to ECS.

4.3 Results

4.3.1.1 SOLiD sequencing data acquisition

SOLiD Fragment DNA library preparation was performed initially for a DNA sample from the human random control panel and then for each of the cell lines listed in Table 17 using the standard protocol described in section 2.7.4. Successful ligation of P1 & P2 adaptors to DNA fragments was confirmed by using agarose gel (1.2%) electrophoresis to resolving products. Figure 10. shows addition of adaptors leading to a shift (a~30 bp) in the average size of sheared DNA fragments (lane 7) compared to sheared DNA without adaptor ligation (lane 5). Purified ligated products were then separated using a 2% agarose Library Size Selection Egel to select fragments of 200-230bp for further nick translation and amplification by PCR. Successful final library preparation was confirmed by further agarose gel (1.2%) analysis shown in Figure 11. This shows nick translated & amplified libraries of different fragment sizes 200-230bp (lanes 3, 4 and 5) compared to a control with no nick translation and absent product (lane 2). Once library optimisation was completed using human random control DNA, the same methods (section 2.4) were applied to DNA extracted from each cell line described in Section 4.1.3.

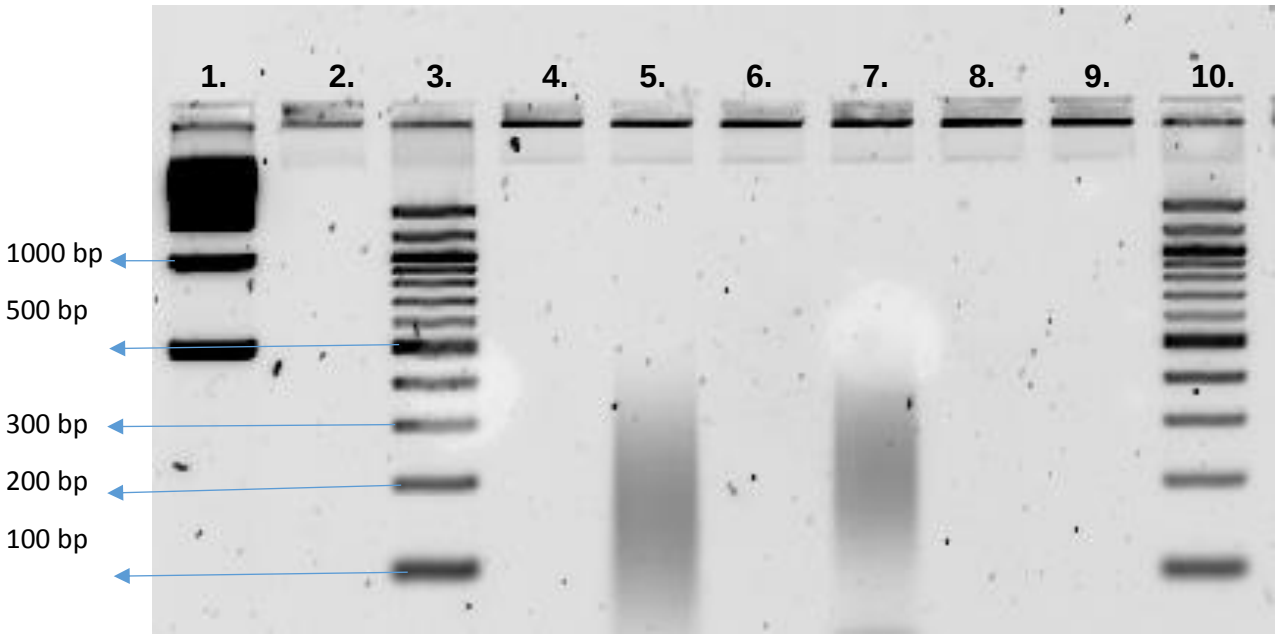

Figure 10. Agarose gel showing products of P1 and P2 Adaptors ligated to Human Control DNA. ***Lane key:*** *1=1 Kb DNA Ladder, 3,10=100bp DNA Ladder, 5=Sheared Control DNA, 7= Sheared Control DNA + Adaptor + Ligase, 8=Negative control*

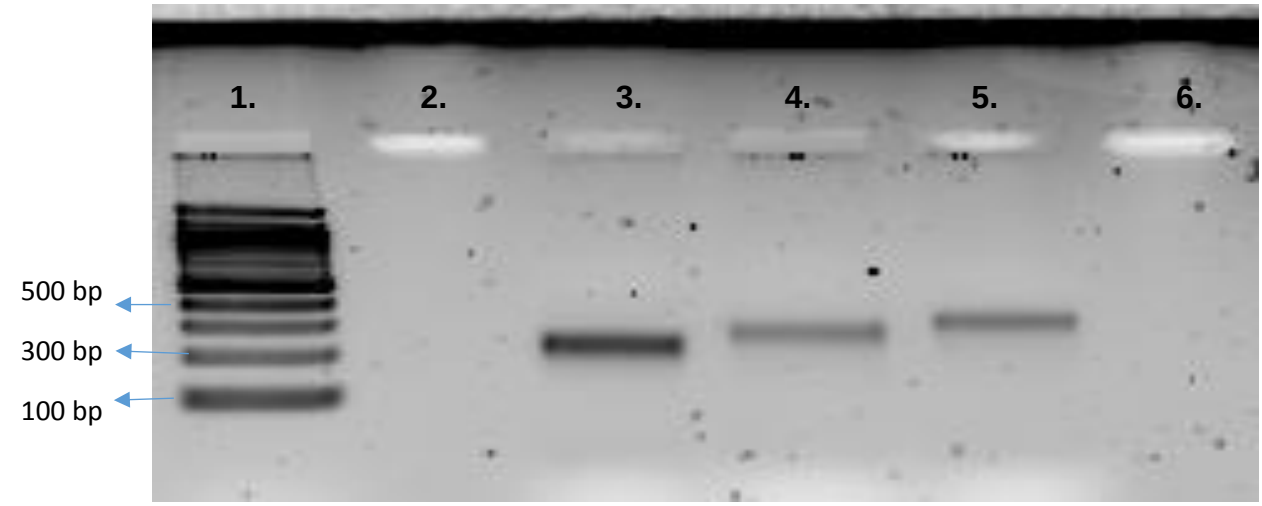

Figure 11. Agarose gel of nick translated and amplified final Human Control DNA SOLiD libraries. ***Lane key:*** *1=100bp DNA Ladder, 2=No nick translated DNA 3,4,5= Nick translated and amplified libraries from 200-230bp in size, 6=Negative control*

4.3.1.2 Quality control of OSCC cell line libraries

The OSCC cell line fragment libraries were analysed for size distribution and quality using an Agilent 2100 Bioanalyser, see Figure 12., prior to being pair-end sequenced using the SOLiD™ sequencer in the Centre for Genome Research, Liverpool. The average sizes of the fragments were 300bp as required for SOLiD sequencing.

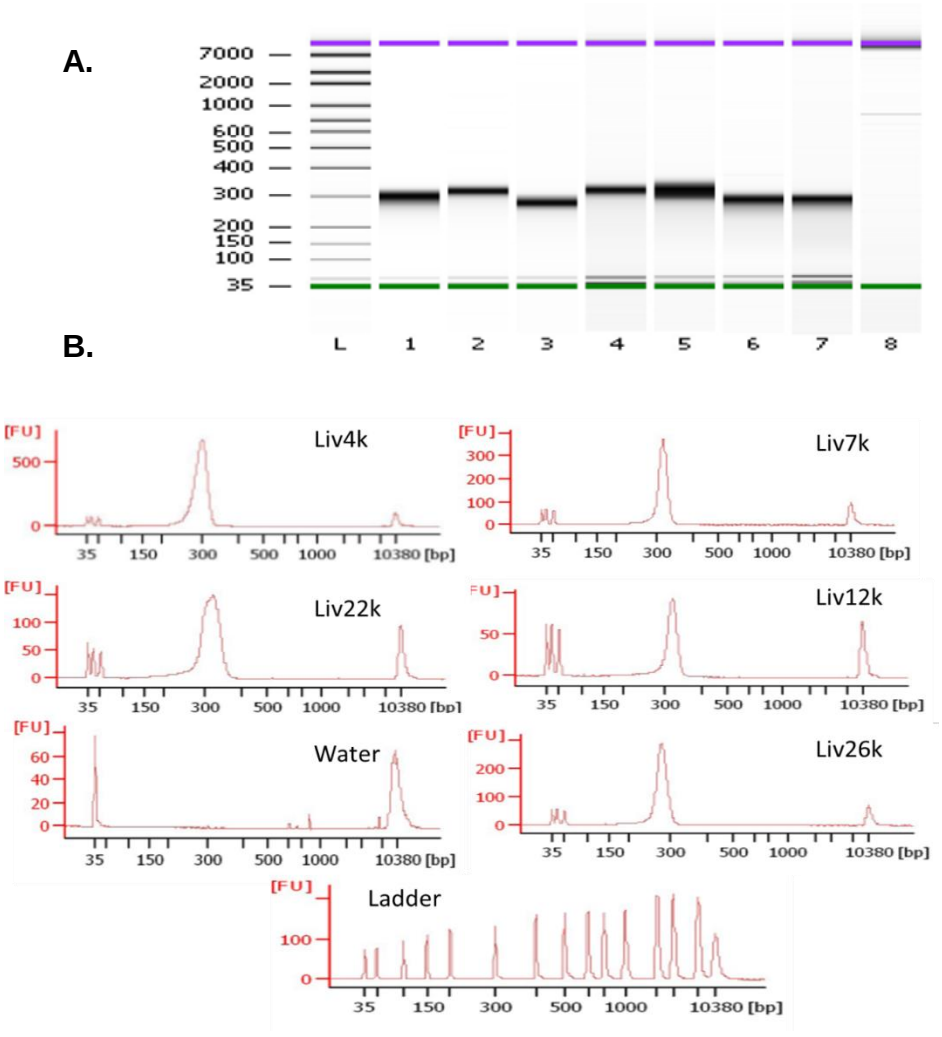

Figure 12. Size distribution of primary OSCC cell line SOLiD libraries A. Bioagilent electrophoresis gel image of fragment size of libraries generated from different cell lines (L=ladder, lane 1=Liv4K, lane 2=Liv7K, lane 3=Liv12K, lane 4=Liv22K, lane 5=Liv26K) compared to positive control DNA libraries (lanes 6 & 7) and water as a negative control (lane 8). **B.** Bioagilent fluorescence (FU) traces of 5 cell line libraries compared to water and ladder.

4.3.2.1 Overall Coverage generated from OSCC cell lines

Sequences from the libraries were aligned to human genome (HG19) and initial data analysis was performed within the Centre for Genomic Research (CGR) using BWA (Burrows Wheeler Aligner) as described in section 2.8.3. Table 18., shows the sequence summary for each OSCC sample. The total number of reads ranged from 811 x 10^6 - 962 x 10^6 and percentage of mapped data against HG19 reference data ranged from 89.39%-92.86%, indicating that the data was suitable for further analysis. However, the mean overall coverage for each sample was low (<10) in unique sequences with regions showing 30× average depth ranging from 57.21%-72.63% across samples. This suggested that while some conclusions could be confidently made from the data, it would be important to assess individual variant coverage before coming to firm conclusions.

Table 18. HG19 coverage by sequence data from SOLiD sequencing of OSCC cell line samples.

Sample	Total Reads	Mapped	HG19 covered %	Mean Cov (sampling every 1k bases)	% sample with 30× average depth	No.of variants	No. of INDELs
Liv4K	861,811,594	78.48%	89.84%	8.55	66.74%	1,833,275	211,272
Liv7K	924,316,936	79.15%	89.10%	8.14	62.98%	1,805,627	48,421
Liv12K	854,142,556	77.06%	92.86%	8.39	65.43%	1,792,593	212,130
Liv22K	962,374,430	78.27%	90.05%	9.6	72.63%	1,979,526	244,099
Liv26K	810,808,638	64.60%	89.39%	6.61	57.21%	972,840	98,281

4.3.2.2 Analysis of Cross contamination between OSCC samples

To check for cross contamination between samples, specific gene variant profiles were considered using haplotype data. The large gene TTN was selected for comparison between samples. Although there were a number of common variants observed across all the different OSCC cell lines, the majority of variants identified were clearly sample specific variants. Furthermore, statistical analysis using a Chi-square test comparing two different outcome groups (numbers of homo and heterozygote variants) for each cell line combination in turn demonstrated significant differences ($p<0.05$) between all cell line comparisons. This indicated each cell line has their own unique genetic landscape of germ line variants. Thus, we concluded that although there was low coverage in certain regions, cross-contamination was not present and regions with satisfactory coverage around 30 for variant calling, based on the established literature [168, 169], could be used to screen for variants in support of further analysis.

4.3.2.3 Identification of novel INDELs

We therefore performed VEP (variant effect predictor) analysis on the SOLiD sequencing data. Only variants with coverage 30× average depth were considered for further analysis. Table 19. summarises the VEP analysis for each cell line. Of variants with potentially deleterious effects (stop gain/loss, start loss, missense variants and coding sequence variants), node -ve cell lines displayed marginally a greater number (12,582 and 13,525) than ECS+ve cell lines (11,621; 12,025 & 7,243), but small sample size precluded any further interpretation. Figure 13. summarises the total for each cell line with a breakdown of variants between samples according to ECS status.

The complete list is tabulated in the supplementary CD (supplementary Table 1). 3313 INDELs with severe consequences were identified. Furthermore, the affected genes harbouring these INDELs were then compared to produce a list of 149 genes affected in one or more cell lines. These genes were then arranged in a binary chart to highlight the distribution across ECS +ve and -ve cell lines (Figure 14). No novel INDELs were

identified for CCND1 and TP53 consistent with the assessment of allelic imbalance in the previous chapter 3. No INDELs identified were present in both ECS -ve cell lines, however 6 genes were identified with INDELs in two or more ECS +ve cell lines; CDK11A (present in all three ECS+ve cell lines tested), FRG1B, TRIT1, MUC17, SLC16A7 and ZNF720. If these genes are considered; FRG1B is a pseudogene [170], TRIT1 modifies tRNA [171], MUC17 codes for a membrane bound mucin [172], SLC16A7 belongs to the monocarboxylate transporter family [173] and ZNF720 codes for a zing finger protein involved in nucleic acid binding [174].

Table 19. Types of Variants identified by VEP analysis in Primary OSCC cell lines.

Variants with potentially deleterious consequences are highlighted in red

	Liv4k	Liv7k	Liv12k	Liv22k	Liv26k	Total
Splice acceptor variant	105	203	209	203	59	779
Splice donor variant	183	118	114	127	118	660
Stop gained	103	114	104	115	49	485
Stop lost	42	31	45	36	22	176
Start lost	40	17	39	39	17	152
Missense variant	12393	11458	11835	13314	7154	56154
Splice region variant	5673	5367	5664	6218	3292	26214
Incomplete terminal codon variant	2	23	1	1	3	30
Synonymous variant	14195	13694	14235	15930	8350	66404
Coding sequence variant	4	1	2	21	1	29
Mature miRNA variant	14	14	14	3	6	51
5 prime UTR variant	4909	4618	4782	5719	2874	22902
3 prime UTR variant	28840	27336	27289	31099	16112	130676
Noncoding transcript exon variant	37688	35788	36274	41896	21663	173309
Intron variant	3495332	3405018	3428187	3765968	1934695	16029200
NMD transcript variant	279527	271680	274004	300717	158166	1284094
Non coding transcript variant	937624	914838	920450	1008214	519333	4300459
Upstream gene variant	441876	429659	433067	487939	248765	2041306
Downstream gene variant	454851	444310	446113	502712	258346	2106332
TF binding site variant	2462	2418	2415	2896	1524	11715
Regulatory region variant	233462	223469	228228	252415	135903	1073477
Intergenic variant	696777	682938	669123	732465	361960	3143263
Total	6646102	6473112	6502194	7168047	3678412	30467867

.

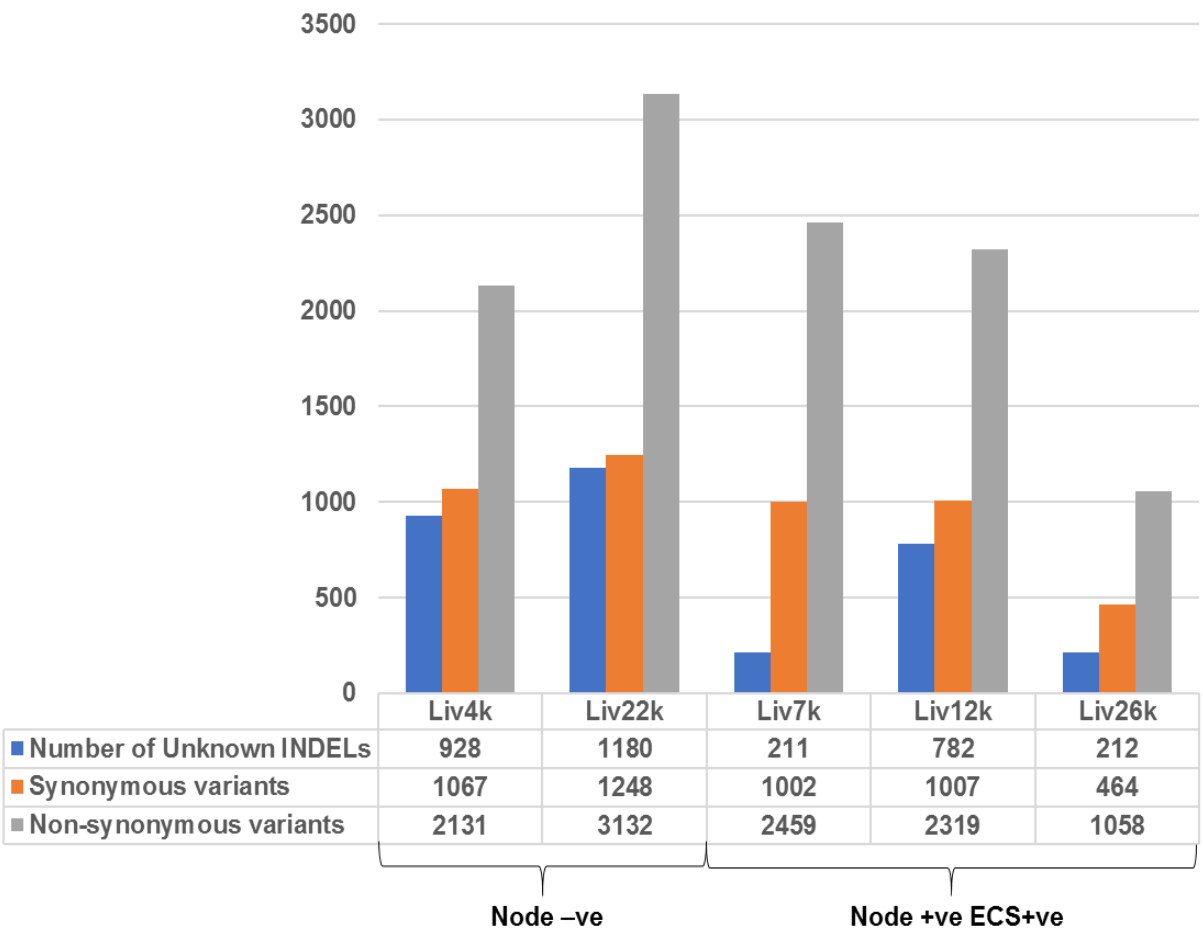

Figure 13. Summary of total numbers of variants and INDELs with potentially severe consequences on the basis of ECS status in each OSCC primary cell line. Cell lines are grouped according to ECS status.

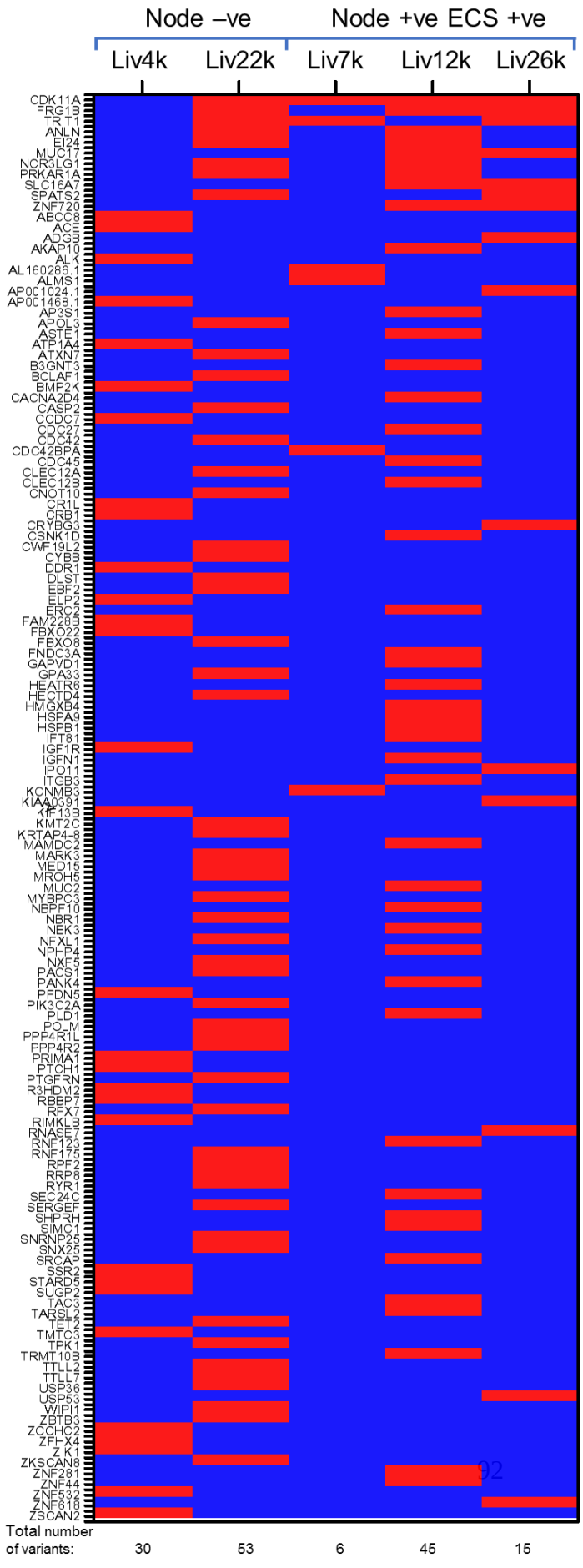

Figure 14. Summary of genes harbouring novel INDELs with potential severe consequences on the basis of ECS status in OSCC primary cell lines. The Binary chart demonstrates the presence (red) or absence (blue) of gene INDELs within each OSCC cell line distributed on the basis of ECS status. Genes were sorted on the basis of number of cell lines harbouring an INDEL with a greater number placed superiorly within the chart.

Of particular interest is CDK11A, found in all 3 ECS +ve cell lines, coding for a cyclin dependant kinase having multiple roles in cell cycle progression, cytokinesis and apoptosis [175]. Although the function of CDK11a in OSCC remains unclear, translocation or deletion of CDK11 has been reported in neuroblastoma [176].

4.3.2.4 Identification of Novel single nucleotide variants

Novel synonymous/non-synonymous variants with 30× average depth for all the primary OSCC cell lines were extracted from the VEP analysis (completely listed in the supplementary CD Table 2 and 3). A total of 4,788 Synonymous variants and 11,099 Non-synonymous variants were identified with the distribution across all OSCC cell lines according to ECS status summarised in Figure 13. The genes affected by non-synonymous variants with potentially severe consequences (defined as missense, stop gain or loss) were identified and analysed across all of the OSCC samples. In total 448 genes displayed a potentially deleterious variant in one or more cell line, however to further focus on key gene variants we produced a list of 27 genes with variants in two or more cell lines from which we compiled a binary chart to assess the distribution on the basis of ECS status (Figure 15). CDC27 stood out as having variants in all 3 ECS +ve cell lines as well as one Node -ve cell line. CDC27 is an anaphase-promoting complex (APC) component that as a result may influence the timing of cellular mitosis [177]. Overexpression of CDC27 has been observed in colorectal cancer and is significantly correlated with tumour progression and poor patient survival [178], however no association has been found with head and neck cancer to date.

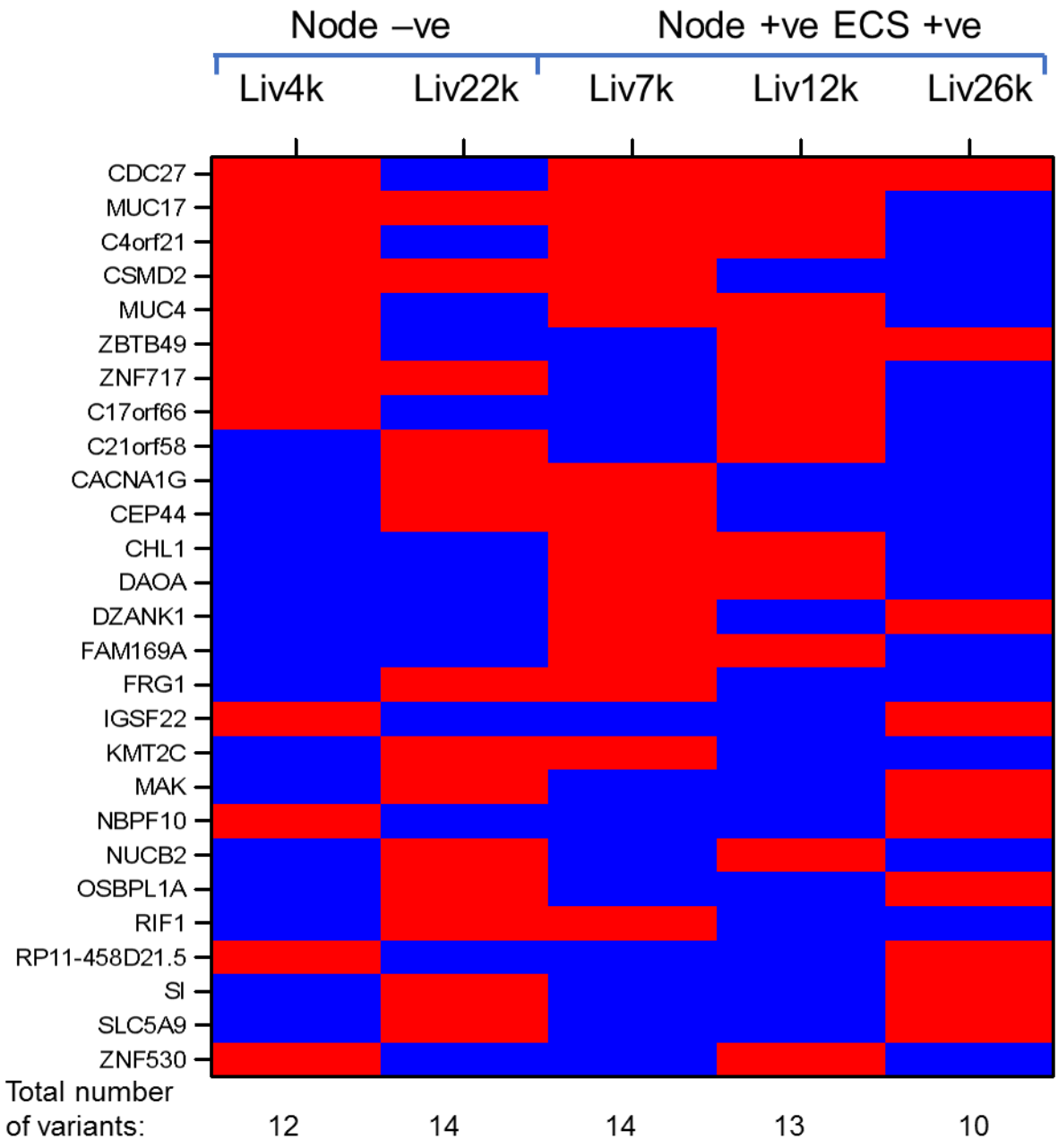

Figure 15. Genes affected by novel non-synonymous variants with potentially severe consequences in more than one OSCC primary cell line. 27 genes represented in this binary chart were identified from the VEP analysis to have missense, stop gain or loss variants identified by WGS. Cell lines are arranged according to ECS status and genes have been sorted to demonstrate those with most gene variations superiorly. Presence of gene variant shown by red square, absence by blue.

4.3.2.5 Pathway analysis for genes having potentially severe functional consequences in primary OSCC cell lines

All variants with potentially severe consequences from more than one cell lines were identified to produce a list of 176 genes (summarised in supplementary CD Table 4). This corresponding Refseq list of genes was used for pathway analysis using REACTOME software. 849 pathways were found to be effected (Supplementary CD Table 5). Of these, 14 pathways were significant as determined by p values <0.05, listed in Table 20. The majority of pathways effected concerned signal transduction. Notably, the only other disease pathway affected was Notch pathway signaling within these OSCC samples with p values of 0.040 as a result of an FBXW7 potentially deleterious variant identified in an ECS+ve (Liv26k) cell line (Table 21.).

Table 20. Pathways significantly affected by potentially deleterious gene variants identified by SOLiD sequencing of primary OSCC cell lines. 14 signaling pathways were identified by Reactome to be significantly effected (p values <0.05) in our primary OSCC cell lines.

Pathway Name	p value
Calmodulin induced events	0.005425621
CaM pathway	0.005425621
Ca-dependent events	0.010781121
Cam-PDE 1 activation	0.012481681
ABC-family proteins mediated transport	0.018570859
DAG and IP3 signaling	0.019213819
Downregulation of ERBB4 signaling	0.022366303
PLC-gamma1 signaling	0.023612493
EGFR interacts with phospholipase C-gamma	0.023612493
PLCG1 events in ERBB2 signaling	0.026052535
PLC beta mediated events	0.036486897
G-protein mediated events	0.039389011
FBXW7 Mutants and NOTCH1 in Cancer	0.040063907
Loss of Function of FBXW7 in Cancer and NOTCH1 Signaling	0.040063907

Table 21. An ECS+ve FBXW7 potentially deleterious single nucleotide variant. A missense FBXW7 variant found in a node positive ECS positive cell line (Liv 26K)

Cell line	Variant	Coverage	Consequence	Gene
Liv26K	chr4: _153273729_A/G	28.9	Missense variant	FBXW7

4.3.2.6 Comparison with other sequencing studies

We then compared our variant analysis observations to the established literature, particularly the next generation sequencing studies discussed earlier and summarised in Table 22 [21, 101, 102, 136, 165]. On specific comparison of our data with the list of significantly mutated genes produced from these studies we observed many similarities with 7/9 significantly mutated genes also found to have variants in our study. The gene with the highest variant frequency was CASP8, found to have variants in three cell lines including two ECS+ve cell lines. This was then followed by TP53 and CDKN2A.

Table 22. Presence of significantly mutated genes in OSCC from the literature in our WGS data. Variant frequencies of nine key genes are shown for all OSCC cell lines along with their ECS status.

Gene	Stransky et al.	Agrawal et al.	Pickering et al.	India ICGC	TCGA	Number of cell lines with variant	Phenotype of cell line
TP53	62%	47%	66%	62%	72%	2/5	N-ve, ECS +ve
CDKN2A	12%	9.20%	0%	2%	22%	2/5	N-ve, ECS +ve
CASP8	8%	3%	10%	34%	9%	3/5	N-ve, 2 x ECS+ve
FAT1	12%	0%	28%	40%	23%	1/5	ECS +ve
NOTCH1	14%	15%	9%	16%	19%	0/5	-
HRAS	5%	4%	9%	12%	4%	1/5	N-ve
PIK3CA	8%	6%	11%	6%	21%	1/5	N-ve
MLL2	11%	0%	9.50%	10%	0%	0/5	-
FBXW7	5.40%	5%	4.70%	2%	5%	1/5	ECS+ve

The missense variant in FBXW7 within an ECS+ve cell line (summarized earlier in table 21), had contributed to our finding on REACTOME analysis that the Notch pathway was a key disease pathway altered within our cell lines. However, no NOTCH1 or MLL2 variants were identified in our sequencing data. Unlike MLL2, with variants absent in two sequencing studies [21, 101], NOTCH1 had significant variants identified by all sequencing studies. Furthermore, considering our overall aim was to assess associations with ECS and aggressive phenotype, NOTCH1 (responsible for cell fate determination, proliferation, differentiation and apoptosis) [179] was theoretically a stronger candidate to pursue for association than CASP8 (primarily responsible for apoptosis) [180] albeit with an absence of variants in our cell lines. Thus, in the light of the importance of Notch signalling highlighted by our pathway analysis, the suggested importance of the NOTCH pathway from the established literature and the absence of NOTCH1 variants within our dataset, we considered whether a NOTCH1 deletion could be present in all cell lines or whether there was a technical reason for absence of variants.

4.3.2.7 Specific NOTCH1 coverage analysis of OSCC Cell lines

A more detailed targeted coverage analysis for NOTCH1 was performed. Reduced coverage (5× average depth) of NOTCH1 in comparison to the other genes was observed suggesting a possible NOTCH1 deletion.

Table 23. Mean Coverage in our series for genes reported to be significantly mutated in the Literature.

Gene	Mean Coverage
TP53	42
CDKN2A	8
CASP8	39
FAT1	32
NOTCH1	5
HRAS	51
PIK3CA	38
MLL2	36
FBXW7	46

However, similar observations across all cell lines raised the concern that this could be an artefact. Therefore, the data was sent for reanalysis by an independent centre (Lifescope). Low NOTCH1 (around 5× average depth) coverage was found in both analyses. We assessed whether this could be a gene specific observation related to the nature of whole genome sequencing rather than an observation specific to this study by comparing NOTCH1 coverage from a complementary parallel project in our lab using SOLiD sequencing of CLL (Chronic lymphocytic leukemia) as well as NOTCH1 coverage downloaded from the Genome 1000 database. Low NOTCH1 coverage evidenced by absence of reads on IGV visualization for both the CLL sample and genome 1000 downloaded data, similar to our OSCC cell line data. Thus, we can conclude that NOTCH1 coverage within the OSCC cell line dataset is low and attributable to the SOLiD platform used. As a result, we were unable to call variants for NOTCH1 gene based on this data.

4.4 Discussion

We set out to screen for genetic alterations associated with ECS in OSCC using next generation SOLiD sequencing and primary OSCC cell lines so as to produce a list of candidate genes that were associated with ECS, a marker of aggressive behaviour in OSCC. Production of SOLiD sequence libraries was successfully optimised as determined by the sizes of fragments generated from control DNA. The method was used subsequently for DNA from five of our primary OSCC cell lines (2 N-, 3 ECS+). SOLiD sequencing of these libraries was performed, Sequence coverage of HG19 was reasonable overall but uneven coverage in specific genes of interest was found, in particular NOTCH1, which we felt was a candidate for association with ECS on the basis of the reported literature and our WGS REACTOME pathway analysis. Although the uneven coverage precluded detailed structural rearrangement analysis, haplotype analysis did confirm an absence of cross-contamination and adequate coverage (30 × average depth) in the majority of regions (58%-73% range in samples) making the data amenable to further analysis. A variety of unknown non-synonymous variants and INDELs with potentially deleterious effects were detected by VEP analysis across all the

cell lines, with a novel FBXW7 variant identified in an ECS+ve cell line. Small numbers of samples analysed precluded statistical significance analysis of association of variants to histopathological subtypes but key directions with regards to further studies were gained.

4.4.1 Factors affecting sequence coverage by SOLiD

SOLiD next generation sequencing has been widely implemented for cancer genome profiling including HNSCC [101] and OSCC [136] and has been found to be comparable to other next generation platforms in terms of fidelity and accuracy [181]. Our observations revealed key regions of absent/low coverage that were similar using a different means of bioinformatics analysis (CGR v LifeScope), type of tumour sample (CLL) as well as an easily accessible reference database (Genome 1000). This suggested that these regions of poor coverage could have been as a result of technique specific drawbacks.

A range of factors could explain the poor coverage observed here. Biases in sample preparation, sequencing, and genomic alignment and assembly can result in regions of the genome that lack coverage and in regions with much higher coverage than theoretically expected [169]. Coverage with NGS has been described in the literature to be poor around areas with high GC content [169]. GC-rich regions, such as CpG islands, are particularly prone to low depth of coverage partly because these regions remain annealed during amplification [182]. Both NOTCH1 and CDKN2A, that are GC-rich, were found in our WGS data to have poor coverage, whereas genes with lower GC content such as TP53 and HRAS had acceptable coverage. Specifically, the NOTCH1 gene is known to have a high GC content (17 of 34 exons contain GC> 65% with exon 1 containing GC >75%). Studies have also shown that despite relatively high read depths across most exons, under representation of coverage in several exons (exon 1-26) or no coverage (exon1) was common [183]. Other types and causes of sequence errors can be diverse with SOLiD sequencing with the *in vitro* amplification steps potentially causing a higher background error rate [184]. Also, beads used in the emulsion PCR step of library amplification if they are carrying a mixture of sequences in close proximity

to one another can create false reads and low quality bases. Furthermore, signal decline, a small regular phasing effect, and incomplete dye removal can result in increasing error as the ligation cycles for sequence read out progress [185] and may have contributed to our poor sample coverage.

4.4.2 WGS identified candidate genomic variants and ECS

Our VEP variant analysis alongside REACTOME pathway analysis of our OSCC cell line data suggested that the Notch pathway could be potentially affected within our samples, being the only disease pathway significantly highlighted in analysis. The missense variant in FBXW7 in a node positive ECS positive cell line also further supported the potential importance of NOTCH pathway. FBXW7 is part of the F-box protein family and constitutes a component of the ubiquitin protein ligase complex [186]. It acts as a tumour suppressor in several tumours and a negative regulator of NOTCH1. Variants in FBXW7 gene can modulate the Notch pathway [187]. FBXW7 mutations have been previously observed in HNSCC samples as summarized in section 4.1.2, suggesting the involvement of the NOTCH pathway in head and neck cancer is likely. We found that the majority of pathways identified in our pathway analysis were signal transduction pathways and this may have great relevance for cancer development. Indeed, signal transduction pathways when considered broadly refer to processes by which regulatory molecules (e.g., extracellular hormones, growth factors, survival cytokines, and specialized proteins) govern the fundamental processes of cell growth, differentiation, and communicate within the cell resulting in cell proliferation. The ultimate end result of signal transduction networks is to regulate cell growth and division [188]. In terms of ECS however, compared to CDK11A (identified from INDEL analysis), CDC27 (identified from SNP variant analysis) and CASP8 (identified from key comparisons of our data with the literature), NOTCH1 remains a much stronger candidate for association based on other factors; the contemporaneous literature, our WGS analysis and the known effects of Notch signalling. Arguably we judged that no other gene than NOTCH1 was so highly consistently mutated in the literature and responsible for a multitude of cellular processes; cell maturation, proliferation,

differentiation and apoptosis, that could prove key to the development of an aggressive phenotype required for ECS [189].

In summary, although we were unable to prove our hypothesis that genetic variants are associated with ECS, pathway analysis of variants identified by whole genome sequencing (WGS) analysis of primary OSCC cell lines suggested Notch signalling could be affected in ECS+ subtypes, however consistent and unexpected deficits of sequence coverage in particular of the NOTCH1 gene, indicated that further investigation was required to be conclusive. Whether the deficit was for technical reasons or because the gene was preferentially lost was unclear (chapter 4, section 4.3.2.7). Comparison to other datasets suggested a technical under-representation but other affected genes like FBXW7 implicated the importance of Notch pathway.

Chapter 5 : Exon sequencing of NOTCH1 from primary OSCC cell line

5.1 Introduction

5.1.1 NOTCH signaling in relation to aggressive forms of cancer

In addition to the NOTCH alterations discussed in Section 4.1.2, the literature further suggests that NOTCH pathway mutations should be expected Upregulation of Notch1, Notch2, , HES1, HEY1 and Jagged1, all genes in the Notch pathway has been reported for OSCC cell lines [190]. Pickering et al. discovered changes in gene copy expression and number for other NOTCH genes including increased copy number in JAG1, JAG2 and in NUMB that were associated with elevated mRNA in HNSCC [136]. Our aCGH data also showed that copy number change in OSCC tumour samples enriched with ECS+ cases (section 2.2) significantly affected Notch signaling ($p = 9.7 \times 10^{-5}$) already discussed in section 3.1.3. ECS only occurs for 25% of OSCC cases, raising the additional possibility that alterations to Notch signaling could be responsible for this aggressive subtype. We therefore decided to further investigate the mutational status of the gene in our primary cell lines firstly to determine whether there was a genuine deficit in copy number or whether the deficit was more likely to be a technical artifact. This would also determine whether Notch mutations were present.

5.1.2 NOTCH1 structure and function

The NOTCH pathway is activated following ligand–receptor binding between two neighboring cells [191] (summarized in Figure 16.). In humans, four NOTCH receptors (NOTCH1–4) have been identified with five known NOTCH ligands from the Jagged family (JAG1 and 2) and Delta family (DLL1, 3 and 4).

The NOTCH heterodimer consists of an extracellular domain (ECD), a transmembrane domain and an intracellular domain (ICD) held together by non-covalent interactions that cause auto-inhibition of the protein [192]. In humans, the ECD is involved in Notch ligand interactions which contain 29 to 36 EGF-like domains [193]. In Notch1 the EGF11–13 region is also capable of binding to its' ligand in a Ca^{2+} dependant manner [194]. Following the EGF-like domains there are three cysteine-rich Notch repeats

(LNRs), followed by a carboxy-terminal hydrophobic region which mediates heterodimerisation (HD). LNR repeats and HD form a negative regulatory region (NRR) which is situated adjacent to the cell membrane. This NRR prevents ligand-independent activation of the Notch receptor [195]. The ICD is composed of conserved protein domains: the RBP-Jκ-associated module (RAM) domain, seven ankyrin (ANK) repeats, a transcription activation domain (TAD) and the C-terminal PEST domain [196].

Engagement of the NOTCH receptor with its ligand induces a conformational change, facilitating S2 cleavage by tumour necrosis factor-a-converting enzyme (TACE), followed by a third cleavage (S3) mediated by the γ-secretase complex. Following this final cleavage, the ICD translocates from the plasma membrane to the nucleus. CSL (RBP-Jκ) complex is then bound and converted from a transcriptional repressor to an activator [197] of hairy enhancer of split (HES) and hairy-related transcription factor (HEY) families [198, 199]. The pathway is also negatively regulated by Fbxw7 (encoded by the FBXW7 gene). this gene targets NOTCH proteins for proteolytic degradation [200]. Multiple laboratories have more recently also reported an alternative non-canonical role for Notch [201-203] antagonizing Wnt/β-catenin signaling, a critical regulator of development and disease, independent of Notch ligand-dependent cleavage or nuclear localization.

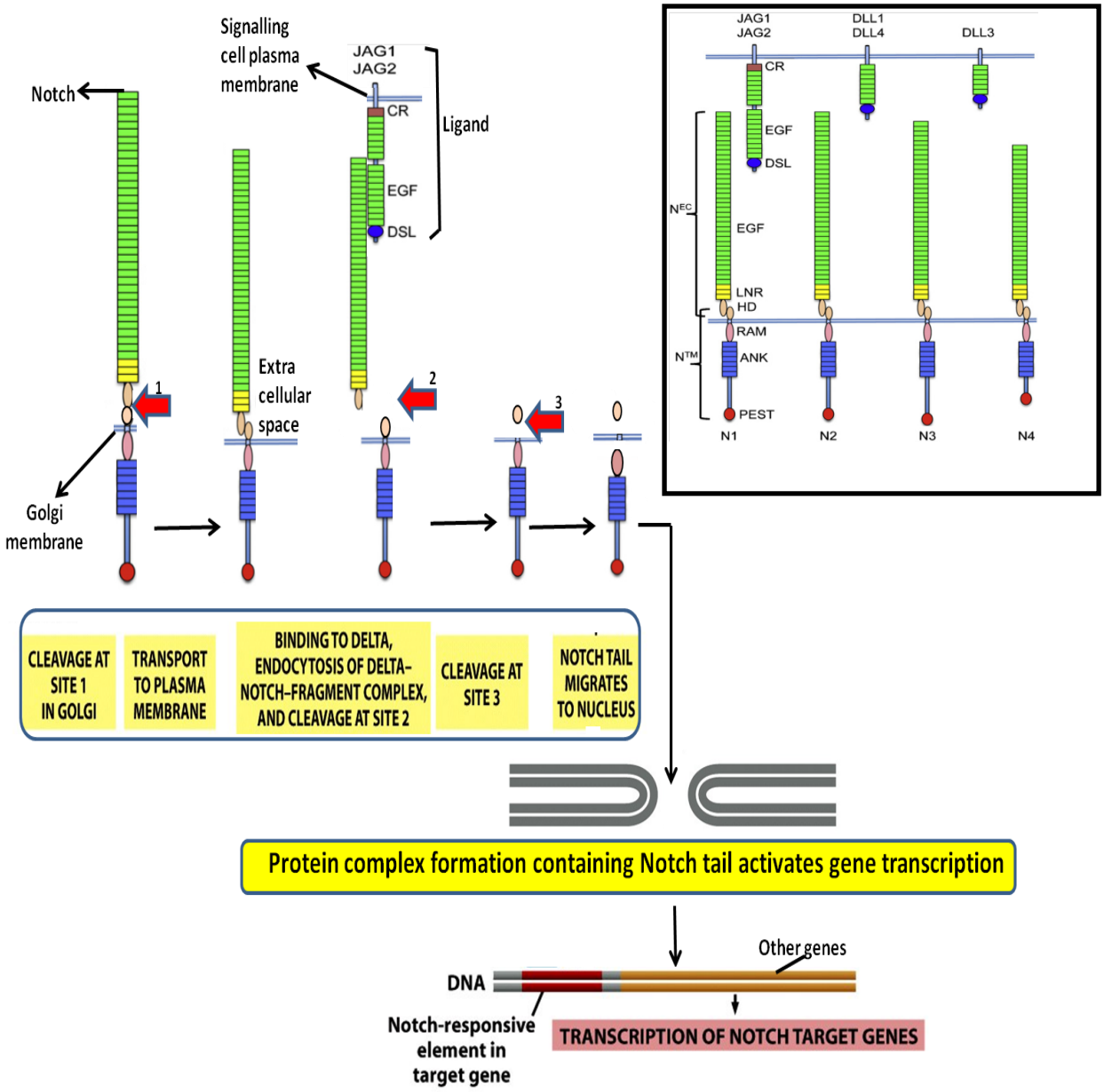

Figure 16. The processing and activation of Notch signaling by ligand binding and proteolytic cleavage. The Structure of different human Notch receptors and ligands are displayed. Out of five Notch ligands expressed in humans, four (JAG1, JAG2, DLL1, and DLL4) can activate Notch receptors. The Ligands have an N-terminal DSL domain, variable numbers of EGF repeats and a juxtamembrane cysteine-rich domain (CR- only for JAG1 and JAG2). There are four Notch receptors, Notch1–4, involved in Notch pathway, comprised of Notch extracellular (NEC) and transmembrane (NTM) subunits. Variable number (29–34) of EGF repeats, 3 Lin12/Notch repeats (LNRs), and the N-terminal portion of the juxtamembrane heterodimerization domain (HD) is present in NEC. NTM is composed of a RAM domain, 7 ankyrin repeats (ANK) repeats and a C-terminal PEST degron domain. (See inset)

(Figure adapted from Source: Molecular biology of The Cell, Bruce Alberts et al.)

5.1.3 NOTCH1 as a potential ‘driver’ of aggressive phenotype in OSCC

A role for NOTCH signaling in human cancer was first demonstrated for T-cell acute lymphoblastic leukaemia, with the identification of a rare t(7:9) (q34:q34.4) translocation [204]. Although it is recognized that NOTCH signaling can be oncogenic, growing evidence has pointed to a tumour suppressor role of this pathway because deletions and inactivating mutations in NOTCH family genes have been identified in a variety of tumour types. In solid tumours, there is evidence for growth suppressive functions for NOTCH signaling in small cell lung cancer [205] and hepatocellular carcinoma [206]. Loss-of-function mutations in NOTCH receptors have also recently been reported in HNSCC, cutaneous and lung SCC [101, 102, 136, 207].

Thus, the specific role of NOTCH1 in OSCC needs to be further clarified; evidence supporting oncogenic and tumour suppressive roles has been published. A cDNA microarray study reported overexpression of receptors (NOTCH1, NOTCH2), ligand (JAG1) and proteins that modulate NOTCH–ligand interactions (manic fringe, lunatic fringe), in microdissected tumour cells from four HNSCC cases [208]. Further reports also found increased expression of NOTCH signaling components in HNSCC. Significantly increased Jagged1 (encoded by JAG1) expression was found in HNSCC compared with dysplasia and normal epithelial tissues [209]. In addition, mRNA levels of several NOTCH pathway genes (NOTCH1, NOTCH2, JAG1, HES1 and HEY1) were also upregulated in oral squamous cell carcinoma (OSCC) compared to normal tongue or oral dysplastic tissues [190, 210].

Clinical correlation studies using tongue cancer samples demonstrated the levels of NOTCH1 protein were positively associated with cervical lymph node metastasis, microvessel density (MVD) and depth of tumour invasion [211, 212]. Localization of the NOTCH1 protein to the invasive front in OSCC [213], has also been observed, allowing for the possibility that NOTCH1 promotes invasion. Additionally, inhibition of the NOTCH pathway by a γ-secretase inhibitor or ‘knocking down’ NOTCH1 significantly reduced cell proliferation and invasion [190, 213-215]. Chemoresistance conferred by NOTCH signaling has been reported for different types of cancer [216], including HNSCC. High expression levels of NOTCH1 protein were found to be strongly associated with

cisplatin resistance in cells isolated from patients with HNSCC [217]. Furthermore, treatment of OSCC cell lines with the pro-inflammatory cytokine, tumour necrosis factor-alpha (TNFα), activated NOTCH1 signaling and led to an increase in tumour sphere-forming ability, stem cell-associated gene expression, chemoresistance and tumourigenicity [210].

However, next generation sequencing studies in HNSCC have suggested a tumour suppressor role for NOTCH signaling, suggesting the NOTCH1 mutations they identified in their respective sequencing studies would lead to loss of function phenotypes [101] [102]. The NOTCH1 mutations identified in both studies were mainly located at or near the epidermal growth factor (EGF)-like ligand-binding domain region within the ECD. Similar results were obtained based on The Cancer Genome Atlas (TCGA) data, in which NOTCH1 mutations were found in 19.3% of HNSCC cases [218]. A further multiplatform analysis revealed the NOTCH1 mutation rate to be 9% but further integrative analysis revealed the NOTCH pathway was defective in 66% (23/35) of cases with alterations in several members of the NOTCH pathway. These included mutations in NOTCH1 and NOTCH2, copy number gains in JAG1, JAG2 and NUMB as well as losses in MAML1 and NOTCH2 [136]. Although JAG1 and JAG2 are generally known to promote NOTCH signaling, they can also inhibit the pathway through cis-inhibition [219]. A contrasting study using a single-molecule DNA sequencing platform analysing a Chinese population reported 42 NOTCH1 non-synonymous mutations in 43% of OSCC cases [220]. A relatively high fraction (29%) of these mutations were located within the heterodimerisation domain within the ECD, where gain-of-function mutations are commonly found. This implies a more complex biological role for NOTCH1 mutations in HNSCC/OSCC than simply being either a tumour suppressor or oncogene. Indeed, studies of Drosophila. have demonstrated that the Abruptex region of the ECD mediates inhibitory Notch activity and thus mutations specific to these regions could result in gains of function [221].

One possible reason for the low sequence coverage of NOTCH1 in our data and samples from other WGS series could be due to its high GC content (17 of 34 exons contain GC> 65% with exon 1 containing GC >75%) of the gene or the large number of

EGF repeats in the ECD. Studies have shown that even with relatively high read depths across most exons, under representation of coverage in several exons (exon 1-26) or no coverage (exon1) was common [183]. We therefore decided to sequence Notch following target enrichment of its exons.

5.1.4 Target enrichment strategies to improve coverage for NOTCH1 Sequencing

Although Sanger sequencing has long been considered the gold standard technique in genetic sequencing it has a number of drawbacks, primarily throughput and lack of versatility for analyzing genes with complex intron / exon organization like NOTCH1. The NOTCH1 gene covers position: base 138000000 to base 138999999 of chr9*;* based on GRCh37 assembly of the human genome) and consists of 34 coding exons. Studies of the NOTCH1 gene would also require large amounts of DNA for assessment of the entire coding regions [74]. Furthermore, obtaining sufficient amounts of DNA for effective sequencing from primary cell line samples can be challenging.

Shotgun sequencing is widely used to sequence very large regions of DNA (often for entire genome) where the DNA is first fragmented into smaller pieces which can be sequenced individually. Sureselect^{TM} is a further method that can enrich targets of interest 300- to 7400-fold dependent upon the size of the targeted region and the number of reads per sequencing run (Agilent). However, one of the weaknesses of the system is that it performs better with large inputs (>5µg) of genomic DNA for library preparation, which is not always available for cancer samples.

Long-range PCR is a fast, efficient and cost-effective choice for enriching candidate genomic region for sequencing, especially when only small sample amounts are available. The method can be combined with next-generation sequencing platforms. Several long-range DNA polymerases are available, able to amplify up to 15 kb or longer genomic DNA [222] with the exons of the NOTCH1 gene comprising ~9.3 kb. Typically, larger amounts of DNA are required for longer targets so as a compromise, multiplex, long range PCR approach for library preparation that was compatible with the next generation Ion Torrent^{TM} sequencing platform, was developed to be applied to each of the locally available OSCC primary cell line samples (n=5) Table 17.

5.2 Hypothesis and aim

Hypothesis: ***Variants in the NOTCH1 gene are associated with ECS in OSCC.***

Aim:

1) To optimize a target enrichment strategy using long range PCR and next generation sequencing to enable sequencing of the exonic regions of NOTCH1 gene.

2) To use this gene specific variant analysis approach for comparing NOTCH1 variants in ECS and non-ECS OSCC primary cell lines for the NOTCH1 gene.

5.3 Results

5.3.1 Optimization of long-range PCR for producing NOTCH1 sequence libraries for Ion Torrent analysis

Long range PCR was performed using human random control DNA and pre specified PCR conditions by Qiagen (see section 2.7.2.1). PCR products (sizes varied between 600 base pairs to 6 kilobases) were analysed by agarose gel electrophoresis (Section 2.5.2). Once PCR Library Preparation was optimized using control DNA, the same procedure was performed with OSCC cell line DNA (see chapter 4, section 4.1.3).

A representative gel image of long range amplicons for exons 13-25 generated from a random control human DNA sample as well as two primary OSCC cell lines (Liv4K, Liv7K), used here as part of optimization, is shown in Figure 17. Expected fragment size for exons 13-18 of ~6000 base pairs and exon 19-25 of ~ 5000 base pairs was seen for both OSCC cell line samples as well as control DNA sample with absence of PCR product in the negative control lane confirming specific amplification and indicating adequate PCR optimization.

The experiments revealed that the primers we had designed for long range PCR were able to amplify successfully exonic regions 5-25 and 31-34 of NOTCH1 but there were particular regions that failed to amplify in a consistent manner (amplicons targeting exons 1-2, 3- 4 and 26-30). The Long Range PCR optimization experiments are summarized in Table 24. To maintain full gene coverage, primers for short range PCR were also subsequently designed to retarget the NOTCH1 exonic regions demonstrating poor coverage by long range PCR. In summary, LR-PCR successfully amplified NOTCH1 exon regions 5-25 and 31-34 covered in only 5 amplicons, with exons 1-4 and 26-30 subsequently successfully targeted by short range PCR. This allowed us to progress on to preparing LR-PCR and SR-PCR amplicon libraries from all five OSCC cell lines (Liv4K, Liv 7K, Liv12K, Liv 22K, Liv26K)**.**

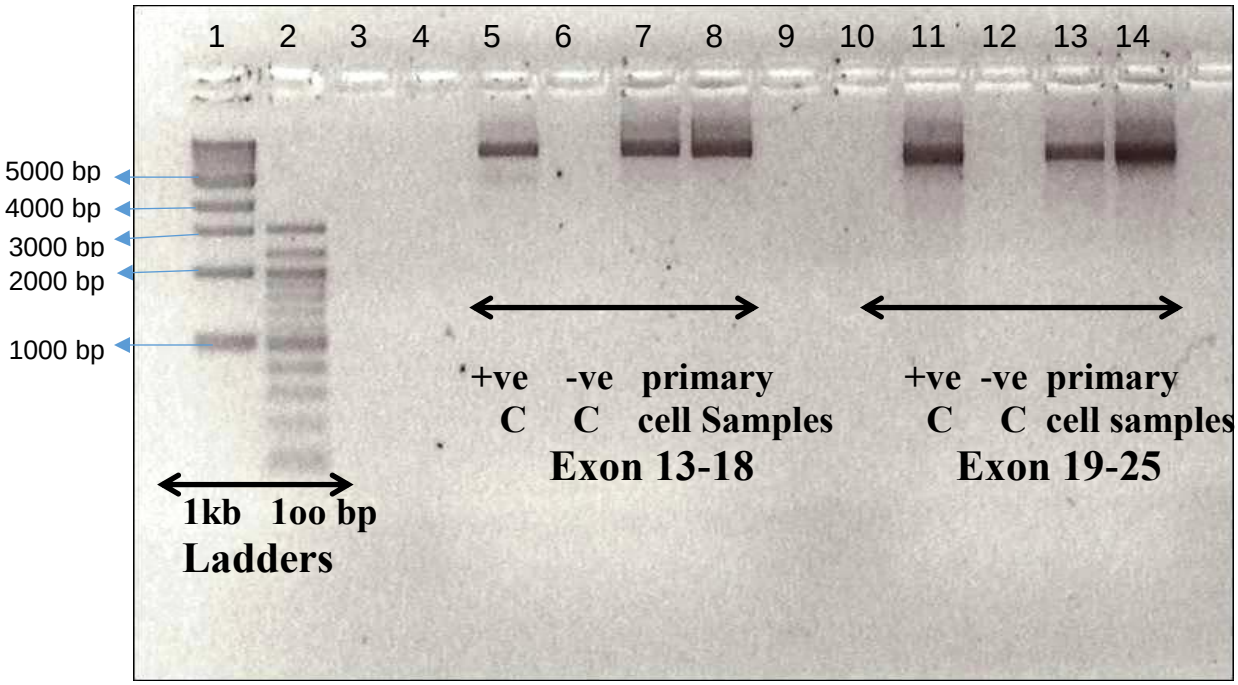

Figure 17. Long range PCR (LR-PCR) amplification of NOTCH1 exons 13-25. Lanes 1 & 2 from the left show 1 kilobase and 100 base pair size markers. Bands are visible in Lanes 5, 7 & 8 at the expected size of 6000 bp corresponding to exons 13-18 and in lanes 11,13 & 14 at the expected size of 5000 bp corresponding to exons 19-25. Lanes 6 & 12 from the left show negative controls with DNA absent. +ve C = Random human control, DNA -ve C = No DNA.

Table 24. Summary of NOTCH1 Long Range PCR optimization.

Amplicon	NOTCH1 Exons	Size (bp)	Long Range PCR Amplification
1	1,2	2002	X
2	3,4	1330	X
3	5	697	✓
4	6,7,8,9,10,11,12	3640	✓
5	13,14,15,16,17,18	5980	✓
6	19,20,21,22,23,24,25	4908	✓
7	26,27,28,29,30	3550	X
8	31,32,33,34	4850	✓

5.3.2 Initial analysis of Ion data using random human control DNA

Ion Torrent^{TM} sequencing of Long and Shorter Range PCR products from the control human DNA library was performed according to the standard protocols as described in Section 2.7. Data was processed using Ion Torrent Suite 2.2 software as described in Section, 2.8. Sequence reads were aligned against human genome assembly Genome Reference Consortium Human Build 37 patch release 13 (GRCh37.p13). The total number of reads generated from the control library was 7,691,887 of which 6,278,086 reads were aligned against chromosome 9; position: 139388896-139440238, which contains NOTCH1 and gave an average NOTCH1 target region coverage of 300-fold. 100 to 200 fold coverage is considered adequate for variant calling. Absence of reads in Exon 1 was evident which may have failed to align due to high GC content.

5.3.3 Variant analysis of NOTCH1 in primary OSCC cell line samples

After confirming adequate coverage for variant calling could be generated from the control library, the libraries made from primary OSCC cell line DNA were also sequenced (see Section 2.7). Reads were aligned against human genome assembly Genome Reference Consortium Human Build 37 patch release 13 (GRCh37.p13). Coverage analysis was performed using Ion Torrent Suite 2.2 software. ~ 77% of the total reads were aligned to NOTCH1. Table 25. shows the mean coverage, which for each cell line was 150-200 fold; adequate for further variant analysis.

Table 25. Summary of sequence coverage from OSCC cell line DNA fragment libraries.

Sample name	Type	Phenotype	Total reads	GRCh37.p13Covered bases against *NOTCH1*	Mean coverage
Liv4K	Primary cell line	Node negative	3,258,560	2,583,993	164.20 fold
Liv7K	Primary cell line	Node negative	3,354,456	2,698,018	199.46 fold
Liv12K	Primary cell line	Node positive ECS+	3,691,887	2,278,086	160.97 fold
Liv22K	Primary cell line	Node positive ECS+	4,279,986	3,455,652	198.52 fold
Liv26K	Primary cell line	Node positive ECS+	4,564,005	3,662,645	153.89 fold

From the exonic/intronic regions of NOTCH1 gene all five cell lines, Ion Torrent Suite 2.2 software (see section 2.8), identified variants already reported either in the Genome 1000 or COSMIC databases. The top 10 most frequent, known variants from the cell lines are summarized in Table 26. Of these variants, four (highlighted in red) have been reported in the literature to be associated with cancer.

Table 26. Common SNPs in OSCC cell lines. The 10 most frequently found variants and their functional consequences are displayed. 3 synonymous codons and one missense variant already described to be associated with cancer are highlighted in red.

SNP-ID	No. of Cell lines	Functional Consequence
rs2229974	5	synonymous codon
rs76371972	5	missense
rs201968456	5	missense
rs184874260	5	synonymous codon
rs114832250	4	missense
rs10521	4	synonymous codon
rs199654211	4	missense
rs35652719	3	missense
rs200053816	3	missense
rs2229971	3	synonymous codon

No novel variants were found in the five OSCC primary cell lines, in one ECS +ve cell line, Liv7K, a rare naturally occuring isogenic NOTCH1 polymorphism was identified in the extracellular domain of NOTCH1 (rs61751543, Minor Allele Frequency 0.01) [223]. This polymorphism has a single nucleotide variation from C>T in exon 23; corresponding to a NOTCH1 extracellular domain variant. The protein residue change due to this polymorphism is R [Arg] ⇒ H [His] at position 1279. To assess the potential

functional consequence of this coding SNP further we performed an evolutionary preservation analysis. NOTCH1 normal protein sequence was used as a control followed by the site and nature of amino acid substitution (corresponding to variant) into the software tool PANTHER-PSEP for predicting non-synonymous genetic variants that may play a causal role in human disease [224]. PSEP (position-specific evolutionary preservation) measures the length of time (in millions of years) a position in current protein has been preserved by tracing back to its reconstructed direct ancestors. The longer a position has been preserved, the more likely that it will have a deleterious effect. PSEP analysis results for this variant based on an Exon 23 NOTCH1 substituition of C>T giving an amino acid change of R64H and a preservation time of 307 million years. This suggests that this could be a damaging variant (450my > time > 200my, corresponding to a false positive rate of ~0.4 as tested on HumVar). The presence of potentially deleterious variant in the ECS+ cell line, supported further investigation of NOTCH1 in relation to ECS.

5.4 Discussion

Setting out to genetically characterise the NOTCH1 gene in five OSCC cell lines and assess whether variants in the NOTCH1 gene are associated with ECS in OSCC we established a target enrichment strategy using long range PCR for next generation sequencing allowing the NOTCH1 exons and flanking intronic sequence to be read with acceptable coverage for variant analysis.

5.4.1 A Long range PCR approach for Ion Torrent^{TM} sequencing of NOTCH1

The key features of an enrichment method have been described as being: enrichment factor, specificity, coverage, evenness of target region coverage, method of reproducibility, overall cost per target base of useful sequence data and required amount of input DNA [225]. One particular challenge for this was the NOTCH1 gene itself. The target region guanine-cytosine (GC) content at extreme values for a target region (<25% or >65%) has a significant impact on enrichment efficiency [226]. This can adversely affect 5'-UTR/ promoter region enrichment and the first exon of genes, which are often GC rich [227]. This was a potential explanation for the poor coverage that we faced in the previous whole genome sequencing part of this study (see Chapter 4, section 4.3.2.7) as well as exon 1 for targeted sequencing.

If PCR product sizes fall within the sequencing length of the applied NGS platform (for example the maximum read length for Ion Torrent: 400 bp); PCR-based enrichment bypasses the need for shot-gun library preparation by using suitably 5'-tailed primers in the final amplification steps. Targeting of very large genome sub-regions is challenging with this approach however (eg. NOTCH1 gene: ~51kb). Alternatively, Long Range PCR has successfully been used to target regions of 5–100 kb [228]. A further advantage of using this approach would be that the error rate of standard Taq DNA polymerase used in other approaches is 1.1×10^{-4} errors/bp [229], although a variety of assay-dependant error rates have been reported. The Long Range PCR Enzyme Mix used here is a mixture of thermostable DNA polymerases which provide both very high extension rates as well as better proofreading ability, allowing up to 5-fold increased fidelity over Taq DNA polymerase [230].

In practice, a short range PCR approach had to be applied for some sections of the NOTCH1 gene (1-2, 3- 4 and 26-30), maybe as a result of nonspecific primer annealing, secondary structures in the DNA template or suboptimal cycling conditions, to maintain full coverage. The subsequent steps were all performed successfully for final DNA library preparation with confirmation achieved by a variety of methods (Bioagilent trace and Ion coverage data generated by Ion Torrent Suite 2.2 software). Specificity was confirmed by high coverage of 150-200 fold for the targeted NOTCH1 sequences in the control library. A clear benefit of the method was the small amount of DNA required from limited supply required to produce results with only 50ng starting amount for long range PCR (comparable to Song et al. 2014), but only 20ng required for short range PCR amplification.

That coverage of NOTCH1 sequences was achieved suggests that the poor coverage observed with WGS data occurred for technical reasons, rather than genetic events leading to reduced copy number.

5.4.2 NOTCH1 Variant analysis for OSCC primary cell lines

Variants that had already been reported as polymorphisms in the Genome 1000 or COSMIC database and these served as incidental positive controls for variant detection by the method. SNPs rs2229974, rs2229971, rs10521, rs76371972) had already reported in the literature to be associated with bladder cancer and acute leukemia [231, 232]. However, there were no unknown non-synonymous variants found in the exonic regions of the primary cell lines assessed. This was surprising considering both the previously reported higher variant false positive calling rate for the Ion torrent method [233] and evidence on 13 HNSCC cell lines that had detected 6 mutations in 4 samples with an overall mutation rate of 31% [220]. These differences may be as a result of cohort or cell line specific effects as they were detected in a variety of head and neck cancers and in a Chinese population. The observation by Pickering et al. that the Notch pathway was defective in 66% of cases in their cohort [136] is at odds with the 25% rate of ECS observed in OSCC [18]. This may be explained by ECS enrichment of the cohort examined by Pickering et al. due to undetected disease in their relatively small

cohort (35 cases) or a cohort specific effect where an excess of mutations were detected in Notch pathway members with an observed low mutation rate of NOTCH1 (9%).

We observed in one node positive ECS positive cell line (Liv7K), a rare NOTCH1 variant (rs61751543) was present. The variant is a missense substitution variation and the resulting protein change would be from Arginine to Histidine. A previous study first reported the rs61751543 NOTCH1 variant, following the profiling of the nonsynonymous genomic variation in 158 genes implicated causally in carcinogenesis. Of the 2,688 allelic variants identified within the cohort, most were very rare, with 75% found in only 1-2 individuals in the European population. Indeed, the rs61751543 NOTCH1 variant allele frequency in the European population has been reported as 0.013% (1000 Genome Project). The variant has been recently identified following whole exome sequencing in two individuals with rectal and one individual with gastric cancer [234] however a complex picture was also observed along with 12 other variants so specific conclusions were difficult to make. Therefore understanding the significance of this finding remains difficult, however generally, it has been observed that only highly aggressive cancers can accumulate genetic changes necessary for unlimited growth *in vitro* and spontaneously become continuous cell lines [235]. Liv7K is a continuous cell line and the patient from where the tissue was obtained died from the aggressiveness of their oral cancer.

There was a relatively small sample size with an absence of normal control cell lines from the same patient to compare to, as well as no Node +ve ECS-ve cell lines available for study (2 x Node -ve ECS-ve cell lines were used). In the light of these, the biological significance of the identified NOTCH1 rare variant is unknown, but warranted further investigation both in terms of functional consequences and as an indicator that NOTCH1 variants should be considered in relation to ECS.

Chapter 6 : TP53 and NOTCH1 Variants in relation to ECS in OSCC

6.1 Introduction

Having established the NOTCH1 sequencing protocol and identified a potentially deleterious variant in an ECS+ primary cell line, we wished to screen for NOTCH1 mutations in our clinical series to see if there was an association between NOTCH1 changes and ECS. At the same time, we wished to screen for TP53 mutations, to see if the low rate of allelic imbalance was matched by low numbers of TP53 mutations in our series and whether there was a relationship with ECS. This required a method of TP53 sequencing to be developed and it was decided to make this compatible with our NOTCH1 sequencing method so they could be used together. The background to NOTCH1 as a candidate 'driver' of ECS has been discussed in the previous chapter (see section 5.1.3), and the role of TP53 in the adverse molecular changes responsible for OSCC has also been previously outlined (section 1.4.2.1).

6.1.1 Sequencing strategy for NOTCH1 and TP53

Prior to the start of the analysis, a power calculation based on dichotomous endpoints (presence or absence of variant) with two independent sample groups was performed and as a result some adjustments were made to the test cohort. Assuming a power of 80%, level of significance of 0.05 and based on the literature expected levels of TP53 and NOTCH1 mutations in non-ECS groups being ~50% and ~15% respectively, to observe an approximate 40% increase in mutational frequency for both genes in ECS cases, a minimum overall sample size of 50 would be required. As a result, some further OSCC cases were added to the series to increase overall numbers (n=50; 19 N-, 10 N+ECS-, 21 N+ECS+). The clinical data for these additional samples is shown in section 2.2.

For calling a base change, ≥100 fold coverage was applied, especially as normal tissue admixture and the heterogeneity of some cancers can decrease variant allele representation [236]. A number of set criteria, established in other studies [237] were also used to ensure good quality variant calling: (i) variant location within an exonic or splicing region (ii) non-synonymous in nature; (iii) absent from European 1,000

genomes variant database (iv) absent from Catalogue of Somatic Mutations in Cancer (COSMIC) database (v) absent from UCSC-SNP137 (University of California Santa Cruz Genome Centre) database (vi) the same variant should not be present in more than 5 samples otherwise it will be considered as an artefact resulting from the sequencing process.

Considering the drawbacks of the Ion Torrent Suite 2.2 software, a further independent bioinformatics analysis of the data was considered necessary. When comparing the Ion Torrent^{TM} platform with other NGS platforms, a higher raw error rate (~1.8% v <0.4%) has been identified. However, if coverage is sufficient, the representation and ability to call SNPs is quite closely matched between these technologies [233, 238]. Along with this, the principal error mode associated with Ion data analysis software is miscalling of homopolymer lengths, which may lead to insertion or deletion errors [238]. Apart from variant detection, haplotype analysis was also necessary to assess for the presence of cross contamination between samples and systematic technical errors associated with the sequencing approaches, which were challenging to analyse in detail using Ion Torrent Suite 2.2 software.

Hence, a BLAST analysis using an Active PERL script for variant detection was performed to reanalyse the Ion Torrent^{TM} platform generated data, allowing for more detailed coverage analysis. Use of PERL in conjunction with BLAST allowed precise positions of variants to be identified, a confounding factor when they occur in homopolymer sequence. The precise methods for the bioinformatics analysis using Active PERL script are documented in section 2.8. Detailed coverage analysis was performed separately for each sample, for both TP53 and NOTCH1 to ensure sufficient read depth both for non-homopolymer and homopolymer regions. This approach was used to obtain coverage data for every PCR amplicon, including fifty bases either side to ensure sufficient amplicon edge coverage. The presence or absence of cross contamination between OSCC samples was verified by haplotype analysis, assessing different locations for common SNPs and sample specific SNPs, with the aim of assessing the uniqueness of sample identity. Candidate variants were then assessed by Sanger sequence to confirm or rule out their presence and produce a final set of

variants. Lastly, association of variants to clinical observations was performed as the final stage of the experimental strategy to assess for any association with ECS, with the detailed clinical data for this set of samples shown in section 2.2.

6.2 Hypothesis and aims

Hypothesis 1: *Variants of NOTCH1 in OSCC are associated with ECS*.

Aim: To use a gene specific NGS approach to identify NOTCH1 variants in an ECS enriched cohort of primary OSCC tumour samples and test association to ECS subtypes.

Hypothesis 2: *Variants in TP53 in OSCC are not associated with ECS.*

Aim: To use a gene specific NGS approach to identify TP53 variants in an ECS enriched cohort of primary OSCC tumour samples and test association to ECS subtypes.

6.3 Results

6.3.1 Long range PCR Amplicon Library Optimization and preparation for TP53 and NOTCH1

Optimisation of Long Range PCR for Ion Torrent™ sequencing was performed for TP53. Each Long-range PCR primer pair for TP53 was designed to allow coverage of the entire exonic region by a minimum number of amplicons and minimal inclusion of intronic sequence. Short introns were included. The PCR product sizes for the TP53 regions varied between ~700 base pairs to ~4.3 kilobases.

Long range PCR Ion library preparation was performed using human random control DNA for NOTCH1 and TP53 gene outlined in section 2.7. Once PCR Library Preparation was optimized using control DNA, the same procedure was performed with OSCC primary tumour DNA (n=50).

6.3.2 Coverage analysis on sequenced OSCC tumour samples

The reads generated from tumour samples were mapped against Genome Reference Consortium Human Build 37 patch release 13 using TMAP aligner (The Torrent Mapping Alignment Program for Ion Torrent). Coverage analysis was performed using Torrent Suite 2.2 software. The coverage data generated by the software demonstrated that the majority of reads (>85%) were mapped onto chromosome 9 and 17 which corresponds to NOTCH1 and TP53 genes respectively (base 138000000 to base 138999999 of chr9 for NOTCH1 and base 7000000 to base 7999999 of chr17 for TP53; based on GRCh37 assembly of the human genome) confirming that the libraries generated encompassed the TP53 and NOTCH1 genes as expected.

When comparing the Ion Torrent platform with other NGS platforms, it is known to have a higher raw error rate (~1.8%) than others platforms (<0.4%) due to miscalling within homopolymer regions, however if there is sufficient coverage, the representation and ability to call variants is quite closely matched between these technologies (Quail et al., 2012). The, more detailed strict coverage analysis was therefore performed using Active

PERL (for full description see Section 2.8). This revealed 40 samples where overall average read depth was ≥100. 10 samples showed either very low coverage (where mean coverage <50) or low coverage (mean coverage >50 but <100). These 10 samples were excluded from further analysis to maintain stringency in variant calling. Their clinical identity with regards to nodal status is tabulated with sample ID (see Table 26).

Table 27. OSCC tumour samples identified by Active PERL script to have poor coverage

Sample Identity	Sample ID	NOTCH1 mean coverage	TP53 mean coverage
N+ve+ECS+	3167	43	37
N+ve+ECS+	3230	25	28
N-ve	3240	44	48
N+ve+ECS+	3241	49	42
N-ve	3272	41	40
N+ve+ECS-	3274	32	29
N-ve	3315	38	32
N+ve+ECS-	3481	27	29
N+ve+ECS+	3251	77	97
N+ve+ECS-	3254	85	64

Figure 18A. shows mean coverage data from combined tumour samples (n=40) generated by Ion Torrent™ sequencing of NOTCH1 and bioinformatics analysis by Active PERL algorithm on the basis of coordinate position with corresponding exonic and intronic positions highlighted on the same axis. From the previous chapter (5) we had previously identified that a low read depth was observed in some regions (ex 1-4,

26-30) of NOTCH1 using long range PCR amplicons. Therefore, short-range PCR primers were used to retargeting those regions. Using both long and short-range amplicons adequate coverage was achieved suitable for variant calling (see chapter 5, section 5.3.1). Detailed PERL script coverage analysis generated from each exon showed adequate coverage (average read depth 100X) for exonic regions 2-34 of the gene (position- chr9: 139,438,476- 139,390,523; based on Genome Reference Consortium Human Build 37 patch release 13) but exon 1 of NOTCH1 (position: 139,440,178- 139,440,238) was missing from our output, as also reported in the results to the previous chapter 5 (section 5.3.1). Short range PCR amplicons (exons 2, 3, 4, 5, 26, 27, 28, 29, 30) produced much greater mean read depth from 300 to 1400 compared to the long range PCR amplicons, mean read depth ~ 150.

Figure 18B. shows coverage data of combined tumour samples (n=40) generated from TP53 gene based on coordinate position with corresponding exonic and intronic positions highlighted on the same axis. Detailed coverage analysis showed adequate reads (above 100 mean coverage) for variant calling for all exonic and intronic regions (position- chr17: 7,590,694-7,571,719).

A.

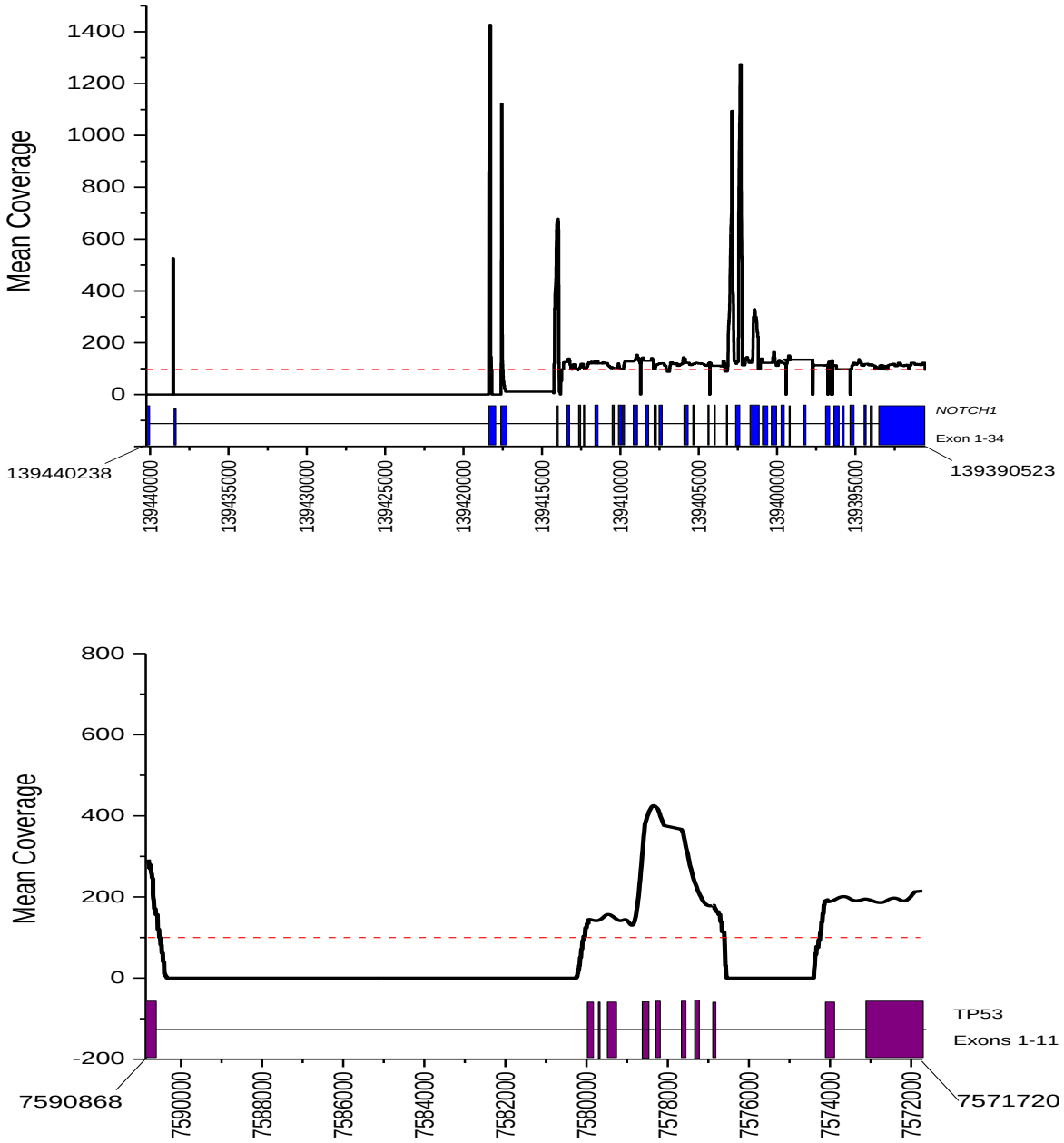

Figure 18. Mean combined coverage of OSCC tumour samples (n=40) following Ion TorrentTM sequencing and ActivePerl Bioinformatic Analysis. 18A. Mean Coverage is shown on the basis of coordinate position of NOTCH1 (139390523-139440238 GRCH 37.13) with precise corresponding exonic and intronic positions (exons 1-34) shown on the same axis. Exon 1 showed no coverage. **18B.** Mean Coverage is shown on the basis of coordinate position of TP53 (7571720-7590868 GRCH 37.13) with precise corresponding exonic and intronic positions (exons 1-11) shown on the same axis. Dashed red line indicates mean coverage above our cut off of 100

6.3.3 Examination of Cross-contamination between OSCC tumour samples

Use of PCR could have resulted in cross-contamination between samples and was assessed by comparing the profile of known polymorphisms generated from each OSCC tumour sample. The principle used was that if different samples had exactly the same known polymorphism profile then the possibility of having cross contamination in between these samples was high. Variants identified using the analysis performed by ActivePerl algorithm were correlated with the UCSC-SNP137 (University of California Santa Cruz Genome Centre: http://hgdownload.soe.ucsc.edu /goldenPath /hg19/snp137Mask/) source sequence database to identify known SNPs. A Latin Square of samples, versus SNPs, sorted according to SNP frequency in the samples overall was created. 3 groups of positive controls were also used in the analysis for each gene (1 group of 3 and then 2 groups of 2) where the same samples had been intentionally repeated, with a similar SNP profile expected each time the same sample was analysed and this was observed, confirming the reproducibility of the results. Unique SNP profiles were observed for both NOTCH1 and TP53 in all of the other samples, suggesting the absence of cross contamination between tumour samples in both cases. Furthermore, statistical analysis using Chi-square tests comparing of variants for each sample combination in turn found significant differences ($p = 0.001 - 0.042$) between in each case.

6.3.4 NOTCH1 variants in primary OSCC tumour samples

Ion AmpliSeq sequencing data initial variant calling was generated using Torrent Suite Software v2.2 with the plug-in “variant caller v2.2” program as described in method section 2.8. For unknown exonic variant detection the filtering criteria outlined previously in Section 6.1.1 was applied. A total of 9/50 samples were found to have a total of 10 NOTCH1 exonic variants (see Table 28.). All variants were single base substitutions with two variants (NOTCH1 Exon 20 139402707 G>C & NOTCH1 Exon 23 139401209 G>A) observed twice and one sample displayed two variants in both exon 5

and exon 20 (sample 3366). The overall variant frequency was 18% (9/50), consistent with previously published rates (11-19%).

Reanalysis of the sequencing data using the Active PERL script was also used to identify variants. Their coverage and frequency in the samples was also determined. 10 sample excluded for poor coverage. Rather than apply cut off filters to set inclusion criteria for possible variants, frequency versus coverage was plotted for the entire data so that allowance could be made for the context of candidate variants. analysed and plotted to allow the identification of significant variants. Previously known SNPs were used as controls for positive selection of variants. Homopolymer and non-homopolymer regions were considered. This demonstrated that at 10% coverage frequency cut off, significant numbers of known SNPs could be missed, thus risking exclusion of novel variants. At a 5 % frequency cut off, the majority of known SNPs and therefore novel variants were retained. This was accepted as the level that maximised true positives, whilst minimising false positives. Other stringent criteria (section 6.1.1) were applied to produce a list of unknown variants. 5 novel variants were each detected by both types of analysis. 5 detected in the first analysis were undetected in the second analysis and the latter gave 9 additional variants (see Table 28). Of the 14 unknown variants that were discovered in the second analysis, 13 were in different samples, and 2 variants were detected twice within 2 different samples (G>C 139402707 exon 20, G>A 139401209 exon 23). One sample displayed two different types of unknown variant (sample 3286).

Sanger revalidation of all identified variants was then performed independently through Source Bioscience. Sequence Scanner v2. software was used to examine the sequences as chromatograms and the status of the expected novel variants determined. All 5 variants detected by both by both methods were validated but only 5/10 (50%) of the variants from the first analysis and 7/14 (50%) were validated.

Table 28. Unknown non-synonymous exonic NOTCH1 variants identified in OSCC tumour cases using Ion Torrent Suite 2.2., PERL script and subsequent Sanger validation. All variants along with position and corresponding amino acid changes are displayed. Histopathological status of the primary tumour sample is also shown. 5 variants detected by both Ion Torrent and PERL analysis were validated by Sanger sequencing, with a further two variants only detected by PERL also validated by Sanger (marked by red). 2 variants were detected twice within 2 different samples (G>C 139402707 exon 20, G>A 139401209 exon 23) and one sample displayed two different types of unknown variant (sample 3286).

Sample Identity	Clinical identity	Base position	Exon position	Base change	Protein change	Protein domain	Ion Torrent	PERL	Sanger
3366	n+ecs+	139402707	ex20	TGT>TCT	CYSTEINE>SERINE	extracellular domain	yes	yes	yes
3455	n+ecs+	139402707	ex20	TGT>TCT	CYSTEINE>SERINE	extracellular domain	yes	yes	yes
3484	n+ecs+	139393368	ex33	GAT>TAT	ASPARTIC ACID>TYROSINE	transmembramne subunit	yes	yes	yes
3499	n+ecs+	139401209	ex23	CGC>CAC	ARGININE>HISTADINE	extracellular domain	yes	yes	yes
3270	n+ecs+	139401209	ex23	CGC>CAC	ARGININE>HISTADINE	extracellular domain	yes	yes	yes
3345	n+ecs+	130404185	ex18	AGG>ATG	ARGININE>LYSINE	extracellular domain	yes	no	no
3233	n+ecs+	139412208	ex8	ATG>ATT	METHIONINE>ISOLEUCINE	extracellular domain	yes	no	no
3366	n+ecs+	139413951	ex5	GGT>GAT	GLYCINE>ASP	extracellular domain	yes	no	no
3300	n+ecs-	139400047	ex25	GGT>GTT	GLYCINE>VALINE	extracellular domain	yes	no	no
3238	n-	139407961	ex14	AGT>GGT	SERINE>GLYCINE	extracellular domain	yes	no	no
3483	n+ ecs-	139400017	ex25	CTG>INSERTION	INSERTION	extracellular domain	no	yes	yes
3286	n+ecs+	139400300	ex25	CGT>AGT	ARGININE>SERINE	extracellular domain	no	yes	no
3286	n+ecs+	139391840	ex34	AAC>AAG	ASPARAGINE>LYSINE	transmembramne subunit	no	yes	no
3233	n+ecs+	139412197	ex8	TCG>TTG	SPLICE SITE	extracellular domain	no	yes	yes
3345	n+ecs+	139395297	ex31	GGC>CGC	GLYSINE>ARGININE	transmembramne subunit	no	yes	no
3461	n+ecs+	139402420	ex21	GGC>INSERTION	INSERTION	extracellular domain	no	yes	no
3340	n0	139400300	ex25	CGT>AGT	ARGININE>SERINE	extracellular domain	no	yes	no
3242	n0	139399949	ex25	TGC>CGC	CYSTEINE>ARGININE	extracellular domain	no	yes	no
3383	n0	139393707	ex32	CTG>INSERTION	INSERTION	transmembramne subunit	no	yes	no

The false positive rate of variant detection for NOTCH1 was unlikely to have been as a result of admixture of normal cells with the tumour cells because tumour content had been previoiusly been reported by a histopathologist at a high average, typical for this type of cancer, of ~70% for our samples and the minimum tumour cellularity required for heterozygote variant calling has been reported as ~20% [239]. Technical, algorithmic issues reasons are more likely to be the principle source of error because the two different methods used were not in agreement and the principal type of error reported is miscalling of homopolymer lengths, which may lead to insertion or deletion errors (Salipante et al., 2014). Indeed, only one of the three insertions identified was found to be validated. Importantly, the use of two different methods is likely to have minimized the false negative rate. If the majority of the novel variants found had been validated, this would have suggested that the criteria for inclusion were too strict.

Known base substitution variants already reported either in COSMIC, in UCSC-SNP137 or in European 1,000 genomes variant database were also identified to obtain a complete variant profile for our OSCC cases. A cut off value of base read depth was set at ≥100 and variant frequency ≥5%. Table 29. shows the most commonly identified known variants present in exonic regions of NOTCH1 and from both homo and non-homopolymer regions of the gene. The proportion discovered in ECS cases along with already known cancer associations, reported in the ClinVar database, are also summarised (n=40).

Table 29. Most common known exonic SNPs observed in NOTCH1. Assessment was made from both homopolymer and nonhomopolymer regions (n=40). The ClinVar database was interrogated to assess the presence of any existing cancer associations and the proportion of ECS samples identified with the SNP present is also shown.

SNP	Cancer Association	Total cases / 40	ECS cases / 14
rs201358664	-	39	13
rs114479009	-	34	12
rs61751535	-	34	12
rs2229974	-	31	12
rs2229971	-	29	8
rs201968456	-	29	8
rs184874260	-	25	7
rs10521	-	25	6
rs76371972	-	20	5

For known INDEL detection, the cut off value for base read depth was also set to ≥100 with ≥5% variant frequency before identifying any variant as a true INDEL and only exonic INDELs were taken into consideration. For NOTCH1, three exonic deletions were observed in more than half of the OSCC samples (rs140371192, rs146350322, rs41309766).

6.3.5 NOTCH1 variant functional significance

Most of the NOTCH1 mutations identified by the two seminal HNSCC NGS studies (Agrawal et al., 2011, Stransky et al., 2011) were observed in the extracellular EGF-like repeat domain and are summarized in Figure 19. Mutations within this region could potentially have great functional significance as it codes for the EGF-like repeats 10–13 (coded by exons 6-9) considered to be the ligand binding region of the NOTCH1 protein. Studies specifically focused on OSCC have demonstrated mutation rates up to 10% for the ligand binding region of NOTCH1 with a predominance of non-synonymous G>A transitions [240].

The 7 novel NOTCH1 validated by Sanger sequencing within our ECS enriched OSCC cohort were found within exons 8, 20, 23, 25 and 33. Furthermore, if NOTCH1 variants found in OSCC cases were structurally distributed and organized by corresponding exon and protein domain then most of the identified mutations would be clustered into the N-terminal, extracellular EGF like domain, with a few mutations falling within regions coding for the transmembrane subunit of NOTCH protein, see Figure 19. While a splice site variant was observed in exon 8 and an insertion in exon 25 the remaining 5 variants were all single base substitutions (G>C, G>T, G>A) with 2 variants (exon 20 G>C, exon 23 G>A) identified in two different samples. To assess the functional consequences of these coding SNPs further we performed an evolutionary preservation analysis entering normal protein sequence as a control followed by site and nature of amino acid substitution (corresponding to variant) into PANTHER-PSEP [224]. PSEP analysis results for three different coding SNP variants are summarised in the Table 30.

Table 30. Summary of NOTCH1 coding SNP PSEP analysis results. All three NOTCH1 variants were found likely to have potentially deleterious effects on PSEP analysis. my=million years, Interpretation thresholds: "probably damaging" (time > 450my, corresponding to a false positive rate of ~0.2 as tested on HumVar database), "possibly damaging" (450my > time > 200my, corresponding to a false positive rate of ~0.4) and "probably benign" (time < 200my).

Variant	Substitution	Preservation Time / my	Interpretation
Exon 20 G>C	C446	1036	probably damaging
Exon 23 G>A	R72H	307	possibly damaging
Exon 33 G>T	D27Y	797	probably damaging

This analysis revealed that these three coding SNPs, were likely to have significant functional consequences and alongside the other variants identified we can hypothesize that all the variants identified would lead to impaired NOTCH signaling either as a result of impaired ligand binding (Exon 8-25 variants) or intracellular NOTCH pathway activation (Exon 33 variant) and this would certainly warrant further investigation with more specific functional studies.

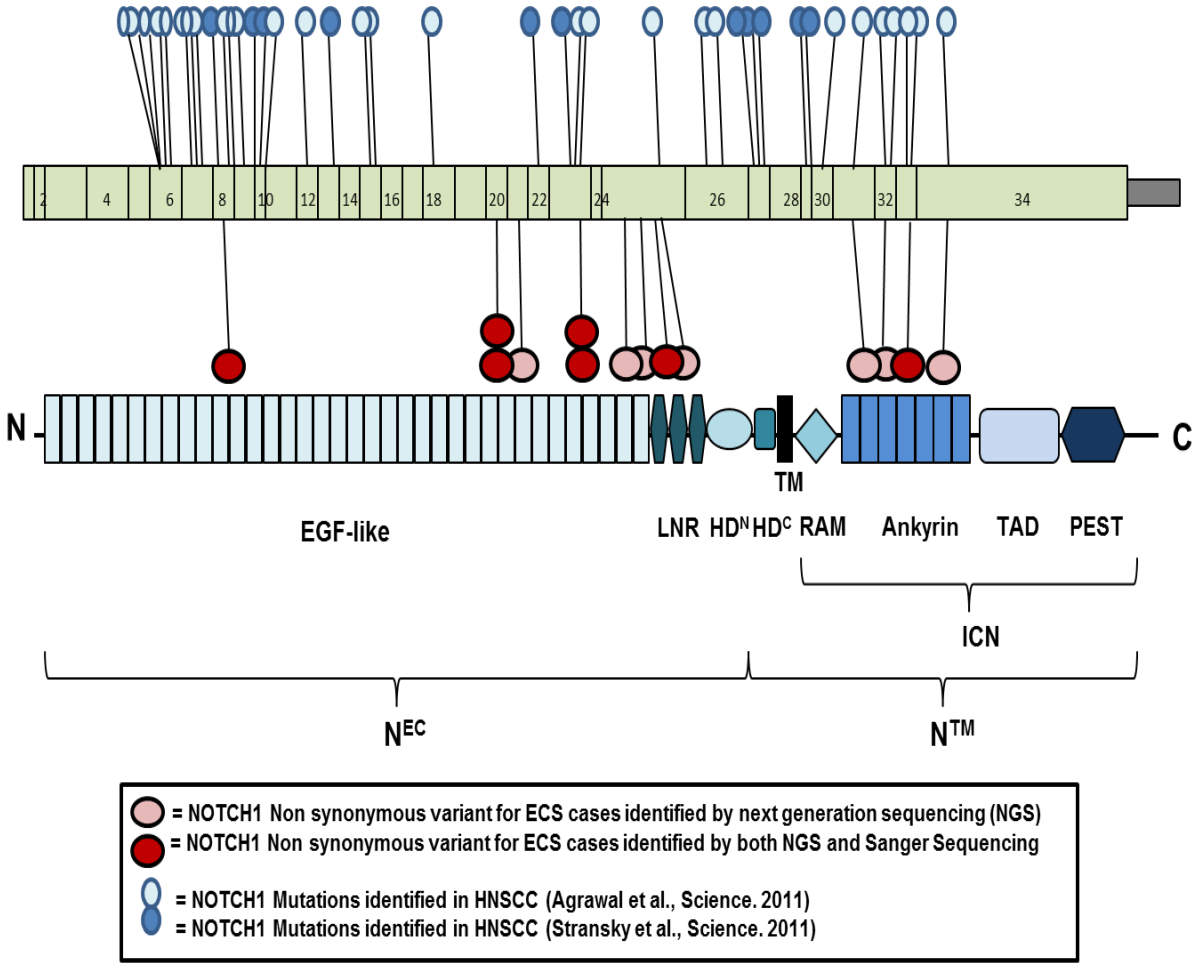

Figure 19. Structural distribution of NOTCH1 variants found in OSCC cases organized by corresponding exon and protein domain. All NGS ActivePerl algorithm identified variants (pink) and Sanger validated variants (red) are shown in relation to coding exon position and corresponding NOTCH1 protein domian. Sites of mutations found from two seminal HNSCC sequencing studies (Agrawal et al, 2011 and Stransky et al. 2011) are shown for comparison (dark and light blue circles). HD = Heterodimerization domain (N and C terminus), PEST = Peptide sequence rich in Proline, Glutamate, Serine and Threonine amino acids, N^{EC}= Extracellular domain, N^{TM} =Transmembrane subunit, RAM = RBP-Jkappa-associated module, TAD = Trans-activating domain , LNR = Lin-12/Notch repeat domain, ICN = intracellular Notch domain

6.3.6 Novel NOTCH1 variants in association with OSCC histopathological status

Histopathological sample identity with regards to nodal status and presence of ECS was correlated with the presence of validated NOTCH1 variants in 2 x 2 contigency tables with ECS+ve and ECS-ve (N- & N+ve ECS-ve) groups or N-ve and N+ve (N+ve ECS-ve & N+ve ECS+ve) and significance determined by a two-tailed Fisher's exact test. The resulting p values are summarised in Table 31. A significant association (p=0.0109) was observed between the presence of NOTCH1 variants and nodal status. NOTCH1 variants also had a significant association (p=0.0044) with the presence of ECS. The validated NOTCH1 variant frequency of 17.5% (7/40) corresponded to mutational frequencies observed in the literature (Agrawal et al. 2011, Stransky et al. 2011) and the association of NOTCH1 variants with poor histopathological prognostic indicators in OSCC also corresponded with previous evidence showing NOTCH1 mutations to be associated with significantly shorter overall and disease-free survivals (Song et al., 2014). Our observation of a significant association of NOTCH1 variants with ECS was a novel finding and taken in conjunction with the potential deleterious functional effects of the NOTCH1 variants described earlier, our results suggest NOTCH1 may be a 'driver' of clinically aggressive forms of OSCC.

Table 31. Summary of NOTCH1 variants according to histopathological subtype. Fisher's exact test was performed comparing NOTCH1 variants in ECS +ve to ECS-ve cases with p values shown, demonstrating a significant association between NOTCH1 gene variants and ECS status.

		NOTCH1 mutation		
		Discovery (By Ion software)	**Discovery (By ActivePerl Algorithm)**	**Sanger Validation**
N-ve	18	1	3	0
N+ve ECS-	8	1	1	1
N+ve ECS+	**14**	**7**	**9**	**6**
p value N+ vs N-		**0.0266**	**0.0896**	**0.0109**
p value ECS+ vs ECS-		**0.0044**	**0.0036**	**0.0044**

6.3.7 TP53 variants in primary OSCC tumour samples

TP53 sequences from our OSCC tumour samples were analysed as above, Section 6.3.4. The coverage versus frequency graphs for each position as for NOTCH1 showed that at a cut off level of 5% variant frequency level, false positive variants were minimised whilst candidate, novel variants were retained.

Table 32. shows the unknown variants (insertions, deletions and SNPs) identified using this method in both homopolymer and non-homoplymer regions for TP53. 7 variants, were identified including 5 previously identified by Ion Torrent analysis alone and only the 5 variants found by both methods were confirmed by Sanger sequencing. The table also summarises TP53 non-synonymous exonic variants along with the expected resulting conformational change. A total of 5/50 samples were found to have TP53 exonic variants with the majority (3/10) found within Node -ve cases. Four of the variants were single base substitutions (TP53 Exon 4 7579370 G>T, TP53 Exon 4 7579473 C>G, TP53 Exon 5 7578473 C>G) with one insertion also identified (TP53 Exon 6 7578185 G>GCT).

Table 32. Validated, unknown non-synonymous exonic TP53 variants identified in OSCC tumour cases. All variants along with position and corresponding amino acid changes are displayed. Histopathological status of the primary tumour sample is also shown. 5 variants detected by both Ion Torrent and PERL analysis were validated by Sanger sequencing (marked by red), with a further two variants only detected by PERL script.

Sample Identity	Clinical identity	Base position	Exon position	Base change	Protein change	Ion Torrent	PERL	Sanger
3345	n+ecs+	7578473	Ex4	CCC>GCC	PROLINE>ALANINE	yes	yes	yes
3300	n+ecs-	7579473	ex3	CCC>CGC	PROLINE>ARGININE	yes	yes	yes
3258	n-	7579370	ex3	AGC>ATC	SERINE>ISOLEUCINE	yes	yes	yes
3461	n-	7578185	ex5	G>GCT	INSERTION	yes	yes	yes
3242	n-	7576855	ex8	CAG>TAG	GLUTAMINE>STOP CODON	yes	yes	yes
3533	n+ecs+	7577105	ex7	CCT>CTT	PROLINE>LEUCINE	no	yes	no
3483	n+ecs-	7577563	ex6	T->-	DELETION	no	yes	no

6.3.8 Functional Significance of TP53 variants

p53 protein is active as a homotetramer, comprising 4 × 393 amino acid residues, functioning primarily as a transcription factor. Functionally, mutations in different regions of the gene would potentially have differential effects; exons 2 and 3 affecting transactivation and the interaction of p53 with CBP, CREB, MDM2 and p300, exons 5 to 8 affecting DNA binding and exons 9 and 10 affecting oligomerization. The five TP53 validated by Sanger sequencing within our OSCC cohort were found within exons 3, 4, 5 and 8. The variants discovered in DNA binding domain coding region exon 5 was an insertion and exon 8 a single base substitution resulting in a stop codon, both likely to profoundly affect protein structure and function (i.e. frameshift mutation and truncated protein). The two single base substitutions (both C>G) observed in transactivation domain coding exon 3 as well as a single base substitution (G>T) in exon 4, while not in the commonly mutated DNA binding domain, may still have had detrimental effects on protein function and were similar to those previously described in OSCC [241, 242]. To assess the functional consequences of these three coding SNPs further we performed PANTHER-PSEP analysis. PSEP analysis results for the three variants are summarised in the Table 33.

Table 33. Summary of TP53 coding SNP PSEP analysis results. All three TP53 variants were found likely to not have deleterious effects on PSEP analysis. my=million years, Interpretation thresholds: "probably damaging" (time > 450my, corresponding to a false positive rate of ~0.2 as tested on HumVar database), "possibly damaging" (450my > time > 200my, corresponding to a false positive rate of ~0.4) and "probably benign" (time < 200my).

Variant	Substitution	Preservation Time / my	Interpretation
Exon 3 C>G	P40R	176	Probably benign
Exon 3 G>T	S74I	91	Probably benign
Exon 4 C>G	P28A	176	Probably benign

This revealed that these three coding SNPs were less likely to have significant functional consequences compared to the other two variants detected, again highlighting the seemingly low level of TP53 abnormalities in our cohort.

6.3.9 Association of novel TP53 variants with histopathological status

Histopathological subtypes with regards to nodal status and presence of ECS were assessed for the presence of TP53 variants. 2 x 2 contigency tables were compiled and analysed for significance by two-tailed Fisher's exact test (Table 34.). No association of TP53 variants was observed with either nodal status or ECS. The low TP53 variant rates (Sanger validated 12.5% (5/40)) compared to the literature ~50% (Agrawal et al. 2011, Stransky et al., 2011) within our ECS enriched cohort is a factor. Lack of association with both nodal status and ECS agrees with the negative results observed previously in this study with regards to a-CGH and SNP genotyping for TP53 allelic imbalance (see chapter 3, section 3.3.4).

Table 34. Summary of TP53 variants in association with histopathological subtype. Variant numbers validated by Sanger Sequencing are shown. Fisher's exact p values are shown and no significant associations were found.

		TP53 variant		
		Discovery (By Ion software)	**Discovery (By ActivePerl Algorithm)**	**Sanger Validation**
N-ve	18	3	3	3
N+ve ECS-	8	1	2	1
N+ve ECS+	**14**	**1**	**2**	**1**
p value N+ vs N-		**0.6419**	**1.0000**	**0.6419**
p value ECS+ vs ECS-		**0.6404**	**1.0000**	**0.6404**

6.3.10 Known TP53 variants in association with OSCC histopathological status

The very low frequency of TP53 alterations found in our study compared to published studies raises the significance of possibly deleterious known variants because these could make up the difference compared to other studies. Known variants were identified for TP53 gene in the cases, using the same criteria as above. Table 35. summarises the known SNPs, found in the exonic regions of TP53 gene both homo and non-homopolymer regions of the genes. The proportion discovered in ECS cases along with already known cancer associations, reported in the ClinVar database, is also summarised. Similarly, for INDEL detection, the cut off value of read depth of the base was set to ≥100 with a variant frequency ≥5%. Only one TP53 exonic insertion (rs397516437) was observed.

Table 35. Common known exonic SNPs observed in TP53. Assessment was made from both homopolymer and nonhomopolymer regions (n=40). The ClinVar database was interrogated to assess the presence of any existing cancer associations. Number of cases harbouring the variant on the basis of ECS is shown, and the fisher's exact test p value comparing variant levels between ECS +ve and ECS -ve groups are shown. No differences between ECS -ve and ECS +ve levels of known potentially deleterious variants was observed.

SNP	Existing Cancer Association	ECS -ve cases / 26	ECS +ve cases / 14	Fisher's exact p ECS+ v ECS-
rs1042522	Li-Fraumeni syndrome	25	12	0.2763
rs28934874	Liver, Malignant melanoma, Multiple myeloma, HNSCC, Adenocarcinoma of lung, Li-Fraumeni syndrome, Squamous cell carcinoma of lung, Brain, Breast, Pancreatic adenocarcinoma, Bladder TCC, Oesophageal, Colorectal, Stomach, Ovarian ,Uterine	20	8	0.2808
rs72661119	Li Fraumeni syndrome	12	4	0.0958
rs141402957	Hereditary cancer predisposing syndrome	12	5	0.7385
rs56275308	Li Fraumeni syndrome, Hereditary cancer predisposing syndrome	12	5	0.7385
rs144340710	Li Fraumeni syndrome, Hereditary cancer predisposing syndrome, Rhabdomyosarcoma	11	4	0.5024
rs11540652	Liver, CLL, Medulloblastoma, Malignant melanoma, Multiple myeloma, HNSCC, Small cell lung cancer, Adenocarcinoma of lung, Li-Fraumeni syndrome, Squamous cell carcinoma of lung, AML, Myelodysplastic syndrome, Brain, Breast, Glioblastoma, Pancreatic adenocarcinoma, Skin SCC, Bladder TCC, Glioma, Oesophagus, Colorectal, Stomach, Ovary, Uterus, Prostate	9	3	0.2992
rs56184981	Li Fraumeni syndrome	9	2	0.0897
rs121912651	Liver, CLL, Medulloblastoma, Malignant melanoma, Multiple myeloma, HNSCC, Small cell lung cancer, Adenocarcinoma of lung, Squamous cell carcinoma of lung, AML, Myelodysplastic syndrome, Brain, Breast, Glioblastoma, Pancreatic adenocarcinoma, Skin SCC, Bladder TCC, Brainstem glioma, Oesophagus, Colorectal, Stomach, Ovary, Uterus, Prostate	8	2	0.1570
rs28934875	Li-Fraumeni syndrome	7	3	1.000

Analysis of known deleterious variants revealed that a significant excess of these variants within ECS cases did not exist. Hence, these variants could not explain the shortfall in the overall low rate of TP53 variation observed in our cohort of cases. This implied that we may genuinely have a subset of cases with lower than reported TP53 variation but carrying NOTCH1 mutations as summarized in Table 36. Only one of the cases was ECS -ve. This sample was from the oropharynx and had been identified as having mutant TP53 by initial NGS analysis but this was not confirmed by Sanger sequencing. The remaining samples were all from the oral cavity with wild type TP53.

Table 36. Samples with Sanger validated NOTCH1 mutants and their TP53 status.

Sample	Site	TP53 status	HPV Status	ECS Status	Mutation type
3366	OC	WT	-	+	Nonsense
3455	OC	WT	-	+	Nonsense
3484	OC	WT	-	+	Indel,Missense
3499	OC	WT	-	+	Missense
3270	OC	WT	-	+	Missense
3483	OP	WT (Mutant in PERL)	-	-	Missense
3233	OC	WT	-	+	Missense

OC=Oral cavity, OP=Oropharynx, WT=Wild type

To investigate this further we re-assessed the literature to see whether a similar subset of cases with low TP53 variation and NOTCH1 mutations had been observed in other studies, specifically in OSCC cases. Table 37 summarizes all the OSCC cases discovered by Agrawal et al. displaying NOTCH1 mutations with their TP53 status, HPV status, Site and Type of mutation shown.

Table 37. OSCC Samples with NOTCH1 mutations in Agrawal et al. and TP53 status.
WT=wildtype, OC=oral cavity, OP=Oropharynx, +=HPV present, -=HPV absent, u= HPV unknown

Sample	Site	TP53 status	HPV Status	Amino acid	Mutation type
HN12PT	OC	WT	-	p.W1843X	Nonsense
600	OC	WT	u	Q290X	Nonsense
HN 105 PT	OC	WT	-	fs, p.G812W	Indel,Missense
478	OC	WT	-	p.G484V	Missense
HN 102 PT	OC	WT	-	p.E450K	Missense
HN 117 PT	OC	WT	-	p.G310R, p.C456R	Missense
HN 208 PT	OC	Mutant	-	p.R365C, p.R1280C	Missense
385	OC	Mutant	-	p.R1984X	Nonsense
HN 107 PT	OC	Mutant	-	p.S402X	Nonsense
HN 115 PT	OC	Mutant	-	fs,p.G1340A	Indel,Missense
HN 183 PT	OC	Mutant	-	p.Q1958X	Nonsense
HN 194 PT	OC	Mutant	-	fs	Indel
HN14PT	OC	Mutant	-	p.P391S	Missense
HN 245 PT	OP	WT	+	fs	Indel
HN 142 PT	OP	WT	+	p.R353H, p.M2011R	Missense
HN 251 PT	OP	WT	+	p.F1292L p.V2039L	Missense
HN 255 PT	OP	WT	u	p.V2039E	Missense
HN 148 PT	OP	Mutant	u	p.G1638V, p.P2272S	Missense
HN 139 PT	OP	Mutant	u	p.C554X	Nonsense

Interrogation of the data revealed 19 samples with NOTCH1 mutants that were OSCC cases. Of these 10 had wildtype TP53 and 9 had mutant TP53. Most NOTCH1 mutant samples identified by Agrawal et al. were from the oral cavity (13/19) with the HPV staus either +ve or unknown for all oropharynx samples. Considering the oral cavity samples we then considered the 12/13 samples that were confirmed as HPV -ve, to draw an appropriate comparison with our study of HPV -ve cases. Of these 6/12 displayed wildtype TP53, suggesting our subset could be similar in nature to these samples. Of note, the ECS status of the samples identified by Agrawal et al. was not documented for comparison.

6.4 Discussion

Assessing the hypotheses that variants in NOTCH1 but not TP53 were associated with ECS in OSCC by using a gene specific NGS mutational analysis approach in an ECS enriched cohort of primary OSCC tumour samples, genetic characterisation was performed after optimising Ion Torrent^{TM} sequencing for TP53 having previously optimised the method for NOTCH1 sequencing. Through this approach we identified both known and unknown variants in both genes with a significant association of NOTCH1 detected unknown variants with ECS, but no association of TP53 variants found. The benefits and drawbacks of the method used are discussed, followed by the significance of known and unknown variants for both genes, focussing on potential functional consequences.

6.4.1 Long range PCR for NOTCH1 and TP53 Ion Torrent^{TM} sequencing target enrichment

Long range PCR followed by Ion amplicon library preparation for both NOTCH1 and TP53 were initially optimized using random control DNA samples in a manner similar to that previously discussed in Chapter 5, Section 5.3. This approach allowed for full exonic coverage of both genes and minimised use of limited quantities of tumour DNA. We were aware that although analysis by Ion torrent software 2.2 was able to identify variants, the Ion torrent sequencer has been previously reported as being prone to miscounting the number of bases in homopolymer reads potentially due to both software and hardware/chemistry concerns [238] and the possibility of strand-specific errors remained after initial bioinformatics analysis. In the light of this, to further improve the confidence of variant calling we performed and independent form of bioinformatic analysis using an ActivePerl script that called Blast to identify sequence alignments and then, located the actual position of variants within alignments, allowing for errors like homolymer misreads. This analysis demonstrated persistent poor coverage in 10 samples that were then excluded from analysis as a result. The analysis also confirmed the absence of cross-contamination between samples for both genes and detected 9

NOTCH1 and 2 aTP53 variants, that had been missed by the Ion Torrent Suite. Variants called originally by the Ion Torrent Suite were also found. This confirmed the benefit of performing such an analysis while demonstrating the limitations of the Ion Torrent Suite Software.

However, only 46% of total variants identified by either Ion software or ActivePerl (or both) were confirmed by Sanger Sequencing. Although a number of studies have used the Ion torrent platform for sequencing samples, some studies have reported that its specificity may be low due to a high false positive rate, especially for INDEL detection [243, 244]. Due to the nature of the sequencing chemistry involved with the Ion Torrent platform, homopolymer regions have a greater indel calling error rate than other genomic regions. The literature suggests that it remains challenging to eliminate these false positives by defining a threshold based on mutation frequency [244] indicating the importance of using other methods for re-validation.

Furthermore, cancer specimens from biopsy specimens may contain significant proportions of normal tissue; blood cells, stromal components, and inflammatory cells [239]. This would result in dilution of tumour DNA with normal DNA. Somatic mutations may therefore only be at low frequencies. High coverage by Next Generation Sequencing can nevertheless detect such variants (<10% allelic frequency) [245, 246] but they are less easily detected by Sanger sequencing which relies on the rationing of electropherograms. We therefore used a cut off of 5% for either allele for variant calling on the basis of a 10% cut off excluding a significant number of variants (see section 6.3.4). In practice, the percentage of tumour cells was estimated by pathologists on hematoxylin and eosin (H&E)-stained slides with the samples used for this analysis demonstrating an average 70% tumour to normal tissue ratio based on histopathological reports as previously described in Chapter 3. This would be considered a high ratio compared to most standards [239]. However, evidence suggests that this means of assessment may be prone to bias and over-estimation [247] with significant inter-pathologist variation reported. More sophisticated methods have been described utilizing transcriptomic expression data [248] or copy number data [249] neither of which

was available as would be required for our samples. Furthermore, PurityEst has been described as a means of using next generation sequencing data to infer tumour purity level from the allelic differential representation of heterozygous loci with somatic mutations in human tumour samples but requires matched normal tissue which was not available in this instance [250]. Future studies evaluating genetic drivers of aggressive phenotype in OSCC would significantly benefit from the inclusion of such matched controls.

6.4.2 Low levels of TP53 nonsynonymous variants in OSCC ECS

Our study revealed 5 Sanger validated TP53 variants spread across the three histopathological subtype groups assessed (3 N-, 1 N+ECS-, 1 N+ECS+) with no significant associations found with either nodal status or ECS. 2 of the variants were in the DNA binding domain coding region of the p53 protein suggesting potential deleterious effects, however, three of the variants were coding SNPs that were deemed 'probably benign' after PSEP functional analysis. Indeed, an overall particularly low rate of TP53 variance of 12.5% (5/40) of Sanger validated variants was found that seems at odds with the reported HNSCC/OSCC literature with recent NGS studies finding 47-66% [21]. Multiple explanations for this difference are possible.

That our careful analysis detected more variants than were validated supports an excess of false positive variants rather than significant numbers of false negatives. It remains possible that sample admixture with a significant proportion of normal cellular DNA in relation to tumour DNA in certain samples, may have contributed to the low identified TP53 variant rate. However, if these technical drawbacks were prevalent and consistent, one would expect a similar effect on NOTCH1 variant frequency, which in contrast was actually found to reflect the rate from previous studies published in the literature. Also in support of this being a novel result was the earlier observation from the same sample cohort with regards allelic imbalance and a lack of copy number change associated with TP53. Therefore, an alternative explanation was considered.

The shortfall in detected TP53 variation could be due to selection of a genetically predisposed group, by assessing known variants within our cohort. Of the known SNPs already reported in the literature we identified a number that had been previously described associated with a range of cancers; stomach, colon, glioma and breast cancer [251-254]. The most commonly identified TP53 SNP, rs1042522, a missense polymorphism found to be present in 37 OSCC samples, has previously been reported to be associated with numerous cancers [255-257]. The SNP corresponds to exon 4 of TP53 causing an arginine to proline substitution at position 72 which resides in the SH3 binding domain of the p53 protein and could bear functional significance [258]. Following assessment of the most commonly identified variants, the distribution of deleterious variants across cases and any statistical significance associated with histopathological status was assessed. No association was found with ECS, confirming the presence of a subset of cases with low TP53 variation and NOTCH1 mutations. Assessment of the findings of Agrawal et al. suggested that a significant number of cases existed in their series with NOTCH1 mutations but wildtype TP53 [101]. As correlation of histopathological status with these samples was not discussed, we can speculate that on the basis of our findings that a significant number of these samples may have been from ECS +ve cases. Of note when considering our subset of cases with wildtype TP53 and mutant NOTCH1, other studies have demonstrated that in the presence of wildtype TP53, Notch1 activation increased the invasiveness of hepatocellular carcinoma cells and was correlated with advanced grade III HCC [259].

Low TP53 mutation rates have previously been reported in the literature to be associated with a number of tumour-specific, demographic or environmental factors. It has been previously reported that in HPV positive HNSCC TP53 mutation is uncommon with one study reporting only a 7% TP53 mutation rate amongst HPV+ve OSCC samples [260]. Furthermore, a recent meta-analysis found that the TP53 codon 72 polymorphism (rs1042522), found in 25/40 of our OSCC cases, was associated with HPV-related OSCC susceptibility cases but there was no association between this polymorphism and non-HPV OSCC cases, suggesting that the polymorphism may be a marker of HPV-related OSCC [261]. Our samples were all considered to be HPV-ve as detected by routine PCR from the regional clinical pathology laboratories but on the

basis of the discussion above, one theory to explain the low TP53 rate could be undetected HPV infection. However, HPV-DNA PCR remains the most accurate method for detection at present [262] bar specific HPV genotyping and sequencing.

Other possible factors that may have interplayed to create the low TP53 rate could be patient demographics, tumour site and smoking habits. Asian populations have been previously reported to demonstrate low TP53 mutation rates (Zannaruddin et al., 2013), with a mutation rate of 27% in a Malaysian cohort but with similar mutational frequencies of PIK3CA and HRAS to Caucasian populations [263]. A Taiwanese study similarly revealed a frequency of 32%. In contrast, other Asian cohort studies of TP53 mutational frequency have revealed higher rates akin to Caucasian reported rates [165] suggesting that demographic effects alone will not suffice to explain the low rates, particularly for our own Caucasian cohort. Site may influence TP53 mutation rate with another study on an Asian cohort revealing a 10.6% TP53 mutational frequency specifically in Tongue OSCC [264]. A further interplay with environmental factors is suggested by evidence that amongst an American cohort of never-smoker tongue OSCC showed a TP53 mutation rate of 19.6% [265]. There were 20 out of 50 samples in our cohort originating from the tongue but smoking data was unavailable at the time of analysis and may have provided further insight into possible reasons for the low level of TP53 variation observed. The possibility remains that a yet undetermined cohort specific factor/s may also have contributed. Further detailed multivariate clinical analysis with a larger sample size would be required to fully determine the factors involved.

6.4.3 NOTCH1 nonsynonymous variants associated with ECS

NOTCH1 variants ((7/40) *100=17% of overall cases) discovered by this study following both bioinformatics and sequencing validation were significantly associated with ECS ($p=0.0044$). Our reported rates of NOTCH1 variants are consistent with the recent literature for HNSCC / OSCC with rates reported from 9-16% [101, 102]. Furthermore, on detailed assessment of Agrawal et al., the potential for a similar subset of OSCC cases carrying wildtype TP53 but mutant NOTCH1 was identified.

The 7 NOTCH1 variants were located within exons 8, 20, 23, 25 and 33 with all bar the exon 33 variant being clustered within the N-terminal, extracellular EGF like domain in a similar manner to previous observations described for HNSCC (See Figure 19.). PANTHER-PSEP analysis of the 5 coding SNP variants, with 4 within these EGF like domains and one within the NICD portion of the protein, suggested likely deleterious effects on protein function. A further insertion was observed in exon 25 and a splice site nonsynonymous variant that was observed in a region specifically coding for the ligand binding domain (exons 8-11) of the EGF like repeats. 3 of our detected variants corresponded to the Abruptex region (exons 19–23) of EGF like repeats. Although little is known about EGF repeat contribution to NOTCH1 function, it has been shown that the integrity of the Abruptex region (EGF repeats 24–29) is required for suppression of NOTCH1 activity [221] and that variants within this region enhance NOTCH1 signaling [266]. Although EGF repeats facilitate ligand-mediated receptor activation, variants in the Lin-12/Notch repeat domain (LNR) coded by exon 25 in which we discovered 2 variants can lead to ligand-independent activity [267].

A large number of known single base substitutions and indels were identified in our cohort of OSCC samples for both genes. For NOTCH1, the most frequently identified SNP was rs201358664 that was found in 39 / 40 samples. Of note however, there were other SNPs (e.g. rs2229974, rs2229971, rs10521, rs76371972) found in the present OSCC case series, that although were not reported in ClinVar, have been reported in the literature to be associated with bladder cancer and acute leukemia [231, 232] but have not been described to be associated with OSCC. Although the majority of these SNPs correspond to missense substitutions, there were exonic known NOTCH1 deletions identified, such as; rs140371192, rs146350322 and rs41309766, potentially bearing significant functional consequences.

Our observations should be considered in the context of the spectrum of NOTCH1 mutations observed in HNSCC by NGS studies, which have been fundamentally different from those identified in haematopoietic tumours [268], as the majority have been (including our study) in the N-terminal EGF-like ligand binding domain and were predicted to alter the protein N-terminal to the transmembrane region.

The location and nature of these alterations suggests that NOTCH1 may be acting as a tumour suppressor in HNSCC. Although subsequent studies confirmed NOTCH1 inactivating mutations and showed NOTCH1 tumour-suppressor activity in OSCC cell lines [136], NOTCH1 dysregulation is likely to be more complex than simple loss-of-function [215]. Furthermore, approximately one third of HNSCCs displayed evidence of increased NOTCH1 pathway activation as compared with normal mucosa [215].

Rettig et al. were able to provide a further interesting insight into the potential role of NOTCH1 in OSCC by demonstrating that cleaved NOTCH1 expression (NICD immunohistochemistry) patterns correlated with NOTCH1 mutations and ECS in a cohort of HNSCC samples but interestingly did not observe a significant association between NOTCH1 mutation and ECS itself [269]. It is already established from a considerable body of evidence that Notch signalling is likely to be linked to the progression and maintenance of a tumour; however, Notch signalling alone is unlikely to directly cause the cellular processes that take place within a tumour without crosstalk between Notch signalling and other signals.

6.4.4 The Putative role of Notch in ECS

Key supporting evidence for either role of NOTCH1 in OSCC as a tumour suppressor or oncogene is summarised in Table 38. In the present study, our results of an association of NOTCH1 variants with ECS as well as the knowledge that significant copy number aberrations of NOTCH1 observed in the CGH data obtained earlier in our group (unpublished data, Lloyd B.), would suggest that NOTCH1 is acting as an oncogene in the ECS-enriched cohort of samples analysed. Some available evidence from the literature would also support this role. Yoshida et al. demonstrated that Notch1 expression correlated with both the T-stage and the clinical stage in OSCC, with loss of Notch1 expression correlated with the inhibition of cell proliferation and TNF-α-dependent invasiveness in an OSCC cell line and a γ-secretase inhibitor prevented cell proliferation and TNF-α-dependent invasion of OSCC cells i*n vitro* [270].

Table 38. Evidence supporting contrasting roles for NOTCH in OSCC. Literature exist to support both a tumour suppressor and oncogene role in OSCC. Our results support a more oncogenic role for NOTCH1 being a potential driver for ECS in OSCC. A dual role may be possible considering tumour heterogeneity and the complexity of Notch signalling and interaction with parallel signalling pathways within the tumour cell.

Oncogene	Tumour Suppressor
• NOTCH1 variants predominantly in the extracellular domain and may inhibit either trans-activation or cis-inhibitory ligand binding (present study)	
• aCGH analysis showing NOTCH1 amplification in our OSCC cohort (B.Lloyd unpublished data)	• WGS studies in HNSCC finding similar extracellular location to the majority of NOTCH1 mutations
• NOTCH1 variants significantly associated with ECS (Present study)	• NOTCH1 expression inhibited cell growth (Pickering et al., 2013)
• EGFR, inhibitory to Notch signalling, copy number loss in our a-CGH data (B. Lloyd unpublished)	• EGFR, inhibitory to Notch signalling, amplification associated with ECS (Michikawa et al., 2011)
• Notch1 expression correlates with advanced stage and inhibition prevented cell proliferation (Yoshida et al., 2013)	• NOTCH1 mutations not associated with ECS (Rettig et al., 2015)
• Intracellular domain of the Notch1 (NICD1) expression associated with ECS (Rettig et al., 2015)	

Rettig et al. found that intracellular domain of the Notch1 (NICD1) expression was associated with ECS in OSCC, supporting a more oncogenic role for NOTCH1 and Notch signalling [269].

Supporting a tumour suppressive role, Pickering et al. expressed the cleaved/activated form of NOTCH1 (ICN1) in five cell lines that harbour missense or truncating NOTCH1 mutations [136]. They found the relative fraction of ICN1-expressing cells decreased to less than 40% of its original value in 16 days. Furthermore, by using mouse xenografts in an orthotopic tongue cancer model they demonstrated both ICN1 and full-length

Notch (NFL) caused a dramatic reduction in tumour size compared with the vector control. Although Rettig et al. reported NICD1 expression to be associated with ECS, however did not find any association of NOTCH1 mutation with ECS in a cohort of OSCC cases suggesting a tumour suppressive or dual functional role for NOTCH1 variation [269].

Further insights have been gained from EGFR signalling. Notch signalling does promote terminal differentiation of keratinocytes and is negatively regulated by the EGFR pathway [271]. The epidermal growth factor receptor (EGFR) is a transmembrane tyrosine kinase receptor belonging to the HER/erbB family and is overexpressed in up to 90% of HNSCC [31]. High EGFR gene copy number has been reported in 10% to 58% of HNSCC [135]. In HNSCC activating EGFR mutations are rare [21]. Overexpression of EGFR and high EGFR gene copy number are associated with poor prognosis and radio resistance [272] . Importantly a study demonstrated EGFR copy number aberration to be associated with ECS [58]. These observations would support NOTCH1 acting as a tumour suppressor gene in certain circumstances. However, on retrospective analysis of the a-CGH data from our cohort (chapter 3, Figure 4.) EGFR was observed to be within a region of loss with no variants with severe consequences also observed within the gene in our whole genome sequencing OSCC cell line analysis. Bearing in mind our results and the other conflicting evidence that exists, a dual role may be possible.

Finally, evidence suggests that the four Notch proteins that have been identified have distinct activities and outcomes depending on cellular context, although it is thought that the mechanistic details of action are similar [273]. As the majority of the variants we found in OSCC as well as in previously published studies, were in the extracellular domain of NOTCH1 [101], more detailed investigation of the functional effect of extracellular domain variant in oral cancer needs to be performed. These studies would focus on the possible impact of perturbations in structure at this site on function and may provide more evidence as to whether Notch is acting as a tumour suppressor or oncogene in contributing to the formation of aggressive phenotype.

Chapter 7 : The effect of Notch inhibition on aggressive phenotype in OSCC cell lines

7.1 Introduction

Notch signaling has been discussed previously in this study (see section 5.1.2); given our detection of NOTCH1 ECD variants associated with ECS in OSCC (see section 6.3.6), as well as considering their predicted protein structural effects (see section 6.3.5), a number of other aspects are worthy of consideration. 1. What role do NOTCH1 ECD variants have to play in the pathogenesis of ECS in OSCC? 2. Could inhibitory Notch signaling affect key aggressive cellular characteristics and be of potential therapeutic use in OSCC? and 3. The Notch family of proteins consists of 4 transmembrane proteins (NOTCH1, NOTCH2, NOTCH3, & NOTCH4) and each could have an impact on tumourigenesis [274] through changes at different levels of NOTCH signaling; their ligands, the receptors themselves or their downstream targets. Evidence suggests that the biologic effects of Notch signaling vary in cell type–dependent fashions [275]. Furthermore, microarray analysis of Notch signaling pathway member expression amongst a wide variety of tumours has revealed potentially differential roles for each Notch pathway member [276]. Could differential effects exist in OSCC?

Considering these points as well as technical drawbacks at this stage of the study, as the detailed characterization of the role of ECD variants would require larger genetic characterization studies of multiple cell lines alongside comprehensive ligand-receptor binding studies, we elected to focus on the latter two points. We set out to perform an initial screen of the functional effect of NOTCH1 inhibition in oral cancer to better understand the biology of aggressive phenotype in OSCC by assessing key *in vitro* markers (proliferation, migration and Notch expression) using both invasive and non-invasive OSCC cell lines and pharmacological inhibition of Notch signaling.

7.1.1 Notch pathway inhibition in HNSCC/OSCC

Notch Signaling has been previously demonstrated to be associated with an aggressive metastatic phenotype *in vitro* cancer models [277, 278]. Yoshida et al. reported Notch1 and intracellular Notch domain (NICD) upregulation in OSCCs and that expression was characteristically localized at the invasive tumour front [213]. They also demonstrated that γ-secretase inhibitor (GSI) prevented cell proliferation and TNF-α-dependent

invasion of OSCC cells *in vitro*. Furthermore, on clinicopathological analysis, Notch1 expression correlated with both the T-stage and the clinical stage [213]. In addition, Weaver et al. reported a gene expression RNA analysis of 19 OSCC samples and demonstrated that Notch pathway upregulation was significantly correlated with patient mortality status and with expression of the pro-invasive gene FGF1. Furthermore, they demonstrated that *in vitro* HNSCC cell Notch activation increased FGF1 transcription and induced an increase in cell migration and invasion, abrogated by FGF1 knockdown [279].

Some further insight can be gained from the consideration of cancer stem-like cells (CSCs) that are likely to play a role in tumour recurrence and confer resistance to anti-cancer therapy treatment [280]. Examining isolated CSCs (oralspheres) from HNSCC cells for tumourigenicity and in immunodeficient mice, aggressive tumour growth and higher expression of Notch1 protein was observed as compared to parental cells [281].

7.1.2 Genetic manipulation of Notch pathway signaling

Sun et al. in 2014 performed an integrated analysis on a cohort of HNSCC tumours assessing expression, copy number, methylation, and mutation analyses. They observed copy number levels of JAG1 and JAG2 ligands and the NOTCH3 receptor were significantly increased in HNSCC, and that no NOTCH pathway gene promoters were significantly hypomethylated in HNSCC tumours compared with normal mucosa. To assess overall NOTCH pathway activation, they analyzed downstream effectors HES1 and HEY1 finding ~32% of HNSCC tumours showed HES1 and/or HEY1 expression consistent with NOTCH activation. On performing exonic sequencing on these HNSCC tumours, 5 novel inactivating NOTCH1 mutations in 4 of the 37 tumours analyzed were identified, but none of the mutations were associated with HES1 or HEY1 activation and thus they were able to characterize a subset of wild-type NOTCH1 HNSCC tumours with increased HES1/HEY1 expression. The authors suggested a bimodal NOTCH pathway pattern of HNSCC alterations, with a small subset harbouring inactivating NOTCH1-specific mutations, and a larger subset having other NOTCH pathway alterations, including increased receptor or ligand expression or gene copy number, as well as downstream pathway activation [215]. They went on to demonstrate

that in NOTCH1 wild-type HNSCC cells, siRNA NOTCH1 or HEY1 inhibition significantly decreased cell growth. Furthermore, application of GSI-XXI, a NOTCH pathway inhibitor, to HNSCC cells also inhibited cell growth quantified by a CCK-8 colorometric assay. Consistent with this study, Pickering and colleagues discovered other NOTCH pathway gene changes in copy number and expression, including increased NOTCH receptor ligand JAG1, JAG2 and MUMB copy number changes that were associated with elevated mRNA levels [136] although overall protein levels were not assessed in the study.

Wang et al. demonstrated in pancreatic cancer cell lines that down-regulation of Notch-1 expression by siRNA inhibited cell growth, induced apoptosis and induced cell cycle arrest in G_0-G_1 phase, while overexpression of Notch 1 by cDNA transfection promoted cell growth and inhibited apoptosis [282]. NOTCH1 and NOTCH2 siRNA mediated knockdown in AML cell lines resulted in a suppression of proliferation, induction of apoptosis and a reduction of mTOR protein with similar findings observed in T-ALL cell lines [283]. siRNA NOTCH1 knockdown in glioma stem cells (GSCs) as part of an investigation of glioblastoma multiforme, demonstrated inhibitory effects on proliferation. Furthermore, Notch-1 silenced GSCs engrafted on Balb/c nude mice showed a reduction in oncogenicity, with direct intratumoural injections of Notch-1-siRNA significantly delaying the growth of pre-established tumours in nude mice [284].

7.1.3 Pharmacological inhibition of Notch pathway signaling

Methods of chemical inhibition target essential Notch pathway signaling steps: expression of ligands, ligand ubiquitination and trans-endocytosis, expression of Notch receptors, ligand-receptor binding, heterodimer dissociation during Notch activation, ADAM-mediated cleavage of Notch, subsequent ubiquitination and endocytosis of the γ-secretase substrate, γ-secretase-mediated cleavage of Notch, assembly of the coactivator complex with Notch and CSL, heterodimerization of Notch transcriptional complexes, Notch post-translational modifications and finally expression of Notch targets [285]. Gamma-secretase inhibitors are commonly employed *in vitro* and *in vivo* preventing the release of the notch intracellular domain (NICD) [286]. Monoclonal

antibodies directed against Notch receptors or their ligands, Notch receptor or ligand decoys and blocking peptides inhibit Notch mediated transcription have also been used.

7.1.4 γ-secretase inhibitors

γ-secretase acts upon a variety of substrates including amyloid precursor protein (APP) [287]. γ-secretase is a high molecular weight protein complex found on the cell membrane with a varying estimated size between 250 and 2000 kDa, including presenilin (PS), nicastrin, Aph-1, and Pen-2 parts [288]. Increasing evidence suggests that assembly of the active PS / γ-secretase complex takes place in a chronologic order. Firstly, Aph-1 and nicastrin forms an intermediate complex. After that, Aph-1 and Pen-2, in turn, function as chaperones assisting in PS / γ-secretase complex assembly, trafficking and PS endoproteolysis [289]. PS contains the catalytic site of the PS / γ–secretase complex and has two forms, namely PS1 and PS2, encoded by two different genes, PSEN1 and PSEN2, where Presenilin 1 plays an important role in proteolytic processing [290]. PS is endoproteolyzed into two subunits- N-terminal fragment (NTF) and C-terminal fragment (CTF), and these fragments are believed to represent the active form of PS [288]. Presenilin appears to be required for the transmembrane cleavage of Drosophila Notch and biochemical studies suggest similar for mammalian Notch proteins [291]. Knockout of presenilin-1 in mice reduces proteolysis of the Notch transmembrane region. Presenilins auto cleave to form the active site of γ-secretase [292].

Chemical inhibitors of γ-secretase (GSIs) mostly interact with the active site of the protease GSIs are classified into two types,: (1) aspartyl proteinase transition-state analogs as peptide isosteres mimicing the transition state of substrate cleavage by γ-secretase binding competitively to presenilin catalytic active sites; and (2) small molecule non-transition-state inhibitors with a different binding site from the active site, possibly at the interface of the γ-secretase complex dimer [293]. [223] and have been explored as therapies for diseases including: Alzheimer’s disease [294] and leukemias [295] [296]. Gamma-secretase (γ-secretase) inhibitors are able to inhibit cell growth and induce cell apoptosis [297].

7.1.5 Notch pathway inhibition in OSCC

DAPT, N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester, is a widely used non-transition state analog, dipeptide inhibitor of the benzodiazepine type [298],

reducing proliferation with an ADAM cleavage site dependency [299]. For Head and Neck cancers, DAPT inhibited the growth of human tongue carcinoma, human nasopharyngeal carcinoma drug-resistant cell line (CNE2) and Tca8113 cells by inducing G0-G1 cell cycle arrest and apoptosis [300] and enhanced cisplatin-sensitivity of Tb3.1 cells [301]. Notch signalling in NPC cells is down-regulated by DAPT increasing radio sensitization, [302]., DAPT reduces head and neck cancer stem cell (CSC) populations and tumour self-renewal *in vitro* and *in vivo* [303]. Given the complexities of NOTCH signaling both within and between cells, the observation of unexpected or adverse outcomes is unsurprising. Off-target Notch-associated effects in the early GSI trials for Alzheimer's disease led to widespread gastrointestinal toxicity [304]. Nevertheless, DAPT holds great utility in the laboratory setting.

The primary OSCC cell lines used previously, particularly including ECS positive lines, provided opportunities to investigate the biology of aggressive subtypes of OSCC. NGS studies of NOTCH1 on these cell lines had demonstrated one particular primary node+ ECS+ cell line containing the rs61751543 SNP. This is a rare non synonymous NOTCH1 extra cellular domain (NECD) polymorphism previously reported to be associated with other cancers [234] and found to be 'possibly damaging' on our earlier evolutionary functional analysis (section 5.3.3). This cell line (Liv7K) was chosen to examine further with regards its cellular biology related to aggressive behavior including cellular migration and proliferation. DAPT was chosen as a means of NOTCH pathway inhibition due to its ready availability with extensive supporting data in the literature.

7.2 Hypotheses and aim

Hypothesis: ***Inhibition of the Notch pathway by DAPT, a γ-secretase inhibitor, would result in decreased cell proliferation and migration in an ECS +ve cell line.***

Aim: The aim was to assess whether Notch pathway interference mediated by γ-secretase inhibition impedes cell growth and migration in both invasive and non-invasive OSCC cell lines.

7.3 Results

7.3.1 Measurement of DAPT toxicity on OSCC cell lines

Inhibitory concentration 50 (IC50) experiments to determine specific DAPT cell line toxicity was carried out as described in section 2.10.1 using the Liv7K cell line and DAPT incubation for either 24 hours or 72 hours with concentrations ranging from 10nM to 50μM in DMSO (DAPT solvent) at 0.5%. Absorbance values were then recorded using a spectrophotometer to assess cell viability. Figure 20. demonstrates the mean % absorbance compared to DMSO control readings from three repeats with standard deviation bars also shown. All concentrations of DAPT tested, over both time periods, provided absorbance readings that were not significantly different from each other with overall reading variation likely to be introduced by experimental error from preparation of serial dilutions. Cell cytotoxicity would have led to significant drops in recorded absorbance which was not observed. The mean absorbance for all obtained values for the 24-hour treatment was lower compared to the 72 hour treatment, consistent with a higher level of growth and proliferation with time giving a higher absorbance rate.

7.3.2 The effect of DAPT on OSCC Cell proliferation

Based on IC50 values and evidence from the literature of DAPT gamma secretase inhibitory efficacy over a prolonged period [305], a crystal violet cell proliferation assay was set up over 4 weeks with or without a DAPT concentration of 25 μM (in culture media that was replaced weekly) on OSCC cell lines. The optical density at 560nm was recorded each week (week 0, 1, 2, 3, 4) in triplicates (standard error of the mean shown), following adjustment for background, as a measure of cell proliferation being directly proportional to the number of viable cells at each time point. Experiments were performed in triplicate with mean optical densities calculated for each time point. Figure

21. demonstrates that both invasive cell lines (BHY and Liv7K) showed impaired proliferation when incubated with DAPT from 2 weeks with final optical densities after 4 weeks of 1.98 and 1.92 respectively (compared to controls of 2.33 and 2.32). On the other hand, H357, the non-invasive cell line, showed no significant difference compared to control at 4 weeks (2.32 v 2.31).

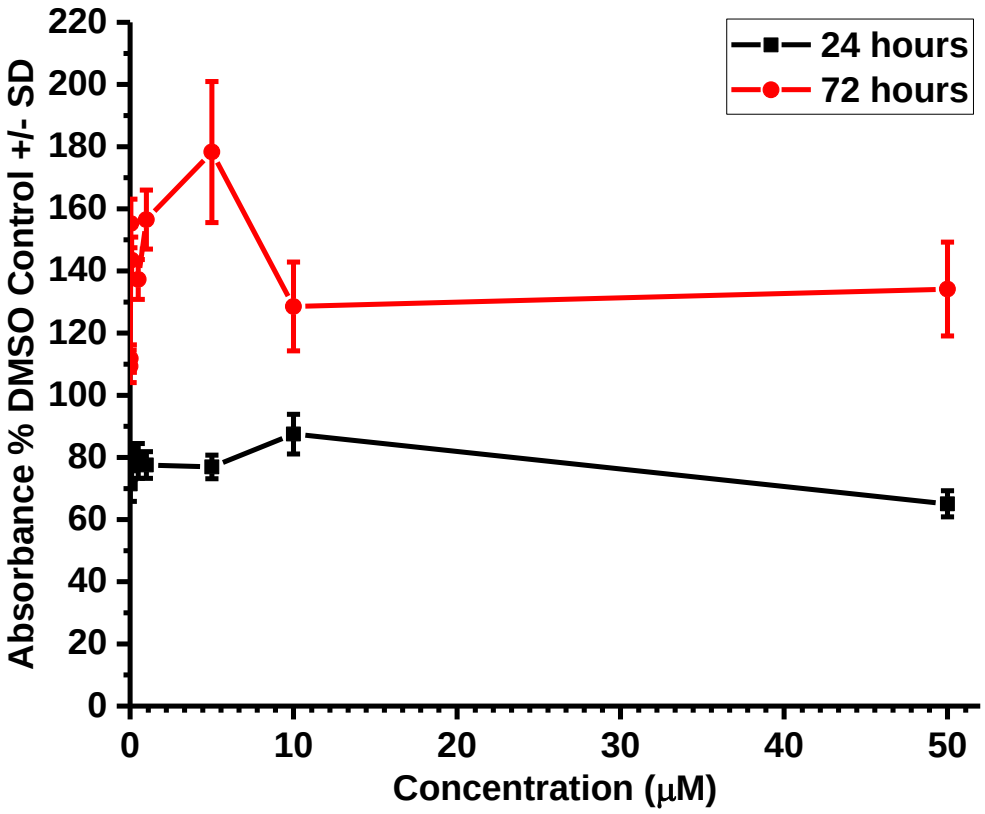

Figure 20. Liv7K Cell Viability following exposure to varying concentrations of DAPT over different time periods. Graph represents values of DAPT concentration (μM) against mean % absorbance of DMSO control values obtained following 24 hour incubation (black) and 72 hour incubation (red). Concentration values from left to right correspond to 50, 10, 5, 1, 0.5, 0.1, 0.05 and 0.01 μM DAPT dissolved in 0.5% DMSO.

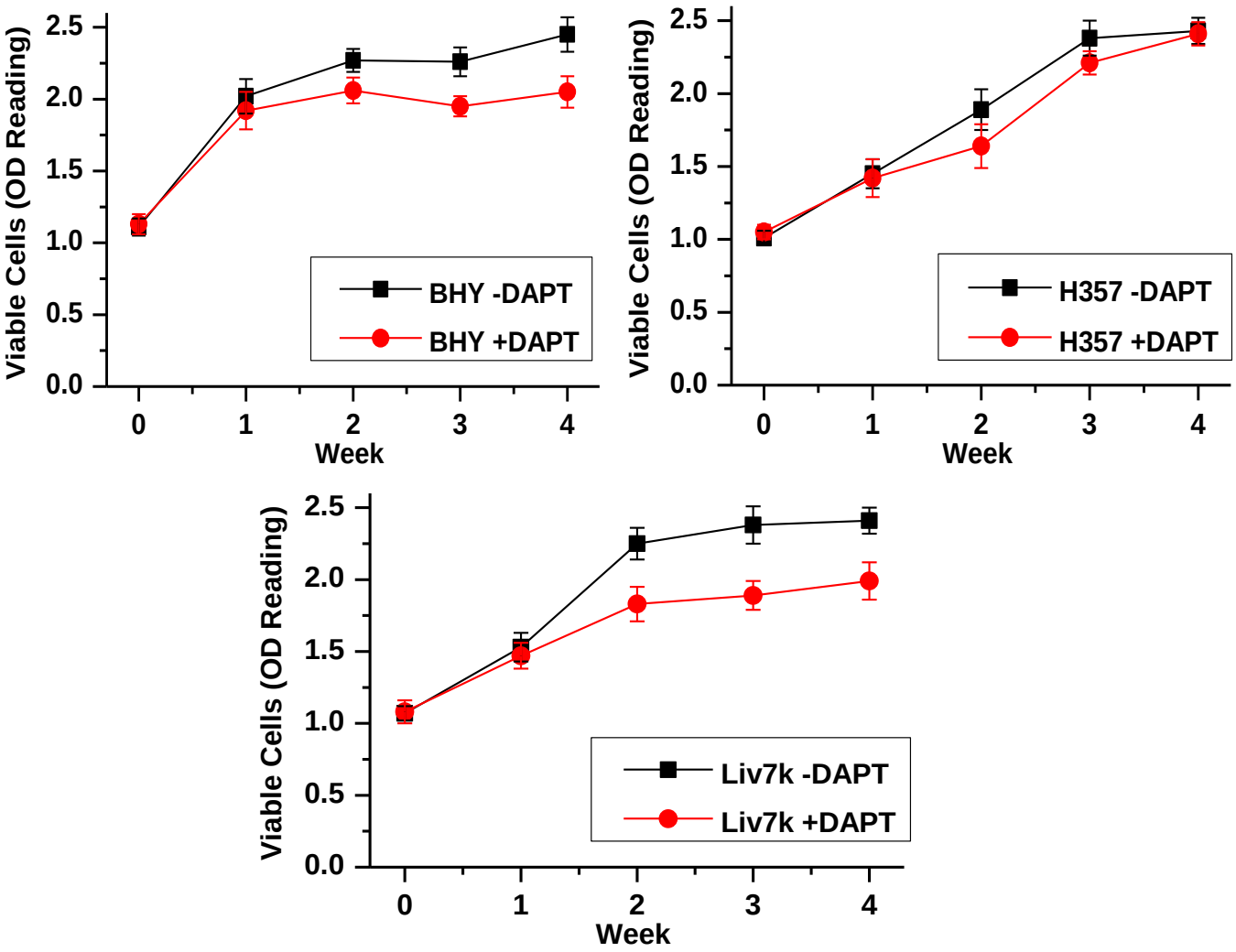

Figure 21. BHY(invasive), H357 (non-invasive) and Liv7k cell proliferation in the absence or presence of DAPT. The graphs illustrate crystal violet proliferation assay results following incubation of BHY cells for 4 weeks in the presence or absence of 25μM DAPT (replenished on a weekly basis). Optical density at 560 nm readings were taken in triplicate and the mean (+/-SEM) is displayed. Recordings made with DAPT incubation are shown by the red line and control by black.

7.3.3 The effect of DAPT treatment on NOTCH receptor expression

To explore the effects of Notch inhibition on Notch receptor expression on OSCC cell lines, western blots were performed for Notch1-4 (method described on section 2.11.2), extracted from the 3 cell line lysates (BHY, Liv7k and H357) on a weekly basis for 4 weeks, in the presence or absence of prolonged DAPT treatment (25 μM) administered weekly.

Western blots for Notch1-4 on control cell lysates without DAPT treatment demonstrated the presence of NOTCH 1(size 125 kDa),3 (size 244 kDa) & 4 (size 210 kDa) in BHY, Liv7K and only 3 & 4 in H357 lysates, but a complete absence of NOTCH 2 in all three cell lines (Fig. 22). In the light of this surprising finding, an instant blue stain was performed on the polyacrylamide electrophoresis gel prior to membrane transfer confirming presence of adequate separated protein products as well as a Ponceau Red stain of PVDF membrane confirming adequate and uniform protein transfer across different cell lines. To test NOTCH2 antibody functionality, HELA cell lysates in two different dilutions (1/500 and 1/1000) were used as a control alongside lysates from all 3 cell lines in the same experiment, confirming the ability of the antibody to detect NOTCH2 (Figure 23.). This suggested a possible deletion of NOTCH 2 in all cell lines examined. Of the detected Notch proteins (1, 3 & 4) in these control conditions, no variation was observed in expression with time over the four-week period.

Following DAPT treatment however, western blots for Notch1-4 on OSCC cell line lysates revealed a response only in invasive cell lines. A decrease in expression of NOTCH 1, 3 & 4 was observed in Liv7K cells, evidenced by absence of NOTCH1 and 3 bands by week 3 and NOTCH 4 by week4 (Figure 22.). BHY cell lysates revealed an absence of NOTCH 4 expression by week 2 but no significant change in NOTCH 1 or 3 (Figure 22.). The non-invasive cell line H357 showed no difference in NOTCH 3 or 4 protein expression following DAPT treatment (Figure 22.).

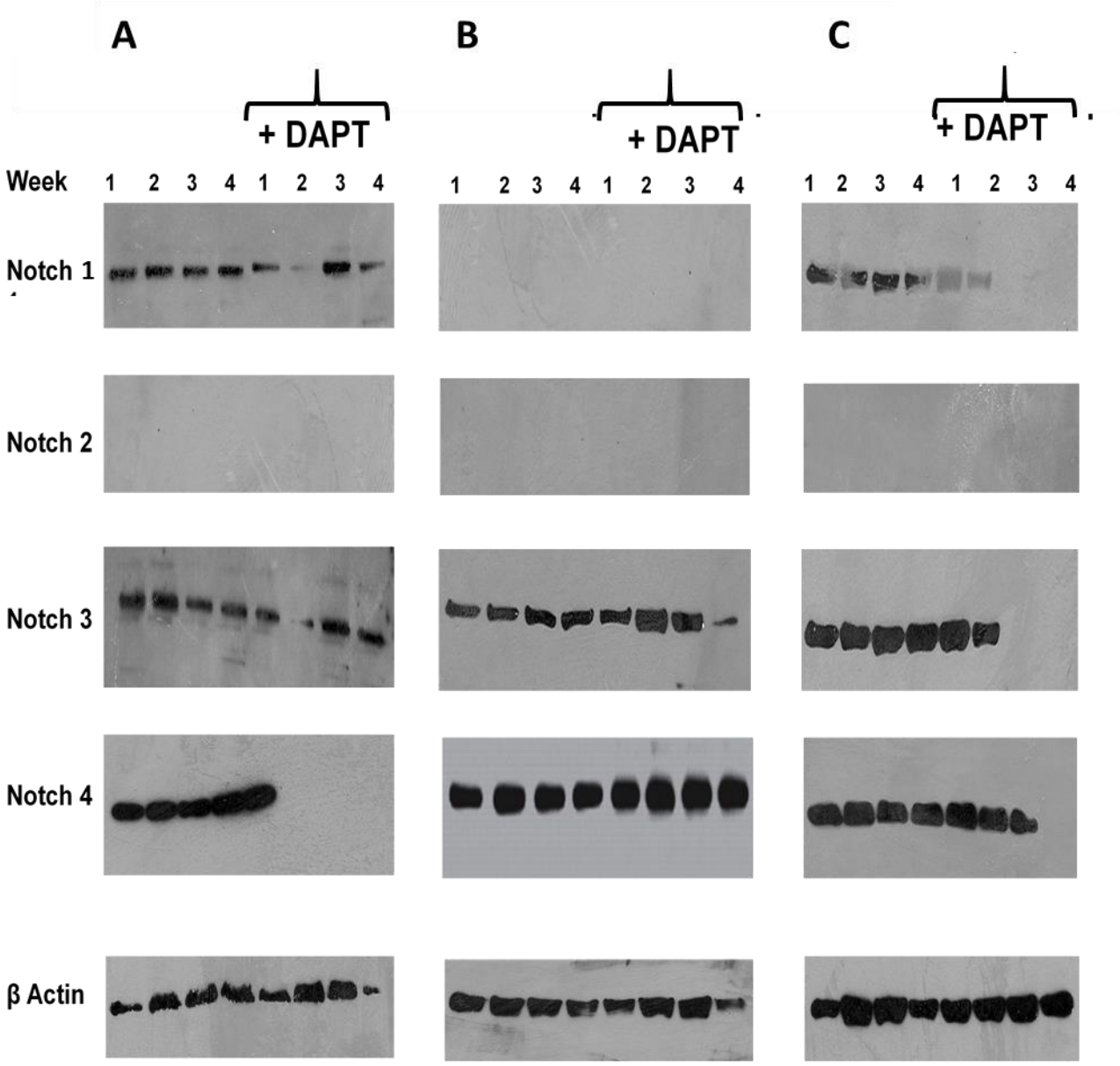

Figure 22. Western blot of NOTCH 1–4 in A. BHY cell line B. H357 C. Liv7K cell line in relation to the presence/absence of DAPT. Western blots were performed on cell lysates in a conventional manner (section 2.11.2). Protein lysates were extracted weekly for upto 4 weeks duration with the cells either incubated with or without DAPT (25 μM). Experiments were performed in triplicate with representative blots shown above. β-actin was used as a loading control.

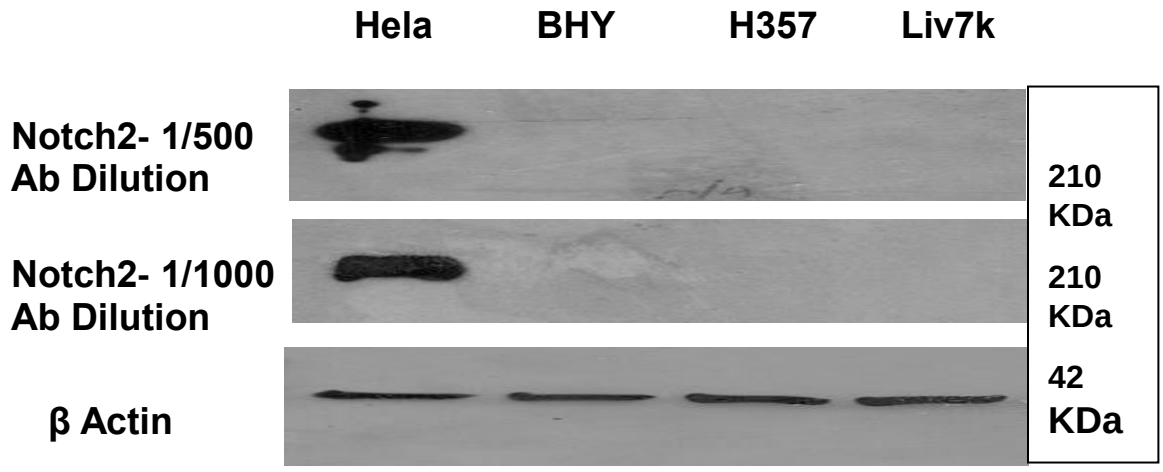

Figure 23. Western blot of NOTCH 2 in HELA cells and OSCC cell lines at differing antibody dilutions. Experiments were performed in triplicate with representative blots shown above. β-actin was used as a loading control.

These results demonstrate that DAPT inhibits NOTCH expression in invasive OSCC cell lines so it was assessed whether this was taking place through the expected inhibitory effects on the y-secretase. Western blots were performed in a similar manner to previously on all OSCC cell line lysates over 4 weeks but also over the shorter time frame of 24 hours, in the presence or absence of DAPT but using antibodies against amyloid precursor protein (APP). APP is an alternative protein target cleaved by y-secretase (described previously in section 7.1.4. and expected to increase in expression if DAPT was exerting its inhibitory effects on NOTCH through y-secretase. APP western blot results for protein lysates from Liv7K cells are shown in Figure 24. as a representative example. Control HELA cell lysates confirmed the detection of APP at the expected fragment sizes (95 & 110 kDa fragments). Liv7K protein lysates in control conditions in the absence of DAPT showed no change in APP levels with time. DAPT treated cells also showed no difference in APP expression over 24 hours incubation, however a gradual APP increase did occur from weeks 2 - 4 compared to control. This supports the conclusion that the effects of the DAPT inhibitor observed on the Notch

pathway in earlier experiments are likely to be as a result of an inhibitory action on y-secretase.

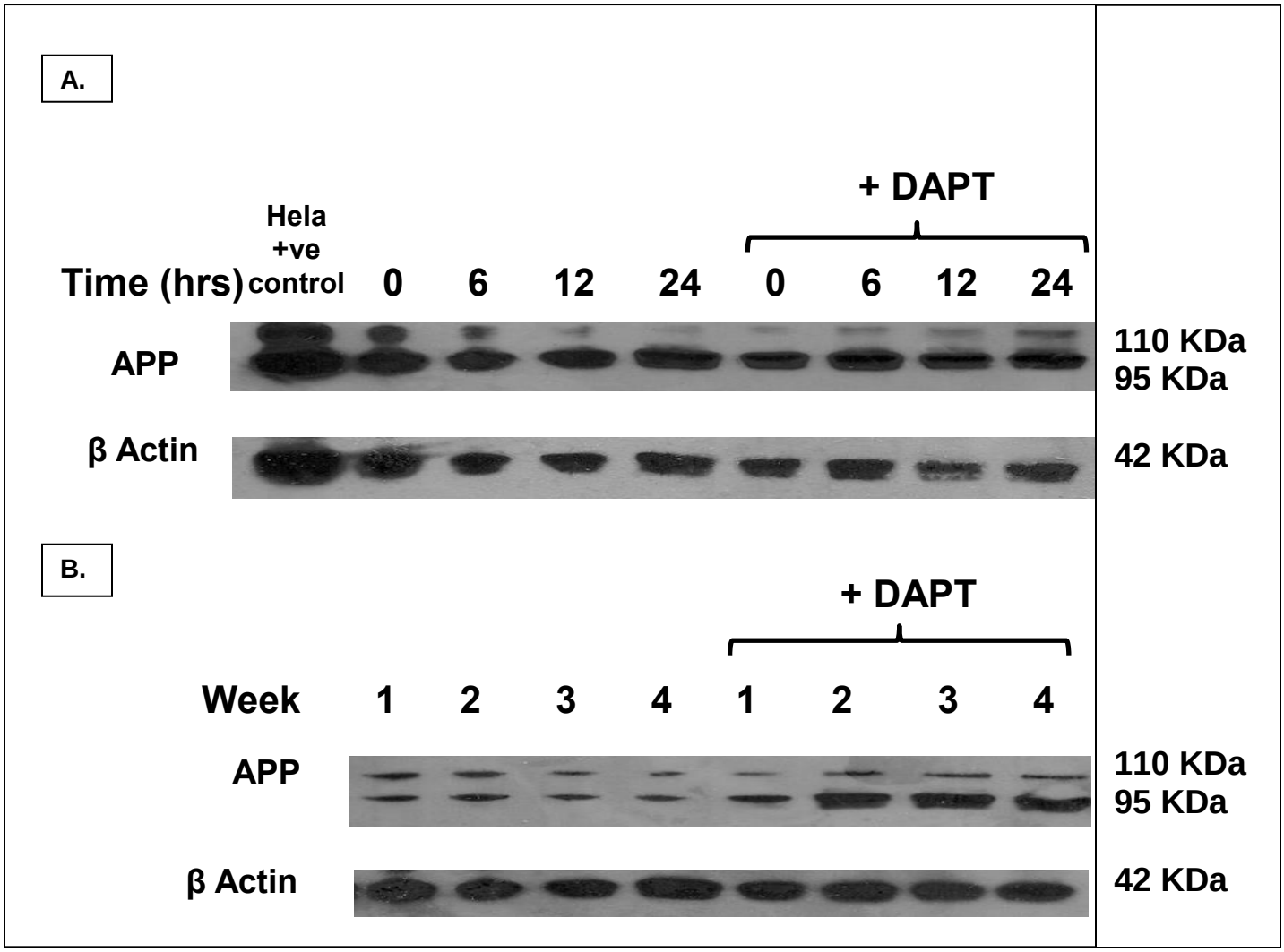

Figure 24. Western blot of amyloid precursor protein (APP) on protein lysates from Liv7K cells in the presence or absence of DAPT. A. Western blots of Liv7K cell protein lysates (+/- DAPT 25 µM incubation) performed in a standard manner on lysates extracted from 0, 6, 12, 24 hour time points. Hela cells were used here as a positive control. **B.** Western blots of Liv7K cell protein lysates (+/- DAPT 25 µM incubation) performed in a standard manner on lysates extracted from weekly time points up to 4 weeks. β-actin was used as a loading control for both blots.

7.3.4 The effect of DAPT on OSCC Cell line migration

To assess the effect of NOTCH variation and pharmacological NOTCH inhibition on cell migration, a cell motility / migration assay was set up in the standard manner using a 'scratch assay' monitored by phase contrast live imaging as previously described in section 2.10.6. BHY, H357 and Liv7K cells were assessed over a 60-hour period, either in the presence or absence of incubation with 25μM DAPT. Images were taken at 1-hour intervals and initially inspected to assess qualitative differences. Images from the start of the experiment and end were then assessed by ImageJ software analysis for quantification of the extent of migration by calculating the mean % wound closure (triplicate experiments) for each cell line, in DAPT present and absent conditions (Figure 25.). Liv7K showed the greatest final migration (80.4 %), followed by BHY (39.8%), (both invasive cell lines), followed by the non-invasive cell line H357 (9.1%). Liv7K cells were able to nearly completely close the 'wound' at the end of the 60-hour course of imaging, however a substantial decrease in cell migration was observed following DAPT incubation (Figure 25.) of Liv7K. Indeed, quantification of the effect of DAPT incubation on Liv7K migration revealed significant migratory inhibition (30.2% v 80.4%; $p<0.05$ on student's t-test) but no significant difference was observed with DAPT on BHY or H357 cells.

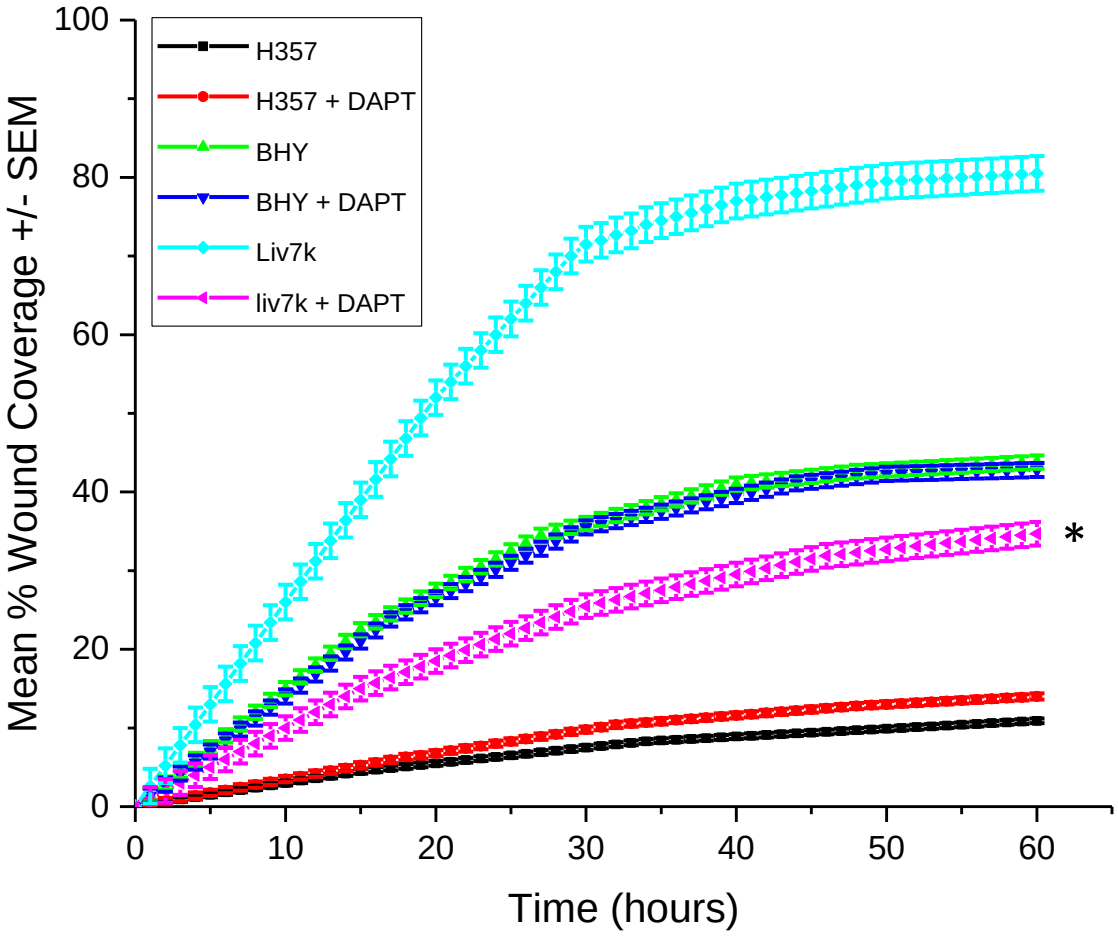

Figure 25. Migration assay observations for OSCC cell lines in the presence or absence of DAPT. Mean % wound closure with both treated (DAPT in 25 µM concentration) and untreated OSCC cell line cells (BHY, H357 and Liv7k) from 0 to 60 hours is demonstrated. ImageJ quantification of % wound closure was performed on images recorded from all groups over 1 hour time points. Mean % wound coverage +/- SEM shown from 3 replicates. * = $p<0.05$ on Student's t test compared to Liv7K without DAPT.

7.4 Discussion

We compared invasive (BHY), non-invasive (H357) primary OSCC cell lines and a primary OSCC cell line with representative histopathology (node positive ECS positive) found previously by this study to have a rare form of extra cellular domain non-synonymous polymorphism (rs61751543). The main aim was to assess whether Notch signaling inactivation, mediated by γ-secretase inhibition, impedes cell growth and migration by invasive and non-invasive OSCC cell lines and how Notch may thus play a role in the development of aggressive phenotype. Previous animal studies demonstrated that continued DAPT treatment at a concentration of 10μM maximally blocked Notch signalling after 28 days [305]. Pilot studies in our lab showed that DAPT at a 25 μM concentration for 72 hours did not significantly inhibit Notch activation in OSCC cell lines. Therefore, DAPT was administered here at 25 μM concentration for 28 days. Insufficient opportunity was available for investigating the exact role of NOTCH1 variants. However, OSCC cell proliferation, migration and Notch receptor expression following notch inhibition with DAPT, only decreased in invasive cell lines, suggesting the importance of Notch pathway activation for their behaviour.

7.4.1 NOTCH1 variation and aggressive OSCC

Proliferation assays in standard cell culture conditions (-DAPT) and for the time period examined, revealed no significant differences between BHY, Liv7K and H357 cells with all cell lines proliferating to the same level after a 4-week period. Migration was significantly different between the cell lines however with Liv7K displaying a significantly greater rate of migration compared to BHY, which was in turn greater than H357. Liv7K was already known to be associated with more aggressive behavior using migration assays [306]. As DAPT had no effect on BHY cell line migration, we speculated that this could either be due to the absence of NOTCH1 variants within the cell line or that the cell line may have been from an ECS-ve patient. The Notch genetic profile of BHY was undetermined and future studies sequencing NOTCH1 in both these cell lines would allow more definitive comparisons to be drawn. Western blot analysis in the absence of DAPT also revealed differential NOTCH receptor profiles between invasive and non-

invasive cell lines with Liv7K and BHY demonstrating NOTCH 1, 3 and 4 expression but H357, although a non-invasive cell line, demonstrating an absence of NOTCH1, with NOTCH 2 absent from all cell lines. The absence of NOTCH1 within our negative controls confounded further interpretation of the effect of NOTCH1 variants. The absence of NOTCH1 from H357 that we detected would be in keeping with the literature as *Yap et al.* demonstrated a frameshift mutation in H357 which causes an eight-nucleotide deletion, making NOTCH1 undetectable [189]. Given the phenotypic difference between H357 and Liv7K, there may be other pathways that contribute to their contrasting phenotypes.

Surprisingly, Notch2 protein was absent from all of our cell lines. Protein lysate generated from Hela cells and the three cells were used for western blot using different NOTCH 2 antibody concentrations. Previous published studies with the same Notch2 antibody had found normal detection by the antibody in Hela cells [307]. A strong band of Notch2 protein was produced from Hela cells using two different concentrations confirming absence of Notch2 from the oral cancer cell lines. Further investigation, such as by investigating whether NOTCH2 deletion is present within the DNA of these cell lines to account for the lack of Notch 2 protein observed, would need to be performed before coming to a definite conclusion. Interestingly, studies in the literature further highlight the possibility of opposing functions of NOTCH1 and NOTCH2. Using a Kras-driven endogenous non-small cell lung cancer mouse model demonstrating NOTCH1 deficiency led to a reduced early tumour formation and lower activity of MAPK compared with the controls but unexpectedly, NOTCH2 deletion resulted in dramatically increased carcinogenesis and increased MAPK activity [106]. A complex balance may exist in terms of differential Notch signalling in OSCC.

The data here does not conclusively demonstrate the effects of NOTCH1 but provides a basis for further work. Rettig et al. combined genetic and expression analysis of HNSCC demonstrated three distinct staining patterns when using an antibody specific for Notch intracellular domain 1 (NICD1): positive/peripheral (47%), positive/nonperipheral (34%), and negative (19%) and NICD1 expression was found to be associated with NOTCH1 mutation status ($P < 0.001$). Most NOTCH1-wild-type tumours were peripheral (55%),

whereas mutated NOTCH1 tumours were most commonly negative (47%). Interestingly nonperipheral tumours were more likely than peripheral tumours to have extracapsular spread. A further study of NICD1 by IHC in our ECS enriched OSCC cohort may prove useful in understanding any functional effects of NOTCH1 genetic variation in OSCC.

7.4.2 Relation of Notch signalling to aggressive phenotype

The importance of Notch signaling in contributing to an aggressive phenotype was supported by our experiments inhibiting Notch signaling with DAPT. Proliferation assays revealed a substantial decrease of both Liv7K and BHY cell growth by treatment with DAPT (25 µM) for four weeks compared to their respective control non-treated cell lines. Migration assays also showed differences with Liv7K cells (derived from an ECS +ve patient) displaying significant decreased cell migration as a result of DAPT treatment. However, BHY, an invasive cell line, was not affected. Furthermore, western blot analysis of NOTCH 1-4 following DAPT incubation revealed decreased expression of Notch 1 and 3 by week 3 and Notch 4 by week 4 in Liv7K cells but only Notch 4 by week 2 in BHY cells (H357 revealing no change in Notch receptor expression from DAPT). The differential findings between the two invasive cell lines, BHY and Liv7K, with regards both proliferation and Notch receptor expression could be due to either the level of Notch pathway activation within the cell line itself and/or the efficacy of the DAPT administered to actually inhibit Notch signaling. DAPT may not have worked as well on BHY cells. As discussed earlier, H357 demonstrated an absence of NOTCH1 on western blot analysis, potentially meaning DAPT may not have been able to display its full inhibitory effect on phenotypic assays (even in the light of NOTCH3 and 4 still being present).

To conclude that DAPT was having its observed effects through the γ -secretase pathway and that our observed phenotypic effects were as a result of Notch signaling inhibition would require further experiments. Inhibition of γ –secretase activity, by DAPT was assessed by western blots analysis with an antibody directed against APP at two different time points. No significant differences between treated and untreated cells

were observed at early times but there was an increase in APP protein in treated cells by week 2 onwards suggesting successful *γ*-secretase inhibition by DAPT at this time point, corresponding to when effects were observed with regards proliferation and Notch receptor expression. Although there are previous studies which suggest that DAPT inhibition requires a prolonged time course of incubation, such as 28 days [308], other studies report a more rapid time frame for efficacy with one study reporting a 15 minute 10μM incubation of DAPT in primary rat peritoneal mesothelial cells blocking transforming growth factor (TGF)-β1 induced Notch signaling activation [309]. In our study, inhibitory DAPT activity may have still been present in the early stages of DAPT incubation to explain our migratory assay observations (significant difference at 60hours) but may have simply been of too low a magnitude to quantify by the resolution of western blotting.

In summary, these experiments support but do not conclusively prove our original hypothesis that inhibition of the Notch pathway by a γ-secretase inhibitor would result in decreased cell proliferation and migration and support a role for Notch signaling in producing aggressive forms of OSCC. Although some studies suggest that when DAPT is administered in high concentration, cell death results [308] this was not our experience. Other studies have also previously revealed that down-regulation of Notch1 expression by DAPT substantially inhibited cell growth, induced G1 cell cycle arrest and induced apoptosis in an ovarian cancer and lung squamous cell carcinoma cell lines in a dose and time dependent manner [216, 310]. This raises the question of whether cell cycle arrest was related to the cell growth inhibition observed in the OSCC cell lines in the present study and to address this question further, studies need to be performed, possibly using flow cytometry for cell cycle analysis, as well as potentially TUNEL assay to explore the mechanisms of apoptosis induced by DAPT in OSCC cell lines further.

7.4.3 Future directions related to Notch

Future studies would potentially benefit from the availability of more primary OSCC cell lines and their genetic characterization. Alternative inhibitory strategy to GSIs i.e. siRNA knockdown or monoclonal antibodies, to rule out off-target gamma secretase inhibitory

effects could also be performed. As downstream Notch receptor signaling targets have been demonstrated to be differentially expressed in HNSCC [136] and only Notch receptor expression levels were assessed here a valuable extension to the study could be Real-time PCR performed on the downstream NOTCH1 genes, HES1 and HEY1, as further markers of Notch pathway activation. This could be of particular significance as a recent study demonstrated that HES1 and HEY1 were overexpressed in HNSCCs (~31.8% cases) in comparison to expression levels in normal mucosa [215].

Rather than being an exhaustive assessment of the relation of NOTCH1 to aggressive phenotype, our study was primarily a genomic study that explored the functional consequences of Notch signalling inhibition following the identification of a significantly greater number of NOTCH1 variants amongst an ECS-enriched cohort. Further transcriptomic, epigenetic and proteomic analysis of the same sample cohort could allow a wider ranging integrative means of analysis to better understand the role of Notch variants in ECS but also identify other potential drivers of aggressive phenotype in OSCC. These other drivers could either interact with Notch signalling or be independent of it.

Chapter 8 : Summary, Discussion and Future Directions

8. Overview

Extra-capsular spread from lymph nodes is a poor prognostic factor and one of the key hallmarks of aggressive phenotype in OSCC. Our SNP genotyping study in an ECS-enriched local cohort found an association of TP53 and CCND1 allelic imbalance with nodal status but not ECS. With overall allelic imbalance rates being comparable with the literature for CCND1 (41%) but not TP53 (36%) which was lower than expected, suggesting the possibility of a unique subset of cases with cohort specific effects and the need for further mutational assessment. This was initiated by whole genome sequencing on primary OSCC cell lines which identified a number of key genes found to be significantly mutated within these samples (CDK11A, CDC27, CASP8). However, some uneven coverage in certain key gene areas (NOTCH1) was also observed. Contemporaneous sequencing studies as well as evidence that notch signaling may play a key role in the characteristics required for aggressive phenotype alongside our own WGS data pathway analysis, suggested NOTCH1 may be of importance. Furthermore, detection of a potentially deleterious FBXW7 missense variant suggested that Notch signalling was significantly affected in the cell lines assessed. Specific Ion Torrent^{TM} sequencing for NOTCH1 and TP53 variants in primary OSCC tumour samples revealed 7 NOTCH1 novel nonsynonymous variants and 5 TP53 variants with overall mutational frequencies for NOTCH1 (17.5%) but not TP53 (12.5%) comparable with the literature. Furthermore, NOTCH1 variants were significantly associated with ECS ($p<0.05$) and the majority of variants corresponded to the ECD region of the NOTCH1 protein. Preliminary functional characterization showed that migration, proliferation and NOTCH isoform expression in an ECS+ve cell line, was significantly decreased by gamma secretase Notch pathway inhibition.

8.1. Importance of CCND1 for ECS

Frequent gain of CCND1 has been observed exclusively in OSCC patients with cervical lymph node metastasis [130]. A study of 164 OSCC cases showed that CCND1 allelic imbalance is an independent predictor for worse survival in OSCC in HPV negative

cases and a promising biomarker for predicting lymph node metastasis in patients with clinically Stage I–II OSCC. [154]. However, a study using FISH to assess CCND1 and EGFR copy number aberration in 127 primary OSCC cases demonstrated EGFR copy number aberration, but not CCND1, to be associated with ECS [58]. Other studies have demonstrated CCND1 amplification detected by q-PCR to be present in 50% of OSCC samples but not associated with lymph node metastasis [153]. In our study, we did observe an association with nodal status but not ECS.

One explanation for lack of association with ECS could be as a result of undetected micro metastases present in node negative samples [311]. Alternatively, as CCND1 is required for progression through the G1 phase of the cell cycle [312] and predominantly affects cellular proliferation it may be less vital to drive invasion. CCND1 amplification is likely to be an early event in oncogenesis with a FISH study on OSCC tumour samples, lymph node metastases, oral leukoplakia (OLK) and oral lichen planus (OLP), finding CCND1 amplification in a higher proportion of OSCC and lymph node samples than in OLK and OLP samples, but no significant difference between OSCC & lymph node samples [135].

8.2 Importance of TP53 for ECS

TP53 copy number alteration was associated with nodal status of OSCC, but next generation sequencing demonstrated that TP53 genetic variants are rare in our subset of OSCC cases in contrast to the literature. TP53 deletion [135] and loss of heterozygosity (LOH) [155] have been previously reported to be prevalent in OSCC with NGS studies revealing it to be consistently the most heavily mutated gene in OSCC [21]. OSCC patients carrying TP53 mutations have reduced survival compared to those with a wild-type status [263, 313] and furthermore the carriage of missense mutations in the TP53 DNA binding domain was associated with decreased disease-specific survival compared with wild-type TP53 in a cohort of 345 OSCC patients [314]. Our lack of TP53 variation was borne out through analysis of both known and unknown TP53 variants and further study of the existing literature supports the possibility of a genuine subset of

NOTCH1 mutant TP53 wild-type cases as observed within our ECS enriched cohort. Indeed, a recent systematic review of TP53 as a prognostic factor in HNSCC's four main anatomical subsites (larynx, oropharynx, hypopharynx, oral cavity) demonstrated large heterogeneity in study-level and patient-level characteristics, making it difficult to ascertain a clear picture [315]. The significance of the low level of TP53 variation amongst our cohort of ECS-enriched cases remains unclear and further sequencing studies with larger sample sizes may shed further light.

8.3 Importance of Notch for ECS

We found significant association of Notch pathway signaling by WGS of primary OSCC cell lines alongside significantly greater NOTCH1 variants associated with ECS. Variants were predominantly clustered in regions coding for the extracellular domain of the NOTCH1 gene. This finding was initially observed on the basis of Ion Torrent^{TM} sequencing but was further confirmed by bioinformatics re-analysis and validation of detected variants by Sanger sequencing, suggesting that Notch could be having a key role in the development of ECS. DAPT inhibition of proliferation, migration and differential Notch protein expression suggested that Notch signaling may be important for the behavior of these cells. Our observations indicate that NOTCH1 is more likely to be acting as an oncogene in our patient cohort, but further studies to elucidate the functional effects of NOTCH1 ECD variants are required.

The majority of variants identified in this study were found within the EGF repeat domain, as already evidenced in other squamous cell carcinoma settings in the literature [102, 268], and have been predicted to be mostly inactivating by disruption of ligand binding. However, insights gained on Notch-ligand interactions from drosophila studies may help explain our oncogenic findings. Studies support contrasting canonical Notch–ligand interaction modes. Trans ligand-receptor binding may occur to initiate Notch signaling [316]. Alternatively, cis inhibitory ligand-receptor binding as a result of ligand expression on the same cell surface as the receptor may limit Notch activity [317]. Although these differing modes of interaction are established, the molecular basis

for activating and inhibitory complexes is poorly understood. Notch EGF 11–12 and the DSL domain residues interact to promote Notch trans-activation, however, the molecular requirements for cis-inhibitory interactions are less clear. Early studies proposed that the Abruptex region of the Notch receptor (EGF24–29) might be involved in cis interactions [221], but a further study demonstrated that the Notch EGF 10–12 region is also required for cis-inhibition [318]. Three of our variants fell within exonic regions encoding these protein domians of the Notch receptor so it remains possible that our ECD detected variants may have interfered with Notch-ligand cis-inhibition in a manner that allowed the predominant activation of Notch signaling.

There remains a need for improved patient prognostication and assessment for OSCC management. Treatment for OSCC depends on the location of the primary tumour, the stage of the disease, and the expected oncological and functional outcomes [24]. Early-stage (I/II) HNSCC only requires single-modality therapy (i.e. surgery or radiotherapy), whereas locally advanced disease (AJCC stage III/IV) requires various combinations of radiotherapy, surgery, and chemotherapy or cetuximab [4]. Present survival rates for all patients with HNSCC are approximately 40% to 60% at 5 years [319] and there remains a significant under staging of disease by current imaging modalities [306]. Notch analysis may help as a biomarker for prognostication, with patients detected to have Notch mutations undergoing more stringent and multi-modal imaging assessment as well as further genetic characterisation, such as TP53.

Activating mutations in classic oncogenes are relatively rare in OSCC, most genetic alterations occurring in tumour suppressor genes, limiting opportunities for targeted therapies and restoring the activity of altered tumour suppressor genes like p53 or CDKN2A is extremely challenging [320]. While studies directly inhibiting Notch signaling in OSCC e.g. by GSIs remain to be reported, targeting non-canonical Notch signaling may hold great promise for novel treatment strategies. A US group are presently conducting (NIH National Cancer Institute NCT02649530) a Phase II clinical trial of WNT974 in patients with recurrent, metastatic or advanced stage HNSCC, with one of the tertiary objectives being to explore the response to therapy based on NOTCH

receptor mutational status, as well as assess expression of NOTCH and WNT pathway gene sets prior to therapy. As previous studies have demonstrated an enriched rate of response to WNT974 in cell lines harboring a NOTCH1 mutation. Thus, selection of appropriate management strategies on the basis of NOTCH variant status may potentially benefit treatment options in a number of ways. Only a minority of patients benefit from anti-EGFR monoclonal antibodies, and the objective response rate in monotherapy is between 6% and 13% [321]. Due to the negative regulation of Notch signalling by EGFR, cetuximab may show greater response in tumours carrying NOTCH variants associated with poor outcome. Alternatively, combination treatments of cetuximab with targeted Notch therapy may improve survival. Finally, information provided by both TP53 and NOTCH variant status may allow the more effective use of combinations of multi-modal therapy or more aggressive single mode therapy for cases of disease that would have otherwise been undertreated without genomic information predicting ECS and aggressive phenotype.

8.4 Limitations and Future Directions

Our initial sample size was determined by the availability of samples locally and power calculations based on expected mutational frequencies from the literature, however our study would benefit from increased sample size and specifically the inclusion of more node positive ECS negative cases. In our assessment of allelic imbalance, 24 samples for TP53 and 9 samples for CCND1 had to be excluded because of homozygosity and 10 samples were excluded from NGS analysis of NOTCH1 and TP53 due to resultant low sequence coverage. Additional samples may potentially indicate differences between all 3 histopathological subgroups, increase significance of notch to ECS and identify whether there are extracellular domain hotspots of mutation. Matched normal DNA would provide the required information to call ‘mutations’ as opposed to ‘variants’. Our analysis would also have benefited from more primary OSCC cell lines with a greater variety of novel ECD variants as well as further node negative and node positive control cell lines to compare.

Ligand binding assays and/or co-immunoprecipitation assays on OSCC cell lines with known NOTCH1 ECD variants would allow a further understanding of the mechanistic changes that variants could have on the notch receptor-ligand interaction. Transfection studies of wild type or mutant NOTCH1 (carrying an ECD variant) into primary OSCC cell lines would be beneficial to demonstrate the functional effects of variants. Specifically, co-expression of mutant and wild type notch with N or C terminal fluorescent protein tags in the same cell in vivo could be used to compare rates of processing.

8.5 Conclusion

Extracapsular spread from lymph nodes is a key poor prognostic marker in OSCC and present imaging modalities are too poor to identify it reliably pre-operatively. This study has identified a unique subset of ECS positive OSCC cases harboring a low rate of TP53 mutation along with significantly elevated levels of NOTCH1 variants within the extracellular domain. Our findings suggest that the downstream effects of NOTCH1 ECD mutations and their potential role in OSCC should be further explored as Notch may be a key driver of ECS in OSCC.

Bibliography

1. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.
2. Ridge, J.A., *head and neck tumors*. 2008: Cancer Management: A Multidisciplinary Approach.
3. CRUK. *Cancer Research UK 'Cancerstats'*. CancerStats Incidence 2009 – UK 2015.
4. Lydiatt, W.M., et al., *Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual.* CA Cancer J Clin, 2017. **67**(2): p. 122-137.
5. Bagan, J., G. Sarrion, and Y. Jimenez, *Oral cancer: clinical features.* Oral Oncol, 2010. **46**(6): p. 414-7.
6. Warnakulasuriya, S., *Global epidemiology of oral and oropharyngeal cancer.* Oral Oncology., 2009. **45**: p. 309–316.
7. Marsh, D., et al., *Stromal features are predictive of disease mortality in oral cancer patients.* J Pathol, 2011. **223**(4): p. 470-81.
8. Wang, B., et al., *The recurrence and survival of oral squamous cell carcinoma: a report of 275 cases.* Chin J Cancer, 2013. **32**(11): p. 614-8.
9. Johnson, N.W., P. Jayasekara, and A.A. Amarasinghe, *Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology.* Periodontol 2000, 2011. **57**(1): p. 19-37.
10. Gupta, B., N.W. Johnson, and N. Kumar, *Global Epidemiology of Head and Neck Cancers: A Continuing Challenge.* Oncology, 2016. **91**(1): p. 13-23.
11. Liu, B., et al., *Synergistic effects of betel quid chewing, tobacco use (in the form of cigarette smoking), and alcohol consumption on the risk of malignant transformation of oral submucous fibrosis (OSF): a case-control study in Hunan Province, China.* Oral Surg Oral Med Oral Pathol Oral Radiol, 2015. **120**(3): p. 337-45.
12. Anantharaman, D., et al., *Population attributable risk of tobacco and alcohol for upper aerodigestive tract cancer.* Oral Oncol, 2011. **47**(8): p. 725-31.
13. Syrjanen, K., et al., *Morphological and immunohistochemical evidence suggesting human papillomavirus (HPV) involvement in oral squamous cell carcinogenesis.* Int J Oral Surg, 1983. **12**(6): p. 418-24.
14. zur Hausen, H., *Papillomaviruses in the causation of human cancers - a brief historical account.* Virology, 2009. **384**(2): p. 260-5.
15. Miller, C.S. and B.M. Johnstone, *Human papillomavirus as a risk factor for oral squamous cell carcinoma: a meta-analysis, 1982-1997.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2001. **91**(6): p. 622-35.
16. Leemans, C.R., B.J. Braakhuis, and R.H. Brakenhoff, *The molecular biology of head and neck cancer.* Nat Rev Cancer, 2011. **11**(1): p. 9-22.
17. Wang, F., et al., *A systematic investigation of the association between HPV and the clinicopathological parameters and prognosis of oral and oropharyngeal squamous cell carcinomas.* Cancer Med, 2017.
18. Shaw, R.J., et al., *Extracapsular spread in oral squamous cell carcinoma.* Head Neck, 2010. **32**(6): p. 714-22.
19. Maxwell, J.H., et al., *Extracapsular spread in head and neck carcinoma: impact of site and human papillomavirus status.* Cancer, 2013. **119**(18): p. 3302-8.
20. Lassen, P., et al., *Effect of HPV-associated p16INK4A expression on response to radiotherapy and survival in squamous cell carcinoma of the head and neck.* J Clin Oncol, 2009. **27**(12): p. 1992-8.
21. Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-82.

22. Scully, C. and J. Bagan, *Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications.* Oral Dis, 2009. **15**(6): p. 388-99.
23. Radford, A.D., et al., *Application of next-generation sequencing technologies in virology.* J Gen Virol, 2012. **93**(Pt 9): p. 1853-68.
24. Shah, J.P. and Z. Gil, *Current concepts in management of oral cancer--surgery.* Oral Oncol, 2009. **45**(4-5): p. 394-401.
25. Jadhav, K.B. and N. Gupta, *Clinicopathological Prognostic Implicators of Oral Squamous Cell Carcinoma: Need to Understand and Revise.* N Am J Med Sci, 2013. **5**(12): p. 671-679.
26. Sobin L.H., G.M.K., Wittekind C., *TNM classification of malignant tumours*. 2009.
27. Mamelle, G., et al., *Lymph node prognostic factors in head and neck squamous cell carcinomas.* Am J Surg, 1994. **168**(5): p. 494-8.
28. Tong, X.J., et al., *The impact of clinical prognostic factors on the survival of patients with oral squamous cell carcinoma.* J Oral Maxillofac Surg, 2014. **72**(12): p. 2497 e1-10.
29. Anneroth, G., J. Batsakis, and M. Luna, *Review of the literature and a recommended system of malignancy grading in oral squamous cell carcinomas.* Scand J Dent Res, 1987. **95**(3): p. 229-49.
30. Fagan, J.J., et al., *Perineural invasion in squamous cell carcinoma of the head and neck.* Arch Otolaryngol Head Neck Surg, 1998. **124**(6): p. 637-40.
31. Larsen, S.R., et al., *The prognostic significance of histological features in oral squamous cell carcinoma.* J Oral Pathol Med, 2009. **38**(8): p. 657-62.
32. Moore, C., M.B. Flynn, and R.A. Greenberg, *Evaluation of size in prognosis of oral cancer.* Cancer, 1986. **58**(1): p. 158-62.
33. Woolgar, J.A., *T2 carcinoma of the tongue: the histopathologist's perspective.* Br J Oral Maxillofac Surg, 1999. **37**(3): p. 187-93.
34. Massano, J., et al., *Oral squamous cell carcinoma: review of prognostic and predictive factors.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2006. **102**(1): p. 67-76.
35. Shaw, R.J., et al., *Prognostic importance of site in squamous cell carcinoma of the buccal mucosa.* Br J Oral Maxillofac Surg, 2009. **47**(5): p. 356-9.
36. Bello, I.O., Y. Soini, and T. Salo, *Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I).* Oral Oncol, 2010. **46**(9): p. 630-5.
37. Ritchie, J.M., et al., *Human papillomavirus infection as a prognostic factor in carcinomas of the oral cavity and oropharynx.* Int J Cancer, 2003. **104**(3): p. 336-44.
38. Woolgar, *Diagnosing and staging oral cancer: Controversies, potential pitfalls and refining histopathological protocols.* Italian Journal of Maxillofacial Surgery, 2011. **22**(1 Supp): p. 5-7.
39. Alakus, H., et al., *Extracapsular lymph node spread: a new prognostic factor in gastric cancer.* Cancer, 2010. **116**(2): p. 309-15.
40. Lagarde, S.M., et al., *Extracapsular lymph node involvement in node-positive patients with adenocarcinoma of the distal esophagus or gastroesophageal junction.* Am J Surg Pathol, 2006. **30**(2): p. 171-6.
41. Gruber, G., et al., *Extracapsular tumor spread and the risk of local, axillary and supraclavicular recurrence in node-positive, premenopausal patients with breast cancer.* Ann Oncol, 2008. **19**(8): p. 1393-401.
42. Horn, L.C., et al., *Extracapsular extension of pelvic lymph node metastases is of prognostic value in carcinoma of the cervix uteri.* Gynecol Oncol, 2008. **108**(1): p. 63-7.
43. Bennett, S.H., et al., *Prognostic significance of histologic host response in cancer of the larynx or hypopharynx.* Cancer, 1971. **28**(5): p. 1255-65.
44. Noone, R.B., et al., *Lymph node metastases in oral carcinoma. A correlation of histopathology with survival.* Plast Reconstr Surg, 1974. **53**(2): p. 158-66.

45. Johnson, J.T., et al., *The extracapsular spread of tumors in cervical node metastasis.* Arch Otolaryngol, 1981. **107**(12): p. 725-9.
46. Snow, G.B., et al., *Prognostic factors of neck node metastasis.* Clin Otolaryngol Allied Sci, 1982. **7**(3): p. 185-92.
47. Lewis, J.S., Jr., et al., *Extracapsular extension is a poor predictor of disease recurrence in surgically treated oropharyngeal squamous cell carcinoma.* Mod Pathol, 2011. **24**(11): p. 1413-20.
48. Zoller, M., M.L. Goodman, and C.W. Cummings, *Guidelines for prognosis in head and neck cancer with nodal metastasis.* Laryngoscope, 1978. **88**(1 Pt 1): p. 135-40.
49. Vaidya, A.M., et al., *Patterns of spread in recurrent head and neck squamous cell carcinoma.* Otolaryngol Head Neck Surg, 2001. **125**(4): p. 393-6.
50. Johnson, J.T., et al., *Cervical lymph node metastases. Incidence and implications of extracapsular carcinoma.* Arch Otolaryngol, 1985. **111**(8): p. 534-7.
51. Vankhanen, V.D., *[Nutritional hygiene as a science and a field of professional activity at the present time].* Gig Sanit, 1986(9): p. 38-41.
52. Myers, J.N., et al., *Extracapsular spread. A significant predictor of treatment failure in patients with squamous cell carcinoma of the tongue.* Cancer, 2001. **92**(12): p. 3030-6.
53. de Juan, J., et al., *Inclusion of extracapsular spread in the pTNM classification system: a proposal for patients with head and neck carcinoma.* JAMA Otolaryngol Head Neck Surg, 2013. **139**(5): p. 483-8.
54. Greene FL, P.D., Fleming ID, Fritz AG, Balch CM, Haller DG, et al., *AJCC cancer staging handbook.* New York: Springer-Verlag., 2002.
55. Ernani, V. and N.F. Saba, *Oral Cavity Cancer: Risk Factors, Pathology, and Management.* Oncology, 2015. **89**(4): p. 187-95.
56. Cooper JS, P.T., Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al., J.Ferlay et al, *Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck.* N Engl J Med, 2004. **350**: p. 1937-1944.
57. Bernier J, D.C., Ozsahin M, Matuszewska K, Lefèbvre JL, Greiner RH, *Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer.* N Engl J Med, 2004. **350**: p. 1945-1952.
58. Michikawa, C., et al., *Epidermal growth factor receptor gene copy number aberration at the primary tumour is significantly associated with extracapsular spread in oral cancer.* Br J Cancer, 2011. **104**(5): p. 850-5.
59. Greenberg, J.S., et al., *Disparity in pathologic and clinical lymph node staging in oral tongue carcinoma. Implication for therapeutic decision making.* Cancer, 2003. **98**(3): p. 508-15.
60. Miyamoto, R., et al., *Potential marker of oral squamous cell carcinoma aggressiveness detected by fluorescence in situ hybridization in fine-needle aspiration biopsies.* Cancer, 2002. **95**(10): p. 2152-9.
61. Henry, N.L. and D.F. Hayes, *Cancer biomarkers.* Mol Oncol, 2012. **6**(2): p. 140-6.
62. Yong-Deok, K., et al., *Molecular genetic study of novel biomarkers for early diagnosis of oral squamous cell carcinoma.* Med Oral Patol Oral Cir Bucal, 2015. **20**(2): p. e167-79.
63. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.
64. Bignell, G.R., et al., *Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution.* Genome Res, 2007. **17**(9): p. 1296-303.
65. Volik, S., et al., *End-sequence profiling: sequence-based analysis of aberrant genomes.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7696-701.

66. Ruan, Y., et al., *Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs).* Genome Res, 2007. **17**(6): p. 828-38.
67. Alberts, B.M., *The molecular biology of cell*. 2008.
68. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-67.
69. Choi, S. and J.N. Myers, *Molecular pathogenesis of oral squamous cell carcinoma: implications for therapy.* J Dent Res, 2008. **87**(1): p. 14-32.
70. Rousseau, A., et al., *Frequent cyclin D1 gene amplification and protein overexpression in oral epithelial dysplasias.* Oral Oncol, 2001. **37**(3): p. 268-75.
71. Califano, J., et al., *Genetic progression model for head and neck cancer: implications for field cancerization.* Cancer Res, 1996. **56**(11): p. 2488-92.
72. Puisieux, A., et al., *Selective targeting of p53 gene mutational hotspots in human cancers by etiologically defined carcinogens.* Cancer Res, 1991. **51**(22): p. 6185-9.
73. Greenblatt, M.S., et al., *Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis.* Cancer Res, 1994. **54**(18): p. 4855-78.
74. Saunders, M.E., et al., *Patterns of p53 gene mutations in head and neck cancer: full-length gene sequencing and results of primary radiotherapy.* Clin Cancer Res, 1999. **5**(9): p. 2455-63.
75. Cabelguenne, A., et al., *p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma: a prospective series.* J Clin Oncol, 2000. **18**(7): p. 1465-73.
76. Temam, S., et al., *p53 gene status as a predictor of tumor response to induction chemotherapy of patients with locoregionally advanced squamous cell carcinomas of the head and neck.* J Clin Oncol, 2000. **18**(2): p. 385-94.
77. Joerger, A.C. and A.R. Fersht, *The tumor suppressor p53: from structures to drug discovery.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000919.
78. Prives, C. and P.A. Hall, *The p53 pathway.* J Pathol, 1999. **187**(1): p. 112-26.
79. Kato, S., et al., *Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8424-9.
80. Masica, D.L., et al., *Predicting survival in head and neck squamous cell carcinoma from TP53 mutation.* Hum Genet, 2015. **134**(5): p. 497-507.
81. Gudkov, A.V. and E.A. Komarova, *The role of p53 in determining sensitivity to radiotherapy.* Nat Rev Cancer, 2003. **3**(2): p. 117-29.
82. Burke, H.B., et al., *Predicting response to adjuvant and radiation therapy in patients with early stage breast carcinoma.* Cancer, 1998. **82**(5): p. 874-7.
83. Chen, S.J., et al., *Ultra-deep targeted sequencing of advanced oral squamous cell carcinoma identifies a mutation-based prognostic gene signature.* Oncotarget, 2015. **6**(20): p. 18066-80.
84. Rivera, C., et al., *Clinicopathological and immunohistochemical evaluation of oral and oropharyngeal squamous cell carcinoma in Chilean population.* Int J Clin Exp Pathol, 2014. **7**(9): p. 5968-77.
85. Braakhuis, B.J., et al., *TP53 mutation and human papilloma virus status of oral squamous cell carcinomas in young adult patients.* Oral Dis, 2014. **20**(6): p. 602-8.
86. Chang, Y.S., et al., *Combined mutational analysis of RAS, BRAF, PIK3CA, and TP53 genes in Taiwanese patients with oral squamous cell carcinoma.* Oral Surg Oral Med Oral Pathol Oral Radiol, 2014. **118**(1): p. 110-116 e1.
87. Cuevas Gonzalez, J.C., et al., *p53 and p16 in oral epithelial dysplasia and oral squamous cell carcinoma: A study of 208 cases.* Indian J Pathol Microbiol, 2016. **59**(2): p. 153-8.

88. Lee, G., et al., *Characterization of novel cell lines established from three human oral squamous cell carcinomas.* Int J Oncol, 2002. **20**(6): p. 1151-9.
89. Irimie, A.I., et al., *Knocking down of p53 triggers apoptosis and autophagy, concomitantly with inhibition of migration on SSC-4 oral squamous carcinoma cells.* Mol Cell Biochem, 2016. **419**(1-2): p. 75-82.
90. Jithesh, P.V., et al., *The epigenetic landscape of oral squamous cell carcinoma.* Br J Cancer, 2013. **108**(2): p. 370-9.
91. Fidler, I.J., *The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited.* Nat Rev Cancer, 2003. **3**(6): p. 453-8.
92. Simple, M., et al., *Cancer stem cells and field cancerization of oral squamous cell carcinoma.* Oral Oncol, 2015. **51**(7): p. 643-51.
93. Cho, R.W. and M.F. Clarke, *Recent advances in cancer stem cells.* Curr Opin Genet Dev, 2008. **18**(1): p. 48-53.
94. Lobo, N.A., et al., *The biology of cancer stem cells.* Annu Rev Cell Dev Biol, 2007. **23**: p. 675-99.
95. Yachida, S., et al., *Distant metastasis occurs late during the genetic evolution of pancreatic cancer.* Nature, 2010. **467**(7319): p. 1114-7.
96. Hammerman, P.S., D.N. Hayes, and J.R. Grandis, *Therapeutic insights from genomic studies of head and neck squamous cell carcinomas.* Cancer Discov, 2015. **5**(3): p. 239-44.
97. Brennan, J.A., et al., *Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck.* N Engl J Med, 1995. **332**(11): p. 712-7.
98. Koch, P. and F.A. Bahmer, *Oral lesions and symptoms related to metals used in dental restorations: a clinical, allergological, and histologic study.* J Am Acad Dermatol, 1999. **41**(3 Pt 1): p. 422-30.
99. Gillison, M.L., et al., *Evidence for a causal association between human papillomavirus and a subset of head and neck cancers.* J Natl Cancer Inst, 2000. **92**(9): p. 709-20.
100. Wiest, T., et al., *Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control.* Oncogene, 2002. **21**(10): p. 1510-7.
101. 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.
102. Stransky, N., et al., *The mutational landscape of head and neck squamous cell carcinoma.* Science, 2011. **333**(6046): p. 1157-60.
103. Liao, C.T., et al., *Outcome analysis of patients with pN2 oral cavity cancer.* Ann Surg Oncol, 2010. **17**(4): p. 1118-26.
104. Rothenberg, S.M.e.a., *The molecular pathogenesis of head and neck squamous cell carcinoma.* J Clin Invest, 2012. **122**(6): p. 1951-7.
105. Kutyavin, I.V., et al., *3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures.* Nucleic Acids Res, 2000. **28**(2): p. 655-61.
106. Rashid, M., et al., *Cake: a bioinformatics pipeline for the integrated analysis of somatic variants in cancer genomes.* Bioinformatics, 2013. **29**(17): p. 2208-10.
107. Ye, K., et al., *Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.* Bioinformatics, 2009. **25**(21): p. 2865-71.
108. Pruitt, K.D., et al., *NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy.* Nucleic Acids Res, 2012. **40**(Database issue): p. D130-5.
109. McLaren, W., et al., *Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor.* Bioinformatics, 2010. **26**(16): p. 2069-70.
110. Eilbeck, K., et al., *The Sequence Ontology: a tool for the unification of genome annotations.* Genome Biol, 2005. **6**(5): p. R44.

111. von Hansemann, D., *Ueber asymmetrische Zelltheilung in epithel Krebsen und deren biologische Bedeutung.* Virchow's Arch. Path. Anat., 1890. **119**: p. 299.
112. Heim S, M.F., *Cancer Cytogenetics.* 2010.
113. Nowell, P.C., *Discovery of the Philadelphia chromosome: a personal perspective.* J Clin Invest, 2007. **117**(8): p. 2033-5.
114. Kallioniemi, A., et al., *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.* Science, 1992. **258**(5083): p. 818-21.
115. Hester, S.D., et al., *Comparison of comparative genomic hybridization technologies across microarray platforms.* J Biomol Tech, 2009. **20**(2): p. 135-51.
116. Solinas-Toldo, S., et al., *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.* Genes Chromosomes Cancer, 1997. **20**(4): p. 399-407.
117. Pinkel, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nat Genet, 1998. **20**(2): p. 207-11.
118. Albertson, D.G., et al., *Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene.* Nat Genet, 2000. **25**(2): p. 144-6.
119. Bocker, T., J. Ruschoff, and R. Fishel, *Molecular diagnostics of cancer predisposition: hereditary non-polyposis colorectal carcinoma and mismatch repair defects.* Biochim Biophys Acta, 1999. **1423**(3): p. O1-O10.
120. Weir, B.A., et al., *Characterizing the cancer genome in lung adenocarcinoma.* Nature, 2007. **450**(7171): p. 893-8.
121. Jen, J., et al., *Allelic loss of chromosome 18q and prognosis in colorectal cancer.* N Engl J Med, 1994. **331**(4): p. 213-21.
122. Bown, N., et al., *Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma.* N Engl J Med, 1999. **340**(25): p. 1954-61.
123. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of urothelial bladder carcinoma.* Nature, 2014. **507**(7492): p. 315-22.
124. Chuang, A.Y., et al., *Presence of HPV DNA in convalescent salivary rinses is an adverse prognostic marker in head and neck squamous cell carcinoma.* Oral Oncol, 2008. **44**(10): p. 915-9.
125. Partridge, M., et al., *The prognostic significance of allelic imbalance at key chromosomal loci in oral cancer.* Br J Cancer, 1999. **79**(11-12): p. 1821-7.
126. Huang, Q., et al., *Genetic differences detected by comparative genomic hybridization in head and neck squamous cell carcinomas from different tumor sites: construction of oncogenetic trees for tumor progression.* Genes Chromosomes Cancer, 2002. **34**(2): p. 224-33.
127. Smeets, S.J., et al., *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.
128. Oga, A., et al., *Comparative genomic hybridization analysis reveals 3q gain resulting in genetic alteration in 3q in advanced oral squamous cell carcinoma.* Cancer Genet Cytogenet, 2001. **127**(1): p. 24-9.
129. Lin, S.C., et al., *Chromosomal changes in betel-associated oral squamous cell carcinomas and their relationship to clinical parameters.* Oral Oncol, 2002. **38**(3): p. 266-73.
130. Sugahara, K., et al., *Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma.* Int J Oncol, 2011. **39**(4): p. 761-9.
131. Yoshioka, S., et al., *Genomic profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization.* PLoS One, 2013. **8**(2): p. e56165.
132. Patel, S.G., et al., *Lymph node density in oral cavity cancer: results of the International Consortium for Outcomes Research.* Br J Cancer, 2013. **109**(8): p. 2087-95.

133. Smeets, S.J., et al., *Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis.* Cell Oncol, 2009. **31**(4): p. 291-300.
134. Roy, D.M., et al., *Integrated Genomics for Pinpointing Survival Loci within Arm-Level Somatic Copy Number Alterations.* Cancer Cell, 2016. **29**(5): p. 737-50.
135. Martin-Ezquerra, G., et al., *Multiple genetic copy number alterations in oral squamous cell carcinoma: study of MYC, TP53, CCDN1, EGFR and ERBB2 status in primary and metastatic tumours.* Br J Dermatol, 2010. **163**(5): p. 1028-35.
136. 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.
137. Ciriello, G., et al., *Emerging landscape of oncogenic signatures across human cancers.* Nat Genet, 2013. **45**(10): p. 1127-33.
138. Miyamoto, R., et al., *Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas.* Oral Oncol, 2003. **39**(6): p. 610-8.
139. Fan, C.C., et al., *Expression of E-cadherin, Twist, and p53 and their prognostic value in patients with oral squamous cell carcinoma.* J Cancer Res Clin Oncol, 2013. **139**(10): p. 1735-44.
140. Subramonia-Iyer, S., et al., *Array-based comparative genomic hybridization for investigating chromosomal abnormalities in patients with learning disability: systematic review meta-analysis of diagnostic and false-positive yields.* Genet Med, 2007. **9**(2): p. 74-9.
141. Oostlander, A.E., G.A. Meijer, and B. Ylstra, *Microarray-based comparative genomic hybridization and its applications in human genetics.* Clin Genet, 2004. **66**(6): p. 488-95.
142. Fukuda, Y., et al., *Monoallelic and unequal allelic expression of the HTR2A gene in human brain and peripheral lymphocytes.* Biol Psychiatry, 2006. **60**(12): p. 1331-5.
143. Chen, X., et al., *Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk.* Hum Mol Genet, 2008. **17**(9): p. 1336-48.
144. Mattarucchi, E., et al., *Different real time PCR approaches for the fine quantification of SNP's alleles in DNA pools: assays development, characterization and pre-validation.* J Biochem Mol Biol, 2005. **38**(5): p. 555-62.
145. Lind, K., et al., *Combining sequence-specific probes and DNA binding dyes in real-time PCR for specific nucleic acid quantification and melting curve analysis.* Biotechniques, 2006. **40**(3): p. 315-9.
146. Armour, J.A., et al., *Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats.* Nucleic Acids Res, 2007. **35**(3): p. e19.
147. Li, W. and M. Olivier, *Current analysis platforms and methods for detecting copy number variation.* Physiol Genomics, 2013. **45**(1): p. 1-16.
148. Chen, C., et al., *Genome-Wide Loss of Heterozygosity and DNA Copy Number Aberration in HPV-Negative Oral Squamous Cell Carcinoma and Their Associations with Disease-Specific Survival.* PLoS One, 2015. **10**(8): p. e0135074.
149. Gonzalez, M.V., et al., *Loss of heterozygosity and mutation analysis of the p16 (9p21) and p53 (17p13) genes in squamous cell carcinoma of the head and neck.* Clin Cancer Res, 1995. **1**(9): p. 1043-9.
150. Huijsmans, C.J., et al., *Allelic imbalance at the HER2/TOP2A locus in breast cancer.* Diagn Pathol, 2015. **10**: p. 56.
151. Parisi, F., et al., *Detecting copy number status and uncovering subclonal markers in heterogeneous tumor biopsies.* BMC Genomics, 2011. **12**: p. 230.
152. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res, 2002. **30**(12): p. e57.

153. Liu, H.S., et al., *Detection of copy number amplification of cyclin D1 (CCND1) and cortactin (CTTN) in oral carcinoma and oral brushed samples from areca chewers.* Oral Oncol, 2009. **45**(12): p. 1032-6.
154. van Kempen, P.M., et al., *Clinical relevance of copy number profiling in oral and oropharyngeal squamous cell carcinoma.* Cancer Med, 2015. **4**(10): p. 1525-35.
155. Ribeiro, I.P., et al., *Genetic gains and losses in oral squamous cell carcinoma: impact on clinical management.* Cell Oncol (Dordr), 2014. **37**(1): p. 29-39.
156. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. **446**(7132): p. 153-8.
157. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers.* Science, 2006. **314**(5797): p. 268-74.
158. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers.* Science, 2007. **318**(5853): p. 1108-13.
159. Jones, S., et al., *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.* Science, 2008. **321**(5897): p. 1801-6.
160. Frattini, V., et al., *The integrated landscape of driver genomic alterations in glioblastoma.* Nat Genet, 2013. **45**(10): p. 1141-9.
161. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. **474**(7353): p. 609-15.
162. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization of squamous cell lung cancers.* Nature, 2012. **489**(7417): p. 519-25.
163. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
164. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-9.
165. India Project Team of the International Cancer Genome, C., *Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups.* Nat Commun, 2013. **4**: p. 2873.
166. Iwai, S., et al., *Mutations of the APC, beta-catenin, and axin 1 genes and cytoplasmic accumulation of beta-catenin in oral squamous cell carcinoma.* J Cancer Res Clin Oncol, 2005. **131**(12): p. 773-82.
167. Kozaki, K., et al., *PIK3CA mutation is an oncogenic aberration at advanced stages of oral squamous cell carcinoma.* Cancer Sci, 2006. **97**(12): p. 1351-8.
168. Black, P.A., et al., *Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates.* BMC Genomics, 2015. **16**: p. 857.
169. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nat Rev Genet, 2014. **15**(2): p. 121-32.
170. An, Q., et al., *Heterogeneous breakpoints in patients with acute lymphoblastic leukemia and the dic(9;20)(p11-13;q11) show recurrent involvement of genes at 20q11.21.* Haematologica, 2009. **94**(8): p. 1164-9.
171. Spinola, M., et al., *Identification and functional characterization of the candidate tumor suppressor gene TRIT1 in human lung cancer.* Oncogene, 2005. **24**(35): p. 5502-9.
172. Gum, J.R., Jr., et al., *MUC17, a novel membrane-tethered mucin.* Biochem Biophys Res Commun, 2002. **291**(3): p. 466-75.
173. Koukourakis, M.I., et al., *Comparison of metabolic pathways between cancer cells and stromal cells in colorectal carcinomas: a metabolic survival role for tumor-associated stroma.* Cancer Res, 2006. **66**(2): p. 632-7.

174. Houchens, C.R., et al., *The dhfr oribeta-binding protein RIP60 contains 15 zinc fingers: DNA binding and looping by the central three fingers and an associated proline-rich region.* Nucleic Acids Res, 2000. **28**(2): p. 570-81.
175. Trembley, J.H., et al., *Cyclin dependent kinase 11 in RNA transcription and splicing.* Prog Nucleic Acid Res Mol Biol, 2004. **77**: p. 263-88.
176. Lahti, J.M., et al., *Alterations in the PITSLRE protein kinase gene complex on chromosome 1p36 in childhood neuroblastoma.* Nat Genet, 1994. **7**(3): p. 370-5.
177. Jin, L., et al., *Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex.* Cell, 2008. **133**(4): p. 653-65.
178. Qiu, L., et al., *Downregulation of CDC27 inhibits the proliferation of colorectal cancer cells via the accumulation of p21Cip1/Waf1.* Cell Death Dis, 2016. **7**: p. e2074.
179. Yuan, X., et al., *Notch signaling: an emerging therapeutic target for cancer treatment.* Cancer Lett, 2015. **369**(1): p. 20-7.
180. Kominami, K., et al., *The molecular mechanism of apoptosis upon caspase-8 activation: quantitative experimental validation of a mathematical model.* Biochim Biophys Acta, 2012. **1823**(10): p. 1825-40.
181. Liu, L., et al., *Comparison of next-generation sequencing systems.* J Biomed Biotechnol, 2012. **2012**: p. 251364.
182. Veal, C.D., et al., *A mechanistic basis for amplification differences between samples and between genome regions.* BMC Genomics, 2012. **13**: p. 455.
183. Wooderchak-Donahue, W.L., et al., *A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective.* BMC Med Genomics, 2012. **5**: p. 50.
184. Kircher, M. and J. Kelso, *High-throughput DNA sequencing--concepts and limitations.* Bioessays, 2010. **32**(6): p. 524-36.
185. Dimalanta ET, Z.L., Hendrickson CL, *Increased Read Length on the SOLiD Sequencing Platform.* SOLiDTM System., 2009.
186. Takeishi, S. and K.I. Nakayama, *Role of Fbxw7 in the maintenance of normal stem cells and cancer-initiating cells.* Br J Cancer, 2014. **111**(6): p. 1054-9.
187. Erbilgin, Y., et al., *Prognostic significance of NOTCH1 and FBXW7 mutations in pediatric T-ALL.* Dis Markers, 2010. **28**(6): p. 353-60.
188. Bianco, R., et al., *Key cancer cell signal transduction pathways as therapeutic targets.* Eur J Cancer, 2006. **42**(3): p. 290-4.
189. Yap, L.F., et al., *The opposing roles of NOTCH signalling in head and neck cancer: a mini review.* Oral Dis, 2015. **21**(7): p. 850-7.
190. Hijioka, H., et al., *Upregulation of Notch pathway molecules in oral squamous cell carcinoma.* Int J Oncol, 2010. **36**(4): p. 817-22.
191. Gordon, W.R., K.L. Arnett, and S.C. Blacklow, *The molecular logic of Notch signaling--a structural and biochemical perspective.* J Cell Sci, 2008. **121**(Pt 19): p. 3109-19.
192. Kovall, R.A., *More complicated than it looks: assembly of Notch pathway transcription complexes.* Oncogene, 2008. **27**(38): p. 5099-109.
193. Egloff, A.M. and J.R. Grandis, *Molecular pathways: context-dependent approaches to Notch targeting as cancer therapy.* Clin Cancer Res, 2012. **18**(19): p. 5188-95.
194. Hambleton, S., et al., *Structural and functional properties of the human notch-1 ligand binding region.* Structure, 2004. **12**(12): p. 2173-83.
195. Sanchez-Irizarry, C., et al., *Notch subunit heterodimerization and prevention of ligand-independent proteolytic activation depend, respectively, on a novel domain and the LNR repeats.* Mol Cell Biol, 2004. **24**(21): p. 9265-73.

196. Chillakuri, C.R., et al., *Notch receptor-ligand binding and activation: insights from molecular studies.* Semin Cell Dev Biol, 2012. **23**(4): p. 421-8.
197. Kao, H.Y., et al., *A histone deacetylase corepressor complex regulates the Notch signal transduction pathway.* Genes Dev, 1998. **12**(15): p. 2269-77.
198. Bailey, A.M. and J.W. Posakony, *Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity.* Genes Dev, 1995. **9**(21): p. 2609-22.
199. Jarriault, S., et al., *Delta-1 activation of notch-1 signaling results in HES-1 transactivation.* Mol Cell Biol, 1998. **18**(12): p. 7423-31.
200. Wu, G., et al., *SEL-10 is an inhibitor of notch signaling that targets notch for ubiquitin-mediated protein degradation.* Mol Cell Biol, 2001. **21**(21): p. 7403-15.
201. Kwon, C., et al., *Notch post-translationally regulates beta-catenin protein in stem and progenitor cells.* Nat Cell Biol, 2011. **13**(10): p. 1244-51.
202. Acosta, H., et al., *Notch destabilises maternal beta-catenin and restricts dorsal-anterior development in Xenopus.* Development, 2011. **138**(12): p. 2567-79.
203. Hayward, P., et al., *Notch modulates Wnt signalling by associating with Armadillo/beta-catenin and regulating its transcriptional activity.* Development, 2005. **132**(8): p. 1819-30.
204. Ellisen, L.W., et al., *TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms.* Cell, 1991. **66**(4): p. 649-61.
205. Sriuranpong, V., et al., *Notch signaling induces cell cycle arrest in small cell lung cancer cells.* Cancer Res, 2001. **61**(7): p. 3200-5.
206. Qi, R., et al., *Notch1 signaling inhibits growth of human hepatocellular carcinoma through induction of cell cycle arrest and apoptosis.* Cancer Res, 2003. **63**(23): p. 8323-9.
207. Wang, N.J., et al., *Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma.* Proc Natl Acad Sci U S A, 2011. **108**(43): p. 17761-6.
208. Leethanakul, C., et al., *Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays.* Oncogene, 2000. **19**(28): p. 3220-4.
209. Zeng, Q., et al., *Crosstalk between tumor and endothelial cells promotes tumor angiogenesis by MAPK activation of Notch signaling.* Cancer Cell, 2005. **8**(1): p. 13-23.
210. Lee, S.H., et al., *TNFalpha enhances cancer stem cell-like phenotype via Notch-Hes1 activation in oral squamous cell carcinoma cells.* Biochem Biophys Res Commun, 2012. **424**(1): p. 58-64.
211. Joo, Y.H., et al., *Relationship between vascular endothelial growth factor and Notch1 expression and lymphatic metastasis in tongue cancer.* Otolaryngol Head Neck Surg, 2009. **140**(4): p. 512-8.
212. Zhang, T.H., et al., *Activation of Notch signaling in human tongue carcinoma.* J Oral Pathol Med, 2011. **40**(1): p. 37-45.
213. Yoshida, R., et al., *The pathological significance of Notch1 in oral squamous cell carcinoma.* Lab Invest, 2013. **93**(10): p. 1068-81.
214. Yao, J., et al., *Gamma-secretase inhibitors exerts antitumor activity via down-regulation of Notch and Nuclear factor kappa B in human tongue carcinoma cells.* Oral Dis, 2007. **13**(6): p. 555-63.
215. Sun, W., et al., *Activation of the NOTCH pathway in head and neck cancer.* Cancer Res, 2014. **74**(4): p. 1091-104.
216. Wang, M., et al., *Down-regulation of Notch1 by gamma-secretase inhibition contributes to cell growth inhibition and apoptosis in ovarian cancer cells A2780.* Biochem Biophys Res Commun, 2010. **393**(1): p. 144-9.
217. Gu, F., et al., *Expression of Stat3 and Notch1 is associated with cisplatin resistance in head and neck squamous cell carcinoma.* Oncol Rep, 2010. **23**(3): p. 671-6.

218. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-9.
219. D'Souza, B., A. Miyamoto, and G. Weinmaster, *The many facets of Notch ligands.* Oncogene, 2008. **27**(38): p. 5148-67.
220. Song, X., et al., *Common and complex Notch1 mutations in Chinese oral squamous cell carcinoma.* Clin Cancer Res, 2014. **20**(3): p. 701-10.
221. de Celis, J.F. and S.J. Bray, *The Abruptex domain of Notch regulates negative interactions between Notch, its ligands and Fringe.* Development, 2000. **127**(6): p. 1291-302.
222. Jia, H., et al., *Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer.* Sci Rep, 2014. **4**: p. 5737.
223. Wolfe, M.S., *γ-Secretase Inhibitors as Molecular Probes of Presenilin Function.* Journal of Molecular Neuroscience, 2001. **17**.
224. Tang, H. and P.D. Thomas, *PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation.* Bioinformatics, 2016. **32**(14): p. 2230-2.
225. Mertes, F., et al., *Targeted enrichment of genomic DNA regions for next-generation sequencing.* Brief Funct Genomics, 2011. **10**(6): p. 374-86.
226. Tewhey, R., et al., *Enrichment of sequencing targets from the human genome by solution hybridization.* Genome Biol, 2009. **10**(10): p. R116.
227. Antequera, F. and A. Bird, *Number of CpG islands and genes in human and mouse.* Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11995-9.
228. Harismendy, O. and K. Frazer, *Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology.* Biotechniques, 2009. **46**(3): p. 229-31.
229. Barnes, W.M., *The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion.* Gene, 1992. **112**(1): p. 29-35.
230. Cline, J., J.C. Braman, and H.H. Hogrefe, *PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases.* Nucleic Acids Res, 1996. **24**(18): p. 3546-51.
231. Rampias, T., et al., *A new tumor suppressor role for the Notch pathway in bladder cancer.* Nat Med, 2014. **20**(10): p. 1199-205.
232. Lee, S.H., et al., *Mutational analysis of NOTCH1, 2, 3 and 4 genes in common solid cancers and acute leukemias.* APMIS, 2007. **115**(12): p. 1357-63.
233. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.
234. Thutkawkorapin, J., et al., *Exome sequencing in one family with gastric- and rectal cancer.* BMC Genet, 2016. **17**: p. 41.
235. Sharma, S.V., D.A. Haber, and J. Settleman, *Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents.* Nat Rev Cancer, 2010. **10**(4): p. 241-53.
236. Arsenic, R., et al., *Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer.* BMC Clin Pathol, 2015. **15**: p. 20.
237. Sutton, L.A., et al., *Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting.* Haematologica, 2015. **100**(3): p. 370-6.
238. Salipante, S.J., et al., *Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling.* Appl Environ Microbiol, 2014. **80**(24): p. 7583-91.
239. Kulkarni S., P.J., *Clinical Genomics : A guide to clinical next generation sequencing.* 2014.
240. Aoyama, K., et al., *Frequent mutations in NOTCH1 ligand-binding regions in Japanese oral squamous cell carcinoma.* Biochem Biophys Res Commun, 2014. **452**(4): p. 980-5.

241. Wong, Y.K., et al., *p53 alterations in betel quid- and tobacco-associated oral squamous cell carcinomas from Taiwan.* J Oral Pathol Med, 1998. **27**(6): p. 243-8.
242. Saranath, D., et al., *p53 inactivation in chewing tobacco-induced oral cancers and leukoplakias from India.* Oral Oncol, 1999. **35**(3): p. 242-50.
243. Costa, J.L., et al., *Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting.* Hum Mutat, 2013. **34**(4): p. 629-35.
244. Yeo, C.D., et al., *Detection and comparison of EGFR mutations in matched tumor tissues, cell blocks, pleural effusions, and sera from patients with NSCLC with malignant pleural effusion, by PNA clamping and direct sequencing.* Lung Cancer, 2013. **81**(2): p. 207-12.
245. Spencer, D.H., et al., *Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data.* J Mol Diagn, 2014. **16**(1): p. 75-88.
246. Thomas, R.K., et al., *High-throughput oncogene mutation profiling in human cancer.* Nat Genet, 2007. **39**(3): p. 347-51.
247. Smits, A.J., et al., *The estimation of tumor cell percentage for molecular testing by pathologists is not accurate.* Mod Pathol, 2014. **27**(2): p. 168-74.
248. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nat Commun, 2013. **4**: p. 2612.
249. Carter, S.L., et al., *Absolute quantification of somatic DNA alterations in human cancer.* Nat Biotechnol, 2012. **30**(5): p. 413-21.
250. Su, X., et al., *PurityEst: estimating purity of human tumor samples using next-generation sequencing data.* Bioinformatics, 2012. **28**(17): p. 2265-6.
251. Palacio-Rua, K.A., et al., *Genetic analysis in APC, KRAS, and TP53 in patients with stomach and colon cancer.* Rev Gastroenterol Mex, 2014. **79**(2): p. 79-89.
252. Jha, P., et al., *TP53 polymorphisms in gliomas from Indian patients: Study of codon 72 genotype, rs1642785, rs1800370 and 16 base pair insertion in intron-3.* Exp Mol Pathol, 2011. **90**(2): p. 167-72.
253. Chitrala, K.N. and S. Yeguvapalli, *Computational screening and molecular dynamic simulation of breast cancer associated deleterious non-synonymous single nucleotide polymorphisms in TP53 gene.* PLoS One, 2014. **9**(8): p. e104242.
254. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.* Science, 1990. **250**(4985): p. 1233-8.
255. Saadatian, Z., et al., *Association of rs1219648 in FGFR2 and rs1042522 in TP53 with premenopausal breast cancer in an Iranian Azeri population.* Asian Pac J Cancer Prev, 2014. **15**(18): p. 7955-8.
256. Fan, C., et al., *The functional TP53 rs1042522 and MDM4 rs4245739 genetic variants contribute to Non-Hodgkin lymphoma risk.* PLoS One, 2014. **9**(9): p. e107047.
257. Gomes, C.C., et al., *TP53 single nucleotide polymorphism rs1042522 in salivary gland neoplasms.* Head Neck, 2014. **36**(12): p. 1685-8.
258. Orsted, D.D., et al., *Tumor suppressor p53 Arg72Pro polymorphism and longevity, cancer survival, and risk of cancer in the general population.* J Exp Med, 2007. **204**(6): p. 1295-301.
259. Lim, S.O., et al., *Notch1 differentially regulates oncogenesis by wildtype p53 overexpression and p53 mutation in grade III hepatocellular carcinoma.* Hepatology, 2011. **53**(4): p. 1352-62.
260. Maruyama, H., et al., *Human papillomavirus and p53 mutations in head and neck squamous cell carcinoma among Japanese population.* Cancer Sci, 2014. **105**(4): p. 409-17.
261. Zeng, X.T., et al., *Association between the TP53 codon 72 polymorphism and risk of oral squamous cell carcinoma in Asians: a meta-analysis.* BMC Cancer, 2014. **14**: p. 469.

262. Pannone, G., et al., *Evaluation of a combined triple method to detect causative HPV in oral and oropharyngeal squamous cell carcinomas: p16 Immunohistochemistry, Consensus PCR HPV-DNA, and In Situ Hybridization.* Infect Agent Cancer, 2012. **7**: p. 4.
263. Zanaruddin, S.N., et al., *Common oncogenic mutations are infrequent in oral squamous cell carcinoma of Asian origin.* PLoS One, 2013. **8**(11): p. e80229.
264. Tan, D.S., et al., *Tongue carcinoma infrequently harbor common actionable genetic alterations.* BMC Cancer, 2014. **14**: p. 679.
265. Heaton, C.M., et al., *TP53 and CDKN2a mutations in never-smoker oral tongue squamous cell carcinoma.* Laryngoscope, 2014. **124**(7): p. E267-73.
266. Pei, Z. and N.E. Baker, *Competition between Delta and the Abruptex domain of Notch.* BMC Dev Biol, 2008. **8**: p. 4.
267. Malecki, M.J., et al., *Leukemia-associated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes.* Mol Cell Biol, 2006. **26**(12): p. 4642-51.
268. Agrawal, N., et al., *Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.* Science. **333**(6046): p. 1154-7.
269. Rettig, E.M., et al., *Cleaved NOTCH1 Expression Pattern in Head and Neck Squamous Cell Carcinoma Is Associated with NOTCH1 Mutation, HPV Status, and High-Risk Features.* Cancer Prev Res (Phila), 2015. **8**(4): p. 287-95.
270. Yoshida, R., et al., *The pathological significance of Notch1 in oral squamous cell carcinoma.* Lab Invest. **93**(10): p. 1068-81.
271. Kolev, V., et al., *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.
272. Ratushny, V., et al., *Targeting EGFR resistance networks in head and neck cancer.* Cell Signal, 2009. **21**(8): p. 1255-68.
273. Sun, Y., et al., *Differential Notch1 and Notch2 expression and frequent activation of Notch signaling in gastric cancers.* Arch Pathol Lab Med, 2011. **135**(4): p. 451-8.
274. Carson, C., B. Murdoch, and A.J. Roskams, *Notch 2 and Notch 1/3 segregate to neuronal and glial lineages of the developing olfactory epithelium.* Dev Dyn, 2006. **235**(6): p. 1678-88.
275. Roy, M., W.S. Pear, and J.C. Aster, *The multifaceted role of Notch in cancer.* Curr Opin Genet Dev, 2007. **17**(1): p. 52-9.
276. Andersson, E.R. and U. Lendahl, *Therapeutic modulation of Notch signalling--are we there yet?* Nat Rev Drug Discov, 2014. **13**(5): p. 357-78.
277. Mu, X., et al., *Notch Signaling is Associated with ALDH Activity and an Aggressive Metastatic Phenotype in Murine Osteosarcoma Cells.* Front Oncol, 2013. **3**: p. 143.
278. Suman, S., T.P. Das, and C. Damodaran, *Silencing NOTCH signaling causes growth arrest in both breast cancer stem cells and breast cancer cells.* Br J Cancer, 2013. **109**(10): p. 2587-96.
279. Weaver, A.N., et al., *Notch Signaling Activation is Associated with Patient Mortality and Increased FGF1-mediated Invasion in Squamous Cell Carcinoma of the Oral Cavity.* Mol Cancer Res, 2016.
280. Chinn, S.B., et al., *Cancer stem cells: mediators of tumorigenesis and metastasis in head and neck squamous cell carcinoma.* Head Neck, 2015. **37**(3): p. 317-26.
281. Shrivastava, S., et al., *Identification of molecular signature of head and neck cancer stem-like cells.* Sci Rep, 2015. **5**: p. 7819.
282. Wang, Z., et al., *Down-regulation of Notch-1 contributes to cell growth inhibition and apoptosis in pancreatic cancer cells.* Mol Cancer Ther, 2006. **5**(3): p. 483-93.
283. Okuhashi, Y., et al., *NOTCH knockdown affects the proliferation and mTOR signaling of leukemia cells.* Anticancer Res, 2013. **33**(10): p. 4293-8.

284. Wang, J., et al., *siRNA targeting Notch-1 decreases glioma stem cell proliferation and tumor growth.* Mol Biol Rep, 2012. **39**(3): p. 2497-503.
285. Espinoza, I. and L. Miele, *Notch inhibitors for cancer treatment.* Pharmacol Ther, 2013. **139**(2): p. 95-110.
286. Curry, C.L., et al., *Gamma secretase inhibitor blocks Notch activation and induces apoptosis in Kaposi's sarcoma tumor cells.* Oncogene, 2005. **24**(42): p. 6333-44.
287. Wolfe, M.S., *Structure, mechanism and inhibition of gamma-secretase and presenilin-like proteases.* Biol Chem, 2010. **391**(8): p. 839-47.
288. Sun, L., et al., *Structural basis of human gamma-secretase assembly.* Proc Natl Acad Sci U S A, 2015. **112**(19): p. 6003-8.
289. Francis, R., et al., *aph-1 and pen-2 are required for Notch pathway signaling, gamma-secretase cleavage of betaAPP, and presenilin protein accumulation.* Dev Cell, 2002. **3**(1): p. 85-97.
290. Kovacsb, A.H.a.D.M., *The Many Substrates of Presenilin/-Secretase.* Journal of Alzheimer's Disease, 2011. **25**: p. 3-28.
291. Adachi, G.S.a.A., *Requirements for Presenilin-Dependent Cleavage of Notch and Other Transmembrane Proteins.* Molecular Cell, 2000. **6**: p. 625–636.
292. Morohashi, Y., et al., *C-terminal fragment of presenilin is the molecular target of a dipeptidic gamma-secretase-specific inhibitor DAPT (N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester).* J Biol Chem, 2006. **281**(21): p. 14670-6.
293. Clarke, E.E., et al., *Intra- or intercomplex binding to the gamma-secretase enzyme. A model to differentiate inhibitor classes.* J Biol Chem, 2006. **281**(42): p. 31279-89.
294. G. Evin, M.F.S., C.L. Masters, *Inhibition of gamma-secretase as a therapeutic intervention for Alzheimer's disease: prospects, limitations and strategies.* CNS Drugs, 2006. **20**: p. 351–372.
295. Grosveld, G.C., *Gamma-secretase inhibitors: Notch so bad.* Nat Med, 2009. **15**(1): p. 20-1.
296. Masuda, S., et al., *Dual antitumor mechanisms of Notch signaling inhibitor in a T-cell acute lymphoblastic leukemia xenograft model.* Cancer Sci, 2009. **100**(12): p. 2444-50.
297. Z. Wang, Y.Z., Y. Li, S. Banerjee, J. Liao, F.H. Sarkar,, *Down-regulation of Notch-1 contributes to cell growth inhibition and apoptosis in pancreatic cancer cells.* Mol. Cancer Ther, 2006. **5**: p. 483–493.
298. Dovey, H.F., et al., *Functional gamma-secretase inhibitors reduce beta-amyloid peptide levels in brain.* J Neurochem, 2001. **76**(1): p. 173-81.
299. Robinson, D.R., et al., *Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer.* Nat Med, 2011. **17**(12): p. 1646-51.
300. Grottkau, B.E., et al., *DAPT enhances the apoptosis of human tongue carcinoma cells.* Int J Oral Sci, 2009. **1**(2): p. 81-9.
301. Zhang, Z.P., et al., *Correlation of Notch1 expression and activation to cisplatin-sensitivity of head and neck squamous cell carcinoma.* Ai Zheng, 2009. **28**(2): p. 100-3.
302. Yu, S., et al., *Down-regulation of Notch signaling by a gamma-secretase inhibitor enhances the radiosensitivity of nasopharyngeal carcinoma cells.* Oncol Rep, 2011. **26**(5): p. 1323-8.
303. Zhao, J., et al., *TSG attenuates LPC-induced endothelial cells inflammatory damage through notch signaling inhibition.* IUBMB Life, 2016. **68**(1): p. 37-50.
304. Imbimbo, B.P., *Therapeutic potential of gamma-secretase inhibitors and modulators.* Curr Top Med Chem, 2008. **8**(1): p. 54-61.
305. Fengxin Zhu, T.L., Fanghua Qiu, Jinjin Fan, Qin Zhou, Xuebing Ding, Jing Nie, and Xueqing Yu, *Preventive Effect of Notch Signaling Inhibition by a -Secretase Inhibitor on Peritoneal Dialysis Fluid-Induced Peritoneal Fibrosis in Rats.* The American Journal of Pathology, 2010. **176**.
306. Dhanda, J., et al., *SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma.* Br J Cancer, 2014. **111**(11): p. 2114-21.

307. Liu, Y.F., et al., *Silencing of MAP4K4 by short hairpin RNA suppresses proliferation, induces G1 cell cycle arrest and induces apoptosis in gastric cancer cells.* Mol Med Rep, 2016. **13**(1): p. 41-8.
308. D. Williams Parsons, e.a., *An Integrated Genomic Analysis of Human Glioblastoma Multiforme.* science, 2008. **321**.
309. Zhu, F., et al., *Preventive effect of Notch signaling inhibition by a gamma-secretase inhibitor on peritoneal dialysis fluid-induced peritoneal fibrosis in rats.* Am J Pathol, 2010. **176**(2): p. 650-9.
310. Cao, H., et al., *Down-regulation of Notch receptor signaling pathway induces caspase-dependent and caspase-independent apoptosis in lung squamous cell carcinoma cells.* APMIS, 2012. **120**(6): p. 441-50.
311. Fan, S., et al., *A review of clinical and histological parameters associated with contralateral neck metastases in oral squamous cell carcinoma.* Int J Oral Sci, 2011. **3**(4): p. 180-91.
312. Baldin, V., et al., *Cyclin D1 is a nuclear protein required for cell cycle progression in G1.* Genes Dev, 1993. **7**(5): p. 812-21.
313. Tanuma, J., et al., *FGFR4 polymorphism, TP53 mutation, and their combinations are prognostic factors for oral squamous cell carcinoma.* Oncol Rep, 2010. **23**(3): p. 739-44.
314. Lapke, N., et al., *Missense mutations in the TP53 DNA-binding domain predict outcomes in patients with advanced oral cavity squamous cell carcinoma.* Oncotarget, 2016.
315. Tandon, S., et al., *A systematic review of p53 as a prognostic factor of survival in squamous cell carcinoma of the four main anatomical subsites of the head and neck.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(2): p. 574-87.
316. de Celis, J.F. and S. Bray, *Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the Drosophila wing.* Development, 1997. **124**(17): p. 3241-51.
317. Micchelli, C.A., E.J. Rulifson, and S.S. Blair, *The function and regulation of cut expression on the wing margin of Drosophila: Notch, Wingless and a dominant negative role for Delta and Serrate.* Development, 1997. **124**(8): p. 1485-95.
318. Becam, I., et al., *A role of receptor Notch in ligand cis-inhibition in Drosophila.* Curr Biol, 2010. **20**(6): p. 554-60.
319. Pulte, D. and H. Brenner, *Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis.* Oncologist, 2010. **15**(9): p. 994-1001.
320. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
321. Mehra, R., R.B. Cohen, and B.A. Burtness, *The role of cetuximab for the treatment of squamous cell carcinoma of the head and neck.* Clin Adv Hematol Oncol, 2008. **6**(10): p. 742-50.

Appendix 1:

Abstract arising from this thesis presented at the **National Cancer Research Institute Conference**, Liverpool, November 2016

& also published in the ***European Journal of Surgical Oncology*** (November 2016 edition)

NOTCH1 Mutations and Extracapsular Spread from Oral Squamous Cell Carcinoma Lymph Node Metastasis

Bhattacharya P. [1], Risk J.M. [1], Dhanda J. [1], Shaw R.J. [1,2], Lloyd B.[1], Sibson R.[1]

[1]Dept Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, L69 3GA. U.K.; [2] Regional Maxillofacial Unit, Aintree University Hospitals NHS Foundation Trust, Liverpool,UK

Background: The most adverse prognostic predictor for Oral Squamous Cell Carcinoma (OSCC) is Extracapsular Spread (ECS)) from lymph node metastasis, which needs further molecular characterization.

Methods: Next Generation Sequencing compared TP53 and NOTCH1 mutations in OSCC primary tumour samples (n=50; 21 node negative [N-], 11 [N+ECS-], 18 [N+ECS-]). Sanger Sequencing independently validated candidate variants. Primary OSCC cell lines were chosen according to NOTCH1 variant status. Their cell growth and migration were compared following treatment with the γ-secretase inhibitor DAPT (N-[N-(3,5-difluorophenacetyl)- L-alanyl]-S-phenylglycine t-butyl ester), used for Notch inhibition.

Results: Data from 40/50 OSCC samples was included following quality control screening for average coverage. Known intronic and exonic variants (1000 Genomes) were found plus 14 NOTCH1 non-synonymous variants in 13 OSCC samples and 7 non-synonymous TP53 variants. 9/14 [N+ECS+] samples had NOTCH1 variants, compared to TP53 (2/14 of [N+ECS+] samples). Sanger Sequencing confirmed 57% of all variants with a rate of 7/40 (17.5%) for NOTCH1. 6/14 (43%) ECS+ samples had confirmed NOTCH1 variants compared to 1/26 (3.8%) ECS-, Fisher's exact test p = 0.0044. Variants clustered in the N-terminal, extracellular EGF-like repeat region.

In vitro, decreased cell growth and migration was observed after DAPT (25 μM) treatment for four weeks compared to controls. Western blot analysis for Notch proteins revealed absence of Notch1 and 3 after 3 weeks and Notch4 by week four.

Conclusion: Overall rates of NOTCH1 mutations in our OSCC cohort were consistent with recent literature and NOTCH1 may be a key driver of ECS in OSCC.

Appendix 2:

Supplementary CD	Table number	File name
Supplementary data CD	**1**	Complete list of Unknown INDELs with potential functional effects
	2	Complete list for unknown synonymous variants
	3	Complete list of Unknown non-synonymous SNPs
	4	Complete list of genes for pathway analysis
	5	Pathway analysis data

Appendix 3:

Reagent / Kit	Manufacturer	Catalogue Number
AllPrep DNA/RNA/Protein Mini Kit	Qiagen	80004
QIAquick PCR Purification Kit	Qiagen	28104
Agarose Seakem LE	Fisher Scientific	BMA 50002
SYBRsafe	Invitrogen	S33102
1kb Ladder	NEB	N3232S
100 bp Ladder	NEB	N3231S
DNA 1000 kit	Agilent	50671504
Taqman assay universal PCR mastermix Amplitaq (R) Gold DNA polymerase	Biorad	4304437
Dulbecco's Modified Eagle's Medium	Sigma Aldrich	D5796
TrypLE Express	GIBCO	12406-013
Cell freezing medium	Sigma	C6295
DAPT (N-[N-(3,5-difluorophenacetyl)- L-alanyl]-S-phenylglycine t-butyl ester)	Sigma-Aldrich	D5942
Phosphate buffered saline	Sigma	D8537
Long Range PCR kit	Qiagen	206402
Ion shear plus reagents -	Life Technologies	4471248
Hot star Taq plus DNA polymerase enzyme	Qiagen	203605
Platinum PCR Super mix high fidelity	Invitrogen	12532-016
OneTouch Reagent Mix	Life Technologies	4468660
The Ion PGM 200 Sequencing Kit	Life Technologies	447400
EXOSAP-IT	Usb	**78200**
DYEnamic ET dye terminator kit	GE healthcare	
SOLiD Fragment Library Construction Reagents	Applied Biosystems	4443713
Egel	Invitrogen	G661002

Reagent / Kit	Manufacturer	Catalogue Number
EZ4U Cell Proliferation and Cytotoxicity Assay	Biomedica Gruppe	BI-5000
Halt Protease Inhibitor cocktail	Thermo Fisher Scientific	JI120584
Bradford ULTRA	Novexin	BFU05L
10% precast polyacrylamide gel	Bio-Rad	456-1033EDU
NOTCH1 primary antibody	Abcam	ab52627
NOTCH2 primary antibody	Abcam	ab137665
NOTCH3 primary antibody	Abcam	ab23426
NOTCH4 primary antibody	Abcam	ab134831
Horseradish peroxidase conjugated rabbit secondary antibody	Amersham, GE Healthcare	NA934
Horseradish peroxidase conjugated mouse secondary antibody	Amersham, GE Healthcare	NA931
Amyloid beta precursor protein antibody	Abcam	ab32136
β-actin antibody	Abcam	ab8227
Chemiluminescence	GE Healthcare	RPN2108