MOLECULAR PROFILING OF PRIMARY HEAD AND NECK SQUAMOUS CELL CARCINOMA AND LYMPH NODE METASTASES

Neeraj Sethi

MBChB, MRCS, DOHNS

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy (PhD)

The University of Leeds

School of Medicine – Leeds Institute of Cancer & Pathology

September 2015

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Publications:

1. **Sethi N**, Wright A, MacLennan K, Wood HM, Rabbitts PH. The past and future impact of next generation sequencing in head and neck cancer. **Head and Neck.** 2015: doi: 10.1002/hed.24085.
2. **Sethi N**, Wright A, Wood HM, Rabbitts PH. MicroRNAs and head and neck cancer: Reviewing the first decade of research. **European Journal of Cancer.** 2014; 50: 2619-35.

Acknowledgements

Abstract

INTRODUCTION

The presence of lymph node metastases and/or extracapsular spread (ECS) has a significant impact on patient survival in Head and neck squamous cell carcinoma (HNSCC). Little is known about the molecular mechanisms associated with metastasis. A marker that could predict metastasis from primary tumour sampling could be of great clinical benefit for patients. Similarly in oropharyngeal squamous cell carcinoma (OPSCC), the molecular changes associated with human papilloma virus are incompletely understood. The impact of viral load has not been well explored and could help identify molecular markers associated with Human papillomavirus (HPV)-driven OPSCC.

METHODS

Tissue samples were identified from Leeds Pathology Archive and nucleic acid extracted from these. This was processed into sequencing libraries and analysed for copy number alteration (CNA) and microRNA (miRNA) profiles in clinicopathologic groups relating to metastasis and HPV viral load.

RESULTS

A panel of 14 CNAs was identified as associated with nodal metastasis and loss of 18q21.1-q21.32 was associated with ECS. The fraction of genome altered (FGA) was also increased in metastatic primary tumours. A panel of 19 CNAs was identified as associated with no detectable viral load and the FGA was found to be increased in this group of OPSCC. Twelve miRNAs were identified as associated with nodal metastasis.

DISCUSSION

The CNA and miRNA profile of primary tumours was found to be largely similar, though not identical, highlighting the need to use metastatic tissue to attempt discovery of metastatic molecular markers. Integrating miRNA and CNA data suggested miRNA expression is not governed by CNA. Potentially translational marker for metastasis and OPSCC with no viral load have been identified.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°C (suffix)	degrees Celsius
aCGH	Array Comparative Genomic Hybridisation
cDNA	complementary DNA
CGH	Comparative Genomic Hybridisation
Chr	Chromosome
CN	Copy number
CNA	Copy number alterations
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
ECS	Extracapsular Spread
FFPE	Formalin-fixed paraffin-embedded
GISTIC	Genomic Identification of Significant Targets in Cancer
GO	Gene ontology
H&E	Hematoxylin and Eosin
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
IHC	Immunohistochemistry
ISH	In-situ hybridization
KEGG	Kyoto Encyclopedia of Genes and Genomes
lincRNA	Large intergenic non-coding RNA
LOH	Loss of heterozygosity

ml	Millilitre
mRNA	Messenger RNA
miRNA	microRNA
miRNASeq	microRNA sequencing
ng	Nanogram
NGS	Next Generation Sequencing
OP	Oropharyngeal
OPSCC	Oropharyngeal squamous cell carcinoma
OSCC	Oral squamous cell carcinoma
PCR	Polymerase chain reaction
PI3K	phosphoinositide 3-kinase
RNA	Ribonucleic acid
RNASeq	RNA sequencing
RNAseq	RNA sequencing
rRNA	Ribosomal RNA
RT	Room temperature
RT-PCR	Reverse transcription PCR
SCC	Squamous cell carcinoma
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
TSG	Tumour suppressor gene
WHO	World Health Organization

Chapter 1

INTRODUCTION AND AIMS OF THESIS

1.1 Head and neck squamous cell carcinoma

1.1.1 Epidemiology

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide and affects over 600 000 people annually (Ferlay J, 2013). In 2012, in England the overall age-standardised incidence of head and neck cancer was 26.9 per 100 000 for men and 10.9 per 100 000 for women per annum (see Table **1-1**) (Statistics, 2012).

Sub-site	Total new cases (2009)		Age-standardised rate per 100,000 per year	
	Male	Female	Male	Female
Oral, Lip & Pharynx	4335	2274	18.9	8.8
Nasal Cavity & Middle Ear	153	125	0.7	0.5
Paranasal sinuses	94	59	0.4	0.2
Larynx	1521	355	6.9	1.4

Table 1-1: New cases and age-standardised incidence in 2012 in England (Office of National Statistics, 2012)

HNSCC is often described as a heterogeneous group of cancers, in that they are biologically similar (approximately 95% are squamous cell carcinoma) but they are clinically distinct dependent on the primary tumour subsite. HNSCC refers to malignant squamous disease of any subsite of the upper aerodigestive tract. These represent a challenging group of tumours to treat

due to their location and the often disfiguring and functionally destructive treatments, whether surgical or chemoradiotherapy (and frequently both). The anatomical classifications of the various subsites are demonstrated in Table **1-2**.

Subsite	Anatomical areas within subsite
Oral Cavity	Lip Oral (anterior 2/3) tongue Buccal mucosa Alveolar ridge Hard palate Retromolar trigone
Oropharynx	Soft palate Palatine tonsils Tonsillar fossa and pillars Base (posterior 1/3) of tongue Posterior wall pharynx
Nasopharynx	Superior, posterior and lateral walls Fossa of Rosenmüller
Larynx	Supraglottis Glottis Subglottis
Hypopharynx	Pyriform sinus/fossa Postcricoid area Posterior pharyngeal wall
Nasal Cavity	Septum Floor Lateral wall
Paranasal sinuses	Maxillary sinus Ethmoid sinus Frontal sinus Sphenoid sinus

Table 1-2: Subsites and the anatomical areas within each subsite

There is a wide geographic variation in the incidence of HNSCC (approximately 20-fold). High incidence areas include South and Southeast Asia (e.g. India, Taiwan), parts of Eastern Europe, Latin America, the Caribbean and Pacific regions (e.g. Papua New Guinea) (Warnakulasuriya, 2009). In high risk countries, such as India, HNSCC is the most common cancer in men and contributes up to 25% of all new cancers (Warnakulasuriya, 2009). This variation is attributed to the high incidence of betel nut chewing in addition to potential dietary factors and social deprivation (see Figure **1-1**).

Trends in survival in HNSCC have shown little improvement over the last three decades as demonstrated in the Oxford Cancer Intelligence Unit report (Price G, 2009). Improvements that have been achieved are put down to generally more aggressive treatments, at the cost of increased toxicity to the patient. Treatment failures most commonly manifest as distant metastases (Ferlito *et al.*, 2001). The ability to stratify patients and their disease to identify those most at risk of recurrence or requiring combined modality therapy has been at the forefront of HNSCC research for decades with little progress.
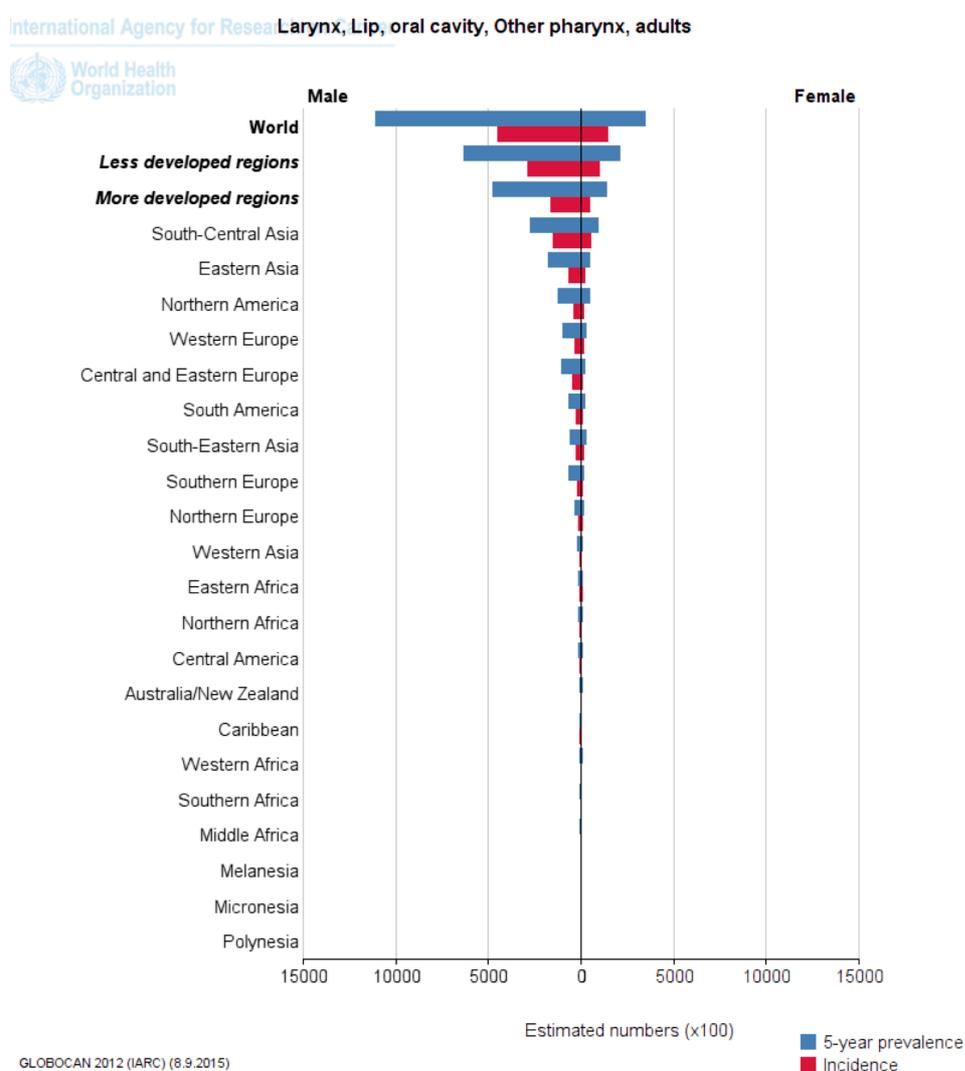
Figure 1-1: Bar chart to show the global variation in incidence/prevalence of HNSCC, produced from WHO GLOBOCAN 2012 data (reproduced with kind permission WHO, 2012).

1.1.2 Aetiology

The most well known source of carcinogenic exposure in HNSCC is cigarette smoking and alcohol consumption. The relative risk of developing HNSCC is 3 to 12 times higher for smokers than non-smokers (Lewin *et al.*, 1998). The risk increases with the amount of tobacco smoked and the duration of smoking (Elwood *et al.*, 1984). Alcohol and smoking have a synergistic effect in inducing malignancy, though in populations who have never smoked there is no conclusive evidence that alcohol is an independent risk factor for HNSCC (though this particular patient group is always small in epidemiological studies on HNSCC) (Licitra *et al.*, 2006, Brennan *et al.*, 2004, Menvielle *et al.*, 2004a).

In non-Western populations, local habits such as betel quid chewing are responsible for a great deal of HNSCC (e.g. India). The betel nut (also known as the areca nut) is harvested from the tropical palm, *Areca catechu*, which is widely cultivated throughout India, Sri Lanka, Thailand and the Philipines and consumed throughout Asia (Croucher and Islam, 2002). The fruit is harvested and the hard, dry buts removed. The nut is often cut into small pieces and chewed either alone or mixed with a variety of additives, which are often wrapped in the leaf of the *Piper betle* plant and referred to as betel quid. The additives vary from region to region and can include tobacco, slaked lime amongst other herbs, fruits and perfumes (Sharan *et al.*, 2012). Betel nut chewing is a socially acceptable and widespread practice amongst Asian and Arabic populations (even in children) and is highly addictive (Yusuf and Yong, 2002).

There is a strong, causal association between betel nut chewing (with and without tobacco) and HNSCC, as well as the precancerous conditions oral leukoplakia, erythroplakia and oral submucous fibrosis (Sharan *et al.*, 2012, Yusuf and Yong, 2002). The betel nut represents a significant cash crop. It is commercially prepared both as a fresh or dried preparation, separately or within mixed preparations. The nut is exported from Asia and available for purchase in retail outlets in Western countries including the UK, albeit largely

within immigrant communities. Other uses such as toothpaste or shampoo have also been developed (Croucher and Islam, 2002).

A diet deficient in fruit and vegetables is also a recognised risk factor for HNSCC (Esteve *et al.*, 1996). A reduced risk of HNSCC has been associated with high intake of vitamins A, C, E and a high polyunsaturated/saturated fatty acids ratio (Esteve *et al.*, 1996, Brennan *et al.*, 2004).

Social inequalities have long been recognised in these cancers. Significantly elevated risks have been observed amongst lower socioeconomic groups, with some evidence suggesting socioeconomic deprivation also has a significant detrimental impact on survival in patients with head and neck cancer (Elwood *et al.*, 1984, Hwang *et al.*, 2013). Explanations for this have often cited alcohol and tobacco consumption though a proportion could also be attributable to occupational exposures (Menvielle *et al.*, 2004b).

In addition to these traditional risk factors human papilloma virus (HPV) has emerged as an independent risk factor for developing HNSCC (particularly HPV-16). HPV-positive HNSCC is genomically, histologically and clinically distinct from HPV-negative HNSCC (Hennessey *et al.*, 2009). Authors have reported cancers from the oral cavity and oropharynx as collectively being "oral cancer". This has led to confusion in interpreting both the incidence of these cancers as well as the incidence of HPV in these cancers (da Silva *et al.*, 2011, Chaturvedi *et al.*, 2008). This subtype of HNSCC is most prevalent in the oropharynx, where there is a preponderance of lymphoid tissue relative to other sub-sites. Subsites outside the oropharynx appear to have a low incidence of HPV-driven disease (Upile *et al.*, 2014).

Epstein Barr virus (EBV) is implicated in the pathogenesis of nasopharyngeal cancer (NPC). EBV has been detected in dysplastic lesions and invasive NPC. Antibodies to EBV are reported as prognostic in EBV-associated cancer (Licitra *et al.*, 2006, Lin *et al.*, 2004).

A large body of cancer research has focused on the ability to identify features of different clinicopathologic groups that have prognostic

significance. The aim of this work is to stratify patients into different risk groups and tailor their treatment accordingly.

1.1.3 Prognostic features of the primary tumour

Increasing size of primary tumour is well recognised to be associated with poorer outcome and as such forms the basis of both the Union for International Cancer Control (UICC) and American Joint Committee on Cancer (AJCC) staging classifications for head and neck cancer according to the Tumour Nodes Metastases (TNM) system (Sobin LH, 2009). This varies according to site and subsite of primary tumour but ultimately size is the important factor (see Table **1-3**).

TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ
T1	Tumour 2cm or smaller in greatest dimension
T2	Tumour larger than 2cm but 4cm or smaller in greatest dimension
T3	Tumour larger than 4cm in greatest dimension
T4a	Tumour invades the larynx, deep/extrinsic muscles of the tongue, medial pterygoid, hard palate or mandible
T4b	Tumour invades lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, or skull base or encases carotid artery

Table 1-3: T-staging for oral squamous cell carcinoma (OSCC)

The histological differentiation of the primary tumour (well, moderate or poor) is a standard pathological diagnostic parameter described by Broders in 1920 (Broders, 1920). This is based on the resemblance of the tumour to the tissue of origin with well-differentiated tumours resembling their tissue of origin more closely than poorly differentiated tumours. Well-differentiated tumours contain orderly stratified cells and heavy keratinisation in a pear formation whilst moderately differentiated tumours have prickle cells, some stratification and less keratinisation. Poorly differentiated tumours have

prominent nuclear pleomorphisms and atypical mitoses though they are still recognisable as a squamous cell carcinoma (SCC) (Mao *et al.*, 2004).

Though the histologic differentiation is a necessity for every histology report its prognostic significance has been analysed with varying results. This may be due to the subjective nature of the grading system. As scoring methods have been introduced to minimise this subjectivity and improve accuracy, studies have shown a trend towards worsening differentiation being associated with worse overall survival and higher risk of metastases (Roland *et al.*, 1992, Janot *et al.*, 1996).

Depth of tumour invasion has been shown to be associated with a higher risk of locoregional recurrence and nodal metastasis (Alkureishi *et al.*, 2008, Shah *et al.*, 1976). Other negative histological prognosticators of the primary tumour include the presence of perineural invasion, blood vessel invasion and lymphatic vessel invasion (Miller *et al.*, 2012, Yilmaz *et al.*, 1998, Michikawa *et al.*, 2012). These have all been shown to be associated with poorer overall survival and increased locoregional recurrence, though none of these is an absolute solitary indicator for adjuvant radiotherapy.

1.1.4 Prognostic features of lymph nodes

A lymph node is defined as an encapsulated collection of lymphoid tissue, of any size, which possesses a peripheral sinus (Ferlito *et al*., 2002). The UICC and AJCC have developed staging classifications for nodal disease (see Table 1–4) (Wittekind and Oberschmid, 2010).

For decades the single most important negative prognostic indicator has been considered to be the presence of cervical lymph node metastases. Finding this at the time of initial presentation is traditionally described to reduce the 5-year survival by approximately 50% regardless of the primary sub-site (Hahn *et al.*, 1987). This is of huge importance considering that lymph node metastasis can be demonstrated in 30–60% of patients with HNSCC across all sub-sites and occult metastatic disease is found in up to

30% who present with no clinical or radiological evidence of disease (Snow *et al.*, 1982, Coatesworth *et al*, 2002).

NX	Regional Lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in a single ipsilateral lymph node, 3cm or less in greatest dimension
N2a	Metastasis in a single ipsilateral lymph node, more than 3cm but not more than 6cm in greatest dimension
N2b	Metastasis in multiple ipsilateral lymph nodes, none more than 6cm in greatest dimension
N2c	Metastasis in bilateral or contralateral lymph node, none more than 6cm in greatest dimension
N3	Metastasis in a lymph node more than 6cm in greatest dimension

Table 1-4: N-staging for regional lymph nodes

This high risk of occult cervical lymph node metastasis has meant that many patients require treatment to their cervical lymph nodes even if preoperatively their clinical nodal stage is N0. This most commonly takes the form of surgery to remove the cervical lymph nodes at risk (referred to as a neck dissection). This surgery is not without considerable morbidity to the patient including risk of facial, vagus, phrenic, hypoglossal and accessory nerve injury, shoulder weakness, chylous fistula, infection and haemorrhage (Ferlito *et al.*, 2006a, Ferlito *et al.*, 2006b, Ferlito *et al.*, 2006c). 70% of those with a preoperatively N0 neck do not have any pathological evidence of metastases after surgery. Combining this with the risks of surgery means that a considerable number of patients are undergoing surgery without any tangible benefit.

Adopting an approach of observation for those with an N0 neck has been investigated in the past with significantly improved survival being demonstrated in patients with clinically N0 neck who have undergone neck dissection compared to those who did not receive treatment for the cervical nodes (Haddadin *et al.*, 1999, Mashberg and Meyers, 1976, Kligerman *et al.*, 1994). Therefore treatment for the cervical lymph node is recommended to

patients deemed to have a risk of occult cervical node metastases of 20% or more (i.e. neck dissection or radiotherapy).

Approximately 30% of lymph nodes containing metastatic SCC have been found to be 3mm or less in diameter. Given that pathologists often use techniques that only allow them to comment on lymph nodes 3mm or more in diameter there is potential for microscopic disease to be missed even in patients who are undergoing neck dissection (Buckley and MacLennan, 2000, Coatesworth and MacLennan, 2002).

Pathological features of the lymph nodes have also been examined for prognostic features. Numerous studies have found that increasing size of largest node is associated with decreased survival (Sessions, 1976, Richard *et al.*, 1987, Spiro *et al.*, 1974). Similar trends have been demonstrated with increasing number of positive nodes and anatomical level of nodal metastasis (Noone *et al.*, 1974, Snow *et al.*, 1982). These findings have not been uniform across all studies: neither Schuller *et al* nor Johnson *et al* found any significance in the impact of number of positive nodes or size of largest metastatic node (Schuller *et al.*, 1980, Johnson *et al.*, 1981). Survival in patients with HPV-positive OPSCC has been shown to be significantly better than those with HPV negative disease (Ang *et al.*, 2010). Previous studies contained tumours of mixed subsites and do not stratify by HPV-status, which may be responsible for the conflicting reports.

A vital feature of cervical lymph node metastases in HNSCC and arguably the most important marker for aggressive disease is the presence of extracapsular spread (ECS) in lymph node metastasis.

1.1.5 Extracapsular spread (ECS)

ECS refers to the extension of cervical lymph node metastasis beyond the capsule of the lymph node (see Figure **1**-**2**) (Coatesworth and MacLennan, 2002). Willis first described this in HNSCC in 1930 (Willis, 1930). Though this was subsequently well recognised as a finding in cervical metastases and Toker published theories as to the mechanism of deposition of tumour

that resulted in ECS, it was not until 1971 that Bennett *et al* first described an association between the presence of extranodal spread and reduction in patient survival (Toker, 1963, Bennett *et al.*, 1971)

Lymph node with ECS (arrows)   Lymph node without ECS
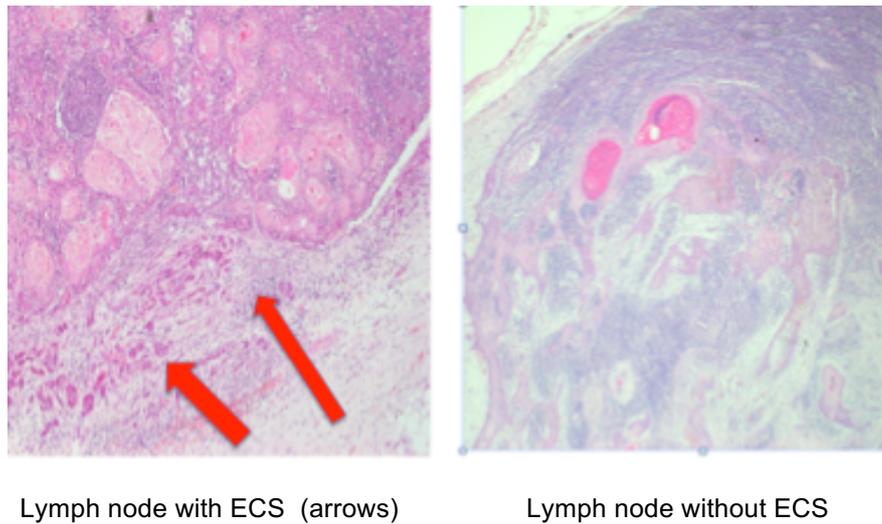
Figure 1-2 – Haematoxylin and eosin histological slide image of extracapsular spread

Since then a number of studies have demonstrated the strong association between extracapsular spread and reduced overall survival (Noone *et al.*, 1974, Sessions, 1976, Shah *et al.*, 1976, Kalnins *et al.*, 1977, Schuller *et al.*, 1980, Johnson *et al.*, 1981, Snow *et al.*, 1982). The incidence of ECS in cervical lymph node metastases varies considerably in the literature from 21% to 85% (Carter *et al.*, 1987, Jose *et al.*, 2003, Hirabayashi *et al.*, 1991, Mamelle *et al.*, 1994). There is also evidence of interobserver variation between pathologists in assessing the presence of ECS. Brekel *et al* found inter-observer variability kappa of 0.44 between pathologists (indicating moderate agreement) (van den Brekel *et al.*, 2012). This is likely due to the lack of a universally accepted definition for ECS. In the last 15 years a proposed definition, which appears to be widely accepted is: tumour extension beyond the lymph node capsule with a desmoplastic stromal response (Coatesworth and MacLennan, 2002). The acknowledgement of desmoplasia associated with true ECS reduces the chances of false

positives due to mechanical handling of the lymph nodes during surgery or subsequent processing in the pathology department.

The variation in incidence and assessment is thought to be strongly related, once again, to the lack of a universally accepted definition for ECS (van den Brekel *et al.*, 2012). Similarly, though several studies have suggested the likelihood of ECS increases with increasing nodal stage and nodal size this relationship is also not universally described (Jose *et al.*, 2003, Brannan *et al.*, 2011, Hirabayashi *et al.*, 1991, Mamelle *et al.*, 1994). A large study utilising the Surveillance, Epidemiology and End Results (SEER) database did not find a linear relationship between ECS and nodal size (Brannan *et al.*, 2011). This may also be related to variation in definition and agreement but it may also be due to mixing HPV-16 positive and negative disease.

Many studies have either included oropharyngeal cancers in their analysis or not clearly distinguished oral from oropharyngeal primary tumours. More recent studies evaluating prognostic factors in HPV-16 positive oropharyngeal cancer have found that ECS does not appear to have a negative impact on disease-specific survival or overall survival (Sinha *et al.*, 2012, Lewis *et al.*, 2011, Haughey and Sinha, 2012). Larger, more recent studies of oral cavity primary tumours (a subsite recognised to have a very low incidence of HPV) show ECS to have the largest negative impact on patient outcome (including overall and disease-free survival and distant metastases) (Shaw *et al.*, 2010, Goldstein *et al.*, 2013).

The impact of ECS on survival is so marked that when studies have stratified patients into those that are pathologically N0 (pN0), those with node metastases (N+) without ECS and those N+ with ECS a clear difference is seen. Survival in pN+ patients without ECS is worse than pN0 patients but significantly better than patients found to be pN+ with ECS (see Figure **1-3**) (Jose *et al.*, 2003, Shaw *et al.*, 2010). ECS also closely correlates with other adverse histological parameters such as perineural invasion and poor differentiation, though not closely enough to be accurately predictive of ECS (Shaw *et al.*, 2010, Lanzer *et al.*, 2014).
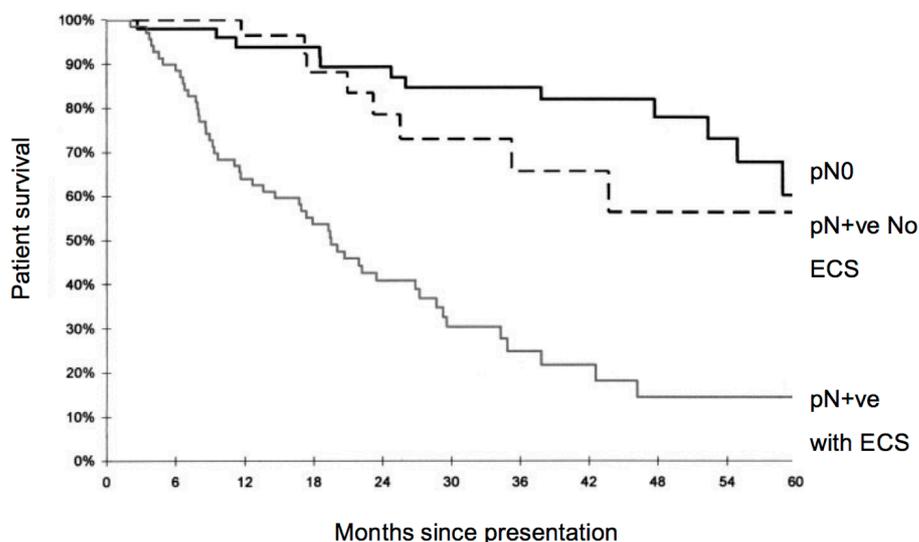
Figure 1-3 - Kaplan-Meier plot of survival comparing patients with HNSCC and differing pathological neck nodal status. (Reproduced with kind permission Jose *et al*, 2003)

The histological approach may also impact on the reported incidence of ECS. Macroscopic ECS has been shown to be associated with significantly reduced survival (Johnson *et al.*, 1981). Microscopic extracapsular spread (defined as that which is not visible macroscopically) was initially regarded as insignificant (Brasilino de Carvalho, 1998). Jose *et al* and Coatesworth *et al* suggested that it could be as important as macroscopic ECS reducing 5-year survival, putting patients at a higher risk of loco-regional recurrence and distant metastases (Coatesworth and MacLennan, 2002, Jose *et al.*, 2003). Greenberg *et* al and Woolgar *et al* demonstrated clearly that microscopic ECS had as significant negative impact on survival as macroscopic ECS (see Figure 1-4) (Woolgar *et al.*, 2003, Greenberg *et al.*, 2003).

Variations in specific laboratory protocols for processing the neck dissection specimens may be responsible for the fact that previously microscopic ECS was found to be insignificant. Some advocate the palpation of the specimen for lymph nodes and subsequent dissection of these palpable lymph nodes from the specimen to be examined. In the Leeds studies, the entire specimen was embedded and sliced for examination (Jose *et al.*, 2003, Coatesworth and MacLennan, 2002). This obviously removes any possibility

of missing smaller lymph nodes and potentially increases the sensitivity for microscopic extracapsular spread but is more labour intensive as it produces more sections for examination. Despite using the palpation and dissection pathologic technique, in the largest, more recent series Shaw *et al* also found no significant difference between 5-year overall survival in patients with macroscopic or microscopic ECS (Shaw *et al.*, 2010). Indeed the Royal College of Pathologists advises there is no need to differentiate between microscopic and macroscopic ECS when reporting on HNSCC specimens (Helliwell T, 2013).
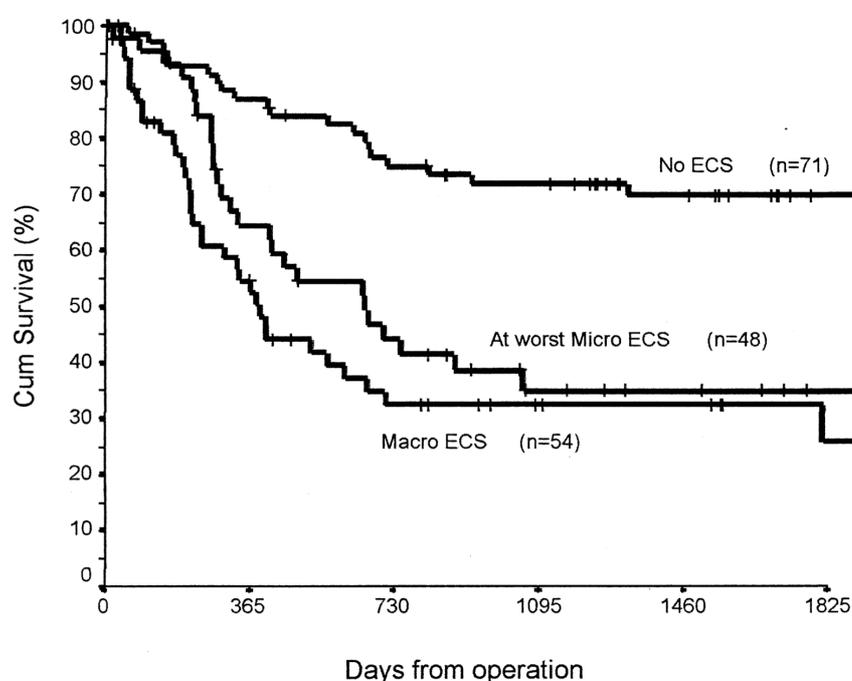
Figure 1-4 - Kaplan-Meier survival classified by presence and extent of extracapsular spread (reproduced with kind permission Woolgar et al, 2003).

It is important to remember that ECS is seen in small, single lymph nodes (including macroscopic ECS) and therefore that this is not purely a sign of late disease; it is a sign of aggressive disease. This also reflects a failing of the UICC and AJCC cancer staging system, which does not include it at all in the nodal staging system. The fact that patients in whom ECS is identified are at higher risk of local recurrence again suggests that this is a sign of biologically aggressive disease.

Again it should be highlighted that the risk of occult neck node metastasis in patients with HNSCC is the reason that the cervical lymph nodes are treated. Where the primary tumour is to be treated with radiotherapy then the neck can also be treated with radiotherapy. Though this has the advantage of only subjecting the patient to a single modality of treatment it is also associated with considerable life-long morbidity (Bentzen *et al.*, 2001, Silverman Jr, 2003). When the primary tumour is treated surgically, the lymph nodes are generally treated surgically (a neck dissection is performed). This has the added advantage of allowing pathologic examination of the lymph nodes enabling the identification of patients with ECS. The latter has been established as an absolute indication for postoperative adjuvant therapy (radiotherapy with or without chemotherapy) as these patients are at double the risk of local recurrence and have triple the incidence of regional relapse (Shaw *et al.*, 2010, Bartelink *et al.*, 1983, Peters *et al.*, 1993). However, this does mean that the majority of patients with a preoperatively N0 neck are undergoing surgery due to the inability to recognise a primary tumour that will metastasise causing ECS.

Several studies have attempted to assess the utility of imaging in detecting ECS. These have involved computed tomography (CT) and magnetic resonance imaging (MRI). Overall they suggest these imaging modalities to be moderately accurate in terms of sensitivity and specificity (Yousem *et al.*, 1992, Zoumalan *et al.*, 2010, Souter *et al.*, 2009, Kimura *et al.*, 2008, Hao and Ng, 2000, King *et al.*, 2004, Lodder *et al.*, 2013). Across a number of studies the CT has been found to have a sensitivity ranging from 50–100% and specificity of 85–100% (Yousem *et al.*, 1992, Souter *et al.*, 2009, King *et al.*, 2004, Lodder *et al.*, 2013). MRI has been found to have a sensitivity ranging from 50–77% and specificity of 90–93% (King *et al.*, 2004, Yousem *et al.*, 1992, Kimura *et al.*, 2008). This may improve as the quality of images produced by newer generations of scanners improves, but is not currently at an acceptable, consistent level for images to be used to guide treatment in this way. More recently 18-fluorodeoxyglucose positron emission tomography (FDG-PET) has been evaluated for the ability to predict ECS in

HNSCC. This has been found to have a sensitivity of 80% and specificity of 74–88% (Joo *et al.*, 2013b, Joo *et al.*, 2013c, Joo *et al.*, 2013d).

In terms of predicting occult cervical node metastases a meta-analysis by Liao *et al* comparing different imaging modalities, including CT, MRI, PET and ultrasound found that all had similar diagnostic accuracy (Liao *et al.*, 2012). All were found to have higher negative predictive probabilities compared to positive probabilities but again none were consistently high enough to warrant widespread acceptance of using this as a method to avoid treating the clinically N0 neck. The situation of having to upstage a radiologically N0 patient to pN+ with ECS transitioning them from the most favourable prognostic group to the least favourable is well recognised (Shaw *et al.*, 2010).

It could be advantageous to recognise or predict ECS accurately with a molecular marker because imaging is not sufficiently sensitive and specific. The poor prognosis of patients with ECS makes them ideal candidates to be targeted for neoadjuvant treatments or for additional biological therapies. Conversely being able to reliably identify those without any clinically or radiologically occult nodal metastases could spare those patients unnecessary surgery.

1.1.6 Genomic alterations in HNSCC

Cancer is considered to be a genetic disease with carcinogenesis being a multistep process that reflects genomic alterations that underlie the transformation of normal cells into malignant cells (Hanahan and Weinberg, 2000). These genomic alterations are considered the key drivers behind the cancerous cell gaining the characteristics deemed to be hallmarks of cancer. These include: resisting cell death, evasion of growth suppressors, induction of angiogenesis, sustenance of proliferative signalling, replicative immortality and activation of invasion and metastasis (Hanahan and Weinberg, 2000).

Foulds demonstrated the progressive development of cancer via a number of premalignant states in animal models (Foulds, 1954). Fearon and

Vogelstein established the morphological development of colorectal cancer occurring in parallel with a stepwise, progressive accumulation of genetic alterations (Fearon and Vogelstein, 1990). Vogelstein *et al* also suggested that carcinogenesis is a result of the accumulation of the genetic alterations rather than the specific order in which they occur. Califano *et al* produced a similar study in HNSCC (see Figure **1**-**5**) (Califano *et al.*, 1996). This work used microsatellite analysis to determine which chromosomal losses are associated with different steps of carcinogenesis. Loss of 9p21 or 3p21 was one of the earliest detectable events leading to dysplasia en route to invasive HNSCC.
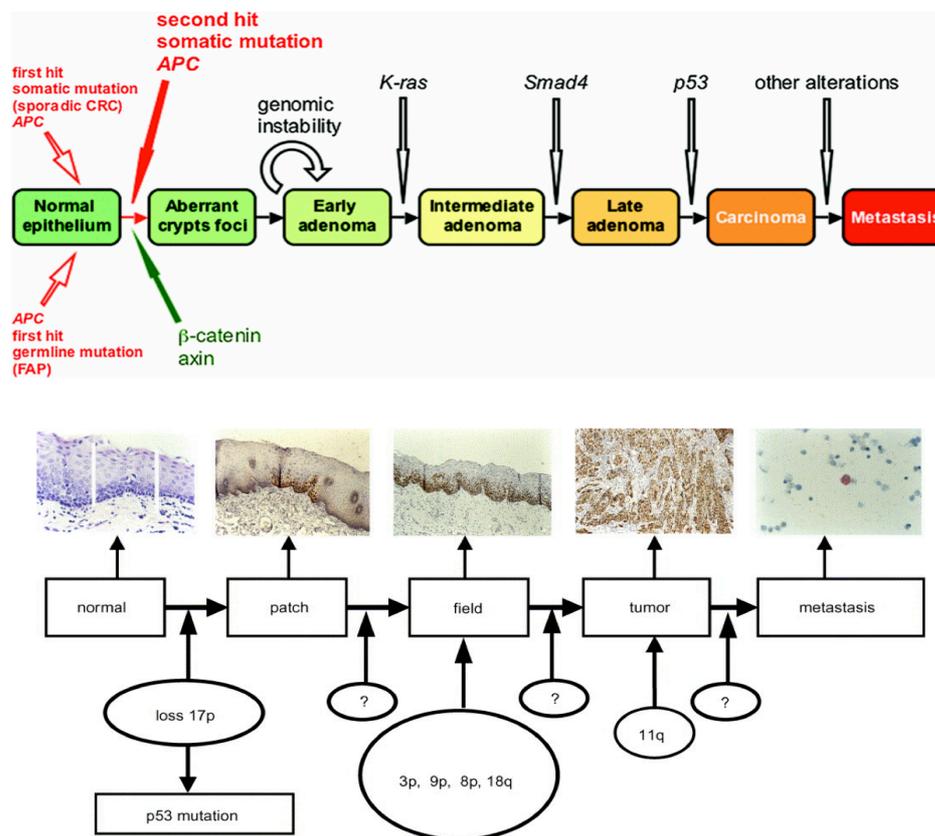
Figure 1-5 - Above is the genetic model for colorectal tumourigenesis and below the HNSCC model (Reproduced with kind permission Fearon and Vogelstein, 1990 and Califano et al, 1996).

Subsequent technological advances in the ability to understand and catalogue the human genome have allowed the cancer genome to be

understood with much greater resolution. The concept of sequential genomic alterations rather than isolated events have been substantiated by the wide field of work in cancer genomics (Tamborero *et al.*, 2013). Since the first somatic mutation in a human cancer gene (G12V in HRAS in human bladder carcinoma) over 660 cancer genes have been curated and included in the Cancer Gene Census (Forbes *et al.*, 2015). Though this is only a fraction of the 20 000 genes in the human genome, a challenge has been the identification of genes that drive tumourigenesis as the overall list of genes altered in cancer may be in the thousands (Ezkurdia *et al.*, 2014). "Driver" genes endow the tumour with a growth advantage over surrounding cells when altered whilst the remaining alterations can be regarded as being within "passenger" genes (Vogelstein and Kinzler, 2015).

Single nucleotide variants and copy number alterations of chromosomal regions have been identified as two of the primary alterations in human cancer (Schroeder *et al.*, 2014). The genes affected by these alterations are termed oncogenes (genes that increase or change their function upon somatic variants) and tumour suppressor genes (genes whose product tend to lose function) dependent on their role in cancer development. Copy number gain or loss may heighten or reduce the function of these and this may highlight a way of identifying driver genes in cancer.

1.1.7 Oncogenes and tumour suppressor genes in HNSCC

Some of the specific genes associated with the hallmarks of cancer listed below have been elucidated.

1.1.7.1 Replicative immortality

The p53 and Rb pathways have been demonstrated to be crucial to this characteristic of HNSCC. *TP53* is an established tumour suppressor gene in HNSCC, with mutations being identified in over 85% of HPV-negative HNSCC and inactivation of *TP53* being found in HPV-positive tumours, via HPV-16 E6 oncoprotein (Poeta *et al.*, 2007, Stransky *et al.*, 2011, Agrawal *et al.*, 2011). As such, *TP53* represents an important driver gene in HNSCC as

it is one of the few driver genes identified as altered in a high proportion of HNSCC. In health p53 is produced in response to cellular insult or deoxyribonucleic acid (DNA) damage and induces cell cycle arrest and/or apoptosis (Gleich and Salamone, 2002). Inducing overexpression of a dominant-negative mutant of *TP53* or inducing inactivation via ectopic HPV-16 E6 in cell lines has been shown to extend lifespan (Opitz *et al.*, 2001, Rheinwald *et al.*, 2002). p53 can be inactivated by HPV-16 E6, which binds p53 and targets it for degradation, upregulation of *MDM2*, which also target p53 for degradation or loss of *CDKN2A*, which downregulates *MDM2* (Rothenberg and Ellisen, 2012). *CCND1* lies on 11q13. Its product phosphorylates Rb leading to cell cycle progression (Gleich and Salamone, 2002). 11q13 is one of the most commonly amplified regions in HNSCC. *CDKN2A* encodes for p16, which slows G1/S phase progression (Nobori *et al.*, 1994). It is frequently inactivated by genetic (e.g. mutation or deletion) or epigenetic alterations (e.g. methylation) in HNSCC. These genes act within the Rb pathway and these abnormalities contribute to HNSCC cell immortalisation (Smeets *et al.*, 2011).

1.1.7.2 Evasion of growth suppression and sustenance of proliferation

EGFR is an important gene in HNSCC. It encodes the protein EGFR which is involved in the MAPK signalling and PI3K/PTEN/AKT signalling pathways as well as inducing *CCND1* (Leemans *et al.*, 2011). *EGFR* is amplified in up to 30% of HPV-negative HNSCC, whilst mutations are rare (Seiwert *et al.*, 2015, Sheu *et al.*, 2009). Overexpression of *EGFR* has been reported in up to 80% of HNSCC, though reported figures vary (Zimmermann *et al.*, 2006). Ectopic expression of *EGFR in vitro* transforms oral keratinocytes (Goessel *et al.*, 2005). Once activated EGFR initiates several downstream signal transduction cascades modulating DNA synthesis, cell adhesion, migration and proliferation (Zimmermann *et al.*, 2006). The pleiotropic nature of *EGFR* may underlie the fact is commonly expressed (Leemans *et al.*, 2011).

TGFβ is another signalling pathway responsible for inhibiting cell growth. Defective TGFβ signalling results in hyperproliferation and reduced

apoptosis (White *et al.*, 2010). These receptors have been reported to be downregulated in HNSCC (White *et al.*, 2010).

1.1.7.3 Resisting cell death

Apoptosis is an essential natural barrier for cells to overcome in order to gain a malignant phenotype (Hanahan and Weinberg, 2011). The PI3K/PTEN/AKT signalling pathway has been suggested to be critical to carcinogenesis. *PIK3CA* and *PTEN* are components of this pathway that function as oncogenes (Qiu *et al.*, 2006). Mutations or deletions of these genes increase the activity of this pathway, allowing cells to resist apoptosis. *EGFR* is also an upstream mediator of this pathway. In HNSCC mutations of *PIK3CA* have been identified in up to 20% of cases, with this incidence rising to 33% of HPV-positive tumours (Lui *et al.*, 2013). Multiple mutations of genes within the PI3K pathway have been associated with advanced stage HNSCC, whilst in a subset of HPV-positive tumour mutations of *PIK3CA* have been found to be the only mutated cancer gene (Lui *et al.*, 2013).

Loss of *TP53* function effectively removes a critical DNA damage detector and allows cells to evade apoptosis (Hanahan and Weinberg, 2011). *BCL2* also regulates apoptosis. Gain of function results in antiapoptotic cellular activity (Danial and Korsmeyer, 2004). Overexpression of *BCL2* has been reported in HNSCC and associated with chemotherapy resistance (Trask *et al.*, 2002, Sharma *et al.*, 2005).

1.1.7.4 Activation of invasion and metastasis

Concrete evidence for specific genes in metastasis in HNSCC is wanting, though many have been purported to be associated with metastasis. Roepman *et al* described a 102-gene panel associated with lymph node metastasis derived using microarray gene expression profiling (Roepman *et al.*, 2005). This high number of genes may reflect both the high level of inter-tumour heterogeneity and the complexity of epithelial-mesenchymal transition (EMT), which is thought to underpin the metastatic phenotype. Walter *et al* reported four molecular subtypes of HNSCC, based on

expression profiling. One subtype was classified as mesenchymal in light of an association with elevated expression of genes associated with EMT including *TWIST1* and *BMI1* (Walter *et al.*, 2013). *TWIST1* is activated via the PI3K/AKT, RAS and WNT signalling pathways. It has been shown to induce EMT in HNSCC cell lines and associated with poor outcome and chemotherapy resistance (Way *et al.*, 2014, Wu and Yang, 2011, Wu, 2011). *BMI1* has been shown to act co-operatively with *TWIST1* to repress levels of E-cadherin and p16INK4a, both of which have tumour suppressor function (Way *et al.*, 2014).

1.1.7.5 Angiogenesis

Neo-angiogenesis is essential to the growth of solid tumours and must be sustained in order to continue growth (Hanahan and Weinberg, 2011). Vascular endothelial growth factor (VEGF) is a member of the platelet-derived growth factor super family and a potent stimulator of new blood vessel growth (Vassilakopoulou *et al.*, 2015). This has been found to be overexpressed in HNSCC and associated with aggressive disease and decreased survival (Kyzas *et al.*, 2005a, Kyzas *et al.*, 2005b). Angiogenesis is induced by different mechanisms, highlighted by a study where anti-VEGF treatment had little effect on tumour growth in those with low levels of VEGF expression (Hasina *et al.*, 2008). Other cytokines including IL-8 and HGF have also been implicated in angiogenesis in HNSCC (Hasina *et al.*, 2008).

1.1.8 The host response in cancer

The concept of host immune surveillance protection against cancer was first postulated by Ehrlich (1909) and further adapted in the 1960s (P., 1909, Bhardwaj, 2007, Burnet, 1964). More recently, evidence that immunodeficient mice were at increased risk of spontaneous tumour development gave further credence to this theory (Shankaran *et al.*, 2001). There are thought to be three phases to this antitumour immunicity response:

(i)   Elimination – during this phase tumour cells are destroyed by elements of the innate and adaptive immune system. Tumour cells express antigens that can be recognised by host T cells, which can lyse tumour cells (Gajewski *et al.*, 2012, Badoual *et al.*, 2010). CD8+ cytotoxic T lymphocytes are capable of directly killing tumour cells. CD4+ T helper type 1 lymphocytes are essential to activating C8+ lymphocytes and type 2 lymphocytes stimulate humoral immunity and eosinophils (Gooden *et al.*, 2011).

(ii)   Equilibrium – in this period the tumour cells persist but do not expand due to "equilibration" by the immune system. A population of immune-resistant cells appears whilst a consistent, effective immunological response is ongoing against non-resistant tumour (Gooden *et al.*, 2011). This phase can last for years (Kim *et al.*, 2007).

(iii)   Escape – the tumour cells actively disable immune recognition and evade immune destruction by co-opting immune cells for growth, angiogenesis and invasion (Bhardwaj, 2007). These escape mechanisms include loss of antigen-presenting machinery, tolerisation of immune cells, induction and recruitment of regulatory immunosuppressive cells (e.g. CD4+ T regulatory lymphocytes actually suppress effector T lymphocytes and are trafficked to tumours by chemokines released by tumour cells) and activation of proinflammatory cytokines that stimulate angiogenesis (Badoual *et al.*, 2010, Gooden *et al.*, 2011). This phase enables processes not just key to cancer development but also to metastasis.

Evasion of immune destruction is regarded as a emerging hallmark of cancer (Hanahan and Weinberg, 2011). Immunosuppressed patients are more likely to develop HNSCC, than immunocompetent patients, as well as more likely to respond poorly to treatment (Engels *et al.*, 2011). In established HNSCC, an endogenous immune response is prognostic. In HPV-positive OPSCC, the presence of tumour-infiltrating lymphocytes such

as CD4+ and CD8+ lymphocytes have been found to be prognostic (Ward *et al.*, 2014). The immune response, reflected by the presence of tumour infiltrating lymphocytes has a important role in the improved survival seen in most HPV-positive tumours. Correlation has also been found between the presence of tumour infiltrating immune cells and circulating immune cells, suggesting promise for a peripheral biomarker to be developed in the future (Green *et al.*, 2013).

1.1.9 Copy number markers for metastasis and ECS in HNSCC

CNAs are a form of structural variation of the genome. The "normal" human genome is considered to have a copy number of two (in that it is diploid). CNA has been defined as a segment of DNA 1 kb or larger that is present at a variable copy number compared to the reference genome (though smaller CNAs have been reported) (Feuk *et al.*, 2006). This can result from duplication or insertion (gain) or a deletion (loss). These regions of gain and loss can represent insertions, deletions and duplications.

Early studies utilised cytogenetic techniques that are often limited to evaluating changes in targeted areas of a chromosome and the breadth of the whole genome is not considered. Loss of heterozygosity (LOH) only enabled assessment of changes in the ratio of length of DNA between normal and cancerous tissue. This was therefore dependent on a candidate driven approach. Though inferences can be made LOH cannot directly discover gain or loss of DNA (copy number alteration, CNA) (Patmore *et al.*, 2005). Newer technologies such as comparative genomic hybridisation allow discovery of CNA.

Traditionally fluorescence in situ hybridisation (FISH) was used to demonstrate CNAs (Stankiewicz and Beaudet, 2007). However FISH has a relatively low resolution (approximately 5-10 Mbp) (Duan *et al.*, 2013). Quantitative real time polymerase chain reaction (qRT-PCR) can be used to evaluate genomic regions, however this requires sequence-specific primers to be designed. Not all regions of DNA will lend themselves to optimal primer design and so precision in genomic region can be lost (Hughes, 2007).

Microarrays can provide higher resolution (array comparative genomic hybridisation (aCGH) or single nucleotide polymorphism (SNP) arrays). The resolution of microarrays are dependent on the number of probes embedded in the array and therefore directly related to cost (Duan *et al.*, 2013). Advantages of these methods include the fact that high resolution can be obtained and that its limitations are well established. These include bias due to the poor sensitivity for detecting copy number gains compared to losses, the inability to detect copy neutral alterations (balanced translocations or inversions) and the fact that only CNAs targeted by the array probes can be interrogated (Stankiewicz and Beaudet, 2007).

The CNA profile of HNSCC is complex. Older studies, using low resolution CGH, identified an average number of alterations per HNSCC sample of 16 compared to 11 in lung carcinoma and 8 for colorectal carcinoma (Gebhart and Liehr, 2000). The chromosomal aberrations associated with metastasis in HNSCC are not well characterised. Though several studies have been conducted into this area variations in findings are common. This is likely due to differing methodologies, heterogeneity of primary tumour subsite and often small numbers of samples affecting individual studies over the years. A summary of the findings of studies into CNA associated with an increased risk of metastasis is shown in Table **1-5**. All of these studies have used technology that is limited to discovering alteration on targeted areas. The concept of using an untargeted, less-biased technology could reveal a more consistent or novel pattern for metastasis and ECS.

The literature on molecular markers for extracapsular spread in HNSCC is relatively sparse. Using FISH, epidermal growth factor receptor (*EGFR*) gene copy number aberrations (both gain and loss) were found to be associated with ECS in OSCC (Michikawa *et al.*, 2011). Another study found EGFR overexpression in 81% of primary tumours associated with ECS, though this study took place in a betel quid prevalent area (Chen *et al.*, 2003a). Correlations to ECS have been found with overexpression of *ERBB1* and *ERBB4* as well as *MMP12* (Kim *et al.*, 2013, Silva *et al.*, 2014).

Author	Tumour subsite	No of patients/ Tissue type	Method	CN Loss assoc with increased risk of lymph node metastasis	CN Gain assoc with increased risk of lymph node metastasis
(Joo et al., 2013a)	OSCC	14/FF	aCGH	3p14.2	
(Yoshioka *et al.*, 2013)	OSCC	25/FFPE	aCGH	1p22.1-13.3, 5q11.2-14.2, 9p13.3-13.1, 19p13.3	7p22.3, 7p22.2, 7p22.1-14.1, 7p13-12.3, 7p12.3-12, 7p12.1-11.2, 8q11.22-12, 8q22.1-24.21, 8q24.22-24.3, 17q24.3-25.1, 17q25.1-25.3
(Ambrosio *et al.*, 2013)	LSCC	32/FFPE	aCGH/ FISH	3q26.2, 18q23	11q13.3
(Xu *et al.*, 2013)	OSCC	37/FF	SNP array	2q31.3, 5p15.3, 7p22.3, 7p22.1, 8q22.1, 8q24, 11q13, 11q23-q24, 17q25.3, 22q11.1, 22q11.21	1p13.1, 3p25-p24, 3p25.3, 3p25, 3p25.1, 3p21.3-p21.2, 3p24.1, 3p22.3, 3p22.2, 3p22.1, 3p21, 3p21.31, 3p21.3, 3p14.1, 3p13, 3p12.3, 3p12-p11.1, 5q11, 5q31.2, 5q31.3, 5q31, 5q32, 5q33.1-q33.3, 5q32-q34, 5q35.3, 9p24.2, 9p24.1, 9p24, 9p22.3, 9p13.2, 9p11, 9q22.32, 9q33.3, 13q12.11, 13q12.2, 5q22.31, 15q, 15q26.1,18q12.3-q21.1, 18q21.31, 18q22, 18q23, 21q22.3
(Fendri *et al.*, 2009)	NPSCC	88/FF	qRT-PCR		3q26.3
(Sticht *et al.*,	OSCC	280/FFPE	FISH		3q25.31

2005)	PSCC LSCC				
(Myo *et al.*, 2005)	OSCC	45/FFPE	FISH		11q13
(Miyaguchi *et al.*, 2012)	OSCC	60/FF	SNP array	8p21.2	
(Bhattacharya *et al.*, 2011)	OSCC	152/FFPE	aCGH		3q24-qter, 8pter-p23.1, 8q12-q24.2, 20pter-qter
(Sugahara *et al.*, 2011)	OSCC	54/FF	aCGH qRT-PCR		11q13
(Chiang *et al.*, 2008)	OSCC	42/FFPE	qRT-PCR		7p11.2
(Fenic *et al.*, 2007)	OSCC PSCC LSCC	33/FF	qRT-PCR		3q26.3
(Xia *et al.*, 2007)	OSCC	33/FFPE	qRT-PCR		11q13.3
(Freier *et al.*, 2007)	OSCC	296/FFPE	FISH		17q25.3
(Noutomi *et al.*, 2006)	OSCC	35/FF	aCGH	4p	
(Lin *et al.*, 2005)	OSCC	66/unclear	qRT-PCR		3q26.32, 3q26.33
(Bockmuhl *et al.*, 2002)	OSCC OPSCC HPSCC LSCC	54/FF	CGH	5q34-q35, 8p12-p22, 10p12, 10q21-qter, 11p14-p15, 11q14, 11q23-qter, 4q21-qter	1q21-q22, 3q24-qter, 6q, 7q11.2, 12q12-q14, 8p11.2
(Welkoborsky *et al.*, 2000)	LSCC	20/Unclear	CGH	18q	11q13, 22q
(Quan and Guo, 1998)	LSCC	32/FF	PCR-PAGE		2p24.3
(Haughey *et al.*, 1992)	OSCC LSCC	8/FF	SB		8q24.21

Table 1-5: DNA copy number gain and loss associated with increased risk of cervical metastasis (FF – Fresh-frozen, FFPE – formalin-fixed paraffin embedded).

CNA has been used to delineate prognostic subgroups in breast cancer as well as lung cancer (Curtis *et al.*, 2012, Kim *et al.*, 2013). These demonstrate that though CNA has been investigated for many years it still presents an exciting avenue of biomarker discovery. This is also in part due to newer technologies, which have ben little explored in this area, such as next generation sequencing (NGS).

1.1.10   Next generation sequencing

The advent of NGS technology represented an epochal moment in cancer genomics. The first major studies in the use of NGS in HNSCC were published in 2011 (Agrawal *et al.*, 2011, Stransky *et al.*, 2011). Together these performed whole exome sequencing on 106 patients with oral, oropharyngeal, laryngeal, hypopharyngeal and sinonasal tumours (both HPV-positive and negative). These studies confirmed the findings of previous genomic work that *TP53* was the most commonly mutated gene in HNSCC (found in 60% of patients) and also discovered the second most commonly mutated gene was *NOTCH1* (in around 15% of patients) (Stransky *et al.*, 2011, Agrawal *et al.*, 2011). This was the first time *NOTCH1* had been implicated in HNSCC.

Interestingly these studies also found that HPV-positive tumours had approximately half the mutation rate of HPV-negative tumours (Agrawal *et al.*, 2011, Stransky *et al.*, 2011). On analysing subgroups they also found smokers had a higher rate of guanosine to thymidine point mutations, in addition to having a higher rate of mutations. On average they found 130 mutated genes per sample. This compares favourably to the TCGA report on 178 lung SCC samples where they found an average of 360 exonic mutations (Cancer Genome Atlas Research, 2012). The surprisingly low proportion of recurring mutations could be related to the mix of subsites reducing the number in each group, but gives a picture that each head and neck tumour is genomically quite different to the next.

In a follow up publication by Lui *et al* in 2013 a further 45 tumours had undergone whole exome sequencing and were combined with the previous

set (Lui *et al.*, 2013). Again a large number of mutated genes were identified per sample and a high degree of inter-tumour mutational heterogeneity observed. Developing their analysis, they focused on specific functional pathways that had previously been identified as targetable in cancer. PI3K pathway mutations were identified in 31% of their cohort. This signalling axis has been shown to have a role in cancer cell growth, survival, motility and metabolism (Samuels *et al.*, 2004, Engelman *et al.*, 2006, Courtney *et al.*, 2010). Lui *et al* found that PI3K-pathway mutated HNSCC contained a higher rate of mutations in known cancer genes and that those with concurrent mutations in *PI3K* pathway genes were all advanced tumours implicating his pathway in HNSCC progression (Lui *et al.*, 2013). This study highlighted the potential for NGS to identify possible therapeutic targets and biomarkers in HNSCC.

Pickering *et al* combined exome sequencing with SNP array copy number, gene expression, miRNA expression and methylation data in 40 OSCC patients (Pickering *et al.*, 2014). Using this integrated approach, they identified four major driver pathways in OSCC including mitogenic signalling, Notch, cell cycle and p53.

The Cancer Genome Atlas (TCGA) has reported preliminary findings from comprehensive genomic sequencing of 279 HNSCC patients. This included whole exome sequencing, whole genome sequencing and whole transcriptome sequencing as well as miRNA, DNA methylation and copy number profiling (Cancer Genome Atlas, 2015). The vast majority were HPV-negative tumours (243/279) and majority of subsites oral cavity and larynx (244/279). They found HPV-positive and negative tumours had different mutation profiles, with HPV-positive tumours exhibiting infrequent mutations in *TP53, CDKN2A, FAT1* and *AJUBA. TP53* mutations were present in 86% of HPV-negative tumours but only 1 of 36 HPV-positive tumours. *PIK3CA* was mutated in both tumour types but a specific mutation of the helical domain of *PIK3CA* was predominant in HPV-positive tumours – an important finding when considering targetable events. *EGFR* was found to be rarely aberrant in HPV-positive tumours compared to HPV-negative

tumours (Cancer Genome Atlas, 2015). This could have serious implications regarding the use of EGFR-inhibitors in these patients.

As the numbers of HNSCC tumours sequenced increases, therapeutic subgroups may be discovered. As the cost of sequencing and bioinformatics drops the concept of truly "personalised medicine" will become widely accessible. NGS is driving a shift away from purely pathological classification of tumours towards integrating clinically relevant genomic subgroups. Gross *et al* utilised TCGA data to identify loss of 3p as a marker for decreased survival. This effect was enhanced by concurrent presence of *TP53* mutation and miR-548k expression (Gross *et al.*, 2014). This demonstrates the importance of CNA, even in the setting of much more detailed mutational data.

Another recent study performed whole exome sequencing on 16 younger non-smokers with oral tongue cancer (< 45 years old) and 28 older smokers (Pickering *et al.*, 2014). The two groups were genomically similar, but on interrogating TCGA data for lung adenocarcinoma, bladder urothelial carcinoma and HNSCC, a smoking mutation signature was generated. Both young and older oral tongue cancers were found to be most similar to a non-smoking mutation profile. Though a small number of patients it demonstrates the accumulative power of NGS.

NGS also has applications for the determination of HPV-status by detecting copies of HPV DNA/RNA within the sample being sequenced. It also has the advantage that all sub-types of HPV can be screened for simultaneously (Conway *et al.*, 2012). This can be achieved with low-coverage and relatively low-cost NGS technology and can be performed as an additional analysis of the same sequencing data being obtained for other purposes at no extra cost. Issues with the use of this technology relate to the fact that detection of a single copy of HPV DNA within the sample does not mean the tumour was driven by HPV and there is no accepted standard for the number of detectable copies that should be regarded as a positive result. Conway *et al* found NGS to be comparable to polymerase chain reaction (PCR) and *p16*

immunohistochemistry with excellent sensitivity and specificity (Conway *et al.*, 2012). It has also been used to screen a large number of oral verrucous carcinoma samples establishing the scarcity of HPV in this type of oral cancer (Samman *et al.*, 2015).

Parfenov et al used NGS to investigate the tumour-host interaction in HPV16 positive HNSCC (Parfenov *et al.*, 2014). They examined whole genome sequencing and DNA methylation profiles in 35 HPV positive tumours and compared these to 270 HPV negative samples from the TCGA cohort. By doing this they were able to identify cancer genes at the sites of HPV DNA integration that were potentially disrupted and involved in the carcinogenic mechanism in virally driven HNSCC.

Intra-tumour heterogeneity has gained increasing prominence recently with landmark studies in renal cell carcinoma using NGS to demonstrate significant mutational difference in different samples from the same tumour (Gerlinger *et al.*, 2012, Gerlinger *et al.*, 2014). Three samples from a single oropharyngeal tumour and two samples from its corresponding cervical metastasis underwent whole genome sequencing in a study by Zhang et al (Zhang *et al.*, 2013). This found only 41% of all somatic point mutations were shared across all five samples. Wood *et al* used exome and copy number sequencing to demonstrate complex clonality within five patients with HNSCC. The clinical picture of these patients varied widely from a patient with multiple tumours over many years to a patient with a single tumour and lymph node metastasis. The recurrent or metastatic clone was identified and traced using their techniques (Wood *et al.*, 2015). This has significant implications for the use of targeted therapies.

The concept of integrative genomics (combining different methods of genomic analysis) is relatively novel. Large scale efforts using only a single genomic method such as exome sequencing analyses have not revealed a universal molecular marker. This likely reflects the complex RNA-genome interactions that exist in reality. A relatively recently discovered component

that could be valuable in the discovery of biomarkers are microRNAs (miRNAs).

1.1.11 MicroRNAs in HNSCC

miRNAs are a class of noncoding RNAs. They are approximately 18-25 nucleotides in length, and modulate gene expression by interacting with messenger RNA (mRNAs) post-transription (Almeida *et al.*, 2011).

miRNAs can bind to the 3'untranslated region (UTR), coding sequences or 5'UTR of the target mRNA and can inhibit translation or target mRNA for degradation (Filipowicz *et al.*, 2008). Since being discovered in 1993 in *Caenorhabditis elegans*, miRNAs have been found to regulate approximately 30% of mRNA transcripts in humans influencing functions such as apoptosis, proliferation and metabolism (Bartel, 2009, Ambros, 2008).

miRNAs were first documented to be conserved in the human genome in 2000 (Pasquinelli *et al.*, 2000) and since then there has been an exponential increase in the rate of discovery of miRNAs, though this may now be slowing.

1.1.11.1 miRNA Biogenesis

miRNAs are encoded in the genome as longer primary transcripts called primary-miRNAs (pri-miRNAs). These are embedded either as separate transcriptional units or within the introns of protein coding genes. These are processed by the RNase III endonuclease Drosha together with DGCR8 (microprocessor complex) in to a structure 60-110nt long called precursor-miRNA (pre-miRNA). These are imperfect stem loop (or 'hairpin') structures that are exported to the cytoplasm by an Exportin-5-dependent mechanism, where the RNase III enzyme DICER-1 cleaves it. This produces a short double-stranded miRNA duplex. A helicase then unwinds this duplex to form the mature miRNA, which is then loaded into the multi-protein structure referred to as RNA-induced silencing complex (RISC). Binding of the RISC to the target mRNA (usually to partly complementary sequences with the 3'UTR) silences its expression (Bartel, 2009, Siomi and Siomi, 2010, Bhayani *et al.*, 2012).

In the normal state miRNAs play a role in biofeedback. They have been shown to figure in differentiation pathways such as skin and smooth muscle (Yi *et al.*, 2009, Yi *et al.*, 2008, Cordes *et al.*, 2009). As a single miRNA can target several hundred mRNAs deregulation of miRNA expression could affect many different signalling pathways and drive cells towards cancerous transformation (Jansson and Lund, 2012). These targets may reside within the same functional pathway or not.

Dysregulation of miRNAs has been widely reported in a plethora of diseases. The exact function of many miRNAs is unclear. Model system studies tend to focus on individual targets and often multiple miRNAs are found have altered expression in different disease states. For instance, miRNA-29 (miR-29) is downregulated in lung cancer whilst it is upregulated in breast cancer (Fabbri *et al.*, 2007, Gebeshuber *et al.*, 2009).

1.1.11.2 miRNAs implicated in HNSCC

Though there has been an exponential increase in the number of miRNAs that have been discovered in the last decade, accurate reports of their function are emerging more slowly (Castoldi *et al.*, 2006). The same miRNA has been found to be up or downregulated in the same cancer in different studies (see Table **1**-**6**). There may be methodological and statistical variations causing this discrepancy and only by performing functional experiments alongside highly robust profiling studies can we be certain of the true significance of each specific miRNA.

1.1.11.3 Oncogenic miRNAs

miR-21 was the first miRNA to be termed an "oncomiR" due to the fact it is found to be overexpressed in many different cancers including lung, breast, pancreatic, brain, lymphoma and colon (Medina *et al.*, 2010, Meng *et al.*, 2008, Papagiannakopoulos *et al.*, 2008, Guo *et al.*, 2008, Frankel *et al.*, 2008, Dillhoff *et al.*, 2008). It has also been identified in many studies related to HNSCC and seems to be the most consistent miRNA (in that it is most commonly identified and always upregulated) (Yu *et al.*, 2010, Reis *et al.*, 2010, Wong *et al.*, 2008b, Scapoli *et al.*, 2010, Liu *et al.*, 2010, Kikkawa *et*

al., 2010, Hui *et al.*, 2010, Childs *et al.*, 2009, Chang *et al.*, 2008, Avissar *et al.*, 2009a, Cervigne *et al.*, 2009, Rentoft *et al.*, 2011, Gombos *et al.*, 2013). miR-21 upregulation has been significantly associated with poor prognosis in HNSCC patients (Li *et al.*, 2009, Avissar *et al.*, 2009b). A large number of significant, oncogenic and tumour suppressor targets have been identified for miR-21 in head and neck and other cancers including *RAS, PDCD4, PTEN, RECK* and *HNRPK* (Reis *et al.*, 2010, Chen *et al.*, 2003b, Jung *et al.*, 2012, Darido *et al.*, 2011, Zhang *et al.*, 2010, Papagiannakopoulos *et al.*, 2008). Other studies have suggested downregulating miR-21, using short-interfering-RNAs (siRNAs), in OSCC cell lines sensitises them to cisplatin (Wang *et al.*, 2012, Bourguignon *et al.*, 2012, Yu *et al.*, 2010).

miR-155 has been recognised as overexpressed in several cancers including HNSCC (Wong *et al.*, 2008b, Kikkawa *et al.*, 2010, Hui *et al.*, 2010, Chang *et al.*, 2008, Cao *et al.*, 2013, Saito *et al.*, 2011, Wang *et al.*, 2010, Gombos *et al.*, 2013). It is involved in haematopoiesis, inflammation and immunity.

Studies on OSCC cell lines have suggested miR-155 promotes proliferation and invasion by down-regulating the *CDC73* gene as well as suppressing the cytokine signalling 1 (SOCS1)-STAT3 pathway (Rather *et al.*, 2013, Zhao *et al.*, 2013). Other important targets of miR-155 include *TP53INP1*, *HIF* and *MAP3K7IP1* (Faraoni *et al.*, 2009).

miR-106b-25 refers to a cluster of miRNAs that includes miR-106b, -93 and -25, which have been found to be overexpressed in HNSCC (Cao *et al.*, 2013, Hui *et al.*, 2010, Ramdas *et al.*, 2009). Knockdown of this cluster has been demonstrated to reduce cell proliferation (Hui *et al.*, 2010). This effect is mediated via targeting the *p21/CDKN1A* pathway (Ivanovska *et al.*, 2008).

miR-130b has been consistently found to be upregulated in HNSCC (Liu *et al.*, 2010, Kikkawa *et al.*, 2010, Avissar *et al.*, 2009a, Cao *et al.*, 2013). Its role in HNSCC is yet to be elucidated but it is related to EMT in endometrial carcinoma (Li *et al.*, 2013). It has also been identified as regulating the tumour suppressor gene RUNX3 in gastric cancer (Lai *et al.*, 2010). Similarly

Study	Tumour subsite/tissue	Upregulated	Downregulated	Methods
(Kozaki *et al.*, 2008)	OSCC/Cell line	miR-374, miR-340, miR-22, miR-10a, miR-140, miR-181a, miR-146a, miR-126, miR-31, miR-9-5p, miR-9-3p	miR-27a, miR-34b, miR-34c, miR-203, miR-302c, miR-23a, miR-27b, miR-34a, miR-215, miR-299, miR-330, miR-337, miR-107, miR-133b, miR-138, miR-139, miR-223, miR-204, miR-370, let-7d, miR-95, miR-302a, miR-36, let-7g, miR-23b, miR-128a, miR-148a, miR-155, miR-200c, miR-302b, miR-368, miR-122a, miR-371, let-7a, miR-26b, miR-30e-5p, miR-96	qRT-PCR
(Hsu *et al.*, 2012)	OSCC, OPSCC, LSCC, HSCC/Plasma	miR-21		qRT-PCR
(Ayaz *et al.*, 2013)	LSCC/Plasma	miR-331-3, miR-60, miR-130, miR-60-5p, miR-212-3p, miR-99b-5p, miR-21-5p, miR-106a-5p, miR-146b-5p, miR-148a-3p, miR-17-5p, miR-194-5p, miR-214-3p, miR-335-5p, miR-483-5p, miR-18a-5p, miR-212-3p, miR-205-5p	miR-19a-3p, miR-25-3p, miR-126-3p, miR-125b-5p, miR-192-5p, miR-203, miR-150-5p, miR-218-5p, miR-451a, miR-601	qRT-PCR
(Maclellan *et al.*, 2012)	Cis, OSCC/Plasma	miR-16, let-7b, miR-26a, miR-17, miR-19a, miR-486-5p, miR-92a, miR-30e, miR-320b, miR-451, miR-7, miR-25, miR-7a, miR-195, miR-624	miR-29a, miR-223, miR-338-3p, miR-142-5p, let-7d	qRT-PCR
(Park *et al.*, 2009)	OSCC/Saliva		miR-125a, miR-200a	qRT-PCR
(Liu *et al.*, 2012)	OSCC/Saliva	miR-31		qRT-PCR
(Liu *et al.*,	OSCC/Tissue	miR-31, miR-21,	miR-100, miR-328,	qRT-PCR

2010)		miR-96, miR-224, miR-182, miR-135b, miR-183, miR-301, miR-147, miR-373, miR-155, miR-223, miR-372, miR-130b, miR-187, miR-371, miR-34b, miR-34c, miR-216, miR-10a, miR-128b, miR-104	miR-99a, miR-124b, miR-149, miR-139, miR-124a, miR-204, miR-211	
(Wong *et al.*, 2008b)	OSCC/Tissue	miR-184, miR-34c, miR-137, miR-72, miR-124a, miR-21, miR-124b, miR-31, miR-128a, miR-34b, miR-154, miR-197, miR-132, miR-147, miR-325, miR-181c, miR-198, miR-155, miR-30a-3p miR-338, miR-17-5p, miR-104, miR-13, miR-213	miR-133, miR-99, miR-19, miR-133, miR-219, miR-100, miR-125b, miR-26b, miR-138, miR-149, miR-195, miR-107, miR-139	qRT-PCR
(Scapoli *et al.*, 2010)	OSCC/Tissue	miR-489, miR-129, miR-23a, miR-214, miR-23b, miR-92, miR-25, miR-210, miR-212, miR-515, miR-146b, miR-21, miR-338	miR-520h, miR-197, miR-378, miR-135b, miR-224, miR-34a	Microarray
(Kikkawa *et al.*, 2010)	HSCC/Tissue	miR-517c, miR-196a, miR-7, miR-196b, miR-650, miR-18a, miR-452, miR-183, miR-432, miR-301a, miR-21	miR-1, miR-375, miR-139-5p, miR-504, miR-125b, miR-199b, miR-100, miR-497, let-7c, miR-30a-3p, miR-218, miR-10b, miR-126-5p, miR-378, miR-328, miR-204, miR-143, miR-126-3p, miR-99a, miR-95, miR-489, miR-203, miR-140-5p, miR-29a, miR-26a, miR-214, miR-30a, miR-26b, miR-30e-3p, miR-30b, let-7b	qRT-PCR

(Hui *et al.*, 2010).	OPSCC, LSCC, HSCC/Tissue	miR-42, miR-9, miR-106, miR-1, miR-20, miR-155, miR-193a, miR-25, miR-92, let-7i, miR-17-5p, miR-19b, miR-223, miR-27a, miR-142-3p, miR-21, miR-106a, miR-15a, miR-21, miR-29b, miR-130b, miR-205, miR-422b	miR-125b, miR-375, let-7a, miR-10a, miR-140, miR-100, miR-143, miR-99a, miR-30c, miR-365, miR-127, miR-199b, let-7c, let-7e, miR-26a	qRT-PCR
(Childs *et al.*, 2009).	OSCC, OPSCC, LSCC, HSCC/Tissue	miR-21, miR-23b	Let-7d, miR-1, miR-133a, miR-205	Microarray
(Chang *et al.*, 2008).	OSCC, OPSCC, LSCC, HSCC/Tissue	miR-21, let-7, miR-18, miR-29c, miR-142-3p, miR-155, miR-146c	miR-494	Microarray
(Avissar *et al.*, 2009a).	OSCC, OPSCC, LSCC/Tissue	miR-21, miR-181d, miR-181b, miR-491, miR-455, miR-18a, miR-130b, miR-221, miR-193b, miR-181a, miR-18b	miR-375	Microarray
(Cao *et al.*, 2013).	LSCC/Tissue	miR-221, miR-222, miR-31, miR-205, , miR-155, miR-663, miR-1826, miR-455-3p, miR-1228-5p, miR-93, miR-181a, miR-193b, miR-21, miR-22, miR-23a, miR-18a, miR-185, miR-25, miR-1207-5p, miR-16, miR-106b, miR-1307, miR-15b, miR-130b	miR-140-3p, miR-125b, miR-145	Microarray
(Ramdas *et al.*, 2009).	OSCC, OPSCC, LSCC/Tissue	miR-7083, miR-7, miR-34b, miR-155, miR-182, miR-21, miR-181c, miR-181a, miR-15b, miR-185, miR-	miR-23b, miR-125a, miR-125b, miR-7029	Microarray

		25, miR-93, let-7i, miR-107, miR-103, miR-221		
(Rentoft *et al.*, 2011).	OSCC/Tissue	miR-659, miR-146b-3p, miR-1301, miR-665, miR-142-5p, miR-7, miR-142-3p, miR-21, miR-936, miR-206	miR-617, miR-29b-2, miR-132, miR-548b-5p, miR-509-5p, miR-22	Microarray
(Gombos *et al.*, 2013).	OSCC/Tissue	miR-21, miR-155, miR-191, miR-221		qRT-PCR

Table 1-6: miRNA profiled when comparing tumour tissue to normal. Reproduced with kind permission, Sethi *et al*, 2014.

miR-223 has been found to be overexpressed in profiling studies of HNSCC (Hui *et al.*, 2010, Lajer *et al.*, 2011). It is recognised as oncogenic in gastric cancer whilst in breast cancer it has been associated with increased cell death and decreased migration (Pinatel *et al.*, 2014, Chen *et al.*, 2012).

1.1.11.4 Tumour suppressor miRNAs

Though miR-31 has been described as upregulated in HNSCC it appears that this relationship is cancer-type specific (Li *et al.*, 2009, Liu *et al.*, 2010, Cao *et al.*, 2013, Scapoli *et al.*, 2010). It has also been observed to be upregulated in hepatocellular carcinoma and colorectal cancer but downregulated in breast, prostate and ovarian cancer (Bandres *et al.*, 2006, Motoyama *et al.*, 2009, Wong *et al.*, 2008a, Yan *et al.*, 2008, Schaefer *et al.*, 2010, Creighton *et al.*, 2010). In pancreatic cancer under and over-expression of miR-31 lead to inhibited migration and invasion in different cell lines, dependent on their endogenous levels of miR-31 (Laurila *et al.*, 2012). As mentioned earlier miR-31 activates the HIF pathway suggesting an important role in the development of HNSCC, however upregulation of miR-31 in breast cancer models causes regression of metastases (Valastyan *et al.*, 2011). This is a prime example of the complicated nature and function of miRNAs.

The miR-99 family consists of miR-99a, -99b and -100. Dysregulation of this family has been reported in prostate and cervical cancer as well as HNSCC (Tsukamoto *et al.*, 2010, Sun *et al.*, 2011, Chen *et al.*, 2012, Kikkawa *et al.*, 2010, Hui *et al.*, 2010, Wong *et al.*, 2008b). miR-99a has been shown to be downregulated contributing to survival of OSCC cells(Yan *et al.*, 2012). Forced restoration of miR-99a and -100 results in suppressed cell proliferation and migration in HNSCC cell lines (Chen *et al.*, 2012). Functionally miR-99a directly targets mTOR, an important signalling pathway in cell survival and growth (Yan *et al.*, 2012).

Low levels of expression of miR-375 correlate to decreased survival and metastasis in HNSCC (Harris *et al.*, 2012). This effect may be mediated via the metadherin pathway, levels of which inversely correlate to miR-375 as well as miR-375 mediated regulation of *MYC* expression (Hui *et al.*, 2011, Jung *et al.*, 2013). Loss of miR-125b has been suggested to play an important role in the development of HNSCC (Nakanishi *et al.*, 2014, Henson *et al.*, 2009). This downregulation has also been implicated in radioresistance by inducing inhibition of *ICAM2*. Forced expression of miR-125b has been shown to decrease proliferation and enhance radiosensitivity of OSCC cells (Shiiba *et al.*, 2013). High levels of *ERBB2*, the proto-oncogene, have been linked to suppression of miR-125b (Scott *et al.*, 2007).

The Let-7 family (the largest) of miRNAs are reported to be downregulated in HNSCC (Maclellan *et al.*, 2012, Kikkawa *et al.*, 2010, Hui *et al.*, 2010, Childs *et al.*, 2009, Yang *et al.*, 2013). Let -7 has been demonstrated to be a direct regulator of *RAS* expression in humans (Johnson *et al.*, 2005). Loss of miRNA-mediated suppression can promote oncogenesis. Let-7a downregulates genes associated with stemness in tumour-initiating cells (Yu *et al.*, 2011). Let-7d has been shown to negatively modulate EMT in OSCC cell lines (Chang *et al.*, 2011). Low levels of Let-7d expression have also been associated with poorer prognosis inpatients with HNSCC (Childs *et al.*, 2009).

1.1.11.5   Mixed role miRNAs

miR-7 has been reported as up and downregulated in HNSCC (Kikkawa *et al.*, 2010, Maclellan *et al.*, 2012, Liu *et al.*, 2009). It has been suggested to

play a role in the keratinisation of OSCC cell lines (Jung *et al.*, 2012). However, it has been demonstrated to inhibit *EGFR* expression and downstream *AKT* activity in lung and breast cancer reducing cell proliferation and survival (Kefas *et al.*, 2008, Webster *et al.*, 2009). HNSCC cell lines have shown the same pattern of *EGFR* regulation and miR-7 has been reported to inhibit erlotinib-resistant cells, highlighting the therapeutic potential of miRNAs (Kalinowski *et al.*, 2012). Tumour-suppressor genes *IGF1R* and *RECK* have also been shown to be regulated by miR-7 in OSCC (Jiang *et al.*, 2010, Jung *et al.*, 2012).

miR-34b is widely reported to be upregulated in HNSCC (Wong *et al.*, 2008b, Ramdas *et al.*, 2009, Liu *et al.*, 2010, Cervigne *et al.*, 2009). This is perhaps counter-intuitive as miR-34b has been shown to act as a tumour-suppressor in a feedback loop with the proto-oncogene *MET* (Wang *et al.*, 2013). Over-expression of the *MET*-axis has been correlated with progression and metastases in HNSCC (Xu and Fisher, 2013).

The first miRNA mimic is currently in early phase trial (a miRNA-34 mimic) in liver cancer demonstrating the potential for miRNA markers to be translated to therapeutic targets in cancer.

1.1.12   Summary and Impact

We know that cancer is a disease characterised by progressive genetic and epigenetic alterations (Hahn and Weinberg, 2002, Jones and Baylin, 2002). The diversity of outcomes in HNSCC is an indicator of the intrinsic biological heterogeneity of the individual tumours (Chung *et al.*, 2006). Evidence that genomic changes drive tumour development through various methods (e.g. mutations, oncogenes, LOH etc.) has accumulated (Hanahan and Weinberg, 2000). Identifying copy number alteration has been shown to be a highly successful method of stratifying a patient's risk of metastasis and overall survival (Desouki *et al.*, 2011). This can be performed at high resolution, without being restricted to known targets, using NGS technology. miRNAs have been shown to be aberrantly expressed in cancer and show great potential for biomarkers and therapeutic targets in cancer management. By examining matched pairs of primary tumours and lymph node metastases (with and without ECS) we may be able to identify a molecular signature in

the primary tumour for lymph node metastasis and ECS. This means we would be able to distinguish between the patient with high-risk lymph node disease (with ECS) and the patient with low-risk lymph node disease (without ECS). This would enable risk stratification of the patient at initial biopsy without the need for cervical node surgery and allow better management planning and reduce morbidity for the patient.

1.1.13 Aim

It is known that cervical lymph nodes with ECS act as an indicator of tumours with a more aggressive phenotype and a poor prognosis. I hypothesised this phenotype is driven by underlying molecular changes and that these should be detectable in the primary tumour. This would provide a biomarker of tumour aggressiveness without the need for cervical node surgery. I planned to use next generation sequencing at low coverage to identify CNA's as a reflection of the molecular changes in cancer and metastasis. I intended to analyse changes at specific chromosomal loci, compare the total number of breakpoints between different groups and apply algorithms to analyse the karyogram complexity (e.g. GISTIC) (Mermel *et al.*, 2011). I then wanted to apply the skills gained from analysing metastatic and non-metastatic molecular profile to evaluating CNA data for oropharyngeal tumours with and without any detectable HPV-viral load. I also set out to obtain miRNA profile data for metastatic and non-metastatic tumours evaluating any differential expression patterns associated with metastasis as well as integrating this with copy number data.

I aimed to answer the following questions with my study:

1. Is the molecular signature, identified in the lymph nodes containing metastatic SCC, a reflection of that in the matching primary tumour samples of individual patients?
2. In patients with a primary head and neck tumour, do the metastatic cervical lymph nodes have a molecular signature that differs from non-metastatic primary tumours?
3. In patients with a primary head and neck tumour, do the metastatic cervical lymph nodes with ECS have a molecular signature that differs from patients in whom metastatic nodes show no ECS?
4. Do patients with a detectable HPV-viral load have a molecular signature that differs from tumours with no detectable HPV-viral load?

Chapter 2

Materials and Methods

2.1 Introduction

Methods used throughout this project are detailed here. Room temperature (RT) is taken to be 21-24°C. Supplier addresses and e-mail addresses are presented in Appendix 8.1. Figure **2-1** below displays an overview of this project.
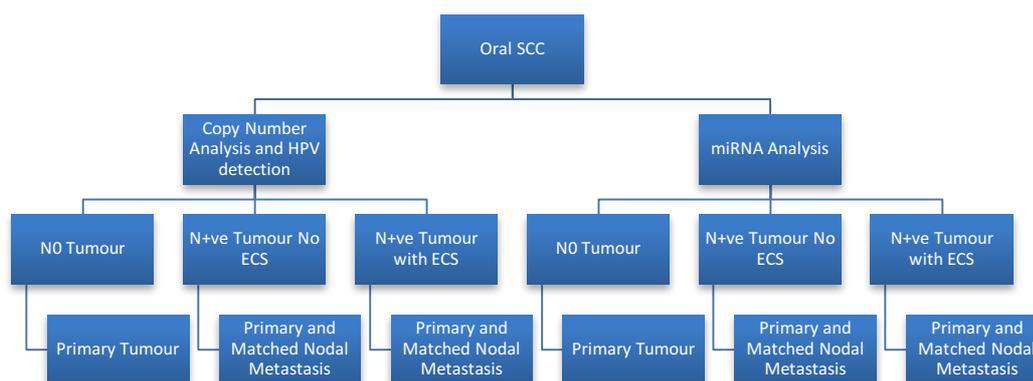
Figure 2-1: General overview of study design

Ethical approval was obtained from Leeds (East) Research Ethics Committee (REC – 07/Q1206/30) for this work.

2.2 Patients

I designed three groups to compare tissue samples as shown in Table **2-1**. Helene Thygesen (CRUK Biostatistician) was contacted to determine the minimum sample size large enough to detect statistically significant difference between groups for this study. Performing simulations (based upon digital karyograms produced by previous CNA studies on HNSCC samples by Prof Rabbits' group), Dr Thygesen determined that

approximately 20 patients per group should be sufficient to identify CNAs that have excess frequencies of 30% in one group relative to the other. Previous studies by Prof Rabbitts' group had found a background rate of CNAs to be at least 5% (gain and loss) and an average of 100 segments per sample were obtained. Using this as the background assumed background CNA rate, Fisher's Exact test was used in the simulations to compare gains to non-gains and losses to non-loss. These simulated p-values were then converted to false discovery rates (or q-values) for multiple testing. Based on these figures CNAs with an excess frequency of 30% were shown to be identifiable with a significant FDR set at < 0.10. This did not mean that CNAs with a lower excess frequency could not be evaluated but should be regarded with a higher degree of caution.

Primary Site	Lymph node status	Tissue to be used
Anterior 2/3 tongue	pN0	Primary and 1 negative node
Anterior 2/3 tongue	pN1-3 no ECS	Primary and 1 positive node
Anterior 2/3 tongue	pN1-3 with ECS	Primary and node with ECS

Table 2-1 - Patient groups

Potential patients were identified from the hospital tissue archive using the CoPath database used by the pathology department in Leeds. This produced a list of over 3000 specimens between 2005 and 2012. The Patient Pathway Manager (PPM) database was then used to identify appropriate patients (many were not HNSCC) from this list that had been diagnosed with squamous cell carcinoma of the anterior two-thirds of the tongue who had undergone surgical resection of their primary tumour with an accompanying cervical lymph node dissection as primary treatment. This search identified potential patients as described in Table 2-1.

These patients were all assigned a study code (e.g. ECS001, ECS002 etc.) and thenceforth treated anonymously. Exclusion criteria included a past medical history of autoimmune disease or drug history of immunosuppressants. These two conditions are known to be associated with

an increased risk of developing cancer and were excluded to attempt to reduce heterogeneity.

Formalin-fixed paraffin embedded (FFPE) tissue blocks for primary tumour and lymph node were obtained (via Andy Clarke, Chief Biomedical Scientist in Bexley Wing Pathology Department) using the block numbers identified from the formal histopathology reports and a file of all histopathology reports was compiled.

Upon receiving the tissue blocks these were then recorded and stored according to block number and study code in accordance with the Human Tissue Act 2004.

2.3 Identification of appropriate tumour and lymph samples

2.3.1 Haematoxylin and eosin staining

The relevant FFPE blocks were stored at 4^oC overnight prior to sectioning. They were then placed on ice ten minutes prior to cutting. The water bath was set at 45^oC and hot plate set at 60^oC. The manual rotatory microtome and forceps were cleaned with Histo-Clear and a new blade inserted.

Sections from each block were cut at 5μm in the vertical position. Slides for these were labelled with the study ID, block number, section number, stain type and thickness. Slides were then placed on the hot plate for 2 hours and either stored overnight at 4^oC or immediately stained for haematoxylin and eosin (H & E).

Staining took place on level 4, Wellcome Trust Brenner Building. This involved deparaffinisation in xylene (Sigma-Aldrich, USA), serial rehydration in graded ethanol (Sigma-Aldrich, USA), staining and dehydration shown in Table **2-2**. Mayer's Haematoxylin (Sigma-Aldrich, USA) and Eosin (BDH, UK) were used for staining.

At completion of the staining protocol a coverslip was mounted onto each slides immediately using DPX mounting medium (Solmedia, UK).

Xylene	3 min x 4
100% ethanol	3 min
90% ethanol	3 min
70% ethanol	3 min
Tap water	2 min
Mayer's Haematoxylin	2.5 min
Tap water	1 min
Scott's tap water	2 min
Tap water	1 min
0.5% Eosin Y	2 min
Tap water	1 min
100% ethanol	15 seconds
100% ethanol	1 min
100% ethanol	5 min x 2
Xylene	3 min x 3

Table 2-2: Protocol followed for H & E staining of FFPE sections.

2.3.2 Slide marking

For each slide, the target areas for dissection (e.g. primary tumour, metastasis) were marked out by a consultant head and neck histopathologist (Dr Preetha Chengot or Professor Kenneth MacLennan). The areas of highest tumour cell content were specifically marked with a visually determined minimum tumour cell content of approximately 70%. The slides with the highest tumour cell content were then selected and tumour-lymph node patient pairs were identified for micro-dissection and nucleic acid extraction.

2.4 Tissue micro-dissection

Seven 10 µm sections were cut from each selected FFPE block for tissue dissection. A further, single 5 µm section was cut immediately after obtaining the seven tissue slides for H & E staining. These were left to air-dry overnight. They were then prepared for dissection by placing on a hot plate

at 60^{0}C for 5 minutes and then dewaxing and rehydrating the slides by serial immersion in Coplin jars according to the following protocol:

Xylene	5 min
100% Ethanol	3 min
90% Ethanol	3 min
70% Ethanol	3 min (and keep each slide immersed until dissection)

Table 2-3: Dewaxing and rehydration protocol for tissue dissection for RNA.

The initial marked H&E slide was used as a guide for dissection. By placing this underneath the rehydrated slides the targeted areas were dissected off precisely using either a size 11 disposable scalpel or a pair of 21 gauge needles. The dissected tissue was immediately placed into a fresh 1.5 ml micro-centrifuge tube. The slide taken following harvesting of the seven sections for dissection was stained with H&E and preserved for marking to ensure accuracy of dissection and if the tissue block was used again. DNA and RNA extraction protocols were then followed as below.

2.5 DNA extraction

According to the area being sampled for extraction one of two Qiagen DNA extraction kits (Qiagen, UK) were used:

2.5.1 Tissue area sampled per slide <5 mm^2

The Qiagen QIAamp DNA micro kit reagents and columns were used for these samples. After transferring the dissected tissue to a 1.5 ml micro-centrifuge tube, 30 µl Buffer ATL and 10 µl Proteinase K solution was immediately added to the tissue. This was then mixed by pulse-vortexing for 15 seconds before spinning down the contents and then incubating the tube in a heat block at 56°C. They were incubated until the sample was completely lysed with daily agitation (up to a maximum of 72 hours with new addition of 10µl Proteinase K solution every 24 hours if visible tissue fragments remained in the tube). After lysis was achieved the sample was then incubated at 90°C for 1 hour on a heat block. Next, 10 µl Buffer ATL and 50 µl Buffer AL were added to the sample, which was mixed by pulse-

vortexing for 15 seconds. Then 50 µl 100% ethanol was then added and mixed thoroughly with pulse vortexing for 15 seconds. This was then incubated at 15-25°C (room temperature) for 5 minutes. The lysate was then transferred and centrifuged through a QIAamp MinElute column at 6,000 x g (8,000 rpm) for 1 minute. The column was then placed in a fresh 2 ml collection tube and 500 µl Buffer AW1 added. This was centrifuged at 6,000 x g (8,000 rpm) for 1 minute and the flow through discarded. 500 µl of AW2 buffer was then added to the column and centrifuged at 20,000 x g (14,000 rpm) for 3 minutes. The flow-through was discarded and the column placed in a fresh 1.5 ml microcentrifuge tube (labelled with the sample's study ID, block ID, DNA, tissue type and date). The DNA was then eluted by adding 30 µl Elution Buffer (EB) to the column. This was incubated for 5 minutes at room temperature before being centrifuged at 20,000 x g (14,000 rpm). The eluate was then reloaded in the same column, allowed to incubate for 5 minutes at room temperature and centrifuged at 20,000 x g (14,000 rpm). This was then stored at -20^{0}C.

2.5.2 Tissue area sampled per slide 5-10 mm^2

The Qiagen QIAamp DNA micro kit reagents and columns (Qiagen, UK) were used for these samples in a similar protocol to that described above. In these dissected tissue samples, 180 µl Buffer ATL and 20 µl Proteinase K solution was immediately added and mixed by pulse-vortexing for 15 seconds. This was then incubated at 56°C until the sample was lysed (with addition of 20 µl Proteinase K solution every 24 hours if tissue fragments remained visible). After lysing was completed the sample was then incubated at 90°C for 1 hour. A further 200 µl Buffer AL and 200 µl 100% ethanol was then added and mixed thoroughly by pulse-vortexing. This was then incubated for 5 minutes at room temperature. The entire lysate was then transferred to a Qiamp MinElute Column and centrifuged in a 2ml collection tube at 6,000 x g (8,000 rpm) for 1 minute. The flow-through was discarded and the bound DNA washed with 500 µl Buffer AW1 and 500 µl Buffer AW2 as in section 2.5.1. The DNA was then eluted in 30 µl of EB as in section 2.5.1 and stored at -20^{0}C.

2.5.3 Tissue are sampled per slide >10 mm^2

The Qiagen QIAamp DNA mini kit reagents and columns (Qiagen, UK) were used for these dissected tissue samples. After dissection, 180 µl Buffer ATL and 20 µl proteinase K were added and mixed by pulse-vortexing for 15 seconds. These were incubated until tissue lysis achieved as in section 2.5.2. After completion of lysis the sample was incubated at 90^{0}C for 1 hour. Then 200 µl Buffer AL was added and mixed by pulse-vortexing for 15 seconds. This was then incubated for 10 minutes at 70^{0}C. Following this, 200 µl 100% ethanol was added and mixed by pulse-vortexing for 15 seconds. The sample was then incubated at room temperature for 5 minutes. The entire lysate was then transferred to the QIAamp Mini spin column and centrifuged at 6,000 x g (8,000 rpm). The flow-through was discarded and 500 µl Buffer AW1 added to the column and centrifuged at 8,000 x g (6,000 rpm) for 1 minute. The flow through was discarded and 500 µl Buffer AW2 added to the column and centrifuged at 20,000 x g (14,000 rpm) for 3 minutes. The column was then transferred to a fresh 1.5 ml microcentrifuge tube (labelled with the sample's study ID, block ID, DNA, tissue type and date) and 70 µl EB added to the column. This was incubated for 5 minutes at room temperature and centrifuged at 20,000 x g (14,000 rpm) for 1 minute. The eluate was reloaded, incubated for 1 minute at room temperature and centrifuged at 20,000 x g for 1 minute. The DNA was then stored at -20^{0}C.

2.6 Total RNA including microRNA extraction

The Qiagen miRNA FFPE kit reagents and columns (Qiagen, UK) were used for these dissected tissue samples. Prior to beginning extraction all equipment and surfaces were cleaned with RNase AWAY (Sigma-Aldrich, USA).

After transferring the dissected tissue to a fresh 1.5 ml micro-centrifuge tube, the sample was placed in a heat block at 55^{0}C for 3-5 minutes to dry the tissue. Immediately following this, 150 µl Buffer PKD were added and mixed by pulse-vortexing for 15 seconds. 10 µl proteinase K was then added and mixed by pipetting. The sample was then incubated at 56^{0}C for up to 2 hours (vortexing every 20 minutes and adding a further 10 µl proteinase K after 1

hour if visible tissue remains). The sample was then incubated at 80^0C for 15 minutes, before being immediately placed on ice for 3 minutes. The sample was then centrifuged at 20,000 x g (14,000 rpm) for 20 minutes. The resulting supernatant was transferred to a fresh 2 ml micro-centrifuge tube, taking care not to disturb the pellet. Following this, 1/10th volume of DNA Booster Buffer (approximately 16 µl) and 10 µl DNase I stock solution was added. This was mixed by inverting the tube and then incubated at room temperature for 15 minutes. Then, 320 µl Buffer RBC was added and mixed by pulse-vortexing for 15 seconds followed by 1120 µl 100% ethanol and mixed by pipetting. The entire lysate was then transferred to an RNeasy MinElute spin column in a 2 ml collection tube and centrifuged at 8,000 x g for 15 seconds. The flow-through was discarded and the column placed in a fresh 2 ml collection tube. Then, 500 µl Buffer RPE was added to the column and centrifuged at 8,000 x g for 15 seconds. The flow through was discarded and wash repeated. The column was then placed in a fresh 2 ml collection tube and centrifuged with the lid open for 5 minutes at full speed. The column was then placed in a fresh 1.5 ml micro-centrifuge tube (labelled with the sample's study ID, block ID, RNA, tissue type and date) and 23 µl RNase-free water added to the column. This was incubated for 1 minute on ice and centrifuged for 1 minute at full speed. Following elution the RNA was stored at -80^0C, if not used immediately.

2.7 Nucleic acid quantification

After extracting the nucleic acid this was then quantified in two ways:

2.7.1 Fluorometry

This used the Qubit dsDNA BR assay kit and Qubit fluorometer (Life Technologies, UK). Initially 1 µl of 200x Qubit dsDNA BR Reagent is diluted to 1x concentration using the BR Buffer provided. Both calibration of the fluorometer and quantification was carried out using the concentrations of standard DNA and sample DNA shown in Table 2-4. For RNA quantification the Qubit RNA BR assay kit (Life technologies, UK). The Qubit RNA BR Reagent 200x concentration provided was diluted to 1:200 in Qubit RNA BR Buffer as with the dsDNA kit. The same Qubit fluorometer was used with the

same concentrations of standard and sample nucleic acid as shown in Table **2**-**4**.

	Standard DNA 1 and 2 (provided with kit)	Sample DNA
Volume of DNA	10 μl	1 μl
Volume of 1x BR Reagent (in BR Buffer)	190 μl	199 μl
Total Volume	200 μl	200 μl

Table 2-4: DNA quantification using Qubit dsDNA BR assay kit.

By following the on-screen instructions the Qubit fluorometer was calibrated each use with the nucleic acid standards provided and 1 μl of the nucleic acid sample was used to determine the concentration within the sample. This concentration was then used to calculate the input for downstream experiments such as sequencing library preparation.

2.7.2 Spectrophotometry

The Nanodrop-1000 was used to determine the concentration of DNA within the sample as well as to give an indication of any protein or chemical contamination using the ultra-violet (UV) radiation absorption. For DNA samples the concentrations were measured against the EB Buffer (Qiagen, UK) the DNA was eluted in whilst for RNA sample the concentration was measured against the RNase-free water provided in the Qiagen miRNA FFPE kit (Qiagen, UK). For each nucleic acid sample 2 μl was used. As well as concentration, the absorbance and the ratio of absorbance at 230 nm, 260 nm and 280 nm (A_{260}:A_{280} and A_{260}:A_{230}) were also recorded.

2.8 DNA Copy number sequencing library preparation

The NEBNext DNA Library Prep Master Mix Set for Illumina and NEBNext Singleplex Oligos for Illumina (New England Biolabs, UK) were used to prepare DNA libraries for sequencing. The protocol described was modified from the commercial protocol supplied with the kits (see Figure **2**-**2** for an overview).
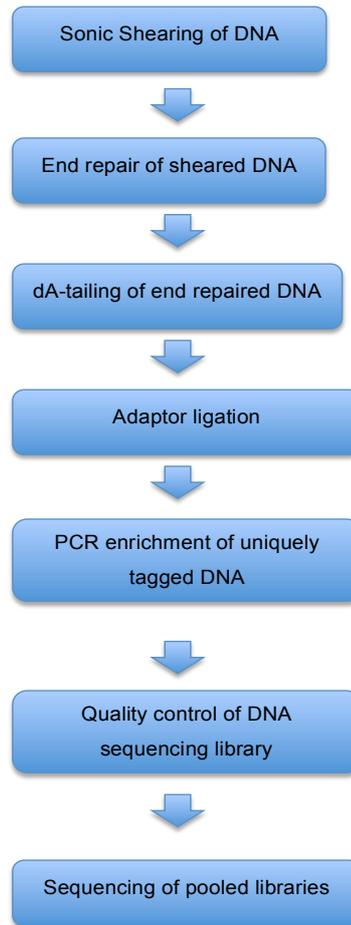
Figure 2-2 - Flow chart of steps within DNA sequencing library preparation

2.8.1 Sample preparation

The input amount of DNA for each sample was approximately 50 – 200 ng. This was obtained by diluting an aliquot from each sample with a concentration >200 ng/µl. The diluent was 1x TE buffer, made up from 100x TE Buffer (Sigma-Aldrich, USA) and nuclease-free water (Life technologies, UK). This input amount was required in a volume of 250 µl and was calculated as per Table **2-5**. Each sample input volume was prepared in a Covaris microTUBE (Covaris Ltd, UK) ready for shearing.

Sample ID	Prep ID	Conc. Of starting material (ng/µl)	Vol. of DNA used (µl)	Vol. of Buffer used (µl)	Total amount of DNA used (ng)	Tag ID	Conc. Of library (ng/µl)
3GL	1	48.8	4	246	195		

Table 2-5: Example of calculation table for input DNA amount and volume for DNA copy number library preparation.

2.8.2 Shearing

The Covaris S2 Focused-ultrasonicator was used to shear the DNA at 19°C in batches of 25 cycles for each sample using the settings shown in Table 2-6.

	Duty Cycle	Intensity	Cycles/burst
1000bp	19.9%	9.9	1000
500cpb	15%	8	500

Table 2-6 - Covaris S2 batch settings

2.8.3 Clean up protocol

This sample was then cleaned using Qiagen MinElute columns with Qiagen PB (binding) buffer and Qiagen PE (wash) buffer (Qiagen, UK). Adding the sonicated DNA sample to 5x volume of PB binding buffer, each sample was added to a Min Elute column and centrifuged at 13,000 rpm for 1 minute, followed by adding 750 µl of PE Buffer to the column and centrifuging this at 13,000 rpm for 1 minute. Following this the flow-through was discarded and the column centrifuged with lid open at 13,000 rpm for 1 minute. The DNA was then eluted by adding 10 µl of EB buffer and centrifuging the column at 13,000 rpm.

2.8.4 Agilent 2200 Tapestation confirmation

Successful shearing was then confirmed using the Agilent 2200 Tapestation D1K Standard Screentape and reagents on the Agilent 2200 Tapestation (Agilent technologies Inc., USA). This was performed using optical tube strips (Agilent technologies Inc., USA). Firstly, 3 µl of D1K ladder was aliquoted into the first tube and mixing 3 µl of D1K Sample Buffer and with 1 µl of each DNA sample in the remaining tubes. These were mixed by pipetting and loaded onto the 2200 Tapestation, which was then run using preloaded software. The resulting digital gel image and electropherogram for each sample are used to confirm successful shearing before continuing with the protocol.

2.8.5 End-repair of fragmented DNA

The fragmented DNA then underwent end-repair to form blunt-ended, 5'-phosphorylated DNA. A master mix was made as shown in Table **2**-**7**.

	Volume per DNA sample/ µl
NEBNext End repair reaction buffer (x10)	5
NEBNext End repair enzyme mix	2.5
Nuclease-free water	33.5

Table 2-7 - End-repair master mix

41 µl of the master mix was added to the 9 µl DNA sample from the section 2.8.4 and incubated at room temperature for 30 minutes. The sample was then cleaned using the Qiagen QiaQuick column with PB buffer and PE buffer (Qiagen, UK), using the same volumes of buffer as proscribed in section 2.8.3. The column was then placed in a fresh 1.5 ml micro-centrifuge tube and the DNA eluted by adding 21 µl EB buffer (Qiagen, UK) to the column and centrifuging this at 13,000 rpm for 1 minute.

2.8.6 dA-Tailing of end-repaired DNA

This step incorporated a tail of deoxyadenosine monophosphate to the 3' ends of blunted dsDNA from the section 2.8.5. A master mix was made as shown in Table **2**-**8**.

	Volume per DNA sample/ µl
NEBNext dA-Tailing Reaction Buffer	2.5
Klenow (3'>5' exo)	1.5

Table 2-8 - dA-Tailing master mix

4 µl of the master mix was added to the 21 µl of DNA from section 2.8.5 and incubated in a heat block at 37°C for 30 minutes. This was then cleaned using a fresh Qiagen MinElute column (Qiagen, UK) according to the protocol in section 2.8.3. The column was then placed in a new 1.5 ml

microcentrifuge tube. The DNA was then eluted by adding 12.5 µl to the column and centrifuging this at 13,000 rpm for 1 minute.

2.8.7 Adaptor ligation to dA-tailed DNA

Adaptors were then ligated to the DNA sample. A master mix was made as shown in Table 2-9.

12.5 µl of master mix was added to the 12.5 µl dA-tailed DNA from section 2.8.6 and incubated at room temperature for 15 minutes. 3 µl of USER enzyme mix (New England Biolabs, UK) was added to the sample and incubated at 37°C for a further 15 minutes. This was then made up to 50 µl by adding 23 µl of EB Buffer (Qiagen, UK).

	Volume per DNA sample/ µl
Quick Ligation Reaction Buffer (x5)	5
NEBNext Adaptor	2.5
Quick T4 Ligase	2.5
Nuclease-free water	2.5

Table 2-9 - Adaptor ligation master mix

This was then cleaned and smaller fragments of DNA were removed using Solid-Phase Reversible Immobilisation (SPRI) beads as described below.

2.8.8 Size-selection

Solid-Phase Reversible Immobilisation (SPRI) beads (Beckman Coulter, UK) were used to perform size selection (removal of smaller, non-target fragments of DNA) on the DNA sample from section 2.8.7. 40 µl (0.8x concentration) of suspended SPRI beads were added to the DNA solution to bind larger fragments of DNA. This was mixed by pipetting and incubated for 5 minutes at room temperature. Then the micro-centrifuge tube was placed in a magnetic tube rack (Life technologies, UK) for 5 minutes to separate the beads. The supernatant was transferred to a new 1.5 ml micro-centrifuge tube and 18 µl (0.2x concentration) SPRI beads were added to bind the DNA targets (approximately 200bp in length). This was mixed by pipetting and then placed in the magnetic rack for 5 minutes. The supernatant was

removed and discarded. The beads containing the bound DNA were washed whilst still in the magnetic rack by adding 200 µl 80% ethanol and incubating for 30 seconds. The ethanol was then removed and the wash repeated with 200 µl fresh 80% ethanol. The ethanol was removed and discarded and the beads allowed to dry for 10 minutes at room temperature. The DNA was then eluted by adding 22 µl of EB buffer (Qiagen, UK) and mixing the beads by pipetting. The magnetic rack was used to separate the beads and the supernatant removed and transferred to a clean 200 µl PCR tube. At this point the sample was split, by storing 10 µl at -20^0C in a 200 µl PCR tube (labelled with study code, DNA, tissue type and date). The remaining 10 µl was used in the following PCR enrichment.

2.8.9 PCR target enrichment

The target adaptor-ligated DNA was then enriched using a 15-cycle PCR protocol. A master mix was made up as per Table **2**-**10**.

	Volume per DNA sample/ µl
NEB High Fidelity 2x PCR master mix	12.5
Universal PCR Primer (25uM)	1.25

Table 2-10 - PCR enrichment master mix

13.75 µl of the master mix was added to remaining 10 µl of DNA from section 2.8.8. 1.25 µl of a separately purchased indexed primer (from a panel of 96 custom-designed primers each containing a unique identifying 6bp tag designed by Dr Henry Wood, Precancer genomics group) was then added (Integrated DNA Technologies, USA). The sample was then transferred to the thermal cycler and cycles of PCR were carried out as described in Table **2**-**11**.

Following 15 cycles the thermal cycler is set to hold the samples at 4^0C, until the samples are cleaned using SPRI beads.

Number of cycles	Time	Temperature
1	30 seconds	98^0C
15	10 seconds	98^0C
	30 seconds	65^0C
	30 seconds	72^0C
	5 minutes	72^0C

Table 2-11: PCR enrichment cycle protocol

2.8.10   Post-PCR SPRI bead clean up

The post-PCR sample from section 2.8.9 was then cleaned by adding 2.5x concentration of SPRI beads (Beckman Coulter, USA) in a fresh 1.5 ml microcentrifuge tube to bind the enriched DNA targets. This was mixed by pipetting and incubated at room temperature for 5 minutes. The tube was then placed in the magnetic rack to separate the beads and the supernatant was then collected and discarded. The beads were then washed with 200 µl 80% ethanol and incubated for 30 seconds. The beads were then separated using the magnetic rack and the ethanol removed and discarded. The wash with ethanol was then repeated. The beads were then left to air-dry from 10 minutes. The DNA was then eluted by adding 40 µl EB buffer (Qiagen, UK) and mixing this by pipetting. After incubating for 2 minutes and separating the beads on the magnetic rack, the supernatant containing the eluted DNA was transferred to a fresh 1.5 ml micro-centrifuge tube (labelled with the sample's study ID, block ID, DNA, tissue type and date).

2.8.11   Library quality control

The DNA libraries were quantified using the Qubit fluorometer (Life Technologies, UK) as detailed in section 2.7.1. They were quality assured for size using the Agilent Tapestation (Agilent technologies Inc., USA). 1 µl of the eluted DNA library sample was examined on the Agilent Tapestation using the D1K Screentape (Agilent technologies Inc., USA). The protocol is identical to that in section 2.8.4. In addition to concentration, the Tapestation also allows confirmation that all excess adaptor oligonucleotide has been removed from the sample (with the appearance of a peak at approximately 115 bp on the electropherogram). If there was contaminating adaptor of an

amount greater than 10% of the target library concentration then the library sample was re-cleaned using as per section 2.8.12. If there was no concern over adaptor contamination then the library sample was stored at -20^{0}C awaiting sample pooling and sequencing. Details of the library stored on our server.

2.8.12   Post PCR sample re-clean with SPRI beads

Firstly 2.5x concentration of SPRI beads (Beckman Coulter, UK) was prepared according to the volume of the sample from section 2.8.11 (38 µl). The beads are prepared as follows: 95 µl of beads are mixed well using vortexing with 38 µl nuclease-free H$_2$O (Life technologies, UK). This was then added to the DNA library sample and mixed well. The sample was incubated for 5 minutes with continuous shaking. The tube was then placed in the magnetic rack for 5 minutes to separate. The supernatant was then removed and discarded. The beads were then washed in 200 µl 70% ethanol and incubated at 30 seconds. The magnetic rack was used to separate the beads and remove the ethanol. The ethanol wash was then repeated. The beads were then air-dried for 10 minutes before eluting the DNA in 42 µl EB buffer (Qiagen, UK). The sample was then quantified and quality assured using the Qubit fluorometer (Life Technologies, UK) and Agilent Tapestation (Agilent technologies Inc., USA) as detailed in sections 2.7.1 and 2.8.11. If there was insufficient DNA in the library sample to proceed this then the pre-PCR sample was removed from -20°C storage (see section 2.8.9) and the post-PCR enrichment clean up process was repeated using this sample.

2.8.13   DNA copy number library sample pooling

To obtain the depth of sequencing coverage required, 40 samples per flow-cell lane. Once 40 samples with unique 6 bp tags had been accrued in storage they were pooled for sequencing.

The DNA concentrations (as measured by the Tapestation) of all the samples to be pooled were compared and a mutually convenient amount to pipette was determined – typically 20 ng, or a volume between 1 and 30 µl.

The samples were then pooled in equal concentrations before being submitted to the sequencing team for running on the sequencer.

The pooled library samples were run on the Illumina HiSeq 2000. This offers 200 million reads per lane. The Illumina HiSeq 2000 produces a FASTQ file. This contains information about each read (including location, size and quality of read) in addition to the raw sequence.

2.9 Copy number analysis data processing

2.9.1 Alignment

The Illumina HiSeq 2000 produces sequencing data in the form of FASTQ files. These are text files that contain the raw nucleotide sequences (sequencing reads) as well as a Phred quality score for each read. Each sample that had been multiplexed in a single sequencing lane was individually tagged with an indexed primer. This tag was identified at the ends of the sequencing reads produced. Using this individual tag the samples were separated into unique files by the sequencing facility prior to delivering the files to our research group.

The adaptor sequences were then removed (or 'trimmed') from the ends of all sequencing reads in all files using software called Cutadapt (http://code.google.com/p/cutadapt/). After removing the adaptor sequences, any remaining nucleotide sequence that was 20 bp or less in length was removed from further analysis (due to inherently poor alignment).

The remaining reads were then aligned to the human reference genome (University of California Santa Cruz version GRCh37/hg19, http;//genome.ucsc.edu) using Burrows Wheeler Alignment (BWA) and to all known HPV subtypes using data downloaded from the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid%20=%201039) (Li *et al*, 2010). The resulting data is stored in the binary alignment map (BAM) format.

2.9.2 Copy number analysis using CNAnorm

CNAnorm is a freely available software package designed specifically for the analysis of copy number data from tumour samples sequenced at low-coverage (approximately 0.01 – 0.5x) (Berri, 2014, Gusnanto *et al.*, 2012). As 40 samples per lane were run on the Illumina HiSeq 2500 in this project, this resulted in approximately 0.033x – 0.33x coverage. The input for this software is the sample BAM file. It divides the reference genome into genomic "windows". These windows can be of a specified length (in base pairs) or can be assigned according to the number of sequencing reads aligned to each window. In this case we instructed CNAnorm to divide the genome into identical length windows (800 Kbp long) for the purposes of downstream analysis.

The normal control sequencing data for this copy number analysis was obtained as a pooled normal sample from 20 individuals available from the 1000 Genomes Project ((ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp). CNAnorm also performs guanine-cytosine (GC) correction of the sequencing reads prior to normalising the sample to normal read ratio (on a window by window basis).

CNAnorm then uses a separate software package called DNAcopy to perform circular binary segmentation of the normalised data (Venkatraman and Olshen, 2007). This designates windows to a segment of genome at the same estimated sample to normal ratio. This essentially estimates changes in copy number. The segmented data is stored as text in a .bed file.

As reads are aligned by genome position the structure of the genome can be analysed. The number of aligned reads per window relative to the number of aligned reads in the same window of the normal control allow inference of the relative copy number of that window. Where segments of DNA change in copy number represent the genomic breakpoints.

2.9.2.1 Determination of HPV status using low-coverage NGS

Conway et al, described the method used to determine viral load (Conway *et al.*, 2012). Briefly, this involves counting the number of reads that align to the human reference genome and those that align to genomes of all HPV-

subtypes. Depth of sequencing is equal to the number of aligned reads per kilobase of the human reference genome. The number of HPV genome reads detected per sample allows the viral load per genome to be calculated using Conway et al's equation:

$$\frac{(\text{no. of viral reads}) \times 6 \ 10^9 \text{ bp diploid human genome})}{(7900 \text{ bp viral genome} \times \text{number of human reads})}$$

2.9.2.2 Visual assessment of digital karyograms

2.9.2.2.1 Individual karyogram

CNAnorm produces a digital karyogram (see Figure **2-1**) as a graph with the chromosomal position represented on the x-axis and the sample to normal ratio on the y-axis. The normalised ratio (relative copy number) of each segment is represented by a black line. Each window is represented by a dot. Those windows with an increased sample to normal ratio (copy number gain) are coloured red. Those with a decreased sample to normal ratio (copy number loss) are coloured blue.
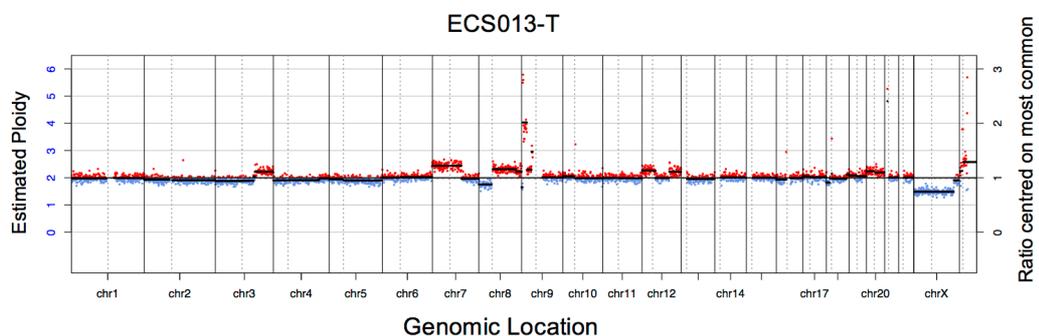
Figure 2-3: Example of digital karyogram produced by CNAnorm.

When visually inspecting these it is possible to estimate both ploidy and tumour cell content of the sample by analysing the karyogram and determining the ratio of loss that represents loss of one copy. By then visually determining the maximum number of copies that are lost at any any genomic region the ploidy can be inferred (e.g. if four copies are lost at a particular point in one sample it must at least tetraploid).

The ratio that represents loss of one copy can then be compared to the ratio that would represent loss of one copy in a homogeneous diploid sample of

100% tumour cell purity and the tumour DNA content (of the dominant clonal population) be inferred from this (see Table **2**-**12**).

Estimated Ploidy	Copy Gain/Loss	Tumour/normal ratio	Tumour DNA Content
2n	1	1.5	100%
2n	1	1.38	75%
2n	1	1.25	50%
2n	1	1.23	25%
2n	-1	0.5	100%
2n	-1	0.63	75%
2n	-1	0.75	50%
2n	-1	0.88	25%
3n	1	1.3	100%
3n	1	1.27	75%
3n	1	1.2	50%
3n	1	1.11	25%
3n	-1	0.66	100%
3n	-1	0.72	75%
3n	-1	0.8	50%
3n	-1	0.89	25%
4n	1	1.25	100%
4n	1	1.21	75%
4n	1	1.17	50%
4n	1	1.1	25%
4n	-1	0.75	100%
4n	-1	0.78	75%
4n	-1	0.83	50%
4n	-1	0.9	25%

Table 2-12: List of assigned ratios to represent loss/gain of 1 copy and inferred tumour DNA content of dominant clone from this.

2.9.2.2.2 Cumulative frequency karyograms

These were created using a script written by Dr Henry Wood (Bioinformatics, Precancer Genomics Group) called seg_compare. This produces a cumulative karyograms of all specified .bed files. This allows estimation of the proportion of samples that have CNA on a genome-wide view as well as a specific individual chromosome view. This analyses the segmented data for each sample and presents a frequency plot. A threshold for designating a

segment as representing a gain or loss must be assigned. This can be done either on a generic basis for all samples or on an individual sample basis.

2.9.2.3 Genomic Identification of Significant Targets in Cancer (GISTIC 2.0)

GISTIC 2.0 is open access software developed by the Broad Institute, USA to provides systematic analysis of CNA, specifically in cancer (Mermel *et al.*, 2011). It provides rapid CNA profiles allowing both broad and focal analysis simultaneously, providing frequency, significance values and heat maps.

GISTIC can be found at www.genepattern.broadinstitute.org. Before running GISTIC the CNAnorm text data files (containing the text data for each sample's genomic windows with raw and segmented copy number data) were converted to the input form using an R script, cnaNorm2.GISTIC.R (Dr Stefano Berri, Precancer Genomic group). This also produced a markers file for each sample based on the genomic co-ordinates. The converted segmented data files were then grouped according to the clinicopathologic parameters set out previously (e.g. metastatic primary tumour, metastasis etc.) and uploaded along with the markers file (which was identical for all samples).

The parameters used when running GISTIC were consistent for all analyses. The algorithm was set to calculate the significance of deletions at a gene level rather than a marker level. All other parameters were left as the default, recommended setting. Of note, the amplification and deletion thresholds were set at 0.1 for all samples (meaning regions with a log2 ratio above or below this were considered amplifications or deletions) as it was not possible to assign an individual sample threshold and the q-value threshold was set to 0.25.

2.10 miRNA sequencing library preparation

The total ribonucleic acid (RNA) samples from section 2.6 were used as input for this protocol. The Illumina TruSeq Small RNA Sample Prep kit (Illumina, USA) was used to create small RNA sequencing libraries. All

equipment and surfaces were wiped with RNase AWAY (Sigma Aldrich, USA) prior to commencement. RNA samples were kept on ice throughout.

2.10.1 Sample preparation

500 ng to 1 µg of the RNA samples from section 2.6 were prepared in 5 µl of RNase-free water (Qiagen, UK) using Table 2-13. It is worth noting that ribosomal depletion is not recommended by Illumina for this protocol. My experience was that samples with < 1 µg of input RNA did not yield a viable library, if subjected to ribosomal RNA depletion.

Sample ID	Prep ID	Conc. Of starting material (ng/µl)	Vol. of RNA (µl)	Vol. of Ultrapure water used (µl)	Total amount of RNA (ng)	Tag ID	Conc. Of library (ng/µl)
47D	1	224	2	3	448	1	

Table 2-13: Example of calculation of input RNA concentration and volume for small RNA library preparation.

2.10.2 3' Adaptor ligation

Adaptors specific for the 3'-hydroxyl group on miRNAs are ligated to the sample by adding 1 µl of RNA 3' adaptor, mixing by pipetting and incubating this at 70^0C for 2 minutes before placing on ice immediately. A master mix (see Table 2-14) was prepared in a 200 µl PCR tube on ice (adding 10% for multiple samples). This required T4 RNA Ligase 2 deletion mutant (Cambio, UK) and mixing by pipetting. 4 µl of this mix was added to the RNA + 3' adaptor sample and mixed by pipetting. This was incubated at 28^0C for 1 hour. At completion, leaving the tube on the thermal cycler, 1 µl Stop Solution was added and incubated for 15 minutes at 28^0C. The tube was then immediately placed on ice.

	Vol per sample (µl)
5x HM Ligation buffer (HML)	2
RNase Inhibitor	1
T4 RNA Ligase 2 deletion mutant	1
Total	4

Table 2-14: Master mix for 3' adaptor ligation.

2.10.3 5' Adaptor ligation

5' adaptors were then added by aliquoting 1.1 x n µl (where n = number of RNA samples) of RNA 5' adaptor in a separate 0.2 µl PCR tube. This was incubated at 70^0C for 2 minutes and then placed on ice. 1.1 x n µl of 10 mM ATP was then added to the aliquoted RNA 5' adaptor tube and mixed by pipetting. A further 1.1 x n µl of T4 ligase was added to the aliquoted RNA 5' adaptor tube and mixed by pipetting. 3 µl of this mix were then added to the sample from section 2.10.2. This was then incubated at 28^0C for 1 hour and then immediately placed on ice.

2.10.4 Reverse transcription

In a separate 200 µl PCR tube 25 mM dNTP was diluted by adding 0.5 µl Ultra Pure Water to 0.5 µl 25 mM dNTP. This was then mixed by pipetting and placed on ice. Then 1 µl of RNA RT Primer was added to the 5' and 3' adaptor-ligated RNA sample from section 2.10.3. This was incubated at 70^0C for 2 minutes and then placed on ice.

In a separate 200 µl PCR tube on ice the master mix shown in Table **2-15** was prepared (adding 10% for multiple samples). This required Superscript II Reverse Transcriptase and 10 mM DTT (Life Technologies, UK) in addition to the other reagents provided in the library kit. It was mixed by pipetting. 5.5 µl of this master mix was added to the adaptor ligated RNA sample and incubated at 50^0C for 1 hour and then placed immediately on ice.

	Vol per sample (µl)
5x First strand buffer	2
12.5 mM dNTP mix	0.5
100 mM DTT	1
RNase Inhibitor	1
Superscript II Reverse Transcriptase	1
Total	5.5

Table 2-15: Reverse transcription master mix

2.10.5 PCR Amplification

In a separate 200 µl PCR tube a PCR master mix was prepared (see Table 2-16). Separate mixes were prepared for each indexed primer. This master mix was then added to the complementary DNA (cDNA) sample from 2.10.4.

	Volume per DNA sample/ µl
Ultra Pure Water	22.5
5x Phusion HF Buffer	10
RNA PCR Primer (RP1)	2
RNA PCR Primer Index (RPIX)	2
25mM dNTP	0.5
Phusion DNA Polymerase	0.5
Total	37.5

Table 2-16: PCR amplification master mix.

The sample was mixed by pipetting and then placed on ice. The thermal cycler was used with the conditions shown in Table 2-17. Upon completion of this protocol the thermal cycler was set to hold the sample at 4^0C.

Time	Temperature	Number of cycles
30 seconds	98^0C	1
10 seconds	98^0C	11
30 seconds	60^0C	
15 seconds	72^0C	
10 minutes	72^0C	1

Table 2-17: Thermal cycler protocol for PCR amplification.

2.10.6 Size selection

The cDNA construct was purified using polyacrylamide gel size selection. Using stock 5x Novex TBE Buffer (Life Technologies, UK). 1l of 1x concentration TBE Buffer was prepared. Pre-cast Novex TBE 6% polyacrylamide gels (Life Technologies, UK) were used in the XCell SureLock Mini-Cell electrophoresis unit (Life Technologies, UK). The ladders were prepared by adding 1 µl of high resolution ladder to 1 µl of 6x DNA loading dye (Life Technologies, UK). 2 µl of Custom ladder was mixed

with 2 µl 6x DNA loading dye. The entire amplified cDNA construct from section 2.10.5 was mixed with 10 µl 6x DNA loading dye. Once the gels were secured in position and the Mini-Cell filled with 1x TBE Buffer (approximately 500 ml), the wells were filled according to Figure **2-4**. The cDNA construct for each sample was divided between two adjacent wells (approximately 30 µl in each). Two wells were left empty in between different patient samples to reduce the risk of cross-contamination. The gel was then run at 146 V until the dye front exited the gel (approximately 45 minutes).

The cassette containing the polyacrylamide gel was then opened whilst submerged in 400 ml of 1x TBE Buffer (Life Technologies, UK). Once opened, add 20 µl of 10 mg/ml ethidium bromide (Life Technologies, UK) was added to the container (to create a 0.5 µg/ml concentration of ethidium bromide solution) and mixed gently for 5 minutes. The gel was then carefully transferred to a UV transilluminator and the bands representing 145–160 bp identified on the High resolution and Custom Ladders. These area within these bands were excised from the lanes containing the patient samples.
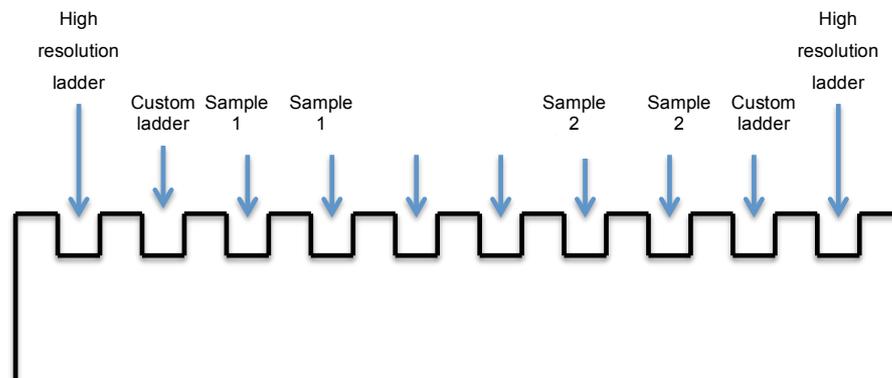
Figure 2-4: Diagrammatic representation of allocation of wells in polyacrylamide gel.

The excised bands from both lanes for each patient sample were transferred to 500 µl gel breaker tubes (Life Technologies, UK). These were placed in a fresh 2 ml collection tube and centrifuged at full speed from 2 minutes. To elute the DNA 200 µl Ultra Pure Water was added to the collection tube and shaken for 2 hours. The eluate was transferred to a new 5 µm filter tube (IST Engineering, USA) and transferred at 600 x g for 10 seconds. The DNA solution was then concentrated to approximately 20 µl by placing it in the

Eppendorf Concentrator 5301 (Eppendorf, Germany) for 20-40 minutes at room temperature (the time needed to concentrate increases with the number of samples in the concentrator on the number of samples in the concentrator).

2.10.7 Small RNA library quality assurance

The concentrated library from section 2.10.6 was then assessed to confirm presence of the correct size nucleic acid chains (adaptor-ligated small RNAs with indexed primers) using the Agilent Bioanalyzer 2100 and Agilent High Sensitivity DNA kit (Agilent, UK).

The kit reagents were allowed to equilibrate to room temperature for 30 minutes. The gel dye mix is prepared by pipetting 15 µl of the blue dye concentrate into the High Sensitivity DNA gel matrix and vortexed for 10 seconds. The gel-dye mix was then transferred to a spin filter in a 1.5 micro-centrifuge tube and centrifuged at 6000 rpm for 10 minutes at room temperature. The filter is then discarded and the tube labelled (Gel-dye mix).

A new High Sensitivity DNA chip (Agilent, UK) was place in the chip priming station and 9 µl of gel-dye mix was pipetted into the well marked Ⓖ. The priming station is then closed and the syringe plunger depressed and held for 1 minute before being released. The priming station was opened and 9 µl of the gel-dye mix pipetted into each well marked Ⓖ. Then 5 µl of High Sensitivity DNA marker was pipetted into the remaining 12 wells. 1 µl of High Sensitivity DNA ladder was pipetted into the well marked with the ladder symbol and 1 µl of library sample from section 2.10.6 into each of the remaining 11 wells. The chip was then vortexed for 1 minute at 2400 rpm. The chip was then placed in the Agilent Bioanalyzer 2100 and the computer set to run the dsDNA High Sensitivity assay.

The output from the Agilent Bioanalyzer 2100 was used to confirm the presence of sequencing library at 145-160 bp and assess the quantity of adaptor dimer or concantamer at 120-138 bp in the library. If excess adaptor contamination is present then the library protocol should be repeated with a new input RNA sample. I found that re-attempting gel size selection of the same sample led to no quantifiable library detectable on gel band or subsequent quantification.

2.10.8   Small RNA library quantification

Each small RNA library sample from section 2.10.6 was then quantified using the Qubit fluorometer and Qubit dsDNA BR assay kit (Life technologies, UK) as detailed in section 2.7.1. This concentration was used when pooling library samples for sequencing.

2.10.9   Small RNA library sample pooling and sequencing

The quantified small RNA library samples were separated into two groups of 12 to ensure that there was no duplication of the indexed primers used to label each sample. Then 8 ng of each library was pooled in a new 200 µl PCR tube (labelled with my name, miRNA library and sequencing lane number). This amount was chosen to ensure that there was sufficient library left in the sample with the lowest concentration that a repeat pooling could be performed in the event of a fault with running the Illumina HiSeq 2500 sequencer.

These were then run on the Illumina HiSeq 2500 Genome Sequencer operating in "Rapid Run Mode", using 50 bp single-end reads (capable of generating 135 million sequencing reads per lane).

2.10.10   miRNA sequencing data processing and analysis

The FASTQ files generated by the sequencer containing the raw sequencing read data were downloaded. Cutadapt was used to remove the adaptor sequences from the ends of each read (http://code.google.com/p/cutadapt/). An open access software called CAP-miRSeq (Comprehensive Analysis Pipeline – microRNA Sequencing) was then used to process the data (Sun *et al.*, 2014). This automatically aligned the trimmed reads to the human reference genome (University of California Santa Cruz version GRCh37/hg19, http://genome.ucsc.edu). After trimming all reads less than 17 bp in length were removed. All RNA types contained in the sequencing library were quantified. CAP-miRSeq generated raw count data for each uniquely mapped miRNA and performed normalisation of these for the number of aligned reads per sample.

CAP-miRSeq then implemented another open access software package called edgeR (Bioconductor, US) to perform further normalisations steps.

These include scaling of the libraries for the number of aligned reads per sample and the dispersion of reads in each sample. It then generated differential expression data according to defined clinicopathologic (Robinson *et al.*, 2010).

2.11 miRNA Nanostring nCounter Expression Assay

The nCounter system (Nanostring Technologies, USA) is a relatively new commercial method for performing expression analysis. It is based on the hybridisation of target molecules to colour-coded, biotinylated probes that are then digitally detected with a raw molecule count being provided, which can allow fold change between samples (Geiss *et al.*, 2008). The nCounter miRNA Expression Assay (Nanostring Technologies, USA) was used to provide an alternative method of miRNA profiling of the samples selected for miRNA sequencing. This was performed to provide validation of any sequencing findings and was specifically chosen as it allowed much a much larger number of miRNAs to be profiled compared to other methods such as qRT-PCR.

Access to the nCounter system was kindly provided by Dr Tudor Fulga (Weatherall Institute of Molecular Medicine, University of Oxford) and training in the protocol was provided by Bruno Steinkraus (PhD Student). Materials and reagents contained in the nCounter Human miRNA Expression Assay Kit and the nCounter Master kit (Nanostring Technologies, USA) were used. All RNA samples were kept on ice throughout this protocol.

2.11.1   Sample preparation

All samples were normalised to 33 ng/µl using RNase-free H_2O (Qiagen, UK) in a volume of 5 µl. All stock RNA samples were immediately returned to -80^0C storage, whilst the diluted RNA samples were kept on ice. Twelve RNA samples were processed at a time.

2.11.2   Annealing protocol

A 1:500 dilution of the miRNA Assay Controls (Nanostring Technologies, USA) was prepared in a new 1.5 ml microcentrifuge using 499 µl RNase-free H_2O (Qiagen, UK).

An annealing master mix was prepared by combining 13 µl of Annealing Buffer, 26 µl of nCounter miRNA Tag Reagent and 6.5 µl of the diluted miRNA Assay Controls. This was mixed well by pipetting. Into 12 new 200 µl PCR tubes, 3.5 µl annealing master mix was pipetted. 3 µl of each RNA sample was added to each tube and mixed by flicking. The tubes were then placed in a thermocycler (see Table **2-18**). After completion of this protocol the samples were held at 48^0C until the ligation protocol was initiated.

Temperature	Time	Cycle
94^0C	1 minute	1
65^0C	2 minutes	1
45^0C	10 minutes	1

Table 2-18: Annealing thermocycler protocol

2.11.3   Ligation protocol

A ligation master mix was prepared by combining 24 µl PEG and 16 µl Ligase Buffer in a fresh 200 µl PCR tube. 2.5 µl of this master mix was then added to the RNA samples held at 48^0C and mixed by flicking. These were then returned to the thermocycler and incubated at 48^0C for 5 minutes.

Temperature	Time	Cycle
48^0C	3 minutes	1
47^0C	3 minutes	1
46^0C	3 minutes	1
45^0C	5 minutes	1
65^0C	10 minutes	1

Table 2-19: Ligation thermocycler protocol

Following this 1 µl Ligase was added directly to each RNA sample without removing them from the thermocycler. The ligation protocol was then initiated on the thermocycler (see Table **2-19**). At completion of this protocol the samples were held at 4^0C on the thermocycler for the purification protocol.

2.11.4 Purification protocol

After completion of the ligation protocol 1 µl Ligation Clean-Up Enzyme was added to each RNA sample from section 2.11.3 and mixed by flicking. The tubes were then returned to the thermocycler and the purification protocol initiated (see Table **2-20**). After completion of this, 40 µl RNAse-free H_2O was added to each sample.

Temperature	Time	Cycle
37^0C	2 hours	1
70^0C	10 minutes	1

Table 2-20: Purification thermocycler protocol

2.11.5 Hybridisation protocol

A master mix was created using 130 µl of the Reporter Codeset and 130 µl of Hybridisation Buffer and mixed by inverting. 20 µl of this master mix was added to the strip of 12 sample tube provided in the nCounter Human miRNA Expression Assay Kit. The RNA samples from section 2.11.4 were denatured at 85^0C for 5 minutes before immediately being placed on ice. An aliquot of 5 µl was transferred from each to a new 200 µl PCR tube (also on ice). The thermocycler was preheated to 65^0C. 5 µl of Capture Probeset was added to each 5 µl aliquot RNA sample and then immediately placed on 65^0C. This was incubated for at least 12 hours before proceeding to post-hybridisation processing.

2.11.6 nCounter prep station protocol

This step was performed by Bruno Steinkraus (Weatherall Institute of Molecular Medicine, University of Oxford) and required the components of the nCounter Master Kit (Nanostring Technologies, USA). After allowing

them to equilibrate to room temperature for 30 minutes the sample tubes from section 2.11.5 were loaded on the prep station, along with the prep plate and cartridge provided. The station was programmed to run on high sensitivity mode and allowed to process the samples. On completion all tubes, plates and waste reagents were disposed of.

2.11.7   Processing and Analysis of nCounter data

The nCounter system produces raw counts of molecules for each sample. These files are immediately input into a software package called nSolver (Nanostring, USA). This is specifically designed to perform quality control and normalisation of each sample. It also performs normalisation to internal positive and negative controls (including five housekeeping genes included in the assay cartridge). The normalised data was then input to the edgeR software used by CAP-miRSeq to perform differential analysis as per designated clinicopathologic groups.

Chapter 3

Clinicopathologic analysis of tissue samples

3.1 Introduction

Head and neck cancer is a non-specific term that can refer to any extracranial malignancy above the level of the clavicles. It is generally used to describe malignant tumours of the upper aerodigestive tract. This extends from the lips and nasal cavity to the trachea and cervical oesophagus.

As has been mentioned previously these tumours are therefore heterogeneous in that they are clinically different but histologically similar in that over 90% of all tumours at all subsites are squamous cell carcinoma. There is also some heterogeneity regarding the aetiological factors most strongly related to each subsite. For instance, in nasopharyngeal squamous cell carcinoma (SCC), the Epstein-Barr virus is implicated as a critical aetiologic agent in the majority of cases (Lo *et al.*, 2004). Smoking tobacco is associated with a significantly increased risk of all head and neck cancer subsites (Ansary-Moghaddam *et al.*, 2009). Smokeless tobacco are all associated with an increased risk of oral and pharyngeal SCC (Stockwell and Lyman, 1986). Even the betel nut is recognised to be an independent risk factor for oral cancer (Merchant *et al.*, 2000).

In most countries (including England), the oral tongue is the most commonly affected subsite of the oral cavity representing up to 42% of OSCC (Krishna Rao *et al.*, 2013, Statistics, 2012). This is most strongly associated with smoking and alcohol use (Ansary-Moghaddam *et al.*, 2009). Floor of mouth SCC is less common than oral tongue cancer in England though is associated with similar risk factors. Trends in the different methods of using tobacco (and related substances) result in variation in the incidence of different subsites, even within the oral cavity. In North America and Western Europe, buccal squamous cell carcinoma accounts for approximately 10% of OSCC, whilst in Southeast Asia buccal SCC is the most common form of

OSCC due to the endemic practice of chewing betel nut (Lubek *et al.*, 2013, Chhetri *et al.*, 2000). Despite these subtle variations there is little difference found in outcomes from subsites of OSCC (Camilon *et al.*, 2014, Shaw *et al.*, 2009).

HNSCC of different subsites do vary somewhat in their behaviour, relating to their local anatomy. For example, laryngeal SCC tend to metastasise to cervical node level II – IV, whereas oral SCC tends to metastasise to levels I – III (Candela *et al.*, 1990a, Candela *et al.*, 1990b, Shah, 1990, Shah *et al.*, 1990, Robson, 2001). The risk of cervical metastases from different stages of different subsites also varies slightly and therefore management for each subsite varies slightly, as does the impact of treatment on the patient.

Despite the general similarities, in order to reduce potential heterogeneity, patients with oral tongue cancers were primarily targeted for this research. These patients are predominantly managed with surgery to remove the primary tumour as well as the cervical lymph nodes. This meant that the histopathology archive should have matched tissue samples for each patient that underwent surgical treatment for oral tongue SCC, making it ideal for obtaining tissue and nucleic acid for analysis from both primary tumour and matched lymph node metastasis.

The initial statistical analysis utilised an estimation of background copy number alteration of 5% gain and 5% loss based on previous analysis from Professor Rabbitts' Precancer Genomics group. Simulations based on this and that an average of 100 segments per sample was obtained in previous studies indicated that 20 patients per groups would be sufficient to reliably identify CNA with an excess frequency of 0.3 in one group compared to the other. This did not mean any CNA with a lower excess frequency was not worthy of analysis but must be regarded with greater caution. This minimum number of patients also appeared an obtainable number of patients from the Leeds histopathology archive.

3.2 Aim

The objective of this study was to describe the results of conducting an extensive retrospective search for suitable patients for this project as well as

their clinicopathologic features. This is important to establish the phenotypic features of the patients included in this study in order to identify any potential bias or flaws in subsequent genomic analysis.

3.3 Results

3.3.1 Sample Identification and procurement

An important aspect of searching for suitable patients was understanding the system currently used in Leeds to catalogue tissue samples (CoPath). This enables a search to be performed between defined dates for certain keywords.

It is also important to remember that tissue samples from the same disease can be described in a multitude of different terms and therefore repeated searches needed to be performed using the keywords listed in Table **3**-**1**. The searches were performed for samples between 01/01/2005 and 31/12/2011. These dates were selected to try and ensure that patients included in the study would have undergone sufficient follow-up to provide a minimum of 2 years follow up data.

CoPath search keywords used:	Tongue carcinoma
	Oral carcinoma
	Glossectomy
	Neck dissection
	Level 1/Level I
	Cervical lymph node

Table 3-1: Search terms used on CoPath system

This produced numerous lists, which in total came to over 3000 patients. These required close inspection of the final diagnosis to determine the site of primary tumour. This provided a list of 91 patients that required further interrogation using the clinical Patient Pathway Manager (PPM) system.

PPM is used to record details of treatment and follow-up for all oncology patients treated at Bexley Wing, St James' Institute of Oncology. Using this system allowed review of each patients clinical background as well as review

of their full histopathology and management to ascertain whether they met the inclusion criteria specified in Chapter 2.2. A number of patients were excluded due to a history of previous HNSCC (n = 5), previous head and neck radiotherapy (n = 1) and a history of autoimmune disease for which they were on immunosuppressants at the time of their diagnosis and treatment (n = 2). This resulted in 71 eligible patients being identified with primary oral tongue cancer that were treated with surgery to the primary tumour and neck dissection. These were assigned to groups according to histopathologic nodal status (N0, N+ECS-, N+ECS+).

A full, anonymised histopathology report of both primary tumour and neck dissection specimens was obtained for every patient. This allowed identification of the specific FFPE tissue blocks within each report that should contain cancerous and metastatic tissue. This is a vital part of the process as a single patient with oral cancer regularly has upwards of twenty tissue blocks created for clinical purposes. Requesting all blocks and subsequently creating H & E slides for examination would have been an unnecessary waste of time and energy.

A pipeline between our research group and Bexley Wing Histopathology Department was established with the approval of Professor MacLennan and Dr Alec High. Andy Clarke, Chief Biomedical Scientist, at Leeds Histopathology department kindly handled all requests for specific tissue blocks via email. These were then retrieved from the archive at which point I was notified to collect them from Bexley Wing.
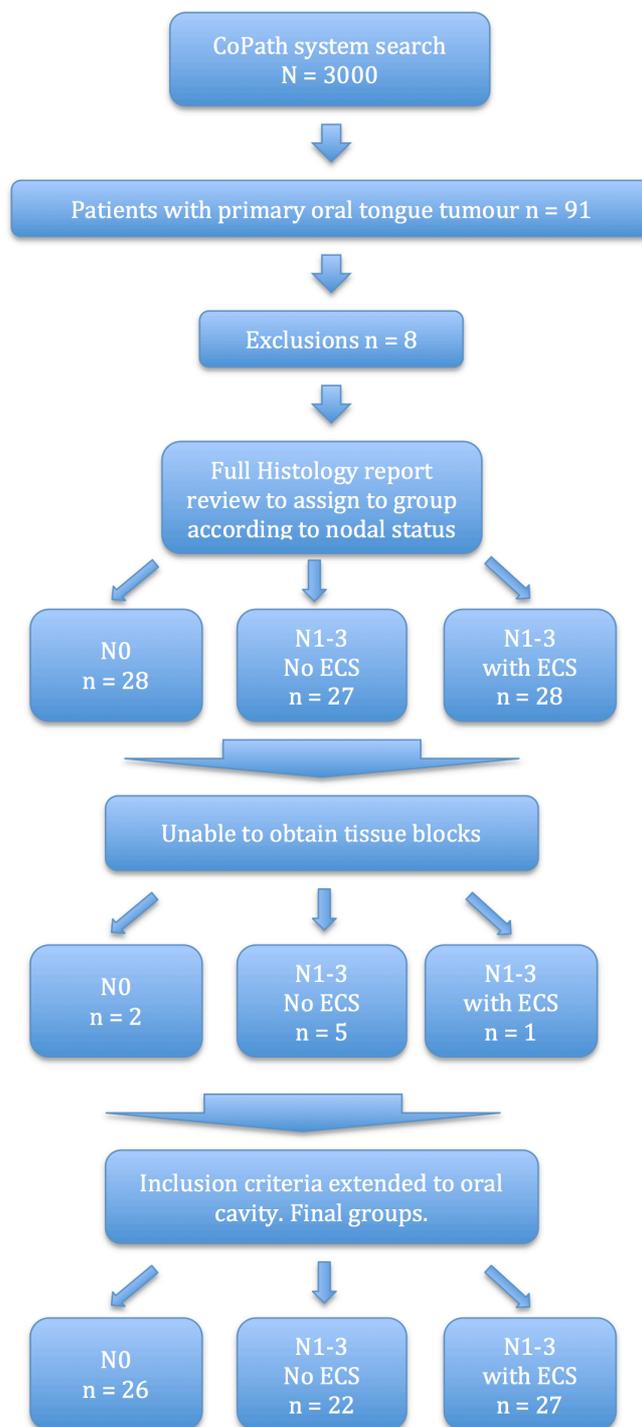
Figure 3-1: Chart to show sample identification and procurement

A significant problem became apparent after requesting several batches of tissue blocks: missing or unobtainable blocks. These missing blocks were either the primary tumour or matched metastasis, or both. Unfortunately, beyond contacting the last person recorded to have possession of the blocks there was little that could be done to overcome this issue. This resulted in

insufficient patient numbers for the patients groups. In order to overcome this issue the original inclusion criteria was broadened from oral tongue to oral cavity. This was carried out as the tissue blocks for several patients in both the ECS+ group and the N+ECS- group appeared to be potentially reducing numbers significantly in these groups. The final numbers of patients obtained are detailed in Figure **3**-**1**.

3.3.2 Clinicopathologic characteristics of samples

In total 75 patients were included in the study for clinicopathologic and subsequent CNA analysis. The diagnosis of SCC and the nodal status was confirmed using the H & E by Dr Preetha Chengot (consultant head and neck histopathologist).

Variable	No. of patients (%)
Sex	
Male	52 (69)
Female	23 (31)
Age/yrs	
Mean (median)	62.5 (60.7)
Range	36.3 – 90.2
Tobacco use	
Current	40 (53)
Ex	18 (24)
Never	17 (23)
Primary Tumour Subsite	
Oral Tongue	62 (83)
Floor of mouth	9 (12)
Buccal	3 (4)
Alveolar	1 (1)

Table 3-2: Summary of clinical characteristics of sample population

All of these cases had DNA successfully extracted for copy number analysis, from both the primary tumour and matched metastasis, as well as a further sub-group being selected for miRNA extraction and sequencing. The clinical and histopathologic characteristics of these patients were summarised in Table **3**-**2** and Table **3**-**3**. More detailed tables of individual patient samples can be found in Appendix 8.2.

The distribution of tumour subsites according to nodal status is shown in Figure **3**-**2**. There male to female ratio was 2.3:1. The preponderance of male

patients is representative of the distribution to be expected in the wider population.

Variable	No. of patients (%)
pT stage	
pT1	29 (39)
pT2	32 (43)
pT3	9 (12)
pT4a	5 (7)
pT4b	0
pN stage	
pN0	26 (35)
pN1	18 (24)
pN2a	3 (4)
pN2b	22 (29)
pN2c	4 (5)
pN3	2 (3)
Extracapsular spread	
Absent	48 (65)
Present	27 (35)
Histopathologic Grade	
Well	24 (32)
Moderate	26 (35)
Poor	25 (33)
Perineural invasion	
Present	37 (49)
Absent	38 (51)
Lymphatic invasion	
Present	10 (13)
Absent	65 (87)
Depth of tumour invasion/mm	
Mean (Median)	12.8 (11)
Range	2.5 – 45

Table 3-3: Histopathologic characteristics of sample population
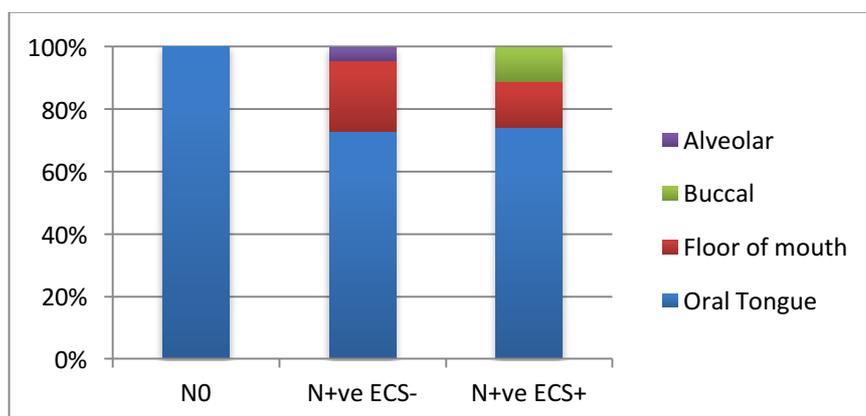
Figure 3-2: Distribution of primary tumour subsite across nodal status grouping

The distribution of T-staging across nodal status groups is shown in Figure **3-3** and the N-staging according to the presence of absence of ECS is

shown in Figure **3-4**. These show a trend towards smaller tumours associated with N0 nodal status, whilst larger tumours tend to have positive nodes and the largest tumours more likely to have ECS. However, 14 patients with ECS (52%) had T1/T2 primary tumours. In terms of N-staging, though disease with ECS is associated with more advanced staging (indicating either greater number or size of nodal metastases), 3 patients (11%) with ECS were staged at N1. This is again consistent with the knowledge that ECS indicates aggressive disease rather than late disease.
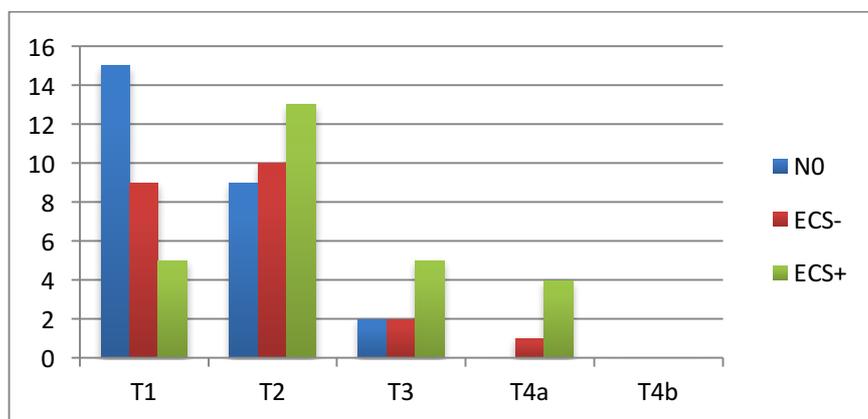
Figure 3-3: Distribution of T-staging across nodal status grouping
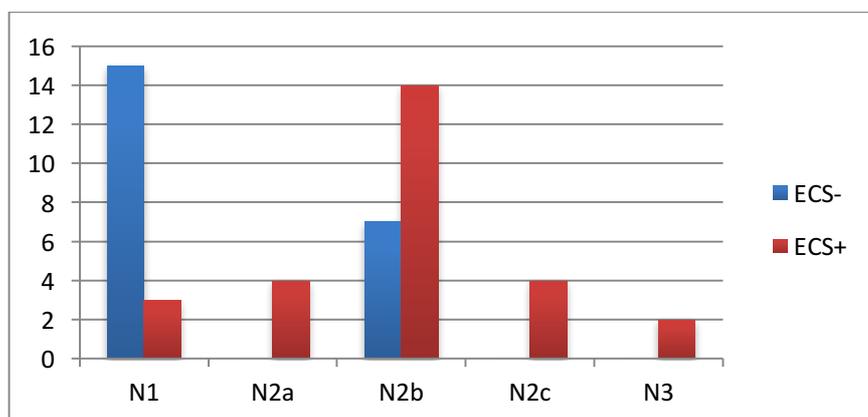
Figure 3-4: Distribution of N-staging according to presence or absence of ECS

Depth of invasion of primary tumour is a histopathological characteristic, previously associated with poorer outcome and metastasis in OSCC (Asakage *et al.*, 1998). In this group of patients depth of invasion was recorded for 64 of 75 patients. The average depth of tumour was higher in the ECS+ group compared to the others and similarly the average depth of

tumour in the N+ECS- was higher than the N0 group. This association was statistically significant using the ANOVA test (p = 0.015).

Univariate analysis was performed of histopathologic variables and their relationship to the presence of ECS or cervical metastases. The chi-squared test was used, except where only two groups were present where Fisher's Exact test was used. Tumour differentiation, lymphatic invasion, perineural invasion and vascular invasion were analysed (see Table **3**-**4**).

Histopathologic characteristic	No. of patients with metastases (%)	p-value	No. of patients with ECS (%)	p-value
Differentiation				
Well	11 (15)	0.016	11 (22)	0.75
Moderate	16 (21)		10 (20)	
Poor	22 (29)		6 (12)	
Lymphatic invasion				
Yes	9 (12)	0.15	7 (14)	0.16
No	40 (53)		20 (41)	
Perineural invasion				
Yes	29 (39)	0.03	17 (35)	0.57
No	20 (27)		10 (20)	
Vascular invasion				
Yes	18 (24)	0.002	9 (18)	0.77
No	31 (41)		18 (37)	

Table 3-4: Univariate analysis of histologic variable and metastasis and ECS

Univariate analysis was performed of histopathologic characteristics and their relationship to recurrence. This was done using chi squared test (and Fisher's Exact test where both variables had only two groups). There was a strong statistically significant relationship between the presence of ECS and locoregional recurrence of disease. Interestingly there was no difference between the N0 group and N+ECS- groups in the rate of locoregional recurrence (p = 0.6). For characteristics such as tumour differentiation, lymphatic invasion, perineural invasion and vascular invasion the relationship with locoregional recurrence is presented in Table **3**-**5**.

Histopathologic Characteristic	No. of patients with recurrence (%)	p value
Differentiation		
Well	6 (8)	0.99
Moderate	6 (8)	
Poor	6 (8)	
Lymphatic invasion		
Yes	5 (7)	0.05
No	13 (17)	
Perineural invasion		
Yes	11 (15)	0.29
No	7 (9)	
Vascular invasion		
Yes	5 (7)	0.77
No	13 (17)	
ECS		
ECS+	14 (52)	0.0001
N+ECS-	1 (5)	
N0	3 (12)	

Table 3-5: Univariate analysis of histopathologic variables and recurrence

3.3.3 Survival Analysis

The range of length of follow-up for any patient alive at the end of the study period was 34 – 111 months (mean – 68.5 months). The presence of ECS was found to have a significant adverse effect on both disease-free survival (p = < 0.0001) and overall survival (p = < 0.0001) (see Figure **3-5** and Figure **3-6**). There was no statistically significant difference in survival between the pN0 and pN+ without ECS groups.
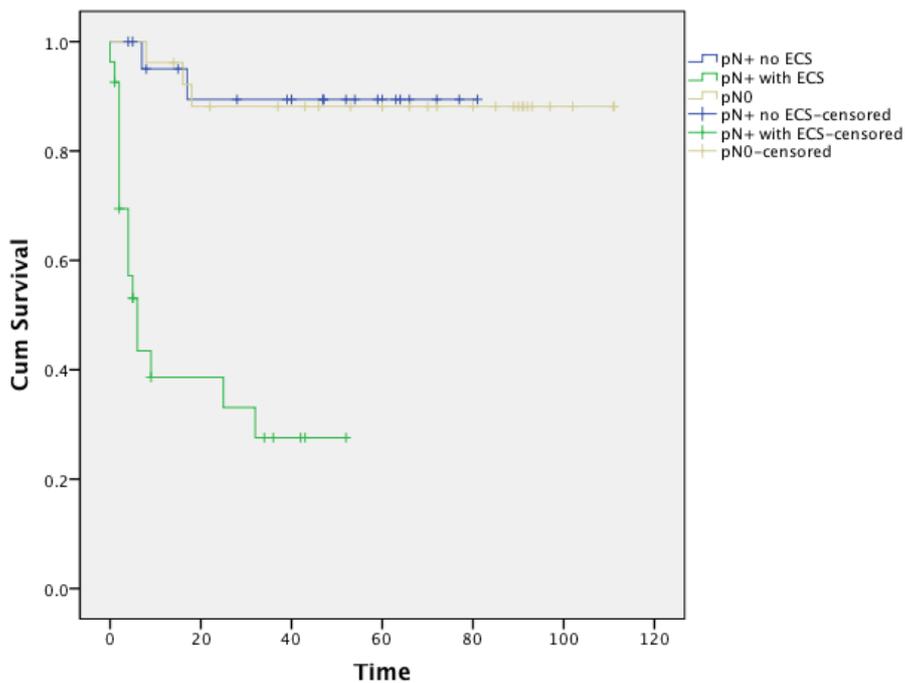
Figure 3-5: Kaplan-Meier disease-free survival curve for pN0, pN+ with and without ECS
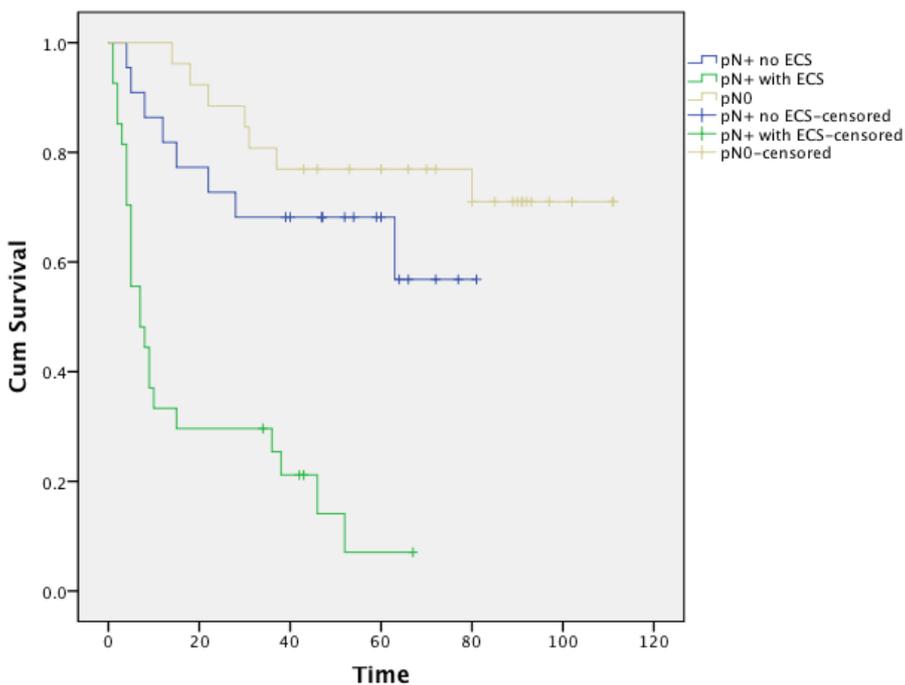
Figure 3-6: Kaplan-Meier overall survival curve for pN0, pN+ with and without ECS

3.4 Discussion

HNSCC is regularly described as a heterogeneous group of cancers. In choosing a group of patients to study I attempted to create as homogeneous a group as possible by targeting a specific primary tumour subsite. The compromise that had to be made with this decision is that the pool of patients from which patients can be selected is reduced, therefore obtaining numbers that could provide statistical significance later became a significant challenge in this study.

A long, laborious search through massive lists of potentially candidates from the CoPath system was required. An important consideration was that prior to 2008, the histopathology department used a number of different systems to catalogue tissue samples due to there being separate IT systems at both Leeds General Infirmary and St James' University Hospital. These have since been unified at Bexley Wing, St James' University Hospital. Samples prior to 2005 have not been transferred to CoPath and whether all samples between 2005 and 2008 have been reliably transferred to the CoPath is unknown. I was unable to gain access to the older systems to search these independently. This did hinder my ability to get more patients.

Tissue blocks that were missing from the Pathology archive had a significant impact on the numbers of patients with oral tongue primary tumours I was able to obtain. In order to combat this I decided to include patients with any subsite of OSCC in order to obtain the minimum group size targeted. Though subtle variation in aetiology has been reported (e.g. smoking tobacco for oral tongue vs. chewing tobacco for buccal) there are conflicting reports regarding the impact of subsite on the clinical behaviour and outcomes of these tumours in terms of survival and recurrence (Camilon *et al.*, 2014, Lubek *et al.*, 2013, Chhetri *et al.*, 2000). In this respect, I needed to bear in mind in later analysis that 13 patients in my study had primary tumour subsites other than oral tongue, particularly in that buccal cancers were only in the ECS+ group. Little work has been reported on subsite-specific molecular signatures in HNSCC. An immunohistochemistry study found simultaneous downregulation of p16 and p21 was found in 47% of oral

tongue against 28% of buccal SCC, suggesting some molecular differences, though these were by no means exclusive (Sathyan *et al.*, 2006).

The number of patients obtained in the original CoPath search per year is shown in Figure **3**-**7**. This demonstrates a fairly even spread of patients between 2006 and 2011, though in 2005 only four patients were obtained. This suggests that not all patients from this time were entered on to the CoPath system, likely contributing to the issues in obtaining samples.
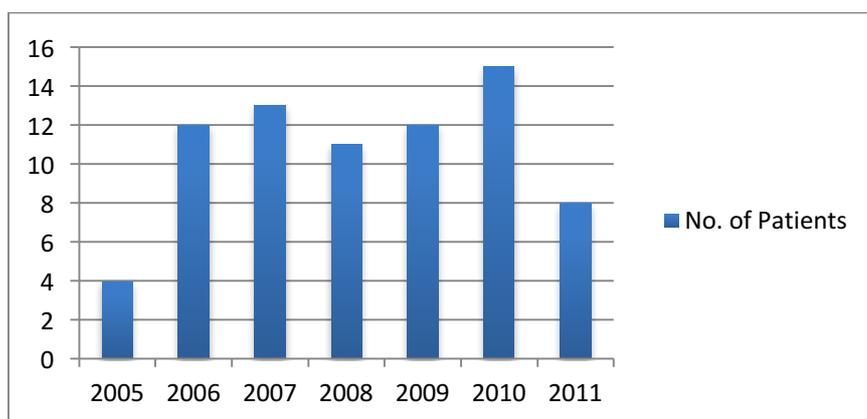
Figure 3-7: Chart to show distributions of patients over time

None of the histologic characteristics analysed in Table **3**-**4** demonstrated any significant association with ECS. Though grade of differentiation, perineural invasion and vascular invasion all showed a significant association with metastasis, it is important to recognise that 20 patients had no perineural invasion yet still were found to have cervical metastases. More patients were found to have cervical metastases in the absence of vascular or lymphatic invasion. These findings highlight the lack of clinical reliability in these histologic characteristics in predicting metastases or ECS. This is in keeping with previous studies examining the relationship of these variables in relation to patient outcome. In a study of 142 patients with mixed subsite HNSCC Fagan *et al* found cervical metastases in 73% of patients with perineural invasion. However they also found metastases present in 43% of patients without perineural invasion (Fagan *et al.*, 1998). Soo *et al* found no association between perineural invasion and cervical metastases (Soo *et al.*, 1986). Similarly studies examining intravascular and lymphatic invasion have

found associations with metastasis in HNSCC, which are statistically significant but not 100% predictive of metastases (Close *et al.*, 1989, Close *et al.*, 1987, Okada, 2010).

In terms of disease recurrence the absence of lymphatic invasion was associated with absence of recurrence of disease on the limits of statistical significance. Though one would expect a strong link between this and poorer outcome this finding is consistent with other studies in OSCC showing some potential association but certainly not as a foolproof marker of recurrence (Goldstein *et al.*, 2013, Olsen *et al.*, 1994). As with many other studies previously discussed the strong link between ECS and recurrence was again shown in this study (p = 0.0001).

This group of patients is a very small population and therefore it is difficult to draw any significant conclusions from analysis of the histologic variables of these samples. However, it is reassuring that in terms of both overall and disease-free survival a massive difference was clearly demonstrated between those patients with and without ECS. In this respect it provides evidence that these patients behaved clinically as would be expected of patients with their histopathologic findings and therefore were a suitable group to utilise in further genomic analysis.

Chapter 4

Copy number analysis of HNSCC and matched cervical metastases to identify characteristics of metastatic and non-metastatic HNSCC and of tumours associated with ECS

4.1 Introduction

The management and prognosis for patients with HNSCC of any subsite is significantly affected by firstly the presence of cervical metastases and in particular by the presence of ECS (Shaw *et al.*, 2010, Coatesworth and MacLennan, 2002). Patients with a clinically N0 neck represent a group in whom treatment to the neck is often warranted on the basis of the risk of occult cervical nodal metastases. Cervical metastases are present in approximately 30% of patients with clinically N0 disease (van den Brekel *et al.*, 1993, Weiss *et al.*, 1994, Alvi and Johnson, 1996, Haddadin *et al.*, 1999). Approximately 60% of these occult cervical metastases have been shown to contain ECS (Coatesworth and MacLennan, 2002). This means that 18% of patients with clinically N0 disease actually have ECS. This would change them from being considered in the group at lowest risk of disease recurrence or progression to the group at highest risk instantly. Patients with clinically undetectable nodal metastases and ECS represent a group that could benefit from being targeted specifically with systemic therapies. However, due to the absence of a reliable marker to identify metastases and ECS, many patients undergo neck dissection to provide histological staging of the neck or have prophylactic radiotherapy to the neck. This means that up to 70% of patients have unnecessary treatment, which is associated with considerable morbidity.

A decision analysis by Weiss *et al* concluded that when the probability of occult metastasis is more than 20% then elective neck treatment would be preferable (Weiss *et al.*, 1994). This benchmark is used widely in clinical

practice and therefore the majority of HNSCC patients with clinically N0 disease undergo treatment to the cervical nodes. Only small T1 oral and glottic carcinomas can potentially avoid this. Retrospective studies comparing observation to elective neck dissection in clinically N0 disease have been performed. Haddadin *et al* found 5-year survival reduced from 80.5% to 44.8% when patients with T1/T2 N0 OSCC were observed instead of undergoing elective neck dissection (Haddadin *et al.*, 1999). Similar findings were reported by Kligerman *et al* (Kligerman *et al.*, 1994).

The only prospective randomised trial with sufficient power regarding the issue of elective neck dissection in early oral cancer was recently published in 2015. This found an absolute 5-year survival benefit in those patients having undergone elective neck dissection as opposed to therapeutic neck dissection (where a watch and wait approach was taken) of 12.5% (D'Cruz *et al.*, 2015). This survival benefit was likely due to the fact that 30% of clinically N0 necks treated with elective neck dissection were found to contain occult metastases, as the neck was the most common site of disease progression/recurrence in the group having undergone therapeutic neck dissection. It is likely that the group of patients with occult cervical metastases with ECS represent those in greatest need of elective treatment. In highlighting the need to treat patients who may have occult cervical metastases we have to remember the patients that undergo treatment who do not have any nodal metastases. These are still the majority receiving treatment.

As discussed in the previous chapter, histopathologic studies for metastasis have failed to produce a reliable marker in the primary tumour for nodal metastases. Genomic studies could provide such a marker. Copy number alterations (CNA) are known to disrupt a larger proportion of the cancer genome than any other type of somatic genetic alteration (Beroukhim *et al.*, 2010, Zack *et al.*, 2013). CNA studies led to the discovery of distinct clinical subsets of cancer patients in lung cancer and significant disrupted pathways in acute lymphoblastic leukaemia (Chitale *et al.*, 2009, Mullighan *et al.*, 2007). CNA studies can allow the identification of the positions of oncogenes and highlight putative targetable pathways (Weir *et al.*, 2007, Zender *et al.*, 2006).

Previous studies in HNSCC have utilised technologies such as FISH, PCR and comparative genomic hybridisation to demonstrate commonly occurring areas of chromosomal gain mapped to 1q, 2q, 3q, 5p, 8q, 11q and areas of chromosomal loss mapped to 1p, 3p, 4p, 11q, 13q, 14q, 17q and 20q (Chen and Chen, 2008, Smeets *et al.*, 2006, Struski *et al.*, 2002). Despite this, relatively little is know about the genomic changes that lead to metastasis. Few studies have specifically aimed to evaluate CNA markers for metastasis in HNSCC primary tumours. A study by Patmore *et al* compared 23 HNSCC primary tumours to matched nodal metastases using CGH, finding no markers for metastasis (Patmore *et al.*, 2004). This was limited by the small number of patients, lack of non-metastatic controls and mixing of primary tumour subsites (in particular mixing oropharyngeal SCC with non-oropharyngeal SCC). Yoshioka *et al* profiled 25 patients with OSCC and their matched nodal metastasis using array-based CGH (Yoshioka *et al.*, 2013). This study found gains of 7p, 8q and 17q were more common in lymph node metastases compared to non-metastatic primary tumours implying they could be involved in metastasis. Though this did include a group of patients that were pN0 there was still a smaller number of patients and they did not take into account those tumours that produced metastases with ECS.

Single nucleotide polymorphism (SNP) arrays have been demonstrated to provide high resolution analysis of CNA within cancer cells, though they have limited power to detect CNA in regions not targeted by the array and any study is limited by the resolution of the array being used (Bignell *et al.*, 2004). NGS can provide copy number data in a way analogous to arrays (Wood *et al*, 2010). Each genomic window consists of 200-900 sequencing reads that uniquely align to that genomic region (using a reference genome to align). Each genomic window can be considered equivalent to one array probe (Wood, 2013). This represents a genomic signal corresponding to a specific genomic location (Zhang and Gerstein, 2010). Whilst array data is calculated from log-ratios of normalised intensity of signal from the test sample to the reference sample, NGS copy number data is calculated from sequencing read counts after accounting for depth of sequencing (Zhang and Gerstein, 2010). Copy number sequencing (CNSeq) has been

demonstrated to be reproducible and reliable (Schweiger *et al.*, 2009, Wood *et al.*, 2010). Wood *et al* demonstrated that the data produced had a very high correlation to an array with an equivalent number of probes (Wood *et al.*, 2010). Hayes *et al* also reported a 100% concordance between CNSeq and aCGH for identifying pathogenic copy number variants, with CNSeq identifying one purported pathogenic variant not detected by aCGH (Hayes *et al.*, 2013).

NGS offers significant advantages in terms of CNA analysis. These include the ability to estimate the tumour-to-normal copy number ratio by counting the read number at specific loci, resolution that can be chosen (pre-sequencing according to aim and cost) but can be manipulated post-hoc. In addition to offering comparable resolution to array-based technologies, CNSeq can be performed reliably and consistently with nanogram quantities of DNA. It can also use degraded sources of DNA such as FFPE tissue avoiding the need for extra preprocessing steps that are often required with arrays (Wood *et al.*, 2010, Gusnanto *et al.*, 2012, Hayes *et al.*, 2013, Gusnanto *et al.*, 2014). The sequencing library can also be reused in multiple experiments at varying depths of sequencing, which can be tailored to budget (Wood, 2013).

One of the most striking features of the literature evaluating CNAs relating to an increased risk of metastasis, is the lack of concordance amongst the different studies (see Table 1-5). Differences in approach could explain some of these discrepancies, particularly when comparing genome-wide approaches to more targeted approaches such as FISH or PCR. Even when comparing apparently the same genome-wide technique, the age of the study (as a surrogate marker of technology generation), issue of resolution, source tissue type (FFPE vs fresh-frozen vs. cell-line) need to be considered as confounding factors. Though the subsite of HNSCC could influence findings it is more likely that the aetiology of the tumour would have a more significant effect on genomic findings for example smoking tobacco, HPV, chewing betel nut. The aetiology could be heavily influenced by the epidemiological background of the samples i.e. betel nut chewing is uncommon in the UK and western Europe but endemic in the Indian

subcontinent. The issue of inter and intra-tumour heterogeneity may affect the concordance of genomic findings between studies.

Beroukhim et al, highlighted the disagreement between studies analysing genomic alterations in the same cancer as a serious issue suggesting differing analytical methods as a potential cause for this (Beroukhim *et al.*, 2007). They also raised the issue that many potential regions of interest put forward by studies could be random events of no biological significance as the background rate of random chromosomal aberrations is not clearly accounted for in many studies. For this reason they developed a systematic approach to analysing chromosomal aberrations in cancer called Genomic Identification of Significant Targets in Cancer (GISTIC) which has been found to significantly improve concordance in genomic alterations found in glioma studies.

This was since updated to a second version (GISTIC 2.0) to provide estimation of focal and broad-level CNAs. Segmented data files are presented to GISTIC 2.0 of identical window size. The copy number profile is then deconstructed into underlying CNAs and GISTIC 2.0 models the background rate of CNAs separating focal and arm-length CNAs on the basis of length. A G-score is then calculated for each CNA (a log ratio of probability of specific CNA to the background rate of CNAs). A "peel-off algorithm" is then applied by GISTIC 2.0, that subtracts segments covering each peak until no significant segments remain on the chromosome. This defines the independent genomic regions undergoing significant CNA. Finally, the G-score is used to define the boundaries of the focal CNAs (Mermel *et al.*, 2011). In addition to providing a standardised, reproducible approach to analysis GISTIC was also produced with the aim of identifying driver genomic event in cancer and ideally using CNA data to localise specific driver genes (Beroukhim *et al.,* 2007).

Though not the only software available to analyse CNA, GISTIC 2.0 has been utilised by all TCGA sub-groups and can be applied to both sequencing and array-generated data, providing segmented data files can be submitted.

4.2 Aims

The work described in this chapter aimed to:

1. Use NGS CNA data to compare genomic features of OSCC primary tumours and their matched cervical node metastases.
2. Use NGS CNA analysis to identify genomic damage patterns within OSCC metastatic and non-metastatic primary tumours.
3. Use NGS CNA analysis to identify genomic damage patterns that are specific to OSCC primary tumours associated with nodal extracapsular spread.

4.2.1 Results 1: Comparison of CNA profiles of OSCC primary tumours to matched cervical node metastases

4.2.1.1 Summary of patient cohort

This initial analysis planned to compare the DNA copy number profile of primary tumours and their matched nodal metastasis. This included 49 patients with a OSCC primary tumour and a paired nodal metastasis (see Table **4**-**1**). In each patient the archived FFPE tissue blocks were obtained. A 5 μm haematoxylin and eosin section was cut from each sample. Professor Kenneth MacLennan or Dr Preetha Chengot (consultant head and neck histopathologists) used this section to confirm the diagnosis and mark out the area of highest tumour cell content. DNA was successfully extracted from all 98 samples (49 tumour-metastasis pairs). These were all processed to copy number libraries (see Chapter 2.8) and multiplexed on the Illumina HiSeq 2500 at 40 samples per lane. The mean number of reads obtained per sample was 10,865,175 (range: 1,448,220 – 86,693,194). Only one sample had a detectable viral load (ECS066-L) which had a barely detectable viral load of 0.2 viral copies per cell.

Primary tumour site	Oral tongue	Floor of mouth	Buccal	Alveolar	Total
No. of patients	36	9	3	1	49

Table 4-1: Table of all OSCC tumour (with matching nodal metastasis) subsite

4.2.1.2 Individual karyograms of metastatic primary tumours

The DNA sequencing libraries were multiplexed on the Illumina HiSeq 2500 at 40 samples per lane. The resulting data was processed to a digital karyogram for each patient (see Chapter 2.9). The karyograms produced using CNAnorm allow a genome wide view of CNA in each sample. Even in visual inspection of the karyograms it is important to exclude specific areas of the genome from analysis. CNAs have been found to be overrepresented in genomic regions close to centromeres and close to telomeres (within 2 Mb) as CNAs do not appear to arise uniformly throughout the genome and are overrepresented in these regions (Nguyen *et al.,* 2006). This is thought

to be related to the repetitive nature of the nucleotide sequence in these regions, which results in reads aligning with poorer quality scores to these regions, making interpretation of them unreliable. As such, CNAs incorrectly appear the be enriched in these regions. Similarly chromosome (chr) 19 has long been regarded as unreliable when assessing CNA and was also interpreted with caution (Kallioniemi *et al.*, 1994). This is due to the wider, normal individual variation existing in chr 19 and the reference genome therefore not being reliable.

For each karyogram an assessment of the tumour cell DNA fraction of the sequenced sample is possible. Though a pathological estimate was made of all tissue blocks to try and obtain the highest dissected tumour cell fraction, this may not be identical to the fraction of DNA obtained from the dissected sample that is actually tumour, and that is subsequently sequenced. Firstly, each individual karyogram was inspected and the segments at which loss/gain of one copy of a genomic region is to be called at for that sample identified, as opposed to background noise due to mixed tumour clone sampling (see Figure **4**-**1**). Simultaneously, an estimation of ploidy can be made from inspecting the karyogram. Once identified, the specific change in ratio of minimum gain/loss to normal can be obtained from the individual .bed file. This threshold can then be used to estimate the sample tumour DNA content by scaling the figure to that which would be produced by a gain/loss of one copy in a sample with 100% tumour DNA content (i.e. +1/-1), accounting for ploidy. The CNA call threshold for each sample is needed for downstream analysis.

Due to a pooled normal being used that was made up of 10 men and 10 women (therefore containing a mean pooled 1.5 chr X and 0.5 chr Y), it was not possible to assess the copy number status of the X and Y chromosomes. This is because every male sample would automatically erroneously be labelled as having loss of chr X and gain of Y and every female sample assigned gain of chr Y and loss of X. In addition to this the level of coverage obtained with my sequencing was arguably not sufficient to reliably determine the status of these chromosomes as they contain large, repetitive regions.

Initial visual inspection of the individual karyograms of the primary tumours obtained from patients with nodal metastases show a number of CNAs in the majority. Losses were most frequently mapped to 3p, 4p, 5q, 8p, 9p, 11q and 18q. Gains were most frequently seen on 3q, 5p, 7p, 8q, 12p, 14q, 20p and 20q. Two samples showed a dramatically reduced number of CNAs.

All individual sample karyograms are shown in Appendix 8.3.
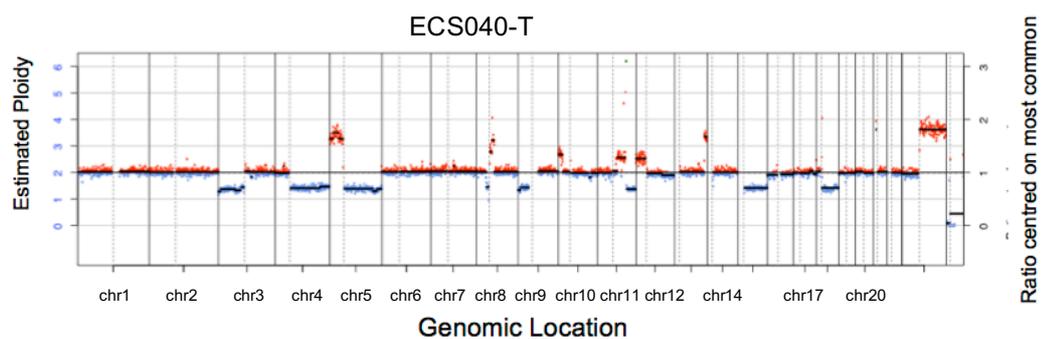
Figure 4-1: An example of a karyogram from an OSCC primary tumour associated with nodal metastasis. Genomic windows with copy number gain are red and those with loss are blue. The tumour DNA fraction of the sample can be estimated from the ratio of the minimum gain/loss to normal. In this case the tumour DNA fraction is approximately 75%.
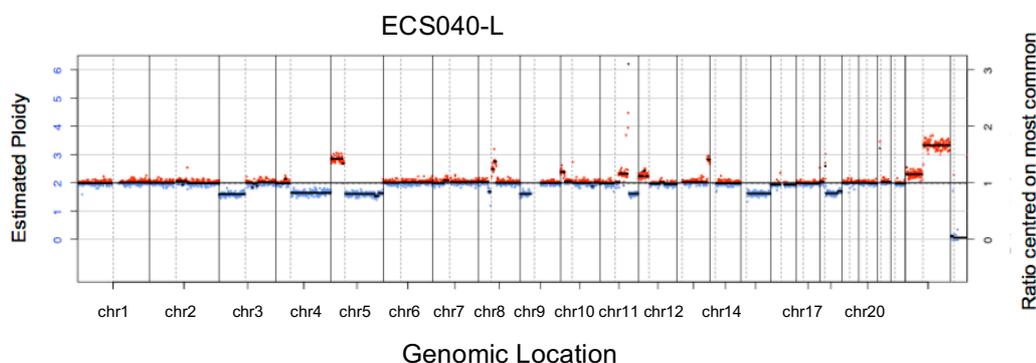
Figure 4-2: The karyogram produced from the matching nodal metastasis from the same patient as fig. 4-1. The difference in height of gains and losses when compared to the primary tumour is due to a lower tumour DNA fraction.

The digital karyograms from the matched nodal metastasis of each primary tumour were also inspected. An example is shown in Figure 4-2. All CNAs and genomic breakpoints within this nodal metastasis were shared with the matched primary tumour suggesting that the clonal cell population sampled from the metastasis was identical to the dominant clonal population sampled from the primary tumour. The level of similarity varied widely amongst tumour-metastasis pairs. Visual inspection and counting of CNAs and

breakpoints remained a slow method of demonstrating and measuring this across the group and so other methods of comparing the primary tumours and varied metastases were evaluated.

4.2.1.3 Cumulative frequency CNA plots of metastatic primary tumours and paired nodal metastases

In order to be able to compare the primary tumours to their matched metastasis as groups cumulative frequency plots were produced. Initial creation of these using seg_compare (Dr Henry Wood, Precancer Genomics group) was attempted using a group-wide minimum threshold for calling areas of gain or loss. This produced cumulative frequency plots that appeared to include a lot of spurious CNAs (e.g. in centromeric or telomeric regions) as well as appearing not to include areas that appeared to be significant on individual karyogram inspection (e.g. loss of 3p) and therefore falsely reducing the proportion of patients with certain regions of gain or loss (see Figure **4**-**3**).

In these plots chromosomal regions of gain are represented in red and regions of loss are highlighted blue. Along the y-axis is the frequency which each gain or loss is seen across the group (as opposed to the tumour to normal ratio presented on the y-axis in the individual karyograms). By assessing each karyogram individually and estimating the tumour DNA fraction and ploidy the minimum threshold for each sample was set for each sample. Cumulative frequency plots were then generated taking into account the individual sample thresholds (see Figure **4-4**). This method of generating cumulative frequency plots reduced spikes of CNA in spurious regions (i.e. close to centromeres and telomeres) and therefore appeared to present a "cleaner" image.

As can be see in Figure **4-4** there are few chromosomal arms where no gain or loss is seen in any sample. Despite this the majority of chromosomal instability is only seen in a minority of samples. For initial assessment I focused on CNAs occurring in at least 50% of the samples.
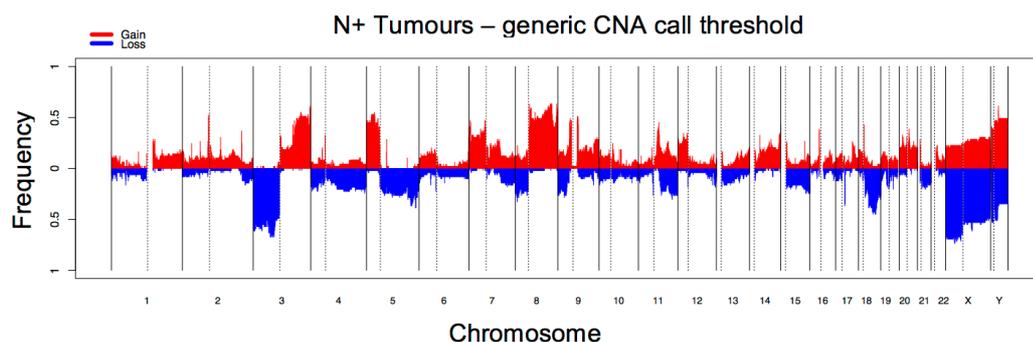
Figure 4-3: Cumulative frequency plot of OSCC primary tumours associated with nodal metastasis. Created using an arbitrary identical CNA threshold for all samples. Regions of gain are represented by red and loss by blue. In this plot approximately 50% of patients are demonstrated to have loss of 3p. This was considerably less than observed on inspecting the individual sample karyograms.
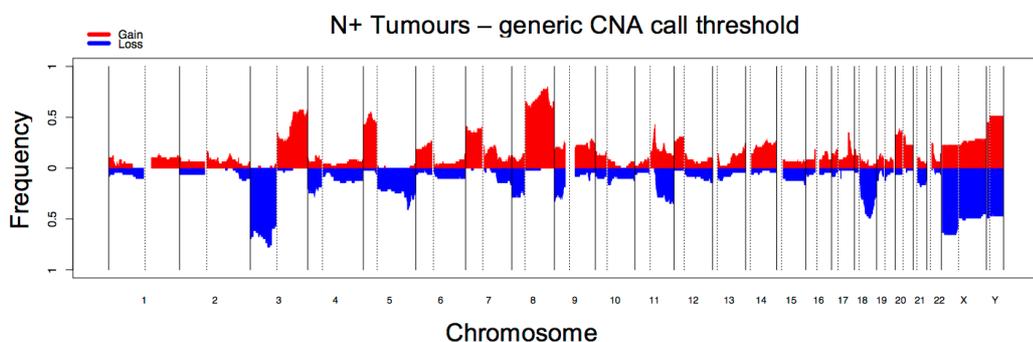
Figure 4-4: Cumulative frequency plot of OSCC metastatic primary tumours. This was created after determining the minimum gain/loss threshold for each sample included in the group. This resulted in approximately 78% of samples having loss of 3p, an accurate reflection of individual karyogram assessment. It also contains far fewer spurious spikes of gain and loss in centromeric and telomeric regions making the overall plot easier to interpret.

The most common CNAs were loss of 3p and gain of 8q, both found in 38/49 (78%) of samples. Gain of 3q was demonstrated in 30/49 (61%) similar to gain of 5p (29/49 (60%)). These are quintessential genomic changes well recognised to be associated with HNSCC and lung SCC (to be expected given the unifying common aetiology is tobacco smoking). Loss of 18q was also shown in 25/49 (50%) of patients.

Regions of gain and loss occurring at lower rates in this group include gain of 11q (21/49 (43%)), loss of 7p (20/49 (40%)), loss of 5q (18/49 (37%)), gain of 20p (18/49 (37%)), loss of 11q (17/49 (34%)), gain of 17p (17/49 (34%)) and gain of 12p (17/49 (34%)).

The seg_compare scipt (Dr Henry Wood, Precancer genomics group) also generates cumulative frequency plots for each chromosome, as well as the genome-wide view above. This allows more specific observations to be made about regions of gain and loss (see Figure **4-5**). These images suggest that most CNAs are broad. This is in keeping with studies of CNAs across a range of human cancers (Beroukhim *et al.*, 2010). However this provides a challenge when trying to identify foci of minimally-altered regions to nominate candidate oncogenes or tumour suppressor genes as broad-level CNAs encompass thousands of genes. For this type of analysis the cumulative frequency karyograms are limited and required different software to be used at a later point. The cumulative individual chromosome plots allow a low resolution view of cytobands and the frequency of CNAs associated with the, across the whole group.

As described above, individual karyograms were generated for each DNA sample obtained from the matching nodal metastasis of each primary tumour in this group (n = 49). This was done using the copy number gain/loss threshold individually determined for each sample as outlined previously (see Figure **4-6**). This demonstrated a very similar pattern to the cumulative frequency plot generated for their respective primary tumours (see Figure **4-4**). A list of the CNAs and frequencies in each group is listed in Table **4-2**.
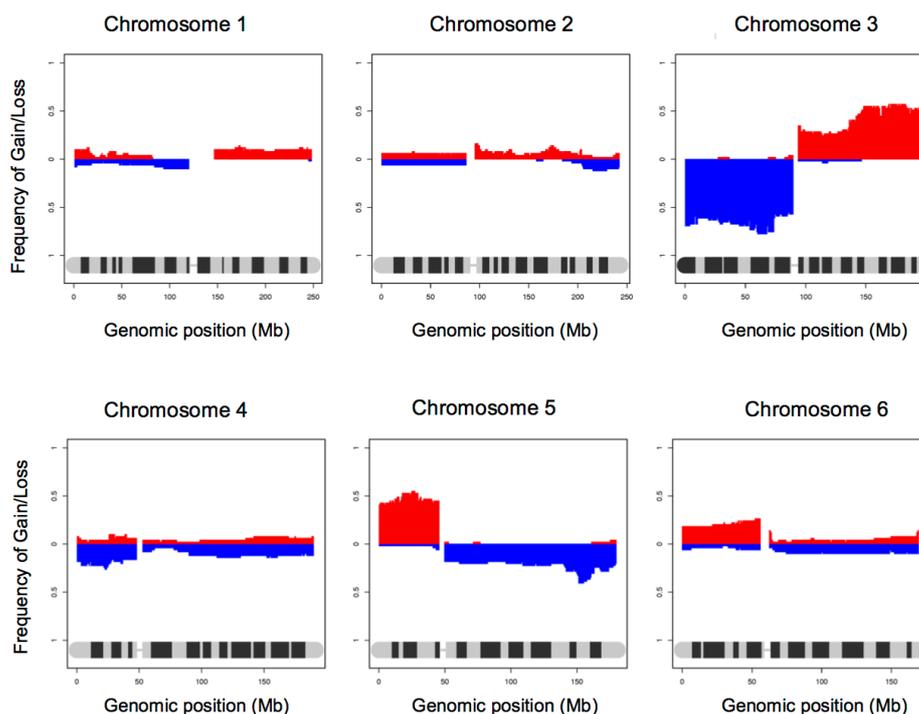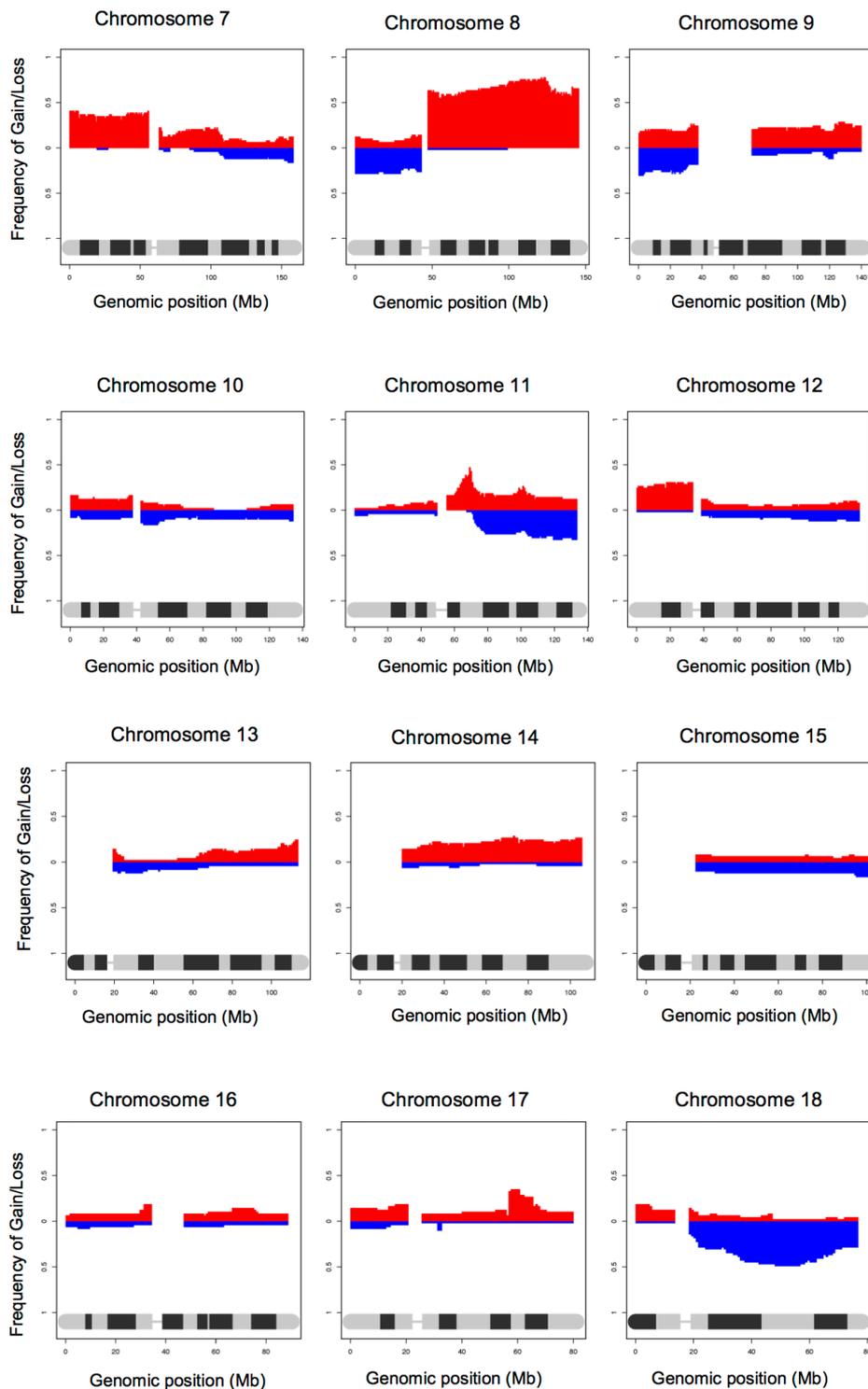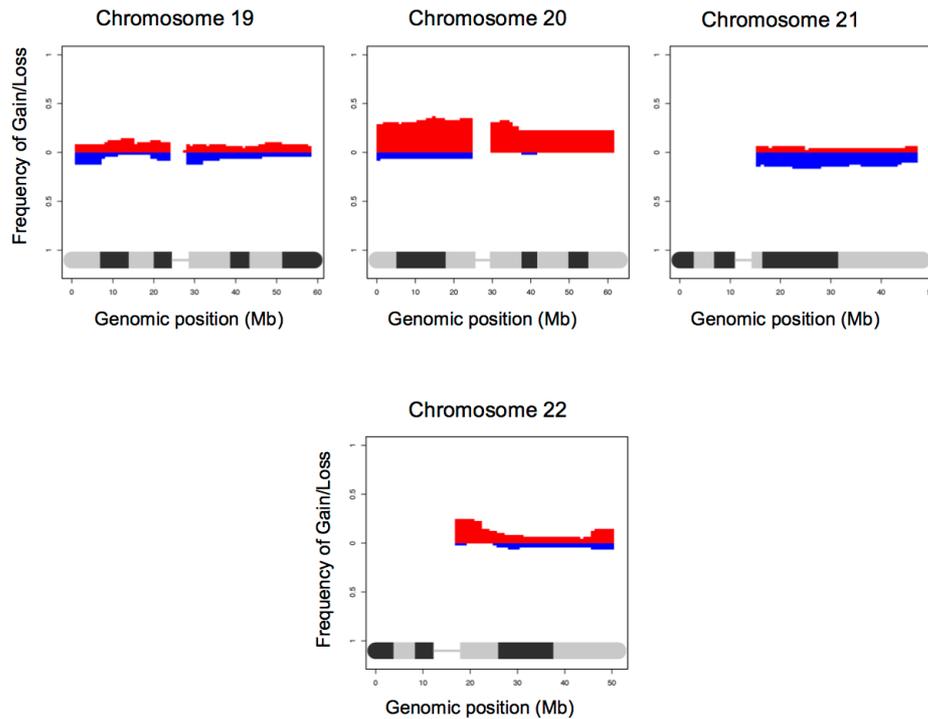
Figure 4-5: Cumulative frequency plots of each chromosome for OSCC primary tumours associated with nodal metastasis.

Visible differences included gain of 6p seen in 13/49 (26%) of metastatic primary tumours compared to 5/49 (11%) of paired metastases. Gain of 12p was seen in 15/49 (31%) of metastatic primary tumours but only in 7/49 (14%) of paired lymph node metastases. Smaller differences were also seen such as loss of the distal portion of 14q, apparent in 2/49 (5%) of primary tumours but absent in their matched metastasis.
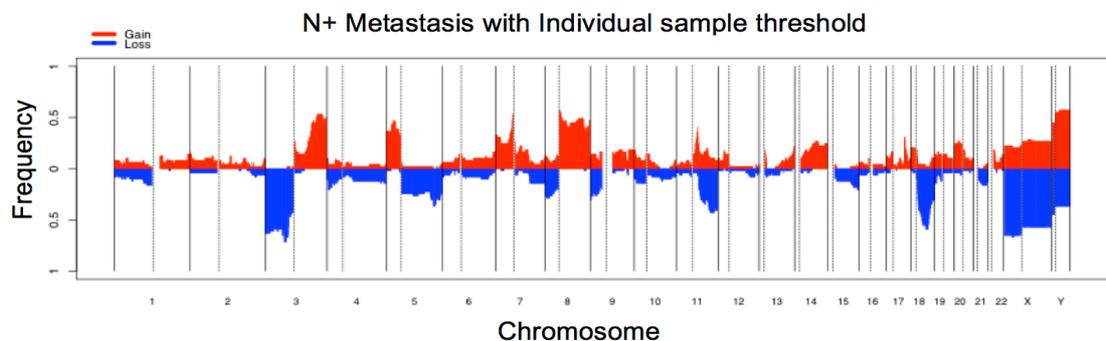
Figure 4-6: Cumulative frequency plot generated from DNA samples from a matching nodal metastasis for each primary tumour in this group.

4.2.1.4 Comparison of CNA between metastatic primary tumours and lymph node metastases

Though visually the cumulative frequency plots for these two groups were very similar a more objective measure of their similarity was needed to establish whether or not a significant genomic difference was present.

4.2.1.4.1 Concordance

Firstly concordance rates for the presence of gain or loss in each window between matched pairs of metastatic primary tumours and their nodal metastasis were calculated. This was done by assigning a binary value to the presence of gain or loss (according to the individual threshold for each sample) in each window in each sample. These were then compared. Each metastatic primary tumour was compared to every other lymph node metastasis as well as every other metastatic primary tumour (see Figure **4**-**7**).

Discrepancies between tumour and paired metastasis could often be ascribed to differences in clonal sampling. In pair ECS040 (see Figure **4**-**7**), though gain of 5p occurs along the whole arm it is not homogeneously gained in both samples. This would not affect the concordance as I have calculated it according to whether gain or loss was present and not the precise level of gain or loss. However, in chr 7 a focal segment of gain is identified in the primary tumour (represented by the black dot). This is not identified in the paired metastasis, though a genomic window of gain is present (represented by red dots). CNAnorm would require > 1 window of gain in the same region to be present in order to then be then assigned a segment (with a black dot/line). Variation like this is responsible for the < 100% concordance. In other samples this may be more obvious when genomic segments are larger.

As expected when compared with itself all samples had 100% concordance of genomic windows. Interestingly the concordance rates between paired tumour and metastasis varied widely from 90% to 10% with a mean of 51% (see Figure **4**-**8**). In no tumour-lymph node pair was 100% concordance found. This is to be expected as variation in factors such as tumour DNA content or mixed clonal sampling could affect concordance. 28/49 (57%) of

pairs had concordance of 50% or greater and only 3 samples had concordance of less than 20%.
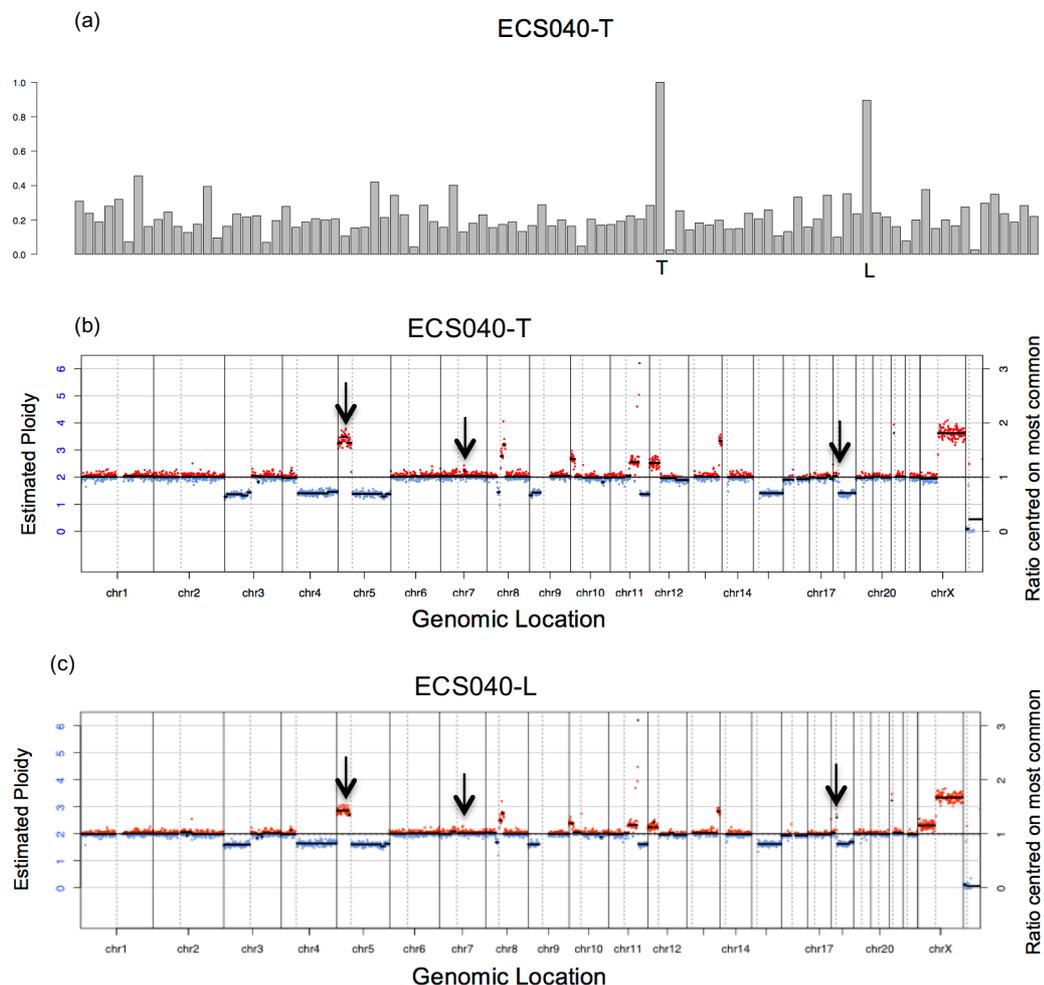
Figure 4-7: (a) Bar chart to show the concordance of copy number alterations for all metastatic primary tumours and all metastases when compared to the primary tumour ECS040. When compared to itself (T) 100% concordance is found. The paired metastasis (L) for ECS040 has a high concordance of 90%. (b) The karyogram from the ECS040 primary tumour and (c) the karyogram from its matched metastasis. The black arrows highlight the minor differences between the two karyograms.

Though tumour DNA content is attempted to be accounted for by assigning each sample a CNA call threshold, clonal heterogeneity is not. However, CNAs contained in small clonal populations may be removed from the analysis by the individual threshold if the levels of gain and loss within this population do not reach the minimum threshold. Larger sub-populations, reaching the threshold will affect concordance.
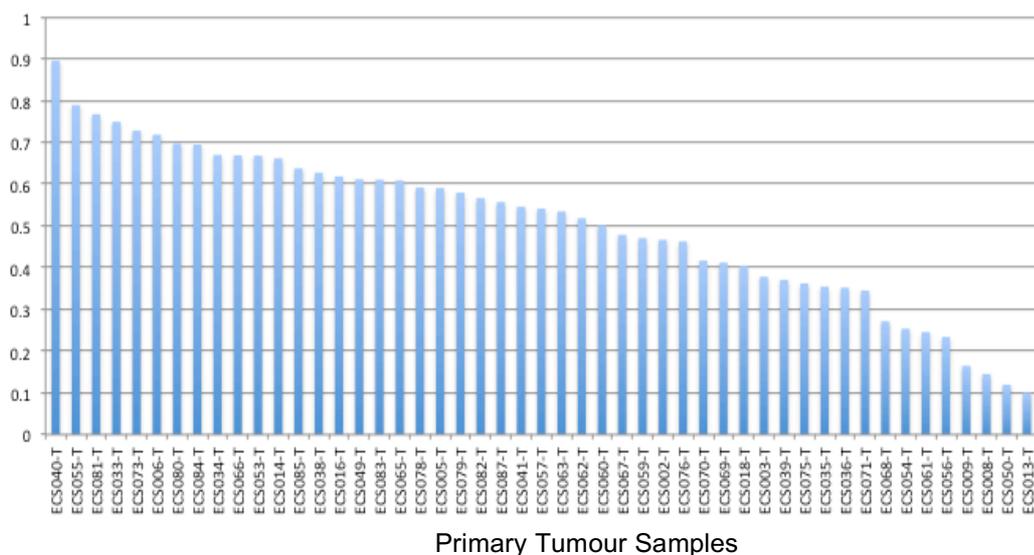
Figure 4-8: Chart of concordance of metastatic primary tumours to their paired nodal metastases.

In 37 of 49 pairs (78%) the paired metastasis had the highest concordance rate of any metastasis compared to it's paired tumour. However, in 12 cases the paired metastasis was not the most concordant sample. Of these 12 tumours, four were within 10% of being the most concordant metastasis. The number of unrelated metastases with higher concordance ranged from 1 to 48. In one case (ECS013) actually had the lowest concordance (10%) of all metastases compared (see Figure **4**-**9**). The variation seen in concordance likely reflects the fact that though certain CNAs are characteristic of SCC there is no universal CNA seen in 100% of samples.

On inspecting the individual karyograms produced such genomic differences are generally apparent. In 6 out of the 12 cases where the matched metastasis was not the most concordant sample, the paired metastasis demonstrated a greatly reduced number of CNA as in ECS013 (see Figure **4**-**9**). This could represent a sampling error due to different clonal populations being present in the metastasis. This could mean that a non-dominant clone was sampled in the metastasis or that the differing environmental selective pressures surrounding the metastasis have led to a different clonal population becoming dominant. It is also possible that the lack of CNAs is due to error in tissue dissection including an excess of non-cancerous cells leading to poor tumour DNA content. Reviewing the size of each metastasis showed that 3 out of 6 were < 5 mm in diameter, which can

be more technically demanding to dissect, however 3 of 6 were > 1 cm in diameter, including one metastasis which was > 3 cm in diameter.
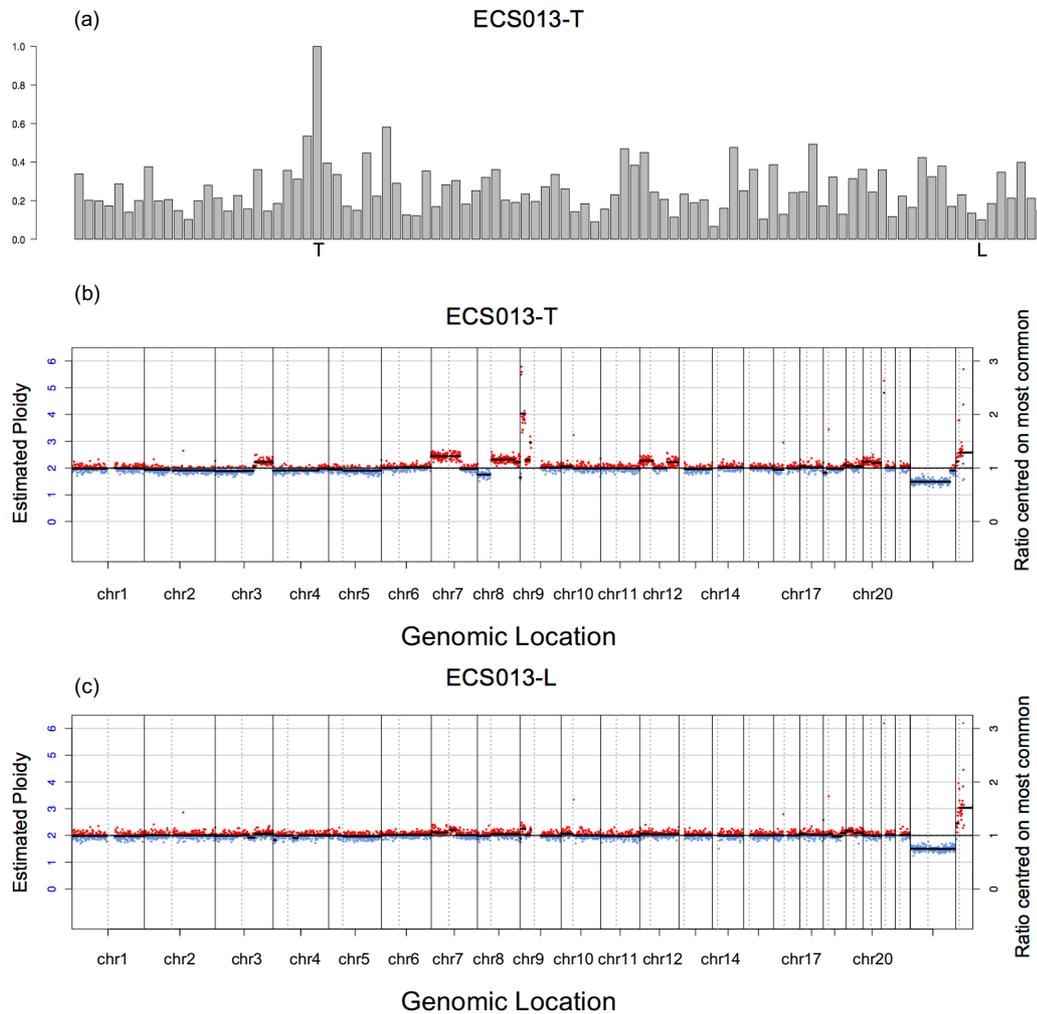
Figure 4-9: (a) Bar chart to show copy number concordance for all metastatic primary tumours and metastases when compared to ECS013 primary tumour. When compared to itself (T) concordance is 100%. The paired metastasis (L) for ECS013 actually has the lowest concordance out of all metastases. (b) Karyogram for the ECS013 primary tumour and (c) karyogram for the ECS013 metastasis. The stark differences in copy number profile are clear.

In 5 out of 12 pairs with poor concordance, the metastasis showed a similar number of CNAs though a different overall genomic profile was seen. This could also be a resulting of ongoing evolution of the clonal population either at the primary tumour or the metastasis. It is also possible that the metastatic clone is dominant in the metastasis but is not in the primary tumour, causing these marked differences in concordance.

In the remaining tumour-metastasis pair (ECS068) with poor concordance (27%) closer inspection of the individual karyograms led to a different conclusion (see Figure **4-10**). As can be seen, though the tumour DNA content is lower in the primary tumour karyogram the areas of gain and loss can still be discerned. The metastasis karyogram demonstrated a completely different CNA profile with almost no regions of chromosomal loss. This is particularly unusual for HNSCC.

An insight into how CNAnorm works is useful here. CNAnorm operates by setting the most common copy number detected as the 'normal'. In the ECS068 metastasis sample, it is apparent that the most common copy number detected is actually a loss and therefore has mistakenly 'reset' the normal baseline masking the segments in blue. The normal baseline for this was manually reset by subtracting the ratio 0.245 from all values in the ECS068-L .bed file to shift the baseline upwards. When concordance was re-calculated using these values it resulted in a concordance of 66% and the paired metastasis had the highest concordance of any sample.

It is still important to note the range in concordance rates calculated in the 37 cases where the primary tumour and paired metastasis had the highest concordance (range: 90% - 35%, mean: 57%). This could be interpreted as showing genomic changes are continually ongoing at either the metastasis or the primary tumour, or that heterogeneity of clonal populations is a factor when sampling. It is also interesting that attempting to use an objective measure of similarity between primary tumour and metastasis highlighted one case of potential error in the CNAnorm karyograms. An alternative method of objectively measuring similarity between pairs could be to use correlation of CNA.
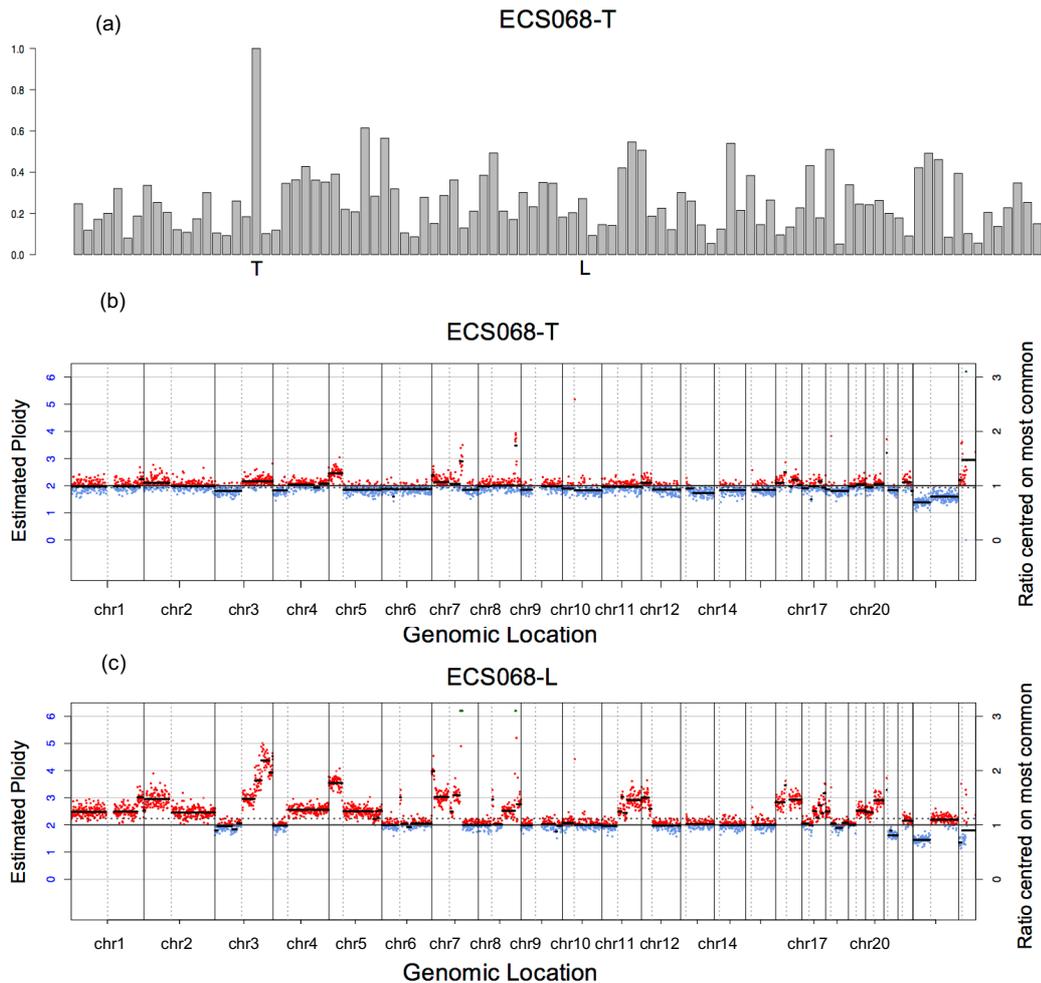
Figure 4-10: (a) Bar chart to show concordance of all metastatic primary tumours and metastases when compared to ECS068 primary tumour. Comparison to itself (T) and its matched metastasis (L) are highlighted (b) Individual karyogram for ECS068 primary tumour and (c) ECS068 paired metastasis.

4.2.1.4.2 Correlation

The correlation of CNA between samples can be calculated by assigning a value to each segment for the presence or absence of CNA for each sample (i.e. -1, 0, 1). This approach was chosen over giving a quantitative value for each CNA was in order to try and reduce noise in the results as well as the fact that the level of gain or loss was not deemed as important as the simple presence of gain or loss in terms of CNA profile. The correlation between the presence or absence of gain or loss can be calculated using Pearson's product moment correlation coefficient (r).

The data produced agreed closely with the concordance data above. In three pairs the matched metastasis was found to have closest CNA

correlation whilst it did not have the highest concordance. In these the rates of CNA on either the primary tumour or the metastasis were significantly reduced. Though visual similarities can be seen on their individual karyograms the differences are clear (see Figure **4**-**11**). It is also worth noting that in these cases the matched metastasis was only marginally more closely correlated to the primary tumour than the next closest metastasis.

It is of interest that the ECS068 tumour-metastasis pair discussed earlier, potentially highlighted a flaw in the CNAnorm programme, was found to have closest correlation in complete contrast to their concordance. This is likely due to the assignation of gain or loss to each segment is related to its position relative to the previous segment and therefore the changes in segment relative to one another are detected rather than the absolute position as assigned by CNAnorm. This may also be the reason why correlation figures appear higher in many cases compared to the concordance rate. This could suggest that, as an isolated measure, correlation may provide a better indication of genomic similarity, rather than concordance.

Otherwise the same tumour-metastasis pairs were found to have closest correlation as with concordance rates. Correlation therefore acted as a second objective technique of comparing how genomically similar primary tumour and matched metastasis were. Advantages and disadvantages of both approaches exist and it seems sensible to suggest using correlation in combination with visual inspection of karyograms will provide a reliable indication of genomic CNA differences between primary tumour and metastasis. Concordance on the other hand allows for genomic regions of high or low concordance to be identified. Ultimately, the majority of tumour-metastasis pairs were genomically similar to each other, regardless of the comparison method.
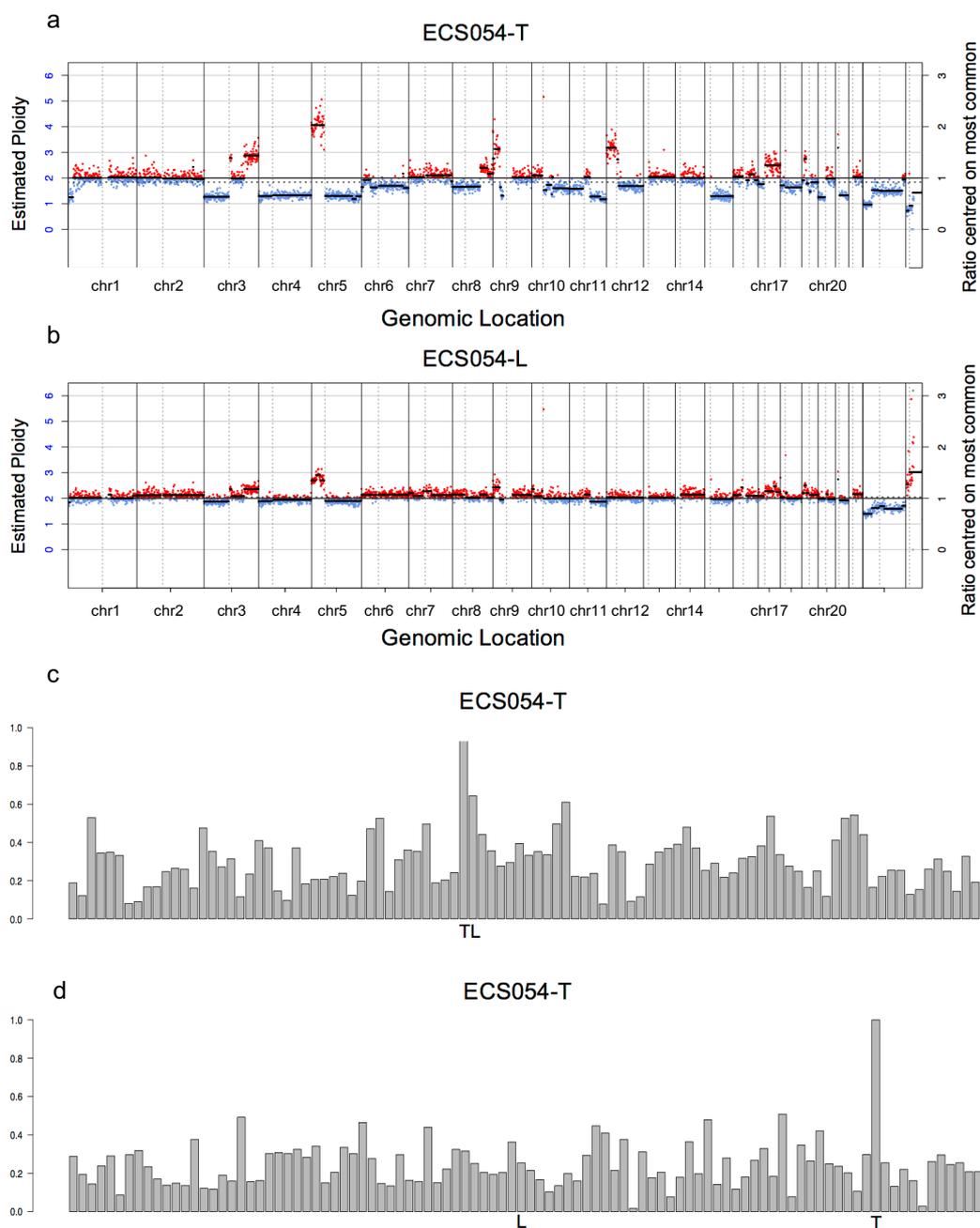
Figure 4-11: (a) Individual karyogram for primary tumour of ECS054 and (b) its matched metastasis. (c) Bar chart comparing correlation of CNA in ECS054 primary tumour (T) to all other tumours and metastases. The closest correlate is the matched metastasis (L). (d) Bar chart comparing concordance rates of CNA in ECS054 primary tumour to all other tumours and metastases. Here the matched metastasis (L) has a poor concordance.

4.2.1.4.3 Use of GISTIC 2.0 for focal analysis

CNAs in cancer are frequently found to encompass the entire length of a chromosomal arm (Zack *et al.*, 2013, Beroukhim *et al.*, 2010). The samples included in this study concur with this finding as seen in the individual and cumulative frequency karyograms above. These arm-level events GISTIC

2.0 is an open access algorithm developed by the Broad Institute that provides a systematic method for analysing chromosomal aberrations in cancer (Mermel *et al.*, 2011). It provides accurate identification of the copy number profile in each sample, providing an indication of significant focal copy number aberrations across a group of samples. It analyses the frequency and amplitude of the CNA as well as assesses the statistical significance of the CNA assigning a q-value to each. GISTIC has been used across numerous cancer studies providing an objective method of analysing samples correlating CNA with clinical outcome and identifying driver CNAs and potentially the driver genes within these regions (Zack *et al.*, 2013, Beroukhim *et al.*, 2010, Bass *et al.*, 2009, Frankel *et al.*, 2014, Xie *et al.*, 2012).

Though cumulative frequency karyograms can be produced at good resolution with CNAnorm files high frequency broad events can mask the lower frequency focal event on the images produced. GISTIC simultaneously runs a focal and broad analysis separately removing any potential masking of focal CNAs. GISTIC also provides lists of amplified and deleted genes within the altered regions as well as heat maps of the copy number data. The GISTIC parameters used are in Chapter 2.9.2.3.

Firstly, an analysis was performed to identify focal chromosomal events. The focal amplification and deletion profiles of the nodal metastases paired to the metastatic primary tumours revealed a strong similarity. It was still important to regard event occurring close to centromeres and telomeres as potentially spurious.

In terms of focal amplifications (see Figure **4**-**12**). The most common shared amplification was gain of 8q24.13, identified in 38/49 (78%) of metastatic primary tumours and 24/49 (49%) of nodal metastases (see Table **4**-**2**). This is closely followed by similar gains of 3q26.2 (present in 57% of metastatic primary tumours and 53% of nodal metastases), 5p14.3, 11q13.3 and 7p11.2.
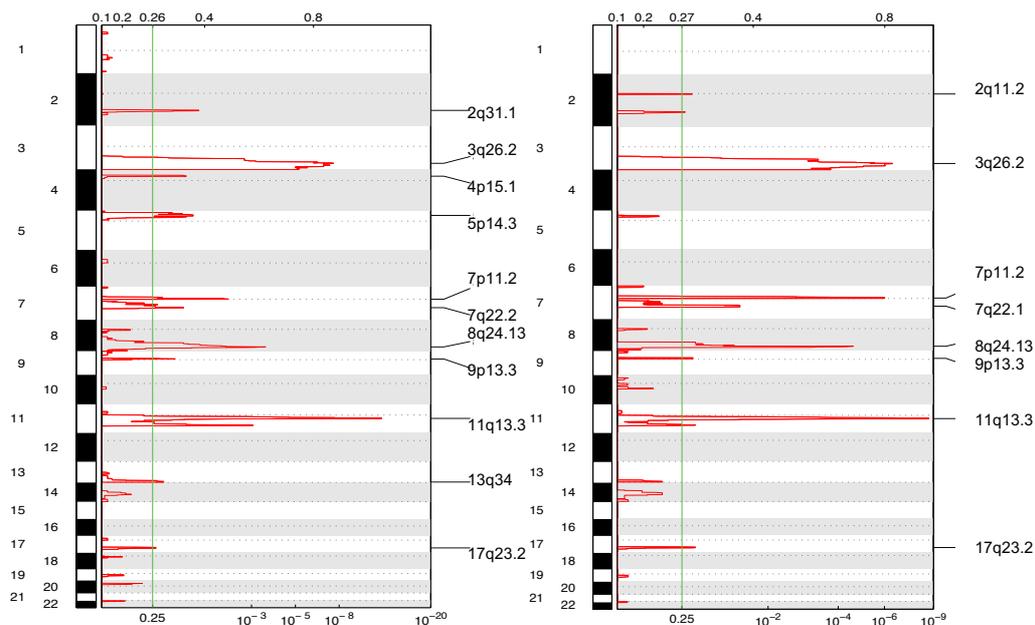
Figure 4-12: Genome-wide focal amplification plots for all metastatic primary tumours (left) and their matched nodal metastases (right). The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line marking the significance threshold (0.25). The top x-axis shows the G-score for the CNAs calculated by GISTIC.

Slight variation was present between groups, in the focal CNA cytobands produced by GISTIC, for example: gain of 7q22.2 reported in primary tumours but gain of 7q22.1 for nodal metastases. The wide peak boundaries, also provided by GISTIC, for these amplifications showed they were very closely overlapping. This was confirmed on inspection of the cumulative frequency plots.

Amplifications			Deletions		
Cytoband	Number of samples in each group (%)		Cytoband	Number of samples in each group (%)	
	Metastatic Primary Tumours (n = 49)	Nodal Metastases (n = 49)		Metastatic Primary Tumours (n = 49)	Nodal Metastases (n = 49)
8q24.13	38 (78)	24 (49)	3p14.2	38 (78)	35 (71)
3q26.2	28 (57)	26 (53)	3p26.1	34 (69)	31 (63)
5p14.3	26(53)	23 (47)	18q21.1	24 (49)	29 (59)
11q13.3	23 (47)	22 (45)	5q33.1	19 (39)	18 (37)
7p11.2	20 (41)	26 (53)	11q24.2	17 (35)	22 (45)
17q23.2	17 (35)	15 (31)	11q21	17 (35)	22 (45)
9p13.3	13 (27)	9 (18)	9p24.3	16 (33)	15 (31)
13q34	12 (24)	9 (18)	8p23.2	14 (29)	14 (29)
7q22.2	11 (22)	9 (18)	4p14	11 (22)	7 (14)
2q11.2	8 (16)	4 (8)	15q26.2	8 (16)	10 (20)
2q31.1	7 (14)	5 (10)	7q36.1	7 (14)	7 (14)
4p15.1	6 (12)	5 (10)	2q34	6 (12)	4 (8)
			19p12	4 (8)	6 (12)

Table 4-2: Frequency of focal CNAs identified by GISTIC 2.0.

Amplifications identified by GISTIC as present in the metastatic primary tumours, but absent in the nodal metastases included gain of 2q31.1, 4p15.1, 5p14.3. When the cumulative frequency karyograms were inspected these CNAs were observed in both groups at similar frequencies (see Table 4-2). Gain of 2q31.1 was present in 7/49 (14%) metastatic primary tumours and 5/49 (10%) nodal metastases. Gain of 4p15.1 was present in 6/49 (12%) metastatic primary tumours and 5/49 (10%) nodal metastases. These were present at low frequencies in both groups, but marginally lower in the nodal metastases, which is likely the reason GISTIC did not identify them as focal CNAs in this group.

Gain of 5p14.3 was present in 26/49 (53%) metastatic primary tumours and 23/49 (47%) of nodal metastases. Despite this much higher and similar frequency in both groups GISTIC only identified this as a focal CNA in the primary tumours. On inspecting the individual karyograms, it is clear that amplifications on 5p occur on a broad level (> half the length of chromosome) in both groups. In the nodal metastases 16/23 amplification on 5p occur across the entire arm, thus reducing the number in which a focal alteration can be identified. The nine focal amplifications common to both

groups, tended to occur at similar frequencies. Gain of 8q24.13 was present in 78% of metastatic primary tumours and 49% of nodal metastases. Seven of the remaining amplifications occur with 8% or less difference between groups. Only gain of 7p11.2 occurred with a higher frequency in the nodal metastases (53% vs. 40% in the metastatic primary tumours).
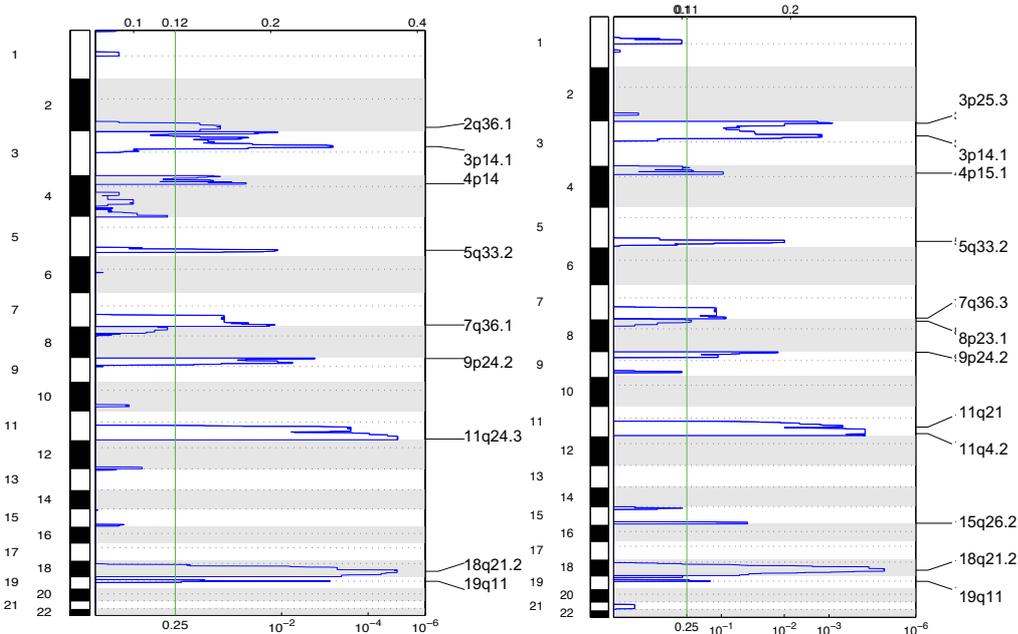
Figure 4-13: Genome-wide focal deletion plots for all metastatic primary tumours (left) and their matched nodal metastases (right). The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line marking the significance threshold (0.25). The top x-axis shows the G-score calculated by GISTIC.

In terms of focal deletion profiles (see Figure 4-13) the primary tumour and their matched metastases are again very similar. The most common shared focal deletion is loss of 3p14.1, occurring in 38/49 (78%) of primary tumours and 35/49 (71%) of nodal metastases. This is followed by loss of 18q21.1 in 24/49 (49%) of primary tumours and 29/49 (59%) of nodal metastases. There is also variation in the focal CNAs reported by GISTIC between groups such as loss of 11q24.3 is present in the metastatic primary tumours, but loss of 11q24.2 in the nodal metastases.

These regions were examined more closely using the individual chromosome cumulative frequency plots created using seg_compare (see Figure 4-14). The close proximity of these region were confirmed using the wide peak boundaries provided by GISTIC revealing these two regions are

overlapping (see Table **4**-**3**). This suggests that the same genes may be lost, though the actual region varies slightly in its co-ordinates. Similarly loss of 4p14 is reported by GISTIC in the primary tumours but loss of 4p15.1 in the nodal metastases. This variation may be due to variation in the exact breakpoints between samples. Therefore slight variation in the exact co-ordinates of overlapping regions in separate groups can occur. The generic CNA call threshold utilised by GISTIC means that alterations in some samples may not be recognised affecting the precise assignation of CNA.
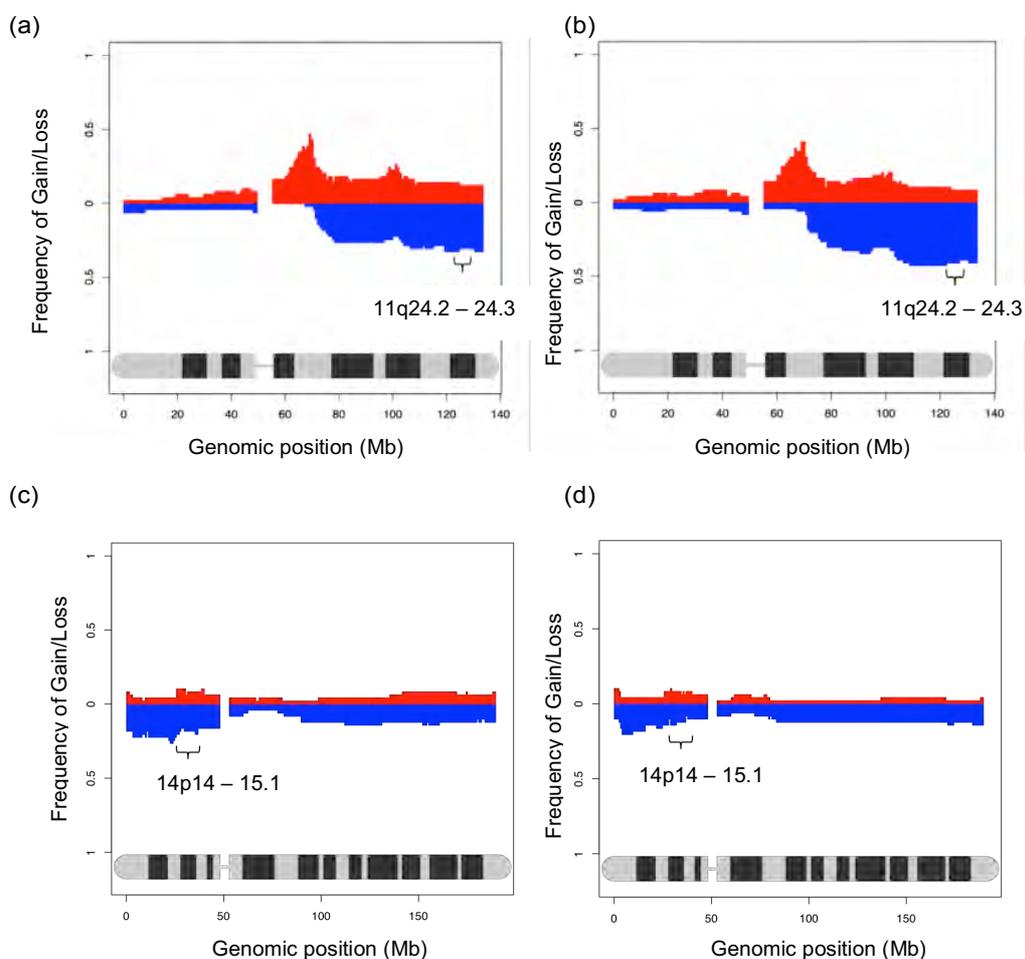
Figure 4-14: Cumulative frequency karyograms for (a) chr 11 in metastatic primary tumours and (b) in matched nodal metastases (c) chr 4 in metastatic primary tumours and (d) in matched nodal metastases.

CNA Region	Wide peak boundaries
11q21	chr11:68000002-135006516
11q24.3	chr11:109600002-135006516
4p15.1	chr4:26400002-38400000
4p14	chr4:30400002-38400000

Table 4-3: Wide peak boundaries of adjacent focal CNAs reveal them to be in fact overlapping rather than adjacent.

The nodal metastases contained four deletions not reported in the metastatic primary tumours by GISTIC. These were again all identified on the inspection of the individual karyograms in both groups at similar frequencies. These included loss of 3p26.1 (in 34/49 (69%) primary tumours and 31/49 (63%) metastases), loss of 8p23.2 (in 14/49 (29%) primary tumours and 14/49 (29%) nodal metastases), loss of 15q26.2 (in 8/49 (16%) primary tumours and 10/49 (20%) nodal metastases) and loss of 2q34 (in 6/49 (12%) primary tumours and 4/49 (8%) nodal metastases). Inspection of the individual karyograms revealed 6/8 losses on 15q included the entire arm, whilst 28/34 losses of 3p and 10/14 losses on 8p also included the entire arm. The high frequency of these broad-level losses may have reduced the sensitivity of GISTIC in identifying focal CNAs in both groups. The generic CNA call threshold used by GISTIC is also a confounding factor when comparing it to CNAnorm karyograms produced using individual sample thresholds.

Again there is a similar rate of focal deletion in both groups (see Table **4-2**) with the nine common deletions having a difference of 10% or less in the proportion of samples the CNA is present. Four of the common deletions were more common in nodal metastases, and one was observed at identical rates in both groups.

Overall it is of interest that the frequency with which focal deletions are observed similar to focal amplifications (see Table **4-2**). In the metastatic primary tumours focal amplifications occur in from 6–38 patients (78%–12%, mean: 36%) whilst in nodal metastases they occur in 5–24 patients (49%–10%, mean: 30%). Focal deletion occur in 4–38 metastatic primary tumours

(78%–8%, mean: 34%), whilst in the nodal metastases they occur in 6 - 35 patients (71%–12%, mean: 35%).

In both groups the majority of focal CNAs are seen in the minority of samples (see Table **4-2**). In terms of focal amplifications only 3/12 (gain of 8q24.13, 3q26.2 and 5p14.3) are present in more than 50% of metastatic primary tumours. Only gain of 8q24.13 and 3q26.2 are present in more than 50% of nodal metastases. Similarly only two focal deletions were present in more than 50% of metastatic primary tumours (loss of 3p14.2 and 3p26.1) and only three present in more than 50% of nodal metastases (loss of 3p14.2, 3p26.1 and 18q21.2).

4.2.1.4.4 GISTIC 2.0 Broad-level analysis

GISTIC generates genome-wide heat maps of CNA on a broad level (see Figure **4-15**).

(a)   (b)
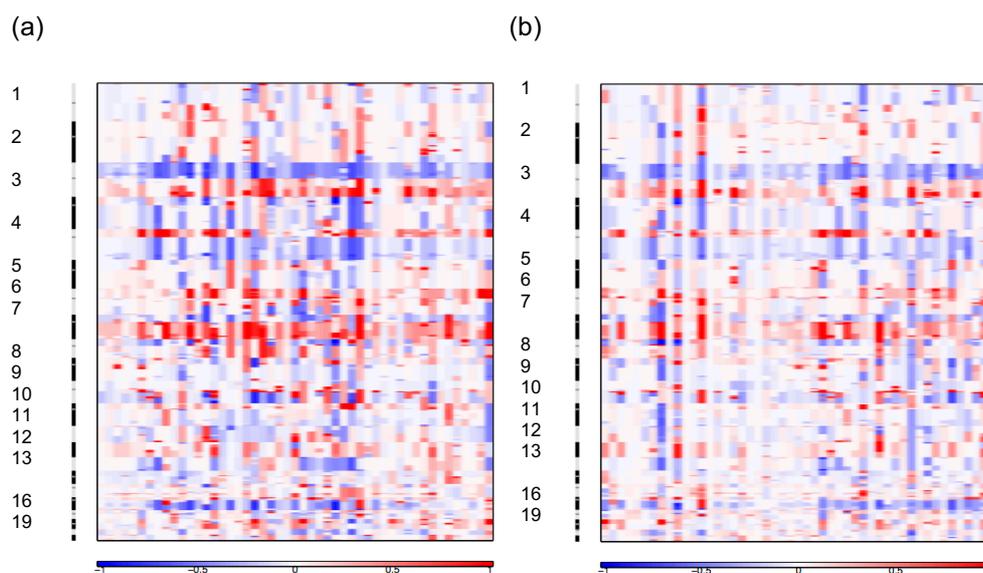
Figure 4-15: GISTIC heat maps for regions of chromosomal gain (red) and loss (blue) are shown above for (a) metastatic primary tumours and (b) their matched nodal metastases. The intensity of colour indicates higher or lower relative CNA.

The GISTIC heat maps are analogous to the cumulative frequency karyograms produced using CNAnorm. They offer a group-wide view and a low resolution individual sample view simultaneously. GISTIC also generates raw data giving the frequency of each arm-level CNA with an associated z-score and q-value (see Table **4-4**). Using a 95% confidence interval, z-scores

outside -1.96 to 1.96 are considered significant. The q-value threshold was set at <0.25, in relation to the false discovery rate due to background chromosomal alterations. GISTIC assigns each of these values to the frequency of arm-level changes to allow identification of potentially key CNAs.

Amplifications						
	Metastatic primary tumours			Nodal Metastases		
Chr Arm	Frequency	z-score	q-value	Frequency	z-score	q-value
3q	0.45	3.44	0.00528	0.24	1.47	0.308
5p	0.5	2.34	0.0628	0.39	2.52	0.0763
7p	0.5	3.35	0.00528	0.26	1.17	0.474
8p	0.33	0.527	0.583	0.34	2.07	0.126
8q	0.64	5.73	1.92E-07	0.49	5.42	1.16E-06
9q	0.36	2.03	0.119	0.2	0.947	0.609
19p	0.29	0.785	0.496	0.29	2.15	0.123
19q	0.26	1.83	0.145	0.23	2.72	0.0644
20p	0.5	2.62	0.0345	0.36	2.25	0.118
20q	0.45	2.75	0.0293	0.29	1.94	0.147
Amplifications						
	Metastatic primary tumours			Nodal metastases		
Chr Arm	Frequency	z-score	q-value	Frequency	z-score	q-value
3p	0.76	8.13	8.66E-15	0.7	10	0
3q	0.29	0.897	0.655	0.39	4.28	0.00012
5q	0.46	4.49	6.98E-05	0.36	4.84	1.24E-05
8p	0.49	2.72	0.0426	0.45	3.86	0.00056
9p	0.39	1.04	0.577	0.38	2.69	0.0247
10p	0.27	-0.578	0.999	0.34	2.08	0.0911
10q	0.16	-0.865	0.999	0.24	2.02	0.0946
15q	0.33	2.03	0.167	0.21	1.62	0.188
18p	0.36	0.266	0.999	0.37	2.12	0.0911
18q	0.43	1.67	0.306	0.37	2.67	0.0247
19q	0.28	2.18	0.142	0.13	0.521	0.629

Table 4-4: List of broad level CNAs identified as significant by GISTIC (z-score outside -1.96-1.96 and q-value < 0.25) highlighted in red.

The frequency of all arm-level CNAs identified by GISTIC is not identical to the rates at which they are identified using the CNAnorm karyograms. This is

due to the fact that GISTIC is unable to accept an individual CNA-calling threshold for each sample and can only assign a generic group-wide threshold. This could potentially include a lot of background "noise", whilst if too high may exclude actual CNAs, and is therefore an important limitation of using GISTIC 2.0.

Interestingly the GISTIC heat maps show similar trends of CNA to the cumulative frequency plots (see Figure **4-4** and Figure **4-6**). The most frequent CNAs were loss of 3p (37/49 (76%) of metastatic primary tumours and 34/49 (70%) of metastases) and gain of 8q (in 31 (64%) of metastatic primary tumours and 24/49 (49%) of metastases). Gain of 5p was present in 25 (50%) of primary tumours and 19/49 (39%) of metastases.

Gain of 11q was identified in 9/49 (19%) of primary tumours and 5/49 (10%) of metastases. In neither of these groups was this amplification found to be at a significant level (q = 0.989). This is a much lower frequency than found using the CNAnorm karyograms (43% in primary tumours). The inconsistency may be due to the fact that GISTIC utilises a generic group-wide threshold, rather than an individual sample threshold, making it less sensitive, to samples with variable tumour DNA content. Gain of 20p was present in 25/49 (50%) of metastatic primary tumours analysed by GISTIC but 18/49 (37%) of tumours analysed using CNAnorm karyograms. This, is still likely due to variation produced by not using an individual sample CNA-call threshold.

Though the frequency of CNAs was similar for the majority of CNAs in primary tumours and metastases their significance level varied. 9/11 Deletions occurred at rates within 10% of each other (see Table **4-4**). Despite this six of them were only found at a level of significance $\quad$ (q < 0.25) in one group. Loss of 18q was present in 18/49 (37%) of nodal metastases (q = 0.00247). Though present at a higher frequency in primary tumours (21/49 (43%)), this was not found to be significant (q = 0.306). This discrepancy may simply reflect that though a deletion on 18q was present in 25/49 (50%) of the individual karyograms, in 13/25 the deletion is focal, not covering the whole arm of the chromosome.

Broad amplifications appear to have more variation between primary tumours and metastases, for example gain of 3q is present in 22/49 (45%) of primary tumours and 12/49 (24%) of nodal metastases. This drop in frequency is likely the cause for it not to be found at a significant level in nodal metastases (q = 0.308). This may reflect the differences due to CNA-call thresholds. However examination of the karyogram reveals gain of part of 3q to be present in 26/49 (53%) of all samples. In 19/26 primary tumours the amplification includes > 50% of the 3q arm whilst in only 7/26 metastases does it cover > 50% of the arm. These differences may also be due to clonal heterogeneity between primary tumour and metastasis.

GISTIC broad analyses are similar to the focal analysis in that the majority of broad CNAs are again seen in the minority of samples. Only 4/10 amplifications are present in ≥ 50% of primary tumours and none in the nodal metastases. Only 1/11 deletions is identified in the primary tumours or nodal metastases. This highlights the inter-tumour heterogeneity amongst HNSCC. An advantage of the heat maps is that they highlight the heterogeneity of CNA. Regions such as 2q are found to have a low frequency of CNA yet a small number of samples have a significant loss. As such GISTIC broad analyses have limitations and need to be interpreted with the value of individual karyograms. All broad CNAs of significance are identifiable on the CNAnorm cumulative frequency plots but GISTIC does provide an estimation of their significance with relation to background CNAs, which is valuable.

4.2.2 Results 2: DNA copy number analysis of non-metastatic OSCC primary tumours

4.2.2.1 Individual karyograms for non-metastatic primary tumours

A total of 26 patients with OSCC were included in this group. All patients underwent surgical resection of both their primary tumour and regional cervical lymph nodes as treatment. The specimens were all examined under standard protocol in Leeds Histopathology department and confirmed to have no lymph node metastases. Tissue blocks from the primary tumours were obtained and the diagnosis confirmed by Professor Kenneth MacLennan or Dr Preetha Chengot (consultant head and neck pathologists).

DNA extraction was carried out and then processed to copy number sequencing libraries as described previously. The samples were multiplexed on the Illumina HiSeq 2500 and the resulting data processed to an individual digital karyogram for each (see Figure **4**-**16**).
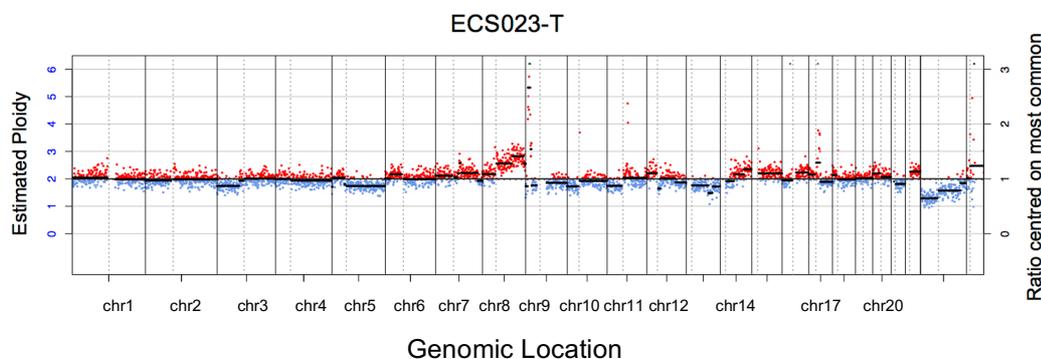
Figure 4-16: An example of an individual karyogram produced from OSCC non-metastatic primary tumour.

In general the karyograms generated from non-metastatic primary tumours showed a different pattern to those from metastatic primary tumours. Ten tumours showed very few, if any CNAs on their karyogram. The remainder demonstrated a lower degree of genomic damage, on visual inspection, compared to those from metastatic primary tumours. Though classical CNAs associated with both HNSCC and lung SCC such as loss of 3p and 5q were still seen this appeared to be in a much lower proportion of samples than in the nodal metastases. Group-wide analysis is easier using cumulative frequency plots.

4.2.2.2 Cumulative frequency plots for non-metastatic primary tumours and comparison to metastatic primary tumours

Cumulative frequency plots were created using seg_compare (Dr Henry Wood, Precancer genomics group). These highlighted the differences in genomic CNAs between non-metastatic primary tumours and the metastatic groups (see Figure **4**-**17**).

Though loss of 3p was seen in non-metastatic primary tumours, it was only apparent in 10/26 (39%) compared to 39/49 (80%) of metastatic primary tumours (using visual inspection of CNAnorm images). Similarly reduced rates of CNA were seen in other typical areas associated with SCC, such as

gain of 3q and 5p and loss of 5q. More details could be discerned from cumulative individual chromosomes (see Figure **4-18**).
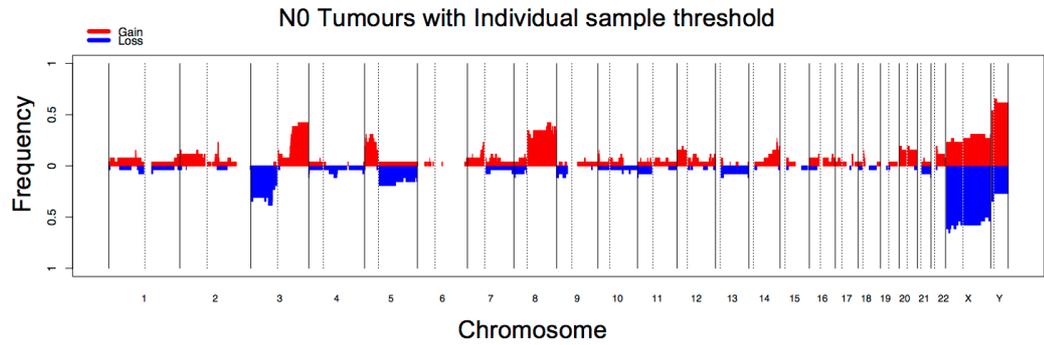
Figure 4-17: Cumulative frequency plot including all non-metastatic OSCC primary tumours.
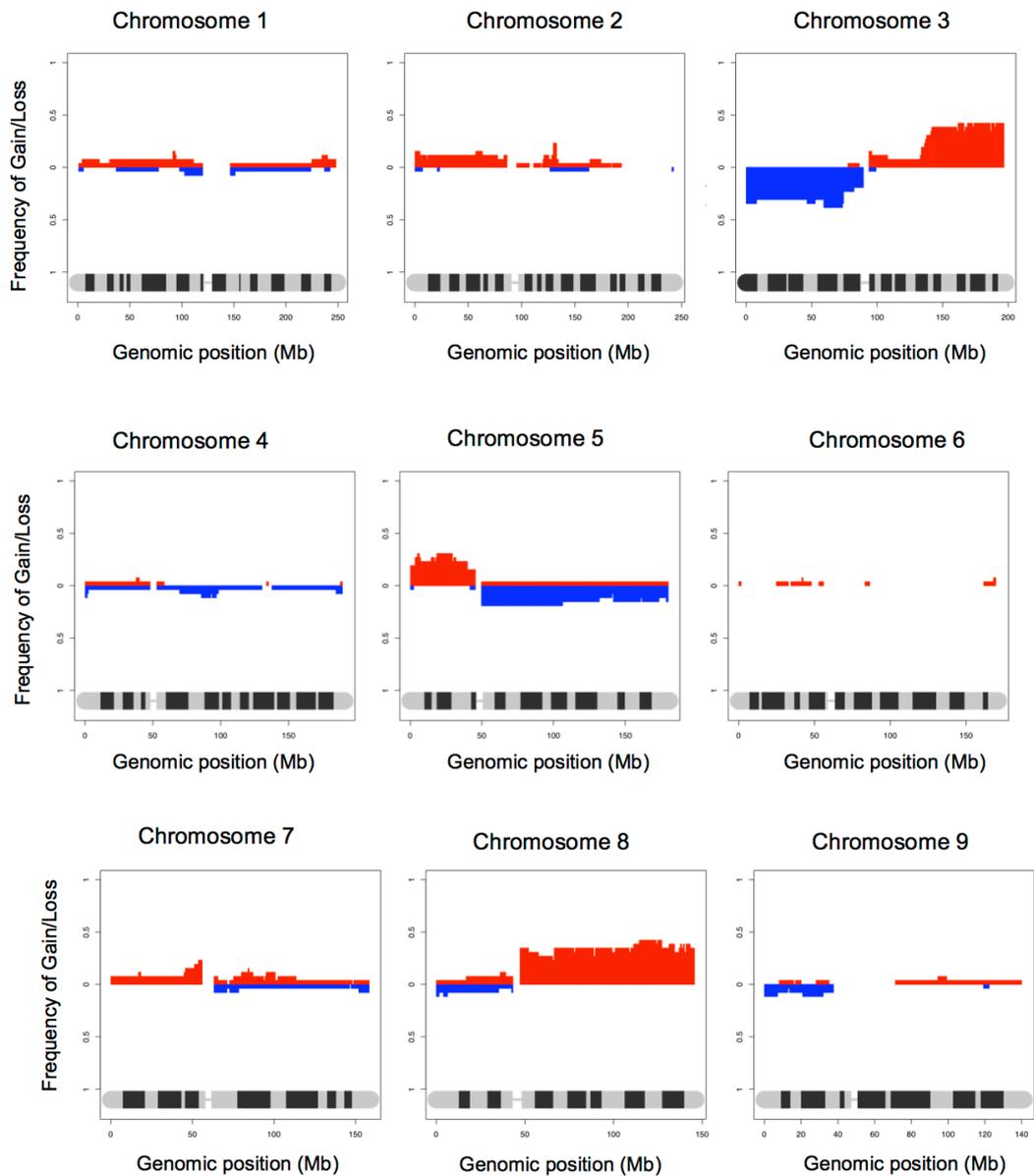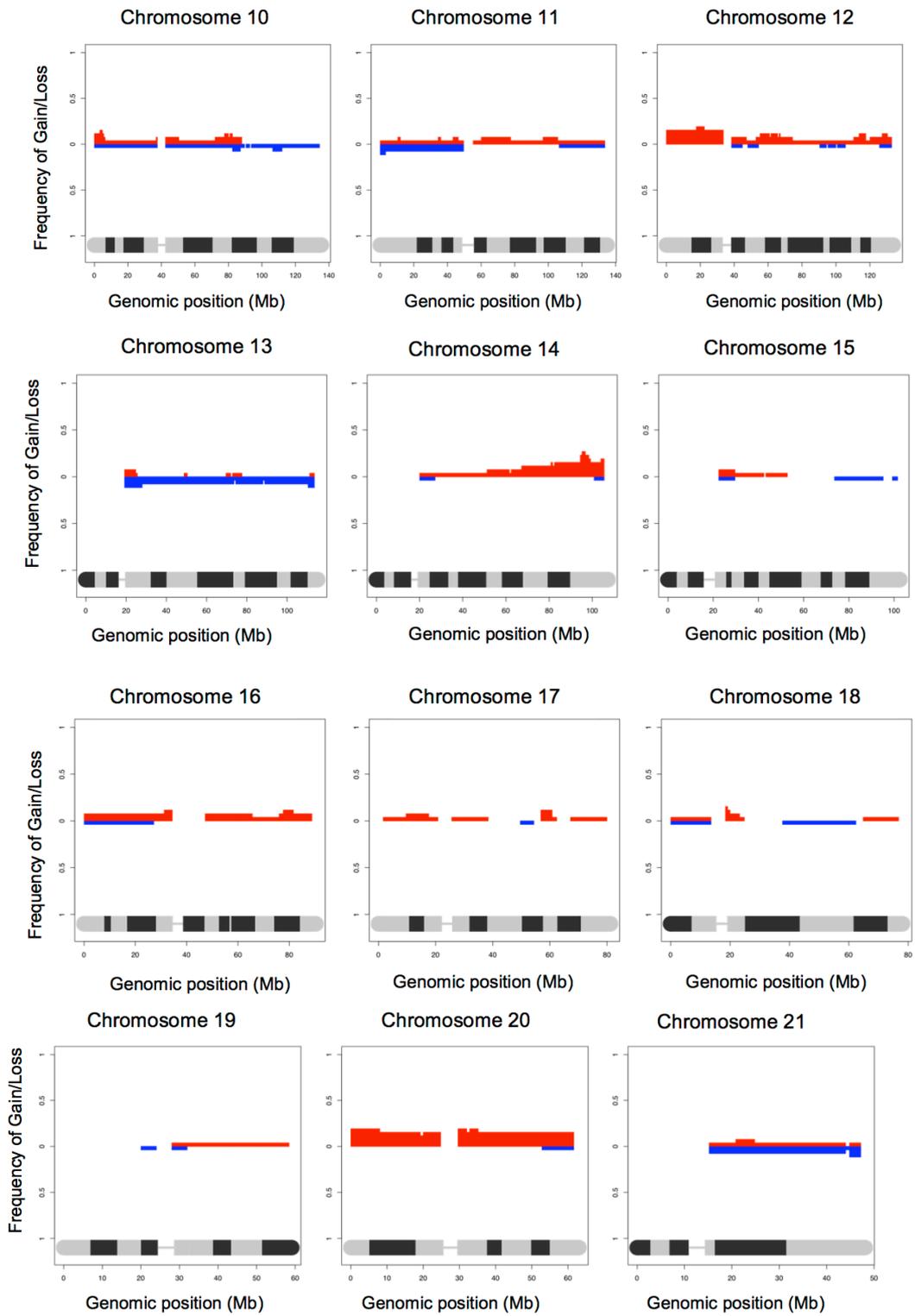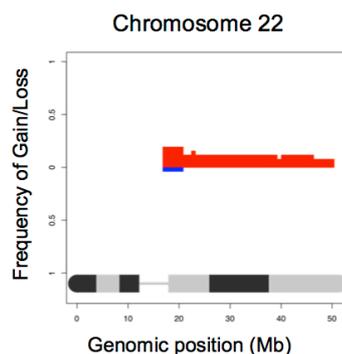
Figure 4-18: Individual cumulative chromosome plots for non-metastatic primary tumours.

Though there were similarities in the regions of CNAs seen on the cumulative karyograms when compared to the metastatic primary tumours there were marked differences in the frequency. Loss of 3p was found in 9/26 (35%) of non-metastatic primary tumours compared to 38/49 (78%) of metastatic primary tumours. Gain of 3q was present in 10/26 (38%) compared to 30/49 (61%) of metastatic primary tumours. Gain of 8q was present in 10/26 (38%) compared to 38/49 (78%) of metastatic primary tumours. Gain of 5p was present in 8/26 (31%) compared to 29/49 (60%) of metastatic primary tumours. Loss of 5q was present in 4/26 (15%) compared to 18/49 (37%) of metastatic primary tumours.

Other CNA differences included loss of 18q, identified in 50% of metastatic primary tumours was only present in 1/26 (4%) of non-metastatic primary tumours and gain of 7p, which was found in 6/26 (23%) of non-metastatic primary tumours but 20/49 (40%) of metastatic primary tumours.

The cumulative frequency plots demonstrated an overall reduced number of CNAs in the non-metastatic tumours compared to the metastatic tumours. They also showed that the CNAs were not uniform in their length and therefore a focal analysis could be more revealing.

4.2.2.3 GISTIC focal analysis of non-metastatic primary tumours

Surprisingly, no significant focal amplifications were identified by GISTIC (see Figure **4**-**19**) in this group. Though numerous events were observed to occur in up to 40% of samples these were not shown to be significantly aberrant when compared to the background rate of alterations (i.e. they did not reach the q-value threshold of 0.25).
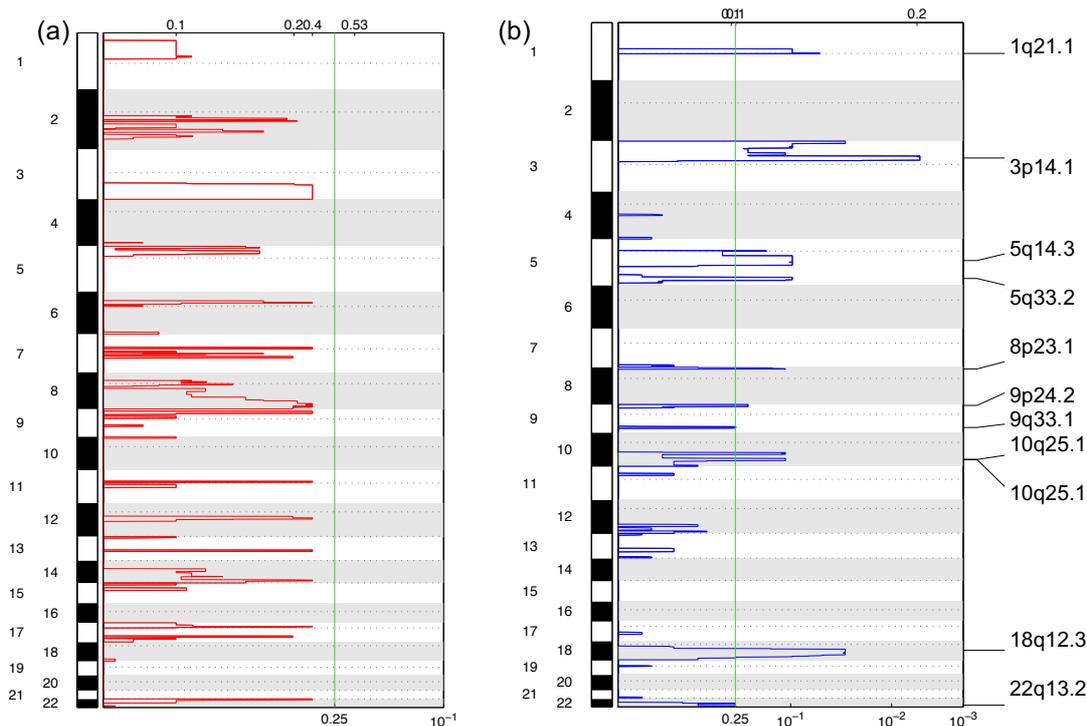
Figure 4-19: Genome-wide focal amplification (red, (a)) and deletion (blue, (b)) plots for non-metastatic tumours. The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line marking the significance threshold (0.25). The top x-axis shows the G-score calculated by GISTIC.

In terms of focal deletions, in general the frequency across the group was much lower than the metastatic samples (see Table **4-5**). The most common deletion present was loss of 3p14.3, in 10/26 (39%) non-metastatic primary tumours, compared to 38/49 (78%) of metastatic primary tumours. This was followed by loss of 5q32 in 4/26 (15%) non-metastatic tumour. The wide peak boundaries for this region reveal it to be present in 19/49 (39%) of metastatic tumours. Loss of 8p23.2 was present in 4/26 (15%) non-metastatic primary tumours and 14/49 (29%) of metastatic primary tumours. Loss of 9p24.3 was present in 3/26 (12%) non-metastatic tumours and 16/49 (33%) of metastatic primary tumours.

The wide peak boundaries for loss of 18q12.2 revealed that this deletion overlapped between metastatic and non-metastatic primary tumours. However it was only present in 1/26 (4%) non-metastatic primary tumours, compared to 24/49 (49%) of metastatic primary tumours.

Cytoband	Frequency in N0 Tumours n = 26 (%)
3p14.3	10 (39)
5q13.1	5 (19)
5q32	4 (15)
8p23.2	4 (15)
1p12	3 (12)
9p24.3	3 (12)
10q24.32	2 (12)
18q12.2	1 (4)
9q33.1	1 (4)

Table 4-5: Frequency of focal deletions identified as reaching a level of significance in non-metastatic primary tumours (q = < 0.25).

Four focal amplifications were identified by GISTIC as occurring at a significant level (q = < 0.25) across the non-metastatic primary tumours, which were not identified in the metastatic primary tumours or metastases. These included loss of 5q13.1, 1p12, 10q24.32 and 9q33.1. All of these were present in 3 or less non-metastatic primary tumours.

Of note, GISTIC identified loss of 22q13.1 as a focal deletion just achieving significance (q = 0.248) in non-metastatic primary tumours. However on inspection of the CNAnorm karyograms no deletion is present in this region (see Figure **4**-**20**). This is likely due to the generic CNA-call threshold that GISTIC uses, rather than an individual sample threshold. By generating the cumulative frequency plots using a generic threshold the deletion at 22q13.1 is visible (see Figure **4**-**20**), demonstrating it is a spurious finding.

The absence of focal amplifications and low incidence of focal deletions suggests that these events are uncommon in non-metastatic tumours. As GISTIC calculates whether these events are likely to be significant in comparison to the background CNA rate, it could reflect the fact that a 10/26 (38%) non-metastatic primary tumours have few or no CNAs. It could however reflect the lower number of samples included in this group. A larger number of samples could reveal more significant CNAs occurring at a low frequency in these patients.

(a) Chr 22, Generic Threshold

(b) Chr 22, Individual Threshold
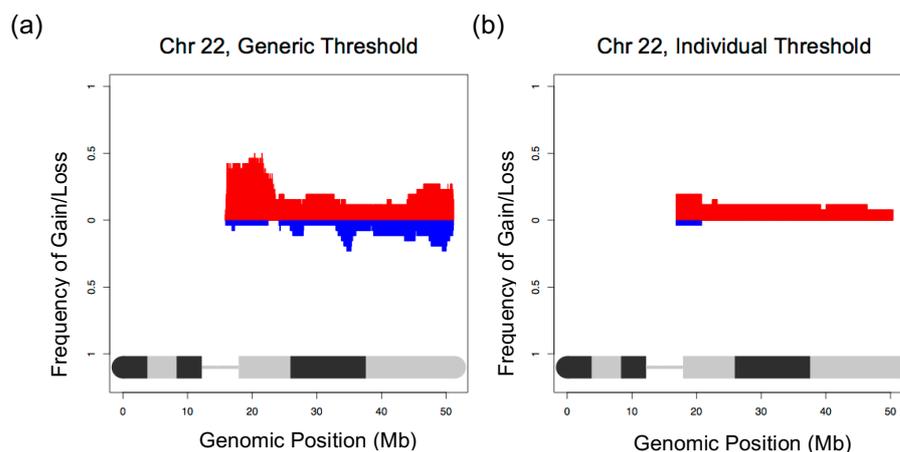
Figure 4-20: Comparison of cumulative individual chromosome plots for chr 22 across non-metastatic primary tumours. A generic CNA threshold (a) shows a deletion of 22q13.1, whilst this is no longer present when applying an individual sample threshold (b).

4.2.2.4 Identification of minimally altered genomic regions associated with metastasis

The CNA profile of nodal metastases and their metastatic primary tumours have been shown to share similarities (see section 4.2.1.4). Disparities may be due to differences in clonal sampling between the two. The metastatic clone is likely to be the dominant clone in the nodal metastasis. Therefore when attempting to identify markers for metastasis it may be more useful to compare the profile of non-metastatic primary tumours to that of the nodal metastases. Any marker found is likely to be identifiable in the metastatic primary tumour as demonstrated by the correlation and concordance analysis.

Though GISTIC identifies potential focal regions of relevance in known clinicopathologic groups, the wide peak boundaries for these are often extremely broad and in order to try and identify potential candidate genes of interest it is necessary to identify the smallest region that is being recurrently copy number altered (the minimally altered region). This is of particular importance when attempting to identify genes of interest, in order to reduce the number of potential candidates. By using open access software called pheatmap (Raivo Kolde, Harvard USA) and combining this with an R-based script called genomeHeat (Dr Henry Wood, Precancer genomics) I was able to create group-wide heatmaps for these samples, utilising the CNAnormout.txt files as input. This software provided greater resolution than

the CNAnorm-generated cumulative frequency plots when attempting to identify the minimally altered region, as well as allowing the ability to simultaneously count the number of samples that any region was altered in. As individual sample CNA thresholds were applied to the input files this was an ideal software.

Firstly the focal region identified by GISTIC in nodal metastases and non-metastatic tumours were analysed to identify the true minimally altered region, corresponding cytoband and frequency in each group. pheatmap was sued to produce sequential heatmaps with increasing resolution to identify the minimally altered region (see Figure **4**-**21**). This process was then repeated for all focal CNAs identified by GISTIC.

Comparison of the frequency of minimally altered regions in nodal metastases and non-metastatic primary tumours is shown in Table **4**-**6**. The focal regions exclusively seen in the nodal metastases could suggest driver genes that play a key role in cellular events enabling metastasis. Focal CNAs shared between the non-metastatic tumours and nodal metastases could represent changes involved in cancer development rather than necessary for metastasis.
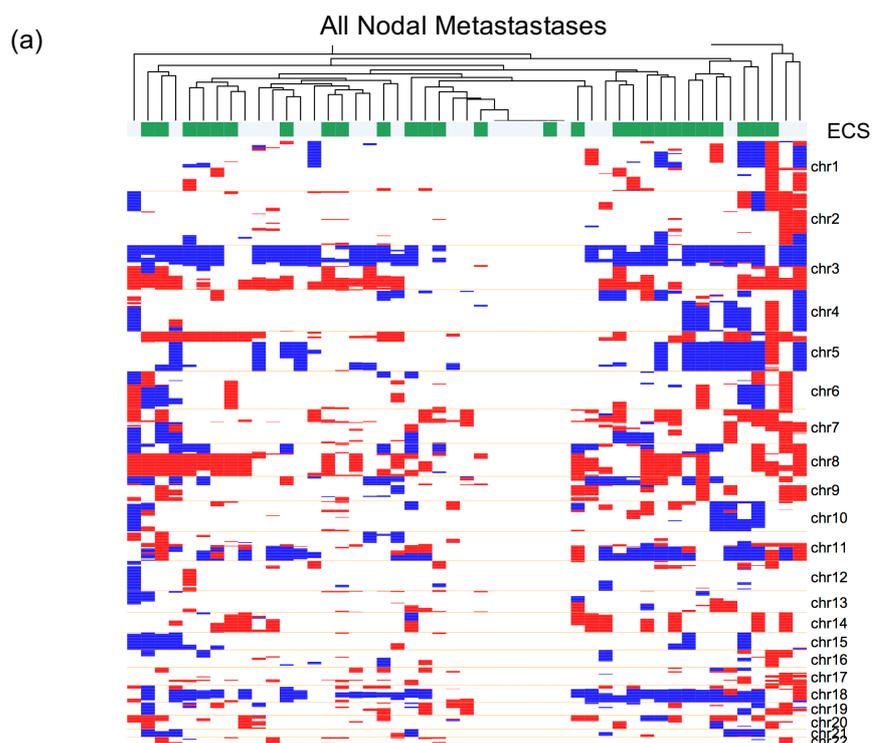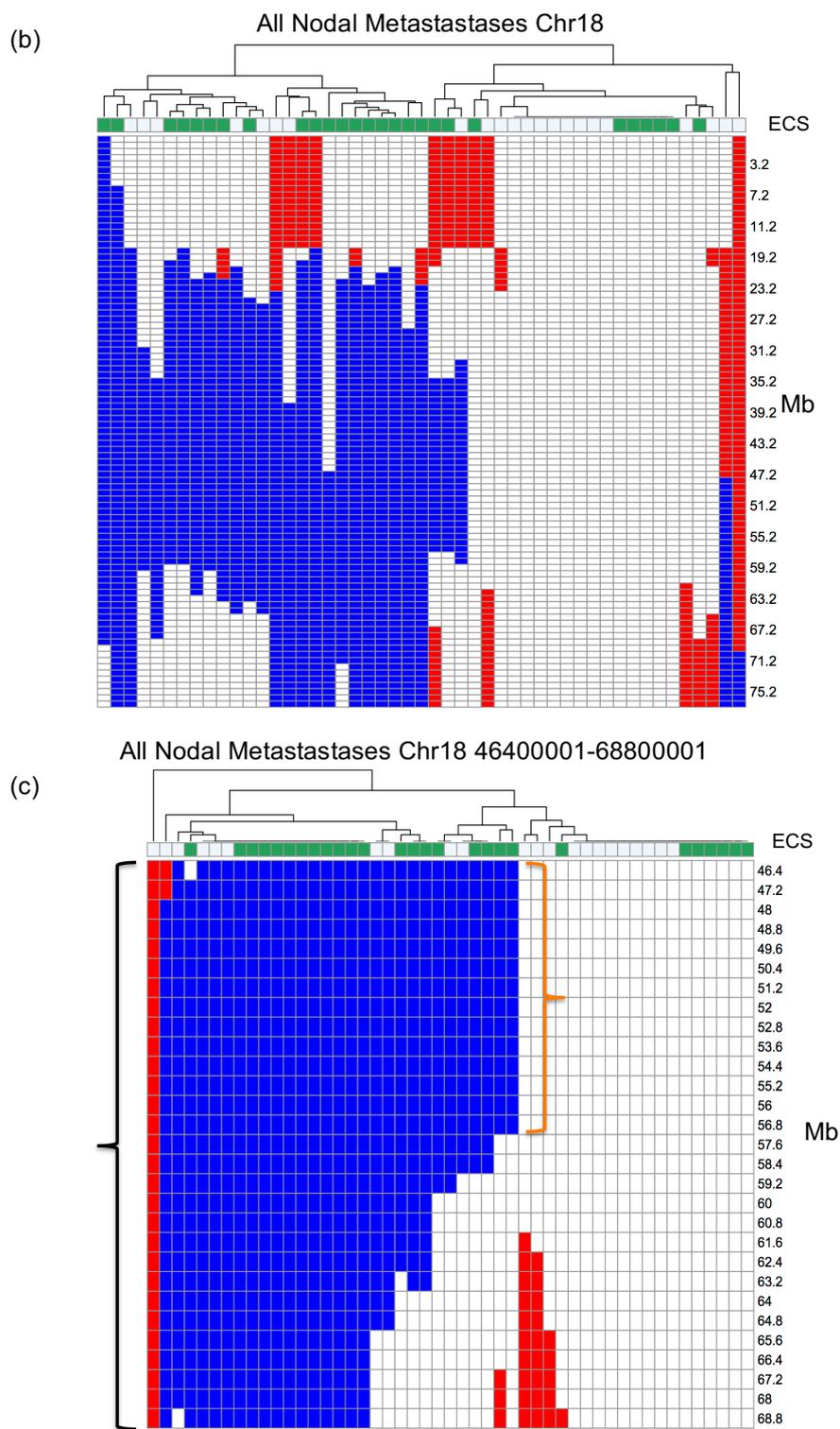
Figure 4-21: Example of sequential heatmaps produced to identify minimally altered region. (a) shows a genome-wide view of the CNA profile of all nodal metastases. (b) demonstrates a view of chr 18q. (c) shows the focal deletion identified using GISTIC is highlighted with black brackets. The actual minimally altered region is highlighted by orange brackets.

Amplifications			Frequency (%)	
Cytoband	GISTIC wide peak boundaries	Minimally altered region	Nodal Metastases (n = 49)	Non-metastatic primary tumours (n = 26)
2q31.1	170400002-176800000	170400002-176800000	5 (10)	2 (8)
3q26.2	167200002-171200000	167200001-184000001	26 (53)	11 (48)
4p15.2-p15.1	24000002-32000000	24800001-31200001	5 (10)	1 (4)
5p14.1-p15.1	15200002-24800000	16000001-24000001	23 (47)	8 (31)
7p12.1-11.2	51200002-64000000	53600000-58000000	26 (53)	6 (23)
7q21.3-q22.3	100000002-106400000	86800001-104800001	9 (18)	2 (8)
8q24.12-q24.13	N/A	119200001-123200001	25 (51)	11 (42)
8q11.21	N/A	48100001-51800001	28 (57)	9 (35)
9p21.1	29600002-36800000	34400001-36800000	9 (18)	1 (4)
11q13.3-p13.2	68000002-71200000	68000001-70400001	22 (45)	2 (8)
13q33.1-q34	103200002-115169878	104000001-112800001	9 (18)	1 (4)
17q22-q23.3	55200002-62400000	56800001-61600000	15 (31)	3 (12)
Deletions				
3p14.3-p14.1	57600002-69600000	58400001-69600001	35 (71)	10 (39)
4p15.1-14	30400002-38400000	34400001-37600001	7 (14)	1 (4)
3p26.3-p26.11	1-9600000	2000000-5600000	31 (63)	9 (35)
5q33.1-q33.3	148000002-159200000	150400001-156800001	18 (37)	4 (15)
7q31.1-q35	107200002-159138663	108800001-145600001	7 (14)	1 (4)
8p23.3-p22	1-25600000	800001-16000001	14 (29)	3 (12)
2q34-q37.1	201600002-243199373	214400001-231200001	4 (8)	0
9p24.3-p24.2	2000001-5600000	800001-4000001	15 (31)	3 (12)
11q23.1-25	109600002-135006516	110400001-133600001	22 (45)	0
15q26.2-q26.3	92800002-102531392	94400001-101600001	10 (20)	1 (4)
18q21.1-q21.32	45600002-69600000	46400001-57600000	29 (59)	1 (4)
19p12	21600002-28800000	20000001-24000001	6 (12)	1 (4)

Table 4-6: Comparison of frequency of minimally altered regions in nodal metastases and non-metastatic primary tumours.

There were no minimally altered amplifications unique to one group. This was surprising, given GISTIC did not identify any focal amplifications in the

non-metastatic primary tumours of significance (q = < 0.25). This was likely related the fact that 10/26 (38%) of this group contained a greatly reduced number of CNAs, leaving fewer samples for GISTIC to compare focal amplification rates to background CNAs.

Several CNAs were found much more frequently in the nodal metastases. In terms of amplifications gain of 11q13.3-q13.2 is present in 22/49 (45%) of nodal metastases and only 2/26 (8%) of non-metastatic primary tumours. Gain of 7p12.1-11.2 is present in 26/49 (53%) of nodal metastases compared to 6/26 (23%) of non-metastatic primary tumours. Gain of 17q22-q23.3 was present in 15/49 (31%) of nodal metastases and 5/26 (19%) of non-metastatic primary tumours. Gain of 9p21.1 was present in 9/49 (18%) of nodal metastases and 1/26 (4%) of non-metastatic primary tumours.

The remaining focal amplification highlighted by GISTIC were found at rates within ≤15% in both groups. Two minimally altered regions were identified at occurring at high frequency in nodal metastases, using pheatmap alone. These were gain of 8q24.12-q24.13 and 8q11.21. They were found in 25/49 (50%) and 28/49 (57%) of nodal metastases and 11/49 (42%) and 9/26 (35%), respectively, of non-metastatic primary tumours.

A number of minimally altered deletions presented a greater contrast between the groups compared to amplifications, with two being unique to the nodal metastases. Loss of 11q23.1-q25 was present in 22/49 (45%) of nodal metastases and absent in non-metastatic primary tumours. Loss of 18q21.1-q21.32 was present in 30/49 (59%) of nodal metastases and only 1/26 (4%) of non-metastatic primary tumours. Loss of 3p14.1-p14.1 was present in 35/49 (71%) of nodal metastases and 10/26 (38%) of non-metastatic primary tumours. Loss of 3p26.3-p26.1 was present in 31/49 (63%) of nodal metastases and only 35% of non-metastatic primary tumours. Loss of 4p15.1 was present in 7/49 (14%) of nodal metastases and in 1/26 (4%) non-metastatic primary tumours.

Out of 22 focal CNAs, pheatmap enabled identification of a minimally altered region (within the wide peak boundaries provided by GISTIC) in 20. In one case the minimally altered region was the same size and in one case it was

actually slightly bigger. This enabled a more focused analysis of potential gene candidates to be undertaken.

4.2.2.5 Identification of genes

The genomic co-ordinates of the minimally altered regions were input into the UCSC Genome Browser to produce lists of all genes contained within (Kent *et al.*, 2002). These were then cross-referenced with gene lists known to be associated with HNSCC. These were comprised of 12 gene pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/), HNSCC genes from the cancer

gene census (http://cancer.sanger.ac.uk/cancergenome/projects/census) and a list of genes identified by Stransky *et al* and the TCGA HNSCC subgroup as harbouring a statistically significant frequency of somatic mutations in HNSCC (Stransky *et al.*, 2011, Cancer Genome Atlas, 2015).

In total 1116 genes were contained within the minimally deleted regions and 641 genes within the minimally amplified regions. After cross referencing these to the lists above, 54 deleted genes and 50 amplified genes were identified (see Table **4**-**7** and Table **4**-**8**). This is consistent with reports that CNA occur across broad regions suggesting the majority of genes are unlikely to be important in cancer or metastasis development. It also highlights the effectiveness of this process of identifying the minimally altered region.

Interestingly, these deleted genes were identified 752 times in nodal metastases compared to 64 times in non-metastatic primary tumours. Amplified genes of interest were identified 688 times in nodal metastases compared to 173 non-metastatic primary tumours. This is consistent with the fact that these CNAs were found more frequently in nodal metastases. It also suggests that in nodal metastases tumour suppressor genes and oncogenes are affected by CNAs at a similar rate.

The higher frequency of amplified genes in non-metastatic primary tumours may indicate that the loss of tumour suppressor genes play a more prominent role in metastasis. However, it is difficult to infer this as these CNAs were selected for their association with nodal metastases. Genomic analyses of HNSCC have suggested there is a surprisingly low level of

crossover in recurrently mutated genes between tumours (Lui *et al.*, 2013, Stransky *et al.*, 2011). For these reasons, pathway analysis rather than gene analysis is a potentially revealing approach to take to understand mechanisms involved in metastasis.

Amplifications	
Cytoband	**Genes**
2q31.1	ITGA6, ZAK, CIR1, CHN1, ATF2
3q26.2	MECOM, TBL1XR1, ZMAT3, PIK3CA, GNB4
4p15.2-p15.1	ANAPC4, SLC34A2, RBPJ, CCKAR
5p14.1-p15.1	CDH12, PRDM9
7p12.1-11.2	EGFR, PHKG1
7q21.3-q22.3	DBF4, ZNF804B, FZD1, AKAP9, AKAP9, GNGT1, GNG11, COL1A2, PPP1R9A, MCM7, GNB2, EPO, SERPINE1, RELN
8q24.12-q24.13	
8q11.21	PRKDC, MCM4
9p21.1	CNTFR, IL11RA, FANCG
11q13.3-p13.2	LRP5, CCND1, FGF19, FGF4, FGF3
13q33.1-q34	COL4A1, COL4A2
17q22-q23.3	CLTC, RPS6KB1, PPM1D, BRIP1

Table 4-7: List of amplified genes of identified in minimally altered regions associated with metastasis.

Deletions	
Cytoband	Genes
2q34-q37.1	ATIC, FN1, PLCD4, STK36, WNT6, WNT10A, FEV, IHH, PAX3, ACSL3, CUL3, IRS1, COL4A4, COL4A3
3p14.3-p14.1	FHIT, PRICKLE2
4p15.1-14	
3p26.3-p26.11	IL5RA, ITPR1
5q33.1-q33.3	ITK
7q31.1-q35	MET, WNT2, WNT16, POT1, GRM8, LEP, FLNC, SMO, CHRM2, DGKI, CREB3L2, KIAA1549, BRAF, WEE2, SSBP1
8p23.3-p22	ANGPT2
9p24.3-p24.2	
11q23.1-25	POU2AF1, PPP2R1B, SDHD, PAFAH1B2, PCSK7, IL10RA, DDX6, CBL, ARHGEF12, HSPA8, EI24, CHEK1, FLI1, KCNJ5, TP53AIP1
15q26.2-q26.3	IGF1R
18q21.1-q21.32	SMAD4, MALT1, PMAIP1
19p12	

Table 4-8: List of deleted genes of identified in minimally altered regions associated with metastasis.

4.2.2.6 Analysis of gene pathways containing CNAs associated with metastasis

Using the lists described in section 4.2.2.5 the distribution of pathways enriched by CNAs in these identified genes was evaluated (see Figure **4-22**). Out of 12 signalling pathways included, 8 contained a higher number of amplified rather than deleted genes.

The pathway containing the most copy number altered genes was the PI3K pathway, well recognised to be associated with proliferation, invasion and metastasis. The EGFR/PI3K/AKT pathway has been identified as an important pathway in oral cancer and immunohistochemistry studies have suggested changes in expression in markers in this pathway are predictive of survival (Chang *et al.*, 2013). Lui et al found this to be the most frequently mutated pathway (31%) in whole exome sequencing data from 151 HNSCC tumours (Lui *et al.*, 2013). CNAs in genes involved in the PI3K pathway were identified in 45/49 (92%) nodal metastases and only 16/26 (62%) non-metastatic primary tumours. It is also worth noting that the ten non-metastatic primary tumours in which no PI3K pathway genes were identified contain very few CNAs.
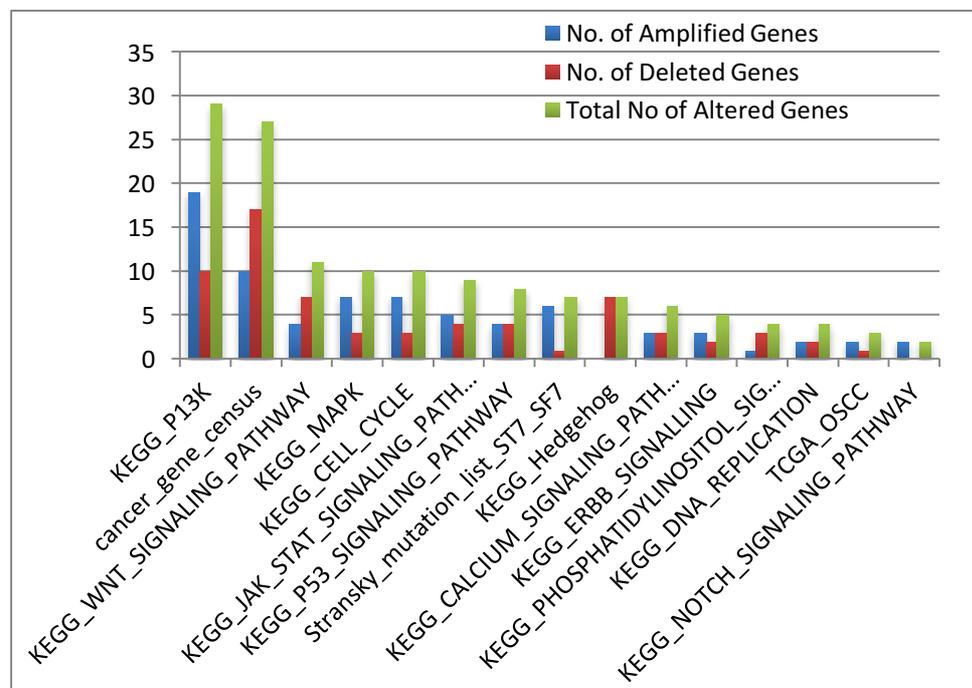
Figure 4-22: Distribution of copy number altered genes within pathways of interest in HNSCC as found in nodal metastases.

The WNT signalling pathway contained the next highest frequency of copy number altered genes, though only 11 hits were obtained compared to 29 in the PI3K signalling pathway. Copy number altered genes were identified at a similar frequency in the MAPK signalling pathway (10), cell cycle pathway (10), JAK/STAT pathways (9) and the p53 signalling pathway (8). Genes identified from the cancer gene census as associated with HNSCC produced the second highest number of hits. This is encouraging that the data is accurately identifying genes involved in cancer. The fact that not all genes are identified within this database is consistent with the fact that the census is based upon mutated genes rather than copy number altered genes, and drawn from all cancer types rather than just HNSCC.

4.2.3 Results 3: DNA copy number analysis of metastases associated with and without extracapsular spread

A important subgroup of HNSCC patients with metastases is those in whom ECS is identified histopathologically. This remains the most significant indicator of biologically aggressive disease and poor outcome in patients. These patients were identified from the metastatic cohort of patients (see Table 4-9) and their low-coverage sequencing data (FASTQ files) processed into digital karyograms as outlined in Chapter 2.9. Cumulative frequency plots were then created from these using seg_compare (Dr Henry Wood, Precancer genomics).

	Metastasis without ECS	Metastasis with ECS
No. of patients	22	27

Table 4-9: Table of patients with nodal metastases with and without ECS.

4.2.3.1 Cumulative frequency CNA plots of metastases with and without ECS

These were created as described in section 4.2.1.3 and inspected visually (see Figure 4-23).

Using this genome-wide view the CNA profile of the two metastatic groups contain a lot of similarities. Gain of 3q was present in 12/22 (55%) of metastases without ECS and 14/27 (52%) of metastases with ECS. Gain of 5p was present in 12/22 (55%) of metastases without ECS and 12/27 (44%) metastases with ECS. Loss of 5q was identified in 10/22 metastases without ECS and 10/27 (37%) metastases with ECS. Loss of 9p is present in 7/22 (32%) of the ECS group and 7/27 (26%) in the ECS group.
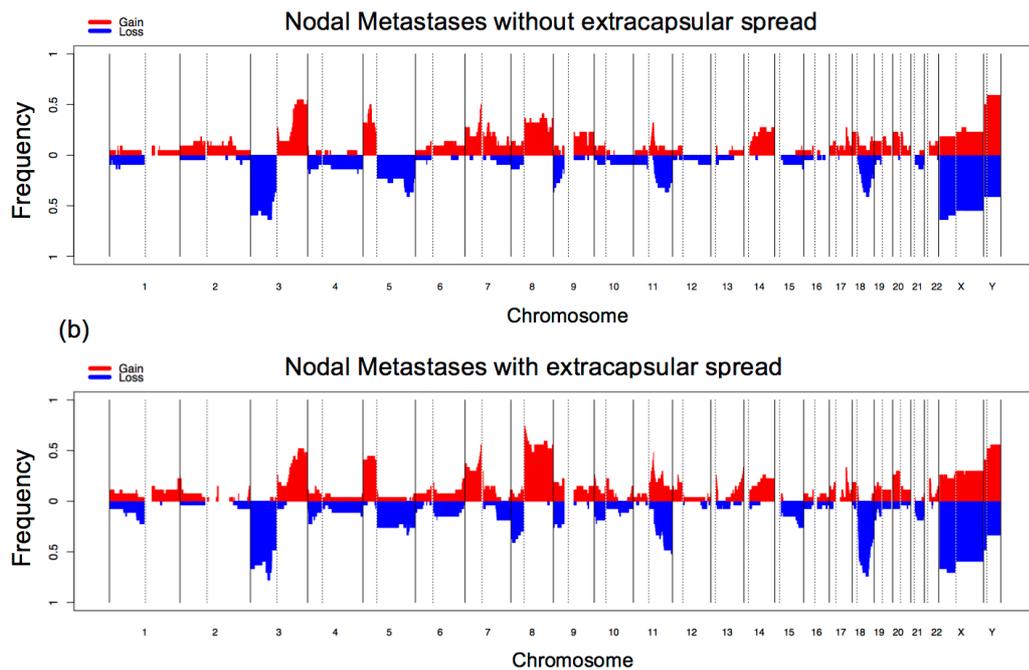
Figure 4-23: Cumulative frequency karyogram plots of (a) nodal metastases with ECS and (b) nodal metastases with ECS.

No broad CNA was unique to one group. Loss of 18q was present in 9/22 (41%) of metastases without ECS and 20/27 (74%) metastases with ECS. Similarly gain of 8q was present in 10/22 (45%) metastases without ECS and 19/27 (70%) in the ECS group. Gain of 7p was present in 8/22 (36%) metastases without ECS and 14/27 (52%) metastases with ECS. Loss of 8p was identified in 3/22 (14%) metastases without ECS and 11/27 (41%) in the ECS group. Loss of 6q was found in 1/22 (5%) metastases without ECS and 5/27 (19%) metastases with ECS. Focal analysis was undertaken to elucidate differences at higher resolution.

4.2.3.2 GISTIC analysis of focal CNA in metastases with and without ECS

Evaluation of focal CNA was performed using GISTIC 2.0 as described previously. GISTIC identified a number of focal amplifications common to both groups of metastases as occurring at a significant frequency (q = < 0.25) (see Figure 4-24). These included gains of 3q26.2, 7p11.2, 8q24 and 11q13.3. Gain of 7q22.1 was observed in the metastases with ECS, whilst gain of 2q11.2 and 9p13.3 were identified in the metastases without ECS by GISTIC. On inspection of the individual karyograms all focal amplifications were present in both groups, at differing frequencies (see Table **4**-**10**).

In terms of focal deletions GISTIC identified a number of differences in the focal alteration profile for these two groups (see Figure **4**-**25**). Nodal metastases without ECS show fewer focal deletions overall (only loss of 4p15.1, 9q24.2 and 18q21.2). Within metastases with ECS GISTIC highlighted losses at 3q26.1, 5q33.2, 11q24.2. Both sub-types of metastasis share loss of 18q21.2. Again the frequency of these focal alterations was evaluated using the individual sample karyograms (see Table **4**-**11**).

(a) Metastases without ECS   (b) Metastases with ECS
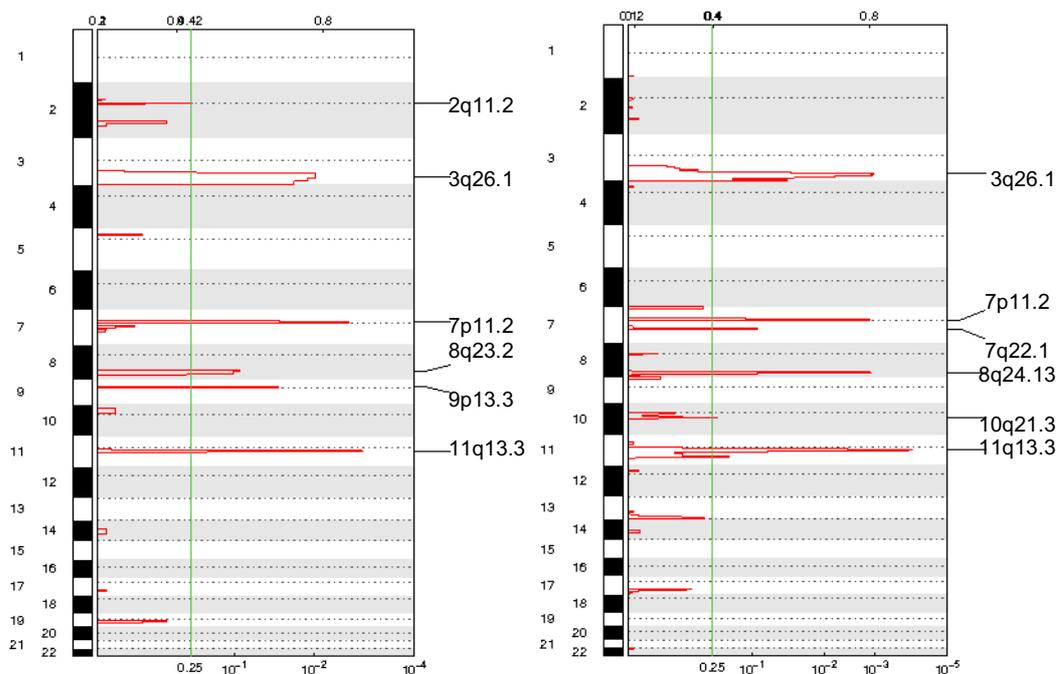
Figure 4-24: GISTIC analysis of focal amplifications in (a) nodal metastases without ECS and (b) with ECS.
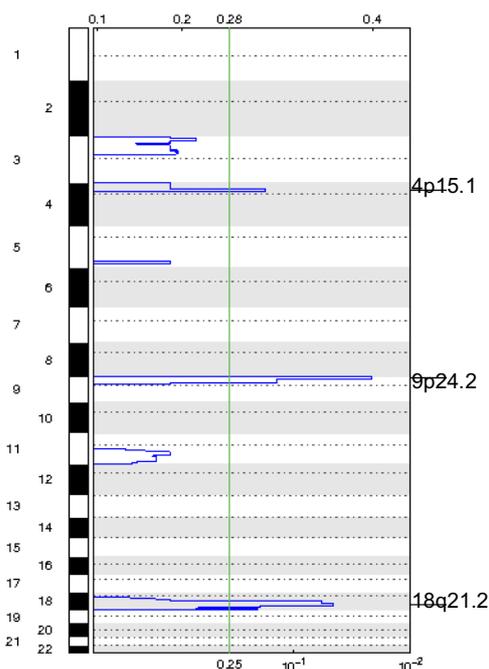
(a) Metastases without ECS   (b) Metastases with ECS
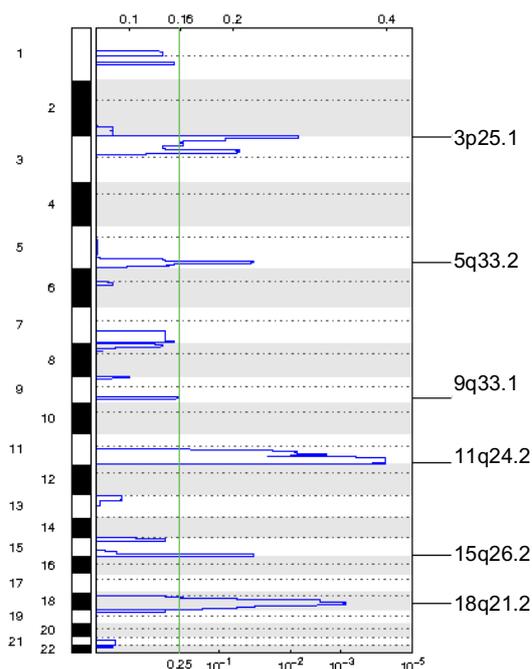
Figure 4-25: GISTIC analysis of focal deletions in (a) nodal metastases without ECS and (b) with ECS.

Using the wide peak boundaries the frequency of all focal CNAs highlighted by GISTIC were evaluated in each group. pheatmap (Raivo Kolde, Harvard USA) was used to identify the minimally altered region within the GISTIC co-ordinates and the frequency of these in each group of metastases was assessed (see Table **4**-**10** and Table **4**-**11**).

No CNA was found exclusively within one group of metastases, though several are identified at a higher frequency in one group. Gain of 8q24.12-q24.13 was present in 17/27 (63%) of metastases with ECS and 8/22 (36%) of those without ECS. Gain of 8q23.3 was present in 16/27 (59%) of metastases with ECS and 9/22 (41%) of those without ECS. Out of nine minimally amplified regions, four were found at a slightly higher frequency in the metastases without ECS. Three of these CNAs (10q21.1-q21.3, 3q26.1-q26.33 and 7q21.3-q22.3) were highlighted by GISTIC as occurring at a significant level (q = < 0.25) in the metastases with ECS. The fact they occurred at a slightly higher frequency in the group without ECS without being highlighted by GISTIC may be related to the lower number of samples in this group.

Amplifications			Frequency (%)	
Cytoband	GISTIC wide peak boundaries	Minimally altered region	Metastases without ECS	Metastases with ECS
2p12-p11.2	85600002-100000000	80800000-87200000	3 (14)	2 (7)
3q26.1-q26.33	166400002-171200000	164800001-182400001	12 (55)	14 (52)
7p12.1-p11.2	50400002-64000000	53600001-56800001	11 (50)	15 (56)
7q21.3-q22.3	94400002-106400000	96800001-106400001	5 (23)	4 (15)
8q23.3	105600002-129600000	112800001-116000001	9 (41)	16 (59)
8q24.12-q24.13	124000002-132000000	119200001-123200001	8 (36)	17 (63)
9p13.3	29600002-36800000	33600001-36000001	3 (14)	6 (22)
10q21.1-q21.3	62400002-68000000	62400002-68000000	2 (9)	2 (7)
11q13.3-q13.4	68000002-71200000	68800001-71200000	8 (36)	14 (52)

Table 4-10: Frequency of amplified minimally altered regions in metastases with and without ECS. The original GISTIC wide peak boundaries are shown next to the revised minimally altered region identified using pheatmap. The cytoband corresponds to the minimally altered region.

Deletions			Frequency (%)	
Cytoband	GISTIC wide peak boundaries	Minimally altered region	Metastases without ECS	Metastases with ECS
3p26.3-p25.3	1-9600000	1-9600000	13 (59)	18 (67)
4p15.31-p15.2	26400002-38400000	20800001-25600001	3 (14)	5 (19)
5q33.1-q33.3	148000002-163200000	150400001-158400001	9 (41)	9 (33)
9p24.3-p24.1	1-12800000	1-4800001	8 (36)	7 (26)
9q33.1-q33.2	116800002-124800000	119200001-123200001	1 (5)	2 (7)
11q22.3-q24.2	103200002-135006516	10800001-125600001	8 (36)	13 (48)
15q26.2-q26.3	91200002-102531392	94400001-102400000	3 (14)	7 (26)
18q21.1-q21.32	33600002-58400000	46400001-57600001	9 (41)	21 (78)

Table 4-11: Frequency of deleted minimally altered regions in metastases with and without ECS. The original GISTIC wide peak boundaries are shown next to the revised minimally altered region identified using pheatmap. The cytoband corresponds to the minimally altered region.

Notably, loss of 18q21.1 was present in 21/27 (78%) of metastases with ECS and 9/22 (41%) of metastases without ECS. Out of nine other focal deletions, seven were found at higher frequency in the metastases with ECS. One of these was identified by GISTIC in the non-ECS group only (loss of 9p24.3). The fact that this and loss of 5q33.1 were present at higher rates in the metastases without ECS suggests they represent genomic change not specifically associated with ECS. Losses at 15q26.2 and 11q24.2 were present in higher frequencies in the ECS group (12% and 11% more respectively), though overall present in less than 50% of samples in each group. This suggests that despite the strong clinical phenotype of ECS (poor prognosis and reduced survival), there is still a large amount of genomic heterogeneity between patients.

4.2.3.3 GISTIC analysis of broad regions of CNA in metastases with and without ECS

As stated previously the GISTIC analysis produces both a focal and broad analysis simultaneously. Heat maps generated from this for nodal metastases with and without ECS are shown in Figure **4-26**.

(a)   (b)
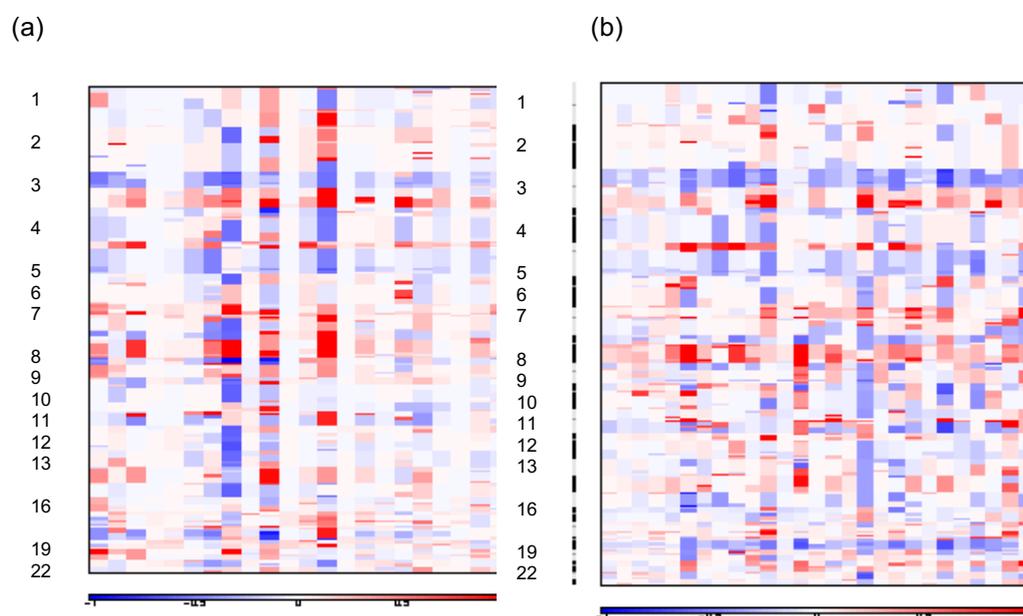
Figure 4-26: GISTIC heat maps for broad analysis of (a) nodal metastases without ECS and (b) with ECS.

They show similar profiles of gain and loss to those seen in the cumulative frequency plots generated using CNAnorm. The shared trend of loss of 3p

and gain of 3q are visible. Loss of 11q and 18q and gain of 8q are seen to occur at higher frequency in the nodal metastases with ECS. It also apparent that two of the samples in the nodal metastases without ECS contain a greatly reduced number of CNAs overall as well as 3 samples which may contain a lower tumour DNA content (inferred from the lower intensity of colour across these samples). This could be due to these metastases being on average smaller than those with ECS. This effect would be reduced by using an individual sample CNA threshold rather than a generic one as used by GISTIC.

4.2.3.4 Identification of genes in metastases with and without ECS

The genes within the minimally altered regions were identified using the UCSC Genome Browser (Kent *et al.*, 2002).

Amplifications	
Cytoband	Genes
2p12-p11.2	TCF7L1
3q26.1-q26.33	SLITRK3, MECOM, TBL1XR1, ZMAT3, PIK3CA, GNB4, SOX2
7p12.1-p11.2	EGFR, PHKG1
7q21.3-q22.3	MCM7, GNB2, EPO, SERPINE1, RELN, ORC5
8q23.3	CSMD3
8q24.12-q24.13	
9p13.3	CNTFR, IL11RA, FANCG, CREB3
10q21.1-q21.3	CDK1
11q13.3-q13.4	CCND1, FGF19, FGF4, FGF3
Deletions	
Cytoband	Genes
3p26.3	IL5RA, ITPR1, OXTR, SRGAP3
4p15.1	ANAPC4, ITK, EBF1
5q33.1	
9p24.3	
9q33.1	TLR4
11q24.2	ATM, DDX10, POU2AF1, PPP2R1B, SDHD, PAFAH1B2, PCSK7, IL10RA, DDX6, CBL, ARHGEF12, HSPA8, EI24, CHEK1
15q26.2	IGF1R
18q21.1	SMAD4, MALT1, PMAIP1

Table 4-12: Genes identified of potential significance within minimally altered regions associated with extracapsular spread.

These lists were then cross-referenced against the KEGG pathways and lists of genes associated with HNSCC (as described in section 4.2.2.5). The final list of genes of potential interest is shown in Table **4-12**).

4.2.3.5 Analysis of gene pathways containing CNAs in metastases with and without ECS

This was generated using the same pathways selected for cross-referencing the gene lists produced from the minimally altered regions (see section 4.2.3.4). The pathways enriched by CNAs in metastases with and without ECS are shown in Figure **4-27**. In terms of signalling pathways there is a higher frequency of amplified genes in each pathways rather than deleted genes. This suggests that oncogenes are more frequently involved in CNAs relating to ECS, rather than tumour suppressor genes being lost. The Cancer Gene Census contained the highest number of hit overall and demonstrates a much higher frequency of deleted genes. This may reflect the fact that the Cancer Gene Census is based on known mutated genes in cancer rather than copy number altered genes.
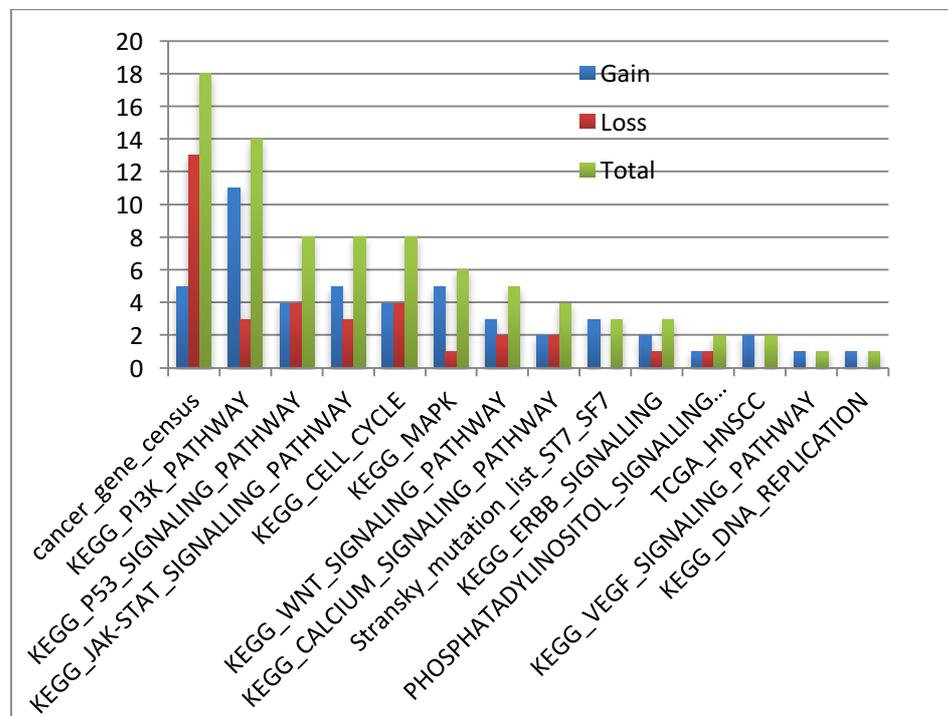
Figure 4-27: Number of genes cross-referenced to each pathway and gene lists.

Across all pathways ECS positive metastases contained a higher number of copy number altered genes (see Figure **4-28**). This is consistent with the fact that only 2 minimally altered regions were found in a higher number of ECS

negative metastases, and even then only in a small number of samples. A similar trend across both types of metastases was seen in terms of frequency of copy number altered gene hits. The pathway most frequently enriched for copy number altered genes was the PI3K pathway (in both ECS positive and negative metastases). This was followed by the p53 signalling pathway, in contrast to the copy number altered gene pathways associated with nodal metastases. In those minimally altered regions the MAPK pathway was the next most highly enriched.
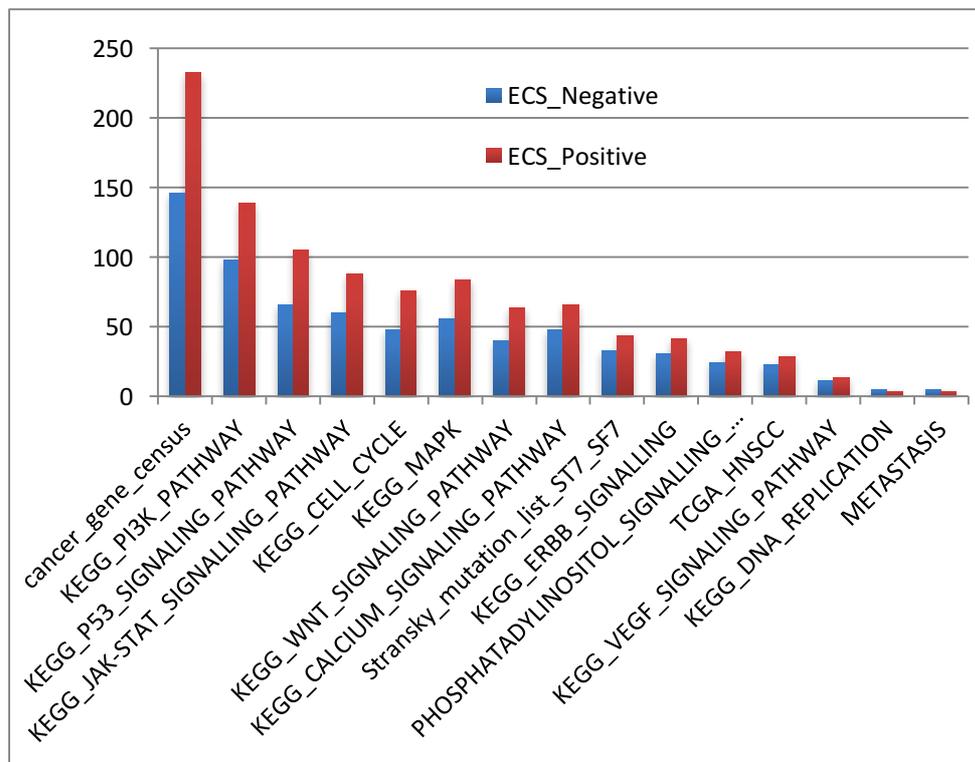
Figure 4-28: Number of times a gene within each pathway was identified in metastases with ECS (red) and without ECS (blue).

However there is a small difference in the overall number of gene hits contained with 171 hits for the p53 signalling pathway (in both ECS positive and negative metastases) and 140 hits for the MAPK pathway. Again there is a similar level of enrichment across the JAK/STAT, MAPK, Cell Cycle and WNT signalling pathways. An important influence on the number of hits obtained per list is the number of genes in each list. The PI3K pathway is very large and this obviously increases the chance of genes being identified as of significance.

4.2.4 A copy number panel associated with nodal metastasis

Using visual inspection of the cumulative frequency karyograms and the minimally altered regions associated with metastases, a panel of CNAs associated with nodal metastasis was identified (see Table **4-13**)

Gain	Loss
4p15.2-p15.1	4p15.1-p14
7p12.1-p11.2	5q33.1-q33.3
9p21.1	7q31.1-q35
11q13.3-q13.2	8p23.3-p22
13q33.1-q34	9p24.3-p24.2
17q22-q23.3	11q23.1-q25
	15q26.2-q26.3
	18q21.1-q21.32

Table 4-13: Selected panel of CNAs associated with nodal metastasis.

All (both metastatic and non-metastatic) samples were then assessed for the presence of these CNAs using the pheatmap (Raivo Kolde, Harvard USA) software. Initially non metastatic primary tumours were compared to all nodal metastases (see Figure 4-29). This heatmap demonstrates the increased frequency of these minimally altered regions in the nodal metastases, but also highlights the fact that no CNA is truly mutually exclusive to one group. Assigning a score of 1 for each CNA present in each sample and 0 for each that is absent in each sample a score was generated. The scores for non-metastatic primary tumours and nodal metastases were analysed using the Shapiro-Wilk test for normality, showing them to be not normally distributed. The Mann Whitney test was used to compare mean CNA scores for each group (see Figure 4-30). The mean CNA score of the nodal metastases was significantly higher than the non-metastatic primary tumours ($p = 0.0001$).

N0 Tumours vs All Metastases
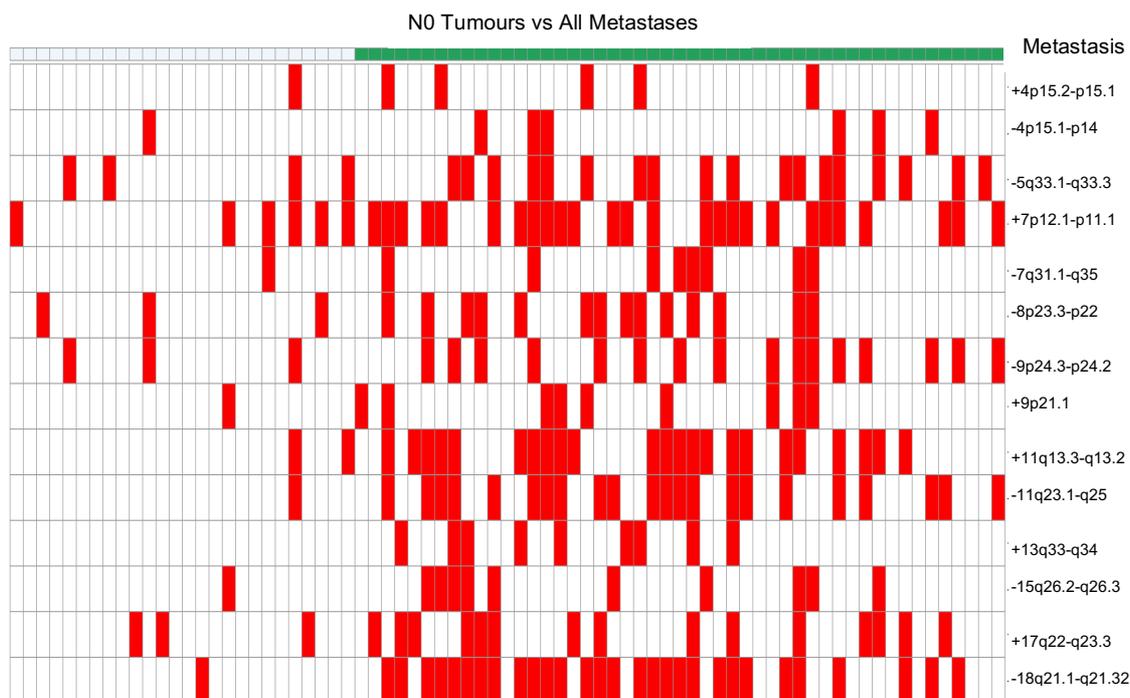
Figure 4-29: Heatmap of presence (red) or absence (white) of panel of CNAs associated with nodal metastases. The non-metastatic primary tumours are represented along the top in white and the nodal metastases in green.
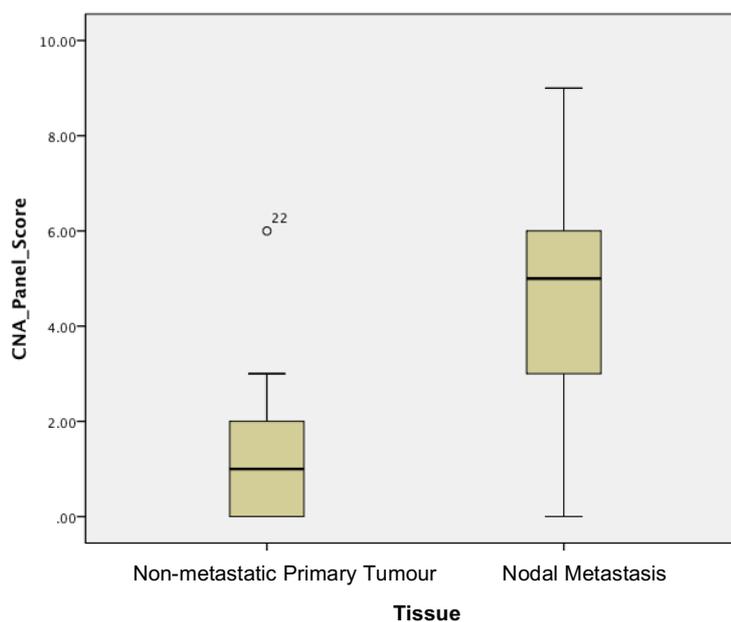
Figure 4-30: Box-plot of the mean CNA scores for non-metastatic primary tumours and nodal metastases.

For clinical benefit, any marker for metastasis needs to be detectable in the primary tumour, in order to potentially spare patients unnecessary surgery. Using the same CNA panel, the non-metastatic primary tumours were compared to the metastatic primary tumours (see Figure **4-31**).

The metastatic primary tumours were also scored for the presence and absence of the CNA markers as described above. The scores for the non-metastatic primary tumours and metastatic primary tumours were then analysed (see Figure **4**-**32**) using the Shapiro-Wilk test which revealed them to be not normally distributed. The mean scores for each group were then compared using the Mann Whitney test, which found the mean score of the metastatic primary tumours to be significantly higher than the non-metastatic primary tumours (p = 0.0001).
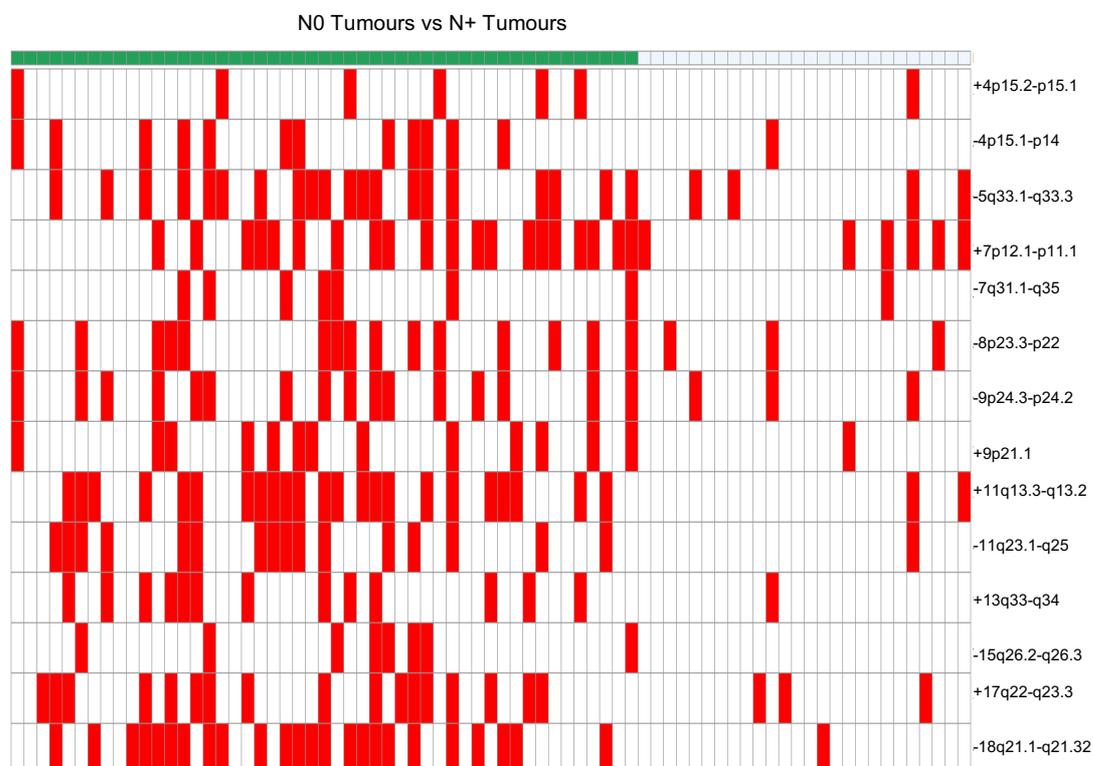
Figure 4-31: Heatmap of presence (red) and absence (white) of CNA panel markers. Non-metastatic primary tumours are shown in white along the top and metastatic primary tumours are shown in green.
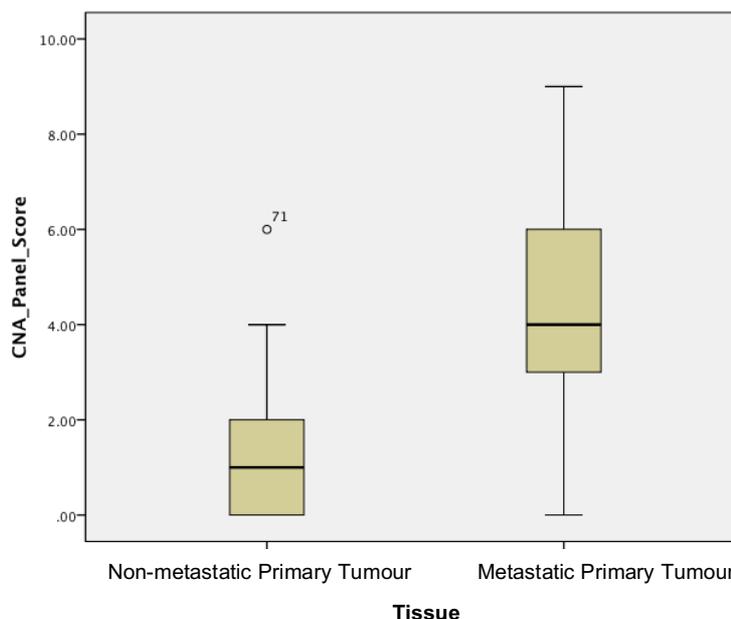
Figure 4-32: Box-plot comparing CNA-scores for non-metastatic primary tumours and metastatic primary tumours.

4.2.5 Analysis of markers of genomic damage

4.2.5.1 Comparison of breakpoints

Though present, the differences in CNAs between the clinicopathologic groups are heterogeneous. Therefore I elected to use a more generic approach to separate samples. The number of segments in the .bed file for each sample reflects the number of breakpoints in each sample. As breakpoints are locations on chromosomes where DNA may be altered or damaged in a number of ways (e.g. amplification, deletion, inversion, translocation) this was hypothesised to be a reasonable estimation of the level of genomic damage contained in each sample. The mean number of segments in metastatic and non-metastatic samples was calculated and compared (see Figure **4-33**). The Shapiro-Wilk test confirmed normality of data and an independent t-test showed no significant difference between non-metastatic tumour and metastatic primary tumours (p = 0.449).
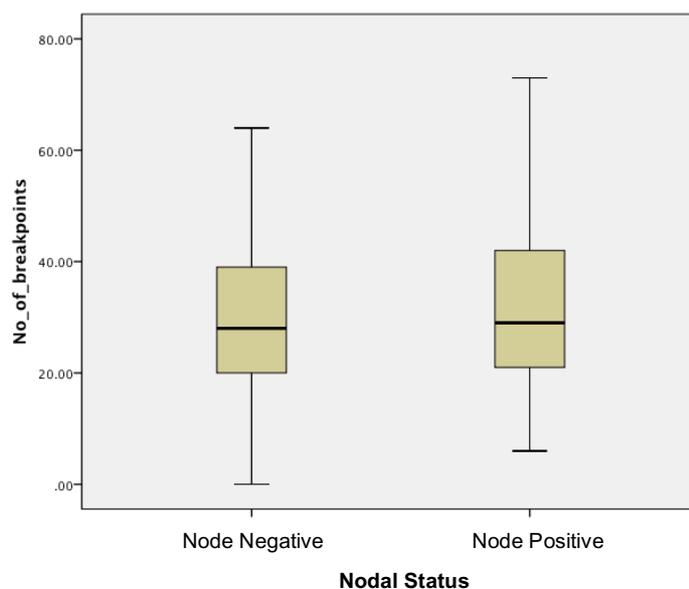
Figure 4-33: Box-plot to compare the number of breakpoints in non-metastatic primary tumours (Node_Negative) and metastatic primary tumours (Node_Positive).

A similar comparison was performed using the ANOVA test (see Figure **4-34**) to compare the number of breakpoints in node negative tumours to tumours associated with and without ECS, again finding no significant difference (p = 0.752). Pearson's correlation (2-tailed) was performed to evaluate any trend of increasing numbers of breakpoints and nodal status. No significant correlation was found (p = 0.513).
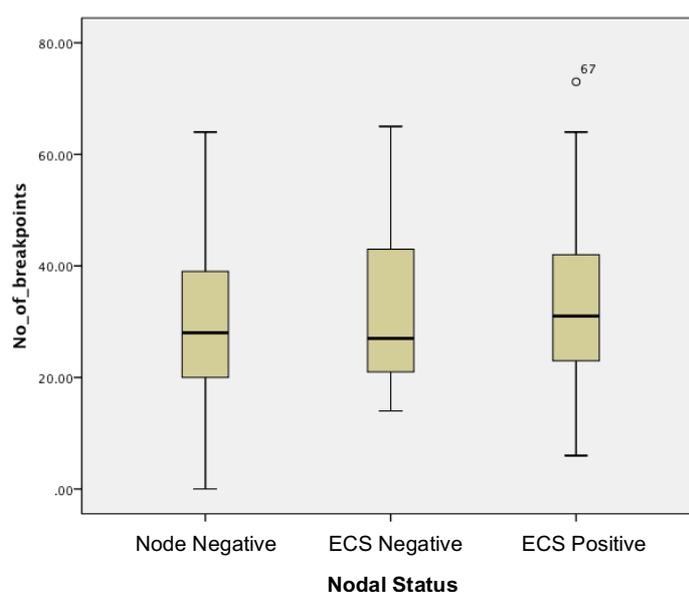
Figure 4-34: Box-plot to compare the number of breakpoints in non-metastatic primary tumours to tumours associated with ECS and without ECS.

4.2.5.2 Fraction of Genome Altered (FGA)

The actual proportion of the genome that was altered in terms of its copy number was calculated as another reflection of the amount of genomic damage in samples from each clinicopathologic group. The FGA is calculated using the number of bases in all segments in each sample (i.e. the genome) as the denominator and number of bases in each sample that are copy number altered as the numerator. These were then compared between groups as with the breakpoints above.
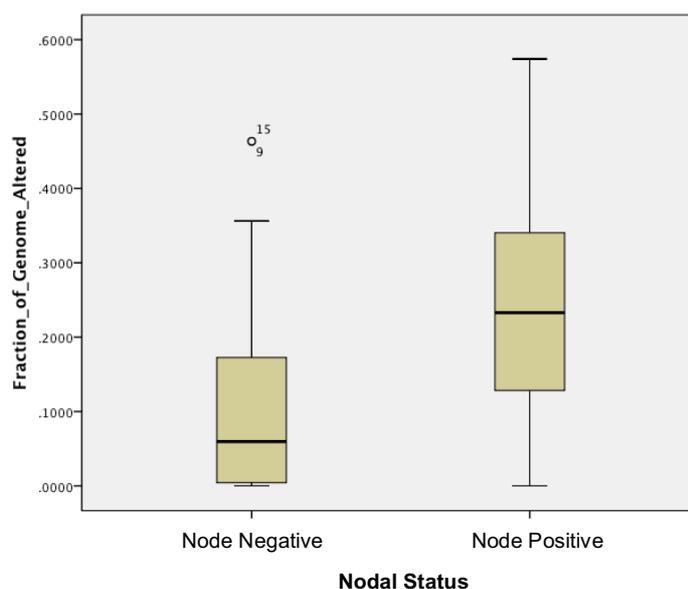
Figure 4-35: Comparison of FGA for non-metastatic primary tumours and metastatic primary tumours.

The FGA for non-metastatic primary tumours and metastatic primary tumours (see Figure 4-35) were analysed using the Shapiro-Wilk test for normality. As it was not normally distributed the Mann-Whitney U test was then used to compare the means of these groups. The mean FGA of the metastatic primary tumours was highly significantly increased compared to the non-metastatic primary tumours (p = 0.0001). The FGA for non-metastatic primary tumours was then compared to metastatic tumours associated with and without ECS (see Figure 4-36). The Mann Whitney test was used to compare these groups and found that both tumours associated with and without ECS had a significantly increased FGA compared to non-metastatic primary tumours (p = 0.001 and p = 0.001, respectively). On comparing the mean FGA between the primary tumours associated with

ECS and those not associated with ECS no significant difference was found (p = 0.673).
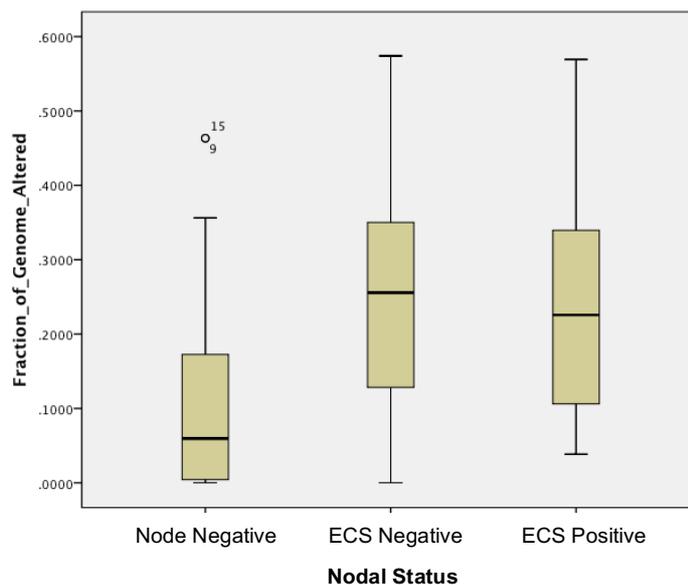
Figure 4-36: Comparison of FGA from non-metastatic primary tumours to primary tumours associated with and without ECS.

4.3 Discussion

4.3.1 Tissue sampling

When considering these results the source of the tissue is important. FFPE tissue has long been regarded as unsuitable for nucleic acid analysis. DNA from this tissue has been damaged by the process of formalin-fixation as well as the period of storage at room temperature (Iwamoto *et al.*, 1996, Greer *et al.,* 1991a, Greer *et al.*, 1991b). The amount of DNA obtained from FFPE samples is likely to also be much lower than compared to fresh-frozen samples (Greer *et al.*, 1991a, Greer *et al.*, 1991b, Farrand *et al.*, 2002). However, by appreciating these issues FFPE tissue can be used consistently. The effects of formalin-fixation on DNA include the formation of crosslinks that maintain morphology but inhibit denaturation and therefore inhibit the action of polymerases (Lu *et al.*, 2010). Extraction techniques need to take this into account, as well as subsequent processing. Using NGS to produce copy number data allowed me to produce sequencing libraries from as little as 50 ng of extracted DNA.

Additionally, the choice to use NGS technology allowed me to multiplex 40 libraries per sequencing lane. This was essential to completing the project within grant budget as well as being well suited to the type of genomic analysis being planned in this study. Low-coverage sequencing such as this would not be appropriate for whole exome or whole genome sequencing (Zeng and Mortazavi, 2012). For copy number analysis it has been demonstrated using lung SCC and HNSCC samples to be consistent and reliable when compared to technical replicates at higher coverage (Wood *et al.*, 2010, Gusnanto *et al.*, 2014). Given that CNAs are often broad in cancer the resolution achievable multiplexing 40 samples per lane was deemed to be an appropriate depth (Taylor *et al.*, 2008, Wood *et al.*, 2010). It could be argued that very low frequency, minimal regions of copy number loss will be missed using low coverage sequencing, but as the original statistical projections suggested a minimum of 20 patients per group it would be difficult to draw strong conclusions from very low frequency CNAs in these small groups.

The issue of tumour DNA content is important to understand. Laser microdissection is often regarded as the gold standard in achieving maximal tumour cell purity (Stoehr *et al.*, 2003). However, attempts to use the PALM laser system within Leeds Institute of Cancer and Pathology were made unsuccessfully. Nonviable yields of DNA were achieved as well as issues with accurate alignment of the laser during use, which was out of service contract. This raised concern over the having to increase the number of cycles of PCR downstream during sequencing library preparation (with the resultant effect of introducing many PCR replicates to the sequencer). The rate of dissection was also extremely slow rate (one sample microdissection taking more than one day). It is well-established that non-laser dissection of tumour samples can be performed with minimal influences on tumour cell purity and nucleic acid profiles (de Bruin *et al.*, 2005, Michel *et al.*, 2003). As I was able to target the highest tumour cell content for dissection and estimate the subsequent ploidy and tumour DNA content from the copy number data I elected to abandon laser microdissection as a non-viable alternative.

Accuracy was ensured by taking a repeat 5 μm section of each sample after the 7 sections had been cut for DNA extraction. This was then stained with haematoxylin and eosin and examined by a consultant head and neck pathologist and the area of highest tumour cell content re-marked. This was used to confirm the accuracy of the area dissected for DNA by comparing it to a random section used for DNA extraction. This also served to confirm that the tumour had not been "cut through" by repeated sectioning.

4.3.2 Advantages and disadvantages of NGS

Several techniques are available to detect CNA. Older techniques include FISH and qRT-PCR. FISH has relatively low resolution, whilst qRT-PCR require specific primers, which may result in suboptimal precision as not all regions of the genome lend themselves to ideal primer design (Duan *et al.*, 2013, Hughes S, 2007).

aCGH provides higher resolution than FISH, though this is dependent on the number of probes embedded in the array and therefore directly related to the cost per sample. To obtain a similar resolution to that achieved in my low coverage sequencing (800 kb) a microarray would cost approximately £600 per sample (Agilent, UK). It only cost approximately £50 per sample to sequence these on the Illumina HiSeq platform. aCGH is well established and widely used, meaning that expertise is easily available. Its limitations are understood, including bias due to relatively lower sensitivity for detecting gains compared to losses and the fact only CNAs targeted by the array probes can be identified (Stankiewicz and Beaudet, 2007). Both aCGH and NGS can provide a genome-wide view of copy number status.

Disadvantages of NGS include the fact it is a much more recent development and therefore expertise in both wet lab and bioinformatics analysis may not be as available. Though bias is avoided in terms of probes, the guanine-cytosine (GC) content of the input DNA affects sequencing reads. GC-poor or GC-rich regions have been reported to lead to uneven coverage of reads with relatively few reads in these parts of the genome (Chen *et al.*, 2013). This needs to be accounted for and any analysis needs to normalise for this effect. An advantage of using CNAnorm is that it

automatically corrects for GC content, which has been shown to result in a more accurate estimation of CNA (Gusnanto *et al.*, 2012).

As NGS does not utilise targeted probes the resolution of CNA is dependent on the depth of sequencing. As sequencing reads are short and may contain errors, when attempting to reveal single nucleotide variants high depth sequencing is required to ensure accuracy. As I was not attempting to identify mutations I did not require high depth sequencing. High depth sequencing cannot overcome alignment issues with highly repetitive regions that are longer than the sequencing read. Using paired end reads (two reads separated by a known distance) can provide confidence in aligning reads to repetitive regions shorter than the distance separating the paired ends (Sims *et al.*, 2014).
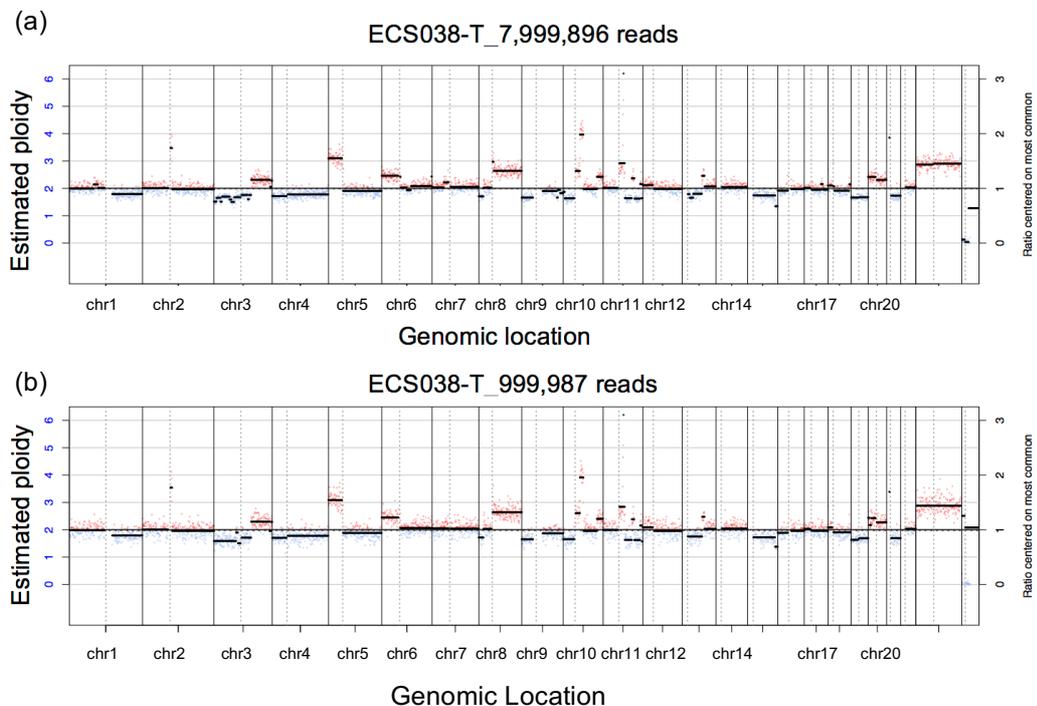
Figure 4-37: (a) shows the digital karyogram generated from the original ECS038-T FASTQ file (7,999,896 sequencing reads). (b) shows the digital karyogram created after the FASTQ file was divided into 8 (999,987 sequencing reads).

Resolution of sequencing data can be reduced and manipulated *post hoc* in contrast to aCGH data. In order to examine the accuracy of CNA profile by sequencing depth a sample was selected (ECS038-T) for *in silico* analysis. ECS038 achieved 7,999,896 sequencing reads when run on the Illuina HiSeq 2500. This FASTQ file was then divided into 8 equal segments,

essential reduced the number of reads for a single sample down to 999,987 reads. Digital karyograms were created for these as for all other samples (see Figure **4**-**37**). These images show strong similarity in CNA profile, demonstrating the accuracy of the depth of sequencing used in my study.

In order to examine the consistency of breakpoint between the original and 'diluted' sample the individual chromosome plots were further evaluated. An example of a chromosome containing multiple breakpoints is shown in Figure **4**-**38**. This was combined with the .bed file containing the genomic locations of the breakpoints (see Table **4**-**14**).
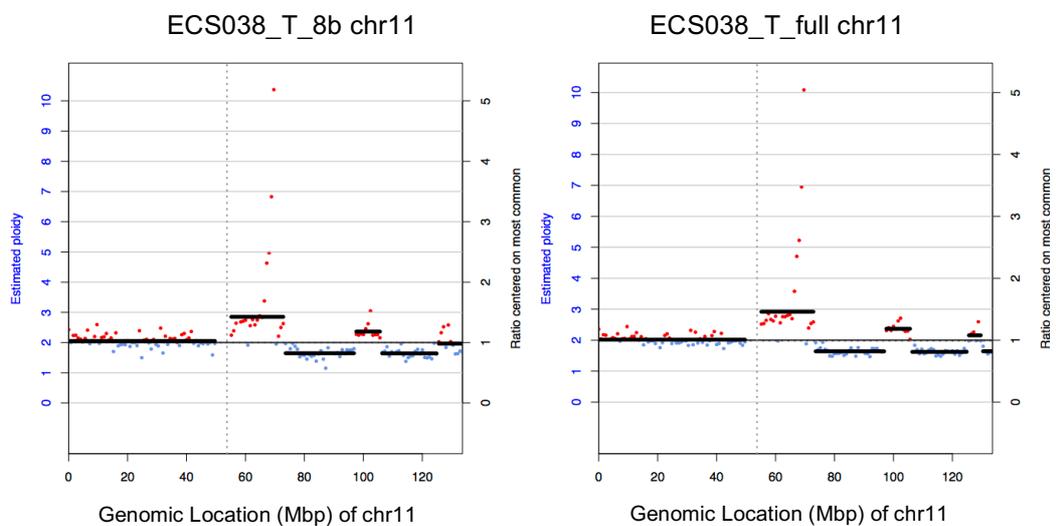
Figure 4-38: Individual chromosome plot of chr 11 (a) from the ECS038-T sample divided into 8 and (b) the full ECS038-T sample.

	Genomic Locations			
	ECS038-T Full Sample		**ECS038-T_Divided into 8**	
	Start	End	Start	End
chr11	1	49600001	1	49600001
chr11	55200001	72800001	55200001	72800001
chr11	73600001	96800001	73600001	96800001
chr11	97600001	105600001	97600001	105600001
chr11	106400001	124800001	106400001	124800001
chr11	125600001	129600001	125600001	133600001
chr11	130400001	133600001		

Table 4-14: Table listing precise genomic locations (breakpoints) of segments as identified in both the full ECS038-T sample and the ECS038-T sample divided into 8.

As can be seen from the individual chromosome plot of chr 11 the overall CNA profile is very similar despite there being multiple breakpoints and segments present. The genomic locations are also very similar with 11/14 breakpoints in the full sample being identical to the divided sample. However, one breakpoint is absent from the divided sample, meaning a small segment of loss from the distal end of 11q is not identified. This reflects the fact no reads contained in the divided sample aligned sufficiently around the distal breakpoint for this to be identified. It highlights the potential for CNAs to be missed at lower levels of sequencing. It is of interest that this breakpoint is close to the telomere of 11q which may also have affected the ability for the breakpoint to be identified as these regions are more repetitive and can present challenges in alignment of reads. It is also in a region where CNAs need to be regarded with caution regardless of depth of sequencing due to the nature of the repetitive genome in these regions.

In total 174 breakpoints were present in the full sample and 134 in the divided sample. This also highlights the potential for CNA alterations to be missed with lower coverage sequencing. However when concordance of the divided sample was calculated for the original ECS038-T sample it was found to be high (95%). This suggests that whilst breakpoints may be missed the regions of gain and loss identified are still very consistent. This likely reflects the fact that breakpoints may be present along a copy number altered segment without change in overall copy number of adjacent segments. Intra-tumour heterogeneity can also increase the number of breakpoints present, whilst not affecting the overall copy number status being called as CNAs within smaller clonal populations may not reach the overall CNA call threshold, and therefore not affect the copy number status.

4.3.3 CNAs in HNSCC

Our understanding of HNSCC as a "genetic disease" has grown exponentially over the last three decades (Vogelstein and Kinzler, 2004). Regions of altered copy number potentially contain proto-oncogenes or tumour suppressor genes. Copy number can correlate with increased or reduced gene expression in both health and disease (McCarroll *et al.*, 2006, Stranger *et al.*, 2007, Sung *et al.*, 2013).

Using a number of techniques (FISH, LOH analysis and aCGH) several recurrent chromosomal alterations have been identified in HNSCC. Some of these could be considered as "classical" (identified in multiple studies in HNSCC) including: loss of 3p, 4, 5q, 8p, 9p, 11, 13q and 18q. Gains of 3q, 5p, 8q, 9q, 11q, 10q and 22q have also been reported (Bockmuhl *et al.*, 1996, Singh *et al.*, 2001, Chen and Chen, 2008, Bockmuhl *et al.*, 1998, Gollin, 2001). In considering the alterations seen in this study the sample groups must be considered in turn.

4.3.3.1 CNAs associated with metastasis

Using CNAnorm generated cumulative frequency karyograms a very similar profile between metastatic primary tumours and their matched metastases can be observed. This similarity is further demonstrated by the rates of concordance and correlation calculated between the metastatic primary tumour and the matching nodal metastasis. These figures do reveal that the DNA copy number signature of metastasis and parent tumour are not identical and that there is variation in their similarity. This is important in trying to identify a marker and gain understanding of the genomic changes associated with metastasis. If a marker is identifiable in the nodal metastasis – is it then identifiable in the primary tumour?

There are relatively few studies using genome-wide approaches, reporting the genomic profile of cervical metastases in HNSCC. Kujawski *et al* analysed 19 pairs of laryngeal primary tumours and metastases using CGH (Kujawski *et al.*, 1999). The majority of metastases were cervical nodes (90%) whilst the others were unspecified metastases. They found losses at 8p, 9q and 13 were more frequent in metastases compared to the primary tumours.

Bockmuhl *et al* analysed 34 metastatic HNSCC primary tumours and their lymph node metastases as well as 20 non-metastatic primary tumours using CGH (Bockmuhl *et al.*, 2002). These were of mixed subsite with the majority being oropharynx and hypopharynx. The proportion of oropharyngeal primary tumours is relevant as this introduces the potentially confounding factor of HPV-associated tumours being included in this cohort. They compared non-metastatic primary tumours to lymph node metastases. They

found few chromosomal regions in the pN0 primary tumours carried an excess of changes compared to the nodal metastases. Deletions of 5q34-q35, 8p12-p22, 10p12, 10q21-qter, 11p14-p15, 11q23-qter and 14q21-qter and gains of 1q21-q22, 3q24-qter, 6q, 7q11.2, 12q12-q12 and 18p11.1 were associated with the nodal metastases.

Patmore *et al* analysed and compared 23 paired primary tumour and lymph node metastases using CGH (Patmore *et al.*, 2004). They found no overriding aberration in the nodal metastases, though gains of 6q (48% c.f. 2%) and 22q (26% c.f. 9%) were found at higher frequencies.

The discordance between studies could be due to several reasons. The generally low number of patients in these reports does limit conclusions as HNSCC is well recognised to be a genomically heterogeneous disease. The different analytical methods used in each study are important and indeed the underlying driver behind the development of GISTIC, which was not available at the time of the previous studies discussed. It is also possible that the aetiology of the tumours is a confounding factor. Alexandrov *et al* demonstrated trends in mutational signatures that were shared between tobacco smoking-associated cancer (lung, head and neck and liver) (Alexandrov *et al.*, 2013). Though likely to be a minority, the unknown proportion of HPV-associated tumours in Bockmuhl *et al's* studies could affect genomic findings (Bockmuhl *et al.*, 2002). Patmore *et al's* also study lacked non-metastatic primary tumours as a comparison (Patmore *et al.*, 2004). These studies were limited in terms of the number of patients included as well as the relatively low resolution of the conventional metaphase CGH technology available at the time.

More recently Yoshioka *et al* used array-based CGH primary tumours and lymph node metastases from 15 OSCC patients and compared these profiles to 10 non-metastatic primary tumours (Yoshioka *et al.*, 2013). This study used technology achieving similar resolution to that obtained using low-coverage NGS. They compared concordance rates of genomic breakpoints to assess clonality between primary tumour and matched metastases. By performing unsupervised hierarchical clustering they were able to identify 12 of 15 matched tumour-lymph node pairs, indicating the

genomic profile of the matched lymph node was most similar to the parent tumour in these cases. Their analysis suggests that similar clonal populations may predominate in primary tumours and their metastases but not exclusively. This concurs with my findings for both concordance and correlation of CNA between tumour-lymph node pairs. They also found gains at 7p, 8q and 17q were differentially detected in lymph node metastases suggesting these regions may be involved in metastasis. A limitation to this data is the lack of focal CNA analysis.

The importance of subclonal populations and their relationship to the metastasis is highlighted in this work. A number of studies in different solid cancers have demonstrated that metastasis can arise from a non-dominant subclone within the primary tumour (Gronwald *et al.*, 1999, Talmadge, 2007). Other work has suggested that the metastatic subclone becomes dominant in the metastasis (Waghorne *et al.*, 1988). As such obtaining genomic profiles of the actual metastasis is vital to obtain mechanistic information abut genomic changes which predispose to metastasis. The lack of homogeneous copy number profiles amongst the metastases in my study suggests that metastases are not mono-clonal, though it is reasonable to hypothesise that the metastatic clone is dominant. However, if metastasis occurs due to a minor subclone in the primary tumour this is an important concept to understand when looking for a biomarker for metastasis.

An important potential confounding factor when considering the genomic profiles is the influence of the underlying aetiology. In the majority of these cases tobacco smoking is the obvious aetiological risk factor. Only 7 of 49 patients (14%) with metastatic HNSCC had never smoked. All other patients were either current or ex-smokers. This clinical data is collected retrospectively from the PPM database so it limited in terms of drawing more detailed data of pack/years and arguably not as accurate as prospectively collected data. As all samples were from the oral cavity rather than the oropharynx, HPV is less likely to predominate and NGS was also used to simultaneously determine the HPV-status by calculating the precise viral load of each sample.

Comparison of CNAs associated with metastasis (as identified using GISTIC in my study) is made in Table **4-15**. The lack of concordance between studies is marked. Though perhaps attributable to the differing technologies and analytical methods used, it is also likely a reflection of the heterogeneous genomic nature of HNSCC. The small number of CNAs in these listed compared to the overall high number of aberrations seen across these tumours suggests that the differences between a metastatic and non-metastatic tumour are relatively small.

An important variation in the use of GISTIC as compared to producing CNAnorm cumulative frequency plots is the use of a generic CNA calling threshold in GISTIC, whilst in developing the cumulative frequency plots individual thresholds were assigned to each sampled to account for variation in tumour DNA content and clonal heterogeneity. As such it may be that GISTIC is not as sensitive as visual inspection of the cumulative frequency plots. However, on comparing the metastatic primary tumours to their matched metastases, little difference is seen. This suggests that if a metastatic genomic signature is visible in the nodal metastasis it would be also be visible in the metastatic primary tumour.

On comparing non-metastatic primary tumours to nodal metastases the differences are more marked. It also allowed CNAs present at significant levels in metastases to be excluded if they were also present at similar levels in non-metastatic primary tumours (e.g. gain of 3q26.2, present in 26/49 (53%) of nodal metastases and 12/26 (48%) of non-metastatic primary tumours).

In examining the genes contained within the altered regions in our studies it is important to remember that whilst many studies have demonstrated correlations between CNA and levels of gene expression they do not correlate universally nor is all altered gene expression uniquely mapped to DNA copy number (Bussey *et al.*, 2006, Cheng *et al.*, 2012, Lu *et al.*, 2011, Yang *et al.*, 2007, Myllykangas *et al.*, 2008, Jarvinen *et al.*, 2008, Xu *et al.*, 2010). Genes recurrently altered within the minimally altered regions may represent driver genes within this sample group. The fact that no gene was exclusively altered in all the samples of one group compared to another

does not preclude them being regarded as driver genes. Even lowly recurrent genes are important, particularly in attempting to identify genes important after early tumour development [Shah clonal and evolutional mutational spectrum Nature 2012].

Sethi		(Yoshioka *et al.*, 2013)		(Patmore *et al.*, 2004)		(Bockmuhl *et al.*, 2002)		(Kujawski *et al.*, 1999)	
Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss
4p15.2-p15.1 5p14.1-p15.1 7p12.1-11.2 7q21.3-q22.3 8q11.21 9p21.1 11q13.3-p13.2 13q33.1-q34 17q22-q23.3	3p14.3-p14.1 4p15.1-14 3p26.3-p26.11 5q33.1-q33.3 7q31.1-q35 8p23.3-p22 2q34-q37.1 9p24.3-p24.2 11q23.1-25 15q26.2-q26.3 18q21.1-q21.32	7p 8q 17q	-	6q 22q	-	1q21-q22 3q24-qter 6q 7q11.2 12q12-q12 18p11.1	5q34-q35 8p12-p22 10p12 10q21-qter 11p14-p15 11q23-qter 14q21-qter		8p 9q 13

Table 4-15: Comparison of CNAs associated with nodal metastases in this study and previous studies that compared metastases to primary tumours.

Gain of 11q13.3-q13.2 was present in 22/49 (45%) of nodal metastases and only 2/26 (8%) of non-metastatic primary tumours. Gain of 11q13.3 is recognized to be a frequent CNA in HNSCC and has been associated strongly with metastasis (Sugahara *et al.*, 2011). Within the genomic co-ordinates for this minimally altered region, five genes were identified as of potential significance: *CCND1*, *FGF3*, *FGF4*, *FGF19* and *LRP5*. *CCND1* encodes for cyclin D1 and previous HNSCC studies have suggested both a role in prognosis and predicting therapeutic response in relation to both expression of cyclin D1 and amplification of *CCND1* (Feng *et al*., 2011, Zhou *et al*., 2009). *CCND1* has been reported to be co-amplified with other genes such as *CTTN (Rodrigo et al., 2009).* This could reflect the fact that many CNAs tend to be large rather than focal. Fibroblast growth factors (FGF)

have been reported to important growth factor pathways in HNSCC cell lines (Marshall *et al.*, 2011). *FGF3* and *FGF4* are considered to have a role in tumourigenesis in HNSCC and have been associated with metastasis (Muller *et al.*, 1997). Interestingly, *FGF3, FGF4* and *FGF19* were reported by Huang *et al* to not be overexpressed in the presence of genomic copy number amplification and therefore suggested not to play a driver role in cancer, in contrast to a study in hepatocellular carcinoma cell lines which found a high correlation between overexpression of *FGF19* and 11q13.3 amplicon (Huang *et al.*, 2006) (Sawey *et al.*, 2011). *LRP5* encodes a receptor involved in the WNT-signalling pathway, which has been identified as a key oncogenic pathway in cancer stem cells (CSCs) in several solid cancers with a potential role in self-renewal of HNSCC CSCs (de Sousa *et al.*, 2011, Vermeulen *et al.*, 2010, Monteiro *et al.*, 2014, Da Forno *et al.*, 2008, Lee *et al.*, 2014).

Loss of 3p14.3-p14.1 was present in 35/49 (71%) of nodal metastases and 10/26 (39%) of non-metastatic primary tumours. This was often co-deleted with another minimally altered region 3p26.3-p26.11. Allelic loss of 3p has been long recognised in HNSCC and associated with poorer survival, tobacco-related disease and nodal status (Dasgupta *et al.*, 2002, Maestro *et al.*, 1993, Gross *et al.*, 2014). Within the genes identified at 3p14.3-p14.1, *FHIT*, represents the most well known tumour suppressor gene. In addition to being one of the largest known genes (~1.5 Mb) it also occurs at the most active common fragile chromosome region (FRA3B) (Pekarsky *et al.*, 2002). It encodes for a histadine triad protein (diadenosine 5',5'''-P1,P4-tetraphosphate ($A_{p4}A$) hydrolase) and is thought to play a key role in purine metabolism, though its precise function is poorly understood (Pekarsky *et al.*, 2002). *FHIT* expression has been found to be reduced or lost in oral, oesophageal, colon, cervical and breast cancer (Huebner and Croce, 2003). It has been associated with reduced survival in HNSCC (Dasgupta *et al.*, 2002). A recent study in lung cancer, using cell lines and murine models found that enforced expression of *FHIT* suppressed metastasis as well as inhibited epithelial-mesenchymal transition (thought to be a critical process during metastasis) (Suh *et al.*, 2014). Though less common in non-metastatic primary tumours, this gene deletion still occurs in 39% and

therefore is less useful as a discriminatory marker for metastasis (hence it's exclusion from the CNA panel in section 4.2.4). It's functions are also unlikely to be purely related to metastasis. This is highlighted by the fact that aberrant FHIT levels have been described in premalignant oral lesions using RT-PCR (Tanimoto *et al.*, 2000).

Loss of 11q23.1-q25 was present in 22/49 (45%) of nodal metastases and not found in non-metastatic primary tumours. Fifteen genes were identified after cross-referencing lists of interest, within this minimally altered region. *TP53AIP1* is a p53 target, which produces 3 isoform transcripts which regulate mitochondrial membrane potential and the mitochondrial apoptotic pathways (Matsuda *et al.*, 2002). As such this gene is a potential mediator of p53-dependent apoptosis, and overexpression of *TP53AIP1* alone induces apoptosis. Given the near universal finding of *TP53* mutation in the TCGA HNSCC subgroups report finding, decreased expression of this has been associated with lymph node metastasis and decreased survival in non-small cell lung cancer (Yamashita *et al.*, 2008). In HNSCC this gene has not been explored specifically. This could represent a targetable event, as demonstrated by a study using a replication defective adenovirus Ad-p53AIP1 (Jiang *et al.*, 2010). This found infection of liver cancer (HepG2) cells induced apoptosis and cell cycle arrest. Any functional role in metastasis specifically has not been elucidated yet.

CHEK1 encodes for checkpoint kinase 1 which belongs to the Ser/Thr protein kinase family. This mediates cell cycle arrest in response to DNA damage. *CHEK1* is regulated by *ATR* forming the ATR/CHEK1 pathway. Upregulation of this pathway has been reported in a subset of HNSCC and hypothesised to occur as a compensatory response to loss of the distal 11q (Sankunny *et al.*, 2014). These findings have been associated to radioresistance in OSCC cell lines, allowing cells to avoid the cell-cycle checkpoint relating to DNA damage. Knocking down the ATR/CHEK1 pathway has subsequently been demonstrated to increase radiosensitivity in vitro (Sankunny *et al.*, 2014).

EI24 is another proapoptotic p53-target gene that plays an important role in negatively controlling cell growth and division. Overexpression of *EI24*

induces apoptosis (Gu *et al.*, 2000b, Gu *et al.*, 2000a). Ectopic *EI24* expression has been shown to inhibit cell colony formation and to also induce apoptosis (Gu *et al.*, 2000a). Though it's precise mechanistic function is unclear recent evidence has shown that EI24 inhibits nuclear protein import via an importin β-binding-like domain. As such it can inhibit p53 nuclear import demonstrating it is a key effector of *TP53* (Lieu *et al.*, 2014). Choi *et al* demonstrated that *EI24*, in cell lines and murine models, acts as a negative regulator of TRAF2 signalling by inducing its degradation leading to NF-κB transcriptional activity (Choi *et al.*, 2013). Decreased expression of *EI24* induced EMT in epithelial cancer cells as well as increasing cell motility, invasiveness and resistance to apoptosis whilst overexpression resulted in the opposite characteristics (Choi *et al.*, 2013). Choi *et al* also found that overexpression of *EI24* and copy number gain of the *EI24* gene correlated with invasiveness and metastasis in breast tumour clinical samples (Choi *et al.*, 2013).

The distal region of 11q clearly contains genes that interact with *TP53*. Given the already established high frequency of *TP53* mutation in non-HPV associated HNSCC it is not surprising that this genomic loss is found frequently in HNSCC. As at least two of the genes are targets of p53 that appear to enhance its proapoptotic effects it is consistent that loss of distal 11q would be associated with metastasis and hence a worse prognosis.

4.3.3.2 CNA associated with ECS

No CNA was found exclusively in metastases with ECS. Loss of 18q21.1-q21.32 was found in 30/49 (59%) of nodal metastases and only 1/26 (4%) of non-metastatic primary tumours. Within the nodal metastases it was present in 9/22 (41%) of metastases without ECS and 21/27 (78%) of metastases with ECS. Three genes of potential interest were identified in this region (*SMAD4*, *MALT1*, *PMAIP1*). *SMAD4* was originally identified as a tumour suppressor gene in pancreatic cancer, though subsequently has been identified in numerous solid cancers (Schutte *et al.*, 1996, Hahn *et al.*, 1996). It plays a key role in mediating signals in both the WNT-signalling pathway and the TGF-β signalling pathway (Nishita *et al.*, 2000). SMAD4 forms complexes with other SMAD proteins which then regulate gene expression

of targets involved in cell death, proliferation and inflammation(Bornstein *et al.*, 2009). TGF-βs are proteins encoded by the TGF-β superfamily of genes that regulate processes including cell proliferation, leucocyte infiltration, regulate the extracellular matrix. The components of this pathway are considered to act as tumour suppressors with SMAD4 being a common mediator to the signalling pathway (Korc, 2009). A study using murine models found that *SMAD4* deletion led to development of spontaneous HNSCC in 74% of mice. They also found in clinical samples decreased expression of *SMAD4* in 86% of HNSCC tissue samples as well as in 67% of the adjacent morphologically normal buccal mucosa (Bornstein *et al.*, 2009). Both of these findings suggested a key role for *SMAD4* as a gatekeeper gene in HNSCC (Korc, 2009). Recently Liu *et al* reported that ablation of *SMAD4* in murine models resulted in activation of the ERBB2 pathway (Liu *et al.*, 2015a). Activation of this pathway has been shown to confer resistance to the EGFR antibody cetuximab in cell lines (Yonesaka *et al.*, 2011). Loss of *SMAD4* has also been associated with metastasis and poor prognosis in colorectal cancer. It has been found to cause alteration of the BMP signalling pathway that causes it to be metastasis promoting in murine models (Voorneveld *et al.*, 2014). An additional study by Liu et al found that in colon cancer cell lines with loss of *SMAD4*, inducing *SMAD4* expression led to reduced migration and invasive ability, reducing VEGF-C secretion and lymphangiogenesis in mouse models (Liu *et al.*, 2015b).

Genes reported to have an association with ECS in HNSCC include *EGFR* and *CCND1*. Increased copy number was found to be significantly associated with 127 OSCC samples using FISH (Michikawa *et al.*, 2011). *SERPINE1* and *SMA* overexpression expression was found to be predictive of ECS in OSCC (81% sensitivity when both found to be upregulated) (Dhanda *et al.*, 2014). Wang et al identified an 11-gene expression signature (*GGH*, *MTFR1*, *CDHN3*, *PSRC1*, *SMIM3*, *CA9*, *IRX4*, *CPA3*, *ZSCAN16*, *CBX7* and *ZFP3*) significantly associated with ECS and associated with worse survival in histologically node negative patients (Wang *et al.*, 2015). No correlations with these gene regions and ECS was found in my samples. In a small study of seven samples using qRT-PCR a three gene expression panel (*CTTN*, *EEF1A1*, *MMP9*) was identified to be associated with ECS

(Zhou *et al.*, 2006). Out of these, in my study, CTTN (at 11q13.3) was amplified in 14/27 (52%) of metastases with ECS and 8/22 (36%) of metastases without ECS. It is interesting that amongst these studies there is also little cross-over in identified associated genes looking for the same clinic-pathologic outcome. This could well be related to the differing methodologies across the three studies. The sample size could also affect consistency between studies. The different geographic origins of the studies, and the inferred underlying aetiologies of OSCC, could influence trends in the genes identified. The recognised clonal heterogeneity of HNSCC could be responsible for the lack of concordance in gene expression signatures. It is also important to remember that metastasis is a complex process involving many cellular pathways and steps (Hanahan and Weinberg, 2000, Hahn and Weinberg, 2002). Combining this knowledge with the fact that recognised heterogeneity of HNSCC it is entirely possible that any genomic marker for such a diverse phenotype could be varied and wider than simply a small panel of genes.

As no single CNA is exclusively found in all samples from one specific group and they often co-occur it is unlikely that these alone represent a reliable marker for lack of metastasis. They could however be combined with the presence of CNAs associated with metastasis to produce a more robust indicator of the presence of nodal metastasis. This is proposed in the 11-CNA panel (see Table **4-13**). When scoring each sample for this signature a significantly higher score was found in metastatic tumours compared to non-metastatic primary tumours (p = 0.0001) but one non-metastatic primary tumour contained six markers from the panel and 6 metastatic primary tumours contained no markers. As such the signature does not discriminate entirely. This may well be due to both inter and intra-tumour genomic heterogeneity. It may also reflect the complex genomic interactions underlying cancer. Combining these factors could mean that using specific genomic markers will not result in a mutually exclusive marker. More detailed genomic analyses such as those performed by the head and neck sub-group of TCGA have failed to identify any new universal mutational signature for HNSCC (Cancer Genome Atlas, 2015). Gene expression

subtypes are yet to find a role in clinical decision making (Walter *et al.*, 2013).

4.3.3.3 Comparison to TCGA HNSCC subgroup data

The TCGA HNSCC subgroups data is available to the public in a post-processed format. Clinical information was downloaded from the UCSC Cancer Genomics Browser (Zhu *et al.*, 2009). This included all samples analysed by the TCGA HNSCC subgroup (n = 530). Samples IDs were obtained from this to identify oral cavity primary tumours that were HPV negative (n = 301). The pathological node positive and pathologically node positive samples out of this sub-cohort were then identified (N0, n = 132 and N+, n = 169). Segmented genomic data was then downloaded from the UCSC cancer genomic browser (N0, n = 116 and N+, n = 142). This data was then input to pheatmap (Raivo Kolde, Harvard USA), to generate genome wide heatmaps for both non-metastatic primary tumours and metastatic primary tumours. The frequency of CNAs identified in the panel associated with metastasis (see Table **4**-**13**) was assessed using the downloaded TCGA data (see Table **4**-**16**).

A similar, consistent trend of increased frequency in metastatic tumours was seen in the TCGA HNSCC data. Loss of 11q23.1-q25 was not present in non-metastatic primary tumours in my samples but is present in 12/116 (10%) of non-metastatic primary tumours in the TCGA data. This highlights the high level of inter-tumour heterogeneity and confirms that no CNA is uniquely exclusive to metastasis. The fact it was not identified in my samples is likely due to the smaller number of samples in my study.

Similarly loss of 18q21.1-q21.32 was present in 25/116 (21%) of TCGA non-metastatic primary tumours, but only 1/26 (4%) of these tumours in my samples. This again likely reflects the smaller number of samples in my study and heterogeneous nature of HNSCC. The overall similar trends in CNAs is reassuring that my data is not spurious given the robust methodology of the TCGA study. A heatmap was generated to visually assess the effectiveness of the CNA panel in selecting metastatic primary tumours (see Figure **4**-**39**).

Cytoband	TCGA cohort N0 (n = 116)		TCGA cohort N+ (n = 142)		Sethi cohort N0 (n = 26)		Sethi cohort N+ (n = 49)	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
+4p15.2-p15.1	3	3	9	6	1	4	5	10
-4p15.1-p14	14	11	30	21	1	4	7	14
-5q33.1-q33.3	16	14	40	28	4	15	18	37
+7p12.1-p11.2	29	25	47	33	6	23	26	53
-7q31.1-q35	7	6	16	11	1	4	7	14
-8p23.3-p22	21	18	69	49	3	12	14	29
-9p24.3-p24.2	15	13	50	35	3	12	15	31
+9p21.1	12	10	29	20	1	4	9	18
+11q13.3-q13.2	27	23	60	42	2	8	22	45
-11q23.1-q25	12	10	45	32	0	0	22	45
+13.q33-q34	9	8	21	15	1	4	9	18
-15q26.2-q26.3	3	3	10	7	1	4	10	20
+17q22-q23.3	5	4	12	9	3	12	15	31
-18q21.1-q21.32	25	21	62	44	1	4	29	59

Table 4-16: Comparison of frequency of CNA panel identified as associated with metastasis in TCGA HNSCC data.
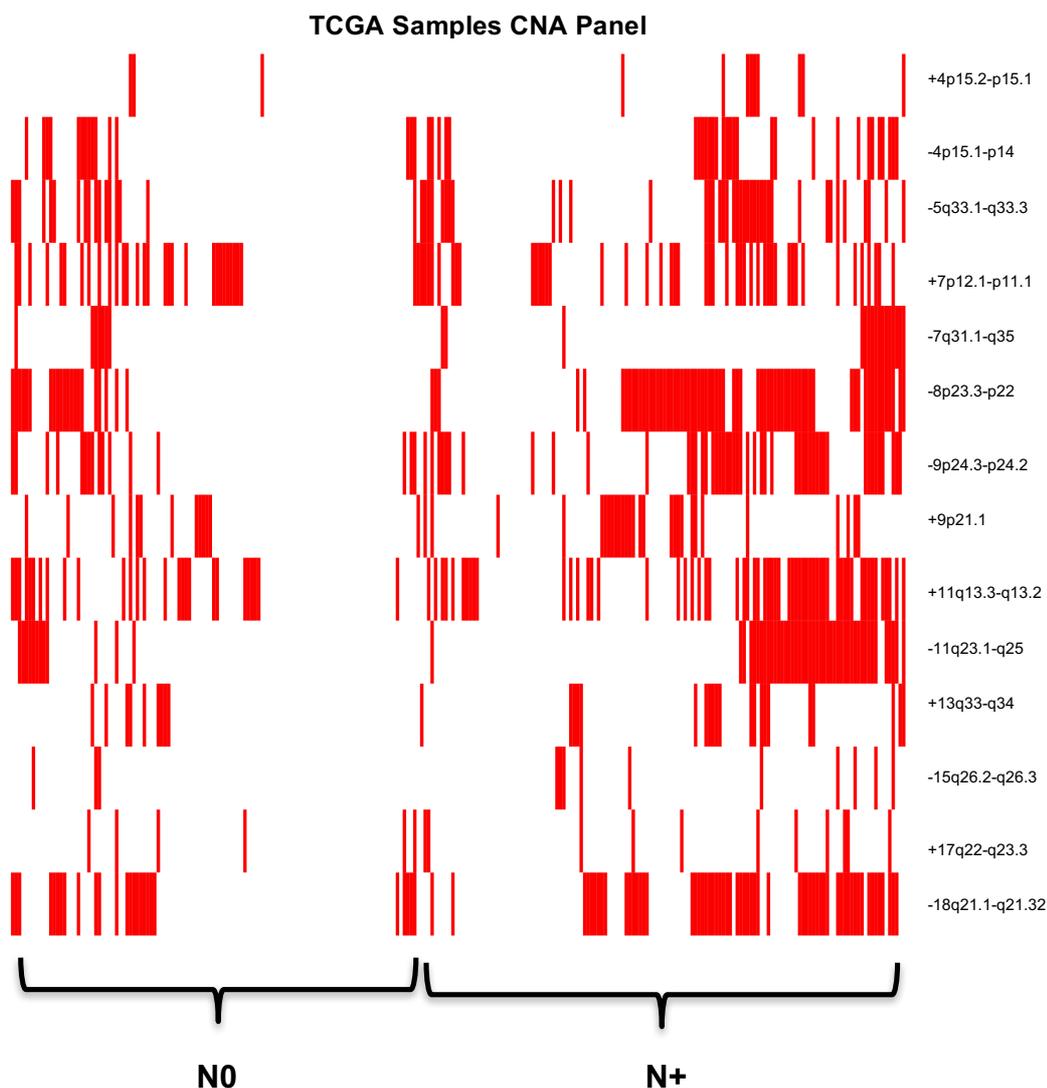
Figure 4-39: Heatmap generated using CNA panel associated with metastasis using TCGA HNSCC subgroup data. Red denotes the presence of the CNA whilst white denotes the absence of the specific CNA.

Inspection of the heatmap generated using the CNA panel suggests it does not separate metastatic vs. non-metastatic primary tumours as strongly as in my samples. The TCGA samples were then scored for the presence/absence of each CNA as previously (present = 1, absent = 0). The scores were analysed using the Shapiro-Wilk test for normality demonstrating they were not normally distributed. Therefore the mean scores for non-metastatic primary tumours and metastatic primary tumours were compared using the Mann Whitney test (see Figure **4-40**). This found the mean score for metastatic primary tumours to be significantly higher (p = <0.0001). Though the mean scores is significantly different in each group this panel fails to discriminate completely between metastatic and non-

metastatic primary tumours. This may reflect the wide inter-tumour heterogeneity in HNSCC.
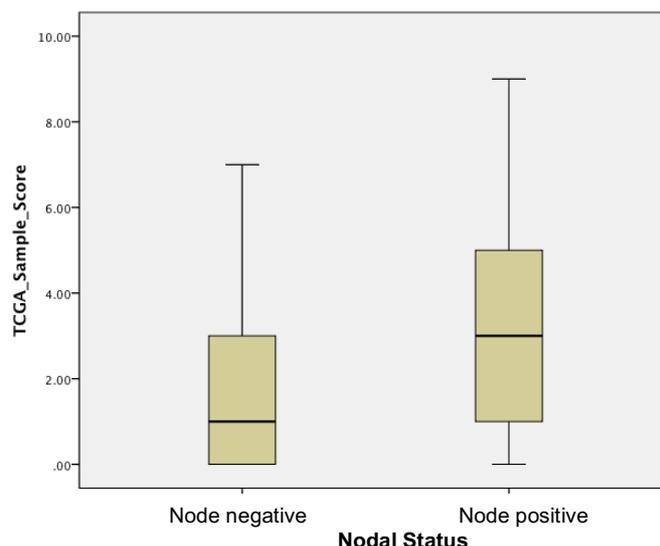
Figure 4-40: Box-plot comparing CNA panel scores for metastatic and non-metastatic primary tumours identified in the TCGA cohort.

One potential confounding factor in using the TCGA data with pheatmap is the fact that a generic, group wide threshold was applied to their data, rather than an individual sample threshold as in my samples. This may increase the amount of background noise being included in the heatmap. However these samples were processed using GISTIC 2.0 by the TCGA group which utilises a generic threshold and as such it is not unreasonable to take this approach. Inspecting the group as a whole it was observed that canonical CNAs such as loss of 3p were present in appropriate proportions in each group.

4.3.3.4 FGA as a biomarker for nodal status

Taking a more generic view of genomic damage may represent a more useful method of stratifying patients risk. No difference was found in the mean number of segments in different groups. Using this a sole inference of genomic damage is actually a flawed approach as CNAs in HNSCC are generally broad. This means that large segments of the genome may be altered with relatively few breakpoints compared to another tumour where an increased number of smaller genomic regions are altered. This could a give

a false impression that there are reduced genomic changes when in fact the converse is possible. Therefore using the FGA is likely to be a more reliable reflection of genomic damage. This approach has been found to predict survival in patients with metastatic colorectal carcinoma (Mehta *et al.*, 2005). In different prognostic subtypes of breast cancer the FGA has been found to vary significantly (Jonsson *et al.*, 2010).

As the mean FGA in metastatic primary tumours was found to be significantly higher than in non-metastatic primary tumours (p = 0.0001) a further analysis was performed to attempt to determine a cut-off for this variable as a predictor for nodal status. An open access web application "Cutoff Finder" was used for this purpose as it provided multiple methods for different data types on one platform (Budczies *et al.*, 2012). Using Fisher's Exact text to calculate the significance of correlation between FGA and nodal status as a binary variable (N0 = 0, N+ve = 1) Cutoff Finder determined the optimal value to be 0.04. This generated a sensitivity of 98% but a specificity of 42.3% (see Figure **4-41**).
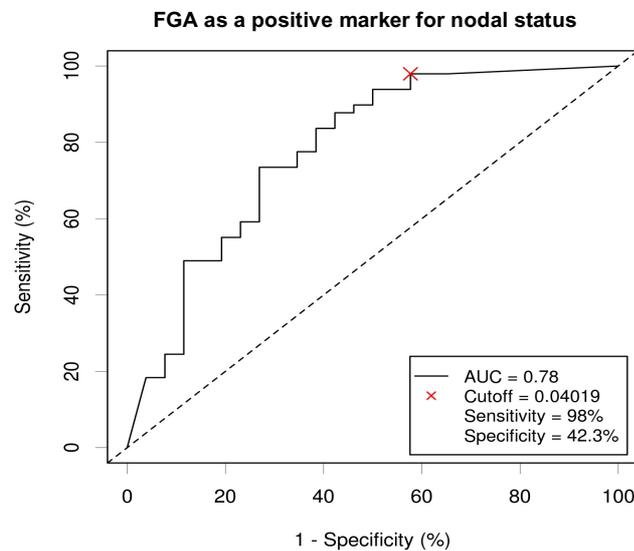
Figure 4-41: ROC for FGA as a marker for nodal metastases.

Cutoff Finder also allowed analysis of the significance of correlation with survival. For disease-free survival it found a significant decrease in disease-free survival using an FGA cut-off of 0.098, with a hazard ratio of 5.21 (1.2-22.69) for an FGA > 0.098 (see Figure **4-42**). For overall survival Cutoff

Finder identified a significant decrease in overall survival for patients with an FGA > 0.135 (p = 0.0059) (see Figure **4-43**).
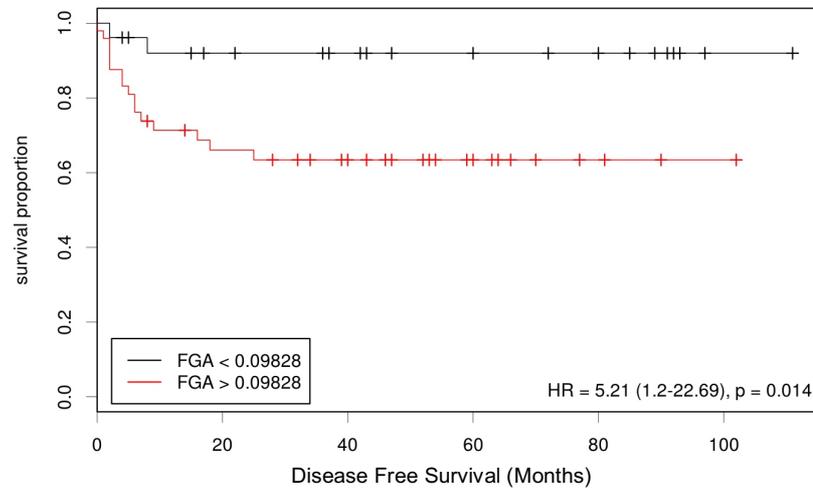
Figure 4-42: Kaplan Meier plot demonstrating reduced disease free survival for patients with an FGA > 0.098 (p = 0.014).
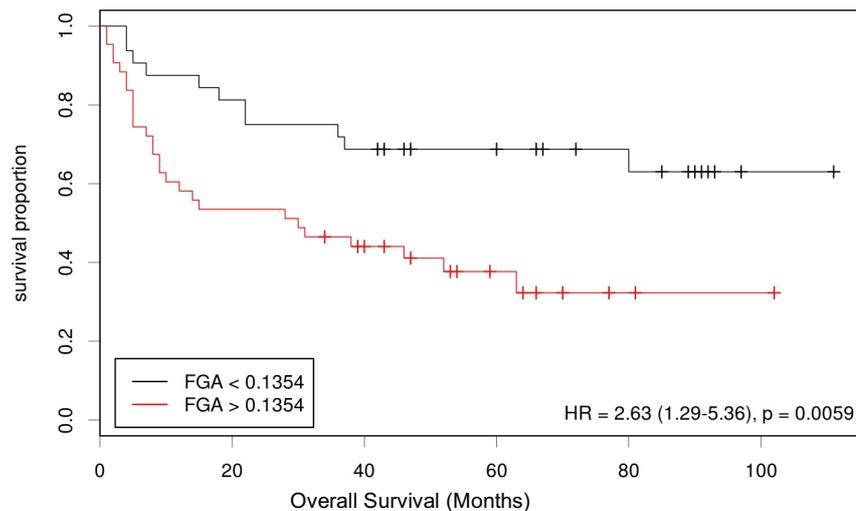
Figure 4-43: Kaplan Meier plot for overall survival demonstrating a significant decrease in patients with FGA > 0.1354 (p = 0.0059).

The potential impact of intra-tumour heterogeneity is difficult to estimate. It is possible that the CNA signature or FGA could vary markedly dependent on the area of tumour sampled for DNA. Clonal heterogeneity has been demonstrated in studies sampling multiple topographically distinct areas of tumour in numerous cancers (McGranahan and Swanton, 2015). Proposed models of metastasis have suggested it could occur from relatively small

clonal populations with the primary tumour supporting the need to analyse metastases so as to identify any metastatic signature (Hynes, 2003). The similarity of nodal metastases to the primary tumour seen in this study suggests that this may not be the situation in every case and that the metastatic clone is often a significant population within the primary tumour. In future work it would be useful to evaluate the impact of this by using multiple sampling and comparing this to the genomic signatures obtained by taking one large, relatively untargeted sample. Indeed the clinical utility of genomic markers relies on the reproducibility and reliability of any signature identified. If these are limited by esoteric technical considerations, then the value of any purported signature is reduced. In taking any work forward to test a genomic signature the impact of intra-tumour heterogeneity must be taken into account.

4.4 Conclusion

In relation to the aims of this chapter:

1. NGS CNA data was used to compare metastatic primary tumours to their matched nodal metastasis. This found that largely metastases were similar to their parent tumour, though not identical.

2. In comparing CNA data between metastatic and non-metastatic primary tumours no single CNA is mutually exclusive in a single clinicopathologic group. A 14-CNA panel was identified as being significantly associated with metastasis. The FGA was found to also be significantly increased in metastatic primary tumours compared to non-metastatic tumours. These measures could be tested as a predictive marker for metastasis.

3. When comparing CNA data between tumours associated with and without nodal ECS, a correlation was found between the presence of ECS and loss of 18q21.1-q21.32. Within this region the gene *SMAD4* represents   a   possible   driver   gene   for   this   phenotype.

Chapter 5

Copy number analysis of OPSCC and comparison to viral load

5.1 Introduction

The association between HNSCC and high-risk human papillomavirus sub-types (most commonly type 16) is well recognised (Gillison *et al.*, 2008, Vidal and Gillison, 2008, Gillison and Lowy, 2004, Gillison, 2004). This association is predominantly in oropharyngeal SCC (OPSCC) with poorly differentiated, basaloid histological characteristics (Gillison, 2004). It is characterised by the detection of HPV DNA, RNA or expression of p16 protein (Venuti and Paolini, 2012). This subgroup of patients has been demonstrated in clinical trials to have a significantly better prognosis compared to patients with HPV-negative tumours (Ang *et al.*, 2010). In fact, HPV-status appears to be the single strongest prognosticator in OPSCC (Rischin *et al.*, 2010, Fakhry *et al.*, 2008). The prevalence of HPV-positive OPSCC is increasing (Mehanna *et al.*, 2013). The prognostic effect of HPV-status is not homogeneous. A small subset of HPV-positive OPSCCs has a poor treatment response and clinical outcome. The reasons for this are unclear and a reliable method for identifying these patients has not been established (Ang *et al.*, 2010). The prevalence of HPV in head and neck cancer subsites outside the oropharynx appears low (< 10%, compared to ~60% of OPSCC) (Upile *et al.*, 2014, Combes and Franceschi, 2014). The beneficial prognostic impact of HPV status also appears to be lost in subsites outside the oropharynx (Lassen *et al.*, 2014).

Targeted exome sequencing revealed a similar mutational burden of HPV-positive and negative HNSCC but differences in the spectrum of mutations (including unique mutations in *DDX3X*, *FGFR2/3*) (Seiwert *et al.*, 2015). Gene expression analysis has shown a different profile for HPV-positive and negative OPSCC, which may well explain the difference in response to

chemoradiation (Lohavanichbutr *et al.*, 2009, Thibodeau *et al.*, 2014). Surprisingly little overlap has been found in gene expression studies demonstrating that even HPV-positive HNSCC is a genomically heterogeneous disease (Weinberger *et al.*, 2006, Slebos *et al.*, 2006, Schlecht *et al.*, 2007, Martinez *et al.*, 2007).

The low-coverage NGS approach taken to provide copy number profiles in my study (see Chapter 4) is much cheaper than mutational analysis or gene expression studies. It also works wells with FFPE-tissue. In addition to providing copy number data it also enables determination of the presence of any viral DNA sequences (Conway *et al.*, 2012, Wood *et al.*, 2012). By then relating the quantity of viral sequences detected to the depth of sequencing coverage the viral load can be determined (Conway *et al.*, 2012). This method was used to determine the HPV-status of the OSCC samples included in my project. Briefly, the number of reads aligning to the human genome was used to calculate read depth in terms of reads per kilobase (Kb). Following this, the number of reads uniquely aligning to viral genomes was counted (this is proportional to the number of Kb viral sequence per human genome, and hence the number of viral genomes per human genome). Given a known number of human genome reads, the possible viral load that could be detected with 95% confidence is:

$$\frac{\text{(no. of viral reads x 6 x } 10^9 \text{ bp diploid human genome)}}{\text{(7900 bp viral genome x number of human reads)}}$$

There is a need for a reliable method of not just identifying HPV-positive HNSCC, but recognising those, which will benefit from the positive prognostic effect of HPV-status (i.e. those which are truly "HPV-driven"). This needs to be a swift and affordable test. The genomic signature of these tumours is still incompletely understood, and may well provide important, translational biomarkers.

The Precancer Genomics group has previously performed low-coverage NGS on a group of HNSCC patients. Part of this group underwent determination of HPV-status by DNA-PCR and p16 immunohistochemistry (IHC) as well as NGS viral load calculation and was published by Conway *et al* (Conway *et al.,* 2012). This paper concluded that NGS was comparable in

sensitivity and specificity to other methods for detecting HPV-status. Another study by Lechner *et al* performed targeted exome sequencing of OPSCC samples finding 100% concordance between sequencing and HPV status determined by p16 IHC and PCR (Lechner *et al.*, 2013). By using the computational techniques I acquired in analysing the copy number profiles of my patient cohort I set out to evaluate if the copy number profile of HPV-positive and negative OPSCC varied according to NGS-determined viral load.

5.2 Aim

The aim of this work was to:

1. Produce a specific genome-wide CNV profile for HPV-positive OPSCC.
2. To evaluate if there is a specific viral load at which this genomic profile becomes apparent.

5.3 Results

5.3.1 Patient Groups

Patients for this study were collected from the Precancer Genomics tissue bank database. The inclusion criteria for these patients were no history of previous HNSCC or radiation and a confirmed oropharyngeal SCC primary tumour. In total 51 patients were identified (see Table **5**-**1**). They included 13 patients that were previously part of a larger separate study published by Conway *et al*, as part of the Precancer Genomics group, University of Leeds (Conway *et al.*, 2012). They compared NGS to DNA-PCR and p16 IHC as a method of assessing HPV-status of a group of HNSCC (Conway *et al.*, 2012).

Primary tumour site	Base of tongue	Tonsil	Posterior wall of pharynx	Soft palate
No. of patients	23	13	4	11

Table 5-1: Table of all OPSCC tumour subsites.

These samples are detailed in Table **5**-**2**. Each primary tumour from these patients underwent DNA extraction according to the same protocol described in Chapter 2.5. They were sequenced either on the Illumina Genome analyser II at 76-bp length reads multiplexing 10 samples per lane or the Illumina HiSeq 2500 at paired end 100 bp length reads. This produced approximately 0.033x – 0.33x coverage of the human genome.

Sample	HPV-16 Viral Load
DH007	0
DH008	0
DH021	0
DH022	0
DH035	0
PG027-T-1	0
PG033-T-3	0
PG092-T-1	0
PG100-T-1	0
PG109-T-1	0
PG113-T-1	0
PG116-T-1	0
PG118-T-1	0
PG129-T-1	0
PG145-T-1	0
PG154-T-1	0
PG161-T-1	0
PG187-T-1	0
PG189-T-1	0
PG190-T-1	0
RG0011_153	0
RG0012_154	0
RG0021B_15G	0
RG0032_15M	0
RG0036_1SU	0
RG026_165	0
RG033_15P	0
RG040WS_169	0
RG056_15Z	0
RG53_16D	0
RG54_16E	0
RG056_1SZ	0
PG133-T-2	0.13
PG096-T-1	0.22
DH034_0S	1.13
PG069-BX2	1.33
PG198-T-1	5.71
RG055_16G	9.00
PG156-T-2	11.63
RG41_16B	12.00
PG139-T-4	21.83
DH001	21.98
PG194-T-1	25.86
DH033	32.26
DH028	37.09
RG0015_156	39.00
DH047	47.54
RG0002B_150	48.00
DH030	66.49
DH004	70.60
DH020	75.42
PG106-BX1	75.94

Table 5-2: List of samples and viral load (calculated as viral genomes per human diploid genome).

5.3.2 Cumulative frequency karyograms

The FASTQ files produced from these sequencing experiments were processed as described in Chapter 2.9. The CNAnorm cumulative frequency karyograms were generated for the zero viral load tumours and those with a viral load > 0 (see Figure **5**-**1**).

In comparing the cumulative CNA profiles for the oropharyngeal tumours with viral load of 0 (n = 31) and those with a viral load > 0 (n = 20), there are several striking differences (see Figure **5**-**1**). Firstly there are fewer overall areas of chromosomal gain and loss seen in tumours with a viral load > 0, suggesting a lower amount of genomic damage and rearrangement.

(a)
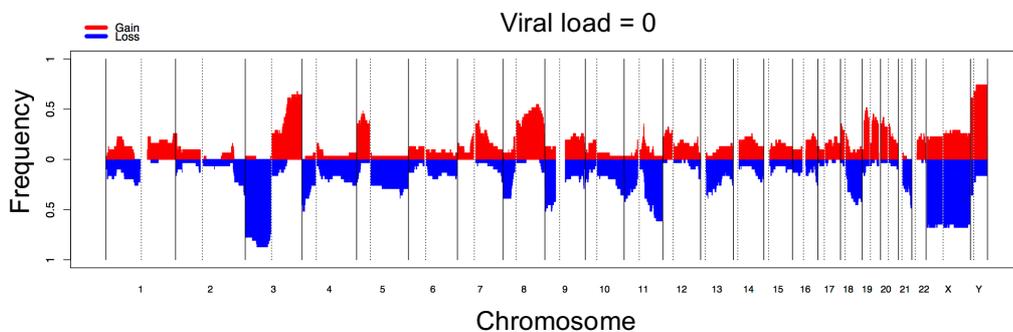
(b)
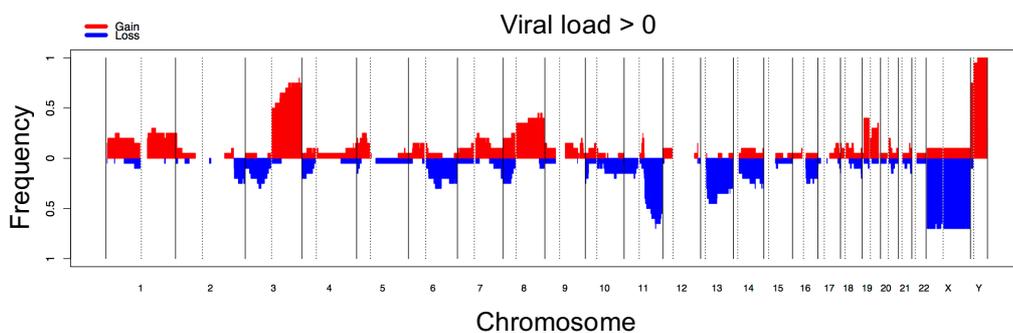
Figure 5-1: Cumulative frequency plots for the (a) Viral load = 0 group and (b) Viral load > 0 group. Regions of gain are represented in red and regions of loss in blue.

Tumours with a viral load of 0 showed a far higher frequency of 3p loss (27/31 (87%)), with few tumours with a viral load > 0 (5/20 (25%)) demonstrating 3p loss. Loss of 9p is seen in 15/31 (48%) of tumours with no viral load and 1/20 (5%) of those with a viral load > 0. Loss of 18q is seen in

13/31 (42%) of tumours with no viral load and 3/20 (15%) of those with a viral load > 0. Other regions of CNA with a visually detectable difference between these two groups are shown in Table **5**-**3**.

	No of samples in each group (%)	
Region	**0 Viral load group (n = 31)**	**Viral load > 0 group (n = 20)**
- 3p	27 (87%)	5 (25%)
- 4q	7 (23%)	1 (5%)
+ 5p	12 (39%)	5 (25%)
- 5q	10 (32%)	1 (5%)
- 9p	15 (48%)	1 (5%)
- 11p	11 (35%)	3 (15%)
+ 12p	10 (32%)	2 (10%)
+ 12q	6 (19%)	1 (5%)
+ 15q	6 (19%)	1 (5%)
- 18q	13 (42%)	3 (15%)
+ 20p	12 (39%)	3 (15%)
+ 20p	12 (39%)	3 (15%)
+ 20q	8 (26%)	1 (5%)
- 21q	12 (39%)	3 (15%)

Table 5-3: List of differential CNA between the two viral load groups found on visual inspection of cumulative frequency karyograms.

5.3.3 GISTIC focal analysis of oropharyngeal tumours with no detectable viral load and those with a detectable viral load

GISTIC was used to analyse the same sample groups to produce a simultaneous focal and broad CNA. The input files for this were created as described in Chapter 2.9.2.3. The results of the focal analysis are shown in Figure **5**-**2** and Figure **5**-**3**). From these analyses the focal amplification profiles can be seen to be largely similar between the two groups. It is important to remember that telomeric and centromeric regions are unreliable. Using the GISTIC focal plots alone suggested the tumours with 0 viral load do possess two unique focal CNAs (gain of 15p15.1 and 8q23.3). However, examination of the sample karyograms revealed that these CNAs are found at differing rates in both groups. Gain of 15p.1 is found in 13/31 (42%) of the 0 viral load group and 3/20 (15%) of the >0 viral load. Gain of 8q23.3 was present 17/31 (55%) of 0 viral load tumours and 9/20 (45%) of the > 0 viral load group.
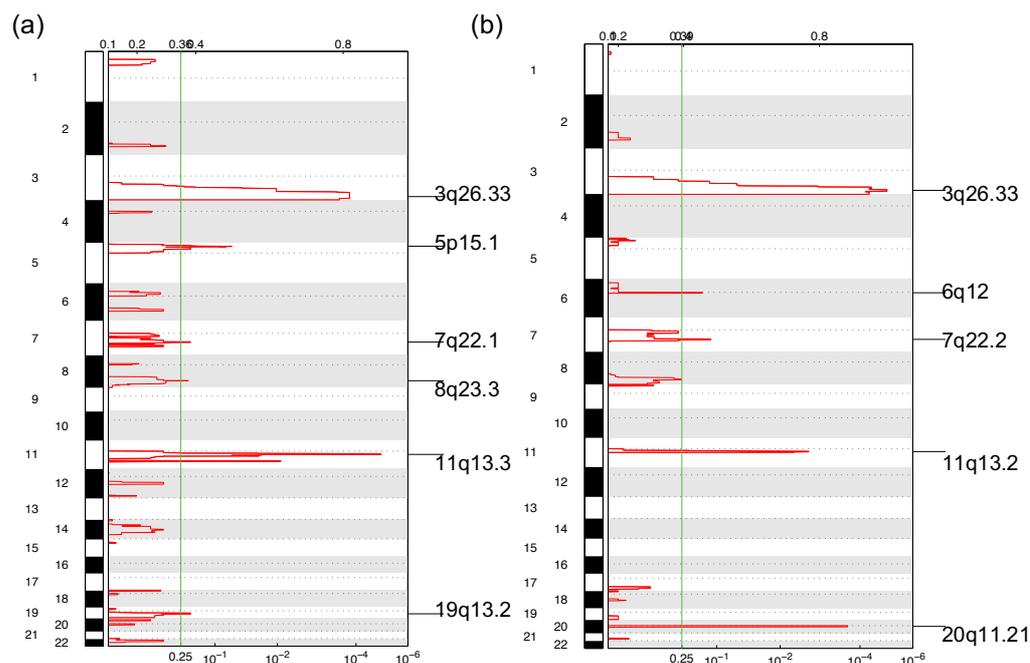
Figure 5-2: Genome-wide focal amplification plots for tumours with 0 viral load (a) and tumours with viral load >0 (b). The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line marking the significance threshold at which the analysis was performed. The top x-axis shows the G-score for each CNA.

Comparison of the focal deletion profiles for these two groups using GISTIC alone reveals a small number of differences. No focal deletions are unique to one group alone. Focal losses that are identified as significant by GISTIC in the tumours with 0 viral load only include loss of 4p16.2 (found in 16/31 (52%) tumours with 0 viral load and 4/20 (20%) tumours with > 0 viral load), loss of 6p22.3 (found in 4 tumours with 0 viral load and 1 tumour with viral load > 0) and loss of 8p22 (found in 12/31 (39%) tumours with 0 viral load compared to 4/20 (20%) tumours with > 0 viral load).

Focal deletions identified by GISTIC as being only in the tumours with viral load > 0 included loss of 6q26. However examination of the individual karyograms found loss of 6q26 in 7/31 (23%) tumours with viral load > 0 and 4/20 (20%) with 0 viral load. Closer inspection of the wide peak boundaries of loss of 11q23.1 in tumours with > 0 viral load and 11q25 (identified by GISTIC in tumours with 0 viral load) found them to actually overlap. Deletion of the region 11q23.1-q25 occurred in 17/31 (55%) of 0 viral load tumours and 14/20 (70%) of the > 0 viral load group.
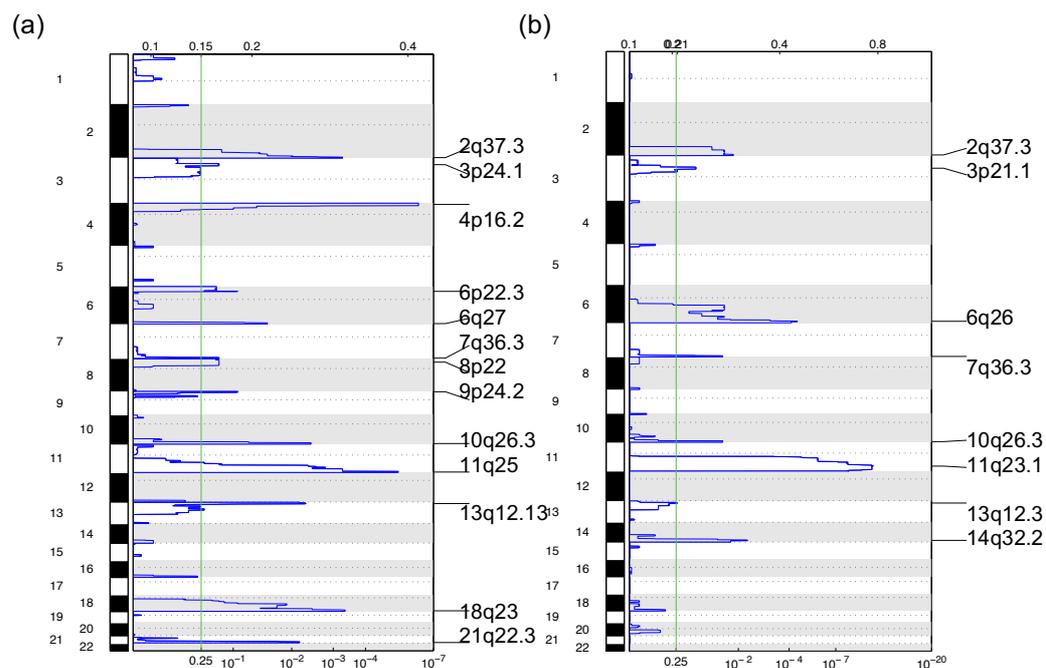
Figure 5-3: Genome-wide focal deletion plots for tumours with 0 viral load (a) and tumours with viral load >0 (b). The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line marking the significance threshold at which the analysis was performed. The top x-axis shows the G-score for each CNA.
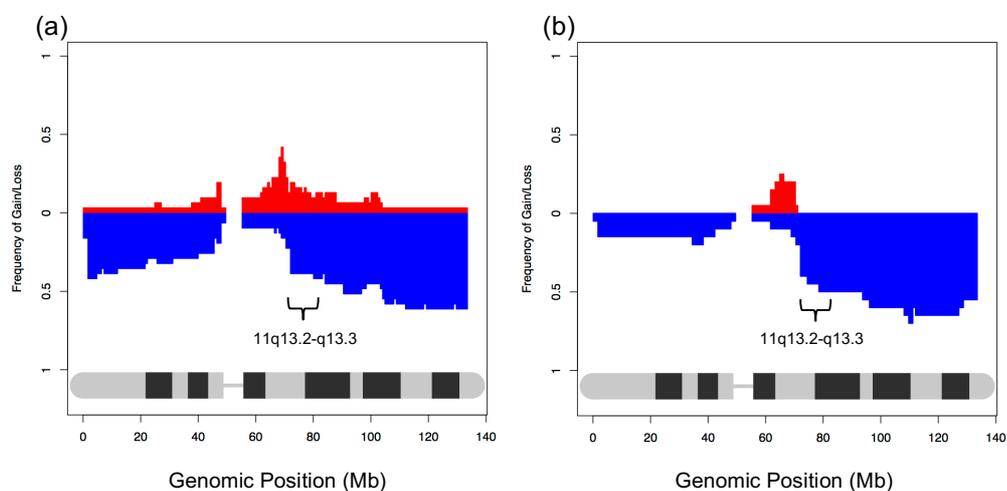
Figure 5-4: Example of two adjacent focal deletions identified using GISTIC. The 0 viral load group is on the left (a) and tumours with viral load >0 on the right (b). Inspection of the CNAnorm generated cumulative karyograms for the individual chromosomes allows the region of deletion to be seen to be actually overlapping the two foci.

GISTIC suggests focal regions of potential interest. However, the wide peak boundaries for these remain large and GISTIC does not provide an accurate reflection of the frequency of these CNAs in one group relative to another. It also does not provide information on adjacent regions in different groups and

if wide peak boundaries overlap between groups. Though CNAnorm cumulative frequency karyograms provide some information on overlapping regions (see Figure **5-4**), resolution is limited. Broad CNAs of similar or higher frequency to focal CNAs may obscure or completely mask the smaller altered region. To reliably identify the minimally altered regions an alternative is required.

5.3.4 Evaluation of minimally altered regions between OPSCC tumours with 0 viral load and viral load > 0

In order to try and identify potential candidate genes of interest it is necessary to identify the smallest region recurrently altered (the minimally altered region). By using open access software called pheatmap (Raivo Kolde, Harvard USA) and combining this with an R-based script called genomeHeat (Dr Henry Wood, Precancer genomics) I was able to create heatmaps for these samples, utilising the CNAnormout.txt files as input. This software provided greater resolution than the CNAnorm-generated cumulative frequency plots when attempting to identify the minimally altered region.

Firstly, I instructed the software to create a heatmap of all samples allowing the software to cluster samples according to CNA profile (see Figure **5-5**).

Then, potential focal regions of interest were identified by visual inspection of the CNAnorm cumulative frequency karyograms. The chromosomal arms with the greatest differential frequency of CNA between the groups were selected. As demonstrated in Table **5-3** loss of 3p occurred with the highest frequency in the 0 viral load group (27/31 (87%)) and also with the largest difference in frequency when compared to the > 0 viral load group (25%). pheatmap allows generation of a sequential cluster heatmaps that can be used to gradually zoom in on the minimally altered region (see Figure **5-6**). These images allow simultaneous assessment of the individual patients and frequency of the minimally altered region as well as the ability to identify samples according to viral load status. Due to the resolution and layout of the image focal CNAs of equivalent or lower frequency to broader CNAs are not obscured by the broader CNA.
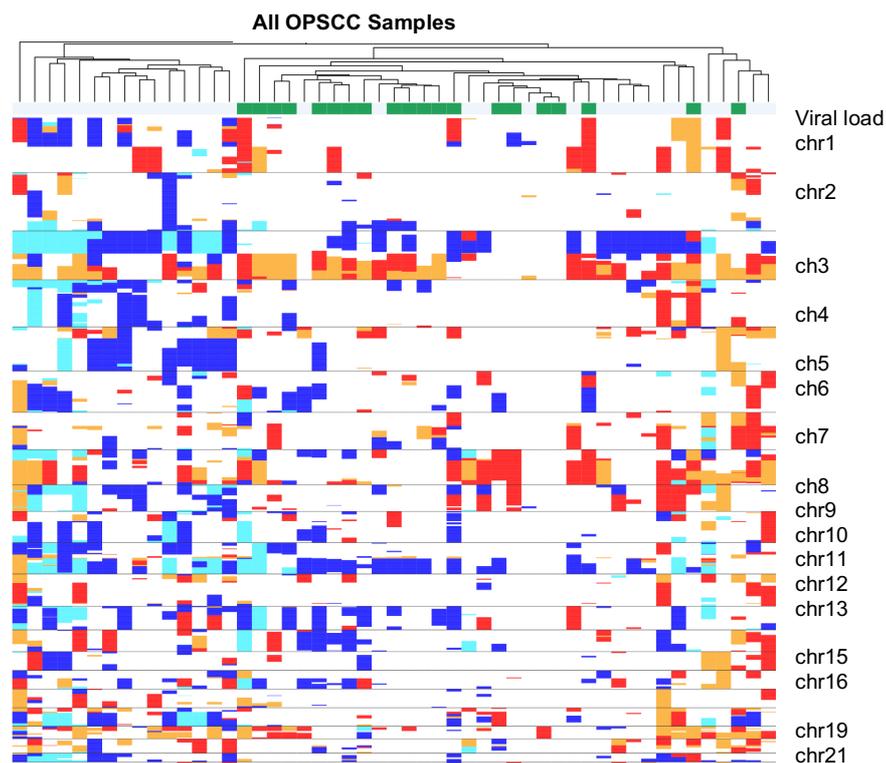
Figure 5-5: Heatmap of all OPSCC tumours generated by pheatmap. Regions of loss of one copy are in dark blue, loss of two or more copies in light blue. Regions of gain of one copy are in red and gain of two or more copies in orange. The chromosome some are numbered along the y-axis. Samples with 0 viral load are represented by white along the top x-axis and samples with viral load >0 represented by green.

Though the cumulative frequency karyograms were inspected to try and identify focal CNAs close correlation was found with the regions highlighted by GISTIC and therefore these were all included. Broader regions identified as on the CNAnorm-generated cumulative frequency karyograms were also assessed. CNAs were selected for attempted identification of minimally altered regions on the basis that they were present in at least 30% of samples in the group with the highest frequency and at least twice as frequently when compared to the other group.

By progressively increasing the resolution of the heatmap and focusing on the smallest region that is recurrently altered, the genomic boundaries for this minimally deleted region amongst tumours with 0 viral load can be obtained. Using loss of 3p24.3 as an example: the wide peak boundaries provided by GISTIC for this are: chr3: 1-84,800,000. Using pheatmap, I was able to identify the actual minimally altered region to be within chr3:

40,000,000-79,200,000. This is still a relatively broad region but suggests the minimally altered region to be loss of 3p22.22-p12.3 and reduces the number of putative genes for downstream analysis from 585 to 390.

(b) **All OPSCC Samples chr3 39200001-78400001**
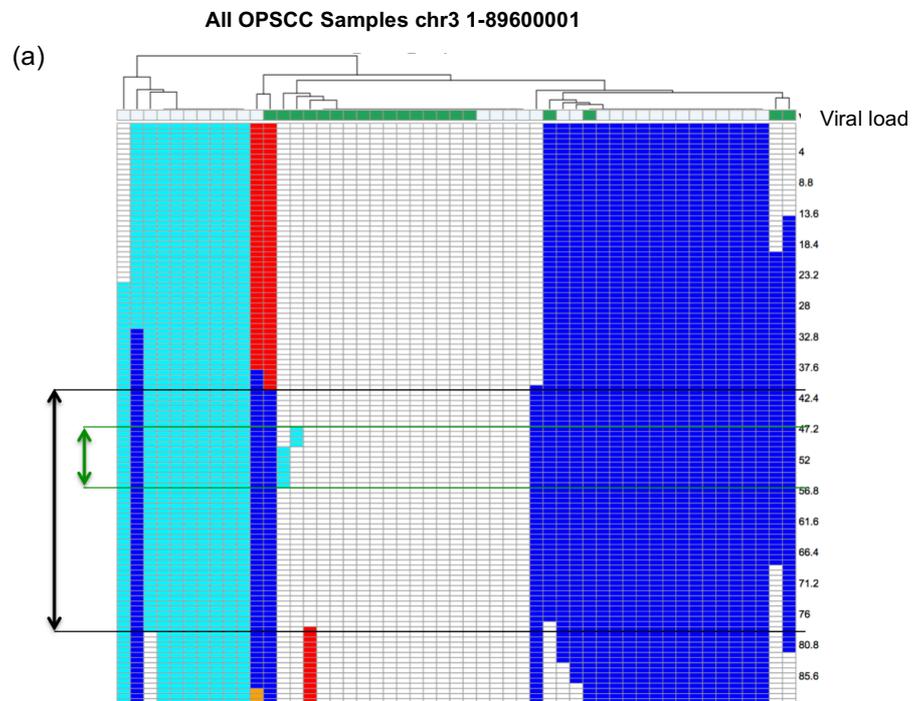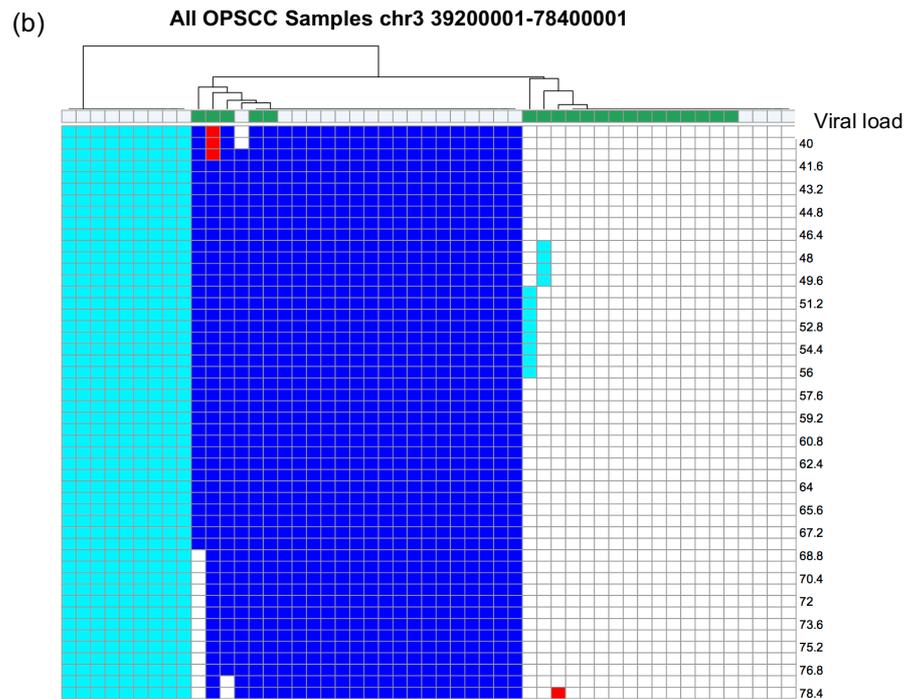
Figure 5-6: (a) Heatmap of all OPSCC tumours based on chr 3p. Regions of loss of one copy are in dark blue, loss of two or more copies in light blue. Regions of gain of one copy are in red and gain of two or more copies in orange. The chromosome some are numbered along the y-axis. Samples with 0 viral load are represented by white along the top x-axis and samples with viral load > 0 represented by green. The smallest, recurrently lost region in the 0 viral load group is highlighted between the black lines, whilst the smallest region lost in the > 0 viral load group is shown between the green lines. (b) Shows this region under higher resolution as obtainable using pheatmap

Relatively broads regions may be of use as a marker, but provide a challenge when trying to identify the underlying key genes that are being affected by CNA. Gain of 5p15.1 was also identified by GISTIC as a focal CNA occurring at a statistically significant rate (q = 0.0589). The wide peak boundaries for this region provided by GISTIC were chr5: 6,400002-24,000000. Using pheatmap to visualise initially chromosome 5, followed by 5p (see Figure **5-7**a & b) allowed identification of two minimally amplified regions within these wider boundaries. By increasing the resolution (see Figure **5-7**c) the genomic windows at which these CNAs begin and end can be identified (5p15.31:7,200000-9,600,000 and 5p14.3: 20,000,000-23,200,000).

The boundaries of the minimally altered regions were then used to obtain the list of all genes within this region from the UCSC Genome Browser (Kent *et al.*, 2002). This process was repeated for all selected regions (see Table **5-4**),

in an effort to specify the most common altered genes. These lists were then further analysed to identify genes of potential significance.
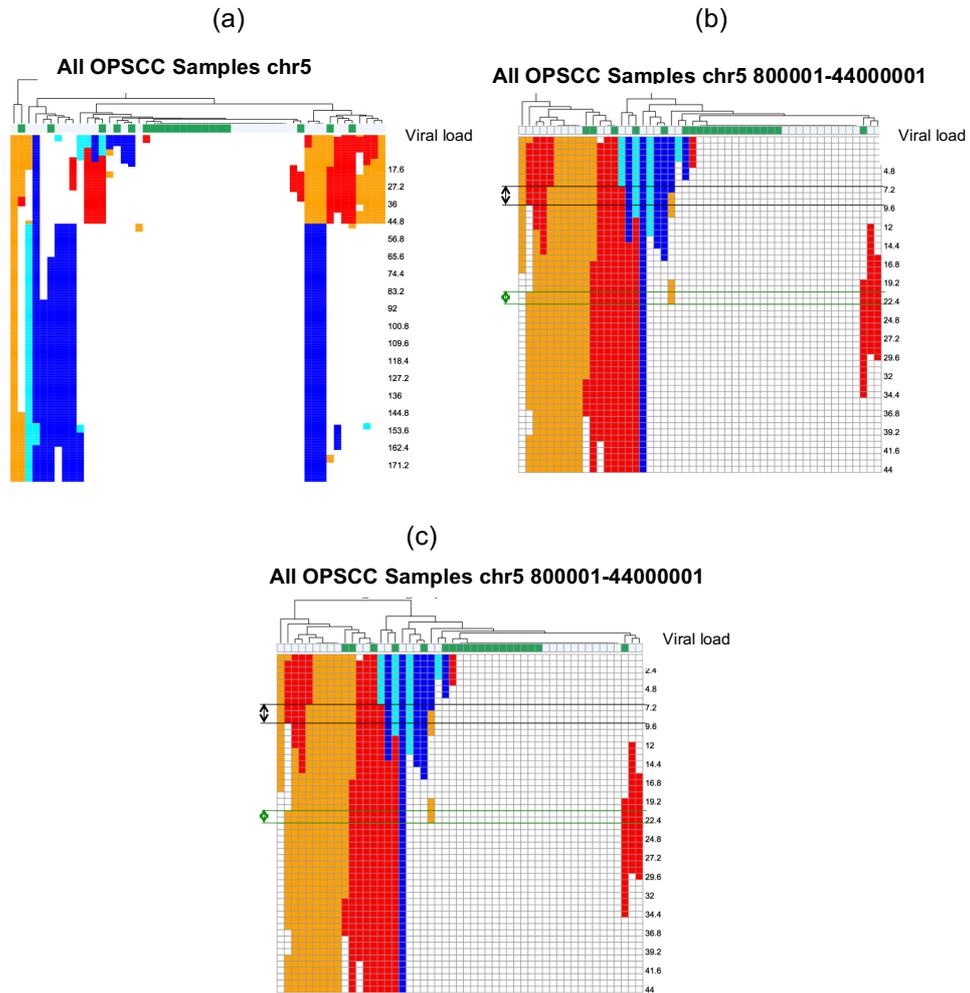
Figure 5-7: (a) Heatmap of all OPSCC tumours based on chr 5. (b) Heatmap based on chr 5p with two recurrently gained regions marked in black and green (c) Heatmap at increased resolution to identify the exact windows representing the boundaries for the minimally altered regions. Regions of gain of one copy are in dark blue, loss of two or more copies in light blue. Regions of gain of one copy are in red and gain of two or more copies in orange. The chromosomes are numbered along the y-axis. Samples with 0 viral load are represented by white along the top x-axis and samples with viral load > 0 represented by green. The smallest, recurrently lost regions in the 0 viral load group are highlighted between the black and black lines.

Deletions				
			No. of samples in each group (%)	
Region	start	end	0 viral load (n = 31)	> 0 viral load (n = 20)
3p22.1-p12.3	40000000	79200000	27 (87)	7 (35)
9p21.3-p21.2	20800001	25600001	16 (52)	1 (5)
9p24.3-p24.1	2000000	4800001	16 (52)	1 (5)
5q33.1-q33.2	150400001	153600001	12 (39)	1 (5)
4p16.3-p16.1	2400000	10400000	16 (52)	4 (20)
8p23.3-p21.1	800001	28800000	12 (39)	5 (25)
18q21.1-22.1	44800001	62400000	14 (45)	2 (10)
18q23.3-q23	70400001	76800000	13 (42)	4 (20)
21q22.2-q22.3	42400001	47200001	15 (48)	3 (15)
21q21.1-q21.3	23200001	28800001	10 (32)	2 (10)
Amplifications				
Region	start	end	0 viral load	> 0 viral load
5p15.31	7200000	9600000	13 (42)	3 (15)
5p14.3	20000000	23200001	15 (48)	5 (25)
7p12.1-p11.2	53600001	56000000	8 (26)	3 (15)
8q24.12-q24.13	120000001	128800001	17 (55)	8 (45)
11q13.2-q13.4	68000001	70400001	12 (39)	5 (25)
19q13.12-q13.2	36000001	41600001	14 (45)	6 (30)
20p12.3-p12.1	7200001	12800001	13 (42)	1 (5)
20q11.21-q11.23	32000001	35200001	10 (32)	4 (20)

Table 5-4: List of minimally altered regions identified in 0 viral load and > 0 viral load tumour groups.

5.3.5 Identification of genes and gene pathways

After using the UCSC Genome Browser to obtain the lists of genes contained within all the minimally altered regions, these lists were then cross referenced against 13 list of genes known to be associated with HNSCC. These included 12 gene pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG, http://genome.jp/kegg/), HNSCC cancer genes identified from the cancer gene census (http://cancer.sanger.ac.uk/cancergenome/projects/census/), and a list of genes gene identified by Stransky *et al* and the TCGA HNSCC subgroup as harbouring a statistically significant frequency of mutation in head and neck cancer (Stransky *et al.*, 2011, Cancer Genome Atlas, 2015).

The final gene lists are shown in Table **5-5**. These represent genes that are copy number altered at a higher frequency in OPSCC tumours with 0 viral load compared to those with viral load > 0. The relatively low number of

genes identified in this stepwise process highlights the potential effectiveness of this progressively focusing analysis.

Amplifications	
Cytoband	Genes
5p15.31	*ADCY2, SEMA5A*
5p14.3	*CDH12*
7p12.1-p11.2	*EGFR*
8q24.12-q24.13	*ZNF572*
11q13.2-q13.4	*LRP5, CCND1, FGF19, FGF4, FGF3*
12p12.3-p12.2	*PIK3C2G, PLCZ1*
19q13.12-q13.2	*PSENEN, RASGRP4, RYR1, MAP4K1, HNRNPL, DLL3, PAK4, AKT2, NUMBL*
20p12.3-p12.1	*PLCB1, PLCB4, PAK7, JAG1*
20q11.21-q11.23	*E2F1*
Deletions	
Cytoband	Genes
3p22.1-p12.3	*CTNNB1, SETD2, CDC25A, SHISA5, LAMB2, RHOA, CACNA2D2, CISH, MAPKAPK3, DUSP7, BAP1, TNNC1, PBRM1, CACNA1D, CACNA2D3, WNT5A, ERC2, FLNB, FHIT, PRICKLE2, MITF, FOXP1*
4p16.3-p16.1	*PPP2R2C, DRD5*
5q32-q33.2	*CSF1R*
8p23.3-p21.1	*ANGPT2, FGF20, PCM1, FGF17, PPP3CC, TNFRSF10B, STC1, PPP2R2A, ADRA1A, PTK2B, FZD3*
9p21.3-p21.2	*IFNB1, IFNW1, IFNA21, IFNA4, IFNA7, IFNA10, IFNA16, IFNA17, IFNA14, IFNA5, IFNA6, IFNA13, IFNA2, IFNA8, IFNA1, IFNE, CDKN2A, CDKN2B*
9p24.3-p24.1	
18q21.1-22.1	*SMAD2, SMAD4, MALT1, PMAIP1, PHLPP1, BCL2, SERPINB5, SERPINB4*
18q23.3-q23	
21q22.2-q22.3	*TMPRSS2, U2AF1*
21q21.1-q21.3	

Table 5-5: Genes identified of potential significance in OPSCC within minimally altered regions.

5.3.6 Analysis of pathways containing CNAs in tumours with 0 viral load vs. > 0 viral load

This was generated using the same pathways selected for cross-referencing the gene lists produced from the minimally altered regions (see section

5.3.5). The pathways enriched by CNAs in tumours with 0 viral load and > 0 viral load are shown in Figure **5-8** and Figure **5-9**).
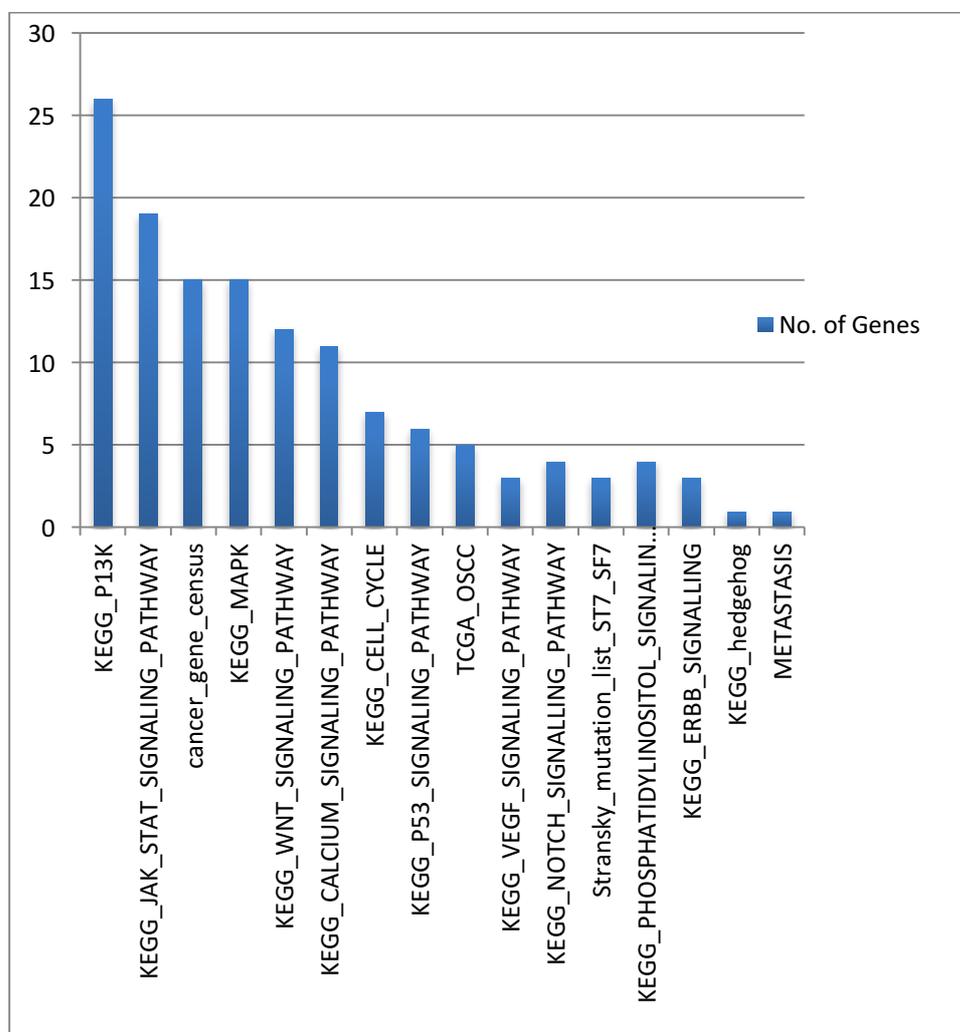
Figure 5-8: Number of genes cross-referenced to KEGG gene pathways and gene lists of potential significance in minimally altered regions (see table 6-4).

The pathway containing the largest number of genes was the PI3K pathway. This is well recognised to be associated with proliferation, invasion and metastasis. Lui *et al* found this to be the most frequently mutated pathway in a study using whole exome sequencing of 151 HNSCC tumours (Lui *et al.*, 2013). The JAK/STAT pathway is widely recognised to be aberrant in haematological malignancy, but is also activated in HNSCC, where members of the pathway contribute to cell survival and proliferation (Lai and Johnson, 2010). Activation of the MAPK pathway has been suggested to

contribute to tumour progression and signal poor prognosis (Leelahavanichkul *et al.*, 2014).
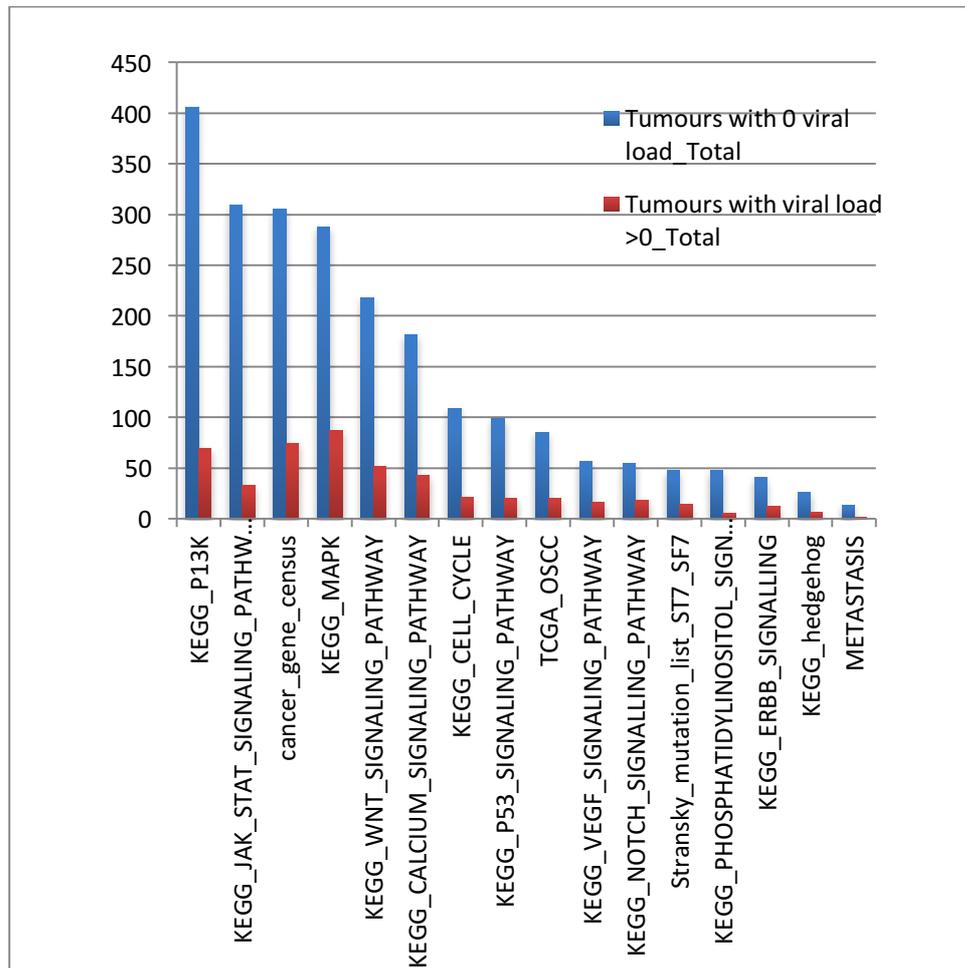
Figure 5-9: Total number of times a gene within each pathway was identified in tumours with 0 viral load (blue) and tumours with viral load > 0 (red).

Comparison of the frequency of CNA of gene within specific pathways is made between the 0 viral load group and > 0 viral load tumours in Figure **5-9**. This demonstrates a much higher frequency of copy number altered gene within the 0 viral load tumours. This is to be expected as the minimally altered regions are present at higher frequency in the 0 viral load group. No minimally altered regions occurred with greater frequency in the >0 viral load group. However, the pathways containing the highest frequency of copy number altered genes differs in the > 0 viral load group.

The pathway most enriched by copy number altered genes in the >0 viral load group was the MAPK pathway, followed by the PI3K signalling and

WNT signalling pathways. This concurs with Siewert *et al*, who found that PI3K signalling pathway aberrations were found in both HPV positive and negative tumours, whilst p53 signalling and cell-cycle pathway aberrations were found more commonly in HPV-negative tumours. It is important to remember that in their study, mutation analysis was used to compare altered networks and therefore used a specific software for this which is not suitable for copy number data, alone.

5.3.7 GISTIC broad analysis of oropharyngeal tumours with no detectable viral load and those with a detectable viral load

Using GISTIC to generate a simultaneous broad analysis of CNA in these two tumour groups produces heat maps (see Figure **5**-**10**) with a similar profile to the cumulative frequency karyograms created using CNAnorm .
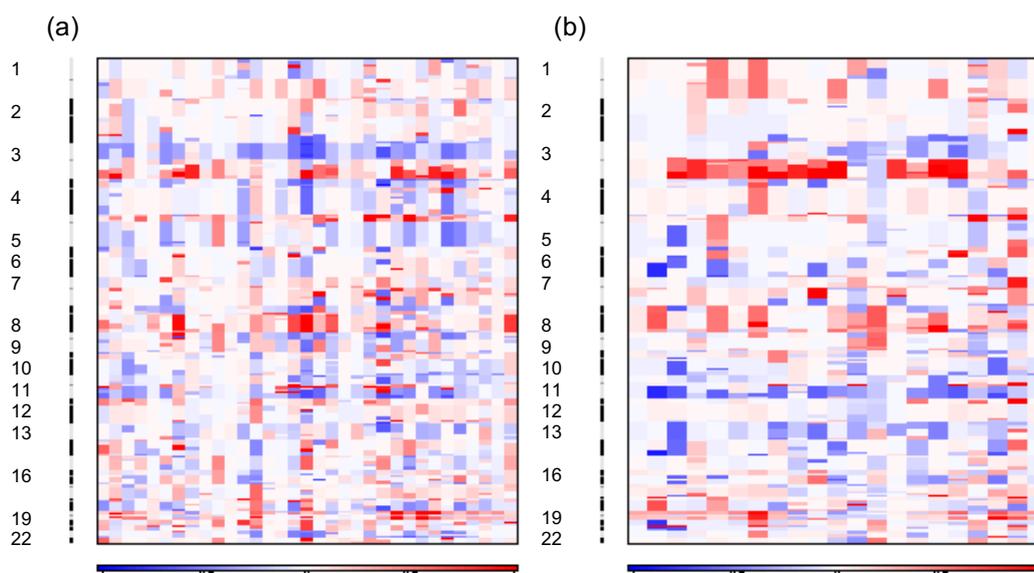
Figure 5-10: Heatmaps generated by GISTIC for CNA in tumours with 0 viral load (a) and viral load > 0 (b). Regions of loss are represented by blue whilst regions of gain are red.

Loss of 3p is seen as a more common event in tumours with 0 viral load as well as the other regions identified by CNAnorm. The advantage of using GISTIC generated heat maps in addition to CNAnorm cumulative frequency plots is the ability to see the group and the individual samples simultaneously. This allows regions that are of lost at a significant level at low frequency such as loss of 5q in tumours with a viral load > 0. This also

allows the inter-tumour heterogeneity of both groups to be visualised. They again suggest there is a reduced overall amount of CNA in the group of oropharyngeal tumours with a viral load > 0.

5.3.8 Analysis of oropharyngeal tumours with intermediate viral load

The groups above were divided purely on the basis of whether or not there is a detectable viral load using NGS. Those with 0 viral load were considered to represent an HPV-negative oropharyngeal tumour. Studies have shown NGS-determined viral load is highly concordant with other techniques such as PCR or IHC (Lechner *et al.*, 2013, Conway *et al.*, 2012). As can be seen in Table **5**-**2** there are a number of samples that had a low or barely detectable viral load. It was decided to analyse the samples with a viral load greater than 0 but no more than 20 as a separate group to those with a viral load greater than 20. This subgroup analysis was performed as the samples with a low viral load (> 0-20) were hypothesised to represent tumour that may not be driven by HPV and therefore possess a CNA profile similar to the 0 viral load group. As not all patients deemed to be HPV positive using established methods are found to have the positive prognosis and treatment response associated with an HPV-positive tumour, this could hold value as an alternative method of stratifying patient risk clinically. This was firstly performed using cumulative karyograms producing using CNAnorm (see Figure **5**-**11**).

The CNA profiles for the intermediate viral load tumours and the high viral load tumours possess a lot of similarities as shown in Figure **5**-**11**. Differences appear to be subtle, with no CNA being exclusively found in one group, or at a large differential level to the other group. Loss of 16q is observed in 5/12 (42%) of high viral load tumours and 2/8 (25%) of the intermediate viral load group. Gain of 7p is seen in 2/8 (25%) of the intermediate viral load tumours and only 1/12 (8%) of the high viral load tumours. Otherwise CNAs appear to occur at similar rates in both groups. This could well be due to the smaller number of patients in each group not allowing differences to be identified. It could also be due to the differences

between the groups being due to low frequency focal CNAs that require a different method to be identified (i.e. GISTIC).
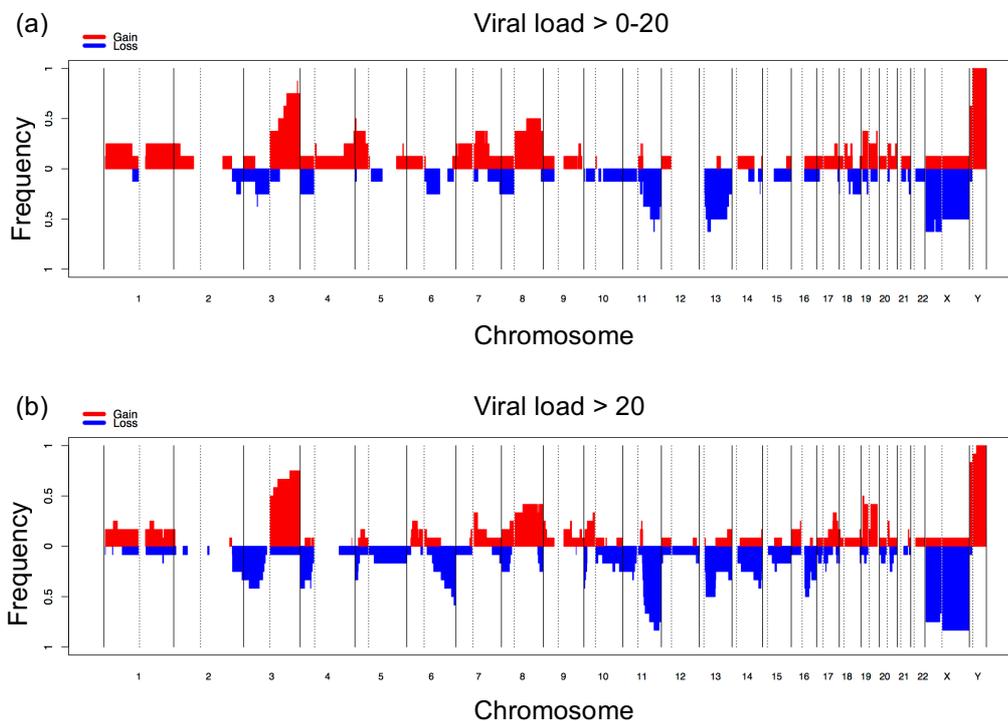
Figure 5-11: Cumulative frequency plots for tumours with viral load > 0-20 (a) and tumours with viral load > 20 (b).

5.3.8.1 GISTIC focal analysis of intermediate viral load group

GISTIC was used to analyse the same sample groups to produce a simultaneous focal and broad CNA. The input files for this were created as described in Chapter 2.9.2.3. The results of the focal analysis are shown in Figure **5**-**12** and Figure **5**-**13**. In reviewing the GISTIC focal analysis it is important to remember the reduced number of patients in each group and to treat telomeric and centromeric CNAs with caution.

The focal amplification profile of the intermediate viral load group (n = 8) is seen to be quite different to the group with viral load > 20 (n = 12) and actually has a greater similarity to the tumours with zero viral load (see Figure **5-2**). The intermediate viral load group shares several focal amplifications identified at a significant level with the zero viral load group (gain of 3q26.3, 7q22.3, 8q24.12, 11q13.12). The tumours with viral load > 20 are seen to contain gain of 20q11.2 at a significant level. In terms of focal deletions (see Figure **5-13**) the intermediate viral load group is again seen to share CNAs with the zero viral load group (gain of 2q36.3, 13q13.2) and have little in common with the viral load > 20 tumours. However, the low frequency of any CNAs in the tumours with viral load > 20 may reflect the lower numbers of patients in this group and therefore the focal regions nominated by GISTIC need clarification by evaluating the minimally altered regions within these focal CNAs.
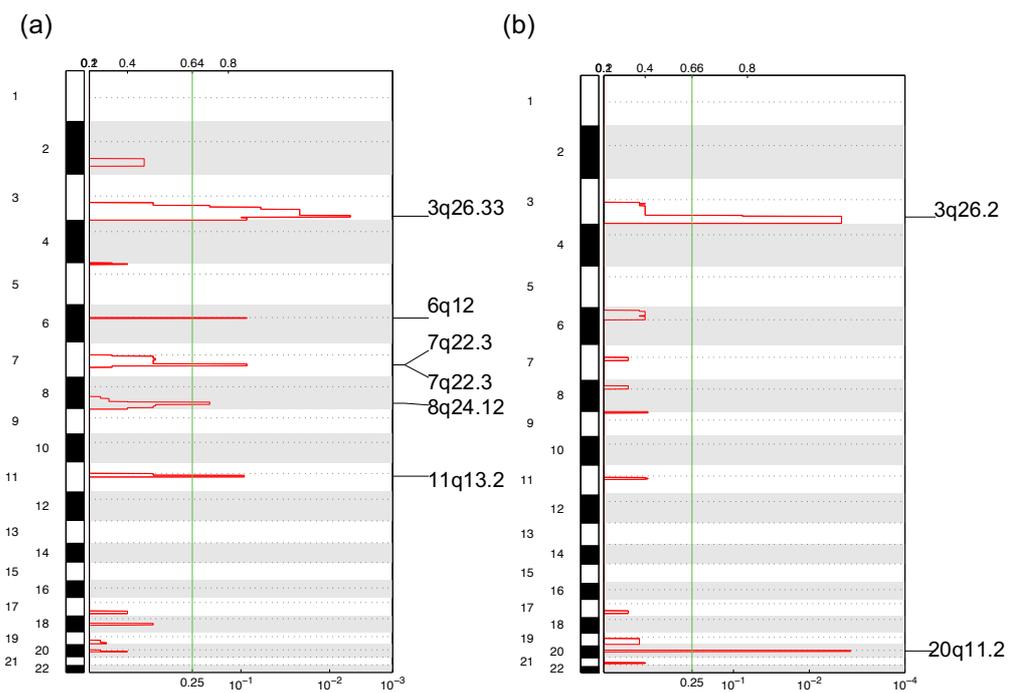
Figure 5-12: Genome-wide focal amplification plots for OPSCC tumours with intermediate viral load > 0-20 (a) and > 20 (b). The chromosomes are numbered along the y-axis. The bottom x-axis indicates the q-value with the green line. The top x-axis shows the G-score for CNAs.
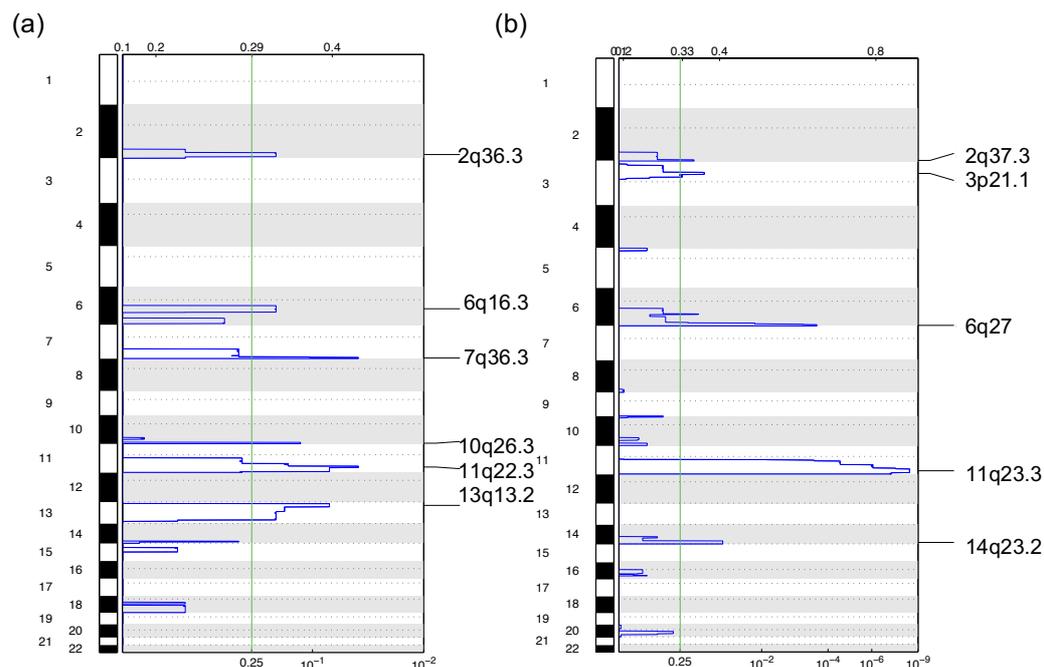
Figure 5-13: Genome-wide focal deletion plots for OPSCC tumours with intermediate viral load > 0-20 (left) and > 20 (right).

5.3.9 Evaluation of minimally altered regions in intermediate and high viral load groups

Using the regions nominated by GISTIC and visual analysis of the CNAnorm karyograms, a number of focal CNAs were analysed to try and identify any minimally altered regions that may differentiate between intermediate or high viral load tumours. This analysis also included using pheatmap software as described in section 5.3.4 (see Table **5-6**). As these groups were smaller, using the regions nominated by GISTIC was preferred rather than simple visual inspection, due to the fact that GISTIC performs statistical analysis of these regions, including adjusting the p-value for multiple testing (q-value) with the level of significance set at 0.25. The focal regions nominated by GISTIC met this criteria.

When evaluating the minimally altered regions across the three groups there are few differences. No CNA is exclusively found in one group. Some do suggest that the intermediate viral load tumour shares similarities with the 0 viral load tumours. Gain of 6q12.1-11.1 is found in 5/31 (16%) of 0 viral load tumours and 1/8 (13%) of intermediate viral load tumours but in 3/12 (25%)

of tumours with viral load > 20. Loss of 6q26-q27 is also identified in 5/12 (42%) of high viral load tumours but 1/8 (13%) of intermediate viral load tumours and 6/31 (19%) of 0 viral load tumours. Gain of 7q11.22-122.2 was present in 14/31 (45%) 0 viral load tumours and 3/8 (38%) intermediate viral load tumours but only 1/12 (8%) high viral load tumours. Gain of 7p12.3-p11.2 is also seen in 8/31 (26%) of 0 viral load tumours, 2/8 (25%) of intermediate viral load tumours and 1/12 (8%) of the high viral load group. Loss of 3p21.31-p21.2 was found much more frequently in 0 viral load tumours compared to tumours with viral load > 0.

When the intermediate and high viral load groups are compared, a similar frequency of loss of 3p is observed (27/31 (87%) in 0 viral load tumours, 3/8 (38%) in intermediate viral load tumours and 4/12 (33%) in high viral load tumours). Loss of 3p has been associated with poor prognosis in HNSCC (Gross *et al.*, 2014). The fact that it occurs at lower frequency in 0 viral load tumours is supportive of Gross *et al's* findings as the prognosis for HPV-16-associated OPSCC is generally improved compared to non-HPV-16-associated OPSCC.

Interestingly some CNAs were seen at a similar rate across all three groups including loss of 11q22.3-q23.2 9 present in 19/31 (61%) of 0 viral load tumours, 5/8 (63%) of intermediate viral load tumours and 10/12 (83%) of high viral load tumours. Gain of 8q24.12-q24.22 was seen in 17/31 (55%) of 0 viral load tumours, 5/8 (63%) of the intermediate group and 6/12 (50%) of the high viral load tumours. These CNAs have been reported previously as frequently occurring in HNSCC (Bauer *et al.*, 2008, Yong *et al.*, 2014). This suggests that these CNAs relate to cancer mechanisms or attributes not specifically influenced by viral status.

No CNAs (focal or broad) appear to differentiate well between the intermediate and high viral load tumours.

Amplifications					
			No. in each group (%)		
Cytoband	start	end	0 viral load n = 31	> 0-20 viral load n = 8	> 20 viral load n = 12
6q12.1-q11.1	55200001	63200001	5 (16)	1 (13)	3 (25)
7q11.22-q22.2	67200000	111800000	14 (45)	3 (38)	1 (8)
8q24.12-q24.22	119800000	132000000	17 (55)	5 (63)	6 (50)
11q13.1-q13.4	64000001	70400001	11 (35)	2 (25)	3 (25)
20q11.1-q11.23	28000000	35200001	10 (32)	3 (38)	2 (16)
7p12.3-p11.2	47200001	56000000	8 (26)	2 (25)	1 (8)
Deletions					
Cytoband	start	end	n in 0 viral load	n in >0-20 viral load	n in >20 viral load
2q35-q37.1	218400001	234400001	7 (23)	2 (25)	3 (25)
2q37.2-q37.3	236800001	241600001	11 (35)	1 (13)	4 (33)
6q15-q22.1	92800001	116800001	6 (19)	2 (25)	4 (33)
7q36.1-q36.3	152000001	158400001	7 (23)	2 (25)	2 (16)
10q26.13-q26.3	125600002	135534747	11 (35)	1 (13)	2 (16)
11q22.3-q23.2	108000000	112800001	19 (61)	5 (63)	10 (83)
13q32.3-q34	100800001	112800001	7 (23)	3 (38)	4 (33)
3p21.31-p21.2	48000000	52000000	27 (87)	3 (38)	4 (33)
6q26-q27	162400001	164800001	6 (19)	1 (13)	5 (42)
14q32.2	96800001	99200001	7 (23)	2 (25)	4 (33)
14q32.31-q32.33	101600001	104000001	6 (19)	1 (13)	5 (42)

Table 5-6: List of minimally altered regions identified in 0 viral load, intermediate and high viral load tumour groups.

5.3.10 Selection of CNA panel to differentiate OPSCC tumours with zero viral load from those with viral load >0

Utilising the minimally altered regions identified in section 5.3.4, a selection were compiled as a panel of CNA markers to potentially separate OPSCC tumours with 0 viral load from those with a viral load >0 (see Table 5-7). This panel was then applied to all OPSCC tumours and a heatmap created as shown in Figure 5-14.

Though this did not separate all tumours with 0 viral load, when applying a binary score of 1 for present to each minimally altered region and 0 for absence of specific gain or loss, a total score for each sample was generated. These scores were then analysed using the SPSS (Version 21, IBM Corp, Armonk NY). The Shapiro-Wilk test for normality was performed finding that data was not normally distributed (p = 0.004). The Kruksal-Wallis

test was then used to compare the mean scores across the three groups (see Figure **5**-**15**) finding that the means scores for the 0 viral load tumours were significantly higher than the other groups (p = <0.001). When comparing the mean scores of the > 0-20 viral load and >20 viral load groups using the Mann Whitney U Test no significant difference was seen (p = 0.624).

CNAs associated with 0 viral load	
Regions of gain	Regions of loss
5p15.31	3p22.1-p12.3
5p14.3	9p21.3-p21.2
8q24.12-q24.13	9p24.3-p24.1
11q13.2-q13.4	5q33.1-q33.2
19q13.12-q13.2	4p16.3-p16.1
20p12.3-p12.1	8p23.3-p21.1
20q11.21-q11.23	18q21.1-22.1
12p12.3-p12.2	18q23.3-q23
7p12.1-p11.2	21q22.2-q22.3
	21q21.1-q21.3

Table 5-7: Panel of CNAs identified in OPSCC tumours associated with 0 viral load.
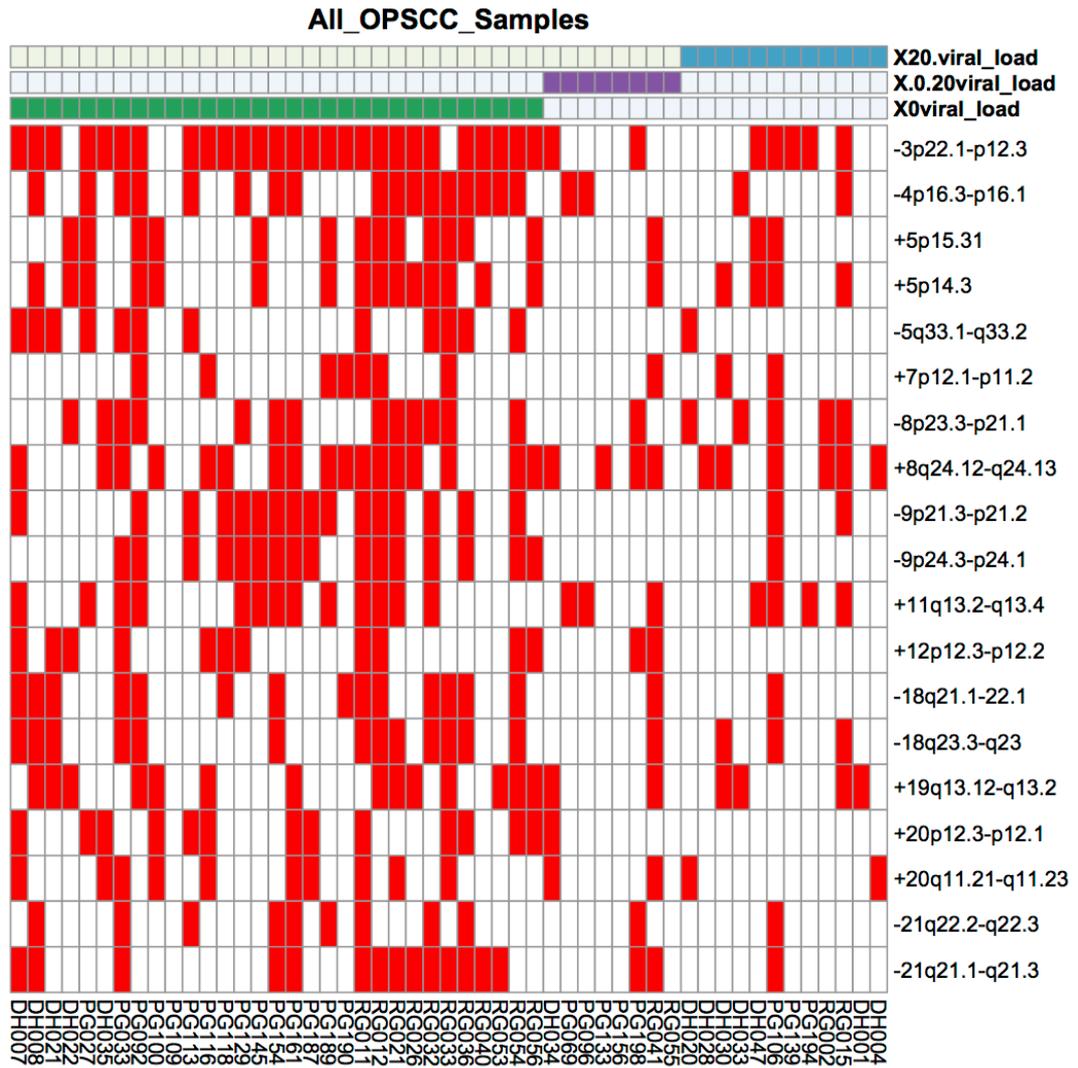
Figure 5-14: Heatmap to show all OPSCC samples when specifically assessed for the presence (red) or absence (white) of the minimally altered regions.
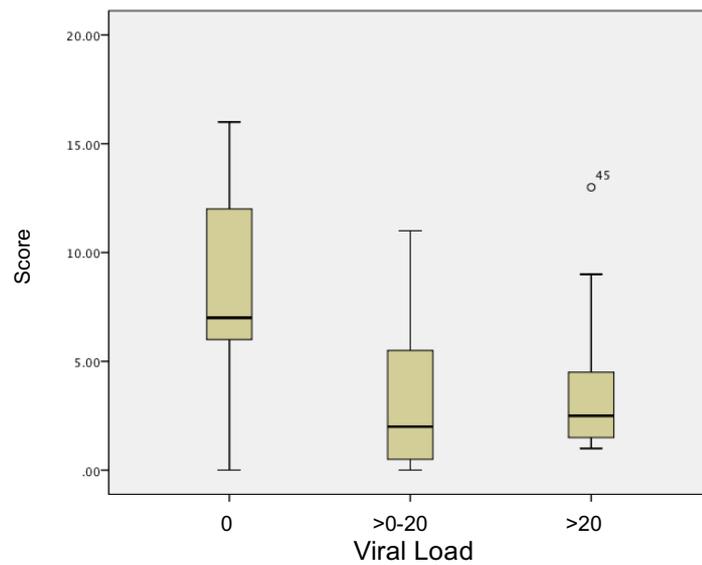
Figure 5-15: Comparison of mean scores for the minimal CNA panel for the 3 groups of viral load.

5.3.11 Comparison of fraction of genome altered (FGA)

The actual proportion of the genome altered in terms of copy number was calculated as a reflection of the underlying amount of altered genomic DNA in each group of tumours. The FGA is calculated using the number of bases in all segments in each sample as the denominator and number of bases in each sample that are copy number altered as the numerator. These were then compared between groups separated according to viral load.

The FGA for OPSCC tumours with 0 viral load and those with viral load >20 were analysed using the Shapiro-Wilk test for normality. As it was normally distributed the two-sample t-test was then used to compare the means of these groups (see Figure **5-16**). The mean FGA of the tumours with 0 viral load (0.328) was found to be significantly higher than the mean FGA of tumours will viral load >0 (0.210, p = 0.006).
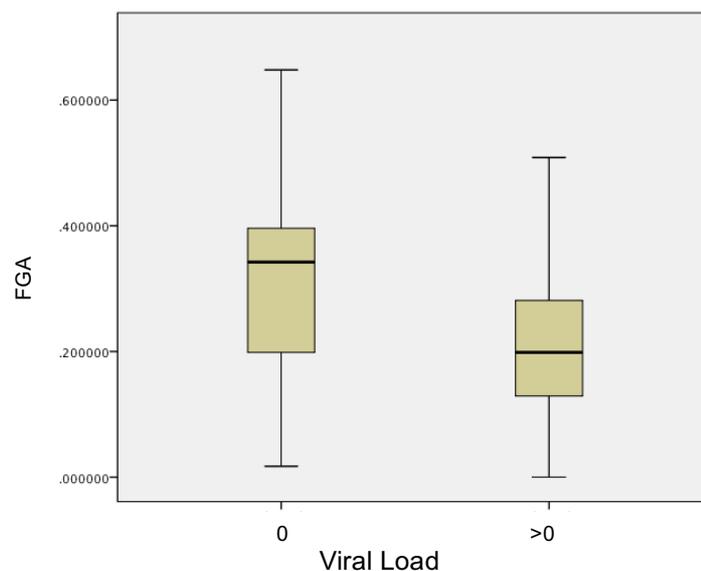
Figure 5-16: Comparison of FGA for tumours with 0 viral load and tumours with viral load >0.

The FGA for the intermediate viral load group and tumours with viral load >20 were also calculated as above. The distribution of these samples was analysed for normality using the Shapiro-Wilk test for normality (see Figure **5-17**). This revealed neither group was normally distributed and thus the

means were compared using the Mann-Whitney U test. The mean FGA of each group was not found to be significantly different (p = 0.910).
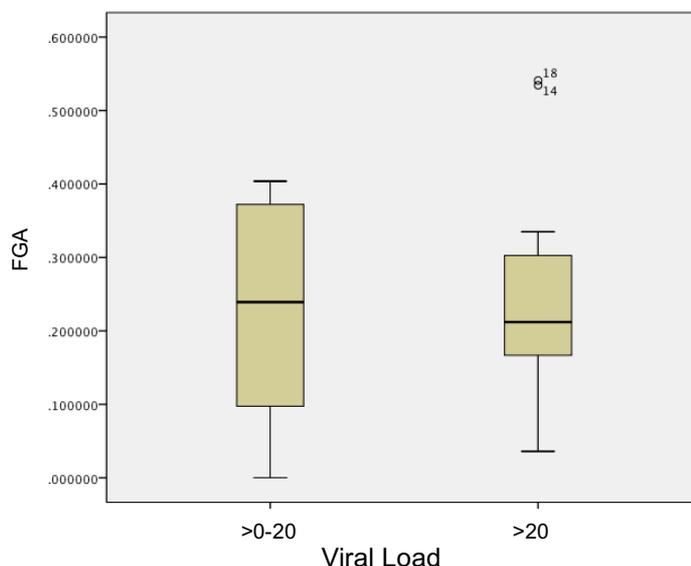
Figure 5-17: Comparison of FGA of tumours with viral load >0-20 and tumours with viral load >20.

5.4 Discussion

5.4.1 CNAs associated with differing viral load in OPSCC

Since Fearon and Vogelstein demonstrated the stepwise accumulation of genetic alterations occurring in parallel with colorectal tumourigenesis, cancer has been regarded as a disease of genomic origin (Fearon and Vogelstein, 1990). There are many ways of examining the cancer genome. CNV represents an important component of genetic damage and variation. It affects a greater fraction of the genome than mutations and has been used to identify novel oncogenes and provide biomarkers for prognostic sub-groups (Curtis *et al.*, 2012, Weir *et al.*, 2007, Nielsen *et al.*, 2008). Despite large genomic studies of HNSCC recently published, there is still a dearth of HPV-positive samples in these, for instance the TCGA cohort is 85% HPV negative (Cancer Genome Atlas, 2015).

The association of 3p loss and OPSCC tumours with zero detectable viral load is of particular interest given the recent findings of the Cancer Genome

Atlas (TCGA) HNSCC project. They reported loss of 3p (particularly in when associated with *TP53* mutation) to be a strong marker for poor prognosis in HPV-16 negative tumours (Gross *et al.*, 2014). In earlier work Stransky *et al* and Agrawal *et al* both found *TP53* mutation to be an almost universal finding in HPV-16-negative tumours (Stransky *et al.*, 2011, Agrawal *et al.*, 2011). Given the positive prognosis associated with the majority of HPV-16 positive tumours it is entirely in keeping that loss of 3p should be largely absent from these patients.

Interestingly gain of proximal 11q and loss of distal 11q is seen with similar frequency in the two groups. This has previously been associated with poor outcome in HNSCC, but was not discriminatory for HPV-16 status in these patients (Ambatipudi *et al.*, 2011).

In general, the HPV-16 negative tumours demonstrate a more "classical" CNV profile in terms of chromosomal changes associated with squamous cell carcinoma (e.g. loss of 3p, gain of 5p). The lack of the more typical alterations associated with SCC likely reflects the aetiology of these tumours and how it differs from the tobacco associated carcinogens.

Seiwert *et al* used their targeted exome sequencing to infer copy number alteration using read depth analysis from 50 HPV-positive and 70 HPV-negative mixed subsite HNSCC samples (Seiwert *et al.*, 2015). As in my study, they found that 3p loss occurred primarily in HPV-negative tumours. They also presented a selection of 25 genes recognised in cancer and the incidence in HPV-positive and negative tumours (see Table **5-8**). Of these 25 genes, gains of *EGFR* and *FGFR1* were observed only in HPV-negative tumours and highlighted by the study as of significance as driver genes of high translational value (Seiwert *et al.*, 2015). In this study they were found in tumours with 0 viral load and those with viral load > 0, though at much lower frequency in tumours with a detectable viral load. It is of interest that there was a reduction in incidence of gain of *EGFR* from intermediate to high viral load. Gain of *FGFR1* did not follow this trend but within Seiwert *et al's* study this was only identified in 1/50 HPV negative samples (Seiwert *et al.*, 2015). Also concurring with Seiwert *et al*, gain of *CCND1* was observed at a rate of 40% in tumours with 0 viral load and 20% in those with viral load > 0

(at a similar frequency in both the intermediate and high viral load tumours). Otherwise the remaining amplified genes highlighted by Seiwert *et al* occur in both 0 viral load tumours and tumours with viral load > 0.

	Gene	Cytoband	Frequency in tumours by viral load (%)		
			0 n = 31	> 0-20 n = 8	> 20 n = 12
Amplified	*EGFR*	7p11.2	8 (26)	2 (25)	1 (8)
	FGFR1	8p11.23-p11.22	7 (23)	1 (13)	4 (33)
	REL	2p16.1	3 (10)	1 (13)	0
	BCL6	3q27.3	19 (61)	7 (88)	8 (67)
	AKT3	1q43-q44	7 (23)	2 (25)	3 (25)
	PIK3CB	3q22.3	13 (42)	5 (63)	8 (75)
	PIK3C2A	11p15.1	1 (3)	0	0
	PIK3CA	3q26.32	21 (68)	7 (88)	8 (75)
	TP63	3q28	20 (65)	7 (88)	8 (75)
	CCND1	11q13.3	13 (42)	3 (38)	1 (8)
	MYC	8q24.21	16 (52)	4 (50)	4 (33)
	MDM2	12q15	6 (19)	0	0
Deleted	*ATM*	11q22.3	13 (42)	6 (75)	7 (58)
	CDKN2A	9p21.3	16 (52)	0	1 (8)
	RB1	13q14.2	10 (32)	4 (50)	5 (42)
	TSC2	16q13.3	3 (10)	0	1 (8)
	CSF1R	5q32	9 (29)	0	1 (8)
	TP53	17p13.1	2 (6)	1 (13)	0
	UBR5	8q22.3	0 (0)	0	0
	SETD2	3p21.31	27 (87)	3 (38)	3 (25)
	NOTCH1	9q34.3	9 (29)	0	0
	SYNE2	14q23.2	2 (6)	1 (13)	4 (33)
	CHD5	1q36.31	6 (19)	0	1 (8)
	MYH6	14q11.2	3 (10)	0	3 (25)
	NF1	17q11.2	1 (3)	0	0

Table 5-8: Prevalence of copy number altered genes highlighted by Seiwert *et al* (Seiwert *et al.*, 2015).

A number of deleted genes occur with greater frequency in tumours with 0 viral load. These include loss of *SETD2*, *CDKN2A*, *CSF1R*, *NOTCH1* and *CHD5*. Seiwert *et al* concurred with the finding that *CDKN2A* occurred

primarily in HPV-negative tumours, but found the other deleted genes at similar rates in both HPV-positive and negative tumours (Seiwert *et al.,* 2015). This could be due to the fact that their cohort consisted of mixed subsites, with only 20/70 of their HPV-negative cohort being oropharyngeal, though the majority of their HPV-positive tumours were oropharyngeal (47/51).

The differences between my study and Seiwert *et al's* may also be related to the combination of methods they used to determine viral status (DNA-PCR, p16 expression and presence of *TP53* mutations).

The advantages and disadvantages of different HPV-testing methods are debated. HPV-16 PCR can be accused of being prone to contamination or being overly sensitive. In-situ hybridisation can also detect a single copy of HPV-16 DNA (Smeets *et al.,* 2007). The problem being that detection of any HPV-16 DNA at all does not necessarily mean the tumour is driven by this aetiological factor and therefore shares the positive prognosis bestowed by this. Immunohistochemistry suffers from inter-observer variation, as well as debate over grading of positivity, as well as its sensitivity (Smeets *et al.,* 2007).

Examining Conway *et al's* data (which compared viral load to the presence of DNA HPV-16 PCR), it can be seen that of 24 patients with 0 viral load 100% concordance with a negative DNA-PCR results was found (Conway *et al.,* 2012). Similarly, all patients with a viral load > 20 were reported to have a positive DNA-PCR result. In 4/6 patients with an intermediate viral load, a positive PCR result and 2 negative results were obtained. This suggests that the intermediate viral load patients are those in whom HPV DNA may be detectable but may well not be the primary aetiological factor behind the development of their OPSCC. This is an important finding as it allows discussion of specific terminology. Though the HPV-16 positive tumour is well-recognised, due to the variety of different tests available the "HPV-driven tumour" cannot yet be reliably recognised. However the use of viral loads alone to group tumours could be improved by adding clinical outcome data for these tumours, which unfortunately was absent.

My study suggests that the genomic profile of OPSCC with an intermediate viral load may share similarities with tumours with 0 viral load or with viral load > 20. This could represent potential for use as a prognostic indicator if viral load could be linked to clinical outcome in the future. This is important in the debate over whether or not to de-escalate treatment regimes for those with HPV16-positive tumours. Hyper-sensitive methods of testing for HPV-16 could lead to patients without a truly HPV-driven tumour having their treatment changed to a less aggressive protocol and potentially being placed at higher risk of recurrence/progression.

However Seiwert *et al's* study demonstrates that there is still considerable inter-tumour genomic heterogeneity (from both a mutational and copy number perspective) (Seiwert *et al.*, 2015). Aside from *TP53* there are no other unifying copy number or mutational aberrations found in one tumour type. This was similarly reported by the TCGA head and neck sub-group, exemplified by the novel recurrent alterations they reported in HPV-positive tumours being deletion of *TRAF3* and amplification of *E2F1* (Cancer Genome Atlas, 2015). Loss of *TRAF3* only occurred in 5/36 (14%) samples, whilst gain of *E2F1* was present in 7/36 samples (19%) (Cancer Genome Atlas, 2015). My study found this to be a particularly heterogeneous region with loss of *TRAF3* in 6/20 (30%) tumours with viral load > 0 and 6/31 (19%) with 0 viral load. Gain of *TRAF3* was present in two tumours with viral load >0 and four tumours with 0 viral load. Again, this may be related to the mixed subsites being included in the TCGA cohort (only 21/36 samples were oropharyngeal), or may simply reflect the heterogeneity of these tumours.

Integration of HPV DNA into the host genome is well recognised as a likely key part of the carcinogenic process in HPV-associated cancers. Parfenov *et al* analysed 35 HPV-positive tumours from the TCGA cohort to determine the integration sites in this cohort (Parfenov *et al.*, 2014). Of the 35 samples, 25 cases had integration of the viral genome into the host identified. This was observed in up to 16 regions of the human genome (per sample), with 103 genomic breakpoints. They found the majority (71%) of integrations occurred within a known gene or within 20kb of a gene. They also found that viral integration was associated with CNA of the integrated region in 82% of cases (Parfenov *et al.*, 2014). After obtaining the list of genes associated

with viral integration I analysed my cohort of samples for CNA of genomic windows within or adjacent to these genes (see Table **5-9**).

| Genes | Cytoband | Frequency in tumours by viral load (%) | | | | | |
| | | 0 n = 31 | | > 0-20 n = 8 | | > 20 n = 12 | |
		Gain	Loss	Gain	Loss	Gain	Loss
RAD51B	14q24.1	6	3	1	1	1	4
PRKRIR	11q13.5	6	12	0	3	0	6
SERPINB4	18q21.33	3	14	0	2	1	0
SERPINB7	18q21.33	3	14	0	2	1	0
LOC100506023	1q25.1	5	1	2	0	3	0
NR4A2	2q24.1	1	2	0	0	0	0
DOLPP1	9q34.11	8	5	2	0	2	0
CRAT	9q34.11	8	5	2	0	2	0
LINC00111	21q22.3	0	15	1	2	1	1
PLGRKT	9p24.1	4	15	1	1	2	0
CD274	9p24.1	4	15	1	1	2	0
ETS2	21q22.2	0	11	1	2	2	0
ZBTB7C	18q21.1	4	0	2	1	1	0
ARHGDIA	17q25.3	3	1	0	1	0	1
CCDC39	3q26.33	21	0	6	0	9	0
TTC14	3q26.33	21	0	6	0	9	0
IQGAP1	15q26.1	4	3	1	1	0	0
CEACAM19	19q13.31	13	0	2	1	3	0
MAGI2	7q21.11	10	0	3	1	1	0
TCTEX1D1	1p31.3	4	6	2	0	2	1
TRPC4AP	20q11.22	10	1	2	1	2	0
GRIK1	21q21.3	0	10	1	1	0	1
PLCXD2	3q13.2	9	5	3	1	8	0
EPSTI1	13q14.11	2	9	0	4	0	5
DNAJC15	13q14.11	2	9	0	4	0	5
LOC100506136	7q21.3	9	1	3	0	1	0
TEAD1	11p15.3-p15.2	1	12	0	1	0	2
ASIC2	17q11.2-q12	5	2	0	0	0	1

Table 5-9: Frequency of copy number alteration in genomic windows containing or adjacent to genes identified at sites of HPV integration by Parfenov *et al* (Parfenov *et al.*, 2014). Those genes found to have altered copy number in a higher proportion of tumours with viral load > 0 are highlighted in red.

CNA alteration was identified at almost all genes highlighted by Parfenov *et al* in both tumours with 0 viral load and tumours with viral load > 0 (Parfenov *et al.*, 2014). *NR4A2* was only found to be copy number altered in tumours with 0 viral load. This may reflect the smaller number of samples in my study or may result from the fact that this gene is smaller than the resolution achievable with my data (800 Kb). The fact that the majority of copy number alterations tend to be larger than a single genomic window would suggest that the sequencing resolution is adequate. However, smaller CNAs are not identifiable without higher resolution (greater depth of sequencing). Germline CNAs tend to be small and these may be of significance in the development of SCC as well as identifying those with a predisposition to harbouring HPV after infection and developing cancer in later life (Park *et al.*, 2015).

In only seven genes was the incidence of CNA found to be proportionally higher in the tumours with viral load > 0 (highlighted in red in Table **5-9**). All seven genes were found to be copy number altered in the 0 viral load group at slightly lower rates, which did not achieve a level of significance after examining the difference using Fisher's Exact test. This suggests that though these genes have been associated with HPV integration they are also altered during carcinogenesis in tumours not caused by HPV.

Parfenov *et al* reported an association specifically with copy number gain at viral integration sites. This trend was followed in this study. Overall, in the 0 viral load tumours, there were 166 amplifications and 171 deletions of the genes in Table **5-9**. In the tumours with viral load > 0 there were 95 amplifications and 58 deletions. Both gain and loss were observed at most genes with deletions not observed at only 6 genes in the > 0 viral load group. These 6 genes were also found to be amplified in the 0 viral load group, demonstrating the heterogeneity to be found in both HPV-associated and non-HPV-associated OPSCC.

5.4.2 Comparison to cervical squamous cell carcinoma

HPV has been associated with cervical adenocarcinoma and squamous cell carcinoma since the early 1980s and to play a key role in the carcinogenesis of these cancers (Bosch *et al.*, 2002). Given the common aetiological factor, comparison of the genomic changes in both OPSCC and cervical squamous

cell carcinoma (CSCC) is appropriate, bearing in mind the potentially significant differences between the two cancers. These include the fact that over 95% of oropharyngeal cancer is histologically SCC, whilst it only comprises approximately 60% of cervical cancer (Vizcaino *et al.*, 2000). HPV-negative cervical squamous cell carcinoma (CSCC) represents up to 10% of CSCC, whilst approximately 35% of OPSCC is HPV-negative (Rodriguez-Carunchio *et al.*, 2015). HPV-16 is associated with 62% of CSCC, with subtypes 18, 31, 33 and 45 making up the majority of the remainder (de Sanjose *et al.*, 2010).

The TCGA cervical cancer subgroup conducted whole exome sequencing of 79 CSCC samples and inferred copy number data from the read depth of this data (Ojesina *et al.*, 2014). All of these were HPV-positive. They utilised GISTIC 2.0 to provide broad and focal CNA analysis. Broad level alterations included gain of 3q, 1p, 1q, 20p, 20q, 14q, 5p, 19q and 8q as well as loss of 3p, 4p, 13q, 3q, 4q, 11q, 17p, 11p, 6q, 8p and 6p (Ojesina *et al.*, 2014). Focal CNAs are listed in Table **5-10**. They demonstrate considerable similarity to the CNA alterations seen in HNSCC. Loss of 3p has been previously reported to be an early event in both HNSCC and CSCC, however in HNSCC it has specifically been associated with poorer outcome (Califano *et al.*, 1996, Wilting *et al.*, 2008, Gross *et al.*, 2014). The fact that loss of 3p was observed in 50% of HPV-positive CSCC in the TCGA cohort is similar to the 35% (7/20) tumours with viral load > 0 in which 3p deletion was present and consistent with the improved outcome associated with both HPV-positive OPSCC and CSCC (Rodriguez-Carunchio *et al.*, 2015, Ojesina *et al.*, 2014, Fakhry *et al.*, 2008).

Several tumour suppressor genes have been identified on 3p, including *FHIT*, found specifically at 3p14.2, a common fragile site (Ingvarsson, 2005). This CNA was seen in 50% of CSCC, 87% of 0 viral load OPSCC and 35% of > 0 viral load OPSCC. *FHIT* has been associated with cellular pathways including apoptosis and cell cycle, in particular inhibiting p53 degradation by *MDM2* (Watanabe *et al.*, 2004). Decreased expression of *FHIT* has also been associated with poor prognosis in a number of cancers including HNSCC (Tai *et al.*, 2004). The fact that deletion of this gene is found much

less frequently in OPSCC with viral load > 0 is consistent with the existing literature.

Focal amplifications	Focal deletions
1p12	1p21.1
2q14.2	2q22.1
2q24.3	2q32.2
3q28	2q36.3
8q24.21	3p26.2
8q24.3	3p25.1
11q13.2	3p14.1
11q22.1	6p25.2
13q34	6q27
15q24.1	7q36.1
16p11.2	8q22.1
17q12	9p13.3
17q24.1	10q11.21
19q13.13	11q22.3
19q13.32	11q25
22q11.23	12q14.1
	13q14.11
	13q21.2
	14q11.2
	15q14
	16p11.2
	16q22.2
	17q25.3
	19p13.3
	19p12

Table 5-10: List of focal amplifications and deletions identified in CSCC using GISTIC 2.0 by Ojesina *et al* (Ojesina *et al.*, 2014).

CDKN2A, found at 9p21.3, is also of particular interest. This encodes for p16, a tumour suppressor protein that plays a role in slowing cell cycle progression specifically from G1-phase to S-phase (Ruas *et al.*, 1999). This gene is normally downregulated by the retinoblastoma tumours suppressor gene product (pRb). In HPV-associated OPSCC the E6 and E7 viral

oncoproteins are expressed. The E7 oncoprotein causes degradation of pRb, thereby promoting cell cycle progression. This results in p16 overexpression in HPV-associated OPSCC (Langendijk and Psyrri, 2010).

Deletion of 9p21.3 was observed in 16/31 (52%) of 0 viral load OPSCC and only 1/20 (5%) of OPSCC with viral load > 0. The TCGA cervical cancer subgroup did not identify loss of 9p21.3 in their CSCC cohort (Ojesina *et al.*, 2014). This is in keeping with the fact that all the TCGA CSCC cohort was HPV-positive. The fact that this gene was only deleted in 1/20 tumour with viral load > 0 is comparable to Seiwert *et al* who identified loss of CDKN2A in only 1/50 HPV-positive sample and 9/70 HPV-negative tumours (Seiwert *et al.*, 2015). p16 overexpression has been reported in up to 100% of HPV-positive HNSCC and 20% of HPV-negative HNSCC (Singhi and Westra, 2010).

Gain of 20p12.2 was identified in 13/31 (42%) of OPSCC tumours with 0 viral load and 1/20 (5%) of tumours with viral load > 0. Evaluation of this region identified the gene *JAG1*, previously identified as a NOTCH ligand. Amplifications of NOTCH pathway genes including *JAG1* have been recently reported to occur in 30% of HNSCC, though the prevalence in HPV-positive or negative HNSCC is not known (Sun *et al.*, 2014). In cervical cancer activation of the NOTCH pathway has been suggested to occur via *JAG1* and has been associated with poorer outcome, though again not specifically with HPV-status (Yousif *et al.*, 2015). The TCGA cervical cancer subgroup did not identify 20p12.2 as a focal amplification in their cohort (Ojesina *et al.*, 2014). This is consistent with the increased frequency of gain of *JAG1* in OPSCC tumours with 0 viral load. *JAG1* could represent further evidence for the therapeutic potential in targeting the NOTCH pathway, particularly in HPV-negative tumours.

Another gene found to be more frequently deleted in tumours with 0 viral load was *CSFR1*, found at 5q32 (deleted in 9/31 (29%) of 0 viral load OPSCC and 1/20 (5%) tumour with viral load > 0). This region is not highlighted as a focal deletion by the TCGA cervical cancer subgroup, nor is broad level loss of 5q (Ojesina *et al.*, 2014). *CSF1R* encodes for the colony-stimulating factor 1 (CSF-1) receptor, which is thought to play a role in

recruiting tumour-infiltrating macrophages in cancer (Pixley and Stanley, 2004). However, blockade of *CSF1R* signalling has been shown to enhance sensitivity to radiotherapy in murine models of prostate cancer (Xu *et al.*, 2013). HPV-positive OPSCC has been shown to have an improved response to chemoradiotherapy compared to HPV-negative OPSCC (Ang *et al.*, 2010). This is in contrast to the findings that deletion of *CSF1R* occurred more frequently in OPSCC with 0 viral load, and may indicate that this pathway does not play a strong role in HPV-positive tumours. In support of this finding is a recent study analysing the relationship of tumour-associated macrophages and head and neck cancer relapse. They found that increased expression of marker of M2 macrophages (CD163) only correlated with clinical outcome (relapse) in HPV-negative tumours (Balermpas *et al.*, 2014). The fact that *CSF1R* is deleted in only 9/31 (29%) of OPSCC tumour with 0 viral load suggests this could be a potential therapeutic target and could be used a marker for treatment selection in a significant number of HPV-negative tumours.

5.4.3 Pathways containing copy number altered genes in tumours of differing viral load

Analysis of pathways allows determination of potentially targetable options from this CNA data. The PI3K signalling pathway contained the most copy number altered genes from the OPSCC tumours with 0 viral load (see Figure 5-9). This pathway is involved in many cellular functions crucial to tumour survival including cell growth, cell movement, metastasis and survival (Cantley, 2002). This in keeping with a study by Lui *et al* which used whole exome sequencing of 151 HNSCC samples and identified *PI3K* mutations in 31% (46/151) (Lui *et al.*, 2013). Only 3 of these tumours containing *PI3K* mutations were HPV-negative, though only 15 tumours out of the entire cohort (n = 151) were HPV-positive. Far more of the genes identified in the PI3K pathway were deleted in this series compared to the number of amplifications (22 genes deleted vs. 5 genes amplified). 42/51 OPSCC tumours were found to contain multiple copy number altered genes in the PI3K pathway analogous to Lui *et al's* study, which found 22% of tumour containing PI3K pathway mutations, harboured multiple genetic alterations (Lui *et al.*, 2013). Of the remaining tumour 3/51 contained no copy number

altered genes in the PI3K pathway and 6/51 contained only 1 gene. Interestingly, 5/6 tumours with only one gene had a viral load > 0, again suggesting this pathway represents greater therapeutic potential in HPV-negative tumours.

The MAPK signalling pathway contained the most genes found to be copy number altered within tumours with > 0 viral load (and the fourth most in tumours with 0 viral load) (see Figure 5-9). In humans there are at least 11 members of the MAPK superfamily, all of which play an essential role in signal transduction in response to the extracellular environment (Leelahavanichkul *et al.*, 2014). Although fewer genes overall were identified to belong to this pathway 46/51 tumours were found to contain copy number altered genes. Of the 4 tumours without any copy number altered MAPK pathway genes 3 were tumours with > 0 viral load. Despite this being the pathway that most copy number altered genes were identified in tumour with > 0 viral load there, there were more hits for these genes within tumours with 0 viral load, indicating this is a significant pathway in both HPV-positive and negative tumours. Interestingly Lui et al found only 8.3% of tumours (12 of 151) to contain mutations in this pathway, but another recent study found overexpression of *p38α* (a MAPK superfamily member) in 79% of 293 OSCC samples examined with IHC (Leelahavanichkul *et al.*, 2014, Lui *et al.*, 2013). They also found it to be a positive regulator of angiogenesis and lymphangiogenesis indicating that though mutations may be infrequent, this pathway is an important one in the context of HNSCC carcinogenesis.

The JAK/STAT signalling pathway has been most extensively studied in haematologic malignancies, however it is recognised to be often activated in cancers of the upper aerodigestive tract, leading to increased cell survival, angiogenesis and immune system evasion (Lai and Johnson, 2010). Inhibition of this pathway has been shown to block cell invasion in HNSCC cell-line models (Lai *et al.*, 2005). This pathway contained the second highest number of copy number altered genes in the 0 viral load tumours, whereas it contained the sixth highest number of genes within the tumours with viral load >0. Interestingly the regions with the greatest clusters of genes within the JAK/STAT pathway were found on 3p and 9p, both of which were more frequently deleted in tumours with 0 viral load (see Table 5-4 and

Table **5-5**). 6/51 tumours were found to contain no copy number altered genes from the JAK/STAT pathway. Of these 5 were tumours with > 0 viral load. Combined with the fact that tumours with > 0 viral load contained far fewer hits for this pathway compared to the 0 viral load group (see Figure **5-9**), this suggests this pathway may represent greater therapeutic potential in HPV-negative tumours. A recent study by Bonner *et al* examining the use of an anti-JAK/STAT3 agent as a radiosensitiser in HNSCC cell-lines, in combination with Cetuximab (Bonner JA, 2015). This found a marked increase in apoptotic events, however these were reduced in the HPV-positive cell-lines compared to the HPV-negative cell-lines.

5.4.4 Indicators of overall genomic damage in tumours with differing viral load

Stransky et al found the mutational burden to be higher in HPV-negative compared to HPV-positive tumours, whilst the TCGA reported a similar overall mutation rate (Cancer Genome Atlas, 2015, Stransky *et al.*, 2011). This difference may relate to other clinical factors, for example Seiwert *et al* found a higher mutational burden in smokers versus nonsmokers (Seiwert *et al.*, 2015). In my series the FGA was found to be significantly lower in the tumours with viral load > 0. This measurement allows inference of genomic damage, but correlation of this to clinical outcomes and smoking status would reveal any clinical utility of this measure in assigning viral status and prognosis.

A similar analogy could be drawn from the frequency of copy number altered genes in pathways of potential significance (see Figure **5-9**). Regardless of the difference in frequency of individual pathways the overall number of hits obtained for each pathway is considerably higher in tumours with 0 viral load compared to those with > 0 viral load. This suggests that the number of genomic CNAs in regions that are significant is higher in tumours with 0 viral load.

5.4.5 Limitations

The most important limitation to this study is the limited number of patients, particularly when attempting to analyse the tumours with intermediate viral load. Considerable inter-tumour heterogeneity is demonstrated in both

tumours with 0 viral load and those with viral load > 0 and this does require greater numbers of samples in order to try and elucidate reliable patterns of genomic damage. Interestingly, however the heterogeneity does also suggest that no unifying genomic marker in HNSCC is present (excluding the almost universal presence of *TP53* mutation in HPV-negative tumours). The challenges in obtaining large numbers of OPSCC samples are clearly demonstrated by the fact that these tumours form the minority of the TCGA head and neck cohort. HPV-positive OPSCC are even more difficult to study in large numbers, primarily because most of these are now treated non-surgically. Therefore tissue samples are generally limited to small biopsies, which are not possible to replicate, as the patients rarely require repeat biopsy or surgery prior to receiving chemoradiation. My study does comprise one of the largest genomic studies of purely OPSCC samples.

Another limitation in this cohort is the lack of detailed clinical information such as smoking status, which could be responsible for different CNAs. It is quite possible that specific CNAs may reflect a different clinicopathologic characteristic of these samples, rather than their HPV-status, as is suggested by the genomic analysis of OSCC conducted by the India Genome Project Team (India Project Team of the International Cancer Genome, 2013). Their samples originate in a region where betel nut chewing is common in contrast to the UK and US. They found mutations specific to their cohort such as *USP9X, MLL4, ARID2, UNC13C*, found in 10-20% of their cohort and found at a rate of less than 10% in the TCGA head and neck cohort (when analysing the TCGA data via cBIOPortal) (Cerami *et al.*, 2012). Seiwert *et al* did report a significant association between smoking and mutations in *TP53, CSMD3, RB1CC1, THSD7A* (Seiwert *et al.,* 2015)*.* Though they did not report any association between smoking and CNA, this could be due to reducing numbers of patients in subgroup analyses. This can only be clarified by evaluating a larger number of patients including detailed clinical outcome data.

With this approach to analysing CNA we are unable to identify copy-neutral (balanced) rearrangements. This is a potential area of interest being missed and one that is little explored in HNSCC. It is also possible that smaller CNAs are present that are not visible at the sequencing resolution utilised in

this study. The fact that most CNAs, even of minimally altered regions, tend to be several genomic windows would suggest that resolution is not an important limitation. However the considerable heterogeneity present between tumours also suggests that genomic patterns that are useful as markers or therapeutic targets may only be present in small numbers of patients, or even in a single patient. If this is approach is to be taken then higher resolution CNA analysis would be useful to identify any and all potentially targetable events, if this information were to be used to try and stratify treatment.

5.5 Conclusion

This study used a low-cost form of NGS to provide genome-wide CNA analyses of FFPE OPSCC of specific HPV-16 viral loads. With reference to the specific aims of this chapter:

1. Producing a CNA profile for HPV positive OPSCC. Though of limited size, this project contained numbers of OPSCC tumours with viral load > 0 comparable to publications produced by the TCGA head and neck sub-group and affiliated groups (Seiwert *et al.*, 2015, Cancer Genome Atlas, 2015). This cohort identified a panel of CNAs associated with 0 viral load. The mean FGA in tumours with 0 viral load was also found to be significantly higher than those with a viral load > 0.

2. In evaluating if there is a specific viral load at which this genomic profile becomes apparent my results suggested that these differences become less distinct in tumours of intermediate viral load. These tumours may represent a different clinical sub-group. A different pattern of copy number altered genes and potentially significant gene pathways were also identified in tumours of differing viral load, highlighting not just potential markers but that these therapeutic targets may be of greater significance in HPV-positive or negative tumours dependent on the pathway. The FGA of tumours with differing viral load was also found to be significantly different suggesting the overall genomic damage to HPV-positive OPSCC tumours is lower, and may also be related to the improved clinical outcomes in these patients. The need to include detailed clinical data for future studies on these genomic alterations is vital to providing greater translational significance.

Chapter 6

MicroRNA profiling of HNSCC and matched cervical metastases to identify characteristics of metastatic and non-metastatic HNSCC

6.1 Introduction

Cancer is thought to progress through the progressive accumulation of genetic and epigenetic changes that allow the cancerous cell to avoid normal cellular controls. Changes to the levels of transcription of a gene into mRNA are referred to as altered gene expression. Alterations in gene expression are well recognised across many cancer types. In HNSCC, an important example of altered gene expression is shown by silencing of *TP53*, due to mutation or inactivation. This has been demonstrated in numerous studies to be the most common single gene to be affected across different studies of head and neck tumours (Califano *et al.*, 1996, Leemans *et al.*, 2011, Agrawal *et al.*, 2011, Poeta *et al.*, 2007). However, the gene expression profile of HNSCC is made up of hundreds of genes, with significant inter-tumour heterogeneity. Chung *et al,* identified four molecular subtypes of HNSCC (basal, atypical, mesenchymal and classical) using gene expression profiling of 60 tumours (Chung *et al.*, 2004).

Gene expression can be altered at an epigenetic, transcriptional, post-transcriptional, translation or post-translational level. miRNAs are a class of non-coding RNAs that regulate gene expression at the post-transcriptional level. Originally identified as "small temporal RNAs" that affected post-embryological development in *Caenorhabiditis elegans* (Lee *et al.*, 1993, Wightman *et al.*, 1993). Although, only discovered in humans around 14 years ago there has been a great deal of research into miRNAs in health and disease. It is estimated that the majority of coding genes are regulated by miRNAs (Lagos-Quintana *et al.*, 2001, Friedman *et al.*, 2009).

Calin *et al* first reported a direct link between miRNAs and cancer, demonstrating loss of miR-15 and miR-16 in chronic lymphocytic leukaemia (Calin *et al.*, 2002). miRNAs have been identified as playing roles in many different cancers since then. They have been found to be differentially expressed by normal and cancerous tissue and to reflect different tissues (Calin and Croce, 2006, Pichler and Calin, 2015). Studies have revealed miRNAs can have an oncogenic (referred to as oncomiRs) and/or a tumour suppressor effect (Sethi *et al.*, 2014). In addition, a single miRNA has been shown to effect the expression of multiple target genes (Jansson and Lund, 2012).

Studies in HNSCC have demonstrated that aberrant expression of miRNAs is associated with tumourigenesis and chemoradiotherapy resistance (Sethi *et al.*, 2014). miRNAs have also shown potential as a biomarker to stratify patient risk in a number of cancers (Mar-Aguilar *et al.*, 2013, Mishra, 2014). In colorectal cancer miRNA profiles have been shown to discriminate between cancer that metastasised to lymph nodes and those, which metastasised to liver (Drusco *et al.*, 2014). One drawback to profiling miRNAs is that our knowledge of them is evolving. Shortly after being first discovered in humans it was estimated that there might be over 100 miRNAs (Bartel, 2004). Over 1000 miRNAs have subsequently been catalogued in the miRNA database (miRBase, www.mirbase.org) with more being added every year (Griffiths-Jones, 2004). Therefore studies reliant upon array-based technology are limited by the fact that they can only identify miRNAs that are present on the array. Using NGS to identify miRNAs allows all known miRNAs to be searched for as well as the fact that the data can be repeatedly interrogated for novel miRNAs without the cost or time involved in re-sequencing the sample.

A challenge when working with RNA is the issue of degradation. It has long been regarded that RNA extracted from FFPE tissue is significantly degraded (Stanta and Schneider, 1991, Lewis *et al.*, 2001). However miRNAs have been shown to be well preserved without degradation, even in archived FFPE tissue samples (Szafranska *et al.*, 2008, Weng *et al.*, 2010, Liu and Xu, 2011, Kolbert *et al.*, 2013). This is thought to be due to the short length of miRNAs meaning they do not undergo degradation to smaller

components readily. In light of this, FFPE tissue archives hold great potential for research into miRNAs.

Of particular interest is the fact that another study by Calin *et al* showed that miRNAs are frequently located at fragile genomic sites associated with CNA in cancer (Calin *et al.*, 2004). Further studies in breast, colon, ovarian and prostate cancer have shown that miRNAs and miRNA-binding sites are significantly overrepresented in genomic regions with copy number gain and underrepresented in regions with copy number loss (Calin *et al.*, 2004, Zhang *et al.*, 2006). Zhang *et al* found that genomic alterations involving miRNAs were highly frequent in epithelial cancer and resulted in changes in miRNA expression. They also found that some deleted miRNAs were not under-expressed and therefore may be being rescued as part of the oncogenic process (Zhang *et al.*, 2006). As a result integrating CNA and miRNA expression data could reveal important miRNAs that could be used as a biomarker for metastasis.

6.2 Aim

This study aimed to:

1. Evaluate and compare the miRNA profile of metastatic primary tumours and a paired lymph node metastasis.
2. It then compared non-metastatic primary tumours to the metastatic miRNA profile to identify any potential markers for nodal metastasis.

6.3 Results

6.3.1 Summary of patient cohort

The patients included in this study were selected from the patient cohort identified in Chapter 3. The clinicopathologic characteristics of these groups were identical to those specified in Chapter 3 and 4 (see Table 6-1). Patients were selected to try and reduce heterogeneity. Inclusion criteria were oral tongue primary tumour, successful DNA extraction for CNA analysis, known current smoking status at the time of surgery. After discussion with Dr Alastair Droop (Bioinfomatics Research Fellow, University of Leeds). It was decided that a minimum of three patients in each group was necessary for the purposes of comparing miRNA profiles. From each patient, in addition to a sample from the primary tumour, a sample of normal epithelium was also used for RNA extraction and sequencing. The metastatic primary tumours had RNA extracted from a single paired lymph node metastasis (resulting in a total of 24 samples from nine patients). The overall number of patients was dictated by the desire to provide a matched normal sample as well as the overall cost of each sample.

	T1-4 N0	T1-4 N1-3 No ECS	T1-4 N1-3 with ECS
Number of patients	3	3	3
Tissue Samples	Normal epithelium, primary tumour	Normal epithelium, primary tumour, lymph node metastasis	Normal epithelium, primary tumour, lymph node metastasis
Sample Codes	ECS026 ECS029 ECS042	ECS033 ECS054 ECS055	ECS040 ECS060 ECS084

Table 6-1: Table to show the number of patients in each group and tissue samples used from each.

All tissue blocks had a fresh 5 μm section stained with haematoxylin and eosin, examined by Professor Kenneth MacLennan or Dr Preetha Chengot (consultant head and neck histopathologists) to mark the area of highest tumour cell content (in both primary tumours and lymph node metastases) as well as identify the normal epithelium on a separate tissue block to those

containing primary tumour. These areas were macrodissected and the total RNA (including miRNA) extracted as described in section Chapter 2.6. These were then all successfully processed to a small RNA sequencing library and multiplexed at 12 samples per lane on the Illumina HiSeq 2500 on rapid mode. This was performed using 50 bp single end reads.

6.3.2 miRNA sequencing (miRNASeq) of HNSCC primary tumours and matching nodal metastases

6.3.2.1 Processing miRNASeq data for differential expression analysis

The FASTQ files generated by the sequencer containing the raw sequencing read data were downloaded. Cutadapt was used to remove the adaptor sequences from the ends of each read (http://code.google.com/p/cutadapt/). Open access software, CAP-miRSeq (Comprehensive Analysis Pipeline – microRNA Sequencing), was then used to process the data (Sun *et al.*, 2014). This automatically aligned the trimmed reads to the human reference genome (University of California Santa Cruz version GRCh37/hg19, http;//genome.ucsc.edu). After trimming all reads less than 17 bp in length were removed. All RNA types contained in the sequencing library were quantified and even data on single nucleotide variants in the coding regions of miRNAs was available. CAP-miRSeq generated raw count data for each uniquely mapped miRNA and performed normalisation of these for the number of aligned reads per sample.

CAP-miRSeq then implemented another open access software package called edgeR (Bioconductor, US) to perform further normalisations steps. These include scaling of the libraries for the number of aligned reads per sample as well as the dispersion of reads in each sample. It then generated differential expression data according to the clinicopathologic groups defined (Robinson *et al.*, 2010).

6.3.2.2 Analysis of read counts

Due to the highly specific nature of the sequencing performed, using the overall coverage of the sequencing performed is a somewhat misleading measurement, in isolation. Small RNA sequencing targets small reads (17 – 22 bp) and has a limited number of overall targets (only those registered as

Homo sapiens miRNAs in miRBase) (Griffiths-Jones, 2004). An average of 5,078,953 (range: 2,199,736 – 10,395,214) sequencing reads were obtained for each sample. Of these an average of 3,598,699 reads (range: 1,478,161 – 7,798,540) aligned to the human reference genome, GRCh38 (UCSC, California, USA). On average 1,706,371 (range: 525,093 – 4,513,312) aligned to the *Homo sapiens* miRNAs registered in miRBase, version 21 (Griffiths-Jones, 2004). Of the reads aligned to miRBase there were on average 526 miRNAs (range: 322 – 714) with at least 5x coverage per sample.

6.3.2.3 Comparison of miRNASeq profile of nodal metastases to matched metastatic primary tumours

Gene expression has been shown to vary by tissue type and amongst individuals (Whitney *et al.*, 2003, Cobb *et al.*, 2005, Palmer *et al.*, 2006). In order to evaluate how similar nodal metastases were to their matching primary tumour, the normalised mature miRNA expression was used to compare metastatic primary tumours to their matched metastases. Pearson's product moment correlation was calculated comparing each nodal metastasis to each metastatic primary tumour using R (Vienna, Austria) (Team, 2015).

Overall this found that the correlation between metastatic primary and matched nodal metastasis was high, ranging from 0.74–0.99 (see

Table **6-2**). However the matching tumour-metastasis pair often did not have the highest correlation. In only 2/6 pairs did the matched metastasis have the highest correlation. In the other 4 pairs another metastasis had a higher correlation. This could reflect the fact that gene expression varies between individuals and is affected by behaviours such as smoking or diet (Jaenisch and Bird, 2003, Landi *et al.*, 2008). It could also be due to the fact that cells within the metastasis are subject to different environmental pressures and develop different expression profiles secondary to this. The similarity between metastatic primary tumours and nodal metastases may also reflect there are similar changes in miRNA expression involved in the progression of cancer across all of the samples. This variation between different

individuals may have obscured the variation due to carcinogenesis or the metastatic process and represent a confounder in these comparisons.
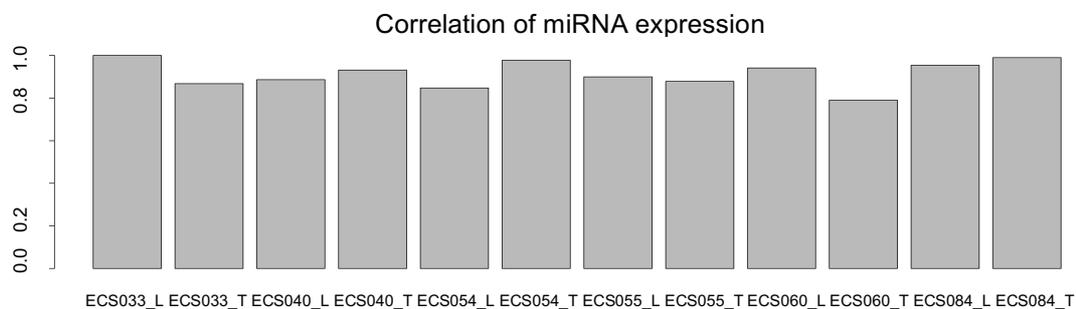
Figure 6-1: Pearson's product moment correlation on comparing ECS033 nodal metastasis (ECS033_L) to itself and every other nodal metastasis and metastatic primary tumour. Correlation with itself is 100%, whilst correlation with its matched metastatic primary tumour (ECS033_T) is 87%. The highest correlation (97%) was found to be with the metastatic primary tumour for patient ECS084.

In order to reduce the background noise produced by the variation in miRNA expression found in different individuals, paired analyses comparing each nodal metastasis and metastatic primary tumour to its matching normal epithelium were performed. This was done using CAP-miRSeq with the significance threshold of 0.25 for the p-value adjusted for multiple testing of each miRNA found to be differentially expressed.

	ECS033 _L	ECS033 _T	ECS040 _L	ECS040 _T	ECS054 _L	ECS054 _T	ECS055 _L	ECS055 _T	ECS060 _L	ECS060 _T	ECS084 _L	ECS084 _T
ECS033 _L	1.00	0.87	0.89	0.93	0.85	0.98	0.90	0.88	0.94	0.79	0.95	0.99
ECS033 _T	0.87	1.00	0.97	0.97	0.87	0.88	0.97	0.94	0.86	0.96	0.94	0.84
ECS040 _L	0.89	0.97	1.00	0.97	0.92	0.89	0.96	0.94	0.88	0.94	0.96	0.86
ECS040 _T	0.93	0.97	0.97	1.00	0.88	0.94	0.96	0.92	0.91	0.91	0.96	0.92
ECS054 _L	0.85	0.87	0.92	0.88	1.00	0.86	0.84	0.81	0.88	0.79	0.92	0.83
ECS054 _T	0.98	0.88	0.89	0.94	0.86	1.00	0.88	0.84	0.98	0.78	0.96	0.98
ECS055 _L	0.90	0.97	0.96	0.96	0.84	0.88	1.00	0.99	0.84	0.97	0.93	0.86
ECS055 _T	0.88	0.94	0.94	0.92	0.81	0.84	0.99	1.00	0.79	0.97	0.91	0.83
ECS060 _L	0.94	0.86	0.88	0.91	0.88	0.98	0.84	0.79	1.00	0.74	0.95	0.95
ECS060 _T	0.79	0.96	0.94	0.91	0.79	0.78	0.97	0.97	0.74	1.00	0.86	0.74
ECS084 _L	0.95	0.94	0.96	0.96	0.92	0.96	0.93	0.91	0.95	0.86	1.00	0.95
ECS084 _T	0.99	0.84	0.86	0.92	0.83	0.98	0.86	0.83	0.95	0.74	0.95	1.00

Table 6-2: Correlation of normalised mature miRNA expression between metastatic primary tumours and nodal metastases. Tumour-metastasis pairs are shaded the same colour. Primary tumours are denoted by the suffix '_T' and lymph node metastases by '_L'.

The miRNA profile for metastatic primary tumours was compared to that of the matching normal epithelium. The miRNA profile of metastatic primary tumours (n = 6) was then compared to the matched normal epithelium (n = 6). Across these sample 1026/2578 miRNAs were not detected in any sample. Comparing the two groups in pairs, in total, 186 miRNAs were identified as significantly differentially expressed (adjusted p-value = < 0.25) between metastatic primary tumours and matched normal (see Table **6-3**). For brevity only the top ten in each analysis is shown, the full lists of each differential expression analysis are in Appendix 8.5.

	log Fold change	p-value	Adjusted p-value
hsa-miR-375	-8.02	2.09E-39	5.40E-36
hsa-miR-196b-5p	5.27	2.14E-17	2.75E-14
hsa-let-7c-5p	-3.45	2.22E-16	1.90E-13
hsa-miR-615-3p	7.54	4.29E-15	2.76E-12
hsa-miR-196a-5p	5.44	9.14E-14	4.71E-11
hsa-miR-1247-5p	-3.54	1.82E-10	7.82E-08
hsa-miR-99a-5p	-3.09	3.33E-10	1.22E-07
hsa-miR-1910-5p	4.23	1.08E-09	3.47E-07
hsa-miR-150-5p	-2.50	3.80E-09	9.81E-07
hsa-miR-125b-5p	-2.41	4.00E-09	9.81E-07

Table 6-3: Top 10 significantly differentially expressed miRNAs between metastatic primary tumours and matched normal epithelium (adjusted p-value = < 0.25).

The miRNA profiles from nodal metastases (n = 6) were then compared to the matching normal epithelium (n = 6) using CAP-miRSeq. In total, 2578 miRNAs were screened for across the samples. Of these, 1070 miRNAs were not detected in any sample. Overall 131 miRNAs were identified as significantly differentially expressed between the nodal metastases and matched normal (see Table **6-4**).

On cross-referencing these two lists of differentially expressed miRNAs, 95 were found to be duplicated in both (with concordant change in expression). This demonstrates that a large proportion of differentially expressed miRNAs are common to both metastatic primary tumours and nodal metastases. The

fact that the miRNA signatures are not identical is in keeping with the fact that HNSCC tumours are made up of molecularly heterogeneous clonal populations. It is also consistent with the hypothesis (maintained from the comparison of CNA profiles in Chapter 4) that the metastatic clonal population is more likely to be dominant within the nodal metastasis tissue and therefore less prone to sampling mixed clonal populations.

	log Fold change	p-value	Adjusted p-value
hsa-miR-375	8.82	6.66E-32	1.72E-28
hsa-miR-196a-5p	-5.93	2.23E-25	2.87E-22
hsa-miR-133b	7.44	2.68E-19	2.30E-16
hsa-miR-1	6.59	6.04E-18	3.89E-15
hsa-miR-615-3p	-7.83	2.06E-17	1.06E-14
hsa-miR-133a-3p	7.15	2.85E-17	1.22E-14
hsa-miR-196b-5p	-4.57	9.76E-17	3.28E-14
hsa-miR-135a-5p	6.24	1.02E-16	3.28E-14
hsa-miR-206	7.76	6.61E-14	1.89E-11
hsa-let-7c-5p	3.35	1.05E-13	2.72E-11

Table 6-4: Top 10 list of significantly differentially expressed miRNAs (adjusted p-value = < 0.25) between nodal metastases and matched normal epithelium.

On this basis, in order to try and identify a molecular profile associated with metastasis it is necessary to sample and analyse the metastatic tissue rather than just the metastatic primary tumour. The molecular markers identified should then be cross-referenced against non-metastatic primary tumours to try and eliminate miRNAs associated with cancer development rather than the metastatic process.

6.3.2.4 Comparison of the miRNASeq profile of non-metastatic primary tumours to nodal metastases

A paired analysis of the miRNA profiles of non-metastatic primary tumours (n = 3) was then compared to that of the matching normal epithelium (n = 3). Across these samples 1221 miRNAs were not detected in any sample.

Overall 69 miRNAs were identified as significantly differentially expressed (adjusted p-value = < 0.25) between non-metastatic primary tumours and the matching normal (see Table **6**-**5**).

	log Fold change	p-value	Adjusted p-value
hsa-miR-1269a	12.80866958	7.52E-17	1.94E-13
hsa-miR-615-3p	8.538011515	2.23E-13	2.88E-10
hsa-miR-1269b	11.44791655	3.48E-11	2.99E-08
hsa-miR-1910-5p	4.231464778	6.29E-09	4.06E-06
hsa-miR-196a-5p	4.921079066	4.54E-08	2.34E-05
hsa-miR-431-5p	3.691385296	7.32E-07	0.000314334
hsa-miR-29c-3p	-2.365685094	5.00E-06	0.001782227
hsa-miR-4713-5p	4.017179645	5.53E-06	0.001782227
hsa-miR-139-5p	-2.979490344	6.49E-06	0.001859884
hsa-miR-4521	-3.095342054	1.57E-05	0.004049336

Table 6-5: Top 10 significantly differentially expressed miRNAs between non-metastatic primary tumours and matched normal epithelium (adjusted p = < 0.25).

This list of differentially expressed miRNAs was cross-referenced against the miRNAs identified in nodal metastases (see Table **6**-**4**). Duplicates were removed to identify miRNAs specifically associated with metastasis rather than carcinogenesis. The unique differentially expressed miRNAs in nodal metastases are more likely to be associated with the metastatic process. This identified 97 miRNAs whose differential expression was uniquely associated with metastasis (see Table **6**-**6**).

	log Fold change	p-value	Adjusted p-value
hsa-miR-133b	7.44	2.68E-19	2.30E-16
hsa-miR-1	6.59	6.04E-18	3.89E-15
hsa-miR-133a-3p	7.15	2.85E-17	1.22E-14
hsa-miR-206	7.76	6.61E-14	1.89E-11
hsa-let-7c-5p	3.35	1.05E-13	2.72E-11
hsa-miR-99a-3p	2.91	4.94E-11	1.06E-08
hsa-miR-99a-5p	3.39	9.21E-11	1.83E-08
hsa-miR-125b-2-3p	2.54	7.71E-09	1.24E-06
hsa-miR-208b-3p	6.52	1.50E-08	2.15E-06
hsa-miR-21-5p	-2.46	1.71E-08	2.32E-06

Table 6-6: Top 10 differentially expressed miRNAs uniquely associated with metastasis.

In order for miRNAs to be of potential patient benefit as a biomarker for metastasis, it is necessary to be able to identify them within the primary tumour. Therefore, the list of miRNAs uniquely associated with metastasis (see Table 6-6) was cross-referenced against the differentially expressed miRNAs identified in metastatic primary tumours (see Table 6-3). This identified 63 miRNAs in common to both lists that could therefore be of potential use as a marker for metastasis within the primary tumour (see Table 6-7).

	log Fold change	p-value	Adjusted p-value
hsa-let-7c-5p	-3.45	2.22E-16	1.90E-13
hsa-miR-99a-5p	-3.09	3.33E-10	1.22E-07
hsa-miR-125b-5p	-2.41	4.00E-09	9.81E-07
hsa-miR-99a-3p	-2.93	5.17E-08	9.11E-06
hsa-miR-4776-5p	-5.59	4.17E-07	4.88E-05
hsa-miR-6842-3p	4.79	6.22E-07	6.68E-05
hsa-miR-10b-5p	-2.38	8.11E-07	8.36E-05
hsa-miR-885-5p	-4.95	1.54E-06	0.000153
hsa-miR-31-3p	3.42	2.00E-06	0.0001913
hsa-miR-21-5p	2.36	3.92E-06	0.000337

Table 6-7: Top 10 differentially expressed miRNA associated with metastasis that are identifiable within the metastatic primary tumour (adjusted p = < 0.25).

6.3.2.5 Comparison of miRNASeq profile of nodal metastases with and without ECS

Paralleling the analysis of CNA performed in Chapter 4 a further breakdown of the metastatic samples was made. In order to try and identify any miRNAs whose differential expression was associated with ECS the nodal metastases were divided into those with and without ECS. A paired analysis was performed comparing the miRNA profiles of nodal metastases with ECS (n = 3) to matching normal epithelium. This identified 131 differentially expressed miRNAs, meeting the significance threshold of < 0.25 for the p-value adjusted for multiple testing (see Table **6**-**8**).

	log Fold change	p-value	Adjusted p-value
hsa-miR-133a-3p	7.90	2.54E-27	6.55E-24
hsa-miR-208b-3p	8.25	4.21E-19	5.42E-16
hsa-miR-1	6.31	2.07E-17	1.77E-14
hsa-miR-375	8.04	4.79E-16	3.09E-13
hsa-miR-615-3p	-7.89	3.47E-15	1.79E-12
hsa-miR-133b	7.59	7.14E-15	3.07E-12
hsa-miR-206	8.66	1.74E-13	6.39E-11
hsa-miR-196a-5p	-5.26	2.68E-13	8.62E-11
hsa-miR-196b-5p	-4.48	3.07E-11	8.78E-09
hsa-miR-135a-5p	6.00	7.24E-11	1.87E-08

Table 6-8: Top 10 significantly differentially expressed miRNAs when comparing nodal metastases with ECS to matching normal epithelium (adjusted p = < 0.25).

A similar paired analysis was performed comparing the miRNA profiles of nodal metastases without ECS (n = 3) to matching normal epithelium. This identified 32 differentially expressed miRNAs using the significance threshold of < 0.25 for the adjusted p-value (see Table **6**-**9**). These two lists were then cross-referenced. Duplicates were removed to leave miRNAs whose differential expression was more likely to be associated with ECS rather than metastasis. The remaining miRNA list was then cross-referenced against the list of differentially expressed miRNAs identified on comparing non-metastatic primary tumours to matching normal epithelium (see Table **6**-**5**).

	log Fold change	p-value	Adjusted p-value
hsa-miR-375	9.30	2.55E-22	6.56E-19
hsa-miR-196a-5p	-6.86	3.27E-11	4.21E-08
hsa-miR-3168	11.09	4.43E-08	3.80E-05
hsa-miR-211-5p	7.45	2.16E-07	0.000139519
hsa-miR-135a-5p	6.39	3.39E-07	0.000174653
hsa-miR-133b	6.74	4.56E-07	0.000195749
hsa-miR-615-3p	-7.87	6.56E-07	0.000224872
hsa-miR-196b-5p	-4.86	6.98E-07	0.000224872
hsa-miR-21-3p	-4.21	8.87E-07	0.000254178
hsa-miR-99a-5p	3.97	1.10E-06	0.000283051

Table 6-9: Top 10 significantly differentially expressed miRNAs when comparing nodal metastases without ECS to matching normal epithelium (adjusted p = < 0.25).

	Log Fold change	p-value	Adjusted p-value
hsa-miR-208b-3p	8.25	4.21E-19	5.42E-16
hsa-miR-206	8.66	1.74E-13	6.39E-11
hsa-miR-133a-5p	7.78	1.54E-10	3.31E-08
hsa-miR-99a-3p	3.22	3.14E-09	5.40E-07
hsa-miR-561-5p	4.28	4.68E-08	7.10E-06
hsa-miR-499a-5p	3.90	1.71E-07	2.45E-05
hsa-miR-146b-5p	-3.09	2.46E-07	3.34E-05
hsa-miR-455-3p	-2.85	4.99E-07	6.44E-05
hsa-miR-2355-5p	-3.42	1.32E-06	0.000149568
hsa-miR-10a-5p	-3.10	2.27E-06	0.000244115

Table 6-10: Top 10 significantly differentially expressed miRNAs associated with ECS.

Duplicates were again removed to avoid miRNAs associated with carcinogenesis alone. This approach discovered 81 differentially expressed miRNAs potentially associated with ECS (see Table **6-10**).

An attempt was then made to validate the miRNASeq findings using the Nanostring nCounter miRNA assay and replicating the sequencing data analyses.

6.3.3 nCounter miRNA assay profiling of HNSCC primary tumours and matched nodal metastases

The Nanostring nCounter expression profiling system is a relatively recent innovation. It uses hybridisation-based technology that can detect specific nucleic acid molecules from very low amounts of input material. It avoids the need for reverse transcription or PCR-based cDNA amplification (as is required in the preparation of miRNA sequencing libraries). Multiplexed probes are hybridised to specific transcripts and generate fluorescent signals, which are counted for each miRNA. Raw molecular counts are obtained (Geiss *et al.*, 2008, Tam *et al.*, 2014). Advantages of this technology include the reduced manipulation of input RNA, avoiding PCR amplification cycles. It only allows testing of 813 miRNAs simultaneously and therefore not all sequencing findings can be validated. The assay used is also unable to identify potentially novel miRNAs. It has also been shown to have a high correlation with miRNA sequencing and microarray-generated data from both cell lines and human tissue (Kolbert *et al.*, 2013, Knutsen *et al.*, 2013, Tam *et al.*, 2014). It was therefore selected as a method of confirming miRNA expression profile generated from NGS.

The same RNA sample from each patient sample used to create miRNA sequencing libraries (see Table **6**-**1**) was used to obtain 150 ng of input RNA for the Nanostring nCounter miRNA Assay. This was performed with Bruno Steinkraus and Tudor Fulga (Weatherall Institute of Molecular Medicine, Oxford), one of only two centres in the UK to have a Nanostring nCounter system.

The raw molecular count data was then processed using a software package called nSolver (Nanostring, Seattle USA) specifically designed to perform quality control and normalisation steps. This analysed the internal positive and negative controls to ensure that ligation worked appropriately. It also performed normalisation of the raw count data to five housekeeping genes included in the assay cartridge. This normalised count data was then input to the edgeR software used by CAP-miRSeq to perform paired analyses as in 6.3.2.1.

6.3.3.1 Correlation of nCounter and miRNASeq profiles

In an effort to compare the two technologies (NGS and Nanostring nCounter assay) a correlation analysis of the expression levels detected by both was performed.

The nCounter miRNA assay contained probes for 813 mature miRNAs. It provides raw molecular counts for these as an output. After normalisation was performed as above the list of miRNAs was ranked by count. The normalised sequencing read counts for the same 813 miRNAs generated from the miRNASeq FASTQ files were also ranked by count. This was performed for each sample (n = 24).
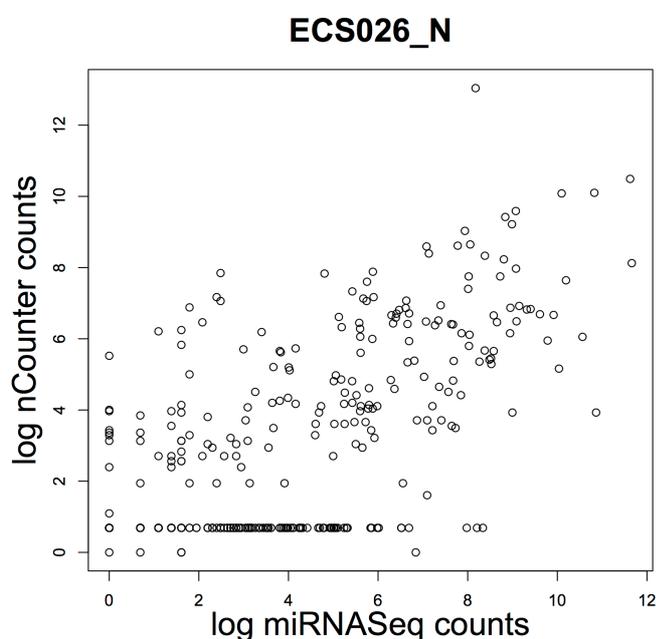
Figure 6-2: Plot of log counts (expression level) of each miRNA for both miRNASeq and nCounter for the sample ECS026_N, rho = 0.533 (p = <0.000001).

 Each list of counts was then tested for normality using the Shapiro-Wilk test. These were all found to be not normally distributed. As such, the log counts for all 813 miRNAs on each platform were plotted against each other and correlation analysed using Spearman's rank correlation coefficient (rho). An example of this is shown in Figure 6-2 for the sample ECS026_N. For this sample rho = 0.533, indicating a moderate correlation. This process was repeated for all samples (see Appendix 8.4 for all correlation scatterplots).

For all samples the correlation coefficients varied from 0.481–0.647, indicating a moderate to strong correlation).

Sample code	Spearman's correlation coefficient (rho)	p-value
ECS026_N	0.533	<0.000001
ECS026_T	0.589	<0.000001
ECS029_N	0.566	<0.000001
ECS029_T	0.612	<0.000001
ECS033_L	0.573	<0.000001
ECS033_N	0.540	<0.000001
ECS033_T	0.647	<0.000001
ECS040_L	0.608	<0.000001
ECS040_N	0.593	<0.000001
ECS040_T	0.581	<0.000001
ECS042_N	0.576	<0.000001
ECS042_T	0.547	<0.000001
ECS054_L	0.559	<0.000001
ECS054_N	0.509	<0.000001
ECS054_T	0.556	<0.000001
ECS055_L	0.589	<0.000001
ECS055_N	0.532	<0.000001
ECS055_T	0.573	<0.000001
ECS060_L	0.481	<0.000001
ECS060_N	0.562	<0.000001
ECS060_T	0.562	<0.000001
ECS084_L	0.592	<0.000001
ECS084_N	0.606	<0.000001
ECS084_T	0.572	<0.000001

Table 6-11: List of Spearman's correlation coefficients calculated for all tissue samples (n = 24).

The correlation scatterplots indicate that both miRNASeq and nCounter fail to detect miRNAs that are reported to have a detectable expression level in the other platform, though this number is higher in miRNASeq. This does not appear to reflect a greater sensitivity of NGS for low abundance of miRNAs but rather each platform has a greater propensity to detect a different spectrum of miRNAs. A previous analysis by Kolbert *et al* compared correlation of miRNA expression levels across a number of platforms (NGS,

microarray, qRT-PCR and nCounter) (Kolbert *et al.*, 2013). This reported a very strong correlation (r = 0.935) between NGS and nCounter miRNA detected expression levels when utilising FFPE tissue. When a subset of miRNAs (37) from each platform was compared to qRT-PCR the nCounter platform demonstrated the highest correlation, but even this was only moderate (r = 0.481). Kolbert *et al* concluded that the choice of platform would depend upon pragmatic factors such as aim of project, finance and time given the commonly interrogated miRNAs were generally similar amongst the platforms (Kolbert *et al.*, 2013). Though my correlation coefficients were lower than Kolbert *et al*, this may reflect variation in tissue type as they examined lung tissue (Kolbert *et al.*, 2013). They also only examined two sample of FFPE tissue and therefore represent a much smaller group. My results suggest that the nCounter miRNA assay is a reasonable choice of technique to use to attempt to validate the miRNASeq data. However, lack of validation by nCounter should not preclude highly significant miRNAs identified using miRNASeq data from being considered in future studies using alternative technology (e.g. microarrays). They also highlight the fact that miRNASeq can screen from over double the number of miRNAs that the nCounter assay can. Therefore miRNAs not validated by nCounter data that are absent from the assay probes should also be considered for investigation in future studies.

6.3.3.2 Comparison of nCounter and miRNA profiles of nodal metastases and matched normal epithelium

The nCounter miRNA profile of nodal metastases was compared to matched normal epithelium, using the significance threshold of < 0.25 for the p-value adjusted for multiple testing. Forty-five miRNAs were found to be differentially expressed (see Table **6-12**).

Of the 45 significantly differentially expressed miRNAs identified using nCounter data, 25 were found in the miRNASeq data.

	log Fold change	p-value	Adjusted p-value
hsa-miR-206	7.47	2.25E-22	1.85E-19
hsa-miR-133a	6.64	8.12E-14	3.34E-11
hsa-miR-375	6.00	5.07E-12	1.39E-09
hsa-miR-424-5p	-4.89	3.04E-11	6.24E-09
hsa-miR-21-5p	-3.35	7.06E-09	1.16E-06
hsa-miR-99a-5p	2.78	2.56E-08	3.51E-06
hsa-miR-503	-4.47	5.99E-08	7.03E-06
hsa-miR-455-5p	-4.37	1.22E-07	1.25E-05
hsa-miR-1246	-5.21	2.91E-07	2.66E-05
hsa-miR-455-3p	-4.37	9.26E-07	7.61E-05

Table 6-12: Top 10 nCounter generated significantly differentially expressed miRNAs between nodal metastases and matched normal epithelium.
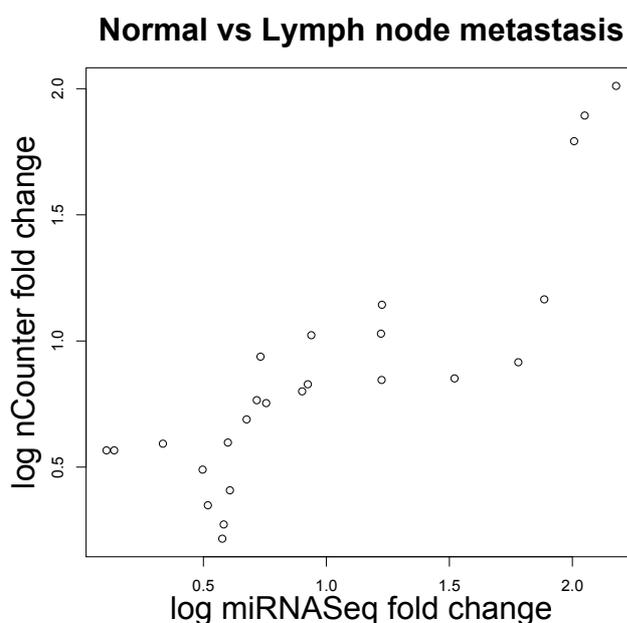
Figure 6-3: Plot of log fold changes generated for 25 miRNAs common to differential expression analysis between lymph node metastases and matched normal epithelium.

The lists of the 25 miRNAs common to both platforms were ranked in order of log fold change (see Table **6-13**) and again plotted against each other (see Figure 6-3). Spearman's correlation coefficient was 1.0 (p = < 0.000001), suggesting a very strong for miRNAs when identified at a significant level (adjusted p-value < 0.25). This again reassures that the nCounter miRNA assay is a reasonable choice of technique to try and confirm miRNASeq data.

Mature miRNA	miRNASeq log Fold change	Mature miRNA	nCounter log fold change
hsa-miR-375	8.82	hsa-miR-206	7.47
hsa-miR-206	7.76	hsa-miR-133a	6.64
hsa-miR-133b	7.44	hsa-miR-375	6.00
hsa-miR-1	6.59	hsa-miR-204-5p	3.20
hsa-miR-204-5p	3.40	hsa-miR-1	3.13
hsa-miR-99a-5p	3.39	hsa-miR-203	2.79
hsa-miR-378i	2.55	hsa-miR-99a-5p	2.78
hsa-miR-100-5p	2.07	hsa-miR-133b	2.55
hsa-miR-125b-5p	2.04	hsa-miR-376c	2.14
hsa-miR-378a-3p	1.82	hsa-miR-125b-5p	1.81
hsa-miR-23b-3p	1.39	hsa-miR-23b-3p	1.80
hsa-miR-130b-3p	-1.11	hsa-miR-378a-3p	1.76
hsa-miR-421	-1.14	hsa-miR-378i	1.76
hsa-miR-455-5p	-1.64	hsa-let-7c	1.63
hsa-miR-135b-5p	-1.67	hsa-miR-100-5p	1.41
hsa-miR-450a-5p	-1.78	hsa-miR-29c-3p	1.24
hsa-miR-424-5p	-1.79	hsa-miR-125a-5p	-1.31
hsa-miR-7-5p	-1.83	hsa-miR-135b-5p	-1.50
hsa-miR-146b-5p	-1.96	hsa-miR-193b-3p	-1.99
hsa-miR-455-3p	-2.12	hsa-miR-630	-2.12
hsa-miR-21-5p	-2.46	hsa-miR-132-3p	-2.22
hsa-miR-31-5p	-2.52	hsa-miR-130b-3p	-2.28
hsa-miR-1246	-3.40	hsa-miR-155-5p	-2.32
hsa-miR-196b-5p	-4.57	hsa-miR-7-5p	-2.34
hsa-miR-196a-5p	-5.93	hsa-miR-28-5p	-2.49

Table 6-13: Table of 25 significant miRNAs (adjusted p-value < 0.25) common to differential expression analysis performed with miRNASeq and nCounter generated data. This compared each lymph node metastasis to matched normal epithelium. The miRNAs are ranked in order of log fold change.

The list of 25 miRNAs common to both platforms demonstrates that, though the correlation of expression is moderate to strong, when miRNAs are identified on both platforms, they are detected at similar levels. Though there are differences in the rank order according to fold change there is consistency in the direction of fold change (increased or decreased expression. Again suggesting the nCounter assay to be a practical choice of validation technique.

6.3.3.3 Comparison of nCounter miRNA profiles of non-metastatic primary tumours to nodal metastases

The nCounter miRNA profiles of non-metastatic primary tumours were then compared to their matched normal epithelium. All 800 miRNA probes were detectable in at least one of the samples. This identified 27 differentially expressed miRNAs (see Table **6-14**).

	Log Fold change	P-value	Adjusted p-value
hsa-miR-221-3p	2.28	2.94E-08	2.42E-05
hsa-miR-4286	2.38	5.17E-07	2.12E-04
hsa-miR-29c-3p	-2.05	3.16E-06	8.66E-03
hsa-miR-424-5p	5.28	1.59E-05	3.27E-03
hsa-miR-183-5p	3.75	5.98E-05	9.83E-03
hsa-miR-182-5p	3.70	8.49E-05	0.012
hsa-miR-455-3p	3.96	0.000104869	0.012
hsa-miR-375	-5.35	0.000157489	0.016
hsa-miR-485-3p	4.05	0.000203723	0.018
hsa-miR-199b-5p	-1.54	0.000273739	0.02

Table 6-14: Top 10 nCounter generated significantly differentially expressed miRNAs (adjusted p-value = < 0.25) between non-metastatic primary tumours and matched normal epithelium.

	Log Fold change	p-value	Adjusted p-value
hsa-miR-206	7.47	2.25E-22	1.85E-19
hsa-miR-133a	6.64	8.12E-14	3.34E-11
hsa-miR-99a-5p	2.78	2.56E-08	3.51E-06
hsa-miR-196b-5p	-3.89	4.87E-06	3.33E-04
hsa-miR-944	-3.57	1.65E-05	0.0010
hsa-miR-31-5p	-3.14	3.88E-05	0.0022
hsa-miR-196a-5p	-3.47	4.80E-05	0.0026
hsa-let-7c	1.63	8.08E-05	0.0040
hsa-miR-146b-5p	-4.75	8.38E-05	0.0040
hsa-miR-28-5p	-2.49	0.000119114	0.0054

Table 6-15: Top 10 differentially expressed miRNAs identified using nCounter data as uniquely associated with metastasis (p = < 0.25).

The two lists of differentially expressed miRNAs (see Table **6**-**12** and Table **6**-**14**) were then cross-referenced to remove duplicates. This allowed miRNAs uniquely associated with metastasis to be identified. This highlighted 30 differentially expressed miRNAs unique to the nodal metastases (see Table **6**-**15**).

6.3.3.4 Validation of miRNASeq data miRNAs associated with metastasis

The list of 30 differentially expressed miRNAs uniquely associated with metastasis (see Table **6**-**14**) identified using the Nanostring nCounter miRNA assay was then compared to the 97 identified using miRNASeq (see Table **6**-**6**).

Using this method the nCounter data validated 12 miRNAs whose differential expression was uniquely associated with metastasis in these samples (see Table **6**-**16**).

It should be noted that 32/97 differentially expressed miRNAs associated with metastases identified using miRNASeq were not contained in the nCounter miRNA probeset, and therefore could not be validated using this technique. These represent miRNAs potentially associated with nodal metastasis.

The average sequencing read count across the patient samples for the differentially expressed miRNAs only detected by miRNASeq (n = 85) was calculated. The median of these was 274 (mean: 80,803, range: 6 – 2,683,169). The average sequencing read count across the patient samples for the miRNAs common to both analyses (n = 12) was then calculated. The median of these was 21,408 (mean: 50,733, range: 66 – 270,827). The large difference in median read count suggests that low abundance miRNAs may be less likely to be detected by the nCounter assay, though the fact that there are miRNAs with relatively high read counts not identified as differentially expressed by nCounter indicates that low abundance is not the only reason that some miRNAs were not validated.

	Log Fold change	p-value	Adjusted p-value
hsa-miR-206	7.47	2.25E-22	1.85E-19
hsa-miR-133a-3p	6.64	8.12E-14	3.34E-11
hsa-miR-99a-5p	2.78	2.56E-08	3.51E-06
hsa-miR-31-5p	-3.14	3.88E-05	0.0022
hsa-let-7c-3p	1.63	8.08E-05	0.0040
hsa-miR-146b-5p	-4.75	8.38E-05	0.0040
hsa-miR-125b-5p	1.81	0.000193726	0.0079
hsa-miR-1	3.13	0.000243149	0.0095
hsa-miR-23b-3p	1.80	0.000359254	0.012
hsa-miR-7-5p	-2.34	0.00051702	0.016
hsa-miR-133b	2.55	0.001873921	0.048
hsa-miR-130b-3p	-2.28	0.002028861	0.049

Table 6-16: miRNAs uniquely associated with nodal metastases validated by nCounter miRNA assay.

6.3.3.5 Validation of miRNASeq metastatic markers identifiable in metastatic primary tumours

The differentially expressed miRNAs uniquely associated with metastasis identified using the nCounter assay (see Table **6**-**15**) were then compared to the miRNAs identified as differentially expressed in the metastatic primary tumours. Duplicates were highlighted as potential miRNA markers of metastasis. These were then cross-referenced against the miRNA markers identified in metastatic primary tumours using miRNASeq data (see Table 6-7). Using this approach seven miRNAs were validated by the nCounter data as being miRNA markers of metastasis identifiable in the primary tumours (see Table **6**-**17**).

	Log Fold change	p-value	Adjusted p-value
hsa-miR-125b-5p	-2.13	1.22E-08	1.11E-06
hsa-miR-130b-3p	4.24	2.01E-08	1.65E-06
hsa-miR-99a-5p	-2.67	3.23E-08	2.04E-06
hsa-miR-146b-5p	3.82	7.63E-07	3.92E-05
hsa-miR-23b-3p	-2.04	9.02E-06	3.70E-04
hsa-miR-7-5p	2.71	0.000450381	0.013
hsa-miR-31-5p	3.09	0.001232828	0.029

Table 6-17: miRNA markers for metastasis identifiable in the primary tumour, validated by nCounter data.

6.3.3.6 Comparison of nCounter miRNA profiling of nodal metastases with and without ECS

In parallel with previous sections the nCounter generated miRNA profiles of nodal metastases with ECS (n = 3) were compared to matching normal epithelium. This identified 29 differentially expressed miRNAs. The nCounter generated miRNA profiles of nodal metastases without ECS (n = 3) was then compared to matching normal epithelium. These two lists were then cross-referenced. Duplicates were removed (leaving miRNAs only associated with metastases with ECS). This revealed 15 differentially expressed miRNAs. These were then cross-referenced against the nCounter generated differentially expressed miRNAs in non-metastatic primary tumours (see Table 6-14). Duplicates were removed (to remove miRNAs associated with carcinogenesis rather than ECS). This approach identified 11 differentially expressed miRNAs associated with ECS (see Table 6-18). On comparing this list to the differentially expressed miRNAs associated with ECS identified from miRNASeq data (see Table 6-10). In this way nCounter failed to validate any miRNAs associated with ECS found using miRNASeq data.

Of the miRNAs highlighted by miRNASeq as associated with ECS (n = 81) only 47 were present on the nCounter miRNA assay probeset. The failure to validate the sequencing data in this situation may be due to the fact not all miRNAs were tested for in the nCounter assay. The small sample size in this study may also be why no common pattern of expression can be seen.

	Log Fold change	p-value	Adjusted p-value
hsa-miR-1	4.80	6.56E-13	1.80E-10
hsa-miR-204-5p	4.98	1.18E-05	0.0013
hsa-miR-378g	4.68	4.54E-05	0.0046
hsa-miR-630	-3.09	0.000320469	0.023
hsa-miR-23b-3p	2.73	0.000722914	0.042
hsa-miR-381	3.98	0.000812441	0.044
hsa-miR-663b	-4.19	0.001002273	0.045
hsa-miR-518b	3.12	0.002070728	0.074
hsa-miR-299-5p	3.33	0.003122911	0.11
hsa-miR-30c-5p	2.94	0.007869778	0.23
hsa-miR-133b	3.12	0.008340578	0.23

Table 6-18: List of nCounter generated differentially expressed miRNAs associated with ECS (after comparison to non-metastatic primary tumours).

The median sequencing read count of miRNAs associated with ECS identified by the nCounter assay (n = 11) was 382 (mean: 10,804, range: 0–70,438). For those only identified by miRNASeq and present on the nCounter assay (n = 47) the median read count was 617 (mean: 16,683, range: 14–270,827). The fact that the median sequencing read count was higher in those miRNAs not highlighted by nCounter data suggests that low abundance of miRNA is not the sole reason that they were not identified by this assay.

6.3.3.7 Integrating CNA and miRNAs associated with metastasis

Identification of miRNAs whose pattern of expression correlates to copy number change could highlight miRNAs of additional use as markers for metastasis. Known genes targets of these miRNAs could also be of interest in future studies of the metastatic process or as additional markers. In order to discover any such miRNAs the log-fold change of differentially expressed

miRNAs was plotted against the genomic copy number of the genomic loci of each miRNA.

Mature miRNA	Precursor miRNA	Cytoband
hsa-miR-125b-5p	hsa-mir-125b-2	21q21.1
	hsa-mir-125b-1	11q24.1
hsa-miR-99a-5p	hsa-mir-99a	21q21.1
hsa-miR-133b	hsa-mir-133b	6p12.2
hsa-mir-206	hsa-mir-206	6p12.2
hsa-miR-23b-3p	hsa-miR-23b	9q22.32
hsa-miR-1	hsa-mir-1-1	20q13.33
	hsa-mir-1-2	18q11.2
hsa-miR-133a-5p	hsa-mir-133a-1	18q11.2
	hsa-mir-133a-2	20q13.33
hsa-let-7c-5p	hsa-let-7c	21q21.1
hsa-let-7c-3p	hsa-let-7c	21q21.1
hsa-miR-146b-5p	hsa-mir-146b	10q24.32
hsa-miR-7-5p	hsa-mir-7-1	9q21.32
	hsa-mir-7-2	15q26.1
	hsa-miR-7-3	19p13.3
hsa-miR-130b-3p	hsa-mir-130b	22q11.21
hsa-mir-31-5p	hsa-mir-31	9p21.3

Table 6-19: List of validated differentially expressed miRNAs associated with metastasis, with corresponding precursor miRNAs and genomic position.

Understanding the biogenesis of miRNAs was key to attempting this analysis. miRNAs are encoded as longer transcripts called primary miRNAs. These are subsequently processed to precursor-miRNAs (approximately 60-100 nt in length). These are stem-loop structures that are cleaved in the cytoplasm to eventually form the mature miRNA. As such, a mature miRNA does not have a corresponding genomic location but a precursor-miRNA does.

Mature miRNAs may be coded for by multiple precursor-miRNAs. This means that simply attempting to match the mature miRNAs to areas of genomic CNA is not possible; the precursor-miRNAs must be used to map

genomic locations, and then interpret which mature miRNAs are descended from them.

Firstly the list of validated differentially expressed miRNAs associated with metastasis (see Table **6**-**16**) was obtained. The potential precursor-miRNAs for each mature miRNA were identified as well as the corresponding genomic locations (see Table **6**-**19**).

The normalised sequencing counts of these precursor-miRNAs were combined with the segmented genomic copy number files for each nodal metastasis as input for a python script called miRNA_CNA (Dr Henry Wood, Precancer Genomics). This plotted the log fold change against the corresponding copy number ratio for each miRNA. This was performed comparing each nodal metastasis (n = 6) to its matching normal epithelium (see Figure **6**-**4**).

By inspecting each plot the miRNAs whose expression correlated to their corresponding genomic loci copy number were then discovered from these plots i.e. overexpression correlating with genomic loci copy number gain and underexpression correlating with genomic loci copy number loss.

As with the CNA analysis the individual threshold for calling CNAs for each sample needed to be accounted for. This was not possible on the expression vs. copy number ratio plots and required visual inspection of each sample's individual chromosomal karyogram, generated using CNAnorm (Gusnanto *et al.*, 2012). This allowed the location of each precursor-miRNA to be evaluated for copy number status (see Figure **6**-**5**).
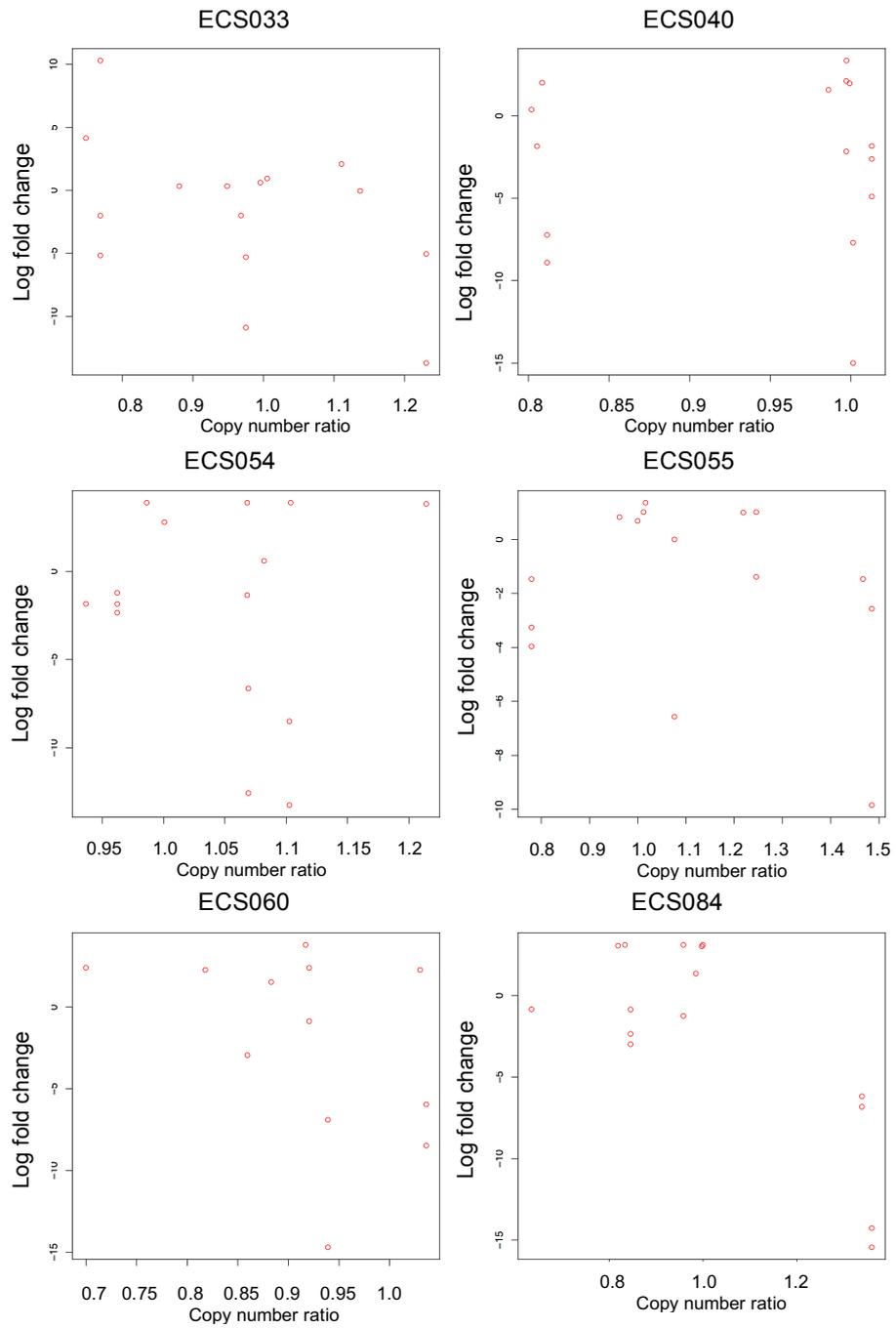
Figure 6-4: Plots of each precursor-miRNA against corresponding genomic copy number ratio.
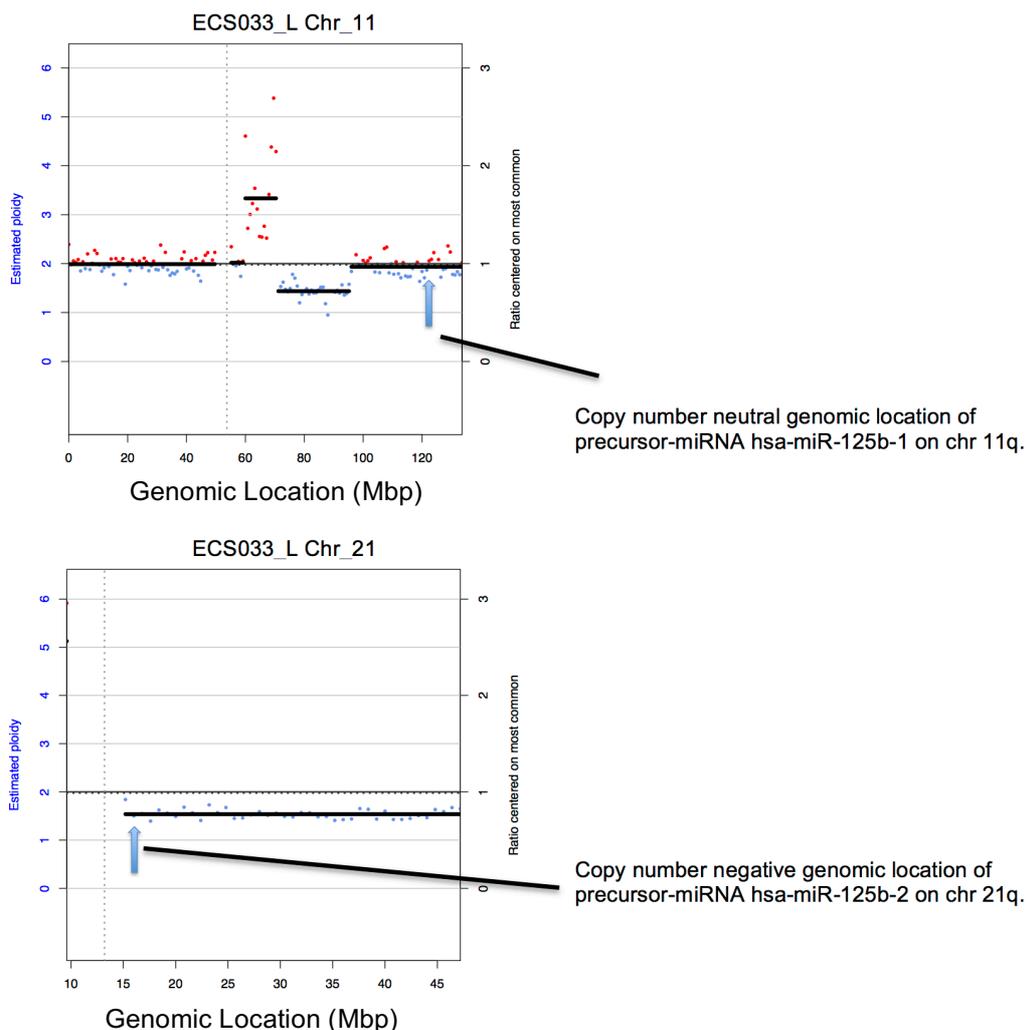
Figure 6-5: Example of using individual chromosomal karyograms to evaluate copy number status of each precursor-miRNA. This example shows how only one of the precursor-miRNAs for the underexpressed mature product (hsa-miR-125b-5p) has a correlating copy number status.

Copy number loss in association with underexpression of miRNAs was observed more frequently than copy number gain and miRNAs (see

Table **6-21** and Table **6-22**). All 6 nodal metastases contained between 1–4 correlating underexpressed mature-miRNAs, compared to only 3/6 containing 1–2 correlating overexpressed miRNAs. Overall, the correlation between expression levels and copy number status was low. For each patient the correlation between fold change and copy number ratio of potential precursor miRNA was calculated using Spearman's correlation

coefficient. This varied from -0.585 - 0.004 (see Table **6**-**20**). For two patients (ECS060 and ECS084) a statistically significant, strong negative correlation was indicated (p = 0.022 and 0.041). For the remaining four patients no statistically significant positive or negative correlation was observed.

Patient	Spearman's correlation coefficient	p-value
ECS060	-0.585	0.022
ECS084	-0.532	0.041
ECS033	-0.309	0.262
ECS040	-0.228	0.413
ECS055	-0.055	0.846
ECS054	0.004	0.99

Table 6-20: List of Spearman's correlation coefficient of precursor miRNA
expression and copy number ratio (significance level set at p = 0.05).

Amongst the mature-miRNAs whose underexpression correlates with genomic copy number loss, hsa-miR-125b-5p is consistently identified in each nodal metastasis. It is encoded by two pre-cursor miRNAs, hsa-miR-125-1 and hsa-miR-125-2). These are located at 21q21.1 and 11q24.1 respectively (see Figure **6**-**5**). At least one of these precursor-miRNAs was located in a deleted region and in one case (ECS084_L) both precursors were deleted. The expression of hsa-miR-99a-5p correlated to copy number loss of its precursor (hsa-miR-99a located at 21q21.1) in 3/6 nodal metastases.

It would be reasonable to hypothesise that the higher the number of precursors known to be associated with the mature miRNA the more likely this is to be to be associated with a corresponding CNA. However, hsa-miR-7-5p is known to be a potential product of three precursor miRNAs (hsa-miR-7-1, hsa-miR-7-2, hsa-miR-7-3) with three different genomic loci (9q21.32, 15q26.1 and 19p13.3). Therefore, potentially this had 18 genomic locations (3 in 6 samples) that could correlate to copy number status. This only correlated overexpression with copy number gain in 2 samples (in 3 precursor-miRNA loci in total).

ECS033		ECS040		ECS054		ECS055		ECS060		ECS084	
Mature	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**
hsa-miR-125b-5p	hsa-mir-125b-1	hsa-miR-125b-5p	hsa-mir-125b-1	hsa-miR-125b-5p	hsa-mir-125b-2	hsa-let-7c-5p	hsa-let-7c	hsa-miR-125b-5p	hsa-mir-125b-1	hsa-miR-125b-5p	hsa-mir-125b-1
		hsa-miR-1	hsa-mir-1-2			hsa-let-7c-3p	hsa-let-7c	hsa-miR-23b-3p	hsa-mir-23b		hsa-mir-125b-2
		hsa-miR-133a-5p	hsa-mir-133a-1			hsa-miR-125b-5p	hsa-mir-125b-2			hsa-let-7c-5p	hsa-let-7c
						hsa-miR-99a-5p	hsa-mir-99a			hsa-let-7c-3p	hsa-let-7c
										hsa-miR-99a-5p	hsa-mir-99a

Table 6-21: List of under expressed miRNAs whose genomic loci correlates to regions of copy number loss in nodal metastases.

ECS033		ECS040		ECS054		ECS055		ECS060		ECS084	
Mature	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Mature**	**Precursor**	**Precursor**	**Mature**
hsa-miR-130b-3p	hsa-mir-130b	Nil	Nil	hsa-miR-7-5p	hsa-mir-7-1	hsa-miR-7-5p	hsa-mir-7-1	Nil	Nil	Nil	Nil
					hsa-mir-7-3						
				hsa-mir-31-5p	hsa-mir-31						

Table 6-22: List of overexpressed miRNAs whose genomic loci correlates to regions of copy number gain in nodal metastases.

miRNAs whose expression correlate with genomic CNA, make up the minority of even the small number of validated miRNAs associated with metastasis. They represent potential markers for metastasis but also can provide insight into the metastatic process.

miRNAs whose dysregulation is independent of CNA are also of interest as a marker for metastasis, which can be used in addition to copy number. Amongst the nodal metastases, miRNAs whose expression is independent of CNA would include hsa-miR-206 and hsa-miR-146b-5p.

6.4 Discussion

6.4.1 Tissue sampling

The transcriptional profile of cells will depend on their specific tissue type. As the level of expression of any gene is effectively unknown until the RNA sample is profiled it is important to obtain as pure a source tissue sample as possible. In order to try and achieve the highest purity of tumour cells, haematoxylin and eosin slides of my samples were marked by a consultant head and neck histopathologist with the specific remit of marking the areas of highest tumour cell content (i.e. not less that 80%).

Using a laser has been shown to achieve high levels of target cell purity from heterogeneous cell populations (Emmert-Buck *et al.*, 1996, Espina *et al.*, 2006, Morrogh *et al.*, 2007). As this method was not available to me due to technical problems with the laser (see Chapter 4.3.1), tissue dissection was performed using the marked slides as a template for the target area for RNA extraction. A further 5 µm slide was cut immediately following the 7 x 10 µm slides cut for extraction. This section was stained with haematoxylin and eosin and marked by the histopathologist to ensure the target area of tumour had not altered. This quality control step was particularly important in smaller samples to ensure that the tumour had not been cut-through.

In terms of preserving tissue samples for prolonged periods, formalin fixation has long been the preferred method due to maintenance of morphology and the compatibility with innumerable immunohistochemical tests. For molecular analysis, the formalin fixation process and long term storage of these

samples represent causes of degradation of nucleic acids. RNA fragmentation occurs during the time between harvesting the sample and fixation as well as during the fixation process and may also occur if the samples are stored at high temperatures (Werner *et al.*, 2000, von Ahlfen *et al.*, 2007). miRNAs have been demonstrated to be well preserved in archived FFPE tissue samples, though there is some evidence that tissue samples subject to extended storages times (over 11 years) may suffer from degradation of miRNAs expressed at low levels (Szafranska *et al.*, 2008). As all tissue samples being used in my study were less than 10 years old and my aim was to study miRNAs FFPE samples were appropriate and obtainable. However, a study by von Ahlfen *et al* did find that formalin-fixation had a negative effect on PCR performed on RNA (von Ahlfen *et al.*, 2007). This was performed on mRNA rather than miRNA. The fact that mRNA from FFPE tissue is degraded may be the cause of this effect, as well as noting they did not assess miRNAs. As PCR is an integral part of creating small RNA libraries this could have lead to bias in the results.

RNA is also subject to degradation by RNases which are almost ubiquitous (Houseley and Tollervey, 2009). In order to try and combat this RNase decontamination was performed before, during and after any handling of RNA in the laboratory.

6.4.2 Nucleic acid manipulation

As detailed in the methods the extracted RNA undergoes manipulation in preparation for both sequencing input and the nCounter assay. During small RNA library preparation key steps include the ligation of 3' and 5' adaptors followed by reverse transcription, PCR amplification and size selection. There is no step for depletion of ribosomal RNA. This was attempted during optimisation of the library preparation protocol but was found to deplete the input RNA to the extent that no viable library was then produced. This may have been related to the fact that the library protocol required a minimum of 1 µg input RNA. This was not always achievable and therefore the input amount was titrated down to 500 ng successfully. When ribosomal depletion was performed with the reduced input amount of RNA no viable library was obtained. As over 80% of total cellular RNA is ribosomal this represents a

significant contaminant to the RNA sample (Lodish H, 2000). The adaptors are highly specific for the chemical structure of miRNAs but the fragmentation of larger RNAs may confound this specificity. Another potential source of error is introduced with PCR. This can occur due to editing errors during the DNA-polymerase-catalysed enzymatic copying and errors due to thermally-induced DNA damage during temperature changes associated with PCR cycling (Pienaar *et al.*, 2006).

 The nCounter miRNA assay preparation protocol did not include any PCR though annealing and hybridisation of tags and probes is performed at varying temperatures, which could introduce error from thermal DNA damage. The nCounter protocol also required less than 10% of the input amount of total RNA and no ribosomal depletion was performed. These could mean that the target miRNAs were in such low abundance as to potentially introduce error.

6.4.3 Profile of sequenced RNA

In terms of the differently types of RNA identified in each sample, miRNA represented the majority of the sequenced library. On average 60% (range: 51 – 71%) of aligned sequencing reads in each sample were miRNAs. This represented by far the largest proportion of aligned reads in any sample, confirming the effectiveness of both the RNA extraction and library preparation techniques. Other types of small RNA identified in the sample included long noncoding RNA (lincRNA), ribosomal RNA (rRNA), mitochondrial RNA, mitochondrial transfer RNA, small nucleolar RNA, and small nuclear RNA (see Figure **6-6**).

On average, over 25% of each sample's aligned reads were made up of non-miRNAs, with the largest proportion being linc-RNAs followed by rRNA. Attempts were made to optimise the Illumina Small RNA protocol by performing rRNA depletion prior to commencing the library protocol. This resulted in a non-viable concentration of library being produced and therefore had to be abandoned. This may have been related to the reduced input amount of total RNA (500 ng rather than 1 μg, as recommended by the manufacturer).
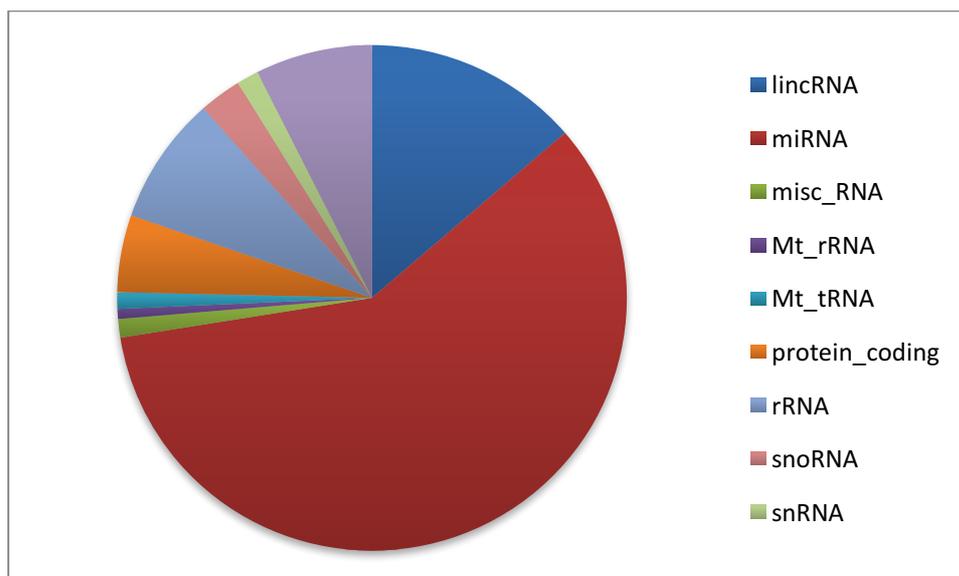
Figure 6-6: Average content of aligned reads in each sample processed for small
  RNA sequencing.

It is unlikely that the other types of RNA could be successfully removed without compromising the miRNA content as the RNA was from an FFPE tissue source and therefore degraded. Longer chains of RNA were therefore fragments to smaller chains and difficult to remove effectively on the basis of size alone. The adaptors in the Illumina protocol were highly specific for the 3'-hydroxyl group on miRNAs and Figure **6-6** demonstrates that they are effective at targeting miRNAs from a sample containing a mix of RNA types.

6.4.4 Intra-tumour heterogeneity

As discussed in Chapter 4, intra-tumour heterogeneity is an important consideration when attempting any molecular analysis. It has been demonstrated clearly in genomic studies of a number of solid cancers (Gerlinger *et al.*, 2014, Gerlinger *et al.*, 2012).

Wood *et al* used deep targeted sequencing and copy number sequencing to demonstrate tumour evolution through clonal populations in HNSCC. They demonstrated through multiple sampling of both a nodal metastasis and the primary tumour from the same patient that two distinct subclonal populations were identifiable in the metastatic primary tumour, one of which gave rise to the metastasis (Wood *et al.*, 2015). This supports the hypothesis behind profiling metastatic nodal tissue: tumours are made up of a heterogeneous

group of clonal populations of cancerous cells (Almendro *et al.*, 2013, Zhang *et al.*, 2013, Gotte *et al.*, 2005). The metastatic clone may not necessarily be the dominant clone within the primary tumour, but is more likely to be dominant within the metastatic tissue (Hong *et al.*, 2015).

The low number of differentially expressed miRNAs between primary tumour and matched metastasis could be due to clonal population sampling difference. However it may also be due to ongoing evolution occurring at the two sites. The molecular changes required for oncogenesis are not sufficient for metastasis to be successful, demonstrated by the reporting of murine models of cancer that do not automatically develop metastases (Minna *et al.*, 2003). However, the molecular aberrations at the primary tumour may allow escape of the metastatic cell, but further evolution may be required for the metastatic clone to thrive at its new site. Regardless, markers for metastasis are more likely to be identifiable when utilising metastatic tissue.

For this reason the nodal metastases were interrogated for differentially expressed miRNAs. To try and remove confounding miRNAs that are involved with carcinogenesis rather than metastasis the differentially expressed miRNAs associated with non-metastatic primary tumours were then removed from the analysis.

6.4.5 miRNA Nomenclature

One important aspect of the rapidly expanding and evolving research base in miRNAs is the corrections to nomenclature that have had to be made as our insight into their biogenesis and function has increased. Unchanged parameters include the first three letter (hsa) which indicate species (*Homo sapiens*) and the numbering of miRNAs, (e.g. hsa-miR-99) which is sequential. Specific precursors are indicated by suffix number (e.g. has-miR-99-1). Suffixes made of letters denote related mature sequences (e.g. hsa-miR-99a, hsa-miR-99b).

Suffixes denoting 3p or 5p indicate which arm of the hairpin-shaped precursor structure the mature sequence originated from (e.g. hsa-miR-99a-5p). Previously, nomenclature attempted to represent which of these arms represented the predominant product from the precursor miRNA by assigning the miRNA name without suffix and the opposing arm product

indicated with the suffix '*'. This was retired in miRBase version 17 for clarity and due to this being an area that has evolved markedly as knowledge about origin, sequence and predominance of miRNAs has changed and continues to change. With each update of miRBase miRNA names are changed and novel ones added. It was important this was acknowledged and traced to ensure accurate literature searches.

miRiadne is an online tool specifically for this purpose and enables tracking of changes in nomenclature to enable swift, reliable referencing (Bonnal *et al.*, 2015). Each miRNA was input into miRiadne to allow identification of previous names for each miRNAs. For example, has-miR-99a-5p was previously registered in miRBase as has-miR-99a until version 18.0 (2011). In this version, the suffix -5p was added and the use of '*' in nomenclature was replaced. This is particularly important to understand when surveying the literature.

6.4.6 miRNAs associated with metastasis

The potential of miRNAs as a biomarker for clinicopathological outcome is highlighted by a recent study by Hur *et al* that identified a four miRNA signature associated with liver metastasis in colorectal cancer (Hur *et al.*, 2015). In my study 12 miRNAs were validated by nCounter data as associated with metastasis (see Table **6**-**16**).

Of there, eight were underexpressed in the nodal metastasis, relative to normal epithelium. Interestingly, four of these were overexpressed in the matched metastatic primary tumours (see Figure 6-7).

It is recognised that a single miRNA may have a wide variety of effects, mediating both upregulation and downregulation of different target genes (Valinezhad Orang *et al.*, 2014). The changing levels of expression relative to the states of tumour and metastasis highlight the complexity of potential roles for miRNAs in regulating processes relating to both carcinogenesis and metastasis.
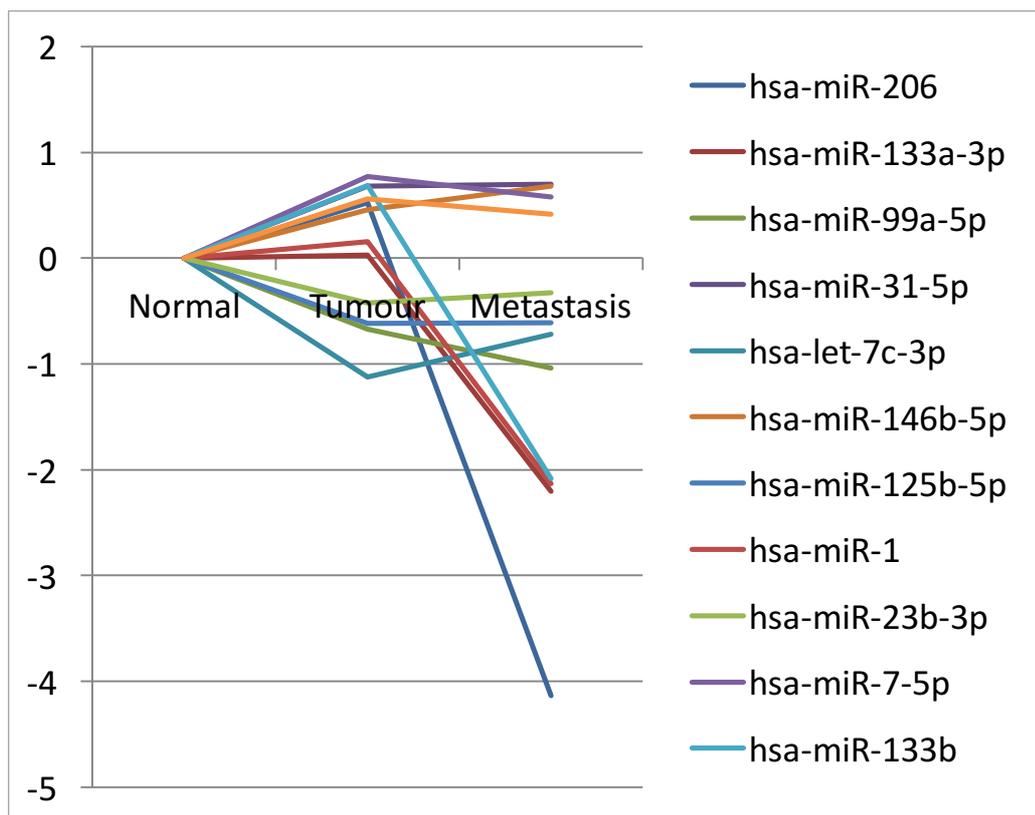
Figure 6-7: Plot of log fold-change of each validated miRNA associated with metastasis, from normal to tumour to metastasis.

The most highly significant (p = 1.85E-19) differentially expressed was has-miR-206. This was overexpressed in the metastatic primary tumours but underexpressed in the nodal metastases. Downregulation of miR-206 has been observed in a number of other cancers (Yan *et al.*, 2009, Song *et al.*, 2009, Vickers *et al.*, 2012). In gastric carcinoma cells loss of miR-206 was described, particularly in those with high metastatic potential (Zhang *et al.*, 2015). The same study also demonstrated that miR-206 levels were significantly decreased in metastatic lymph nodes compared to metastatic primary tumours. Transfection of a miR-206 inhibitor in gastric carcinoma cells with high metastatic potential was shown to increase cell migration and invasion (Zhang *et al.*, 2015). The transcription factor, *PAX3* was confirmed to be a direct target of mir-206, in gastric cancer cells. Overexpression of *PAX3* correlated positively with *MET* expression and predicted poor prognosis in patient samples.

The *MET* signalling pathway is essential for normal embryological development but has been implicated in carcinogenesis and epithelial-mesenchymal transition (EMT), a key step in tumour invasion and the metastatic process (Mazzone and Comoglio, 2006). It is a member of the receptor protein tyrosine kinase family. Overexpression of *MET* has been reported widely in HNSCC, as has a correlation with poor prognosis and nodal metastasis (Xu and Fisher, 2013). Over-activation of the MET pathway has also been implicated in anti-EGFR therapy (cetuximab) resistance (Lau and Chan, 2011).

Similarly, in hepatocellular carcinoma cells *NOTCH3* was revealed to be a target of miR-206. NOTCH3 is a member of the Notch family of transmembrane receptors, whose signalling is involved in regulating cell differentiation and senescence (Cui *et al.*, 2013). Downregulation of *NOTCH3* was observed in response to overexpression of miR-206 resulting in an increase in cell cycle arrest, apoptosis and cellular migration (Liu *et al.*, 2014).

Alteri *et al* demonstrated that *CCND1* was a major target of miR-206, and that miR-206-mediated repression of cyclin D1 is directly coupled to growth inhibition of cancer cell-lines (Alteri *et al.*, 2013). Cyclin D1 promotes progression through the G1 phase of the cell cycle and inhibits the retinoblastoma protein (Du *et al.*, 2013). RAS-mediated pathways induce transcription of cyclin D1 (Hitomi and Stacey, 1999). RAS proteins also play essential roles in signalling pathways that regulate cell proliferation. They are mutated in up to 20% of all cancers (Downward, 2003). As such cyclin D1 lies at the junction of a number of key signalling pathways in cancer. Specifically in HNSCC, cyclin D1 is overexpressed in up to 70% of HNSCC (Thomas *et al.*, 2005). It has also been associated with occult nodal metastases and poor prognosis in HNSCC (Capaccio *et al.*, 2000).

miR-206 has been observed to be downregulated in supraglottic carcinoma, though no relation to metastasis was made (Zhang *et al.*, 2014). This is the first report that it is underexpressed in relation to nodal metastases in HNSCC. The fact that its expression was slightly elevated in the matching primary tumours of these metastases may indicate a tumour suppressive

role prior to metastasis, and that this non-metastatic clone was sampled in the primary tumour. It may also be due to the differing primary tumour subsite evaluated. It may also be related to the fact that in Zhang *et al's* studies, metastatic and non-metastatic tumours were included (Zhang *et al.*, 2014). As such, miR-206 represents a valid potential target in HNSCC. However its role as a biomarker may be limited as it was not identified amongst the metastatic primary tumour as significantly differentially expressed. It may have a role as a serum biomarker, which is yet to be explored.

miR-125b-5p has also been identified as playing a tumour suppressor role in a number of cancers. In Ewing's sarcoma and cervical cancer, miR-125b-5p was found to be underexpressed, as well as targeting the PI3K/AKT/mTOR signalling pathway (Li *et al.*, 2014, Cui *et al.*, 2012). Li *et al* discovered the oncogene *PIK3CD* to be suppressed by overexpression of miR-125b-5p in Ewing sarcoma cells(Li *et al.*, 2014). *PIK3CD* is a key regulator of the PI3K/AKT/mTOR signalling pathway, which is an essential cell cycle mediator. PI3Ks are a family of enzymes that have been linked to regulating cell growth, differentiation, apoptosis, proliferation and migration (Engelman, 2009). Lui *et al* found this pathway contained the highest frequency of mutations in a series of 151 HNSCCs undergoing whole-exome sequencing, illustrating dysregulation of this pathway to be a common aberration in HNSCC (Lui *et al.*, 2013). However, in breast cancer cells, miR-125b overexpression was found to be associated with metastasis in cell lines and murine models. miR-125b-5p was shown to upregulate α-SMA and vimentin, which lead to increased metastatic potential and mesenchymal cell characteristics (consistent with EMT) (Tang *et al.*, 2012).

In HNSCC, decreased expression of miR-125b-5p has been reported in HNSCC studies of mixed subsite (Wong *et al.*, 2008b, Kikkawa *et al.*, 2010). In my study it was shown to be underexpressed in metastatic primary tumours and to be consistently underexpressed in matching nodal metastases. The fact that it was validated as identifiable in the primary tumours highlights the potential for this as a marker for metastasis.

miR-130b-3p was found to be over expressed in both metastatic primary tumours and matching nodal metastases. It has previously been reported to be upregulated in colorectal, renal cell and endometrial carcinoma (Wu *et al.*, 2012a, Colangelo *et al.*, 2013, Li *et al.*, 2013). In endometrial cancer, it was associated with EMT via effects on *DICER1* (Li *et al.*, 2013). Colangelo *et al* identified increased levels of miR-130b-3p in association with more advanced tumour stage and peroxisomal proliferator activated receptor gamma (PPARgamma) as a direct target (Colangelo *et al.*, 2013). This belongs to the family of nuclear hormone receptors and regulates lipid and glucose metabolism(Krishnan *et al.*, 2007). Via this target, deregulation of *PTEN*, E-cadherin and *VEGF* was demonstrated (Colangelo *et al.*, 2013). These are key mediators in invasion and metastasis. *PTEN* has been shown to play a tumour suppressor role in HNSCC and 30% of HNSCC exhibit decreased expression of this (Squarize *et al.*, 2013). This is unsurprising given it is a member of the PI3K/AKT/mTOR pathway.

The upregulation of miR-130b-3p observed in my study is consistent with previous studies profiling miRNAs in HNSCC (Cao *et al.*, 2013, Sethi *et al.*, 2014). It has also been reported as upregulated in plasma of patients with colorectal cancer undergoing chemotherapy and predictive of non-responders to treatment (Kjersem *et al.*, 2014). This draws attention to the potential for miRNAs as predictive markers for metastasis.

miR-99a-5p was underexpressed in both metastatic primary tumours and nodal metastases. The change was greater in the nodal metastases, though this may be due to mixed clonal sampling in the primary tumours, rather than reflective of an actual cellular change. Previous studies in HNSCC have found downregulation of miR-99a contributes to survival of OSCC cells and forced restoration resulted in suppression of cell proliferation and migration (Chen *et al.*, 2012, Yan *et al.*, 2012).

Yan *et al* demonstrated that *mTOR* was directly targeted by miR-99a-5p in OSCC cell lines (Yan *et al.*, 2012). Overexpression of miR-99a-5p was also shown to downregulated expression of *mTOR*. This role for miR-99a-5p was also been demonstrated in breast cancer cells, with some evidence that miR-99a-5p can reverse the breast cancer stem cell phenotype via the

PI3K/AKT/mTOR signalling pathway (Yang *et al.*, 2014). Similarly, in cervical cancer cells miR-99a-5p was shown to inhibit proliferation and invasion via the *mTOR* pathway (Wang *et al.*, 2014). This pathway interaction is consistent with the report that as part of a three microRNA signature miR-99a-5p predicted response to anti-EGFR therapy (given that EGFR can activate the PI3K pathway via phosphorylation of ERBB3) (Cappuzzo *et al.*, 2014).

hsa-miR-146b-5p was overexpressed in both metastatic primary tumours and nodal metastases at similar levels of fold change. This has previously been recognised to be overexpressed in oral and pharyngeal SCC tissue samples (Xiao *et al.*, 2012, Lajer *et al.*, 2012). This study is the first to report an association with metastasis in HNSCC.

In thyroid cancer cells, over-expression of hsa-miR-146b-5p was found to directly target the 3' untranslated region of *SMAD4*, inhibiting expression levels. Geraldo *et al* identified the TGF-β-signalling pathway as the route by which this occurred (Geraldo *et al.*, 2012). *SMAD4* originally identified as a tumour suppressor gene in pancreatic cancer has since been recognized to play a key role mediating both the WNT-signalling and TGF-β-signalling pathway (Hahn *et al.*, 1996, Nishita *et al.*, 2000). These pathways regulate processes included cell proliferation, survival, migration and polarity as well as immune regulation (Anastas and Moon, 2013, Massague, 2008). In HNSCC *SMAD4* deletion was found to lead to spontaneous HNSCC in murine models, with a high proportion (86%) of HNSCC tissue samples showing decreased expression of *SMAD4* (Bornstein *et al.*, 2009). *SMAD4* is located in 18q21.1-q21.32, a region found to be lost in 59% of nodal metastases and only 4% of non-metastatic primary tumours in which copy number sequencing was performed in Chapter 4. Integrating CNA and miRNA expression data like this provides further evidence for a significant role for *SMAD4* in metastasis and suggests potential for hsa-miR-146b-5p as a marker.

hsa-miR-23b-3p was observed to be underexpressed in both metastatic primary tumours and nodal metastases. miR-23b-3p has been found to have differing oncogenic and tumour suppressor effects in different studies.

Pellegrino *et al* demonstrated an inverse correlation between expression and breast cancer metastases and tumour growth *in vivo* (Pellegrino *et al.*, 2013). However, Jin *et al* found suppression of miR-23b-3p led to upregulation of tumour suppressor protein and reduced growth and metastasis *in vivo* (Jin *et al.*, 2013). In renal cell carcinoma and glioma miR-23b-3p has been shown to have an oncogenic role whilst in pancreatic and bladder cancer it has been shown to have tumour suppressor effects (Donadelli *et al.*, 2014).

In bladder cancer increased miR-23b-3p expression positively correlated with improved patient survival. *ZEB1*, a key oncogenic moderator of EMT, was found to be a direct target of miR-23b-3p (Majid *et al.*, 2013, Sayan *et al.*, 2009). Increased expression of *ZEB1* has been identified in HNSCC stem-cell-like cell lines and inhibition of this in murine models reduced tumour growth and metastasis (Chu *et al.*, 2013). *ZEB1* has also been shown to repress transcription of miR-200, which leads to immunosuppression of tumour-infiltrating lymphocytes (CD8(+) T-cells) and metastasis (Chen *et al.*, 2014).

In HNSCC, miR-23b-3p has been reported to be upregulated and downregulated in HNSCC tissue from a range of subsites (Scapoli *et al.*, 2010, Childs *et al.*, 2009, Ramdas *et al.*, 2009). These apparently conflicting roles may reflect the mixed tumour subsites included in previous studies. It may also be due to the fact that miR-23b-3p appears to mediate a number of pathways and given the intertumour heterogeneity of HNSCC, in different tumours different pathways are dysregulated as part of carcinogenesis. A larger, homogeneous cohort may reveal more about the role of this miRNA in HNSCC and potential use as a marker.

hsa-miR-7-5p was overexpressed in metastatic primary tumours and nodal metastases. This is consistent with previous studies that profiled HNSCC tissue samples, from oral, oropharyngeal and hypopharyngeal subsites (Kikkawa *et al.*, 2010, Rentoft *et al.*, 2011, Ramdas *et al.*, 2009, Maclellan *et al.*, 2012). This miRNA has been shown mediate the PI3K/AKT/mTOR pathway in lung, breast and prostate cancer (Webster *et al.*, 2009, Kefas *et al.*, 2008). It directly targets the 3' UTR of the *EGFR* mRNA, inhibiting

expression of this and other subsequent downstream molecules (Kalinowski *et al.*, 2014). This suggests a tumour suppressor function for has-miR-7-5p in these cancers, contrary to the fact several studies found it to be upregulated in HNSCC. A study using HNSCC cell lines found upregulation of miR-7 resulted in decreased expression of *EGFR,* and that exogenous miR-7 inhibited tumour growth (Kalinowski *et al.,* 2012). As such this may have limited use as a marker for metastasis given conflicting evidence for its function role and expression levels in HNSCC.

hsa-miR-31-5p was consistently upregulated in metastatic primary tumours and nodal metastases. This is another miRNA whose expression level appears to be cancer-type specific. It has been reported to be overexpressed in hepatocellular and colorectal carcinoma whilst underexpression has been observed in breast and ovarian cancer (Bandres *et al.*, 2006, Motoyama *et al.*, 2009, Wong *et al.*, 2008a, Yan *et al.*, 2008, Schaefer *et al.*, 2010). Numerous studies in HNSCC have reported overexpression of has-miR-31-5p in tissue samples and cell lines (Xiao *et al.*, 2012, Kozaki *et al.*, 2008, Liu *et al.*, 2010b). Upregulation of miR-31 was found to predict the presence of lymph node metastases in patients with lung adenocarcinoma (Meng *et al.,* 2013). Liu *et al* found miR-31-5p increased oncogenicity of HNSCC cell lines and increased tumour growth in murine models via the hypoxia-inducible factor (HIF) signalling pathway (Liu *et al.*, 2010b). In a separate study they found that plasma levels of hsa-miR-31-5p were elevated in OSCC patients and that these levels decreased markedly after tumour resection (Liu *et al.*, 2010a). This miRNA has not been reported to be related to metastasis in HNSCC previously, Liu *et al's* work demonstrates how markers identified in tissue could translate to less invasive tests.

6.4.7 Genes and gene pathways affected by miRNAs

Literature searches performed on each validated miRNA highlighted recurrently involved pathways. However, performing an analysis of pathways affected by miRNAs is challenging as a single miRNA can affect hundreds of genes and there is still a relative paucity of data on miRNA targets in specific diseases. To formally assess recurrent pathways, an online tool specifically

designed to provide access to up to date data called DIANA-miRPath version 2.0 was used (Vlachos *et al.*, 2012). This incorporated TarBase, the largest manually curated database of experimentally validated miRNA gene interactions (Vlachos *et al.*, 2015). The validated miRNAs associated with metastasis were each input into DIANA-miRPath and the resultant pathways evaluated.

The number of genes known to be targeted by each miRNA ranged from 18 – 1012 (see Table **6-23**). The large differential likely reflects the lack of research into individual miRNAs targeting, rather than a definitive functional difference. Using 15 cell signalling pathways known to be relevant to cancer, the number of miRNAs associated to each pathway by DIANA-miRPath was counted (see Figure **6-8**). The pathway with the highest frequency of miRNA hits was the MAPK signalling pathway (n = 12). This was closely followed by the PI3K/AKT, calcium and apoptosis signalling pathways. The distribution of hits amongst pathways contrasted with the similar analysis performed in Chapter 4, where there was a sharp drop off in hits. This is due to the fact that a single miRNA affects multiple pathways and appears to have a wider range of potential effects than a single gene. The pathways with a lower frequency of miRNA hits include NOTCH and Hedgehog signalling pathway (n = 3 and 1, respectively). This may simply reflect a lack of functional research in genes in these pathways or the relative number of genes known in each pathway.

miRNA	No of genes known to be targeted
hsa-miR-206	18
hsa-miR-133a-3p	51
hsa-miR-99a-5p	13
hsa-miR-31-5p	34
hsa-let-7c-3p	21
hsa-miR-146b-5p	9
hsa-miR-125b-5p	116
hsa-miR-1	1012
hsa-miR-23b-3p	30
hsa-miR-7-5p	417
hsa-miR-133b	19
hsa-miR-130b-3p	245

Table 6-23: List of genes known to be targeted by each miRNA associated with metastasis.
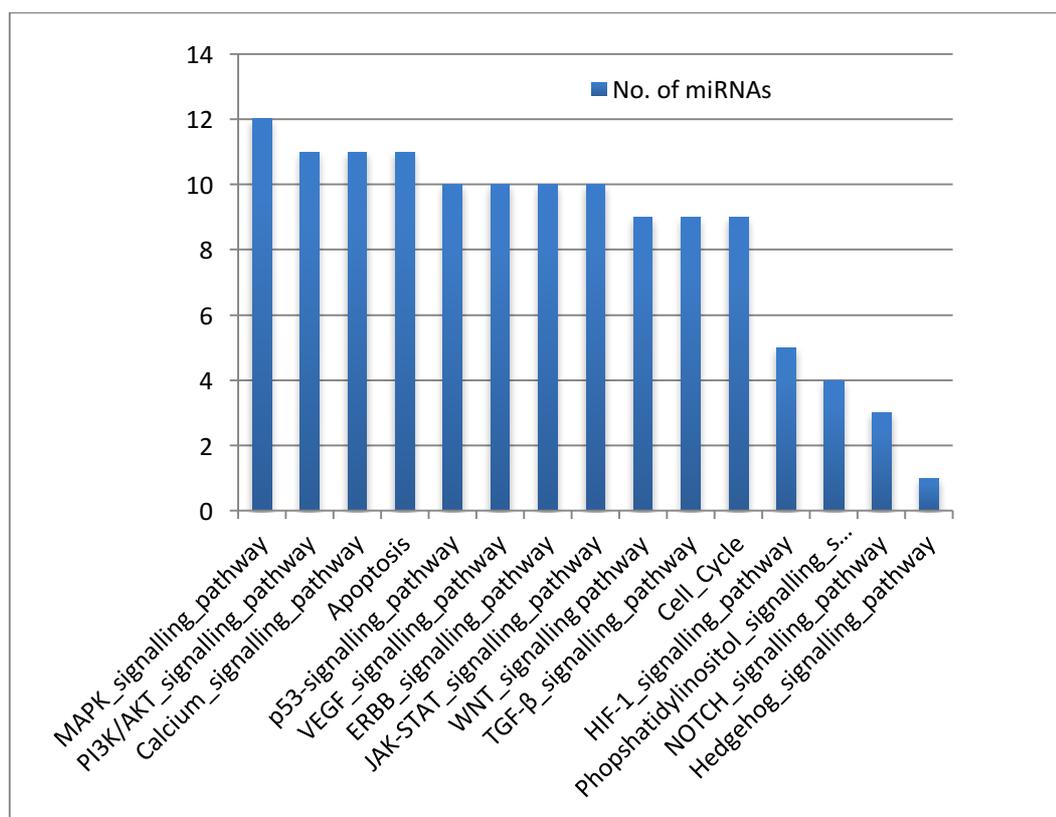
Figure 6-8: No. of validated miRNAs cross-referenced to each signalling pathway by DIANA-miRPath.

6.4.8 Integrating genomic CNA and miRNA expression data

On interrogating miRNA data for associated copy number changes it is essential to remember that a mature miRNA has no genomic location. It is derived from a precursor-miRNA, which does have a genomic location. However, a single mature miRNA may be derived from multiple precursor miRNAs. The precursor-miRNAs may be at different genomic locations. CAP-miRSeq utilises a software package called miRDeep2 (Berlin, Germany) as part of its algorithm to align sequencing reads to miRBase and also to calculate the number of reads for each precursor-miRNA (Friedlander *et al.*, 2008). This is calculated on the basis that each of the RNA products generated after the precursor-miRNA is cleaved by DICER (the 3' and 5' product) has a certain probability of being sequenced. miRDeep maps sequenced RNA reads to both the genomic position of the predicted precursor as well as the reads that correspond to the DICER cleavage products (that make up the precursor-miRNA) (Friedlander *et al.*, 2008). In this way CAP-miRSeq is able to assign reads to the most statistically likely precursor-miRNA. Where it is unable to decide it will divide read evenly amongst the multiple precursor-miRNAs allowing each to be assessed equally. This appeared to happen more commonly when the precursor-miRNAs had adjacent genomic locations, reducing the potential impact of uncertainty in assignation of precursor-miRNA.

In this way each precursor-miRNA associated with the 12 validated mature miRNAs linked to metastasis were plotted against the genomic copy number ratio of their genomic location. Two upregulated (hsa-miR-125b-5p, hsa-miR-99a-5p) and one downregulated (hsa-miR-7-5p) were recurrently associated with copy number gain and loss respectively. Only one was consistently associated with CNA in every sample (hsa-miR-135b-5p) suggesting that the expression of this miRNA is directly related to copy number status of its precursor miRNA.

The fact that miRDeep2 uses statistical probability to assign the reads to corresponding precursor-miRNAs means that they may potentially be prone to error or bias in the sequencing. Where miRDeep2 is unable to identify the specific precursor-miRNA it evenly divides the sequencing reads for the

mature miRNA between the precursors available. This is a useful failsafe position as it allows each precursor to be considered as potentially responsible for the expression of the mature miRNA, but makes it impossible to decide which is most likely. This did not occur in any of the precursors associated with metastasis. This indicates that sufficient reads of the mature miRNAs and products of cleavage of the precursor-miRNA were present to allow identification of the most likely candidate.

hsa-miR-130b-3p and hsa-miR-23b-5p only correlated their expression levels with copy number status of their precursor-miRNAs in a single sample each. As such these miRNAs could hold potential as a marker for metastasis in addition to the CNA marker identified in Chapter 4.

The effect of deleted or amplified miRNAs on the expression of their target genes can only be assessed by evaluating the gene expression profile of the same RNA samples. A previous study by Wu *et al* found that copy number altered miRNAs tend to target a higher number of gene than average (Wu *et al.*, 2012b). Given the fact that a single miRNA can both upregulate and downregulate different genes, and can influence the expression of many genes means that the potential targets and effects are great in number. Taking the additional step of adding mRNA expression data to this existing CNA and miRNA data could reveal a great deal about the mechanisms controlling the metastatic process.

My study suggests strongly that, in general, copy number status does not dictate expression levels of miRNAs in metastasis in head and neck cancer. A recent study in colorectal cancer tissue samples evaluated the relationship between CNA and miRNA expression levels (Gasparini, 2013). This similarly found only 2 miRNAs had altered copy number status. The relationship between copy number status and miRNA expression may also be consistent in health and disease as another recent study evaluated 5 haemopoetic cell lines using microarrays and found little correlation between copy number and miRNA expression (Veigaard and Kjeldsen, 2014).

Though hsa-miR-125b-5p consistently correlated expression to copy number status it is important to remember the inter-tumour heterogeneity of HNSCC. This was clearly shown by the CNA profiling performed in chapter 4 and 5. A

larger number of tumours, with multiple sampling performed per tumour, would need to be interrogated to confirm the consistent nature of this relationship. It was not possible to assess the available TCGA data for the expression levels of the miRNAs identified in my study as they only currently provide miRNA read count data for tumours and not for the matched normal, prohibiting a like-for-like analysis.

6.5 Conclusion

With regards to the specific aims of this chapter:

1. NGS and nCounter generated data was used to compare the miRNA profiles of nodal metastases to non-metastatic primary tumours and identified 12 miRNAs whose differential expression was associated with metastasis.

2. The matching metastatic primary tumours were then used to identify those miRNAs that may have use as predicting metastasis from primary tumour tissue. A panel of seven miRNAs were identified as potential markers for metastasis. Integrating genomic CNA data and miRNA expression data revealed that the expression of very few miRNAs is controlled by copy number status alone.

Chapter 7

Discussion

7.1 Introduction

Metastasis is a potentially catastrophic event for patients with cancer. The majority of deaths from cancer are due to metastatic disease. Models of metastasis vary from the linear progression model, wherein a stepwise accumulation of genomic abnormalities leads to the metastatic clone, to the parallel progression model, where tumour cells leave the primary tumour before the acquisition of the definitive malignant phenotype and undergo further evolution at the site of metastasis (Klein, 2009). Incontrovertible proof does not exist for any model. Regardless of which is preferred, all support the hypothesis that the primary tumour consists of metastatic and non-metastatic clonal populations.

Intra-tumour heterogeneity has been demonstrated in a wide range of solid cancer including HNSCC (McGranahan and Swanton, 2015, Wood *et al.*, 2015). Relatively few studies have performed comparative analysis of primary tumours with matched metastases. Previous studies have shown metastasis may arise from a non-dominant clone within the primary tumour and that the metastatic subclone becomes dominant in the metastasis (Gronwald *et al.*, 1999, Talmadge, 2007, Waghorne *et al.*, 1988).

7.2 Comparison of copy number profile of metastatic primary tumours to metastases

Very few studies have compared genomic profiles of metastatic tissue to matched primary tumour as well as a cohort non-metastatic tumours. Yoshioka *et al* found 12/15 tumour-metastasis pairs of OSCC were identifiable using unsupervised hierarchical clustering of genomic breakpoints, concurring with our findings (Yoshioka *et al.*, 2013). Bockmuhl *et al* utilised concordance and "similarity scores" based on chromosomal changes to compare metastatic primary tumours to nodal metastases. This

allowed identification of up to 24/34 paired samples, though the cohort consisted of mostly oropharyngeal and hypopharyngeal tumours (24/34) (Bockmuhl *et al.*, 2002).

My study demonstrated that the CNA profiles of nodal metastases and their matching primary tumour are not identical. In 37/49 cases the paired metastasis had the closest concordance of CNA profile to its primary tumour. Similar levels of correlation were found between nodal metastases and matching primary tumours (in 40/49 cases was the metastases the closest correlate to its paired tumour). The differences between primary tumour and metastasis support the notion that in order to identify molecular changes associated with metastasis, it is essential to use metastatic tissue.

It would have been revealing to use multiple samples of both the primary tumours and the nodal metastases, as well as including samples from all nodal metastases where there was more than one present. This could have allowed greater insights into the effect of intra-tumour heterogeneity. A more recent study in non-small cell lung cancer found wide variation in somatic mutations between primary tumours and metastases. Vignot *et al* found 94% concordance for recurrent, presumed driver mutations compared to 63% for likely passenger mutations (Vignot *et al.*, 2013). Using whole exome sequencing or transcriptome sequencing could be revealing in future work in HNSCC but presents considerably higher costs. In order to gain a better insight into tumour evolution the concept of multi-region sampling could be taken down to a single cell level. This could be combined with the relatively lower cost approach of low-coverage sequencing to provide CNA data to show the evolution of metastatic HNSCC.

7.3   Identification of CNA associated with metastasis

Comparison of the CNA profile of nodal metastases to non-metastatic primary tumours demonstrated a number of differences, though no single CNA was exclusively present in all samples of one group.   The CNA associated with metastasis identified in this and previous HNSCC studies are listed in Table **7**-**1**. Some of the most striking CNAs associated with metastasis in my study included gain of 11q13.3-q13.2 (present in 45% of metastases and 8% of non-metastatic primary tumours), loss of 3p14.3-

p14.1 (present in 71% of metastases and 39% of non-metastatic primary tumours) and loss of 11q23.1-q25 (present in 45% of metastases and 0% of non-metastatic primary tumours). The high frequency of alterations on 11q in nodal metastases highlighted genes that interact with *TP53* (*TP53AIP1* and *EI24*) which are little explored in HNSCC.

The low level of concordance between the studies shown in Table **7-1** is of concern when attempting to identify a copy number signature for metastasis. HNSCC represents a genomically heterogeneous group of tumours. Though my study had the highest number of samples (49 tumour-metastasis pairs and 26 non-metastatic primary tumours) the low number of samples across the studies means any purported marker may be limited in its application to a larger cohort.

Another limitation to my study is the resolution of sequencing. Though cost-effective the low-coverage sequencing approach taken achieved a genomic resolution of 800 kb. This means any CNA smaller than this would not be discernible in my data. There is no universal definition of what a focal copy number alteration is. GISTIC 2.0 used the cut off of < 98% of the chromosomal arm as the upper limit for focal CNAs, with alterations smaller than this included in the analysis. Numerous studies in different solid cancers have utilised the definition of < 3 Mb in length as the cut off for focal CNA (Brosens *et al.*, 2010, Jones *et al.*, 2008, Leary *et al.*, 2008, Parsons *et al.*, 2008, Weir *et al.*, 2007). Genes vary in their length from 0.2 kb to 2.5 Mb (Strachan T, 1999). In my study CNAs spanning a single genomic window (800 kb) were rare, though there is no way to tell if smaller CNAs were being missed.

After identifying a 14-CNA marker for metastasis in Chapter 4 (see Table 4-13), this was then tested against copy number data downloaded from TCGA HNSCC cohort. Though the difference in scores for this marker was significantly different (p = < 0.001) between metastatic and non-metastatic primary tumours it was not able to discriminate completely between metastatic and non-metastatic tumours. This likely reflects the high level of inter-tumour heterogeneity in HNSCC. This limits the applicability of CNA alone as a reliable marker for metastasis. The TCGA data was also

processed using a generic CNA threshold as opposed to my data, which was analysed to assign each sample an individual threshold. This may have affected the accuracy of the CNA calls. However, generic thresholds are used widely in CNA analysis including in software packages such as GISTIC 2.0 and as such any translational marker should arguably withstand this.

Sethi		(Yoshioka *et al.*, 2013)		(Patmore *et al.*, 2004)		(Bockmuhl *et al.*, 2002)		(Kujawski *et al.*, 1999)	
Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss
4p15.2-p15.1	3p14.3-p14.1	7p	-	6q	-	1q21-q22	5q34-q35		8p
5p14.1-p15.1	4p15.1-14	8q		22q		3q24-qter	8p12-p22		9q
7p12.1-11.2	3p26.3-p26.11	17q				6q	10p12		13
7q21.3-q22.3	5q33.1-q33.3					7q11.2	10q21-qter		
8q11.21	7q31.1-q35					12q12-q12	11p14-p15		
9p21.1	8p23.3-p22					18p11.1	11q23-qter		
11q13.3-p13.2	2q34-q37.1						14q21-qter		
13q33.1-q34	9p24.3-p24.2								
17q22-q23.3	11q23.1-25								
	15q26.2-q26.3								
	18q21.1-q21.32								

Table 7-1: Comparison of CNAs associated with nodal metastases in this study and previous studies that compared metastatic tissue to primary tumours.

Future work should focus on the characteristics of the genes identified from this data, including immunohistochemistry analysis as well as functional analysis of the upregulation and downregulation of these genes *in vitro* and/or *in vivo*.

7.4 Evaluation of CNA profile of metastasis with and without ECS

Though the presence of lymph node metastasis does have a detrimental impact on patient outcomes in HNSCC, the presence of ECS remains the most significant negative prognosticator, reducing overall survival by up to 70% (Shaw *et al.*, 2010, Jose *et al.*, 2003). No single CNA was found exclusively in metastases with or without ECS. Loss of 18q21.1-q21.32 was found in 30/49 (59%) nodal metastases and 1/26 (4%) non-metastatic primary tumours. It further identified in 21/27 (78%) metastases with ECS compared to 9/22 (41%) metastases without ECS. Three genes of potential interest were identified in this region (*SMAD4*, *MALT1*, *PMAIP1*) . *SMAD4* has been highlighted as performing a tumour suppressor function in pancreatic cancer originally (Hahn *et al.*, 1996, Schutte *et al.*, 1996). Since then, loss of *SMAD4* has been observed to lead to spontaneous head and neck tumour development in murine models (Bornstein *et al.*, 2009). Decreased expression of SMAD4 was reported in HNSCC tissue samples as well as the morphologically normal adjacent mucosa (Korc, 2009). The fact that it is underexpressed in adjacent mucosa suggests that loss of function of this gene is an early event in HNSCC, as this tissue represents part of the "cancerised field" in patients with HNSCC (Slaughter *et al.*, 1953, Braakhuis *et al.*, 2003). Previous work has also demonstrated that *TP53* mutations are identifiable in morphologically normal mucosa in HNSCC (Wood *et al.*, 2015, Escher *et al.*, 2009). Being an early event does not preclude it from being an important event in metastasis, particularly in the setting of the parallel progression theory of metastasis. Examining for loss of *SMAD4* in precancer could reveal it to be an early marker for risk of malignant transformation and metastasis.

Relatively few studies have examined a molecular marker for ECS in HNSCC. Increased copy number of *EGFR* and *CCND1* was observed in OSCC tissue using FISH (Michikawa *et al.*, 2011). Overexpression of *SERPINE1* and *SMA* has been found to be predictive of ECS with a high sensitivity (81%) but lower specificity (50%) (Dhanda *et al.*, 2014). Wang *et al* identified a an 11-gene expression signature associated with ECS (Wang *et al.*, 2015). Similarly to genomic copy number studies searching for association with metastasis there is clearly a lack of consistency in findings

between studies. This may be affected by the number of patients included in each study and therefore affected by the inter-tumour heterogeneity. As all these studies were based upon a single tissue sample, intra-tumour heterogeneity may be affecting the ability to identify a consistent marker associated with ECS.

SMAD4 may represent a driver gene in not just HNSCC but specifically with the development of ECS. The low number of potential driver genes identified is also reflected in other cancers such as lung and colorectal carcinoma. Tomasetti et al found only 3 sequential mutations were required for the development of cancer in these two types (Tomasetti *et al.*, 2015).

Future work attempting to elucidate a marker for ECS may required a more integrated approach utilising mutation, copy number and transcriptomic profiling. However, it may also reflect that the underlying molecular processes resulting in the ECS phenotype are too varied to identify a simple marker for all patient with ECS. Therefore a larger number of samples that allow molecular sub-groups to be identified within metastases with ECS may be of value. However, it should be noted that with precise histopathological categorisation significant subgroups of cancer samples can be selected and genomic alterations tracked to these sub groups. Shain *et al* used targeted exome sequencing within a group of 37 primary melanomas and their adjacent precursor lesion to identify driver mutations that track the evolution of melanoma. This was only possible using precise histological classification of each tissue-type sampled along with detailed clinical characterisation of each patient (Shain *et al.*, 2015).

7.5 Fraction of genome altered (FGA) in metastasis

CNAs affect a larger proportion of the genome than any other type of genetic alteration (Zack *et al.*, 2013). In interpreting and visualising the genome and CNAs it is possible to evaluate the actual length of the genome that is affected by CNA. As such, attempts have been made to assess any relationship between the fraction of genome altered (FGA) and clinicopathologic outcomes in patients with cancer. In prostate cancer, this was prognostic for prostate cancer relapse and metastasis (Hieronymus *et al.*, 2014). Genomic evaluation of 359 breast cancer samples identified

different values of FGA to be associated with differing histological subtypes of breast cancer (e.g. triple negative vs. estrogen receptor positive vs. estrogen receptor negative) (Jonsson *et al.*, 2010).

In my study, the mean FGA of metastatic primary tumours was found to be significantly higher than non-metastatic primary tumours (p = 0.0001), though no difference was found between tumours associated with and without ECS. Further analysis identified the cut off for this to be an FGA of 0.04 that identified metastatic primary tumours with a sensitivity of 98% and specificity of 42.3%.

No previous studies have examined this aspect of CNA burden in HNSCC. FGA as a marker for metastasis is appealing as it removes subjective interpretation of the copy number profile of a sample and provides a single figure as a result. It also is less likely to be affected by the genomic resolution of sequencing or array as the effect of small alterations in the exact breakpoint will not substantially change the overall proportion of genome that is copy number altered. Therefore extremely small CNAs that are obscured by broader CNA when performed a group-wide analysis are less likely to impact on this result. However, the potential for false positives is considerable with a specificity of 42.3%. It may also be affected by clonal sampling and therefore future work could evaluate this.

An FGA > 0.098 and > 0.125 was also associated with a significant decrease in disease-free survival and overall survival respectively (p = 0.014 and p = 0.0059 respectively). This could hold potential for stratifying patients into those that could benefit from multimodal therapy. Neoadjuvant chemotherapy has been found to be of limited overall benefit in HNSCC (Garden, 2014). FGA could identify patients, at the point of primary tumour biopsy, that are at increased risk of recurrence and decreased survival. These patients could then be stratified into clinical trials to determine those that may benefit from more aggressive treatment regimes.

7.6   Evaluation of CNA profile of OPSCC and viral load

The association between OPSCC and HPV is well established though incompletely understood (Gillison *et al.*, 2008, Vidal and Gillison, 2008, Gillison and Lowy, 2004, Gillison, 2004). Previous studies using exome

sequencing have revealed a similar mutational burden of both HPV-positive and negative HNSCC tumours, with characteristic differences in the spectrum of mutations. The most striking of these being a low incidence of *TP53* mutations in HPV-positive tumours, whilst it is almost universally present in HPV-negative tumours (Stransky *et al.*, 2011, Agrawal *et al.*, 2011, Seiwert *et al.*, 2015). Both have been suggested to be genomically heterogeneous. NGS has been shown to be an effective method of detecting HPV-DNA and determining viral load (Conway *et al.*, 2012, Lechner *et al.*, 2013).

My study identified a panel of 19 CNAs associated with 0 viral load (see Table 5-6). When scoring each tumour for the presence/absence of these CNAs, the mean score of tumours with 0 viral load was found to be significantly higher than tumours with a viral load >0 (p = <0.0001). No difference in scores was found between tumours with intermediate and high viral loads, suggesting that viral load alone does not dictate the nature of the genomic alterations associated with OPSCC.

Seiwert *et al* analysed 50 HPV-positive and 70 HPV-negative mixed subsite HNSCC samples, revealing considerable inter-tumour genomic heterogeneity (Seiwert *et al.*, 2015). Copy number altered genes associated with HPV-negative tumours included *EGFR, FGFR1, CCND1* and *CDKN2A*. My study concurred with these findings. Loss of *SETD2, CSF1R, NOTCH1* and *CHD5* were also observed with greater frequency in tumours with 0 viral load in my samples, though Seiwert *et al* identified these at similar rates in HPV-positive and negative tumours (Seiwert *et al.*, 2015). The differences between studies could be related to the fact Seiwert *et al's* tumours were of mixed subsites (20/70 HPV-negative and 47/51 HPV positive being oropharyngeal) or they could simply reflect the inter-tumour heterogeneity of these tumours (Seiwert *et al.*, 2015).

The viral load was calculated using the number of reads obtained per sample and the number of reads that aligned to the viral genome. A potential confounding factor to this could be the coverage of the sequencing. As this provided a maximum resolution of 800 kb per sample it is possible that viral genome reads were not identified in some samples and potential labelled as

0 viral load when they actually had a low level of viral DNA present. Another, criticism of this study could be the lack of another method of determining HPV status of each sample in addition to NGS (e.g. p16 immunohistochemistry). This could have provided an alternative method of grouping copy number profiles and evaluating differences. However previous work by Conway *et al* demonstrated that low coverage sequencing had 100% sensitivity compared to DNA PCR or p16 IHC (Conway *et al.*, 2012). This indicates that detection of viral DNA using NGS correlated highly with a positive PCR or IHC test. This suggests that the depth of sequencing was adequate to detect low levels of viral DNA. In order to glean more translational information from the impact of viral load, clinical outcome data including survival and recurrence would be essential for any future work, ideally of a greater number of patients to add power to any sub-groups of viral load.

7.7 Fraction of genome altered (FGA) in OPSCC according to viral load

The FGA was again evaluated as a reflective marker of viral load. This was found to be significantly lower in tumours with viral load >0 (p = 0.006). This is an interesting genomic marker as it requires less visual assessment of each individual karyogram compared to using a panel of CNAs. It also provides a bottom line figure which is attractive as a translational marker. However it is not a definitive binary divider between these two tumour groups, similarly to the CNA panel for metastatic and non-metastatic primary tumours. If this was related to clinical outcomes such as survival or recurrence it may be of use in differentiating HPV-positive tumours that are associated with a beneficial prognosis and treatment response. Future work could specifically evaluate this.

The importance of FGA has not been reported previously in OPSCC but Stransky *et al* did report they found the mutational burden of HPV-negative HNSCC to be higher than HPV-positive (Stransky *et al.*, 2011). Mroz and Rocco developed a measure of intra-tumour heterogeneity based on whole exome sequencing of tumour and matched normal DNA called Mutant Allele Tumour Heterogeneity (MATH) (Mroz and Rocco, 2013). This is based upon

the mutant allele fraction of tumour-specific mutations and assigns a score for heterogeneity. They found that higher intra-tumour heterogeneity (measured by MATH) was related to worse outcome in a series of 74 HNSCC patients (Mroz *et al.*, 2013). This series of patients was primarily HPV-negative (63/74). They went on to examine 305 TCGA HNSCC patients (36 of which were HPV-positive). They found MATH scores were significantly lower in HPV-positive tumours (p = 0.004) (Mroz *et al.*, 2015). The effect of intra-tumour heterogeneity must be considered in future work evaluating the validity of FGA. Subclonal CNAs containing in heterogeneous tumour samples may not reach the CNA threshold set for the sample. As such the FGA for the sample would not necessarily be increased. This is a potential flaw in applying FGA. Multi-region sampling of each tumour may assist with identifying this and accounting for heterogeneity. Mroz *et al's* findings suggest that in HPV-positive tumours, heterogeneity would potentially have less of an effect than in HPV-negative tumours (Mroz *et al.*, 2015). Calculating an average of multiple samples may increase the ability of FGA to separate tumours by HPV-status or by patient outcome.

7.8 Identification of miRNAs associated with metastasis

My study utilised RNA from six nodal metastases and three non-metastatic primary tumours along with matching normal epithelium from all nine patients to identify miRNAs associated with metastasis. To increase the likelihood of these not just being associated with carcinogenesis the miRNA profile of the metastatic primary tumours was used to discover the differentially expressed miRNAs only associated with nodal metastases. These findings were initially discovered using NGS and validated using the nCounter miRNA assay.

Twelve miRNAs were found to be associated with metastasis (see Table 6-17). Of these eight were also identifiable in the metastatic primary tumour (see Table 6-18). Though the full function of these miRNAs is still poorly understood, collectively they represent a great potential as metastatic markers in tumour tissue and have been demonstrated as detectable in the serum of other cancers, such as hsa-miR-130b-3p in the plasma of patients with colorectal cancer (Kjersem *et al.*, 2014).

Though many studies have attempted to comprehensively profile miRNAs in HNSCC, there is a high level of heterogeneity in the miRNAs identified (Sethi *et al.*, 2014). Part of this is likely due to the fact that there was a steep increase in the number of miRNAs registered in miRBase from its inception (<100) to today (1881) (Griffiths-Jones, 2004). The fact that every year novel miRNAs means that array-based methods of detecting miRNAs are limited to searching only for the probes present. miRNASeq offer the advantage of simply sequencing all small RNA sequences present in the sample. These can then be re-aligned to each updated version of miRBase as it is produced, essentially "future-proofing" the data against the discovery of novel miRNAs.

Few studies have actually utilised tissue from nodal metastases in profiling miRNAs in HNSCC. As demonstrated with the copy number analysis performed in Chapter 4, it is necessary to use this tissue to reliably identify any potential markers or targets that are specifically associated with metastasis. Fletcher *et al* found hsa-miR-205 to be overexpressed in 12 metastatic HNSCC primary tumours and their matching nodal metastases, using qRT-PCR (Fletcher *et al.*, 2008). This miRNA has also been found to be underexpressed in HNSCC in a separate study (Fletcher *et al.*, 2008). It was not identified as significantly differentially expressed in my study using both miRNASeq and nCounter approaches. This could reflect inter-tumour heterogeneity or the complexity of miRNA function and the epigenetic influences on their behaviour, which are poorly understood.

The Nanostring nCounter expression profiling system offered the ability to profile and therefore attempt to validate a large number of miRNAs (800) rather than select a smaller number to attempt to validate with another method such as qRT-PCR. It also avoided any PCR cycles as part of the preparation of the RNA, which is another source of potential error. Recent studies have shown it to have a high correlation to miRNASeq and microarray generated data (Kolbert *et al.*, 2013, Tam *et al.*, 2014). Analysis of the sequencing counts of the miRNAs there were and were not validated by nCounter data found no significant difference between them suggesting that sensitivity to low abundance miRNAs was not the reason they were not validated by the nCounter data. The fact the small RNA library preparation

involves 12 cycles of PCR amplification may contribute to miRNAs detected by sequencing but not by nCounter data.

The most important limitation to the work was the limited number of patients included. Given the inter-tumour heterogeneity demonstrated by the copy number analysis performed in Chapter 4 and 5 it is likely that the miRNA profiles of HNSCC will be similarly heterogeneous. It should be noted that 32/97 miRNAs identified using miRNASeq data were not present on the nCounter miRNA probeset, meaning they could not be validated using this technique and therefore represent miRNAs still potentially associated with metastasis. These could potentially be investigated in future work using custom-made probeset or alternative techniques.

7.9   Integration of miRNA and CNA data

In the final analysis I attempted to correlate the expression of miRNAs associated with metastasis to copy number changes. A low level of correlation was found overall. One miRNA (hsa-miR-125b-5p) was consistently identified to be underexpressed in each nodal metastasis. At least one of its precursor-miRNAs was also located at a site of genomic loss in each metastasis. Though other miRNAs were identified in specific samples to correlate expression to copy number, no other miRNA was found share this correlation in every metastasis.

Given the association of miRNAs to fragile genomic locations in other cancers this low level of correlation between expression and CNA strongly suggests that genomic copy number does not dictate expression of miRNAs in metastasis in HNSCC (Calin *et al.*, 2004, Zhang *et al.*, 2006). This concurs with a study examining the expression of miRNAs in colorectal cancer (N. Gasparini, 2013). However miRNAs that are not associated with CNAs hold potential for use as part of a marker that integrates different molecular components (e.g. genomic CNA and specific miRNAs). Future work including more samples could explore this. Gross *et al* found loss of 3p correlated to poorer patient survival in HNSCC. This effect was synergistic with *TP53* mutation and hsa-miR-548k expression (Gross *et al.*, 2014). If metastatic nodal tissue was evaluated in a similar way an integrated marker for metastasis may be identified.

Again this work is limited by the fact only small numbers of patients were included. Combined with the inter-tumour heterogeneity, any marker needs to be tested in a larger cohort of samples. Though the tumour and metastatic RNA was extracted from the same tissue blocks as the DNA was obtained from for copy number analysis, it was not obtained from the same tissue sample. This could introduce variation due to intra-tumour heterogeneity and affect the correlation identifiable from comparing these data. This could be accounted for in future work by utilising kits that allow dual-extraction of DNA and RNA from a single tissue sample (e.g. the QIAgen AllPrep DNA/RNA/miRNA Universal kit).

7.10 Reflections

Though a great deal was learnt during the course of this research, on reflection there are several aspects that could have been enhanced with the benefit of hindsight. Ideally the patients would have been recruited prospectively with detailed clinical information being collected contemporaneously. This is particular importance with respect to smoking or tobacco use. Accurate clinical data can only be reliably collected prospectively using a predetermined set of parameters. Detailed information can allow sub-groups to be evaluated, even within relatively small groups of heterogeneous tumours

Attempting to integrate the host response to cancer by obtaining data such as the presence of tumour infiltrating lymphocytes and using additional samples including blood and saliva would be of great interest. Correlating the host response to genomic characteristics could reveal novel markers. Screening blood and saliva for both genomic (e.g. circulating tumour DNA) or immunological (e.g. immune cells) markers would enable non-invasive biomarkers to be recognised.

Blood may also represent a valuable alternative to using morphologically normal epithelium. As the "normal" epithelium was obtained from the excision specimen it is still within the potentially cancerised field and may well contain carcinogen-induced molecular abnormalities.

It would also have been of benefit to try and account for the effect of intra-tumour heterogeneity. By utilising multiple, topographically separate tissue samples from each individual tumour and nodal metastasis, the real utility of any marker can be evaluated. It would also have been of great interest to evaluate each metastasis (in patients with more than one) and ascertain whether the same clonal population was responsible for each metastasis. When attempting to guide biological therapies it is of limited gain to only target one of the clonal populations present and capable of invasion and metastasis.

7.11 Conclusion

In general my work has revealed genomic differences between metastatic and non-metastatic HNSCC. It has also revealed differing profiles and levels of genomic damage between OPSCC with and without a detectable viral load. miRNAs associated with metastasis were also identified, though little correlation was found between miRNA expression and genomic CNA.

Potentially translational elements of this include a panel of CNAs associated with metastatic primary OSCC tumours and a panel associated with OPSCC with 0 viral load. Statistically significant differences in the FGA of metastatic primary tumours as well as OPSCC with and without detectable viral load were also found and could hold potential for future clinical use. Greater numbers of samples needs to be assessed for these elements as well as the evaluation of the miRNAs associated with metastasis and potential integration of these markers. These genomic features could be combined with traditional histopathological techniques to give an integrated assessment of patient prognosis.

Chapter 8

Appendices

8.1 List of Suppliers

Supplier	Location	e-mail/telephone
Agilent Technologies	Stockport, UK	customercare_uk@agilent.com
Beckman Coulter	High Wycombe, UK	infouk@beckman.com
BDH	Poole, UK	uksales@uk.vwr.com
Cambio	Cambridge, UK	+44 1954 210 200
Covaris Ltd	Brighton, UK	EUcustomerservice@covarisinc.com
Eppendorf AG	Hamburg, Germany	eppendorf@eppendorf.com
Illumina	San Diego, USA	customerservice@illumina.com
Integrated DNA Technologies	Iowa, USA	eutechsupport@idtdna.com
IST Engineering	California, USA	+1 408 586 8889
Life technologies	Paisley, UK	ukorders@lifetech.com
Nanostring Technologies	Seattle, USA	info@nanostring.com
New England Biolabs	Hertfordshire, UK	orders.uk@neb.com
Sigma-Aldrich	St Louis, USA	ukorders@sial.com
Solmedia	UK	labsupplies@soledialtd.com

8.2 Clinicopathologic Data

ECS Code	pT Stage	pN Stage	ECS+/-	Tumour thickness (mm)	Differentiation	Primary Site	Perineural Invasion?	Intravascular Invasion?	Lymphatic Invasion
ECS001	pT2	pN0	N0	5	Well	Oral tongue	No	Yes	No
ECS002	pT1	pN1	ECS-	12	Well	Oral tongue	Yes	No	No
ECS003	pT2	pN2b	ECS-	12	Poor	Oral tongue	Yes	Yes	No
ECS004	pT2	pN0	N0	3.5	Mod	Oral tongue	Yes	No	Yes
ECS005	pT1	pN2b	ECS-	8	Mod	Oral tongue	Yes	No	No
ECS006	pT2	pN2b	ECS+	25	Mod	Oral tongue	No	No	No
ECS007	pT1	pN0	N0	8	Mod	Oral tongue	Yes	No	No
ECS008	pT1	pN1	ECS-	3.6	Mod	Oral tongue	No	No	No
ECS009	pT2	pN1	ECS-	3	Mod	Oral tongue	No	No	Yes
ECS013	pT2	pN1	ECS-	14	Mod	Oral tongue	Yes	Yes	No
ECS014	pT1	pN1	ECS-	2.5	Poor	Oral tongue	No	No	No
ECS015	pT3	pN0	N0	10	Well	Oral tongue	No	No	No
ECS016	pT2	pN1	ECS+	10	Mod	Oral tongue	No	No	No
ECS017	pT2	pN0	N0	6	Mod	Oral tongue	No	No	No
ECS018	pT2	pN2b	ECS+	15	Well	Oral tongue	No	No	No
ECS019	pT1	pN0	N0	-	Well	Oral tongue	No	No	No
ECS020	pT1	pN0	N0	-	Well	Oral tongue	No	No	No
ECS021	pT1	pN0	N0	-	Poor	Oral tongue	Yes	No	No
ECS022	pT1	pN0	N0	5	Well	Oral tongue	No	No	No
ECS023	pT1	pN0	N0	3.1	Mod	Oral tongue	No	No	No
ECS024	pT1	pN0	N0	6	Mod	Oral tongue	No	No	No
ECS025	pT1	pN0	N0	3	Poor	Oral tongue	No	No	No
ECS026	pT1	pN0	N0	16	Well	Oral tongue	No	No	No
ECS027	pT1	pN0	N0	5	Mod	Oral tongue	No	No	No
ECS028	pT1	pN0	N0	8	Mod	Oral tongue	Yes	No	No
ECS029	pT1	pN0	N0	9	Well	Oral tongue	No	No	No

ECS030	pT1	pN0	N0	-	Well	Oral tongue	No	No	No
ECS031	pT1	pN0	N0	3	Well	Oral tongue	No	No	No
ECS032	pT!	pN0	N0	6	Well	Oral tongue	No	No	No
ECS033	pT1	pN1	ECS-	9.5	Poor	Oral tongue	No	No	No
ECS034	pT1	pN1	ECS-	4	Poor	FOM/Oraltongue	Yes	Yes	No
ECS035	pT2	pN1	ECS+	27	Poor	Oral tongue	Yes	No	Yes
ECS036	pT1	pN2b	ECS+	-	Mod	Oral tongue	Yes	No	No
ECS038	pT1	pN2b	ECS+	8	Poor	Oral tongue	Yes	No	No
ECS039	pT1	pN2b	ECS+	14	Mod	Oral tongue	No	No	No
ECS040	pT1	pN2a	ECS+	11	Poor	Oral tongue	Yes	Yes	No
ECS041	pT1	pN1	ECS-	14	Poor	Oral tongue	Yes	Yes	No
ECS042	pT2	pN0	N0	21	Poor	Oral tongue	Yes	No	No
ECS043	pT2	pN0	N0	12	Mod	Oral tongue	No	No	No
ECS044	pT2	pN0	N0	16	Poor	Oral tongue	Yes	No	No
ECS045	pT2	pN0	N0	14	Mod	Oral tongue	No	No	No
ECS047	pT2	pN0	N0	21	Well	Oral tongue	No	No	No
ECS048	pT2	pN0	N0	12	Well	Oral tongue	Yes	No	No
ECS049	pT2	pN1	ECS-	9	Poor	Oral tongue	Yes	No	No
ECS050	pT2	pN1	ECS-	14	Poor	FOM/Oral tongue	Yes	Yes	No
ECS053	pT1	pN2b	ECS+	12	Poor	Oral tongue	Yes	Yes	Yes
ECS054	pT1	pN1	ECS-	22	Poor	Oral tongue	Yes	No	No
ECS055	pT2	pN1	ECS-	25	Poor	Oral tongue	No	No	No
ECS056	pT2	pN2b	ECS-	10	Mod	Oral tongue	No	Yes	No
ECS057	PT2	pN2b	ECS-	-	Well	Oral tongue	Yes	Yes	No
ECS059	pT2	pN1	ECS+	11	Mod	Oral tongue	No	No	No
ECS060	pT2	pN2b	ECS+	-	Well	Oral tongue	Yes	Yes	No
ECS061	pT2	pN2a	ECS+	32	Mod	Buccal	Yes	No	No
ECS062	pT2	pN2b	ECS+	14	Well	Oral tongue	No	No	No
ECS063	pT2	pN2c	ECS+	11	Poor	Oral tongue	No	No	Yes

ECS064	pT3	N0	N0	12	Well	Oral tongue	Yes	No	No
ECS065	pT3	pN1	ECS-	32	Well	Oral tongue	No	No	No
ECS066	pT3	pN2b	ECS+	45	Mod	Oral tongue	Yes	No	No
ECS067	pT3	pN2b	ECS+	15	Poor	FOM/Oral tongue	Yes	No	Yes
ECS068	pT3	pN2c	ECS+	-	Mod	Oral tongue	Yes	No	No
ECS069	pT3	pN3	ECS+	28	Well	Oral tongue	No	No	No
ECS070	pT2	pN2b	ECS-	-	Poor	Oral tongue	No	Yes	No
ECS071	pT2	pN2b	ECS+	-	Poor	Oral tongue	Yes	Yes	Yes
ECS073	pT4a	pN2c	ECS+	40	Poor	Oral tongue	Yes	Yes	No
ECS075	pT2	pN2b	ECS+	7	Mod	Oral tongue	Yes	No	No
ECS076	pT2	pN2b	ECS+	5	Poor	FOM/Oral tongue	Yes	Yes	Yes
ECS078	pT2	pN1	ECS-	8	Mod	FOM/Oral tongue	No	Yes	Yes
ECS079	pT3	pN1	ECS-	11	Well	FOM/Oral tongue	No	No	No
ECS080	pT4a	pN2b	ECS-	8.4	Well	Alveolar	Yes	No	No
ECS081	pT3	pN2a	ECS+	18	Well	Buccal	Yes	Yes	No
ECS082	pT2	pN2b	ECS+	9	Mod	Buccal	No	No	No
ECS083	pT1	pN2b	ECS-	15	Poor	FOM	Yes	No	No
ECS084	pT4a	pN3	ECS+	12	Poor	FOM	No	Yes	No
ECS085	pT4a	pN2c	ECS+	-	Well	Oral tongue	Yes	Yes	Yes
ECS087	pT4a	pN2b	ECS+	11	Poor	FOM	Yes	No	No

Table 8-1: Histopathologic details of patient samples.

ECS Code	Age at diagnosis	Postop XRT?	Disease free F/U? (Months)	Overall Survival (Months)	Recurrence?	Site?	Distant mets?	Alive/Dead
ECS001	50yr 2mth	No	14	14	No	-	No	Dead
ECS002	74yr 7mth	No	64	64	No	-	No	Alive

ECS003	75yr 0mth	Yes	60	60	No	-	No	Alive
ECS004	52yr 7mth	Yes	60	60	No	-	No	Alive
ECS005	60yr 7mth	Yes	15	15	No	-	No	Dead
ECS006	52yr 11mth	Yes	2	3	No	-	Yes	Dead
ECS007	53yr 5mth	No	53	53	-	-	No	Alive
ECS008	72yr 9mth	No	52	52	No	-	No	Alive
ECS009	64yr 9mth	No	7	12	Yes	Neck	Yes	Dead
ECS013	66yr 8mth	No	47	47	No	-	No	Alive
ECS014	52yr 3mth	No	47	47	No	-	No	Alive
ECS015	90yr 2mth	No	46	46	No	-	No	Alive
ECS016	75yr 8mth	Yes	43	43	No	-	No	Alive
ECS017	84yr 11mth	No	43	43	No	-	No	Alive
ECS018	84yr 6mth	Yes	42	42	No	-	No	Alive
ECS019	47yr 2mth	No	111	111	No	-	No	Alive
ECS020	52yr 10mth	No	111	111	No	-	No	Alive
ECS021	58yr 5mth	Yes	90	90	No	-	No	Alive
ECS022	48yr 9mth	No	91	91	No	-	No	Alive
ECS023	48yr 9mth	No	16	30	No	-	No	Dead
ECS024	65yr 4mth	No	89	89	No	-	No	Alive
ECS025	54yr 0mth	No	85	85	No	-	No	Alive
ECS026	56yr 8mth	No	91	91	No	-	No	Alive
ECS027	67yr 3mth	No	97	97	No	-	No	Alive
ECS028	61yr 0mth	No	80	80	No	-	No	Dead
ECS029	53yr 6mth	No	102	102	No	-	No	Alive
ECS030	74yr 8mth	Yes	72	72	No	-	No	Alive
ECS031	67yr 0mth	No	80	80	No	-	No	Alive
ECS032	61yr 1mth	No	37	37	No	-	No	Dead
ECS033	39yr 11mth	Yes	17	22	No	-	Yes	Dead
ECS034	68yr 10mth	No	39	39	No	-	No	Alive

ECS035	80yr 6mth	No	2	4	Yes	Local	No	Dead
ECS036	80yr 7mth	Yes	4	4	Yes	Neck	No	Dead
ECS038	77yr 1mth	Yes	2	5	Yes	Neck	No	Dead
ECS039	55yr 6mth	Yes	2	7	Yes	Neck	No	Dead
ECS040	54yr 10mth	Yes	36	36	No	-	No	Dead
ECS041	70yr 11mth	Yes	81	81	No	-	No	Alive
ECS042	68yr 7mth	No	66	66	No	-	No	Alive
ECS043	56yr 1mth	Yes	93	93	No	-	No	Alive
ECS044	59yr 7mth	Yes	70	70	No	-	No	Alive
ECS045	71yr 0mth	No	92	92	No	-	No	Alive
ECS047	70yr 0mth	Yes	22	22	No	-	No	Dead
ECS048	47yr 0mth	No	18	31	Yes	Oral and Neck	Yes	Dead
ECS049	53yr 8mth	Yes	77	77	No	-	No	Alive
ECS050	57yr 1mth	Yes	72	72	No	-	No	Alive
ECS053	66yr 4mth	Yes	2	67	Yes	Neck	No	Alive
ECS054	56yr 11mth	Yes	66	66	No	-	No	Alive
ECS055	60yr 8mth	Yes	28	28	No	-	No	Dead
ECS056	57yr 1mth	Yes	63	63	No	-	No	Dead
ECS057	60yr 1mth	Yes	8	8	No	-	-	Dead
ECS059	67yr 11mth	No	2	2	No	-	No	Dead
ECS060	52yr 3mth	Yes	52	52	No	-	No	Dead
ECS061	64yr 0mth	Yes	32	38	No	-	Yes	Dead
ECS062	71yr 6mth	Yes	6	9	Yes	Neck	No	Dead
ECS063	74yr 3mth	Yes	9	9	No	-	No	Dead
ECS064	79yr 3mth	No	8	18	Yes	Oral, Neck	No	Dead
ECS065	72yr 10mth	No	5	5	No	-	No	Dead
ECS066	83yr 5mth	Yes	5	5	No	-	No	Dead
ECS067	56yr 7mth	Yes	34	34	No	-	No	Alive
ECS068	74yr 5mth	Yes	9	15	Yes	Oral and Neck	No	Dead

ECS069	77yr 10mth	Yes	4	5	Yes	Oral and Neck	No	Dead
ECS070	54yr 3mth	Yes	59	59	No	-	No	Alive
ECS071	36yr 4mth	Yes	4	4	No	-	Yes	Dead
ECS073	56yr 11mth	Yes	5	5	No	-	No	Dead
ECS075	47yr 2mth	No	1	1	No	-	No	Dead
ECS076	60yr 1mth	Yes	25	46	Yes	pyriform fossa	No	Dead
ECS078	62yr 11mth	Yes	40	40	No	-	No	Alive
ECS079	77yr 10mth	Yes	4	4	No	-	No	Dead
ECS080	60yr 2mth	Yes	54	54	No	-	No	Alive
ECS081	61yr 4mth	Yes	6	7	Yes	Oral and Neck	No	Dead
ECS082	56yr 3mth	Yes	1	2	Yes	Oral and Neck	No	Dead
ECS083	43yr 1mth	Yes	47	47	No	-	No	Alive
ECS084	44yr 1mth	No	5	8	Yes	Neck	No	Dead
ECS085	54yr 1mth	No	0	1	Yes	Oral and Neck	No	Dead
ECS087	61yr 1mth	Yes	2	10	Yes	Neck	No	Dead

Table 8-2: Clinical data for patient samples.

8.3 Digital karyograms for all OSCC primary tumours and lymph node metastases

Primary tumours are denoted by the suffix '–T' and metastases by '–L'.

ECS001-T

ECS002-T

ECS002-L

ECS003-T

ECS003-L

ECS004-T

ECS005-T

ECS005-L

ECS006-T

ECS006-L

288
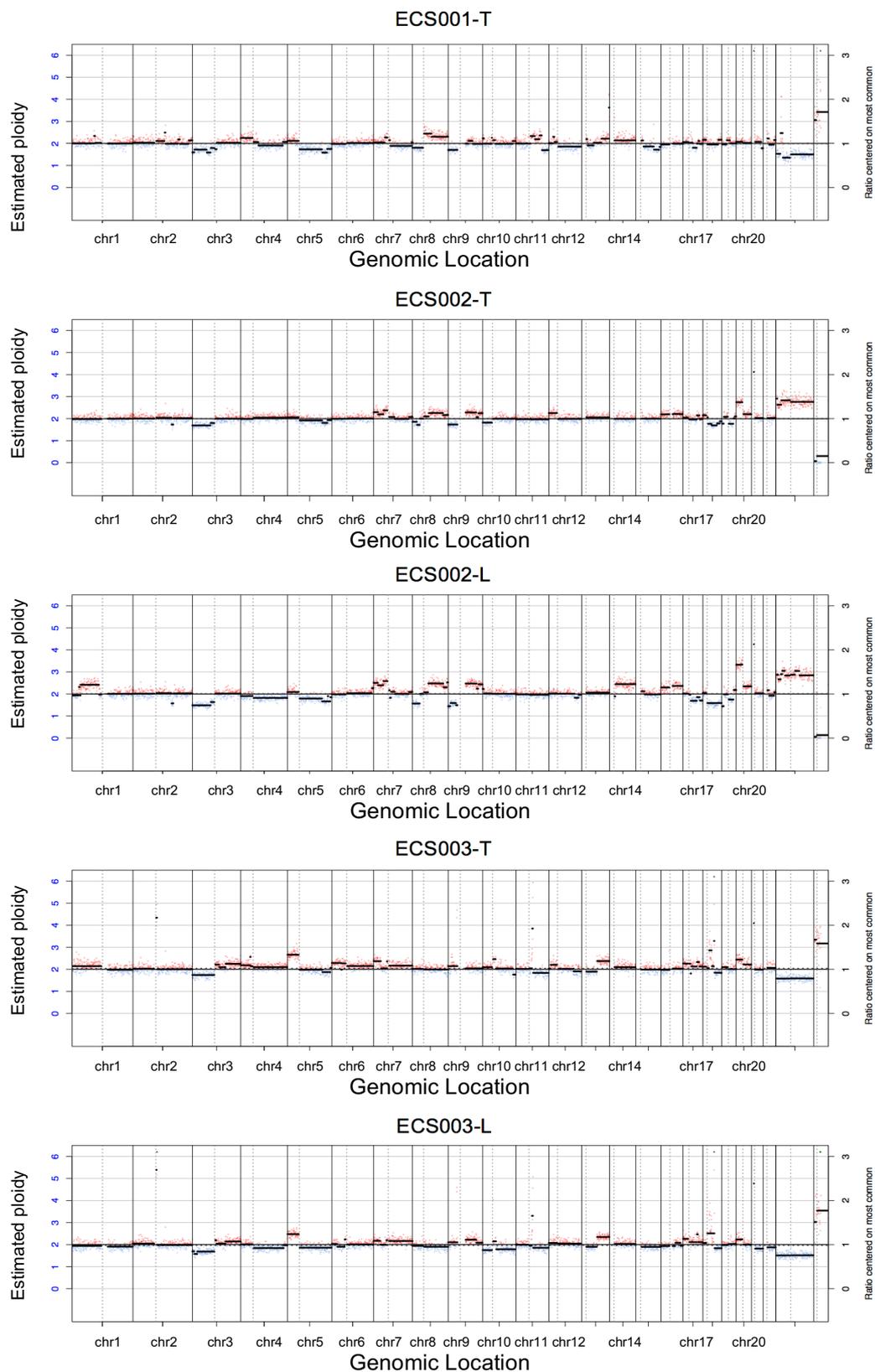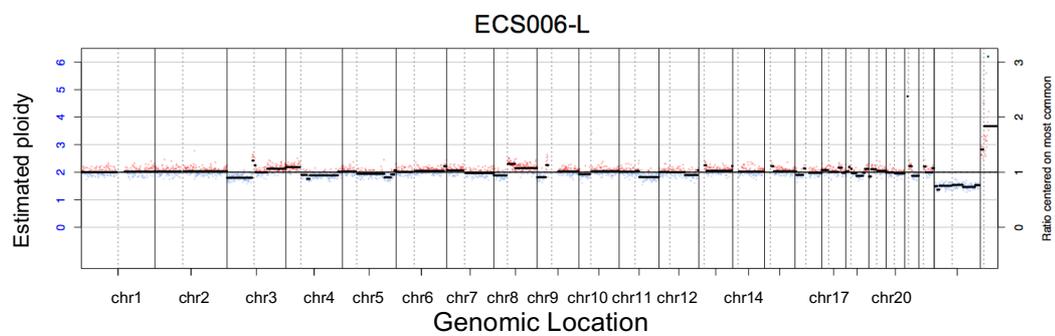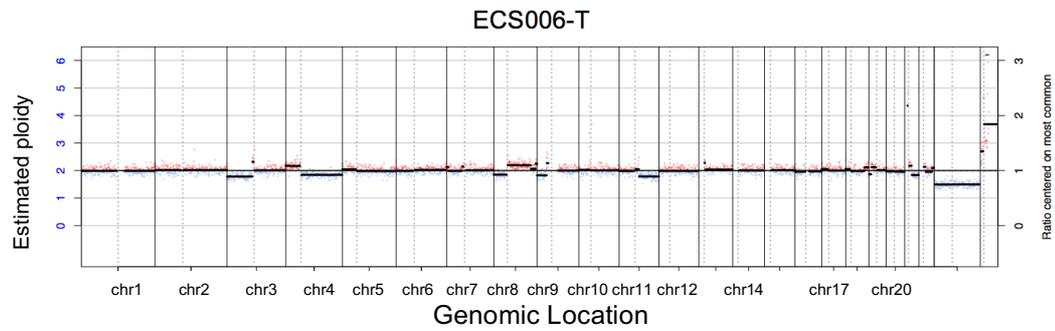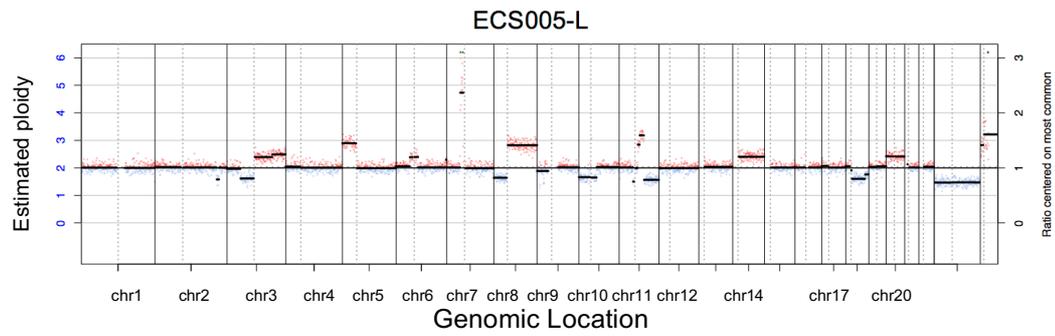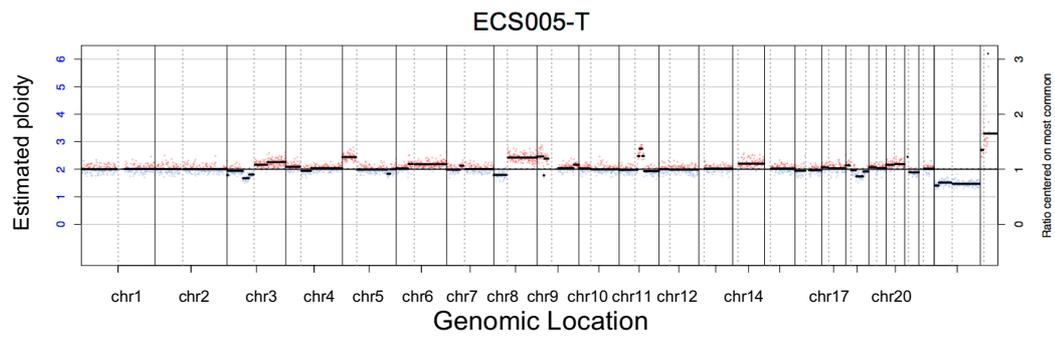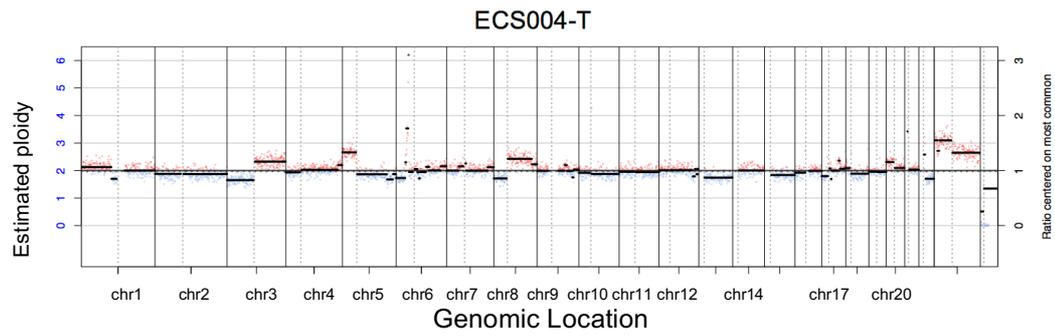
ECS007-T
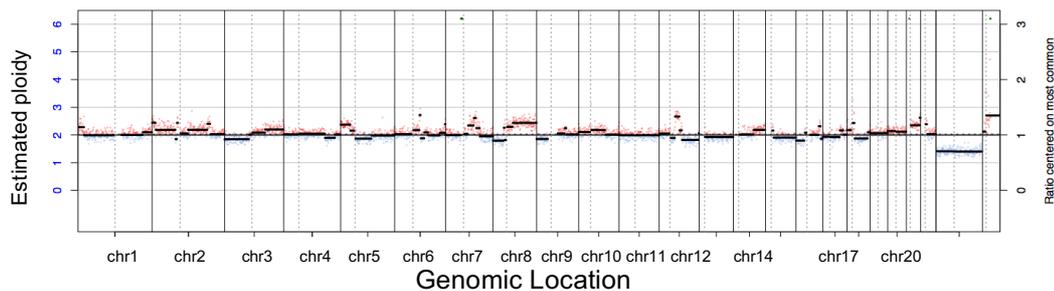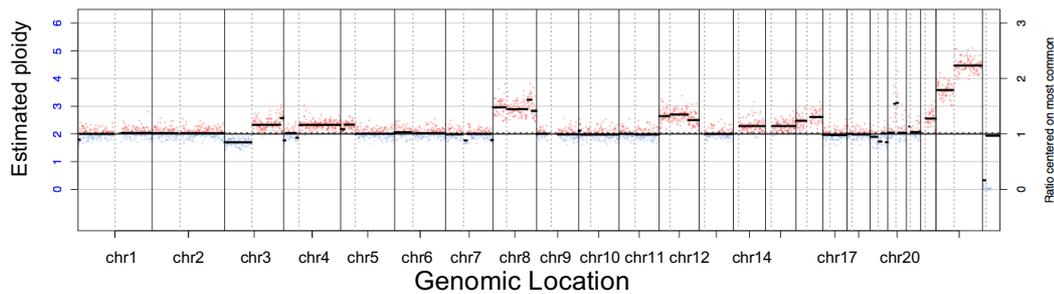
ECS008-T

ECS008-L

ECS009-T

ECS009-L

289
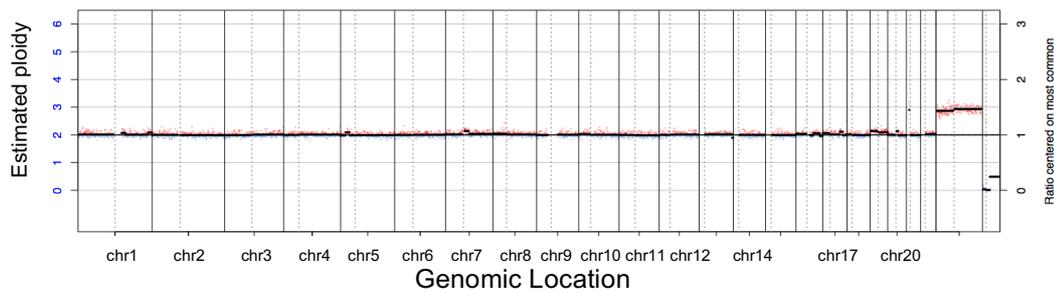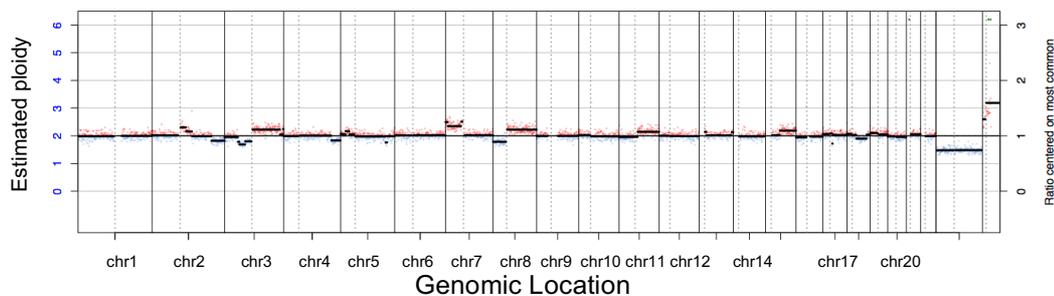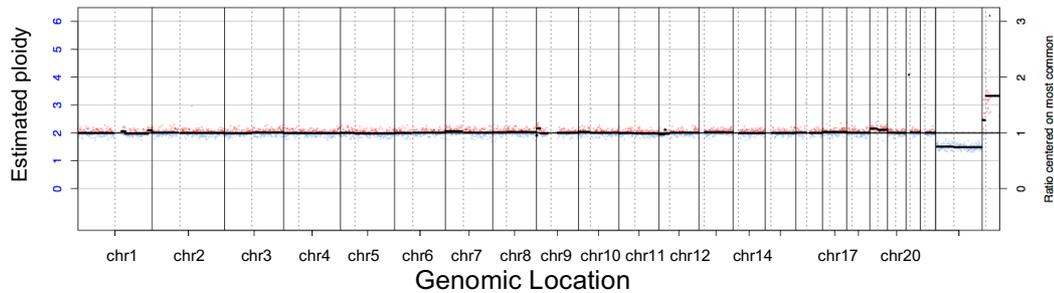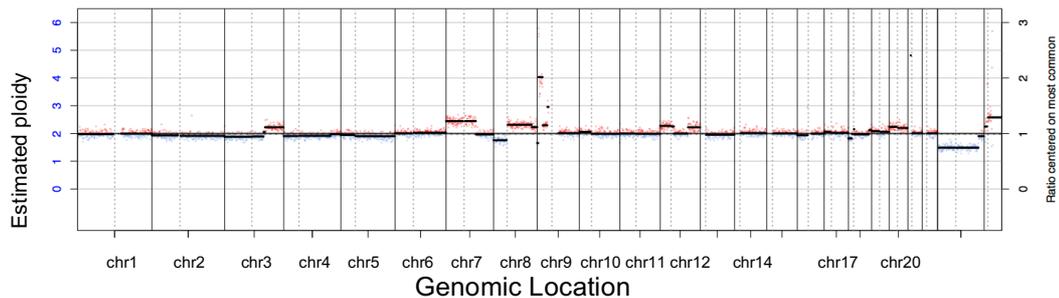
ECS013-T

ECS013-L
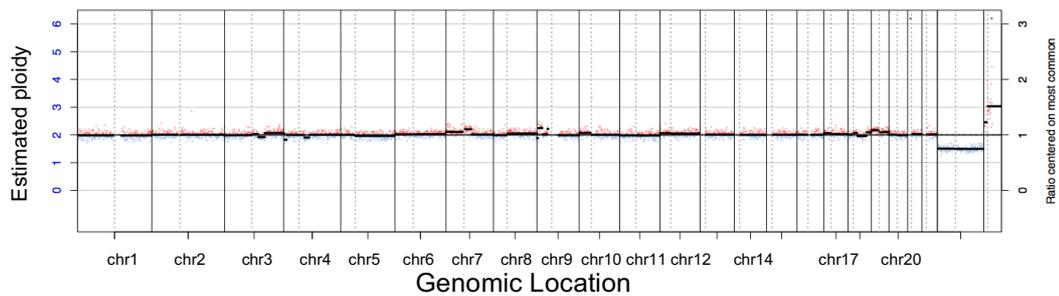
ECS014-T

ECS014-L

ECS015-T

290
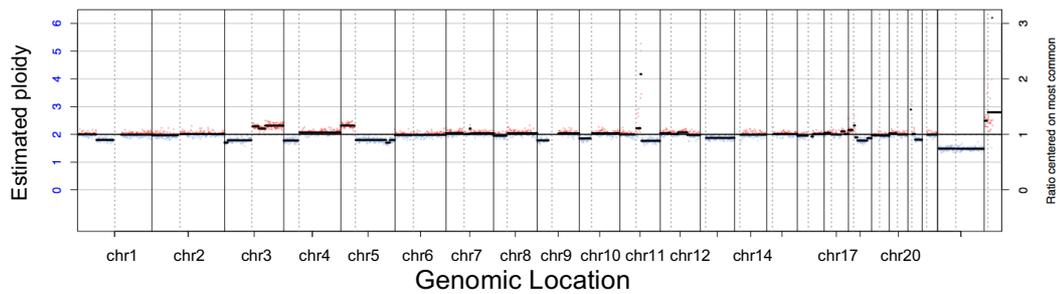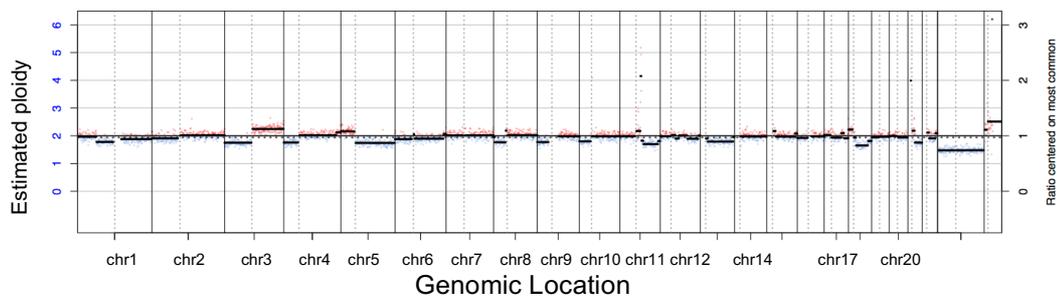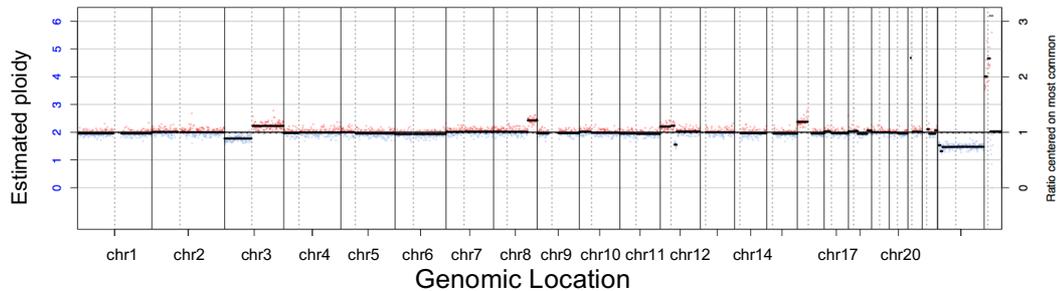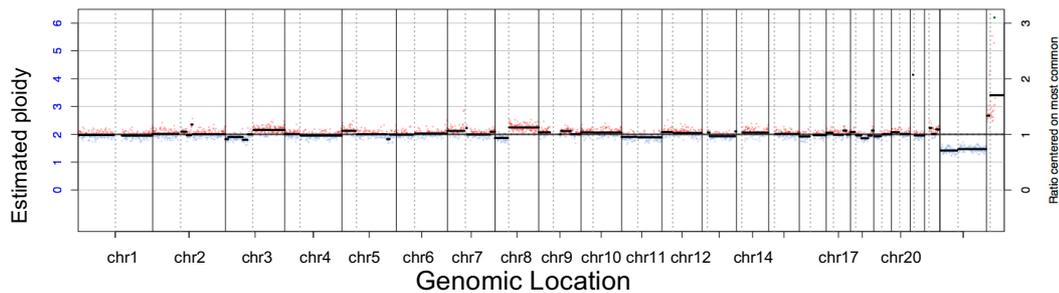
ECS016-T

ECS016-L
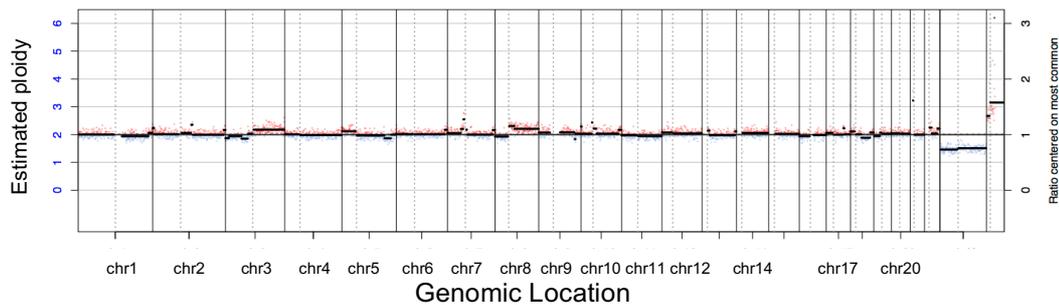
ECS017-T
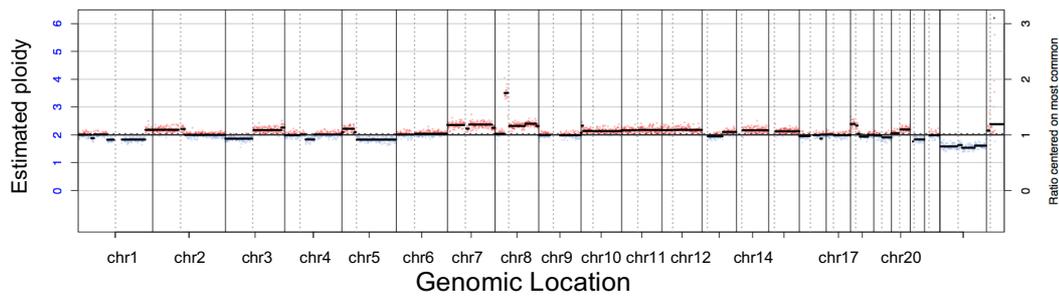
ECS018-T
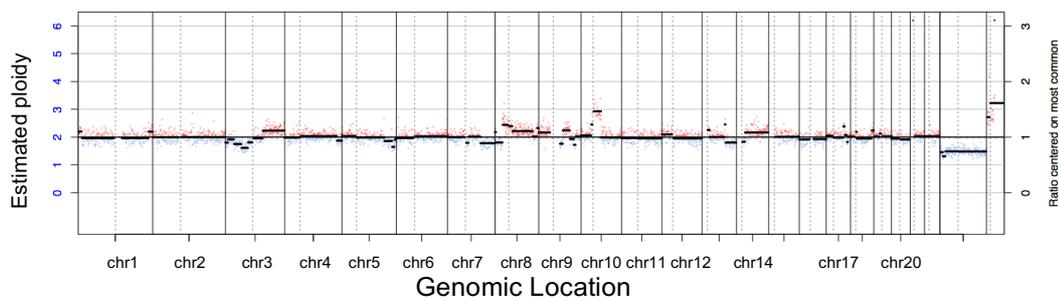
ECS018-L
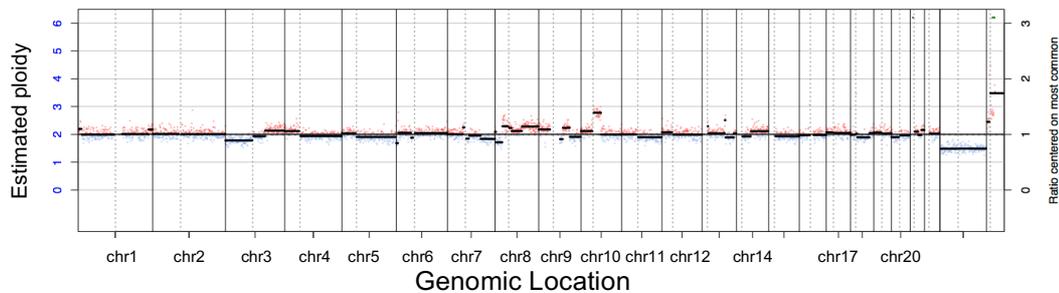
ECS019-T
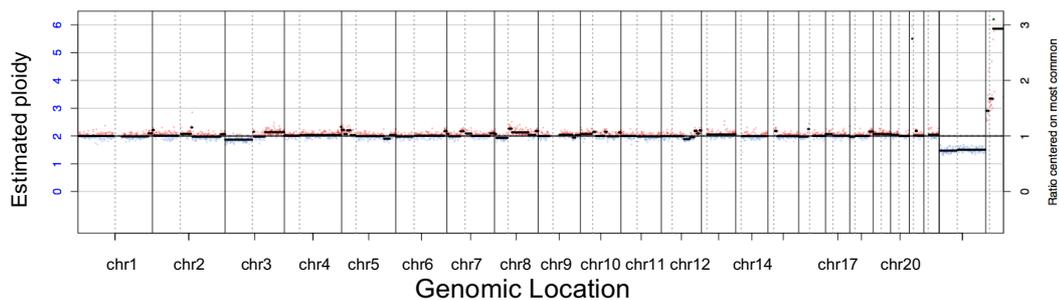
ECS020-T
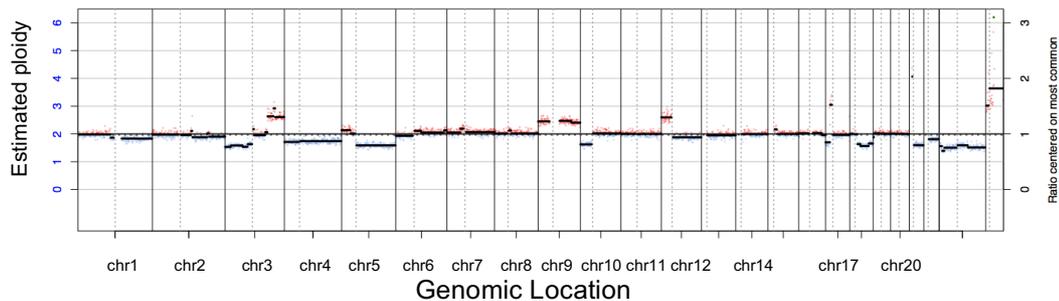
ECS021-T

ECS022-T

ECS023-T

292
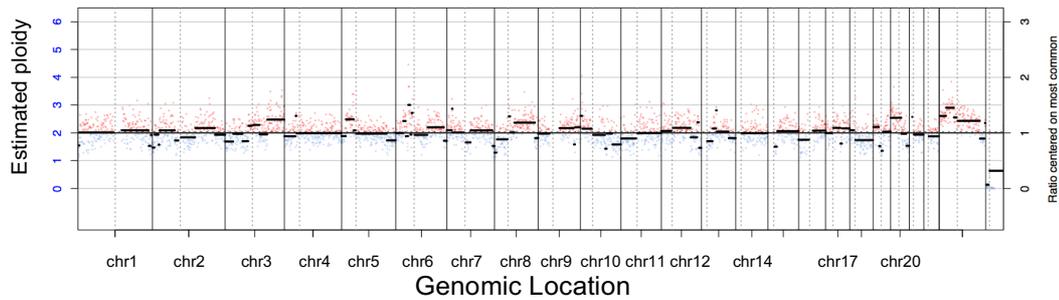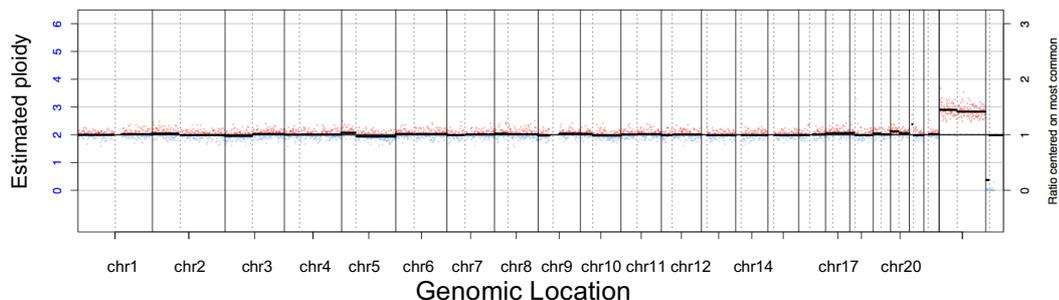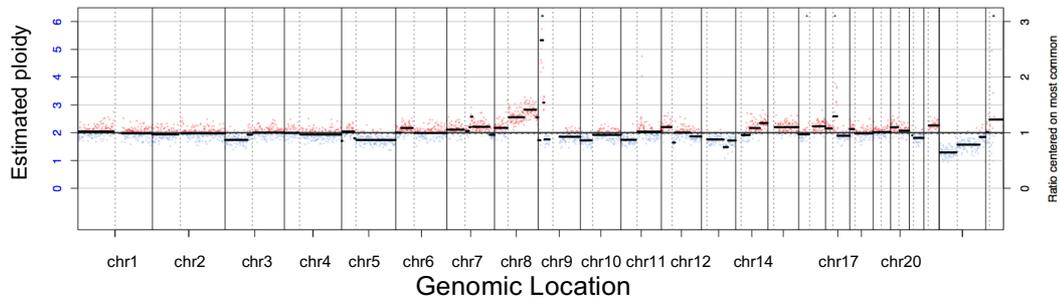
ECS024-T
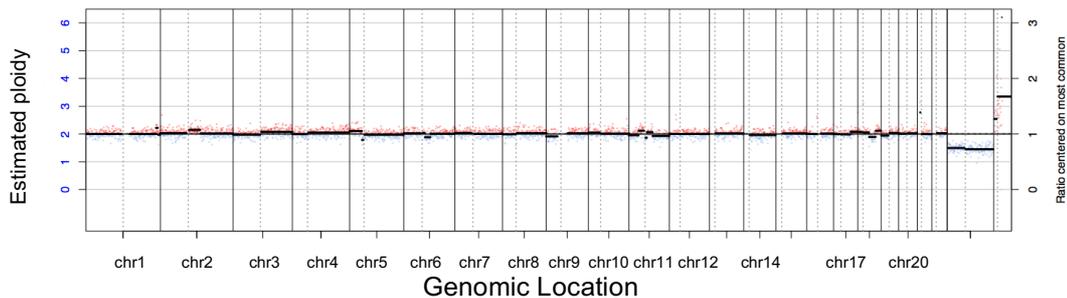
ECS025-T
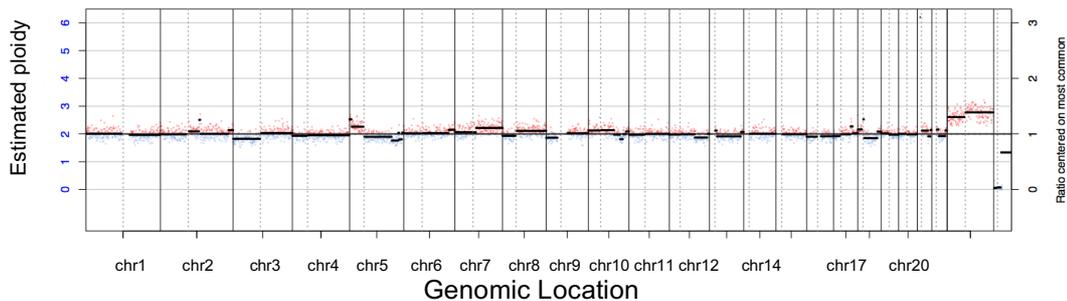
ECS026-T

ECS027-T

ECS028-T

293
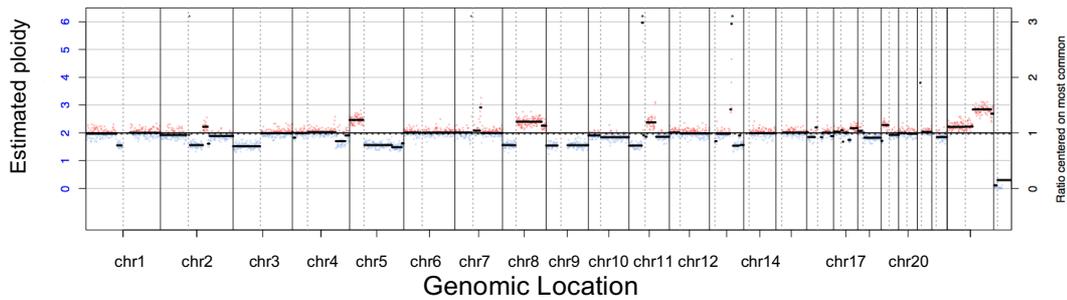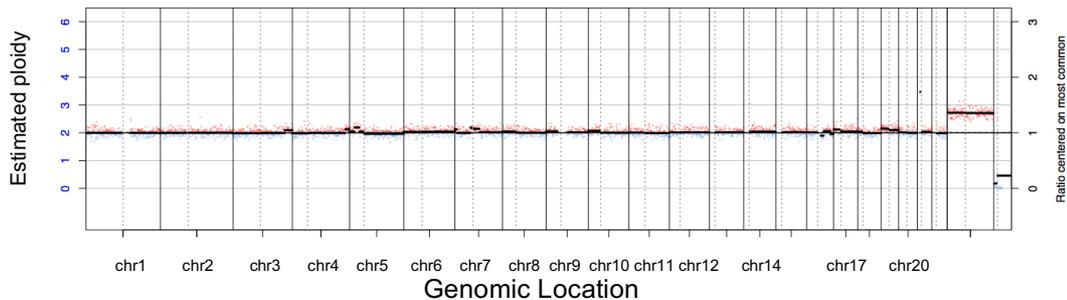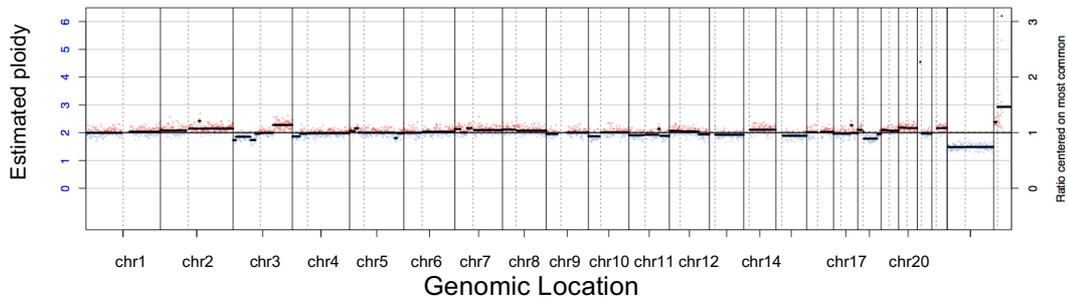
ECS029-T
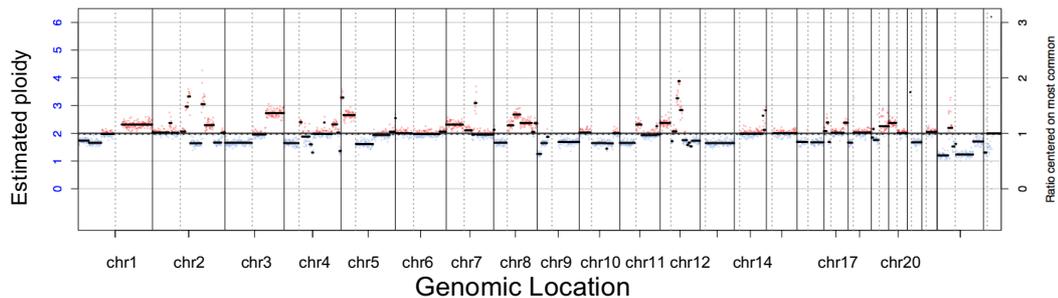
ECS030-T
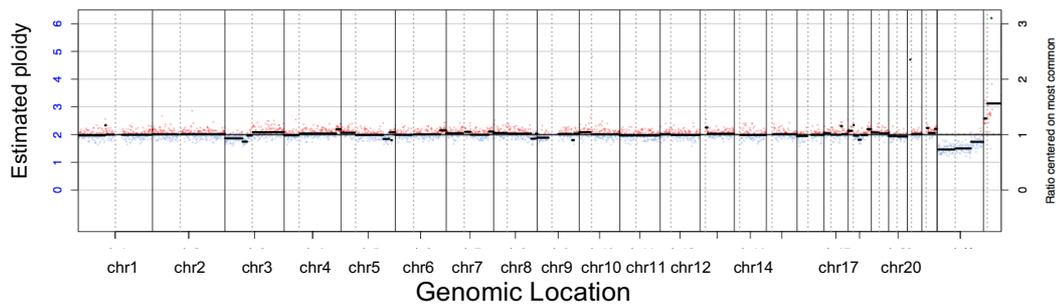
ECS031-T
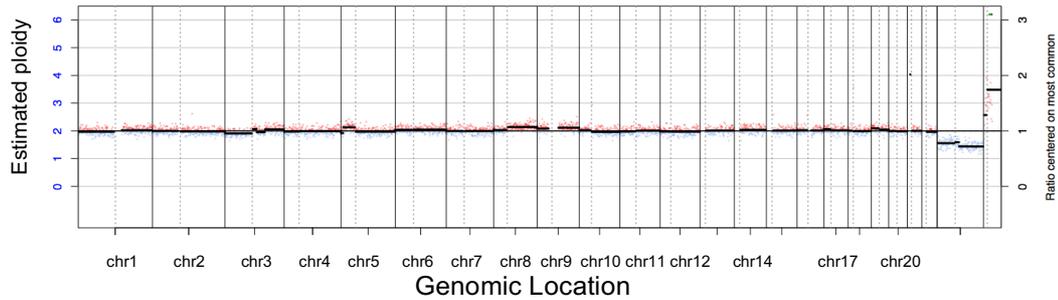
ECS033-T
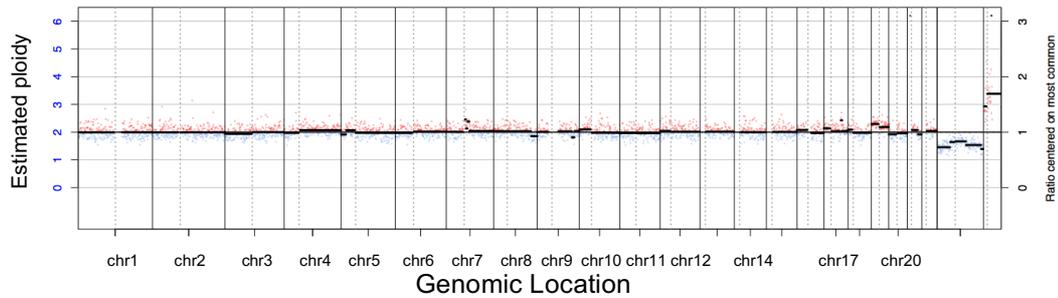
ECS033-T
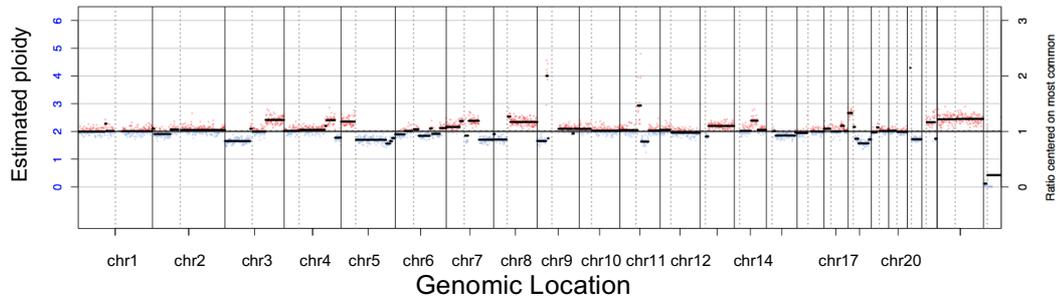
ECS033-L
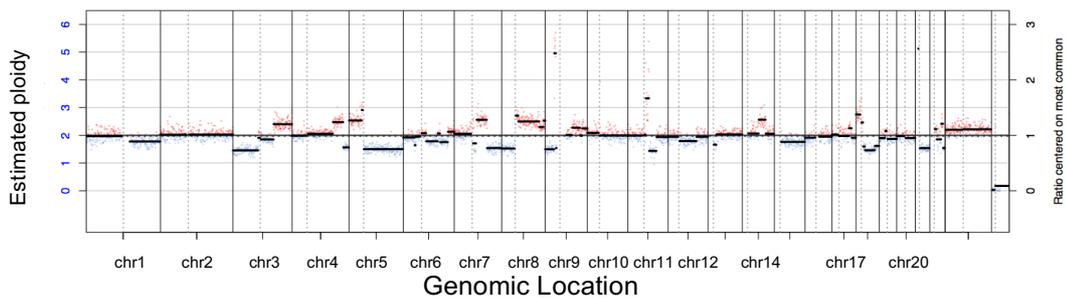
ECS034-T
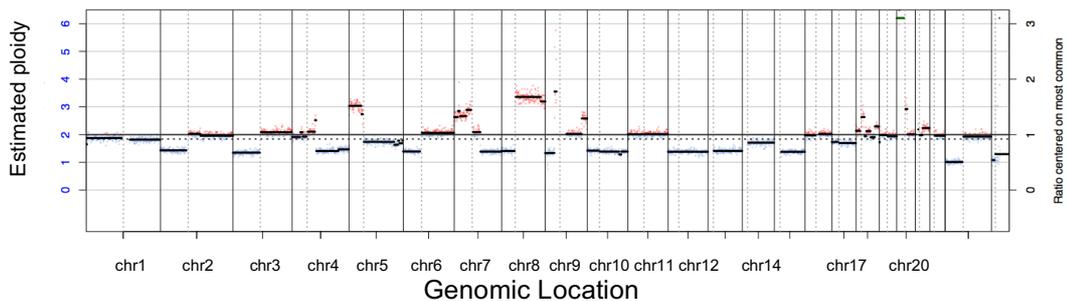
ECS034-L

ECS035-T

ECS035-L

295
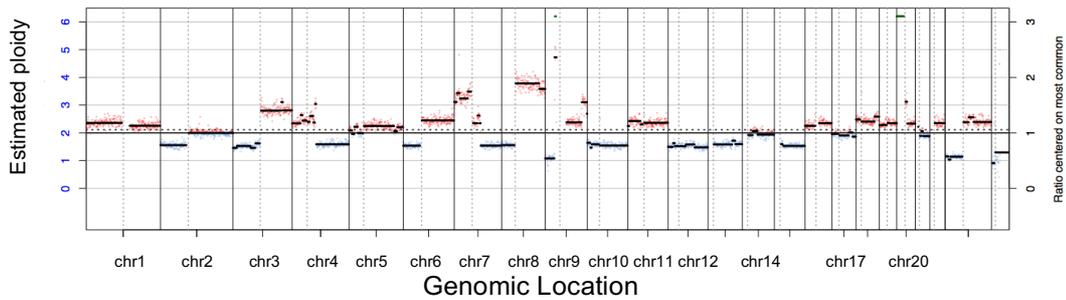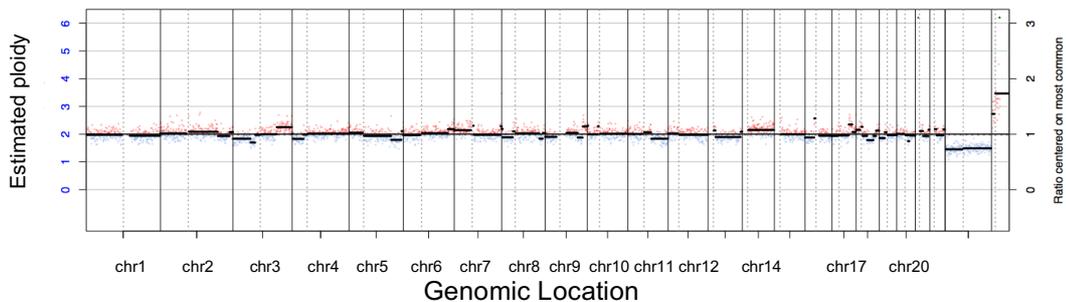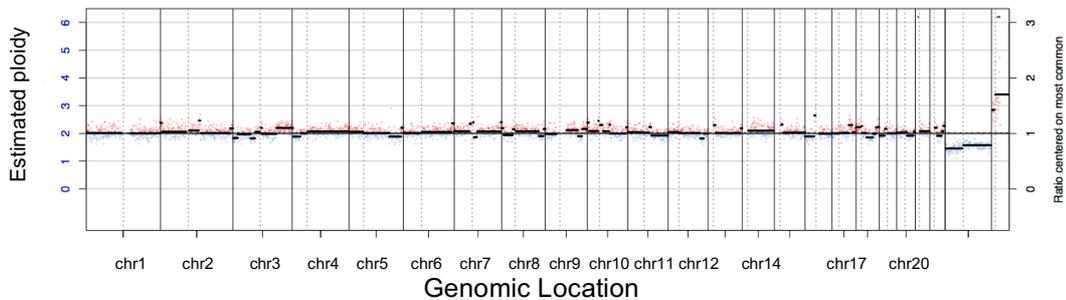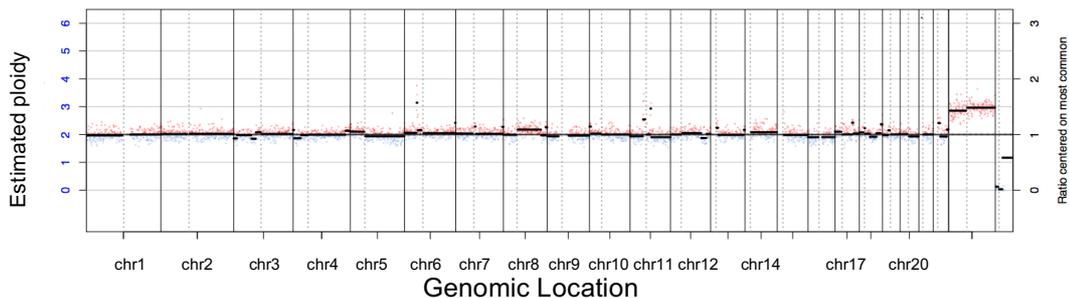
ECS036-T

ECS036-L
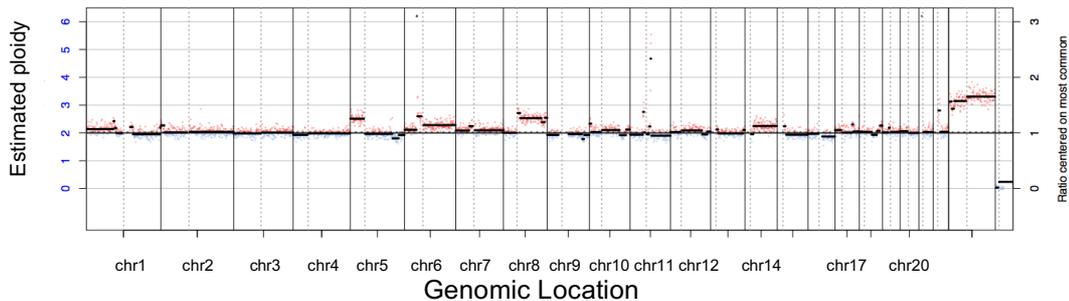
ECS038-T

ECS038-L

ECS039-T

296

ECS039-L

ECS040-T

ECS040-L

ECS041-T

ECS041-L
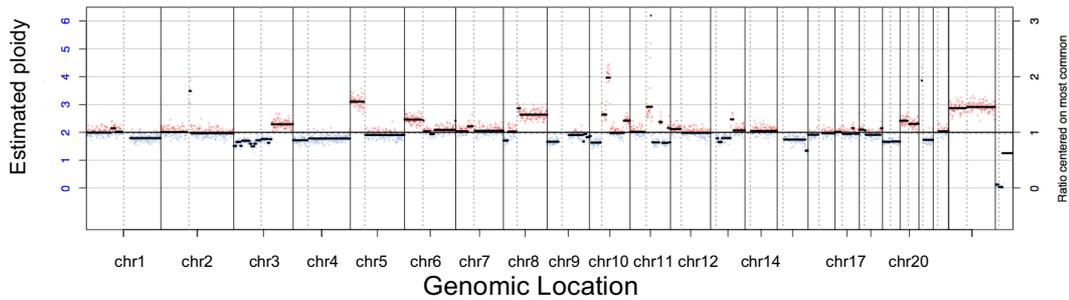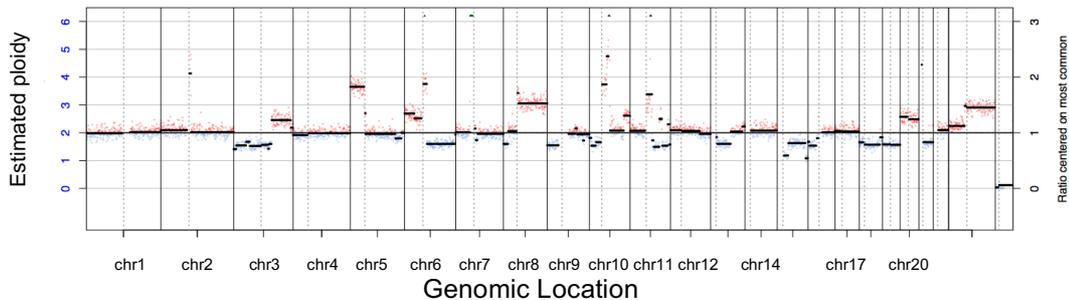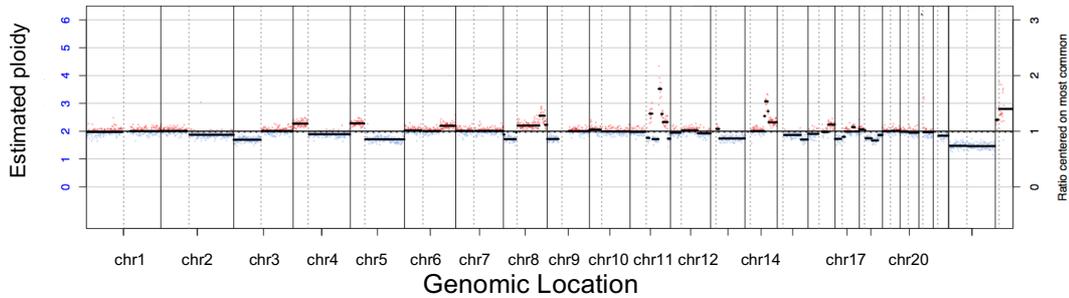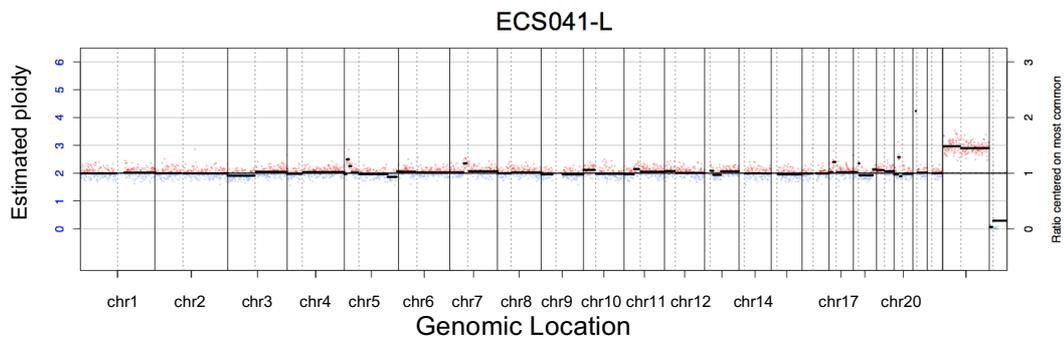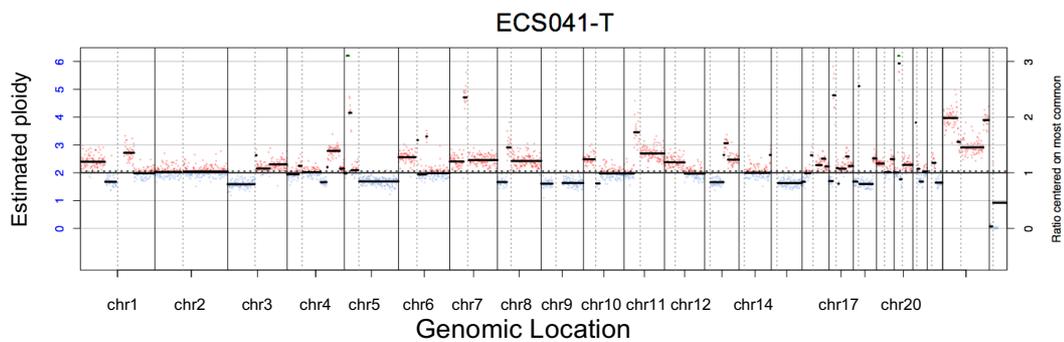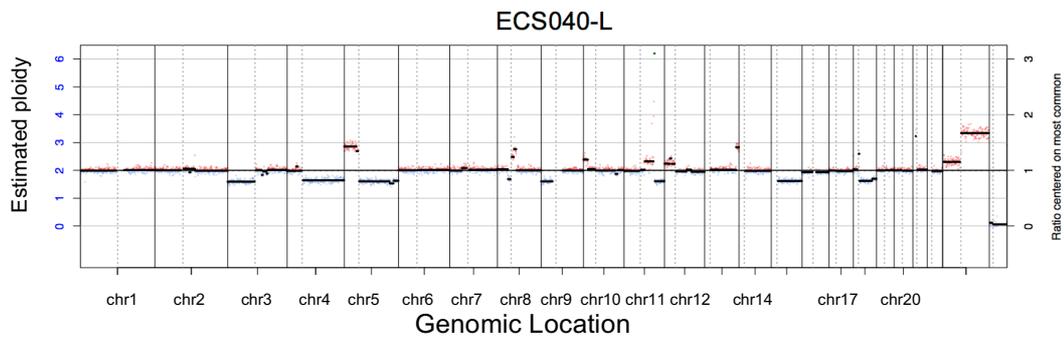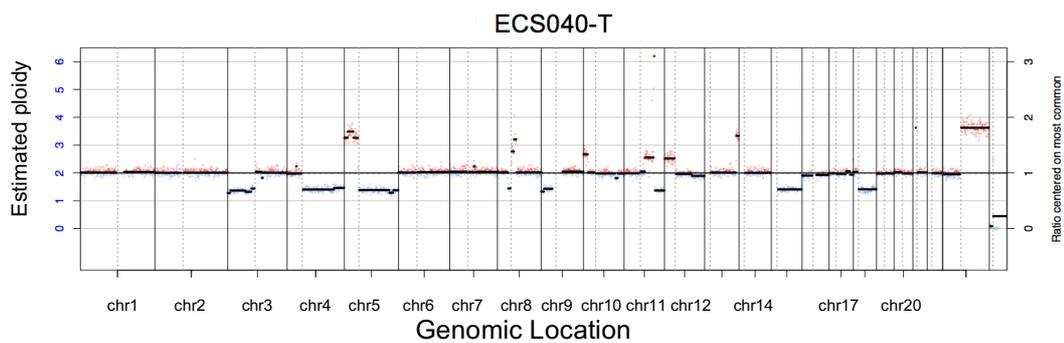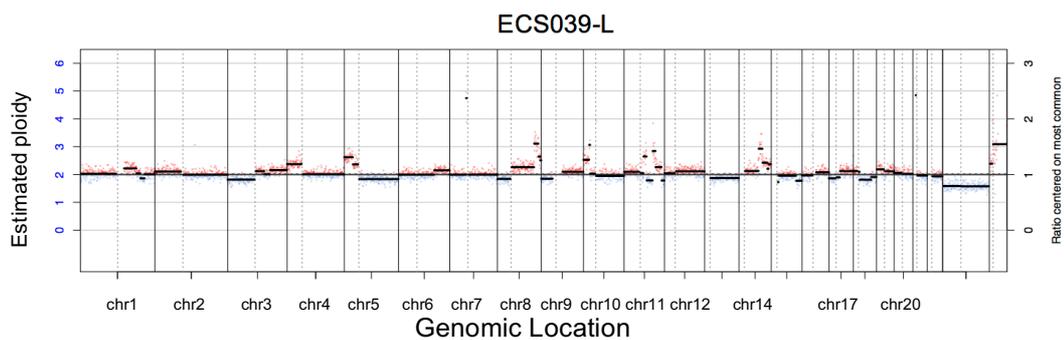
ECS042-T
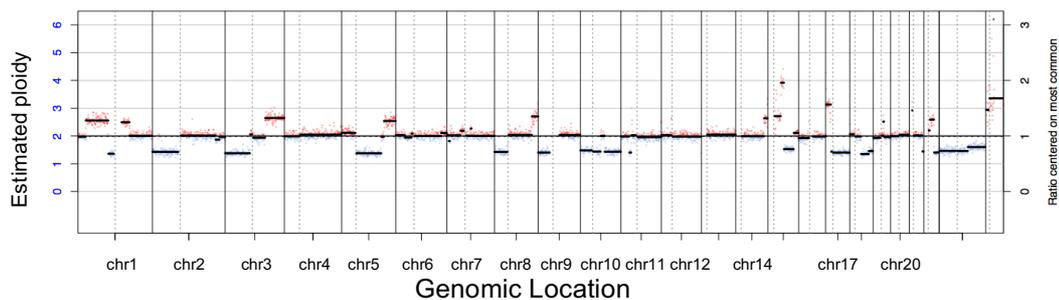
ECS043-T
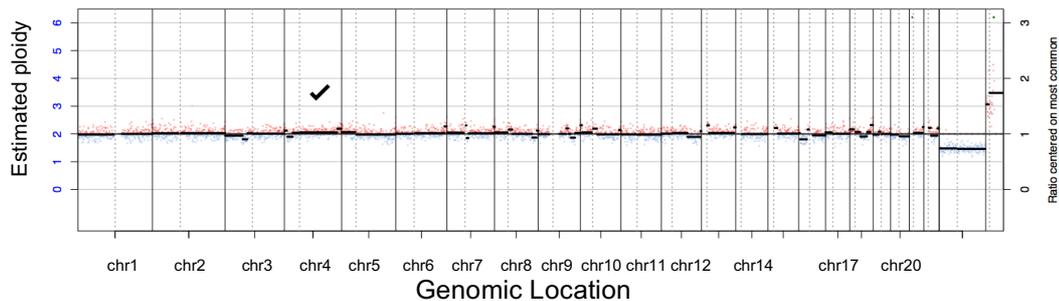
ECS044-T
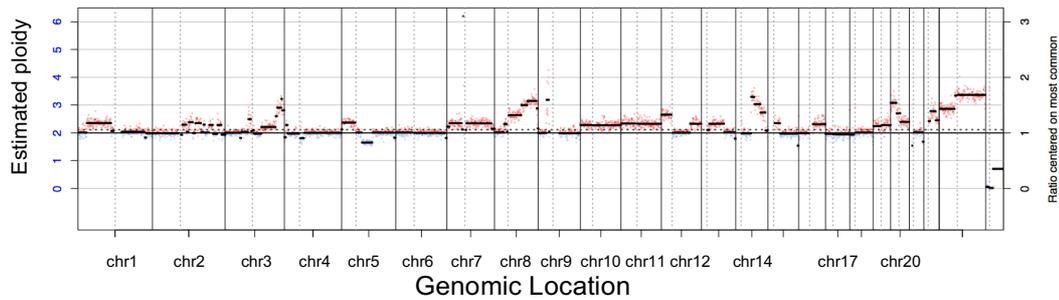
ECS045-T
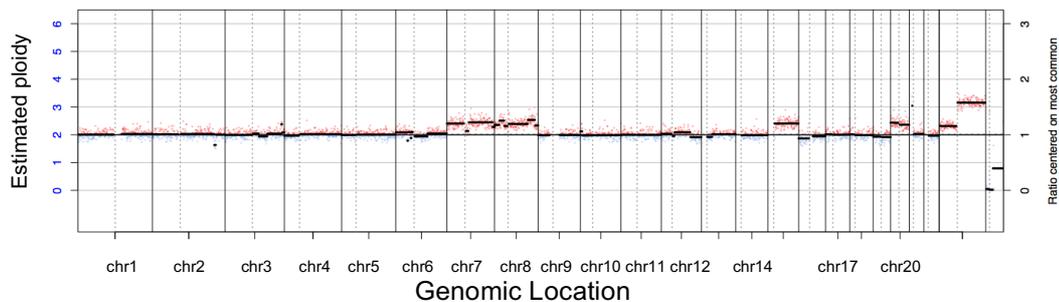
ECS047-T
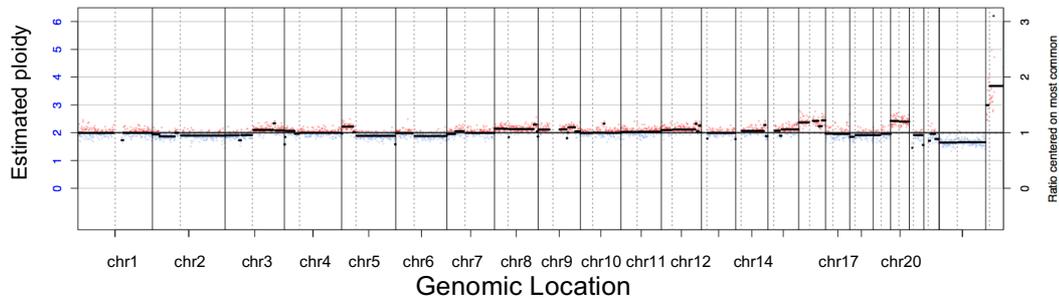
ECS048-T
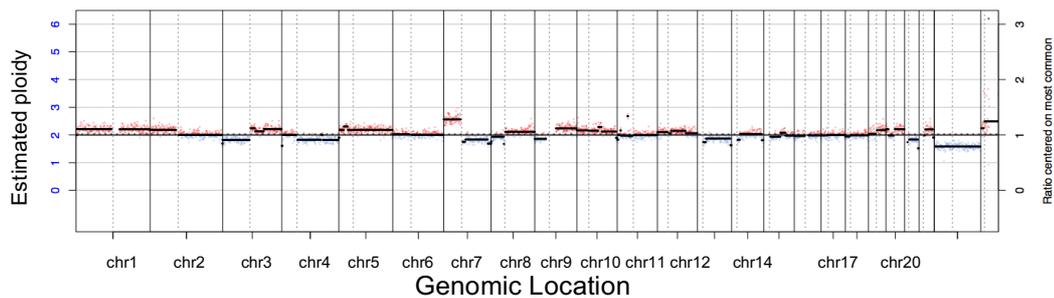
ECS049-T
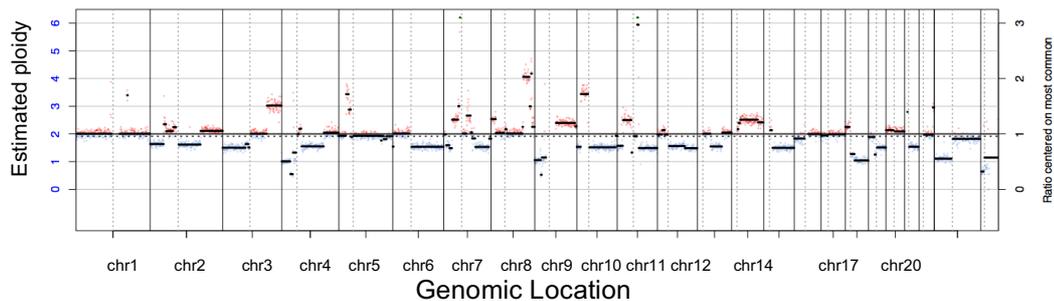
ECS049-L

ECS050-T

ECS050-L

299
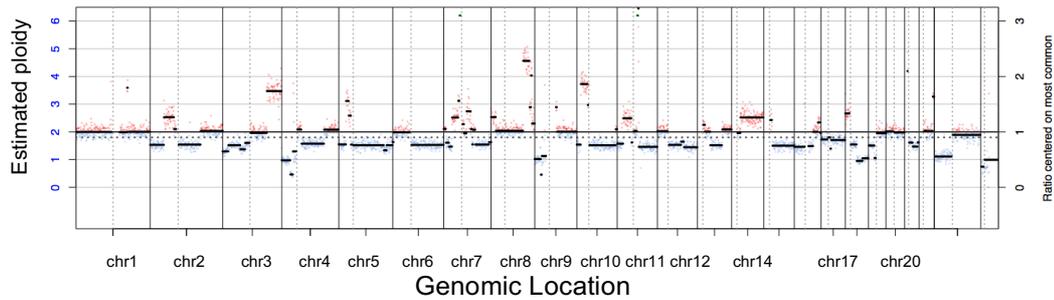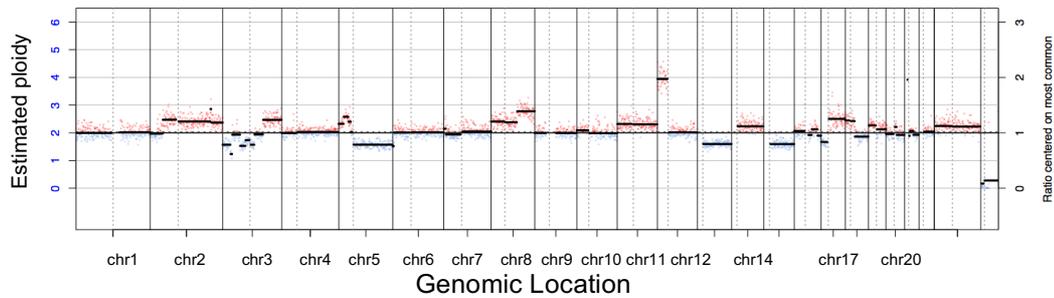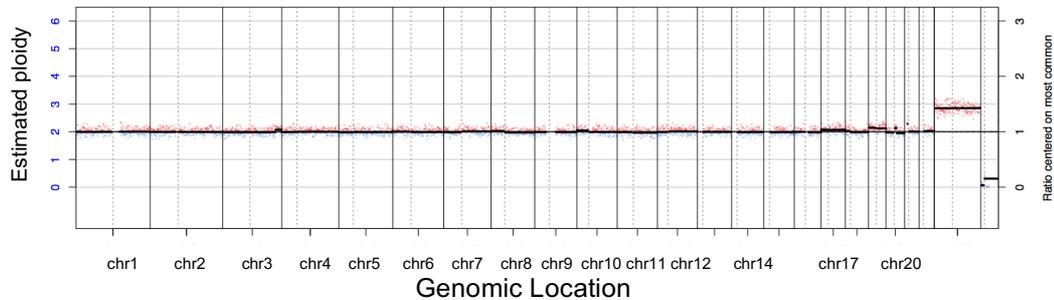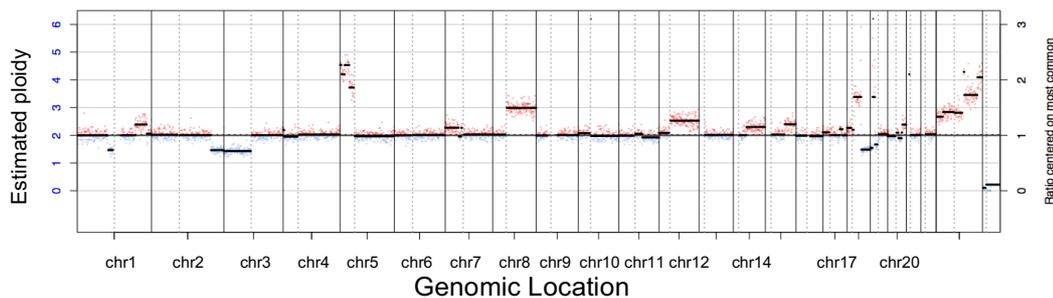
ECS053-T

ECS053-L
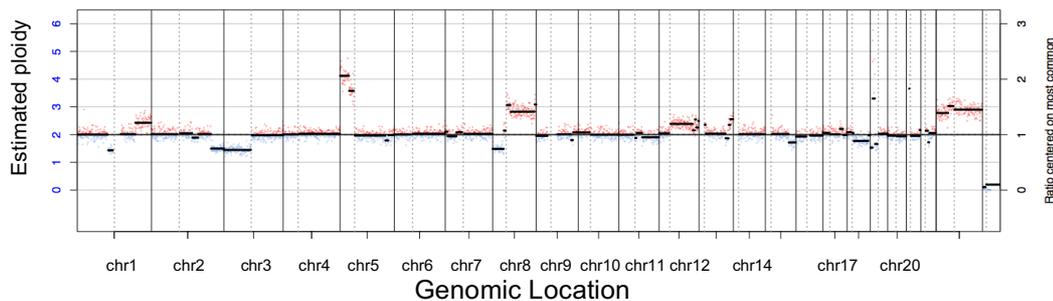
ECS054-T

ECS054-L

ECS055-T

300
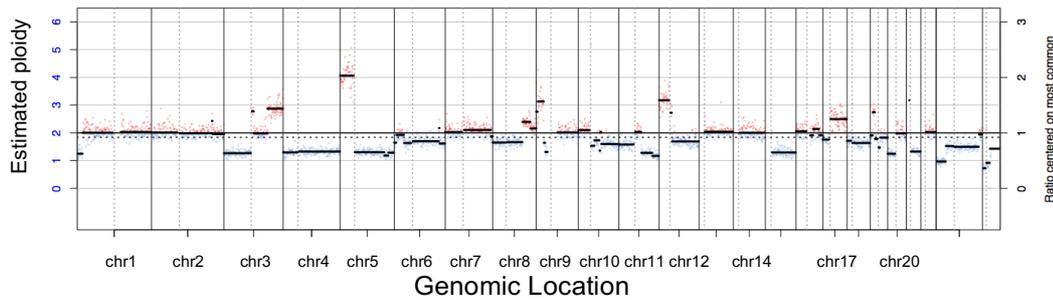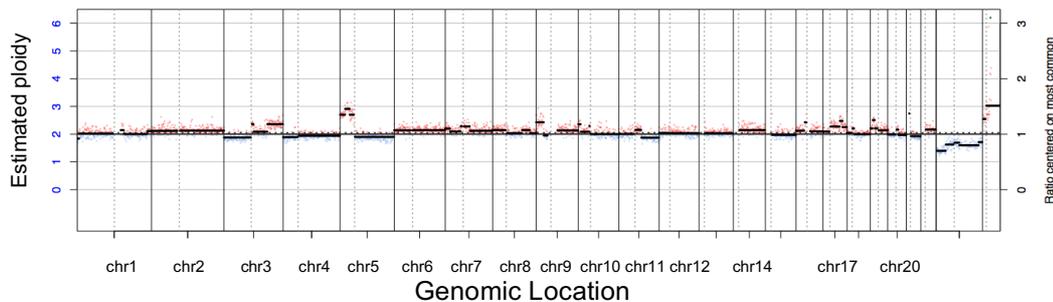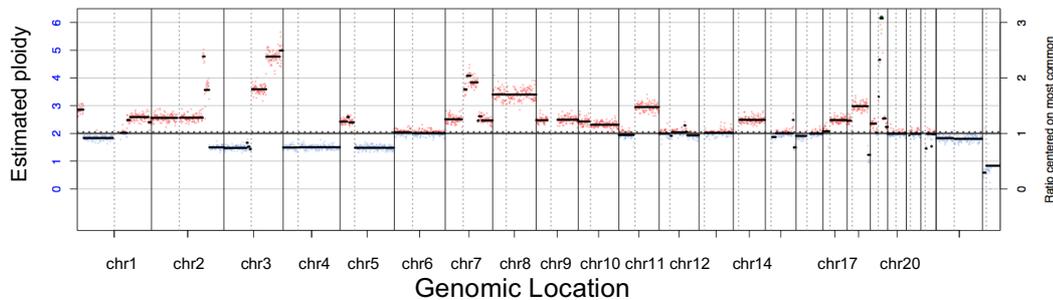
ECS055-L
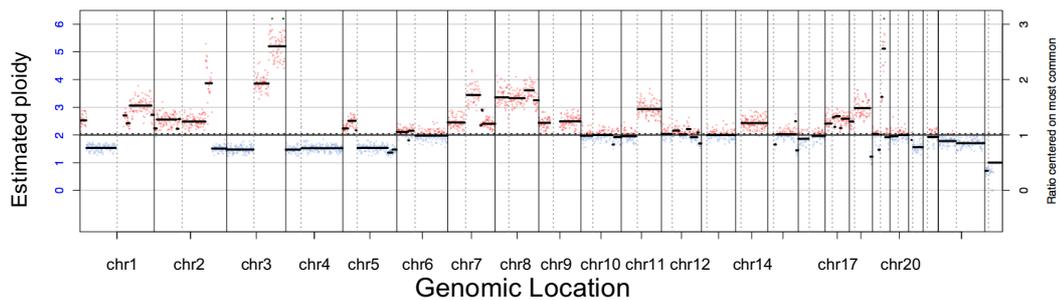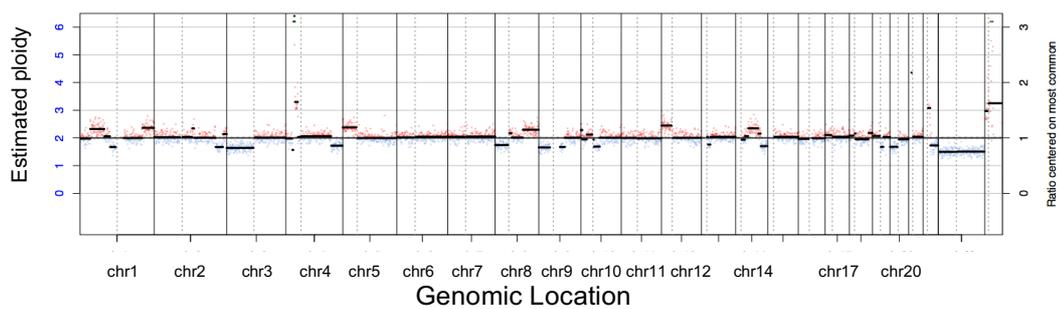
ECS056-T

ECS056-L

ECS057-T

ECS057-L

301
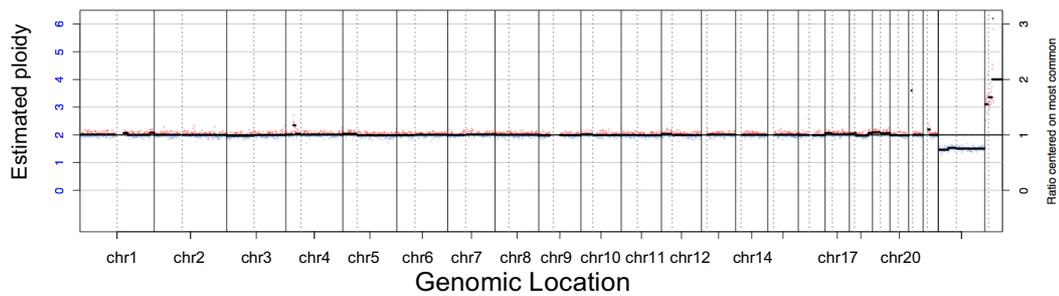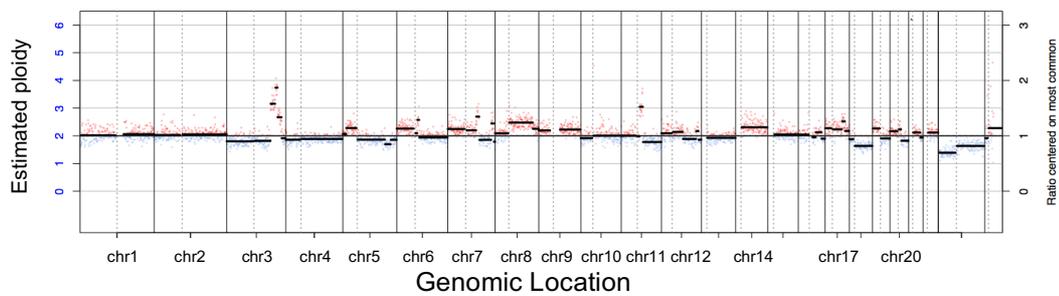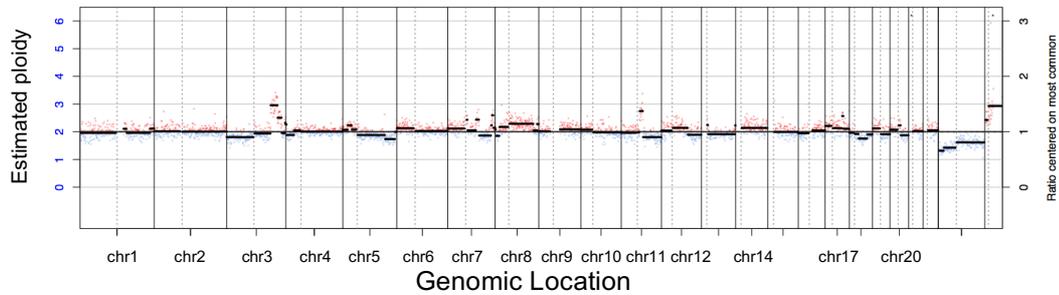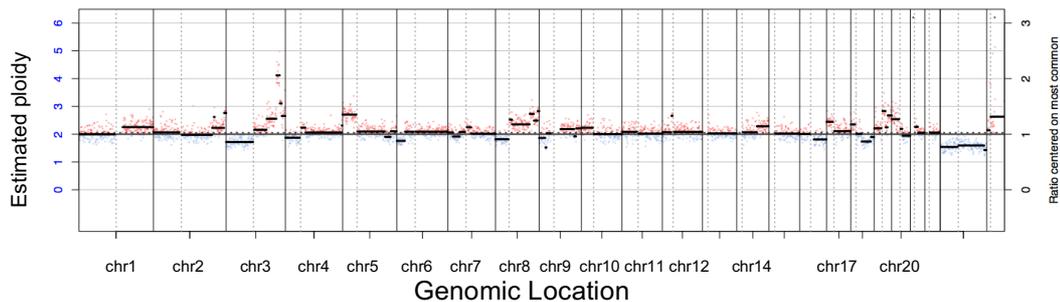
ECS059-T

ECS059-L
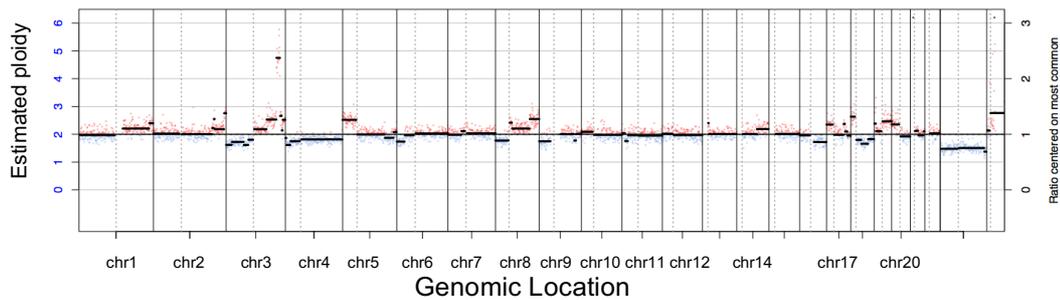
ECS060-T

ECS060-L

ECS061-T

302
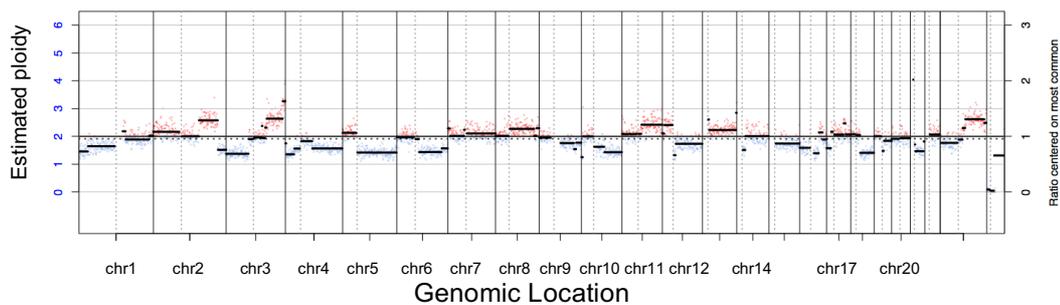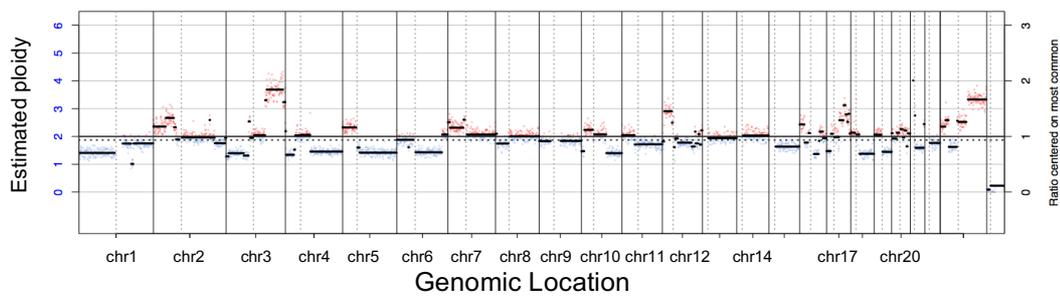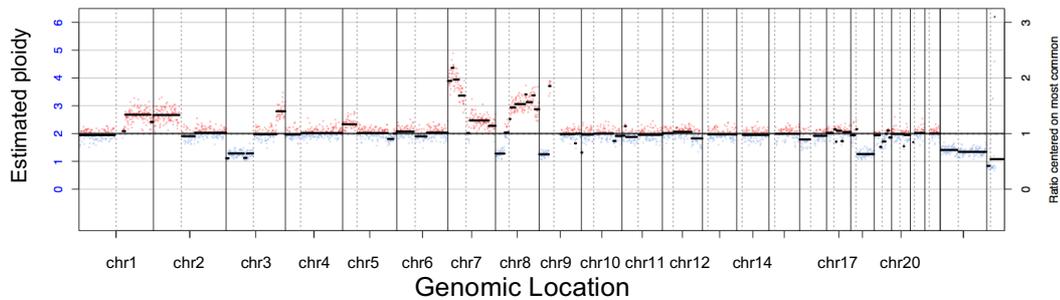
ECS061-L
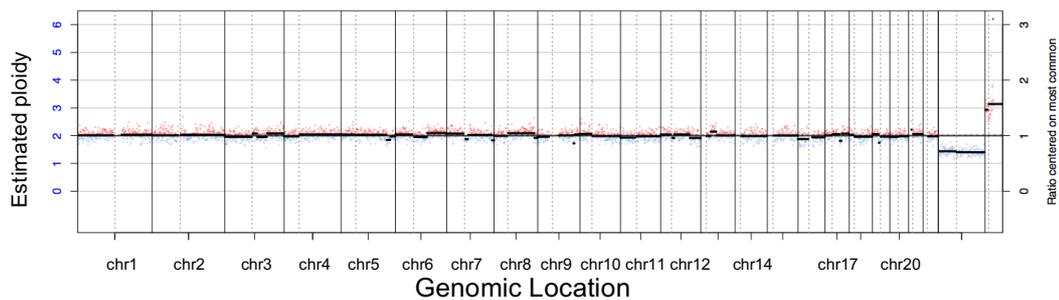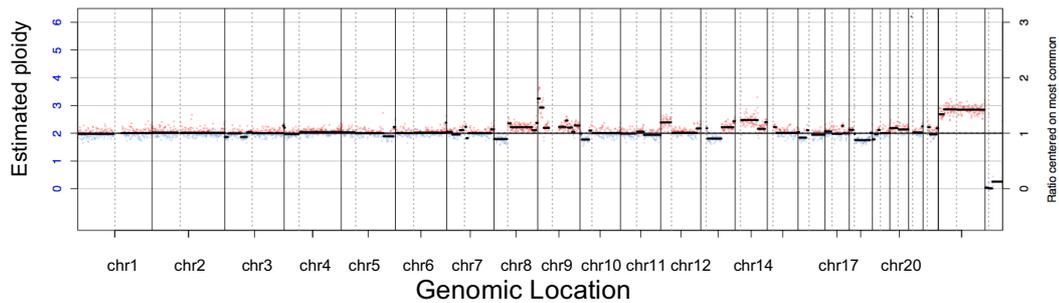
ECS062-T

ECS062-L

ECS063-T

ECS063-L

303
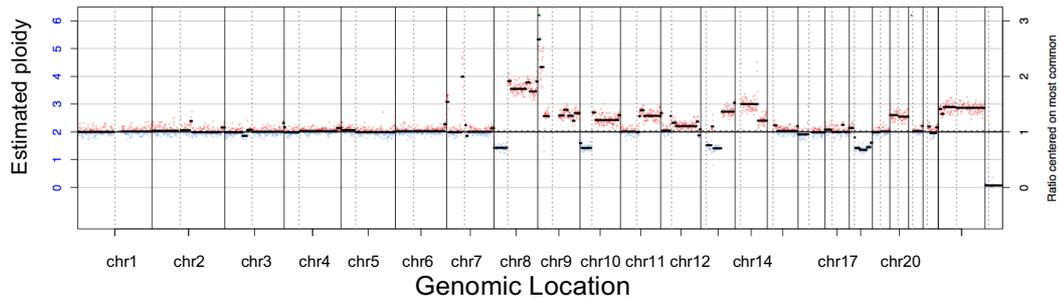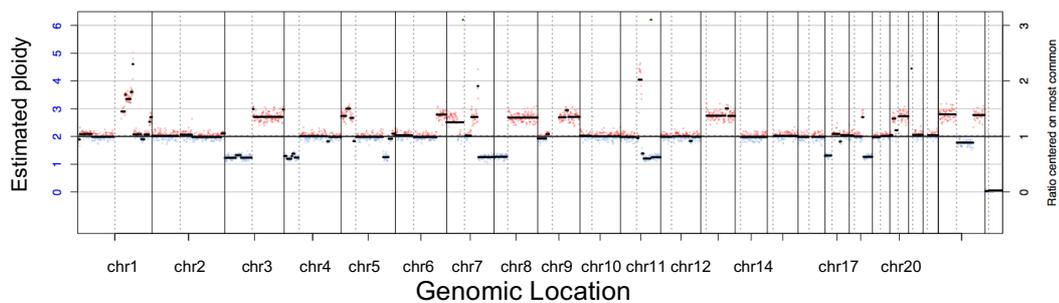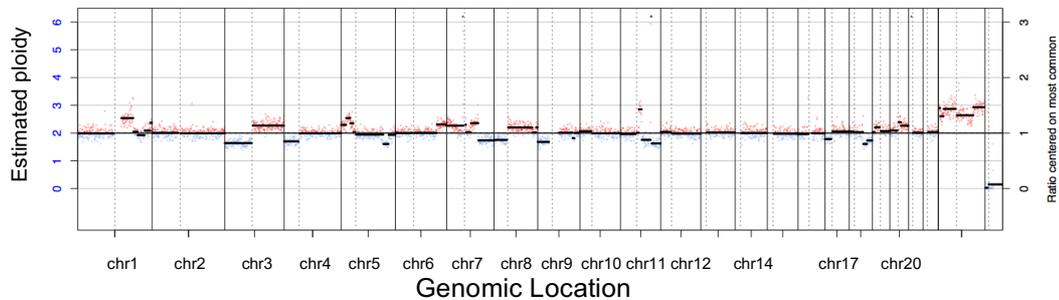
ECS064-T
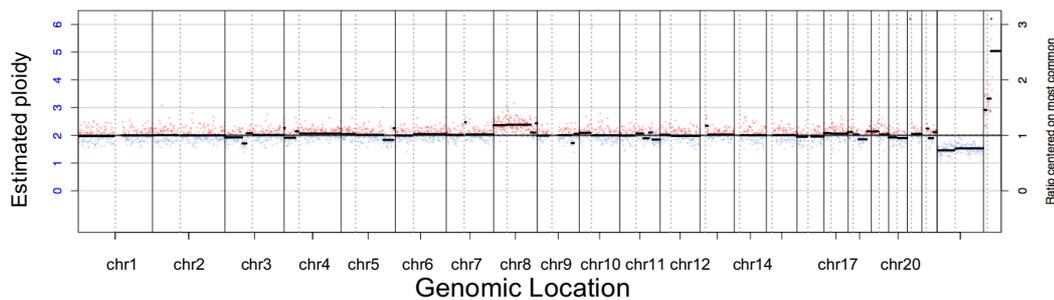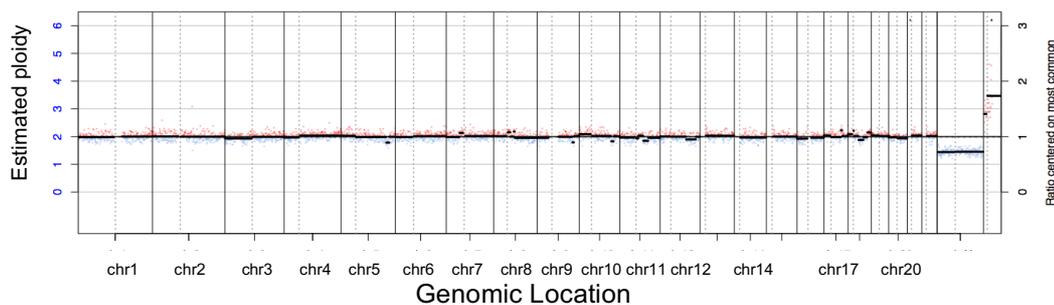
ECS065-T

ECS065-L

ECS066-T

ECS066-L

304
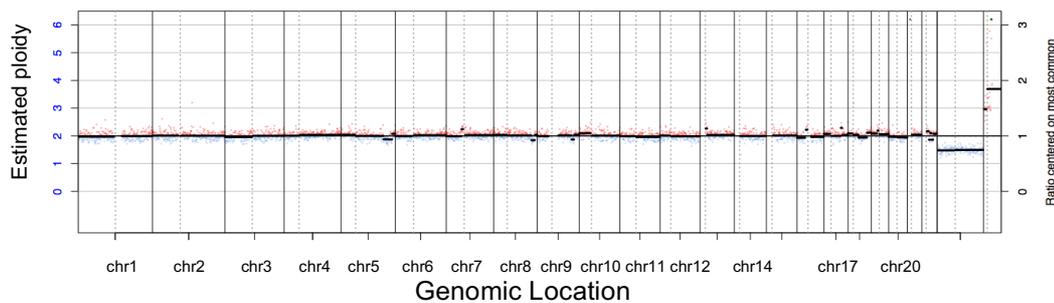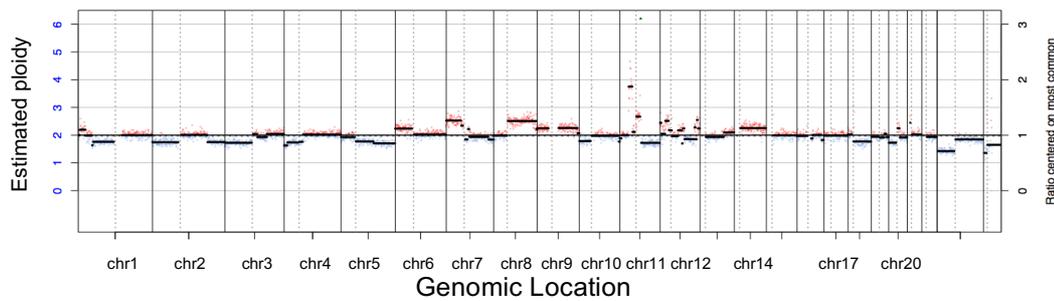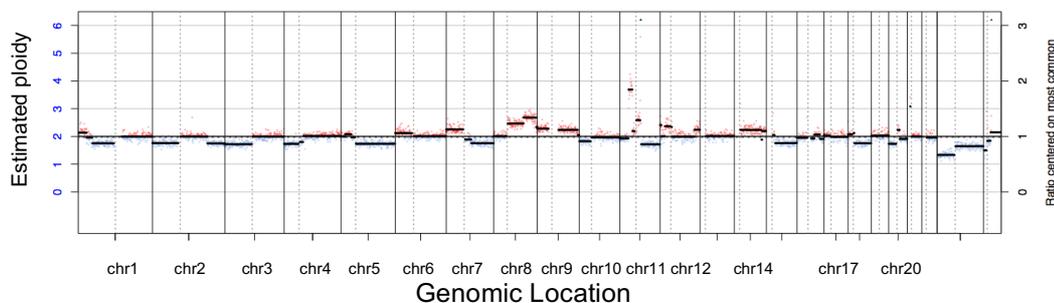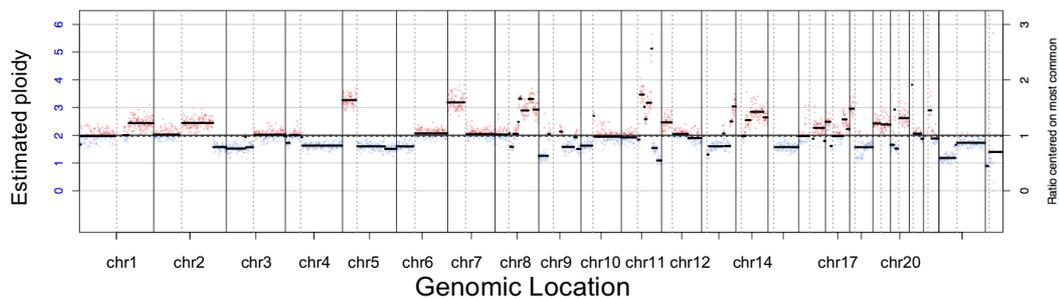
ECS067-T

ECS067-L
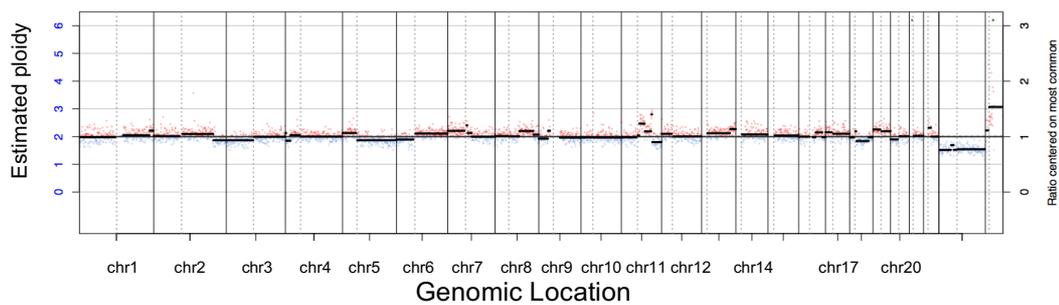
ECS068-T

ECS068-L

ECS069-T

305
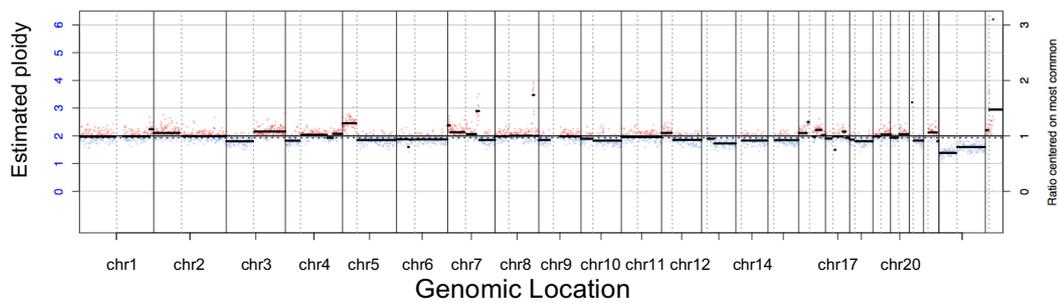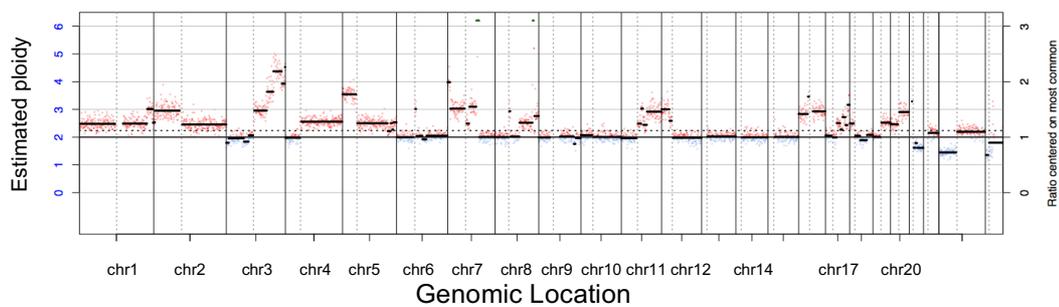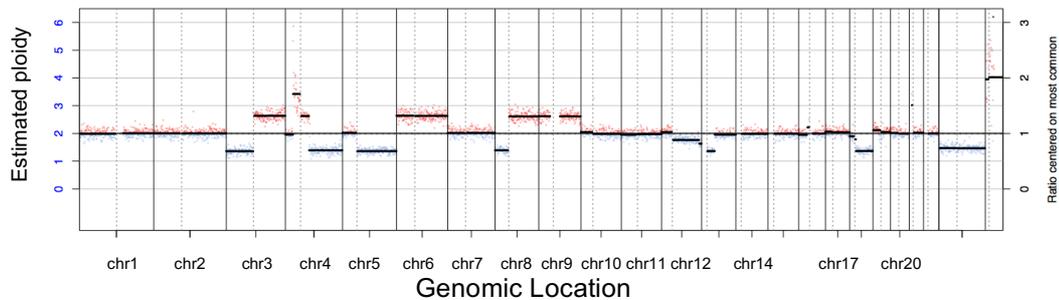
ECS069-L
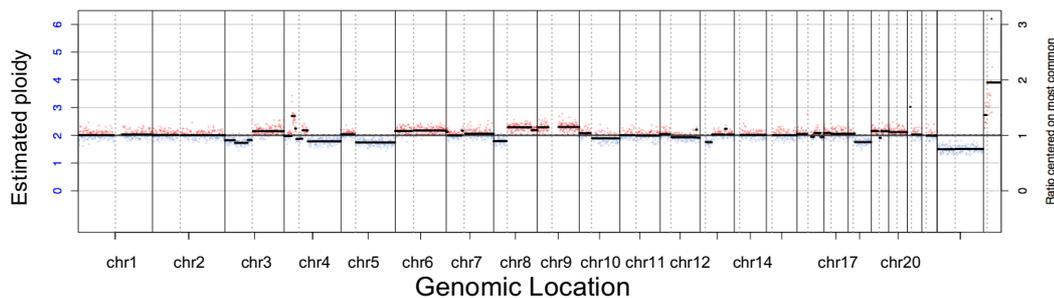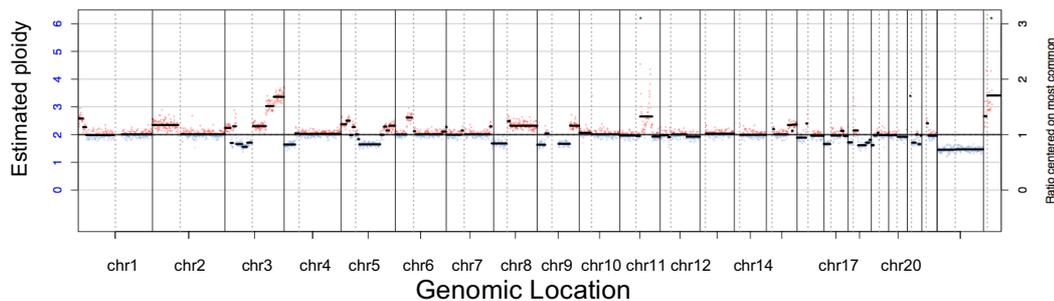
ECS070-T

ECS070-L

ECS071-T

ECS071-L

306
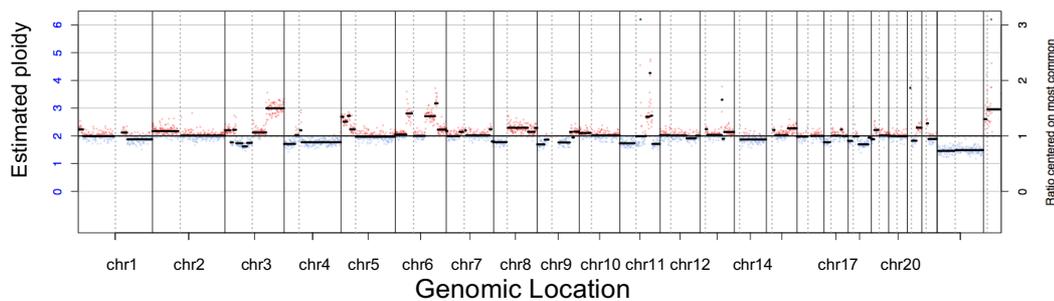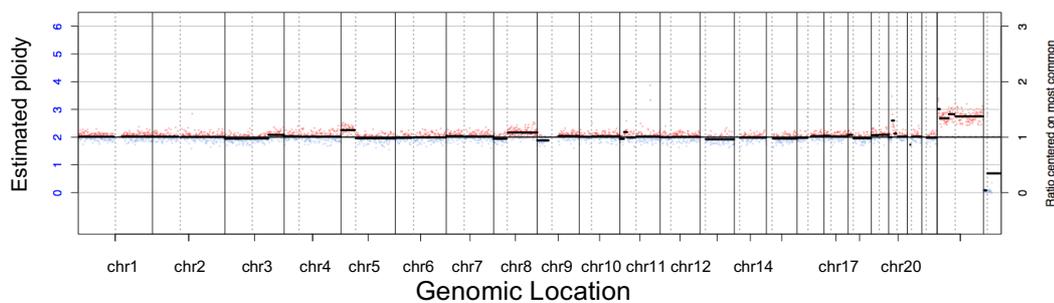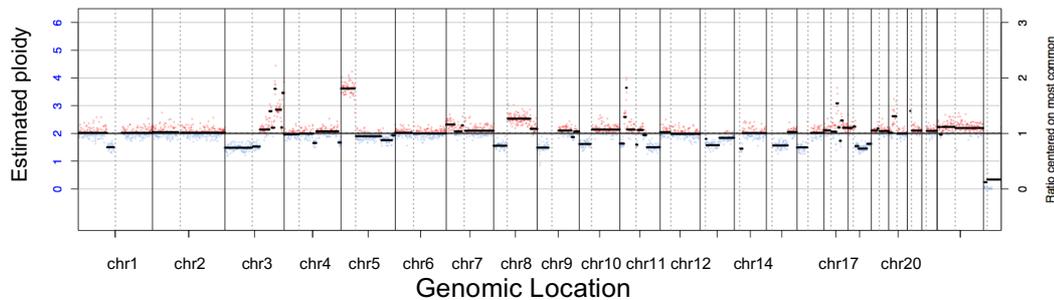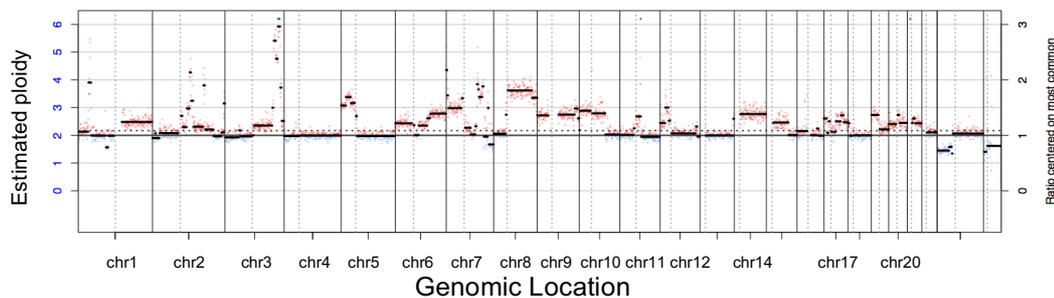
ECS073-T

ECS073-L

ECS075-L

ECS075-L

ECS076-T

307
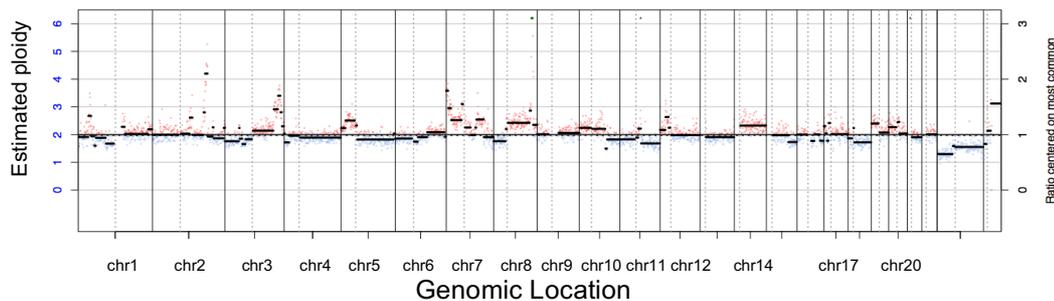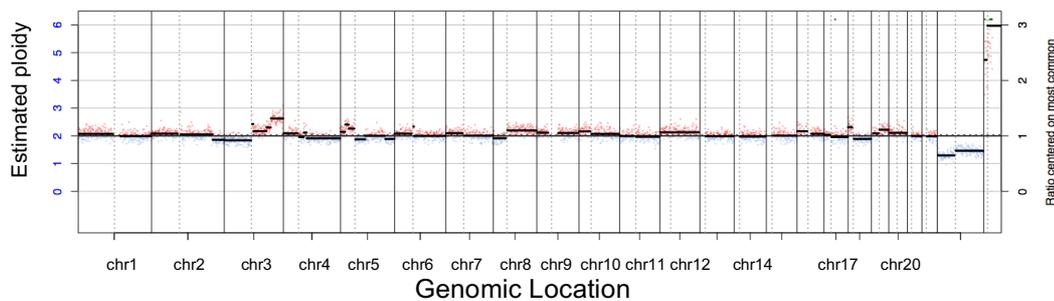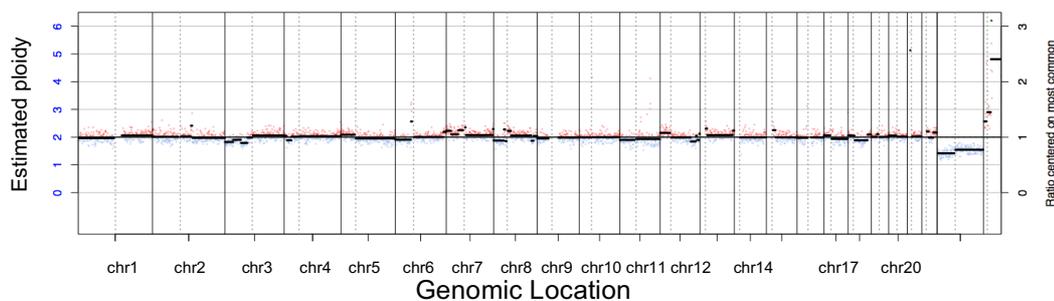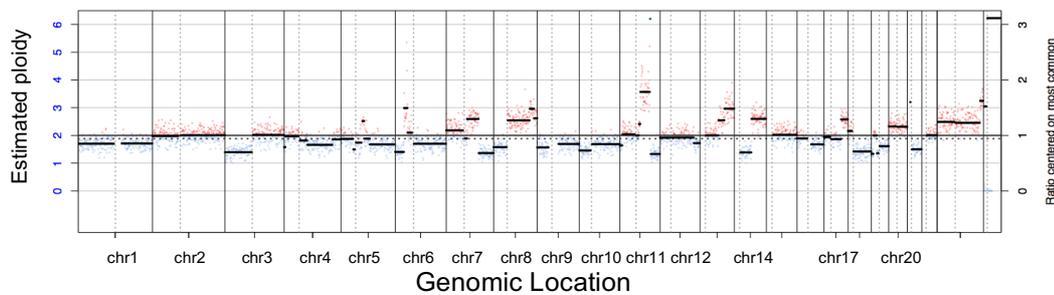
ECS076-L
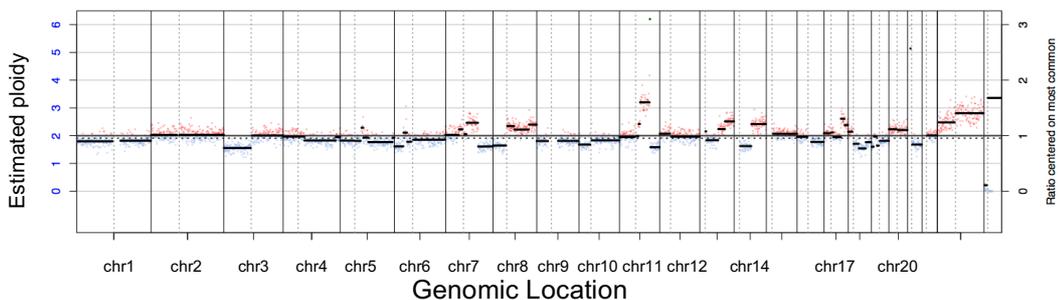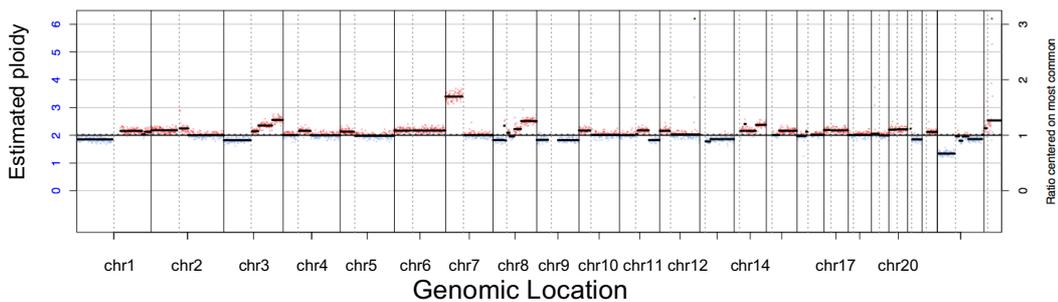
ECS078-T

ECS078-L

ECS079-T

ECS079-L

308
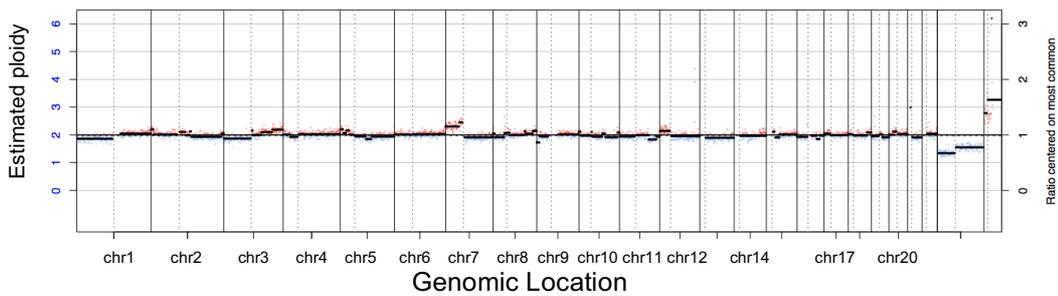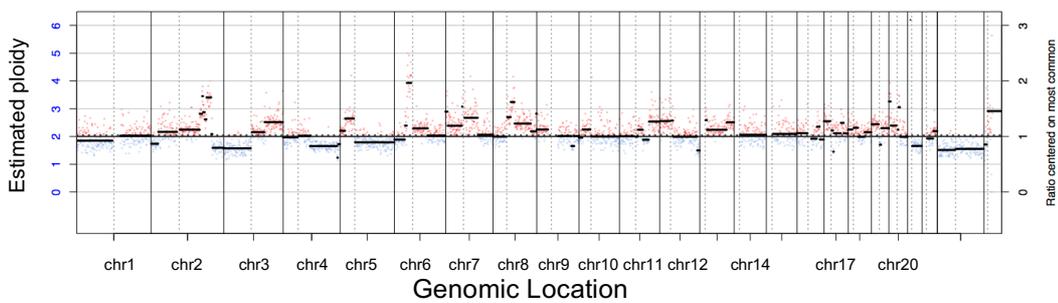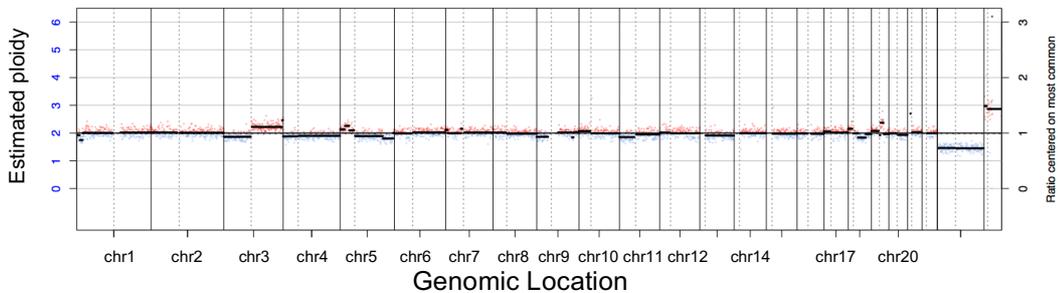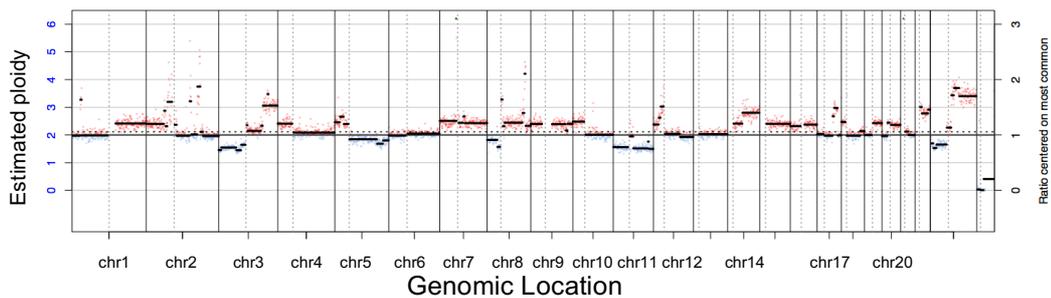
ECS080-T

ECS080-L
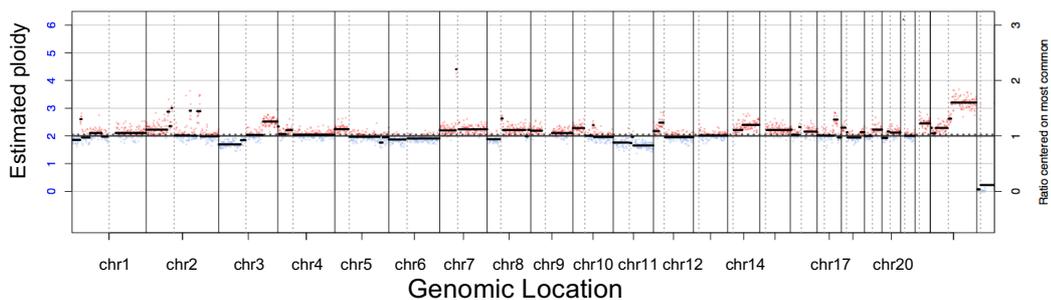
ECS081-T

ECS081-L

ECS082-T

309
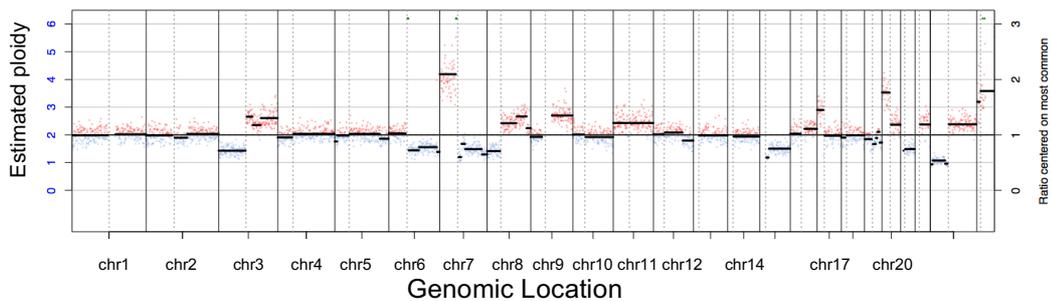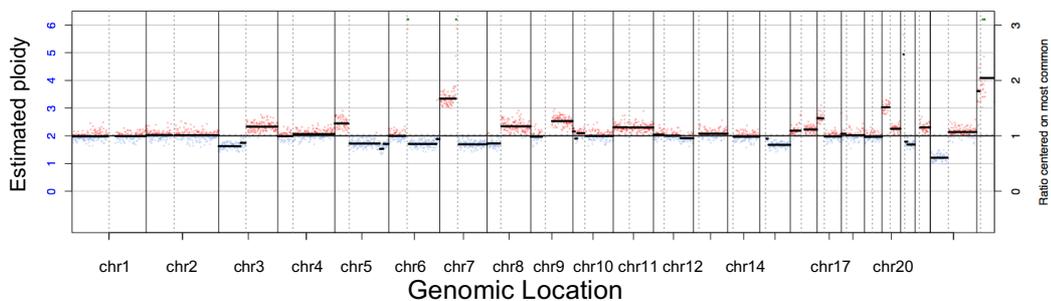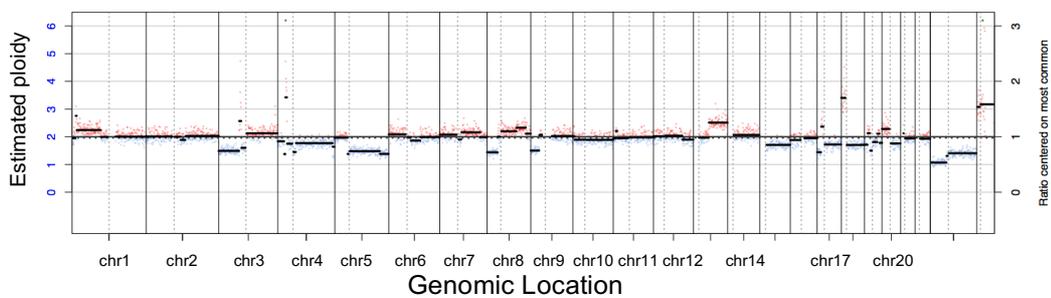
ECS082-L
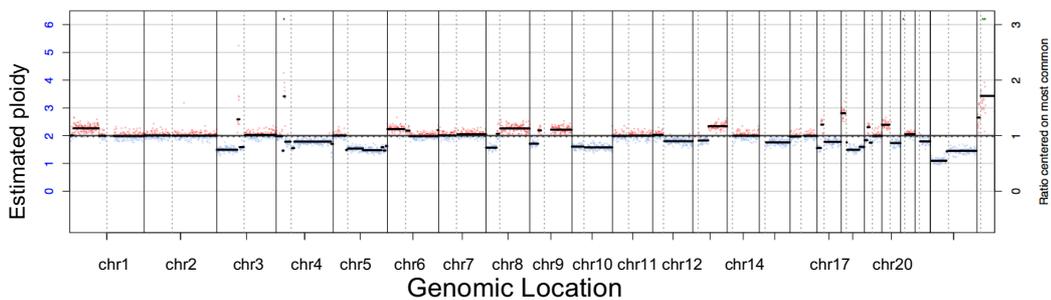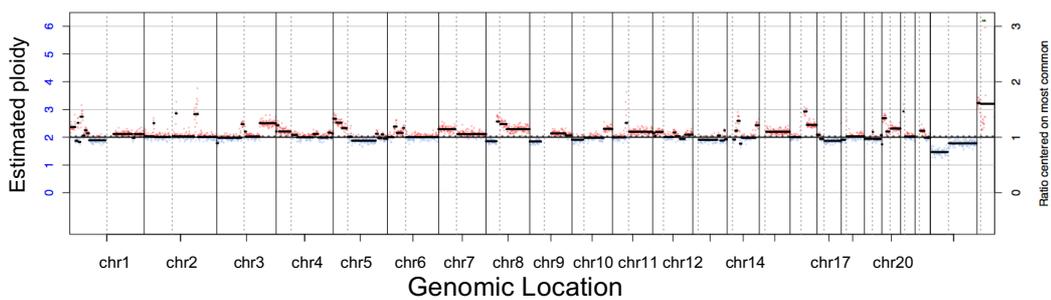
ECS083-T

ECS083-L
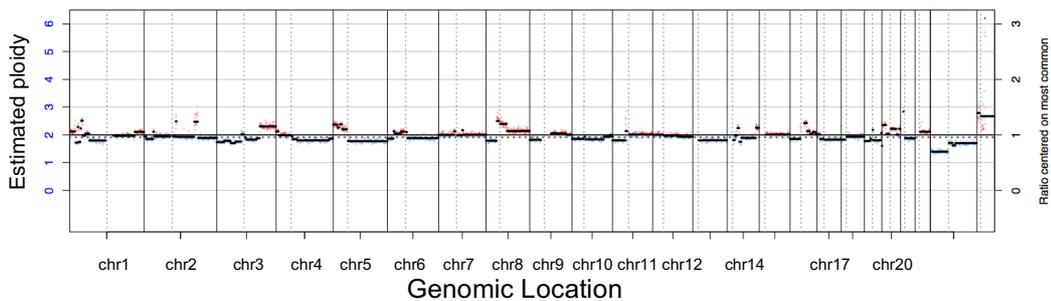
ECS084-T
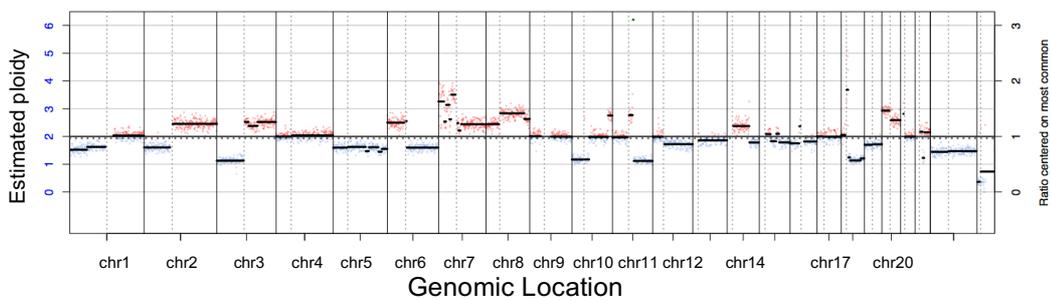
ECS084-L
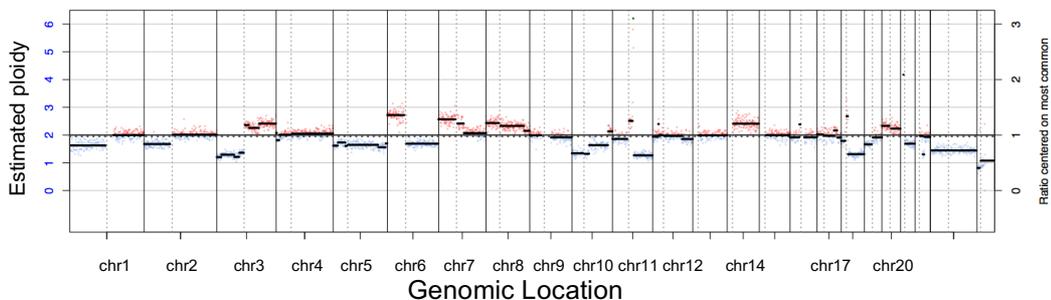
ECS085-T
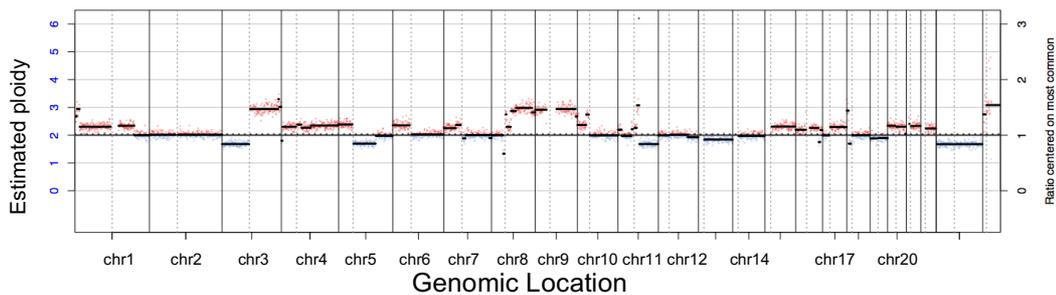
ECS085-L
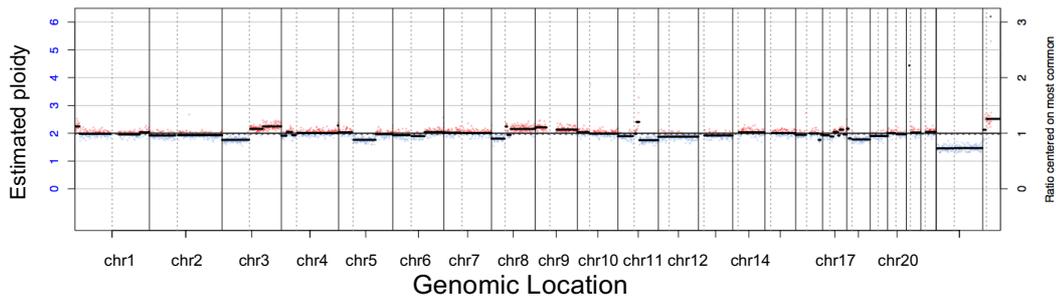
ECS087-T
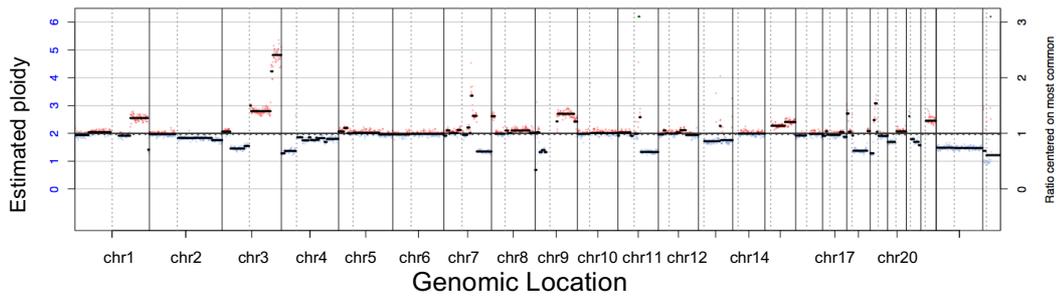
ECS087-L
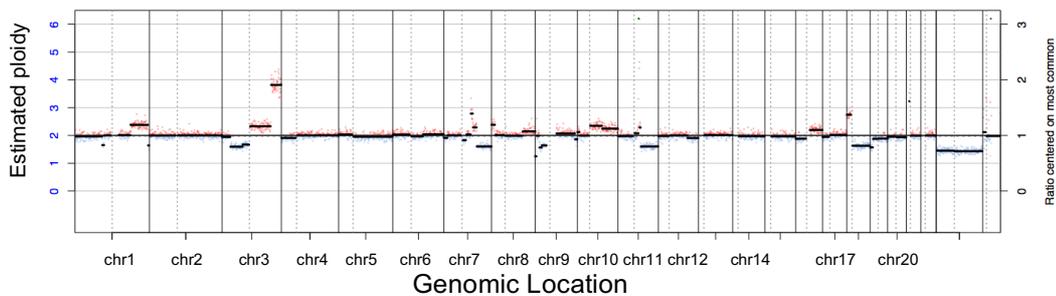
8.4 Scatterplots for miRNASeq vs. nCounter miRNA expression levels
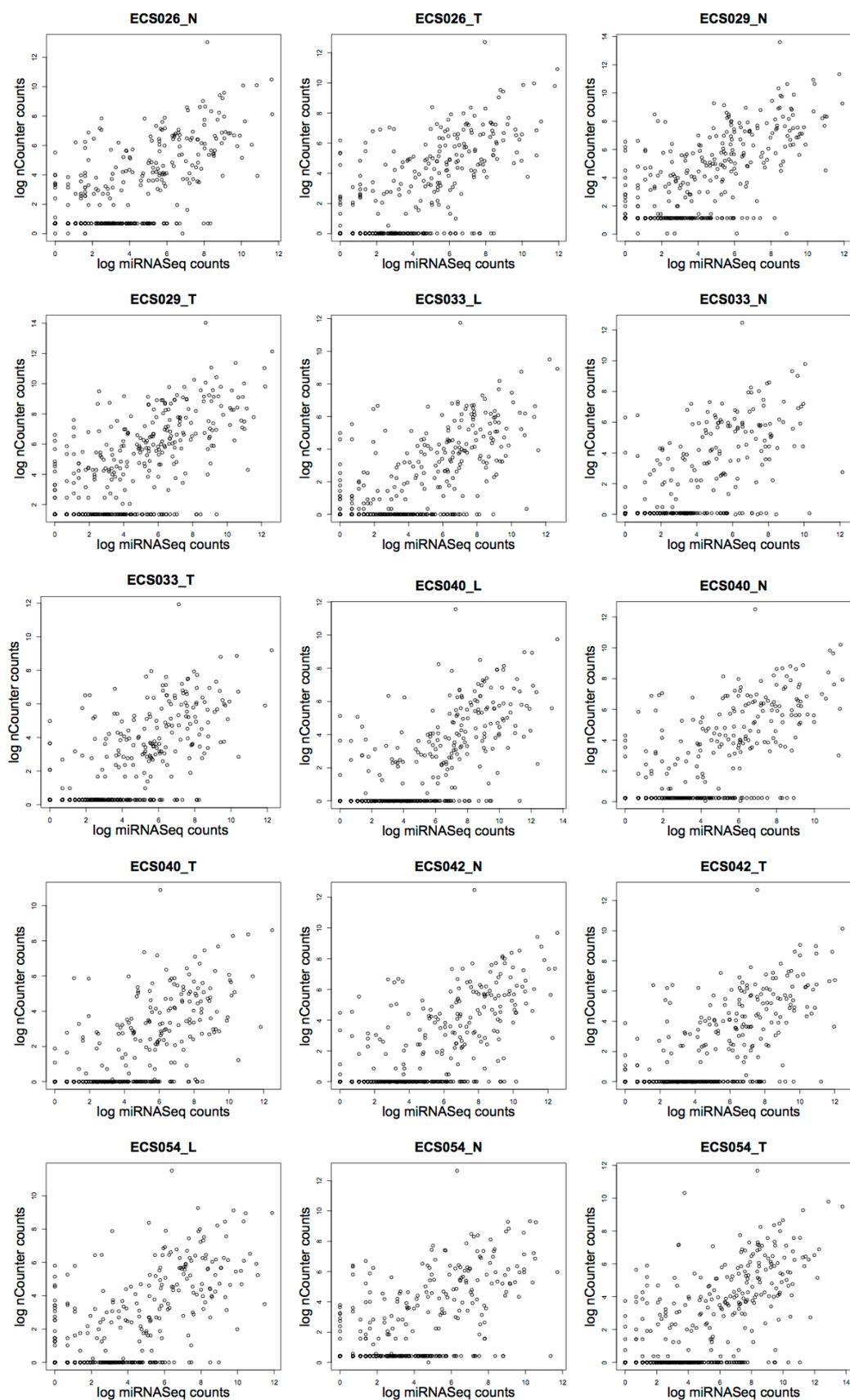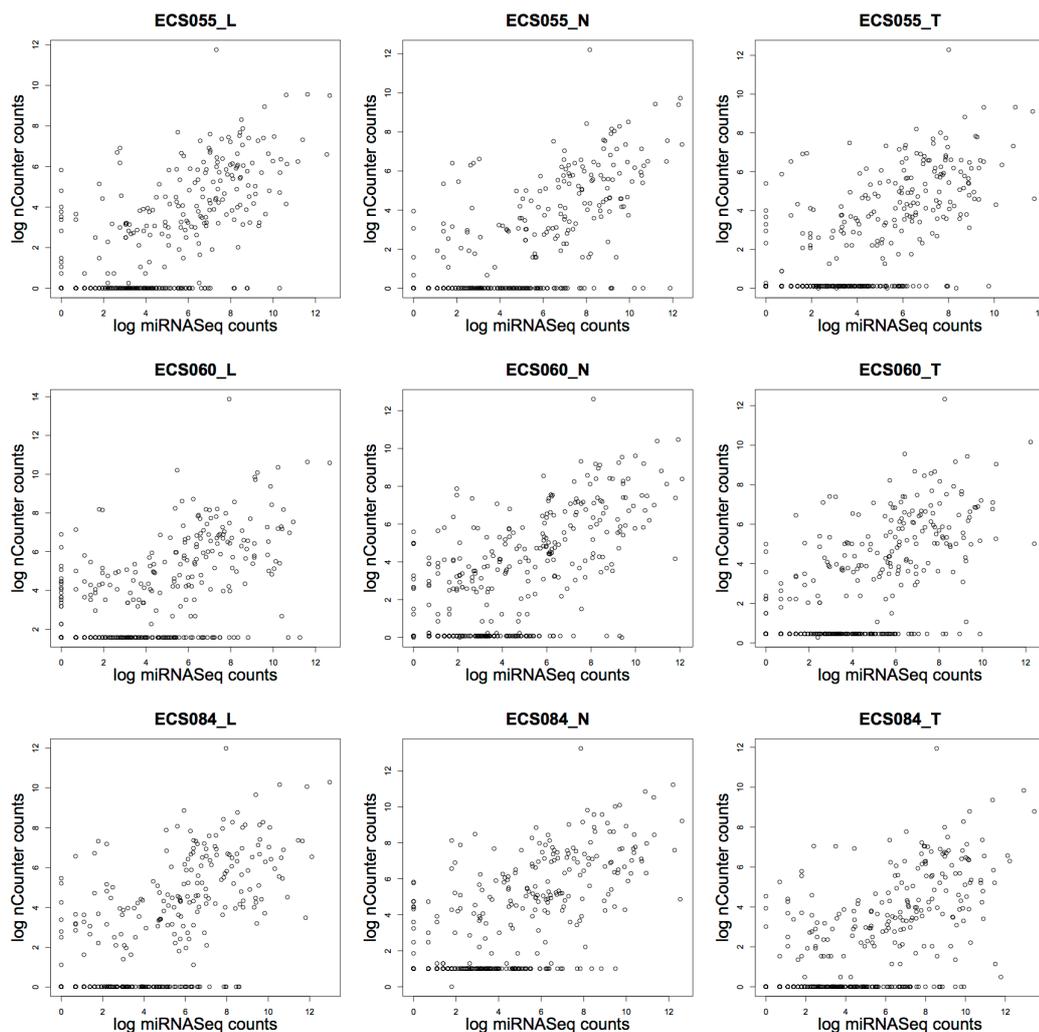
8.5 miRNA Differential expression tables

miRNASeq			nCounter		
Mature miRNA	**log Fold Change**	**Adjusted p-value**	**Mature miRNA**	**Log Fold Change**	**Adjusted p-value**
hsa-miR-375	8.83	1.72E-28	hsa-miR-206	7.47	1.85E-19
hsa-miR-196a-5p	-5.93	2.87E-22	hsa-miR-133a	6.64	3.34E-11
hsa-miR-133b	7.44	2.30E-16	hsa-miR-375	6.00	1.39E-09
hsa-miR-1	6.59	3.89E-15	hsa-miR-424-5p	-4.89	6.24E-09
hsa-miR-615-3p	-7.84	1.06E-14	hsa-miR-21-5p	-3.36	1.16E-06
hsa-miR-133a-3p	7.15	1.22E-14	hsa-miR-99a-5p	2.78	3.51E-06
hsa-miR-196b-5p	-4.58	3.28E-14	hsa-miR-503	-4.47	7.03E-06
hsa-miR-135a-5p	6.24	3.28E-14	hsa-miR-455-5p	-4.37	1.25E-05
hsa-miR-206	7.77	1.89E-11	hsa-miR-1246	-5.21	2.66E-05
hsa-let-7c-5p	3.36	2.72E-11	hsa-miR-455-3p	-4.37	7.61E-05
hsa-miR-6510-3p	3.37	5.79E-09	hsa-miR-421	-3.36	2.70E-04

hsa-miR-99a-3p	2.91	1.06E-08	hsa-miR-196b-5p	-3.89	3.33E-04
hsa-miR-99a-5p	3.39	1.83E-08	hsa-miR-944	-3.58	1.05E-03
hsa-miR-1247-5p	2.74	5.24E-08	hsa-miR-31-5p	-3.15	2.28E-03
hsa-miR-211-5p	6.11	1.28E-07	hsa-miR-196a-5p	-3.47	2.63E-03
hsa-miR-125b-2-3p	2.55	1.24E-06	hsa-let-7c	1.63	4.05E-03
hsa-miR-139-5p	2.48	2.03E-06	hsa-miR-146b-5p	-4.75	4.05E-03
hsa-miR-208b-3p	6.53	2.15E-06	hsa-miR-28-5p	-2.50	5.44E-03
hsa-miR-21-5p	-2.46	2.32E-06	hsa-miR-183-5p	-2.55	6.76E-03
hsa-miR-378i	2.56	2.73E-06	hsa-miR-125b-5p	1.82	7.96E-03
hsa-miR-3168	9.60	1.47E-05	hsa-miR-1	3.14	9.52E-03
hsa-miR-1246	-3.40	1.84E-05	hsa-miR-155-5p	-2.33	1.07E-02
hsa-miR-133a-5p	6.67	3.95E-05	hsa-miR-23b-3p	1.81	1.28E-02
hsa-miR-1269a	-5.98	3.95E-05	hsa-miR-18a-5p	-3.00	1.53E-02
hsa-miR-450b-5p	-2.37	1.20E-04	hsa-miR-7-5p	-2.34	1.70E-02
hsa-miR-4776-5p	5.52	1.97E-04	hsa-miR-450a-5p	-3.06	1.84E-02
hsa-miR-378c	1.94	2.11E-04	hsa-miR-203	2.80	2.23E-02
hsa-miR-125b-5p	2.05	2.25E-04	hsa-miR-135b-5p	-1.50	2.81E-02
hsa-miR-196a-3p	-5.39	4.59E-04	hsa-miR-182-5p	-2.51	4.48E-02
hsa-miR-31-5p	-2.52	4.59E-04	hsa-miR-378a-3p	1.76	4.69E-02
hsa-miR-455-3p	-2.13	5.74E-04	hsa-miR-378i	1.76	4.69E-02
hsa-miR-6842-3p	-4.36	6.22E-04	hsa-miR-133b	2.55	4.81E-02
hsa-miR-561-5p	3.32	7.33E-04	hsa-miR-125a-5p	-1.31	4.91E-02
hsa-miR-1910-5p	-3.34	7.85E-04	hsa-miR-130b-3p	-2.29	4.91E-02
hsa-miR-204-5p	3.41	1.00E-03	hsa-miR-193b-3p	-1.99	5.52E-02
hsa-miR-21-3p	-3.02	1.13E-03	hsa-miR-630	-2.12	9.00E-02
hsa-miR-378a-3p	1.82	1.13E-03	hsa-miR-100-5p	1.42	9.21E-02
hsa-miR-7705	-2.54	1.23E-03	hsa-miR-181b-5p	-3.51	9.23E-02
hsa-miR-378f	2.54	1.28E-03	hsa-miR-181d	-3.51	9.23E-02
hsa-miR-1251-5p	5.05	1.40E-03	hsa-miR-132-3p	-2.23	1.06E-01
hsa-miR-424-3p	-2.43	1.61E-03	hsa-miR-204-5p	3.21	1.26E-01
hsa-miR-10a-3p	-2.60	1.61E-03	hsa-miR-1180	-2.75	1.34E-01
hsa-let-7c-3p	2.77	1.63E-03	hsa-miR-663b	-2.75	1.44E-01
hsa-miR-1247-3p	3.03	2.38E-03	hsa-miR-29c-3p	1.24	1.56E-01
hsa-miR-100-5p	2.08	2.54E-03	hsa-miR-376c	2.15	2.33E-01
hsa-miR-424-5p	-1.79	2.82E-03			
hsa-miR-301b	-2.09	3.30E-03			
hsa-miR-30a-5p	1.62	3.30E-03			
hsa-miR-136-3p	2.42	3.35E-03			
hsa-miR-31-3p	-2.67	3.48E-03			
hsa-miR-203a	3.21	3.48E-03			
hsa-miR-203b-5p	4.34	3.86E-03			
hsa-miR-381-3p	3.42	4.13E-03			
hsa-miR-3129-3p	-2.91	4.71E-03			
hsa-miR-378a-5p	1.77	5.41E-03			
hsa-miR-25-5p	-2.07	5.78E-03			

hsa-miR-455-5p	-1.64	8.23E-03			
hsa-miR-146b-5p	-1.97	1.10E-02			
hsa-miR-454-3p	-1.44	1.18E-02			
hsa-miR-135b-5p	-1.68	1.47E-02			
hsa-miR-92b-3p	-1.50	1.65E-02			
hsa-miR-4485	1.85	1.75E-02			
hsa-miR-2355-5p	-2.27	2.05E-02			
hsa-miR-30a-3p	1.54	2.17E-02			
hsa-miR-23b-3p	1.40	2.22E-02			
hsa-miR-146b-3p	-1.48	2.22E-02			
hsa-miR-450a-5p	-1.78	2.64E-02			
hsa-miR-301a-3p	-1.43	2.79E-02			
hsa-miR-10a-5p	-1.88	3.11E-02			
hsa-miR-338-5p	1.72	3.38E-02			
hsa-miR-503-5p	-2.67	3.51E-02			
hsa-miR-940	-2.08	3.99E-02			
hsa-miR-1271-3p	-4.00	4.34E-02			
hsa-miR-378d	1.66	4.64E-02			
hsa-miR-4510	2.78	4.64E-02			
hsa-miR-183-3p	-1.92	4.67E-02			
hsa-miR-615-5p	-4.08	4.67E-02			
hsa-miR-335-3p	-1.49	4.77E-02			
hsa-miR-96-5p	-1.46	5.12E-02			
hsa-miR-4705	3.21	5.22E-02			
hsa-miR-431-3p	2.52	5.44E-02			
hsa-miR-3065-3p	1.82	5.44E-02			
hsa-miR-5089-3p	3.64	6.12E-02			
hsa-miR-7-5p	-1.84	6.12E-02			
hsa-miR-3687	-2.60	6.12E-02			
hsa-miR-671-5p	-1.54	7.14E-02			
hsa-miR-1301-3p	-2.07	7.14E-02			
hsa-miR-4652-5p	-3.73	7.28E-02			
hsa-miR-421	-1.15	7.31E-02			
hsa-miR-101-3p	1.39	7.41E-02			
hsa-miR-27b-3p	1.19	7.44E-02			
hsa-miR-542-5p	-1.61	7.75E-02			
hsa-miR-149-5p	1.22	8.50E-02			
hsa-miR-1266-5p	3.41	8.87E-02			
hsa-miR-136-5p	2.05	9.14E-02			
hsa-miR-139-3p	2.65	9.64E-02			
hsa-miR-1290	-2.20	9.90E-02			
hsa-miR-203b-3p	1.86	1.02E-01			
hsa-miR-499a-5p	2.37	1.04E-01			
hsa-miR-454-5p	-1.54	1.15E-01			
hsa-miR-431-5p	-1.76	1.23E-01			

hsa-miR-130b-3p	-1.11	1.31E-01
hsa-miR-125b-1-3p	1.25	1.41E-01
hsa-miR-148a-5p	1.33	1.41E-01
hsa-miR-4662a-5p	-1.96	1.43E-01
hsa-miR-370-5p	-3.44	1.44E-01
hsa-miR-422a	1.85	1.45E-01
hsa-miR-4326	-1.56	1.45E-01
hsa-miR-4741	-3.61	1.58E-01
hsa-miR-598-3p	1.40	1.71E-01
hsa-miR-885-5p	2.25	1.72E-01
hsa-miR-4284	3.09	1.72E-01
hsa-miR-29c-5p	1.35	1.80E-01
hsa-miR-486-5p	1.43	1.85E-01
hsa-miR-570-3p	-2.80	1.86E-01
hsa-miR-487b-3p	1.70	1.86E-01
hsa-miR-3196	-2.36	1.86E-01
hsa-miR-3129-5p	-3.49	1.92E-01
hsa-miR-411-5p	1.85	1.93E-01
hsa-miR-331-5p	-1.38	2.00E-01
hsa-miR-22-5p	-1.19	2.07E-01
hsa-miR-519a-3p	-3.13	2.07E-01
hsa-miR-1908-5p	-2.72	2.07E-01
hsa-miR-193b-5p	-1.76	2.08E-01
hsa-miR-18a-3p	-1.40	2.10E-01
hsa-miR-1291	1.23	2.18E-01
hsa-miR-887-3p	1.36	2.22E-01
hsa-miR-199b-5p	1.36	2.28E-01
hsa-miR-30e-3p	1.04	2.36E-01
hsa-miR-515-5p	-3.24	2.37E-01
hsa-miR-10b-5p	1.31	2.39E-01

Table 8-3: Differentially expressed miRNAs between nodal metastases and matched normal epithelium.

miRNASeq			nCounter		
Mature miRNA	log Fold Change	Adjusted p-value	Mature miRNA	log Fold Change	Adjusted p-value
hsa-miR-1269a	12.81	1.94E-13	hsa-miR-221-3p	2.28	2.42E-05
hsa-miR-615-3p	8.54	2.88E-10	hsa-miR-4286	2.38	2.13E-04
hsa-miR-1269b	11.45	2.99E-08	hsa-miR-29c-3p	-2.05	8.67E-04
hsa-miR-1910-5p	4.23	4.06E-06	hsa-miR-424-5p	5.28	3.28E-03
hsa-miR-196a-5p	4.92	2.34E-05	hsa-miR-183-5p	3.75	9.83E-03
hsa-miR-431-5p	3.69	3.14E-04	hsa-miR-182-5p	3.70	1.16E-02
hsa-miR-29c-3p	-2.37	1.78E-03	hsa-miR-455-3p	3.96	1.23E-02
hsa-miR-4713-5p	4.02	1.78E-03	hsa-miR-375	-5.35	1.62E-02

hsa-miR-139-5p	-2.98	1.86E-03	hsa-miR-485-3p	4.05	1.86E-02
hsa-miR-4521	-3.10	4.05E-03	hsa-miR-199b-5p	-1.54	2.25E-02
hsa-miR-195-5p	-2.28	4.35E-03	hsa-miR-378a-3p	-2.12	2.53E-02
hsa-miR-183-5p	2.83	5.08E-03	hsa-miR-378i	-2.12	2.53E-02
hsa-miR-196b-5p	4.44	1.50E-02	hsa-miR-455-5p	3.68	4.37E-02
hsa-miR-182-5p	2.56	1.50E-02	hsa-miR-331-3p	3.49	4.37E-02
hsa-miR-3929	5.07	1.50E-02	hsa-miR-503	3.96	4.37E-02
hsa-miR-450a-5p	2.83	1.58E-02	hsa-miR-450a-5p	3.37	4.37E-02
hsa-miR-21-3p	2.55	1.74E-02	hsa-miR-1246	3.90	5.49E-02
hsa-miR-503-5p	3.19	1.85E-02	hsa-miR-139-5p	-3.74	6.70E-02
hsa-miR-1246	2.83	1.85E-02	hsa-miR-135b-5p	1.48	8.66E-02
hsa-miR-9-5p	2.69	2.54E-02	hsa-miR-1206	3.16	1.04E-01
hsa-miR-34c-5p	2.30	2.73E-02	hsa-miR-34c-5p	4.38	1.06E-01
hsa-miR-212-3p	2.85	2.73E-02	hsa-miR-96-5p	2.40	1.48E-01
hsa-miR-542-3p	2.62	2.73E-02	hsa-miR-21-5p	2.66	1.48E-01
hsa-miR-381-3p	-1.95	2.73E-02	hsa-miR-222-3p	1.28	1.48E-01
hsa-miR-1251-5p	-4.30	2.96E-02	hsa-miR-324-5p	2.21	1.74E-01
hsa-miR-212-5p	3.22	3.91E-02	hsa-miR-421	2.94	1.78E-01
hsa-miR-485-3p	2.31	4.11E-02	hsa-miR-144-3p	-1.25	1.95E-01
hsa-miR-101-5p	-2.03	4.25E-02	hsa-miR-145-5p	-1.55	1.95E-01
hsa-miR-301b	2.34	4.25E-02			
hsa-miR-424-3p	2.35	4.25E-02			
hsa-miR-497-5p	-2.11	4.48E-02			
hsa-miR-183-3p	2.23	4.48E-02			
hsa-miR-450b-5p	2.53	4.96E-02			
hsa-miR-34b-3p	2.79	5.58E-02			
hsa-miR-3607-3p	-2.04	5.70E-02			
hsa-miR-542-5p	2.08	6.02E-02			
hsa-miR-323a-3p	2.22	6.02E-02			
hsa-miR-34b-5p	1.93	7.41E-02			
hsa-miR-378i	-2.04	7.41E-02			
hsa-miR-378a-3p	-1.78	9.07E-02			
hsa-miR-6510-3p	-2.28	9.07E-02			
hsa-miR-92b-3p	1.64	9.36E-02			
hsa-miR-1197	3.39	9.57E-02			
hsa-miR-100-5p	-1.78	9.58E-02			
hsa-miR-4664-3p	3.74	9.58E-02			
hsa-miR-516a-5p	4.81	9.61E-02			
hsa-miR-767-5p	4.66	9.75E-02			
hsa-miR-375	-3.83	1.10E-01			
hsa-miR-101-3p	-1.58	1.23E-01			
hsa-miR-3065-5p	-1.83	1.23E-01			
hsa-miR-455-5p	1.57	1.33E-01			
hsa-miR-29b-2-5p	-2.29	1.33E-01			

hsa-miR-135a-5p	-5.67	1.35E-01			
hsa-miR-211-5p	-3.55	1.39E-01			
hsa-miR-937-3p	3.13	1.43E-01			
hsa-miR-378a-5p	-1.89	1.45E-01			
hsa-miR-758-3p	1.93	1.47E-01			
hsa-miR-380-3p	2.91	1.59E-01			
hsa-miR-424-5p	1.96	1.59E-01			
hsa-miR-548i	4.91	1.59E-01			
hsa-miR-942-5p	-2.29	1.61E-01			
hsa-miR-204-5p	-3.83	1.65E-01			
hsa-miR-105-3p	4.17	1.65E-01			
hsa-miR-3938	4.85	1.77E-01			
hsa-miR-199b-5p	-1.41	1.98E-01			
hsa-miR-1247-5p	-1.75	1.98E-01			
hsa-miR-512-3p	3.35	2.23E-01			
hsa-miR-4443	4.03	2.35E-01			
hsa-miR-338-5p	-1.75	2.42E-01			

Table 8-4: Differentially expressed miRNAs between non-metastatic primary tumours and matched normal epithelium.

miRNASeq			nCounter		
Mature miRNA	log Fold Change	Adjusted p-value	Mature miRNA	Log Fold Change	Adjusted p-value
hsa-miR-375	-8.03	5.40E-36	hsa-miR-375	-6.29	2.38E-14
hsa-miR-196b-5p	5.27	2.75E-14	hsa-miR-424-5p	6.43	4.60E-14
hsa-let-7c-5p	-3.45	1.90E-13	hsa-miR-1180	5.09	5.67E-10
hsa-miR-615-3p	7.54	2.76E-12	hsa-miR-421	4.71	1.12E-09
hsa-miR-196a-5p	5.44	4.71E-11	hsa-miR-503	5.75	3.02E-09
hsa-miR-1247-5p	-3.54	7.82E-08	hsa-miR-21-5p	3.04	7.16E-09
hsa-miR-99a-5p	-3.10	1.22E-07	hsa-miR-196b-5p	4.87	1.08E-08
hsa-miR-1910-5p	4.23	3.47E-07	hsa-miR-663b	4.93	6.32E-07
hsa-miR-150-5p	-2.51	9.81E-07	hsa-miR-125b-5p	-2.13	1.11E-06
hsa-miR-125b-5p	-2.42	9.81E-07	hsa-miR-130b-3p	4.24	1.65E-06
hsa-miR-424-3p	3.08	9.81E-07	hsa-miR-455-5p	4.57	1.71E-06
hsa-miR-125b-2-3p	-3.15	1.24E-06	hsa-miR-4286	3.16	1.71E-06
hsa-miR-503-5p	4.43	6.50E-06	hsa-miR-99a-5p	-2.67	2.04E-06
hsa-miR-99a-3p	-2.94	9.11E-06	hsa-miR-944	4.10	4.10E-06
hsa-miR-1269a	6.81	9.11E-06	hsa-miR-196a-5p	4.29	1.59E-05
hsa-miR-424-5p	2.74	9.11E-06	hsa-miR-146b-5p	3.82	3.92E-05
hsa-miR-135a-5p	-5.02	9.85E-06	hsa-miR-18a-5p	4.09	3.98E-05
hsa-miR-1246	3.69	1.38E-05	hsa-miR-1246	4.33	5.34E-05
hsa-miR-211-5p	-5.66	1.44E-05	hsa-miR-450a-5p	3.71	2.26E-04
hsa-miR-450b-5p	2.84	2.32E-05	hsa-miR-23b-3p	-2.04	3.71E-04
hsa-miR-204-5p	-4.16	3.75E-05	hsa-miR-203	-4.43	8.59E-04

hsa-miR-4776-5p	-5.60	4.88E-05	hsa-miR-150-5p	-2.18	2.49E-03
hsa-miR-450a-5p	2.54	6.68E-05	hsa-miR-135b-5p	1.89	5.65E-03
hsa-miR-6842-3p	4.79	6.68E-05	hsa-miR-455-3p	3.75	6.09E-03
hsa-miR-10b-5p	-2.38	8.36E-05	hsa-miR-708-5p	3.01	6.47E-03
hsa-miR-885-5p	-4.95	1.53E-04	hsa-let-7c	-1.45	1.18E-02
hsa-miR-31-3p	3.42	1.91E-04	hsa-miR-630	3.01	1.32E-02
hsa-miR-139-5p	-2.25	2.59E-04	hsa-miR-7-5p	2.71	1.32E-02
hsa-miR-21-3p	3.25	3.10E-04	hsa-miR-26b-5p	-1.52	1.69E-02
hsa-miR-21-5p	2.36	3.37E-04	hsa-miR-181b-5p	2.95	1.72E-02
hsa-miR-6510-3p	-3.59	5.04E-04	hsa-miR-181d	2.95	1.72E-02
hsa-let-7c-3p	-3.39	8.32E-04	hsa-miR-9-5p	3.29	1.72E-02
hsa-miR-203a	-4.57	8.60E-04	hsa-miR-151a-3p	1.67	2.54E-02
hsa-miR-18a-3p	2.44	8.60E-04	hsa-miR-93-5p	1.38	2.96E-02
hsa-miR-23b-3p	-1.77	9.84E-04	hsa-miR-31-5p	3.09	2.97E-02
hsa-miR-4705	-4.78	1.06E-03	hsa-miR-1915-3p	3.01	2.97E-02
hsa-miR-1468-5p	-2.36	1.61E-03	hsa-miR-28-5p	2.20	3.16E-02
hsa-miR-3196	3.87	2.26E-03	hsa-miR-204-5p	-2.90	3.16E-02
hsa-miR-455-5p	1.76	2.52E-03	hsa-miR-320e	2.37	3.96E-02
hsa-miR-6723-5p	5.12	2.52E-03	hsa-miR-4488	1.76	4.91E-02
hsa-miR-31-5p	2.50	2.88E-03	hsa-miR-199b-5p	-1.46	5.18E-02
hsa-miR-7705	2.48	3.68E-03	hsa-miR-125a-5p	1.25	5.18E-02
hsa-miR-2355-5p	2.52	4.65E-03	hsa-miR-664-3p	-1.74	5.18E-02
hsa-miR-934	5.13	5.11E-03	hsa-miR-497-5p	-2.78	5.18E-02
hsa-miR-25-5p	2.19	5.48E-03	hsa-miR-195-5p	-1.64	5.24E-02
hsa-miR-15b-3p	2.19	5.48E-03	hsa-miR-4508	2.95	6.21E-02
hsa-miR-671-5p	2.05	6.04E-03	hsa-miR-139-5p	-2.27	6.21E-02
hsa-miR-1268b	2.44	6.45E-03	hsa-miR-30c-5p	-1.96	6.27E-02
hsa-miR-3168	-7.45	6.45E-03	hsa-miR-34c-5p	2.80	6.60E-02
hsa-miR-3176	2.20	6.45E-03	hsa-miR-183-5p	2.24	6.60E-02
hsa-miR-196a-3p	4.90	6.84E-03	hsa-miR-96-5p	2.27	6.92E-02
hsa-miR-940	2.30	7.03E-03	hsa-miR-451a	-1.61	7.48E-02
hsa-miR-3117-3p	2.61	7.03E-03	hsa-miR-185-5p	1.53	7.94E-02
hsa-miR-26a-5p	-1.66	7.26E-03	hsa-miR-25-3p	1.10	9.96E-02
hsa-miR-511-3p	2.29	7.26E-03	hsa-miR-518b	-2.11	1.13E-01
hsa-miR-339-5p	1.60	8.09E-03	hsa-miR-221-3p	1.39	1.25E-01
hsa-miR-708-5p	1.62	8.53E-03	hsa-let-7i-5p	1.14	1.25E-01
hsa-miR-542-3p	2.13	8.59E-03	hsa-miR-29c-3p	-1.18	1.30E-01
hsa-miR-125a-5p	-1.49	9.09E-03	hsa-miR-155-5p	1.74	1.33E-01
hsa-miR-431-5p	2.10	9.93E-03	hsa-miR-600	2.01	1.34E-01
hsa-miR-663b	3.98	1.06E-02	hsa-miR-342-3p	-1.11	1.35E-01
hsa-miR-1251-5p	-4.61	1.11E-02	hsa-miR-106b-5p	1.15	1.44E-01
hsa-miR-148a-3p	-1.55	1.15E-02	hsa-miR-19a-3p	1.23	1.48E-01
hsa-miR-130b-3p	1.52	1.16E-02	hsa-miR-29a-3p	1.00	1.48E-01
hsa-miR-1268a	2.38	1.16E-02	hsa-miR-187-3p	2.01	1.48E-01
hsa-miR-101-3p	-1.54	1.18E-02	hsa-miR-125a-3p	1.75	1.49E-01

hsa-miR-3677-3p	2.57	1.36E-02	hsa-miR-26a-5p	-0.95	1.57E-01
hsa-miR-455-3p	1.76	1.36E-02	hsa-miR-130a-3p	1.04	1.58E-01
hsa-miR-3691-5p	3.40	1.36E-02	hsa-miR-331-3p	2.48	1.59E-01
hsa-miR-129-5p	2.69	1.60E-02	hsa-miR-22-3p	0.99	1.69E-01
hsa-miR-1247-3p	-2.97	1.99E-02	hsa-miR-520f	1.54	1.71E-01
hsa-miR-1301-3p	2.15	2.25E-02	hsa-miR-106a-5p	0.91	1.81E-01
hsa-miR-561-5p	-2.07	2.51E-02	hsa-miR-17-5p	0.91	1.81E-01
hsa-miR-7-5p	2.38	2.59E-02	hsa-miR-4532	2.07	1.81E-01
hsa-miR-18a-5p	1.60	2.72E-02	hsa-miR-135a-5p	-2.72	1.81E-01
hsa-miR-4485	-1.80	2.72E-02	hsa-miR-1234	1.68	1.82E-01
hsa-miR-106b-5p	1.26	2.91E-02	hsa-miR-450b-5p	1.50	1.85E-01
hsa-miR-2355-3p	2.50	2.99E-02	hsa-miR-542-3p	1.55	1.85E-01
hsa-miR-30a-5p	-1.36	3.09E-02	hsa-miR-1290	1.60	1.85E-01
hsa-miR-1307-3p	1.29	3.13E-02	hsa-miR-126-3p	-0.91	2.28E-01
hsa-miR-100-5p	-1.54	3.24E-02	hsa-miR-182-5p	2.06	2.32E-01
hsa-miR-27b-3p	-1.47	3.50E-02	hsa-miR-193b-3p	1.74	2.37E-01
hsa-miR-4510	-2.69	3.52E-02	hsa-miR-376c	-1.51	2.40E-01
hsa-miR-136-3p	-1.43	3.52E-02	hsa-miR-520h	1.51	2.40E-01
hsa-miR-4521	-2.54	3.69E-02			
hsa-miR-542-5p	1.79	3.75E-02			
hsa-miR-195-5p	-1.48	3.79E-02			
hsa-miR-29c-3p	-1.29	3.79E-02			
hsa-miR-92b-3p	1.25	3.79E-02			
hsa-miR-301b	2.51	3.85E-02			
hsa-miR-381-3p	-1.95	3.97E-02			
hsa-miR-148a-5p	-1.59	4.83E-02			
hsa-miR-1271-3p	3.99	4.84E-02			
hsa-miR-1296-5p	1.52	4.84E-02			
hsa-miR-96-5p	1.45	4.98E-02			
hsa-miR-30e-3p	-1.24	5.04E-02			
hsa-miR-4741	3.97	5.04E-02			
hsa-miR-19a-3p	1.48	5.75E-02			
hsa-miR-664a-3p	-1.51	5.75E-02			
hsa-miR-486-5p	-1.93	5.75E-02			
hsa-miR-22-5p	1.70	5.75E-02			
hsa-miR-6753-3p	3.90	5.85E-02			
hsa-miR-3619-5p	3.89	6.29E-02			
hsa-let-7a-5p	-1.24	6.37E-02			
hsa-miR-1287-5p	1.96	6.43E-02			
hsa-miR-1908-5p	2.79	6.53E-02			
hsa-miR-1180-3p	1.54	7.25E-02			
hsa-miR-550a-5p	2.11	7.25E-02			
hsa-miR-2467-5p	2.06	7.47E-02			
hsa-miR-3934-5p	1.77	7.70E-02			
hsa-miR-140-3p	-1.21	8.11E-02			

hsa-miR-520f-3p	3.69	8.40E-02			
hsa-miR-4707-3p	2.72	8.66E-02			
hsa-miR-345-3p	3.80	8.66E-02			
hsa-miR-615-5p	3.92	8.66E-02			
hsa-miR-1305	2.52	8.87E-02			
hsa-miR-877-3p	3.74	8.89E-02			
hsa-miR-3687	2.48	8.89E-02			
hsa-miR-139-3p	-3.21	8.89E-02			
hsa-miR-6854-5p	3.69	8.89E-02			
hsa-miR-3129-3p	2.28	8.89E-02			
hsa-miR-3162-3p	3.69	8.89E-02			
hsa-miR-937-3p	1.91	9.16E-02			
hsa-miR-4677-5p	3.39	9.16E-02			
hsa-miR-1273e	3.79	9.16E-02			
hsa-miR-550a-3-5p	2.07	9.21E-02			
hsa-miR-1307-5p	1.11	9.29E-02			
hsa-miR-378a-3p	-1.14	1.02E-01			
hsa-miR-449b-5p	2.75	1.04E-01			
hsa-miR-126-3p	-1.04	1.07E-01			
hsa-miR-454-3p	1.40	1.07E-01			
hsa-miR-5008-3p	2.68	1.07E-01			
hsa-miR-4652-5p	3.86	1.07E-01			
hsa-miR-675-5p	2.67	1.07E-01			
hsa-miR-16-1-3p	1.98	1.08E-01			
hsa-miR-548aq-5p	3.60	1.11E-01			
hsa-miR-378i	-1.36	1.11E-01			
hsa-miR-4516	2.24	1.13E-01			
hsa-miR-4746-5p	2.05	1.18E-01			
hsa-miR-758-3p	1.94	1.20E-01			
hsa-miR-4671-3p	2.78	1.26E-01			
hsa-miR-146b-5p	1.17	1.29E-01			
hsa-miR-26b-5p	-1.23	1.30E-01			
hsa-miR-619-5p	1.83	1.35E-01			
hsa-miR-570-3p	2.93	1.36E-01			
hsa-miR-3653	-2.32	1.40E-01			
hsa-miR-4798-5p	3.22	1.44E-01			
hsa-miR-135b-5p	1.77	1.44E-01			
hsa-miR-331-5p	1.39	1.44E-01			
hsa-miR-30a-3p	-1.10	1.53E-01			
hsa-miR-4443	2.27	1.55E-01			
hsa-miR-215-5p	-1.42	1.55E-01			
hsa-miR-199b-5p	-0.99	1.57E-01			
hsa-miR-335-3p	1.31	1.63E-01			
hsa-miR-128-3p	1.03	1.63E-01			
hsa-miR-4488	3.06	1.63E-01			

Mature miRNA	log Fold Change	Adjusted p-value
hsa-miR-5089-5p	-3.13	1.64E-01
hsa-miR-193b-5p	1.62	1.65E-01
hsa-miR-519a-3p	3.36	1.68E-01
hsa-miR-451a	-1.39	1.68E-01
hsa-miR-3613-5p	2.14	1.78E-01
hsa-miR-185-3p	1.55	1.78E-01
hsa-miR-187-5p	3.30	1.78E-01
hsa-miR-663a	3.32	1.89E-01
hsa-let-7i-3p	1.18	1.89E-01
hsa-miR-590-5p	1.98	1.97E-01
hsa-miR-138-5p	2.21	2.01E-01
hsa-miR-4713-5p	2.33	2.01E-01
hsa-miR-219b-5p	3.26	2.01E-01
hsa-miR-5582-3p	3.28	2.01E-01
hsa-miR-941	1.07	2.05E-01
hsa-miR-454-5p	1.48	2.13E-01
hsa-miR-34c-5p	1.05	2.13E-01
hsa-let-7g-5p	-0.94	2.16E-01
hsa-miR-3145-3p	3.19	2.16E-01
hsa-miR-584-3p	3.19	2.17E-01
hsa-miR-301a-3p	1.23	2.23E-01
hsa-miR-487b-3p	-1.23	2.27E-01
hsa-miR-106b-3p	0.96	2.27E-01
hsa-miR-3065-3p	-1.25	2.27E-01
hsa-miR-150-3p	-1.62	2.30E-01
hsa-miR-642a-3p	2.55	2.31E-01
hsa-miR-192-5p	-0.89	2.32E-01
hsa-miR-491-5p	1.54	2.32E-01
hsa-miR-129-1-3p	3.64	2.40E-01
hsa-miR-3129-5p	3.14	2.42E-01

Table 8-5: Differentially expressed miRNAs between metastatic primary tumours and matched normal epithelium.

miRNASeq			nCounter		
Mature miRNA	log Fold Change	Adjusted p-value	Mature miRNA	log Fold Change	Adjusted p-value
hsa-miR-133a-3p	7.90	6.55E-24	hsa-miR-206	8.33	4.35E-19
hsa-miR-208b-3p	8.26	5.42E-16	hsa-miR-133a	7.81	1.03E-14
hsa-miR-1	6.32	1.77E-14	hsa-miR-1	4.80	1.80E-10
hsa-miR-375	8.05	3.09E-13	hsa-miR-375	5.60	5.09E-04
hsa-miR-615-3p	-7.89	1.79E-12	hsa-miR-424-5p	-5.07	1.07E-03
hsa-miR-133b	7.59	3.07E-12	hsa-miR-99a-5p	3.48	1.39E-03
hsa-miR-206	8.66	6.39E-11	hsa-miR-204-5p	4.98	1.39E-03
hsa-miR-196a-5p	-5.27	8.62E-11	hsa-miR-378g	4.69	4.66E-03

hsa-miR-196b-5p	-4.48	8.78E-09	hsa-miR-503	-4.23	1.80E-02
hsa-miR-135a-5p	6.01	1.87E-08	hsa-miR-1246	-4.07	2.03E-02
hsa-let-7c-5p	3.78	2.48E-08	hsa-miR-630	-3.10	2.39E-02
hsa-miR-133a-5p	7.79	3.31E-08	hsa-miR-455-5p	-4.36	3.35E-02
hsa-miR-1246	-4.42	3.74E-08	hsa-miR-196b-5p	-4.05	4.19E-02
hsa-miR-204-5p	3.64	4.35E-07	hsa-miR-23b-3p	2.74	4.24E-02
hsa-miR-99a-3p	3.23	5.40E-07	hsa-miR-381	3.98	4.45E-02
hsa-miR-381-3p	3.68	9.62E-07	hsa-miR-378a-3p	2.36	4.58E-02
hsa-miR-561-5p	4.28	7.10E-06	hsa-miR-378i	2.36	4.58E-02
hsa-miR-499a-5p	3.91	2.45E-05	hsa-miR-31-5p	-2.86	4.58E-02
hsa-miR-146b-5p	-3.09	3.34E-05	hsa-miR-663b	-4.19	4.58E-02
hsa-miR-455-3p	-2.85	6.44E-05	hsa-miR-944	-3.66	5.52E-02
hsa-miR-6510-3p	3.23	1.11E-04	hsa-miR-29c-3p	2.07	5.52E-02
hsa-miR-2355-5p	-3.42	1.50E-04	hsa-miR-455-3p	-4.09	6.02E-02
hsa-miR-125b-2-3p	2.59	1.50E-04	hsa-miR-4286	-2.57	7.40E-02
hsa-miR-10a-5p	-3.11	2.44E-04	hsa-miR-518b	3.12	7.40E-02
hsa-miR-378i	2.80	2.55E-04	hsa-miR-299-5p	3.34	1.07E-01
hsa-miR-21-5p	-2.98	2.55E-04	hsa-miR-421	-3.02	1.45E-01
hsa-miR-99a-5p	2.73	2.59E-04	hsa-let-7c	1.92	1.69E-01
hsa-miR-378c	2.33	5.01E-04	hsa-miR-30c-5p	2.95	2.33E-01
hsa-miR-1247-3p	3.46	5.25E-04	hsa-miR-135b-5p	-1.74	2.33E-01
hsa-miR-3607-3p	3.60	6.16E-04	hsa-miR-133b	3.13	2.36E-01
hsa-miR-1247-5p	2.64	6.72E-04			
hsa-miR-31-3p	-3.13	8.24E-04			
hsa-miR-211-5p	4.30	9.05E-04			
hsa-miR-1269a	-6.23	9.62E-04			
hsa-miR-4770	4.13	1.29E-03			
hsa-miR-135b-5p	-2.30	1.44E-03			
hsa-miR-450b-5p	-2.24	1.61E-03			
hsa-miR-365a-5p	-3.08	1.84E-03			
hsa-miR-96-5p	-2.32	2.19E-03			
hsa-miR-139-5p	2.50	2.79E-03			
hsa-miR-1251-5p	6.00	2.84E-03			
hsa-let-7c-3p	3.26	3.24E-03			
hsa-miR-1910-5p	-3.27	3.48E-03			
hsa-miR-3129-3p	-3.76	3.77E-03			
hsa-miR-136-3p	2.31	4.77E-03			
hsa-miR-4776-5p	5.64	4.86E-03			
hsa-miR-378f	3.00	5.03E-03			
hsa-miR-1301-3p	-3.07	5.11E-03			
hsa-miR-193b-5p	-2.60	5.41E-03			
hsa-miR-582-3p	2.79	5.75E-03			
hsa-miR-7-5p	-2.60	8.30E-03			
hsa-miR-455-5p	-1.93	9.17E-03			
hsa-miR-208a-3p	5.42	9.45E-03			

hsa-miR-378d	2.21	9.80E-03			
hsa-miR-31-5p	-2.57	1.03E-02			
hsa-miR-941	-2.09	1.16E-02			
hsa-miR-10a-3p	-2.65	1.26E-02			
hsa-miR-7705	-2.55	1.37E-02			
hsa-miR-301b	-1.90	1.40E-02			
hsa-miR-454-3p	-1.82	1.40E-02			
hsa-miR-193b-3p	-2.14	1.43E-02			
hsa-miR-4662a-5p	-2.89	1.43E-02			
hsa-miR-101-3p	2.05	1.55E-02			
hsa-miR-4640-3p	-5.33	2.45E-02			
hsa-miR-335-3p	-2.05	2.58E-02			
hsa-miR-619-5p	-2.64	3.01E-02			
hsa-miR-378a-3p	2.00	3.69E-02			
hsa-miR-92b-3p	-1.87	4.15E-02			
hsa-miR-183-5p	-2.68	4.15E-02			
hsa-miR-130b-3p	-1.64	4.26E-02			
hsa-miR-146b-3p	-1.75	4.54E-02			
hsa-miR-4429	5.28	4.65E-02			
hsa-miR-5701	2.25	4.87E-02			
hsa-miR-3065-3p	2.06	4.97E-02			
hsa-miR-1276	-3.26	5.28E-02			
hsa-miR-6842-3p	-3.63	5.63E-02			
hsa-miR-146a-5p	-2.07	5.63E-02			
hsa-miR-3679-5p	-5.21	5.73E-02			
hsa-miR-421	-1.49	5.97E-02			
hsa-miR-503-5p	-2.84	5.99E-02			
hsa-miR-431-3p	2.68	6.00E-02			
hsa-miR-1271-3p	-4.82	6.95E-02			
hsa-miR-584-5p	-1.73	8.65E-02			
hsa-miR-4521	2.83	9.30E-02			
hsa-miR-22-5p	-1.67	9.30E-02			
hsa-miR-212-3p	-2.07	9.42E-02			
hsa-miR-5089-3p	4.43	9.54E-02			
hsa-miR-542-5p	-1.99	9.61E-02			
hsa-miR-125b-5p	1.71	9.66E-02			
hsa-miR-203b-5p	4.46	9.85E-02			
hsa-miR-3653	2.52	9.85E-02			
hsa-miR-424-5p	-1.64	9.96E-02			
hsa-miR-422a	2.39	1.00E-01			
hsa-miR-512-3p	-3.46	1.00E-01			
hsa-miR-183-3p	-2.32	1.00E-01			
hsa-miR-4713-5p	-3.10	1.02E-01			
hsa-miR-5096	-1.99	1.05E-01			
hsa-miR-3613-5p	-1.92	1.17E-01			

hsa-miR-1290	-2.48	1.18E-01			
hsa-miR-450a-5p	-1.91	1.19E-01			
hsa-miR-187-3p	1.97	1.21E-01			
hsa-miR-3687	-2.79	1.24E-01			
hsa-miR-1262	3.32	1.25E-01			
hsa-miR-424-3p	-2.04	1.25E-01			
hsa-miR-5700	-4.85	1.27E-01			
hsa-miR-25-5p	-1.73	1.30E-01			
hsa-miR-431-5p	-1.69	1.31E-01			
hsa-miR-141-5p	-1.61	1.33E-01			
hsa-miR-671-5p	-1.83	1.33E-01			
hsa-miR-615-5p	-4.66	1.33E-01			
hsa-miR-365a-3p	-1.57	1.37E-01			
hsa-miR-365b-3p	-1.57	1.37E-01			
hsa-miR-296-3p	2.12	1.48E-01			
hsa-miR-4652-5p	-4.56	1.48E-01			
hsa-miR-18a-5p	-1.70	1.60E-01			
hsa-miR-4326	-2.01	1.61E-01			
hsa-miR-6509-5p	-4.44	1.62E-01			
hsa-miR-194-3p	4.19	1.68E-01			
hsa-miR-598-3p	1.69	1.77E-01			
hsa-miR-4449	1.61	1.88E-01			
hsa-miR-30a-5p	1.52	1.94E-01			
hsa-miR-501-3p	-1.38	2.03E-01			
hsa-miR-548a-3p	-3.12	2.06E-01			
hsa-miR-4775	-1.82	2.21E-01			
hsa-miR-203b-3p	2.43	2.25E-01			
hsa-miR-105-3p	-4.49	2.25E-01			
hsa-miR-21-3p	-1.96	2.40E-01			
hsa-miR-149-5p	1.37	2.43E-01			
hsa-miR-30a-3p	1.62	2.43E-01			
hsa-miR-9-3p	-4.17	2.48E-01			
hsa-miR-136-5p	1.85	2.48E-01			

Table 8-6: Differentially expressed miRNAs between nodal metastases with ECS and matched normal epithelium.

miRNASeq			nCounter		
Mature miRNA	Log Fold Change	Adjusted p-value	Mature miRNA	Log Fold Change	Adjusted p-value
hsa-miR-375	9.30	6.56E-19	hsa-miR-1246	-9.64	2.68E-20
hsa-miR-196a-5p	-6.86	4.21E-08	hsa-miR-375	6.34	1.98E-05
hsa-miR-3168	11.10	3.80E-05	hsa-miR-21-5p	-3.87	1.43E-04
hsa-miR-211-5p	7.46	1.40E-04	hsa-miR-146b-5p	-5.37	5.70E-04
hsa-miR-135a-5p	6.39	1.75E-04	hsa-miR-424-5p	-4.64	5.70E-04

hsa-miR-133b	6.75	1.96E-04	hsa-miR-181b-5p	-4.66	7.36E-04
hsa-miR-615-3p	-7.87	2.25E-04	hsa-miR-181d	-4.66	7.36E-04
hsa-miR-196b-5p	-4.86	2.25E-04	hsa-miR-18a-5p	-4.52	1.46E-03
hsa-miR-21-3p	-4.22	2.54E-04	hsa-miR-455-5p	-4.38	2.58E-03
hsa-miR-99a-5p	3.97	2.83E-04	hsa-miR-93-5p	-1.70	3.10E-03
hsa-miR-133a-3p	5.66	3.07E-04	hsa-miR-28-5p	-3.74	3.10E-03
hsa-miR-1	7.27	7.61E-04	hsa-miR-296-5p	-4.30	3.10E-03
hsa-let-7c-5p	2.76	8.37E-03	hsa-miR-125a-5p	-1.77	3.70E-03
hsa-miR-6510-3p	3.35	9.02E-03	hsa-miR-503	-4.79	3.70E-03
hsa-miR-100-5p	2.62	1.75E-02	hsa-miR-99a-5p	2.06	4.30E-03
hsa-miR-203a	3.33	2.15E-02	hsa-miR-455-3p	-4.64	5.26E-03
hsa-miR-196a-3p	-6.18	6.16E-02	hsa-miR-183-5p	-3.58	7.65E-03
hsa-miR-139-3p	4.86	7.13E-02	hsa-miR-106b-5p	-1.54	1.10E-02
hsa-miR-4776-5p	5.24	7.65E-02	hsa-miR-196a-5p	-4.05	1.10E-02
hsa-miR-4770	-6.62	8.98E-02	hsa-miR-182-5p	-4.24	1.10E-02
hsa-miR-6842-3p	-5.68	8.98E-02	hsa-miR-155-5p	-3.01	1.10E-02
hsa-miR-21-5p	-2.14	9.65E-02	hsa-miR-143-3p	-5.35	1.43E-02
hsa-miR-125b-5p	2.19	9.65E-02	hsa-miR-221-3p	-1.65	1.58E-02
hsa-miR-1247-5p	2.67	9.65E-02	hsa-miR-34c-5p	-4.38	2.81E-02
hsa-miR-31-5p	-2.63	1.46E-01	hsa-miR-203	2.67	2.81E-02
hsa-miR-1269a	-5.91	1.46E-01	hsa-miR-421	-3.78	2.81E-02
hsa-miR-187-3p	-2.61	1.74E-01	hsa-let-7e-5p	-1.33	3.33E-02
hsa-miR-125b-2-3p	2.31	1.74E-01	hsa-miR-2682-5p	-3.82	3.33E-02
hsa-miR-424-3p	-3.04	1.74E-01	hsa-miR-454-3p	-3.72	4.03E-02
hsa-miR-139-5p	2.27	2.41E-01	hsa-miR-206	4.75	4.18E-02
hsa-miR-940	-3.73	2.41E-01	hsa-miR-130a-3p	-1.26	4.78E-02
hsa-miR-141-3p	2.11	2.41E-01	hsa-miR-450a-5p	-3.43	5.35E-02
			hsa-miR-25-3p	-1.23	6.00E-02
			hsa-let-7c	1.34	6.28E-02
			hsa-miR-324-5p	-2.73	6.28E-02
			hsa-miR-196b-5p	-3.69	6.28E-02
			hsa-miR-133a	4.15	6.72E-02
			hsa-miR-96-5p	-3.00	8.45E-02
			hsa-miR-944	-3.45	8.54E-02
			hsa-miR-132-3p	-2.87	8.71E-02
			hsa-miR-31-5p	-3.55	8.71E-02
			hsa-let-7i-5p	-1.22	9.91E-02
			hsa-miR-548ah-5p	-3.19	9.99E-02
			hsa-miR-106a-5p	-1.18	1.01E-01
			hsa-miR-17-5p	-1.18	1.01E-01
			hsa-miR-1290	-2.74	1.10E-01
			hsa-miR-15b-5p	-1.27	1.51E-01
			hsa-miR-181a-5p	-1.21	1.51E-01
			hsa-miR-15a-5p	-1.15	1.51E-01
			hsa-let-7d-5p	-1.06	1.56E-01

			hsa-miR-19b-3p	-1.07	1.57E-01
			hsa-miR-7-5p	-2.39	2.05E-01
			hsa-miR-130b-3p	-2.37	2.07E-01
			hsa-miR-1180	-3.14	2.12E-01
			hsa-miR-331-3p	-2.43	2.35E-01
			hsa-miR-125b-5p	1.46	2.35E-01

Table 8-7: Differentially expressed miRNAs between nodal metastases without ECS and matched normal epithelium.

Chapter 9

Bibliography

AGRAWAL, N., FREDERICK, M. J., PICKERING, C. R., BETTEGOWDA, C., CHANG, K., LI, R. J., et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science, 333, 1154-7.

ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A., BEHJATI, S., BIANKIN, A. V., et al. 2013. Signatures of mutational processes in human cancer. Nature, 500, 415-21.

ALKUREISHI, L. W., ROSS, G. L., SHOAIB, T., SOUTAR, D. S., ROBERTSON, A. G., SORENSEN, J. A., et al. 2008. Does tumor depth affect nodal upstaging in squamous cell carcinoma of the head and neck? Laryngoscope, 118, 629-34.

ALMEIDA, M. I., REIS, R. M. & CALIN, G. A. 2011. MicroRNA history: discovery, recent applications, and next frontiers. Mutat Res, 717, 1-8.

ALMENDRO, V., MARUSYK, A. & POLYAK, K. 2013. Cellular heterogeneity and molecular evolution in cancer. Annu Rev Pathol, 8, 277-302.

ALTERI, A., DE VITO, F., MESSINA, G., POMPILI, M., CALCONI, A., VISCA, P., et al. 2013. Cyclin D1 is a major target of miR-206 in cell differentiation and transformation. Cell Cycle, 12, 3781-90.

ALVI, A. & JOHNSON, J. T. 1996. Extracapsular spread in the clinically negative neck (N0): implications and outcome. Otolaryngol Head Neck Surg, 114, 65-70.

AMBATIPUDI, S., GERSTUNG, M., GOWDA, R., PAI, P., BORGES, A. M., SCHAFFER, A. A., et al. 2011. Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome. PLoS One, 6, e17250.

AMBROS, V. 2008. The evolution of our thinking about microRNAs. Nat Med, 14, 1036-40.

AMBROSIO, E. P., SILVEIRA, C. G., DRIGO, S. A., SACOMANO VDE, S., MOLCK, M. C., ROCHA, R. M., et al. 2013. Chromosomal imbalances exclusively detected in invasive front area are associated with poor outcome in laryngeal carcinomas from different anatomical sites. Tumour Biol, 34, 3015-26.

ANASTAS, J. N. & MOON, R. T. 2013. WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer, 13, 11-26.

ANG, K. K., HARRIS, J., WHEELER, R., WEBER, R., ROSENTHAL, D. I., NGUYEN-TAN, P. F., et al. 2010. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med, 363, 24-35.

ANSARY-MOGHADDAM, A., HUXLEY, R. R., LAM, T. H. & WOODWARD, M. 2009. Risk of Upper Aerodigestive Tract Cancer Associated with Smoking with and without Concurrent Alcohol Consumption. Mount Sinai Journal of Medicine, 76, 392-403.

ASAKAGE, T., YOKOSE, T., MUKAI, K., TSUGANE, S., TSUBONO, Y., ASAI, M., et al. 1998. Tumor thickness predicts cervical metastasis in patients with stage I/II carcinoma of the tongue. Cancer, 82, 1443-8.

AVISSAR, M., CHRISTENSEN, B. C., KELSEY, K. T. & MARSIT, C. J. 2009a. MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma. Clin Cancer Res, 15, 2850-5.

AVISSAR, M., MCCLEAN, M. D., KELSEY, K. T. & MARSIT, C. J. 2009b. MicroRNA expression in head and neck cancer associates with alcohol consumption and survival. Carcinogenesis, 30, 2059-63.

AYAZ, L., GORUR, A., YAROGLU, H. Y., OZCAN, C. & TAMER, L. 2013. Differential expression of microRNAs in plasma of patients with laryngeal squamous cell carcinoma: potential early-detection markers for laryngeal squamous cell carcinoma. J Cancer Res Clin Oncol, 139, 1499-506.

BADOUAL, C., SANDOVAL, F., PERE, H., HANS, S., GEY, A., MERILLON, N., et al. 2010. Better understanding tumor-host interaction in head and neck

cancer to improve the design and development of immunotherapeutic strategies. Head Neck, 32**,** 946-58.

BALERMPAS, P., RODEL, F., LIBERZ, R., OPPERMANN, J., WAGENBLAST, J., GHANAATI, S., et al. 2014. Head and neck cancer relapse after chemoradiotherapy correlates with CD163+ macrophages in primary tumour and CD11b+ myeloid cells in recurrences. Br J Cancer, 111, 1509-18.

BANDRES, E., CUBEDO, E., AGIRRE, X., MALUMBRES, R., ZARATE, R., RAMIREZ, N., et al. 2006. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. Mol Cancer, 5, 29.

BARTEL, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281-97.

BARTEL, D. P. 2009. MicroRNAs: target recognition and regulatory functions. Cell, 136, 215-33.

BARTELINK, H., BREUR, K., HART, G., ANNYAS, B., VAN SLOOTEN, E. & SNOW, G. 1983. The value of postoperative radiotherapy as an adjuvant to radical neck dissection. Cancer, 52, 1008-13.

BASS, A. J., WATANABE, H., MERMEL, C. H., YU, S., PERNER, S., VERHAAK, R. G., et al. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet, 41, 1238-42.

BAUER, V. L., BRASELMANN, H., HENKE, M., MATTERN, D., WALCH, A., UNGER, K., et al. 2008. Chromosomal changes characterize head and neck cancer with poor prognosis. J Mol Med (Berl), 86, 1353-65.

BELVEDERE, O., BERRI, S., CHALKLEY, R., CONWAY, C., BARBONE, F., PISA, F., et al. 2012. A computational index derived from whole-genome copy number analysis is a novel tool for prognosis in early stage lung squamous cell carcinoma. Genomics, 99, 18-24.

BENNETT, S. H., FUTRELL, J. W., ROTH, J. A., HOYE, R. C. & KETCHAM, A. S. 1971. Prognostic significance of histologic host response in cancer of the larynx or hypopharynx. Cancer, 28, 1255-65.

BENTZEN, S. M., SAUNDERS, M. I., DISCHE, S. & BOND, S. J. 2001. Radiotherapy-related early morbidity in head and neck cancer: quantitative clinical radiobiology as deduced from the CHART trial. Radiother Oncol, 60, 123-35.

BEROUKHIM, R., GETZ, G., NGHIEMPHU, L., BARRETINA, J., HSUEH, T., LINHART, D., et al. 2007. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A, 104, 20007-12.

BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., et al. 2010. The landscape of somatic copy-number alteration across human cancers. Nature, 463, 899-905.

BERRI, S. 2014. CNAnorm: A package to detect Copy Number Alterations (CNA) from sequencing data. http://www.bioconductor.org.

BHATTACHARYA, A., ROY, R., SNIJDERS, A. M., HAMILTON, G., PAQUETTE, J., TOKUYASU, T., et al. 2011. Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. Clin Cancer Res, 17, 7024-34.

BHARDWAJ, N. 2007. Harnessing the immune system to treat cancer. J Clin Invest, 117, 1130-6.

BHAYANI, M. K., CALIN, G. A. & LAI, S. Y. 2012. Functional relevance of miRNA sequences in human disease. Mutat Res, 731, 14-9.

BIGNELL, G. R., HUANG, J., GRESHOCK, J., WATT, S., BUTLER, A., WEST, S., et al. 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res, 14, 287-95.

BOCKMUHL, U., SCHLUNS, K., SCHMIDT, S., MATTHIAS, S. & PETERSEN, I. 2002. Chromosomal alterations during metastasis formation of head and neck squamous cell carcinoma. Genes Chromosomes Cancer, 33, 29-35.

BOCKMUHL, U., SCHWENDEL, A., DIETEL, M. & PETERSEN, I. 1996. Distinct patterns of chromosomal alterations in high- and low-grade head and neck squamous cell carcinomas. Cancer Res, 56, 5325-9.

BOCKMUHL, U., WOLF, G., SCHMIDT, S., SCHWENDEL, A., JAHNKE, V., DIETEL, M., et al. 1998. Genomic alterations associated with malignancy in head and neck cancer. Head Neck, 20, 145-51.

BONNAL, R. J., ROSSI, R. L., CARPI, D., RANZANI, V., ABRIGNANI, S. & PAGANI, M. 2015. miRiadne: a web tool for consistent integration of miRNA nomenclature. Nucleic Acids Res, 43, W487-92.

BONNER JA, T. H., YANG ES. 2015. The combined use of anti-EGFr and anti-JAK/STAT-3 agents as radiosensitizers in HPV negative and positive head and neck cancer cell lines. Oral Oncology, 51, e50-e51.

BORNSTEIN, S., WHITE, R., MALKOSKI, S., OKA, M., HAN, G., CLEAVER, T., et al. 2009. Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation. J Clin Invest, 119, 3408-19.

BOSCH, F. X., LORINCZ, A., MUNOZ, N., MEIJER, C. J. & SHAH, K. V. 2002. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol, 55, 244-65.

BOURGUIGNON, L. Y., EARLE, C., WONG, G., SPEVAK, C. C. & KRUEGER, K. 2012. Stem cell marker (Nanog) and Stat-3 signaling promote MicroRNA-21 expression and chemoresistance in hyaluronan/CD44-activated head and neck squamous cell carcinoma cells. Oncogene, 31, 149-60.

BRAAKHUIS, B. J., TABOR, M. P., KUMMER, J. A., LEEMANS, C. R. & BRAKENHOFF, R. H. 2003. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. Cancer Res, 63, 1727-30.

BRANNAN, A. G., JOHNSTONE, P. A. & COOPER, J. 2011. Extracapsular tumor extension in cervical lymph nodes: reconciling the literature and seer data. Head Neck, 33, 525-8.

BRASILINO DE CARVALHO, M. 1998. Quantitative analysis of the extent of extracapsular invasion and its prognostic significance: a prospective study of 170 cases of carcinoma of the larynx and hypopharynx. Head Neck, 20, 16-21.

BRENNAN, P., LEWIS, S., HASHIBE, M., BELL, D. A., BOFFETTA, P., BOUCHARDY, C., et al. 2004. Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: a HuGE review. Am J Epidemiol, 159, 1-16.

BRODERS, A. 1920. Squamous-cell epithelioma of the lip. JAMA, 74, 656-664.

BROSENS, R. P., HAAN, J. C., CARVALHO, B., RUSTENBURG, F., GRABSCH, H., QUIRKE, P., et al. 2010. Candidate driver genes in focal chromosomal aberrations of stage II colon cancer. J Pathol, 221, 411-24.

BUCKLEY, J. G. & MACLENNAN, K. 2000. Cervical node metastases in laryngeal and hypopharyngeal cancer: a prospective analysis of prevalence and distribution. Head Neck, 22, 380-5.

BURNET, M. 1964. Immunological Factors in the Process of Carcinogenesis. Br Med Bull, 20, 154-8.

BUDCZIES, J., KLAUSCHEN, F., SINN, B. V., GYORFFY, B., SCHMITT, W. D., DARB-ESFAHANI, S., et al. 2012. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One, 7, e51862.

BUSSEY, K. J., CHIN, K., LABABIDI, S., REIMERS, M., REINHOLD, W. C., KUO, W. L., et al. 2006. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. Mol Cancer Ther, 5, 853-67.

CALIFANO, J., VAN DER RIET, P., WESTRA, W., NAWROZ, H., CLAYMAN, G., PIANTADOSI, S., et al. 1996. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res, 56, 2488-92.

CALIN, G. A. & CROCE, C. M. 2006. MicroRNA signatures in human cancers. Nat Rev Cancer, 6, 857-66.

CALIN, G. A., DUMITRU, C. D., SHIMIZU, M., BICHI, R., ZUPO, S., NOCH, E., et al. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A, 99, 15524-9.

CALIN, G. A., SEVIGNANI, C., DUMITRU, C. D., HYSLOP, T., NOCH, E., YENDAMURI, S., et al. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A, 101, 2999-3004.

CAMILON, P. R., STOKES, W. A., FULLER, C. W., NGUYEN, S. A. & LENTSCH, E. J. 2014. Does buccal cancer have worse prognosis than other oral cavity cancers? Laryngoscope, 124, 1386-91.

CANCER GENOME ATLAS RESEARCH, N. 2012. Comprehensive genomic characterization of squamous cell lung cancers. Nature, 489, 519-25.

CANCER GENOME ATLAS, N. 2015. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature, 517, 576-82.

CANDELA, F. C., KOTHARI, K. & SHAH, J. P. 1990a. Patterns of cervical node metastases from squamous carcinoma of the oropharynx and hypopharynx. Head Neck, 12, 197-203.

CANDELA, F. C., SHAH, J., JAQUES, D. P. & SHAH, J. P. 1990b. Patterns of cervical node metastases from squamous carcinoma of the larynx. Arch Otolaryngol Head Neck Surg, 116, 432-5.

CANTLEY, L. C. 2002. The phosphoinositide 3-kinase pathway. Science, 296, 1655-7.

CAO, P., ZHOU, L., ZHANG, J., ZHENG, F., WANG, H., MA, D., et al. 2013. Comprehensive expression profiling of microRNAs in laryngeal squamous cell carcinoma. Head Neck, 35, 720-8.

CAPACCIO, P., PRUNERI, G., CARBONI, N., PAGLIARI, A. V., QUATELA, M., CESANA, B. M., et al. 2000. Cyclin D1 expression is predictive of occult

metastases in head and neck cancer patients with clinically negative cervical lymph nodes. Head Neck, 22, 234-40.

CAPPUZZO, F., SACCONI, A., LANDI, L., LUDOVINI, V., BIAGIONI, F., D'INCECCO, A., et al. 2014. MicroRNA signature in metastatic colorectal cancer patients treated with anti-EGFR monoclonal antibodies. Clin Colorectal Cancer, 13, 37-45 e4.

CARTER, R. L., BLISS, J. M., SOO, K. C. & O'BRIEN, C. J. 1987. Radical neck dissections for squamous carcinomas: pathological findings and their clinical implications with particular reference to transcapsular spread. Int J Radiat Oncol Biol Phys, 13, 825-32.

CASTOLDI, M., SCHMIDT, S., BENES, V., NOERHOLM, M., KULOZIK, A. E., HENTZE, M. W., et al. 2006. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA, 12, 913-20.

CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov, 2, 401-4.

CERVIGNE, N. K., REIS, P. P., MACHADO, J., SADIKOVIC, B., BRADLEY, G., GALLONI, N. N., et al. 2009. Identification of a microRNA signature associated with progression of leukoplakia to oral carcinoma. Hum Mol Genet, 18, 4818-29.

CHANG, C. J., HSU, C. C., CHANG, C. H., TSAI, L. L., CHANG, Y. C., LU, S. W., et al. 2011. Let-7d functions as novel regulator of epithelial-mesenchymal transition and chemoresistant property in oral cancer. Oncol Rep, 26, 1003-10.

CHANG, K. Y., TSAI, S. Y., CHEN, S. H., TSOU, H. H., YEN, C. J., LIU, K. J., et al. 2013. Dissecting the EGFR-PI3K-AKT pathway in oral cancer highlights the role of the EGFR variant III and its clinical relevance. J Biomed Sci, 20, 43.

CHANG, S. S., JIANG, W. W., SMITH, I., POETA, L. M., BEGUM, S., GLAZER, C., et al. 2008. MicroRNA alterations in head and neck squamous cell carcinoma. Int J Cancer, 123, 2791-7.

CHATURVEDI, A. K., ENGELS, E. A., ANDERSON, W. F. & GILLISON, M. L. 2008. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. J Clin Oncol, 26, 612-9.

CHEN, I. H., CHANG, J. T., LIAO, C. T., WANG, H. M., HSIEH, L. L. & CHENG, A. J. 2003a. Prognostic significance of EGFR and Her-2 in oral cavity cancer in betel quid prevalent area cancer prognosis. Br J Cancer, 89, 681-6.

CHEN, L., GIBBONS, D. L., GOSWAMI, S., CORTEZ, M. A., AHN, Y. H., BYERS, L. A., et al. 2014. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. Nat Commun, 5, 5241.

CHEN, Y. & CHEN, C. 2008. DNA copy number variation and loss of heterozygosity in relation to recurrence of and survival from head and neck squamous cell carcinoma: a review. Head Neck, 30, 1361-83.

CHEN, Y. C., LIU, T., YU, C. H., CHIANG, T. Y. & HWANG, C. C. 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS One, 8, e62856.

CHEN, Y., KNOSEL, T., KRISTIANSEN, G., PIETAS, A., GARBER, M. E., MATSUHASHI, S., et al. 2003b. Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. J Pathol, 200, 640-6.

CHEN, Z., JIN, Y., YU, D., WANG, A., MAHJABEEN, I., WANG, C., et al. 2012. Down-regulation of the microRNA-99 family members in head and neck squamous cell carcinoma. Oral Oncol, 48, 686-91.

CHENG, L., WANG, P., YANG, S., YANG, Y., ZHANG, Q., ZHANG, W., et al. 2012. Identification of genes with a correlation between copy number and expression in gastric cancer. BMC Med Genomics, 5, 14.

CHHETRI, D. K., RAWNSLEY, J. D. & CALCATERRA, T. C. 2000. Carcinoma of the buccal mucosa. Otolaryngol Head Neck Surg, 123, 566-71.

CHIANG, W. F., LIU, S. Y., YEN, C. Y., LIN, C. N., CHEN, Y. C., LIN, S. C., et al. 2008. Association of epidermal growth factor receptor (EGFR) gene copy number amplification with neck lymph node metastasis in areca-associated oral carcinomas. Oral Oncol, 44, 270-6.

CHILDS, G., FAZZARI, M., KUNG, G., KAWACHI, N., BRANDWEIN-GENSLER, M., MCLEMORE, M., et al. 2009. Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma. Am J Pathol, 174, 736-45.

CHITALE, D., GONG, Y., TAYLOR, B. S., BRODERICK, S., BRENNAN, C., SOMWAR, R., et al. 2009. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene, 28, 2773-83.

CHOI, J. M., DEVKOTA, S., SUNG, Y. H. & LEE, H. W. 2013. EI24 regulates epithelial-to-mesenchymal transition and tumor progression by suppressing TRAF2-mediated NF-kappaB activity. Oncotarget, 4, 2383-96.

CHU, P. Y., HU, F. W., YU, C. C., TSAI, L. L., YU, C. H., WU, B. C., et al. 2013. Epithelial-mesenchymal transition transcription factor ZEB1/ZEB2 co-expression predicts poor prognosis and maintains tumor-initiating properties in head and neck cancer. Oral Oncol, 49, 34-41.

CHUNG, C. H., LEVY, S. & YARBROUGH, W. G. 2006. Clinical applications of genomics in head and neck cancer. Head Neck, 28, 360-8.

CHUNG, C. H., PARKER, J. S., KARACA, G., WU, J., FUNKHOUSER, W. K., MOORE, D., et al. 2004. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell, 5, 489-500.

CLOSE, L. G., BROWN, P. M., VUITCH, M. F., REISCH, J. & SCHAEFER, S. D. 1989. Microvascular invasion and survival in cancer of the oral cavity and oropharynx. Arch Otolaryngol Head Neck Surg, 115, 1304-9.

CLOSE, L. G., BURNS, D. K., REISCH, J. & SCHAEFER, S. D. 1987. Microvascular invasion in cancer of the oral cavity and oropharynx. Arch Otolaryngol Head Neck Surg, 113, 1191-5.

COATESWORTH, A. P. & MACLENNAN, K. 2002. Squamous cell carcinoma of the upper aerodigestive tract: the prevalence of microscopic extracapsular spread and soft tissue deposits in the clinically N0 neck. Head Neck, 24, 258-61.

COBB, J. P., MINDRINOS, M. N., MILLER-GRAZIANO, C., CALVANO, S. E., BAKER, H. V., XIAO, W., et al. 2005. Application of genome-wide expression analysis to human health and disease. Proc Natl Acad Sci U S A, 102, 4801-6.

COLANGELO, T., FUCCI, A., VOTINO, C., SABATINO, L., PANCIONE, M., LAUDANNA, C., et al. 2013. MicroRNA-130b promotes tumor development and is associated with poor prognosis in colorectal cancer. Neoplasia, 15, 1086-99.

COMBES, J. D. & FRANCESCHI, S. 2014. Role of human papillomavirus in non-oropharyngeal head and neck cancers. Oral Oncol, 50, 370-9.

CONWAY, C., CHALKLEY, R., HIGH, A., MACLENNAN, K., BERRI, S., CHENGOT, P., et al. 2012. Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. J Mol Diagn, 14, 104-11.

CORDES, K. R., SHEEHY, N. T., WHITE, M. P., BERRY, E. C., MORTON, S. U., MUTH, A. N., et al. 2009. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. Nature, 460, 705-10.

COURTNEY, K. D., CORCORAN, R. B. & ENGELMAN, J. A. 2010. The PI3K pathway as drug target in human cancer. J Clin Oncol, 28, 1075-83.

CREIGHTON, C. J., FOUNTAIN, M. D., YU, Z., NAGARAJA, A. K., ZHU, H., KHAN, M., et al. 2010. Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. Cancer Res, 70, 1906-15.

CROUCHER, R. & ISLAM, S. 2002. Socio-economic aspects of areca nut use. Addict Biol, 7, 139-46.

CUI, F., LI, X., ZHU, X., HUANG, L., HUANG, Y., MAO, C., et al. 2012. MiR-125b inhibits tumor growth and promotes apoptosis of cervical cancer cells

by targeting phosphoinositide 3-kinase catalytic subunit delta. Cell Physiol Biochem, 30, 1310-8.

CUI, H., KONG, Y., XU, M. & ZHANG, H. 2013. Notch3 functions as a tumor suppressor by controlling cellular senescence. Cancer Res, 73, 3451-9.

CURTIS, C., SHAH, S. P., CHIN, S. F., TURASHVILI, G., RUEDA, O. M., DUNNING, M. J., et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 486, 346-52.

D'CRUZ, A. K., VAISH, R., KAPRE, N., DANDEKAR, M., GUPTA, S., HAWALDAR, R., et al. 2015. Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer. N Engl J Med.

DA FORNO, P. D., PRINGLE, J. H., HUTCHINSON, P., OSBORN, J., HUANG, Q., POTTER, L., et al. 2008. WNT5A expression increases during melanoma progression and correlates with outcome. Clin Cancer Res, 14, 5825-32.

DA SILVA, S. D., FERLITO, A., TAKES, R. P., BRAKENHOFF, R. H., VALENTIN, M. D., WOOLGAR, J. A., et al. 2011. Advances and applications of oral cancer basic research. Oral Oncol, 47, 783-91.

DANIAL, N. N. & KORSMEYER, S. J. 2004. Cell death: critical control points. Cell, 116, 205-19.

DARIDO, C., GEORGY, S. R., WILANOWSKI, T., DWORKIN, S., AUDEN, A., ZHAO, Q., et al. 2011. Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. Cancer Cell, 20, 635-48.

DASGUPTA, S., MUKHERJEE, N., ROY, S., ROY, A., SENGUPTA, A., ROYCHOWDHURY, S., et al. 2002. Mapping of the candidate tumor suppressor genes' loci on human chromosome 3 in head and neck squamous cell carcinoma of an Indian patient population. Oral Oncol, 38, 6-15.

DE BRUIN, E. C., VAN DE PAS, S., LIPS, E. H., VAN EIJK, R., VAN DER ZEE, M. M., LOMBAERTS, M., et al. 2005. Macrodissection versus

microdissection of rectal carcinoma: minor influence of stroma cells to tumor cell gene expression profiles. BMC Genomics, 6, 142.

DE SANJOSE, S., QUINT, W. G., ALEMANY, L., GERAETS, D. T., KLAUSTERMEIER, J. E., LLOVERAS, B., et al. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol, 11, 1048-56.

DE SOUSA, E. M., VERMEULEN, L., RICHEL, D. & MEDEMA, J. P. 2011. Targeting Wnt signaling in colon cancer stem cells. Clin Cancer Res, 17, 647-53.

DESOUKI, M. M., LIAO, S., HUANG, H., CONROY, J., NOWAK, N. J., SHEPHERD, L., et al. 2011. Identification of metastasis-associated breast cancer genes using a high-resolution whole genome profiling approach. J Cancer Res Clin Oncol, 137, 795-809.

DHANDA, J., TRIANTAFYLLOU, A., LILOGLOU, T., KALIRAI, H., LLOYD, B., HANLON, R., et al. 2014. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. Br J Cancer, 111, 2114-21.

DILLHOFF, M., LIU, J., FRANKEL, W., CROCE, C. & BLOOMSTON, M. 2008. MicroRNA-21 is overexpressed in pancreatic cancer and a potential predictor of survival. J Gastrointest Surg, 12, 2171-6.

DONADELLI, M., DANDO, I., FIORINI, C. & PALMIERI, M. 2014. Regulation of miR-23b expression and its dual role on ROS production and tumour development. Cancer Lett, 349, 107-13.

DOWNWARD, J. 2003. Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer, 3, 11-22.

DRUSCO, A., NUOVO, G. J., ZANESI, N., DI LEVA, G., PICHIORRI, F., VOLINIA, S., et al. 2014. MicroRNA profiles discriminate among colon cancer metastasis. PLoS One, 9, e96670.

DU, Z., TONG, X. & YE, X. 2013. Cyclin D1 promotes cell cycle progression through enhancing NDR1/2 kinase activity independent of cyclin-dependent kinase 4. J Biol Chem, 288, 26678-87.

DUAN, J., ZHANG, J. G., DENG, H. W. & WANG, Y. P. 2013. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. PLoS One, 8, e59128.

Ehrlich P.,1909. Uber den jetzigen Stand der Karzinomforschung. Nederlands Tijdschrift voor Geneeskunde, 5, 273-90.

ELWOOD, J. M., PEARSON, J. C., SKIPPEN, D. H. & JACKSON, S. M. 1984. Alcohol, smoking, social and occupational factors in the aetiology of cancer of the oral cavity, pharynx and larynx. Int J Cancer, 34, 603-12.

EMMERT-BUCK, M. R., BONNER, R. F., SMITH, P. D., CHUAQUI, R. F., ZHUANG, Z., GOLDSTEIN, S. R., et al. 1996. Laser capture microdissection. Science, 274, 998-1001.

ENGELMAN, J. A. 2009. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer, 9, 550-62.

ENGELMAN, J. A., LUO, J. & CANTLEY, L. C. 2006. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. Nat Rev Genet, 7, 606-19.

ENGELS, E. A., PFEIFFER, R. M., FRAUMENI, J. F., JR., KASISKE, B. L., ISRANI, A. K., SNYDER, J. J., et al. 2011. Spectrum of cancer risk among US solid organ transplant recipients. JAMA, 306, 1891-901.

ESCHER, A., PIOTET, E., WARIDEL, F., IGGO, R. & MONNIER, P. 2009. p53 Mutation in histologically normal mucosa of the aero-digestive tract is not a marker of increased risk for second primary carcinoma in head and neck cancer patients. Eur Arch Otorhinolaryngol, 266, 547-51.

ESPINA, V., WULFKUHLE, J. D., CALVERT, V. S., VANMETER, A., ZHOU, W., COUKOS, G., et al. 2006. Laser-capture microdissection. Nat Protoc, 1, 586-603.

ESTEVE, J., RIBOLI, E., PEQUIGNOT, G., TERRACINI, B., MERLETTI, F., CROSIGNANI, P., et al. 1996. Diet and cancers of the larynx and hypopharynx: the IARC multi-center study in southwestern Europe. Cancer Causes Control, 7, 240-52.

EZKURDIA, I., JUAN, D., RODRIGUEZ, J. M., FRANKISH, A., DIEKHANS, M., HARROW, J., et al. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet, 23**, 5866-78.

FABBRI, M., GARZON, R., CIMMINO, A., LIU, Z., ZANESI, N., CALLEGARI, E., et al. 2007. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A, 104, 15805-10.

FAGAN, J. J., COLLINS, B., BARNES, L., D'AMICO, F., MYERS, E. N. & JOHNSON, J. T. 1998. Perineural invasion in squamous cell carcinoma of the head and neck. Arch Otolaryngol Head Neck Surg, 124, 637-40.

FAKHRY, C., WESTRA, W. H., LI, S., CMELAK, A., RIDGE, J. A., PINTO, H., et al. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst, 100, 261-9.

FARAONI, I., ANTONETTI, F. R., CARDONE, J. & BONMASSAR, E. 2009. miR-155 gene: a typical multifunctional microRNA. Biochim Biophys Acta, 1792, 497-505.

FARRAND, K., JOVANOVIC, L., DELAHUNT, B., MCIVER, B., HAY, I. D., EBERHARDT, N. L., et al. 2002. Loss of heterozygosity studies revisited: prior quantification of the amplifiable DNA content of archival samples improves efficiency and reliability. J Mol Diagn, 4, 150-8.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. Cell, 61, 759-67.

FENDRI, A., KHABIR, A., MNEJJA, W., SELLAMI-BOUDAWARA, T., DAOUD, J., FRIKHA, M., et al. 2009. PIK3CA amplification is predictive of poor prognosis in Tunisian patients with nasopharyngeal carcinoma. Cancer Sci, 100, 2034-9.

FENG, Z., GUO, W., ZHANG, C., XU, Q., ZHANG, P., SUN, J., et al. 2011. CCND1 as a predictive biomarker of neoadjuvant chemotherapy in patients with locally advanced head and neck squamous cell carcinoma. PLoS One, 6, e26399.

FENIC, I., STEGER, K., GRUBER, C., ARENS, C. & WOENCKHAUS, J. 2007. Analysis of PIK3CA and Akt/protein kinase B in head and neck squamous cell carcinoma. Oncol Rep, 18, 253-9.

FERLAY J, S. I., ERVIK M, DIKSHIT R, ESER S, MATHERS C, REBELO M, PARKIN DM, FORMAN D, BRAY, F 2013. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. GLOBOCAN 2012 v1.0. Lyon, France: International Agency for Research on Cancer.

FERLITO, A., RINALDO, A., BUCKLEY, J. G. & MONDIN, V. 2001. General considerations on distant metastases from head and neck cancer. ORL J Otorhinolaryngol Relat Spec, 63, 189-91.

FERLITO, A., RINALDO, A., ROBBINS, K. T. & SILVER, C. E. 2006a. Neck dissection: past, present and future? J Laryngol Otol, 120, 87-92.

FERLITO, A., RINALDO, A., SILVER, C. E., GOURIN, C. G., SHAH, J. P., CLAYMAN, G. L., et al. 2006b. Elective and therapeutic selective neck dissection. Oral Oncol, 42, 14-25.

FERLITO, A., RINALDO, A., SILVER, C. E., SHAH, J. P., SUAREZ, C., MEDINA, J. E., et al. 2006c. Neck dissection: then and now. Auris Nasus Larynx, 33, 365-74.

FEUK, L., CARSON, A. R. & SCHERER, S. W. 2006. Structural variation in the human genome. Nat Rev Genet, 7, 85-97.

FILIPOWICZ, W., BHATTACHARYYA, S. N. & SONENBERG, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet, 9, 102-14.

FLETCHER, A. M., HEAFORD, A. C. & TRASK, D. K. 2008. Detection of metastatic head and neck squamous cell carcinoma using the relative expression of tissue-specific mir-205. Transl Oncol, 1, 202-8.

FORBES, S. A., BEARE, D., GUNASEKARAN, P., LEUNG, K., BINDAL, N., BOUTSELAKIS, H., et al. 2015. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res, 43, D805-11.

FOULDS, L. 1954. The experimental study of tumor progression: a review. Cancer Res, 14, 327-39.

FRANKEL, A., ARMOUR, N., NANCARROW, D., KRAUSE, L., HAYWARD, N., LAMPE, G., et al. 2014. Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. Genes Chromosomes Cancer, 53, 324-38.

FRANKEL, L. B., CHRISTOFFERSEN, N. R., JACOBSEN, A., LINDOW, M., KROGH, A. & LUND, A. H. 2008. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. J Biol Chem, 283, 1026-33.

FREIER, K., PUNGS, S., STICHT, C., FLECHTENMACHER, C., LICHTER, P., JOOS, S., et al. 2007. High survivin expression is associated with favorable outcome in advanced primary oral squamous cell carcinoma after radiation therapy. Int J Cancer, 120, 942-6.

FRIEDLANDER, M. R., CHEN, W., ADAMIDI, C., MAASKOLA, J., EINSPANIER, R., KNESPEL, S., et al. 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol, 26, 407-15.

FRIEDMAN, R. C., FARH, K. K., BURGE, C. B. & BARTEL, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res, 19, 92-105.

GAJEWSKI, T. F., FUERTES, M. B. & WOO, S. R. 2012. Innate immune sensing of cancer: clues from an identified role for type I IFNs. Cancer Immunol Immunother, 61, 1343-7.

GARDEN, A. S. 2014. The never-ending story: finding a role for neoadjuvant chemotherapy in the management of head and neck cancer. J Clin Oncol, 32, 2685-6.

GEBESHUBER, C. A., ZATLOUKAL, K. & MARTINEZ, J. 2009. miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. EMBO Rep, 10, 400-5.

GEBHART, E. & LIEHR, T. 2000. Patterns of genomic imbalances in human solid tumors (Review). Int J Oncol, 16, 383-99.

GEISS, G. K., BUMGARNER, R. E., BIRDITT, B., DAHL, T., DOWIDAR, N., DUNAWAY, D. L., et al. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol, 26, 317-25.

GERALDO, M. V., YAMASHITA, A. S. & KIMURA, E. T. 2012. MicroRNA miR-146b-5p regulates signal transduction of TGF-beta by repressing SMAD4 in thyroid cancer. Oncogene, 31, 1910-22.

GERLINGER, M., HORSWELL, S., LARKIN, J., ROWAN, A. J., SALM, M. P., VARELA, I., et al. 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet, 46, 225-33.

GERLINGER, M., ROWAN, A. J., HORSWELL, S., LARKIN, J., ENDESFELDER, D., GRONROOS, E., et al. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med, 366, 883-92.

GILLISON, M. L. & LOWY, D. R. 2004. A causal role for human papillomavirus in head and neck cancer. Lancet, 363, 1488-9.

GILLISON, M. L. 2004. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. Semin Oncol, 31, 744-54.

GILLISON, M. L., D'SOUZA, G., WESTRA, W., SUGAR, E., XIAO, W., BEGUM, S., et al. 2008. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst, 100, 407-20.

GLEICH, L. L. & SALAMONE, F. N. 2002. Molecular genetics of head and neck cancer. Cancer Control, 9, 369-78.

GOESSEL, G., QUANTE, M., HAHN, W. C., HARADA, H., HEEG, S., SULIMAN, Y., et al. 2005. Creating oral squamous cancer cells: a cellular model of oral-esophageal carcinogenesis. Proc Natl Acad Sci U S A, 102, 15599-604.

GOLDSTEIN, D. P., BACHAR, G. Y., LEA, J., SHRIME, M. G., PATEL, R. S., GULLANE, P. J., et al. 2013. Outcomes of squamous cell cancer of the

oral tongue managed at the Princess Margaret Hospital. Head Neck, 35, 632-41.

GOLLIN, S. M. 2001. Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. Head Neck, 23, 238-53.

GOMBOS, K., HORVATH, R., SZELE, E., JUHASZ, K., GOCZE, K., SOMLAI, K., et al. 2013. miRNA expression profiles of oral squamous cell carcinomas. Anticancer Res, 33, 1511-7.

GOODEN, M. J., DE BOCK, G. H., LEFFERS, N., DAEMEN, T. & NIJMAN, H. W. 2011. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. Br J Cancer, 105, 93-103.

GOTTE, K., TREMMEL, S. C., POPP, S., WEBER, S., HORMANN, K., BARTRAM, C. R., et al. 2005. Intratumoral genomic heterogeneity in advanced head and neck cancer detected by comparative genomic hybridization. Adv Otorhinolaryngol, 62, 38-48.

GREEN, V. L., MICHNO, A., STAFFORD, N. D. & GREENMAN, J. 2013. Increased prevalence of tumour infiltrating immune cells in oropharyngeal tumours in comparison to other subsites: relationship to peripheral immunity. Cancer Immunol Immunother, 62, 863-73.

GREENBERG, J. S., FOWLER, R., GOMEZ, J., MO, V., ROBERTS, D., EL NAGGAR, A. K., et al. 2003. Extent of extracapsular spread: a critical prognosticator in oral tongue cancer. Cancer, 97, 1464-70.

GREER, C. E., LUND, J. K. & MANOS, M. M. 1991a. PCR amplification from paraffin-embedded tissues: recommendations on fixatives for long-term storage and prospective studies. PCR Methods Appl, 1, 46-50.

GREER, C. E., PETERSON, S. L., KIVIAT, N. B. & MANOS, M. M. 1991b. PCR amplification from paraffin-embedded tissues. Effects of fixative and fixation time. Am J Clin Pathol, 95, 117-24.

GRIFFITHS-JONES, S. 2004. The microRNA Registry. Nucleic Acids Res, 32, D109-11.

GRONWALD, J., HADACZEK, P., STORKEL, S., HOLTGREVE-GREZ, H., RABBITTS, P., CREMER, T., et al. 1999. Molecular evidence for derivation of metastatic cells from minor subclones of primary clear renal cell carcinomas. Cancer Detect Prev, 23, 479-84.

GROSS, A. M., OROSCO, R. K., SHEN, J. P., EGLOFF, A. M., CARTER, H., HOFREE, M., et al. 2014. Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. Nat Genet, 46, 939-43.

GU, Z., FLEMINGTON, C., CHITTENDEN, T. & ZAMBETTI, G. P. 2000a. ei24, a p53 response gene involved in growth suppression and apoptosis. Mol Cell Biol, 20, 233-41.

GU, Z., GILBERT, D. J., VALENTINE, V. A., JENKINS, N. A., COPELAND, N. G. & ZAMBETTI, G. P. 2000b. The p53-inducible gene EI24/PIG8 localizes to human chromosome 11q23 and the proximal region of mouse chromosome 9. Cytogenet Cell Genet, 89, 230-3.

GUO, Y., CHEN, Z., ZHANG, L., ZHOU, F., SHI, S., FENG, X., et al. 2008. Distinctive microRNA profiles relating to patient survival in esophageal squamous cell carcinoma. Cancer Res, 68, 26-33.

GUSNANTO, A., TAYLOR, C. C., NAFISAH, I., WOOD, H. M., RABBITTS, P. & BERRI, S. 2014. Estimating optimal window size for analysis of low-coverage next-generation sequence data. Bioinformatics, 30, 1823-9.

GUSNANTO, A., WOOD, H. M., PAWITAN, Y., RABBITTS, P. & BERRI, S. 2012. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. Bioinformatics, 28, 40-7.

HADDADIN, K. J., SOUTAR, D. S., OLIVER, R. J., WEBSTER, M. H., ROBERTSON, A. G. & MACDONALD, D. G. 1999. Improved survival for patients with clinically T1/T2, N0 tongue tumors undergoing a prophylactic neck dissection. Head Neck, 21, 517-25.

HAHN, S. A., SCHUTTE, M., HOQUE, A. T., MOSKALUK, C. A., DA COSTA, L. T., ROZENBLUM, E., et al. 1996. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science, 271, 350-3.

HAHN, S. S., SPAULDING, C. A., KIM, J. A. & CONSTABLE, W. C. 1987. The prognostic significance of lymph node involvement in pyriform sinus and supraglottic cancers. Int J Radiat Oncol Biol Phys, 13, 1143-7.

HAHN, W. C. & WEINBERG, R. A. 2002. Rules for making human tumor cells. N Engl J Med, 347, 1593-603.

HANAHAN, D. & WEINBERG, R. A. 2000. The hallmarks of cancer. Cell, 100, 57-70.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. Cell, 144, 646-74.

HAO, S. P. & NG, S. H. 2000. Magnetic resonance imaging versus clinical palpation in evaluating cervical metastasis from head and neck cancer. Otolaryngol Head Neck Surg, 123, 324-7.

HARRIS, T., JIMENEZ, L., KAWACHI, N., FAN, J. B., CHEN, J., BELBIN, T., et al. 2012. Low-level expression of miR-375 correlates with poor outcome and metastasis while altering the invasive properties of head and neck squamous cell carcinomas. Am J Pathol, 180, 917-28.

HASINA, R., WHIPPLE, M. E., MARTIN, L. E., KUO, W. P., OHNO-MACHADO, L. & LINGEN, M. W. 2008. Angiogenic heterogeneity in head and neck squamous cell carcinoma: biological and therapeutic implications. Lab Invest, 88, 342-53.

HAUGHEY, B. H. & SINHA, P. 2012. Prognostic factors and survival unique to surgically treated p16+ oropharyngeal cancer. Laryngoscope, 122 Suppl 2, S13-33.

HAUGHEY, B. H., VON HOFF, D. D., WINDLE, B. E., WAHL, G. M. & MOCK, P. M. 1992. c-myc oncogene copy number in squamous carcinoma of the head and neck. Am J Otolaryngol, 13, 168-71.

HAYES, J. L., TZIKA, A., THYGESEN, H., BERRI, S., WOOD, H. M., HEWITT, S., et al. 2013. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. Genomics, 102, 174-81.

HELLIWELL T, W. J. 2013. Dataset for histopathology reporting of nodal excisions and neck dissection specimens associated with head and neck carcinomas. In: PATHOLOGISTS, R. C. O. (ed.). London: Royal College of Pathologists.

HENNESSEY, P. T., WESTRA, W. H. & CALIFANO, J. A. 2009. Human papillomavirus and head and neck squamous cell carcinoma: recent evidence and clinical implications. J Dent Res, 88, 300-6.

HENSON, B. J., BHATTACHARJEE, S., O'DEE, D. M., FEINGOLD, E. & GOLLIN, S. M. 2009. Decreased expression of miR-125b and miR-100 in oral cancer cells contributes to malignancy. Genes Chromosomes Cancer, 48, 569-82.

HIERONYMUS, H., SCHULTZ, N., GOPALAN, A., CARVER, B. S., CHANG, M. T., XIAO, Y., et al. 2014. Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci U S A, 111, 11139-44.

HIRABAYASHI, H., KOSHII, K., UNO, K., OHGAKI, H., NAKASONE, Y., FUJISAWA, T., et al. 1991. Extracapsular spread of squamous cell carcinoma in neck lymph nodes: prognostic factor of laryngeal cancer. Laryngoscope, 101, 502-6.

HITOMI, M. & STACEY, D. W. 1999. Cyclin D1 production in cycling cells depends on ras in a cell-cycle-specific manner. Curr Biol, 9, 1075-84.

HONG, M. K., MACINTYRE, G., WEDGE, D. C., VAN LOO, P., PATEL, K., LUNKE, S., et al. 2015. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nat Commun, 6, 6605.

HOUSELEY, J. & TOLLERVEY, D. 2009. The many pathways of RNA degradation. Cell, 136, 763-76.

HSU, C. M., LIN, P. M., WANG, Y. M., CHEN, Z. J., LIN, S. F. & YANG, M. Y. 2012. Circulating miRNA is a novel marker for head and neck squamous cell carcinoma. Tumour Biol, 33, 1933-42.

HUANG, X., GODFREY, T. E., GOODING, W. E., MCCARTY, K. S., JR. & GOLLIN, S. M. 2006. Comprehensive genome and transcriptome analysis of

the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. Genes Chromosomes Cancer, 45, 1058-69.

HUEBNER, K. & CROCE, C. M. 2003. Cancer and the FRA3B/FHIT fragile locus: it's a HIT. Br J Cancer, 88, 1501-6.

HUGHES S, W. R., MOLDOVAN L, SQUIRE JA. 2007. Use of quantitative PCR for the detection of genomic microdeletions or microduplications. In: HUGHES S, M. A. (ed.) PCR: Methods Express. Bloxham, UK.: Scion Publishing Ltd.

HUI, A. B., BRUCE, J. P., ALAJEZ, N. M., SHI, W., YUE, S., PEREZ-ORDONEZ, B., et al. 2011. Significance of dysregulated metadherin and microRNA-375 in head and neck cancer. Clin Cancer Res, 17, 7539-50.

HUI, A. B., LENARDUZZI, M., KRUSHEL, T., WALDRON, L., PINTILIE, M., SHI, W., et al. 2010. Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas. Clin Cancer Res, 16, 1129-39.

HUR, K., TOIYAMA, Y., SCHETTER, A. J., OKUGAWA, Y., HARRIS, C. C., BOLAND, C. R., et al. 2015. Identification of a metastasis-specific MicroRNA signature in human colorectal cancer. J Natl Cancer Inst, 107.

HWANG, E., JOHNSON-OBASEKI, S., MCDONALD, J. T., CONNELL, C. & CORSTEN, M. 2013. Incidence of head and neck cancer and socioeconomic status in Canada from 1992 to 2007. Oral Oncol, 49, 1072-6.

HYNES, R. O. 2003. Metastatic potential: generic predisposition of the primary tumor or rare, metastatic variants-or both? Cell, 113, 821-3.

INDIA PROJECT TEAM OF THE INTERNATIONAL CANCER GENOME, C. 2013. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nat Commun, 4, 2873.

INGVARSSON, S. 2005. Tumor suppressor genes on chromosome 3 and cancer pathogenesis. Cancer Genomics & Proteomics, 2, 247-54.

IVANOVSKA, I., BALL, A. S., DIAZ, R. L., MAGNUS, J. F., KIBUKAWA, M., SCHELTER, J. M., et al. 2008. MicroRNAs in the miR-106b family regulate

p21/CDKN1A and promote cell cycle progression. Mol Cell Biol, 28, 2167-74.

IWAMOTO, K. S., MIZUNO, T., ITO, T., AKIYAMA, M., TAKEICHI, N., MABUCHI, K., et al. 1996. Feasibility of using decades-old archival tissues in molecular oncology/epidemiology. Am J Pathol, 149, 399-406.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet, 33 Suppl, 245-54.

JANOT, F., KLIJANIENKO, J., RUSSO, A., MAMET, J. P., DE BRAUD, F., EL-NAGGAR, A. K., et al. 1996. Prognostic value of clinicopathological parameters in head and neck squamous cell carcinoma: a prospective analysis. Br J Cancer, 73, 531-8.

JANSSON, M. D. & LUND, A. H. 2012. MicroRNA and cancer. Mol Oncol, 6, 590-610.

JARVINEN, A. K., AUTIO, R., KILPINEN, S., SAARELA, M., LEIVO, I., GRENMAN, R., et al. 2008. High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. Genes Chromosomes Cancer, 47, 500-9.

JIANG, L., LIU, X., CHEN, Z., JIN, Y., HEIDBREDER, C. E., KOLOKYTHAS, A., et al. 2010a. MicroRNA-7 targets IGF1R (insulin-like growth factor 1 receptor) in tongue squamous cell carcinoma cells. Biochem J, 432, 199-205.

JIANG, Y., CHEN, H., JIA, H., XU, Y., LIU, G., WANG, Y., et al. 2010. Adenovirus Ad-p53AIP1-mediated gene therapy and its regulation of p53-MDM2 interactions. Exp Ther Med, 1, 363-368.

JIANG, Y., CHEN, H., JIA, H., XU, Y., LIU, G., WANG, Y., et al. 2010b. Adenovirus Ad-p53AIP1-mediated gene therapy and its regulation of p53-MDM2 interactions. Exp Ther Med, 1, 363-368.

JIN, L., WESSELY, O., MARCUSSON, E. G., IVAN, C., CALIN, G. A. & ALAHARI, S. K. 2013. Prooncogenic factors miR-23b and miR-27b are

regulated by Her2/Neu, EGF, and TNF-alpha in breast cancer. Cancer Res, 73, 2884-96.

JOHNSON, J. T., BARNES, E. L., MYERS, E. N., SCHRAMM, V. L., JR., BOROCHOVITZ, D. & SIGLER, B. A. 1981. The extracapsular spread of tumors in cervical node metastasis. Arch Otolaryngol, 107, 725-9.

JOHNSON, S. M., GROSSHANS, H., SHINGARA, J., BYROM, M., JARVIS, R., CHENG, A., et al. 2005. RAS is regulated by the let-7 microRNA family. Cell, 120, 635-47.

JONES, P. A. & BAYLIN, S. B. 2002. The fundamental role of epigenetic events in cancer. Nat Rev Genet, 3, 415-28.

JONES, S., ZHANG, X., PARSONS, D. W., LIN, J. C., LEARY, R. J., ANGENENDT, P., et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science, 321, 1801-6.

JONSSON, G., STAAF, J., VALLON-CHRISTERSSON, J., RINGNER, M., HOLM, K., HEGARDT, C., et al. 2010. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res, 12, R42.

JOO, Y. H., PARK, S. W., JUNG, S. H., LEE, Y. S., NAM, I. C., CHO, K. J., et al. 2013a. Recurrent loss of the FHIT gene and its impact on lymphatic metastasis in early oral squamous cell carcinoma. Acta Otolaryngol, 133, 992-9.

JOO, Y. H., YOO IE, R., CHO, K. J., PARK, J. O., NAM, I. C. & KIM, M. S. 2013c. Extracapsular spread in hypopharyngeal squamous cell carcinoma: diagnostic value of FDG PET/CT. Head Neck, 35, 1771-6.

JOO, Y. H., YOO IE, R., CHO, K. J., PARK, J. O., NAM, I. C., KIM, C. S., et al. 2013b. Relationship between extracapsular spread and FDG PET/CT in oropharyngeal squamous cell carcinoma. Acta Otolaryngol, 133, 1073-9.

JOO, Y. H., YOO, I. R., CHO, K. J., PARK, J. O., NAM, I. C. & KIM, M. S. 2013d. Extracapsular spread and FDG PET/CT correlations in oral squamous cell carcinoma. Int J Oral Maxillofac Surg, 42, 158-63.

JOSE, J., COATESWORTH, A. P., JOHNSTON, C. & MACLENNAN, K. 2003. Cervical node metastases in squamous cell carcinoma of the upper aerodigestive tract: the significance of extracapsular spread and soft tissue deposits. Head Neck, 25, 451-6.

JR, S. S. 2003. Complications of treatment. In: JR, S. S. (ed.) Oral Cancer. Canada: BC Decker Inc.

JUNG, H. M., PATEL, R. S., PHILLIPS, B. L., WANG, H., COHEN, D. M., REINHOLD, W. C., et al. 2013. Tumor suppressor miR-375 regulates MYC expression via repression of CIP2A coding sequence through multiple miRNA-mRNA interactions. Mol Biol Cell, 24, 1638-48, S1-7.

JUNG, H. M., PHILLIPS, B. L., PATEL, R. S., COHEN, D. M., JAKYMIW, A., KONG, W. W., et al. 2012. Keratinization-associated miR-7 and miR-21 regulate tumor suppressor reversion-inducing cysteine-rich protein with kazal motifs (RECK) in oral cancer. J Biol Chem, 287, 29261-72.

KALINOWSKI, F. C., BROWN, R. A., GANDA, C., GILES, K. M., EPIS, M. R., HORSHAM, J., et al. 2014. microRNA-7: a tumor suppressor miRNA with therapeutic potential. Int J Biochem Cell Biol, 54, 312-7.

KALINOWSKI, F. C., GILES, K. M., CANDY, P. A., ALI, A., GANDA, C., EPIS, M. R., et al. 2012. Regulation of epidermal growth factor receptor signaling and erlotinib sensitivity in head and neck cancer cells by miR-7. PLoS One, 7, e47067.

KALLIONIEMI, O. P., KALLIONIEMI, A., PIPER, J., ISOLA, J., WALDMAN, F. M., GRAY, J. W., et al. 1994. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. Genes Chromosomes Cancer, 10, 231-43.

KALNINS, I. K., LEONARD, A. G., SAKO, K., RAZACK, M. S. & SHEDD, D. P. 1977. Correlation between prognosis and degree of lymph node involvement in carcinoma of the oral cavity. Am J Surg, 134, 450-4.

KEFAS, B., GODLEWSKI, J., COMEAU, L., LI, Y., ABOUNADER, R., HAWKINSON, M., et al. 2008. microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. Cancer Res, 68, 3566-72.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M., et al. 2002. The human genome browser at UCSC. Genome Res, 12, 996-1006.

KIKKAWA, N., HANAZAWA, T., FUJIMURA, L., NOHATA, N., SUZUKI, H., CHAZONO, H., et al. 2010. miR-489 is a tumour-suppressive miRNA target PTPN11 in hypopharyngeal squamous cell carcinoma (HSCC). Br J Cancer, 103, 877-84.

KIM, J. M., KIM, H. J., KOO, B. S., RHA, K. S. & YOON, Y. H. 2013. Expression of matrix metalloproteinase-12 is correlated with extracapsular spread of tumor from nodes with metastasis in head and neck squamous cell carcinoma. Eur Arch Otorhinolaryngol, 270, 1137-42.

KIMURA, Y., SUMI, M., SAKIHAMA, N., TANAKA, F., TAKAHASHI, H. & NAKAMURA, T. 2008. MR imaging criteria for the prediction of extranodal spread of metastatic cancer in the neck. AJNR Am J Neuroradiol, 29, 1355-9.

KING, A. D., TSE, G. M., YUEN, E. H., TO, E. W., VLANTIS, A. C., ZEE, B., et al. 2004. Comparison of CT and MR imaging for the detection of extranodal neoplastic spread in metastatic neck nodes. Eur J Radiol, 52, 264-70.

KJERSEM, J. B., IKDAHL, T., LINGJAERDE, O. C., GUREN, T., TVEIT, K. M. & KURE, E. H. 2014. Plasma microRNAs predicting clinical outcome in metastatic colorectal cancer patients receiving first-line oxaliplatin-based treatment. Mol Oncol, 8, 59-67.

KLEIN, C. A. 2009. Parallel progression of primary tumours and metastases. Nat Rev Cancer, 9, 302-12.

KLIGERMAN, J., LIMA, R. A., SOARES, J. R., PRADO, L., DIAS, F. L., FREITAS, E. Q., et al. 1994. Supraomohyoid neck dissection in the treatment of T1/T2 squamous cell carcinoma of oral cavity. Am J Surg, 168, 391-4.

KNUTSEN, E., FISKAA, T., URSVIK, A., JORGENSEN, T. E., PERANDER, M., LUND, E., et al. 2013. Performance comparison of digital microRNA

profiling technologies applied on human breast cancer cell lines. PLoS One, 8, e75813.

KOLBERT, C. P., FEDDERSEN, R. M., RAKHSHAN, F., GRILL, D. E., SIMON, G., MIDDHA, S., et al. 2013. Multi-platform analysis of microRNA expression measurements in RNA from fresh frozen and FFPE tissues. PLoS One, 8, e52517.

KORC, M. 2009. Smad4: gatekeeper gene in head and neck squamous cell carcinoma. J Clin Invest, 119, 3208-11.

KOZAKI, K., IMOTO, I., MOGI, S., OMURA, K. & INAZAWA, J. 2008. Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer. Cancer Res, 68, 2094-105.

KRISHNA RAO, S. V., MEJIA, G., ROBERTS-THOMSON, K. & LOGAN, R. 2013. Epidemiology of oral cancer in Asia in the past decade--an update (2000-2012). Asian Pac J Cancer Prev, 14, 5567-77.

KRISHNAN, A., NAIR, S. A. & PILLAI, M. R. 2007. Biology of PPAR gamma in cancer: a critical review on existing lacunae. Curr Mol Med, 7, 532-40.

KUJAWSKI, M., SARLOMO-RIKALA, M., GABRIEL, A., SZYFTER, K. & KNUUTILA, S. 1999. Recurrent DNA copy number losses associated with metastasis of larynx carcinoma. Genes Chromosomes Cancer, 26, 253-7.

KYZAS, P. A., CUNHA, I. W. & IOANNIDIS, J. P. 2005a. Prognostic significance of vascular endothelial growth factor immunohistochemical expression in head and neck squamous cell carcinoma: a meta-analysis. Clin Cancer Res, 11, 1434-40.

KYZAS, P. A., STEFANOU, D., BATISTATOU, A. & AGNANTIS, N. J. 2005b. Potential autocrine function of vascular endothelial growth factor in head and neck cancer via vascular endothelial growth factor receptor-2. Mod Pathol, 18, 485-94.

LAGOS-QUINTANA, M., RAUHUT, R., LENDECKEL, W. & TUSCHL, T. 2001. Identification of novel genes coding for small expressed RNAs. Science, 294, 853-8.

LAI, K. W., KOH, K. X., LOH, M., TADA, K., SUBRAMANIAM, M. M., LIM, X. Y., et al. 2010. MicroRNA-130b regulates the tumour suppressor RUNX3 in gastric cancer. Eur J Cancer, 46, 1456-63.

LAI, S. Y. & JOHNSON, F. M. 2010. Defining the role of the JAK-STAT pathway in head and neck and thoracic malignancies: implications for future therapeutic approaches. Drug Resist Updat, 13, 67-78.

LAI, S. Y., CHILDS, E. E., XI, S., COPPELLI, F. M., GOODING, W. E., WELLS, A., et al. 2005. Erythropoietin-mediated activation of JAK-STAT signaling contributes to cellular invasion in head and neck squamous cell carcinoma. Oncogene, 24, 4442-9.

LAJER, C. B., GARNAES, E., FRIIS-HANSEN, L., NORRILD, B., THERKILDSEN, M. H., GLUD, M., et al. 2012. The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer. Br J Cancer, 106, 1526-34.

LAJER, C. B., NIELSEN, F. C., FRIIS-HANSEN, L., NORRILD, B., BORUP, R., GARNAES, E., et al. 2011. Different miRNA signatures of oral and pharyngeal squamous cell carcinomas: a prospective translational study. Br J Cancer, 104, 830-40.

LANDI, M. T., DRACHEVA, T., ROTUNNO, M., FIGUEROA, J. D., LIU, H., DASGUPTA, A., et al. 2008. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One, 3, e1651.

LANGENDIJK, J. A. & PSYRRI, A. 2010. The prognostic significance of p16 overexpression in oropharyngeal squamous cell carcinoma: implications for treatment strategies and future clinical studies. Ann Oncol, 21, 1931-4.

LANZER, M., GANDER, T., KRUSE, A., LUEBBERS, H. T. & REINISCH, S. 2014. Influence of histopathologic factors on pattern of metastasis in squamous cell carcinoma of the head and neck. Laryngoscope, 124, E160-6.

LASSEN, P., PRIMDAHL, H., JOHANSEN, J., KRISTENSEN, C. A., ANDERSEN, E., ANDERSEN, L. J., et al. 2014. Impact of HPV-associated

p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol, 113, 310-6.

LAU, P. C. & CHAN, A. T. 2011. Novel therapeutic target for head and neck squamous cell carcinoma: HGF-MET signaling pathway. Anticancer Drugs, 22, 665-73.

LAURILA, E. M., SANDSTROM, S., RANTANEN, L. M., AUTIO, R. & KALLIONIEMI, A. 2012. Both inhibition and enhanced expression of miR-31 lead to reduced migration and invasion of pancreatic cancer cells. Genes Chromosomes Cancer, 51, 557-68.

LEARY, R. J., LIN, J. C., CUMMINS, J., BOCA, S., WOOD, L. D., PARSONS, D. W., et al. 2008. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A, 105, 16224-9.

LECHNER, M., FRAMPTON, G. M., FENTON, T., FEBER, A., PALMER, G., JAY, A., et al. 2013. Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. Genome Med, 5, 49.

LEE, R. C., FEINBAUM, R. L. & AMBROS, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 75, 843-54.

LEE, S. H., KOO, B. S., KIM, J. M., HUANG, S., RHO, Y. S., BAE, W. J., et al. 2014. Wnt/beta-catenin signalling maintains self-renewal and tumourigenicity of head and neck squamous cell carcinoma stem-like cells by activating Oct4. J Pathol, 234, 99-107.

LEELAHAVANICHKUL, K., AMORNPHIMOLTHAM, P., MOLINOLO, A. A., BASILE, J. R., KOONTONGKAEW, S. & GUTKIND, J. S. 2014. A role for p38 MAPK in head and neck cancer cell growth and tumor-induced angiogenesis and lymphangiogenesis. Mol Oncol, 8, 105-18.

LEEMANS, C. R., BRAAKHUIS, B. J. & BRAKENHOFF, R. H. 2011. The molecular biology of head and neck cancer. Nat Rev Cancer, 11, 9-22.

LEWIN, F., NORELL, S. E., JOHANSSON, H., GUSTAVSSON, P., WENNERBERG, J., BIORKLUND, A., et al. 1998. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: a population-based case-referent study in Sweden. Cancer, 82, 1367-75.

LEWIS, F., MAUGHAN, N. J., SMITH, V., HILLAN, K. & QUIRKE, P. 2001. Unlocking the archive--gene expression in paraffin-embedded tissue. J Pathol, 195, 66-71.

LEWIS, J. S., JR., CARPENTER, D. H., THORSTAD, W. L., ZHANG, Q. & HAUGHEY, B. H. 2011. Extracapsular extension is a poor predictor of disease recurrence in surgically treated oropharyngeal squamous cell carcinoma. Mod Pathol, 24, 1413-20.

LI, B. L., LU, C., LU, W., YANG, T. T., QU, J., HONG, X., et al. 2013. miR-130b is an EMT-related microRNA that targets DICER1 for aggression in endometrial cancer. Med Oncol, 30, 484.

LI H., DURBIN R. 2010. Fast and accurate long read alignment with Burrows-Wheeler transform. Bioinformatics, 26, 589-595.

LI, J., HUANG, H., SUN, L., YANG, M., PAN, C., CHEN, W., et al. 2009. MiR-21 indicates poor prognosis in tongue squamous cell carcinomas as an apoptosis inhibitor. Clin Cancer Res, 15, 3998-4008.

LI, J., YOU, T. & JING, J. 2014. MiR-125b inhibits cell biological progression of Ewing's sarcoma by suppressing the PI3K/Akt signalling pathway. Cell Prolif, 47, 152-60.

LIAO, L. J., LO, W. C., HSU, W. L., WANG, C. T. & LAI, M. S. 2012. Detection of cervical lymph node metastasis in head and neck cancer patients with clinically N0 neck-a meta-analysis comparing different imaging modalities. BMC Cancer, 12, 236.

LICITRA, L., ROSSINI, C., BOSSI, P. & LOCATI, L. D. 2006. Advances in the changing patterns of aetiology of head and neck cancers. Curr Opin Otolaryngol Head Neck Surg, 14, 95-9.

LIEU, K. G., SHIM, E. H., WANG, J., LOKAREDDY, R. K., TAO, T., CINGOLANI, G., et al. 2014. The p53-induced factor Ei24 inhibits nuclear import through an importin beta-binding-like domain. J Cell Biol, 205, 301-12.

LIN, J. C., WANG, W. Y., CHEN, K. Y., WEI, Y. H., LIANG, W. M., JAN, J. S., et al. 2004. Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. N Engl J Med, 350, 2461-70.

LIN, S. C., LIU, C. J., KO, S. Y., CHANG, H. C., LIU, T. Y. & CHANG, K. W. 2005. Copy number amplification of 3q26-27 oncogenes in microdissected oral squamous cell carcinoma and oral brushed samples from areca chewers. J Pathol, 206, 417-22.

LIU, A. & XU, X. 2011. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. Methods Mol Biol, 724, 259-67.

LIU, C. J., KAO, S. Y., TU, H. F., TSAI, M. M., CHANG, K. W. & LIN, S. C. 2010a. Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. Oral Dis, 16, 360-4.

LIU, C. J., LIN, S. C., YANG, C. C., CHENG, H. W. & CHANG, K. W. 2012. Exploiting salivary miR-31 as a clinical biomarker of oral squamous cell carcinoma. Head Neck, 34, 219-24.

LIU, C. J., TSAI, M. M., HUNG, P. S., KAO, S. Y., LIU, T. Y., WU, K. J., et al. 2010b. miR-31 ablates expression of the HIF regulatory factor FIH to activate the HIF pathway in head and neck carcinoma. Cancer Res, 70, 1635-44.

LIU, J., CHO, S. N., AKKANTI, B., JIN, N., MAO, J., LONG, W., et al. 2015a. ErbB2 Pathway Activation upon Smad4 Loss Promotes Lung Tumor Growth and Metastasis. Cell Rep.

LIU, W., XU, C., WAN, H., LIU, C., WEN, C., LU, H., et al. 2014. MicroRNA-206 overexpression promotes apoptosis, induces cell cycle arrest and inhibits the migration of human hepatocellular carcinoma HepG2 cells. Int J Mol Med, 34, 420-8.

LIU, X., JIANG, L., WANG, A., YU, J., SHI, F. & ZHOU, X. 2009. MicroRNA-138 suppresses invasion and promotes apoptosis in head and neck squamous cell carcinoma cell lines. Cancer Lett, 286, 217-22.

LIU, Y., SHENG, J., DAI, D., LIU, T. & QI, F. 2015b. Smad4 acts as tumor suppressor by antagonizing lymphangiogenesis in colorectal cancer. Pathol Res Pract, 211, 286-92.

LO, K. W., TO, K. F. & HUANG, D. P. 2004. Focus on nasopharyngeal carcinoma. Cancer Cell, 5, 423-428.

LODDER, W. L., LANGE, C. A., VAN VELTHUYSEN, M. L., HAUPTMANN, M., BALM, A. J., VAN DEN BREKEL, M. W., et al. 2013. Can extranodal spread in head and neck cancer be detected on MR imaging. Oral Oncol, 49, 626-33.

LODISH H, B. A., ZIPURSKY SL, ET AL. 2000. Processing of rRNA and tRNA. Molecular Cell Biology. 4th edition ed. New York: W H Freeman.

LOHAVANICHBUTR, P., HOUCK, J., FAN, W., YUEH, B., MENDEZ, E., FUTRAN, N., et al. 2009. Genomewide gene expression profiles of HPV-positive and HPV-negative oropharyngeal cancer: potential implications for treatment choices. Arch Otolaryngol Head Neck Surg, 135, 180-8.

LU, K., YE, W., ZHOU, L., COLLINS, L. B., CHEN, X., GOLD, A., et al. 2010. Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. J Am Chem Soc, 132, 3388-99.

LU, T. P., LAI, L. C., TSAI, M. H., CHEN, P. C., HSU, C. P., LEE, J. M., et al. 2011. Integrated analyses of copy number variations and gene expression in lung adenocarcinoma. PLoS One, 6, e24829.

LUBEK, J. E., DYALRAM, D., PERERA, E. H., LIU, X. & ORD, R. A. 2013. A retrospective analysis of squamous carcinoma of the buccal mucosa: an aggressive subsite within the oral cavity. J Oral Maxillofac Surg, 71, 1126-31.

LUI, V. W., HEDBERG, M. L., LI, H., VANGARA, B. S., PENDLETON, K., ZENG, Y., et al. 2013. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov, 3, 761-9.

MACLELLAN, S. A., LAWSON, J., BAIK, J., GUILLAUD, M., POH, C. F. & GARNIS, C. 2012. Differential expression of miRNAs in the serum of patients with high-risk oral lesions. Cancer Med, 1, 268-74.

MAESTRO, R., GASPAROTTO, D., VUKOSAVLJEVIC, T., BARZAN, L., SULFARO, S. & BOIOCCHI, M. 1993. Three discrete regions of deletion at 3p in head and neck cancers. Cancer Res, 53, 5775-9.

MAJID, S., DAR, A. A., SAINI, S., DENG, G., CHANG, I., GREENE, K., et al. 2013. MicroRNA-23b functions as a tumor suppressor by regulating Zeb1 in bladder cancer. PLoS One, 8, e67686.

MAMELLE, G., PAMPURIK, J., LUBOINSKI, B., LANCAR, R., LUSINCHI, A. & BOSQ, J. 1994. Lymph node prognostic factors in head and neck squamous cell carcinomas. Am J Surg, 168, 494-8.

MAO, L., HONG, W. K. & PAPADIMITRAKOPOULOU, V. A. 2004. Focus on head and neck cancer. Cancer Cell, 5, 311-6.

MAR-AGUILAR, F., MENDOZA-RAMIREZ, J. A., MALAGON-SANTIAGO, I., ESPINO-SILVA, P. K., SANTUARIO-FACIO, S. K., RUIZ-FLORES, P., et al. 2013. Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. Dis Markers, 34, 163-9.

MARSHALL, M. E., HINZ, T. K., KONO, S. A., SINGLETON, K. R., BICHON, B., WARE, K. E., et al. 2011. Fibroblast growth factor receptors are components of autocrine signaling networks in head and neck squamous cell carcinoma cells. Clin Cancer Res, 17, 5016-25.

MARTINEZ, I., WANG, J., HOBSON, K. F., FERRIS, R. L. & KHAN, S. A. 2007. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. Eur J Cancer, 43, 415-32.

MASHBERG, A. & MEYERS, H. 1976. Anatomical site and size of 222 early asymptomatic oral squamous cell carcinomas: a continuing prospective study of oral cancer. II. Cancer, 37, 2149-57.

MASSAGUE, J. 2008. TGFbeta in Cancer. Cell, 134, 215-30.

MATSUDA, K., YOSHIDA, K., TAYA, Y., NAKAMURA, K., NAKAMURA, Y. & ARAKAWA, H. 2002. p53AIP1 regulates the mitochondrial apoptotic pathway. Cancer Res, 62, 2883-9.

MAZZONE, M. & COMOGLIO, P. M. 2006. The Met pathway: master switch and drug target in cancer progression. FASEB J, 20, 1611-21.

MCCARROLL, S. A., HADNOTT, T. N., PERRY, G. H., SABETI, P. C., ZODY, M. C., BARRETT, J. C., et al. 2006. Common deletion polymorphisms in the human genome. Nat Genet, 38, 86-92.

MCGRANAHAN, N. & SWANTON, C. 2015. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell, 27, 15-26.

MEDINA, P. P., NOLDE, M. & SLACK, F. J. 2010. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. Nature, 467, 86-90.

MEHANNA, H., BEECH, T., NICHOLSON, T., EL-HARIRY, I., MCCONKEY, C., PALERI, V., et al. 2013. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region. Head Neck, 35, 747-55.

MEHTA, K. R., NAKAO, K., ZURAEK, M. B., RUAN, D. T., BERGSLAND, E. K., VENOOK, A. P., et al. 2005. Fractional genomic alteration detected by array-based comparative genomic hybridization independently predicts survival after hepatic resection for metastatic colorectal cancer. Clin Cancer Res, 11, 1791-7.

MENG, F., WEHBE-JANEK, H., HENSON, R., SMITH, H. & PATEL, T. 2008. Epigenetic regulation of microRNA-370 by interleukin-6 in malignant human cholangiocytes. Oncogene, 27, 378-86.

MENG, W., YE, Z., CUI, R., PERRY, J., DEDOUSI-HUEBNER, V., HUEBNER, A., et al. 2013. MicroRNA-31 predicts the presence of lymph node metastases and survival in patients with lung adenocarcinoma. Clin Cancer Res, 19, 5423-33.

MENVIELLE, G., LUCE, D., GOLDBERG, P. & LECLERC, A. 2004b. Smoking, alcohol drinking, occupational exposures and social inequalities in hypopharyngeal and laryngeal cancer. Int J Epidemiol, 33, 799-806.

MENVIELLE, G., LUCE, D., GOLDBERG, P., BUGEL, I. & LECLERC, A. 2004a. Smoking, alcohol drinking and cancer risk for various sites of the larynx and hypopharynx. A case-control study in France. Eur J Cancer Prev, 13, 165-72.

MERCHANT, A., HUSAIN, S. S., HOSAIN, M., FIKREE, F. F., PITIPHAT, W., SIDDIQUI, A. R., et al. 2000. Paan without tobacco: an independent risk factor for oral cancer. Int J Cancer, 86, 128-31.

MERMEL, C. H., SCHUMACHER, S. E., HILL, B., MEYERSON, M. L., BEROUKHIM, R. & GETZ, G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol, 12, R41.

MICHEL, C., DESDOUETS, C., SACRE-SALEM, B., GAUTIER, J. C., ROBERTS, R. & BOITIER, E. 2003. Liver gene expression profiles of rats treated with clofibric acid: comparison of whole liver and laser capture microdissected liver. Am J Pathol, 163, 2191-9.

MICHIKAWA, C., UZAWA, N., KAYAMORI, K., SONODA, I., OHYAMA, Y., OKADA, N., et al. 2012. Clinical significance of lymphatic and blood vessel invasion in oral tongue squamous cell carcinomas. Oral Oncol, 48, 320-4.

MICHIKAWA, C., UZAWA, N., SATO, H., OHYAMA, Y., OKADA, N. & AMAGASA, T. 2011. Epidermal growth factor receptor gene copy number aberration at the primary tumour is significantly associated with extracapsular spread in oral cancer. Br J Cancer, 104, 850-5.

MILLER, M. E., PALLA, B., CHEN, Q., ELASHOFF, D. A., ABEMAYOR, E., ST JOHN, M. A., et al. 2012. A novel classification system for perineural

invasion in noncutaneous head and neck squamous cell carcinoma: histologic subcategories and patient outcomes. Am J Otolaryngol, 33, 212-5.

MINNA, J. D., KURIE, J. M. & JACKS, T. 2003. A big step in the study of small cell lung cancer. Cancer Cell, 4, 163-6.

MISHRA, P. J. 2014. MicroRNAs as promising biomarkers in cancer diagnostics. Biomark Res, 2, 19.

MIYAGUCHI, K., UZAWA, N., MOGUSHI, K., TAKAHASHI, K., MICHIKAWA, C., NAKATA, Y., et al. 2012. Loss of NKX3-1 as a potential marker for an increased risk of occult lymph node metastasis and poor prognosis in oral squamous cell carcinoma. Int J Oncol, 40, 1907-14.

MONTEIRO, J., GASPAR, C., RICHER, W., FRANKEN, P. F., SACCHETTI, A., JOOSTEN, R., et al. 2014. Cancer stemness in Wnt-driven mammary tumorigenesis. Carcinogenesis, 35, 2-13.

MORROGH, M., OLVERA, N., BOGOMOLNIY, F., BORGEN, P. I. & KING, T. A. 2007. Tissue preparation for laser capture microdissection and RNA extraction from fresh frozen breast tissue. Biotechniques, 43, 41-2, 44, 46 passim.

MOTOYAMA, K., INOUE, H., TAKATSUNO, Y., TANAKA, F., MIMORI, K., UETAKE, H., et al. 2009. Over- and under-expressed microRNAs in human colorectal cancer. Int J Oncol, 34, 1069-75.

MROZ, E. A. & ROCCO, J. W. 2013. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol, 49, 211-5.

MROZ, E. A., TWARD, A. D., HAMMON, R. J., REN, Y. & ROCCO, J. W. 2015. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. PLoS Med, 12, e1001786.

MROZ, E. A., TWARD, A. D., PICKERING, C. R., MYERS, J. N., FERRIS, R. L. & ROCCO, J. W. 2013. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. Cancer, 119, 3034-42.

MULLER, D., MILLON, R., VELTEN, M., BRONNER, G., JUNG, G., ENGELMANN, A., et al. 1997. Amplification of 11q13 DNA markers in head and neck squamous cell carcinomas: correlation with clinical outcome. Eur J Cancer, 33, 2203-10.

MULLIGHAN, C. G., GOORHA, S., RADTKE, I., MILLER, C. B., COUSTAN-SMITH, E., DALTON, J. D., et al. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature, 446, 758-64.

MYLLYKANGAS, S., JUNNILA, S., KOKKOLA, A., AUTIO, R., SCHEININ, I., KIVILUOTO, T., et al. 2008. Integrated gene copy number and expression microarray analysis of gastric cancer highlights potential target genes. Int J Cancer, 123, 817-25.

MYO, K., UZAWA, N., MIYAMOTO, R., SONODA, I., YUKI, Y. & AMAGASA, T. 2005. Cyclin D1 gene numerical aberration is a predictive marker for occult cervical lymph node metastasis in TNM Stage I and II squamous cell carcinoma of the oral cavity. Cancer, 104, 2709-16.

N. GASPARINI, T. L., S. ANDREOLI, L. LEMOS, J. LIMA, G. PEREIRA, L. SAKAMOTO, R. PEREIRA, R. POGUE, R. ANDRADE. Copy number variation and the expression of miRNAs in colorectal cancer. American Society of Human Genetics Annual Conference, 2013 Boston, USA.: ASHG.

NAKANISHI, H., TACCIOLI, C., PALATINI, J., FERNANDEZ-CYMERING, C., CUI, R., KIM, T., et al. 2014. Loss of miR-125b-1 contributes to head and neck cancer development by dysregulating TACSTD2 and MAPK pathway. Oncogene, 33, 702-12.

NGUYEN, D. Q., WEBBER, C. & PONTING, C. P. 2006. Bias of selection on human copy-number variants. PLoS Genet, 2, e20.

NIELSEN, K. V., EJLERTSEN, B., MOLLER, S., JORGENSEN, J. T., KNOOP, A., KNUDSEN, H., et al. 2008. The value of TOP2A gene copy number variation as a biomarker in breast cancer: Update of DBCG trial 89D. Acta Oncol, 47, 725-34.

NISHITA, M., HASHIMOTO, M. K., OGATA, S., LAURENT, M. N., UENO, N., SHIBUYA, H., et al. 2000. Interaction between Wnt and TGF-beta

signalling pathways during formation of Spemann's organizer. Nature, 403, 781-5.

NOBORI, T., MIURA, K., WU, D. J., LOIS, A., TAKABAYASHI, K. & CARSON, D. A. 1994. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. Nature, 368, 753-6.

NOONE, R. B., BONNER, H., JR., RAYMOND, S., BROWN, A. S., GRAHAM, W. P., 3RD & LEHR, H. B. 1974. Lymph node metastases in oral carcinoma. A correlation of histopathology with survival. Plast Reconstr Surg, 53, 158-66.

NOUTOMI, Y., OGA, A., UCHIDA, K., OKAFUJI, M., ITA, M., KAWAUCHI, S., et al. 2006. Comparative genomic hybridization reveals genetic progression of oral squamous cell carcinoma from dysplasia via two different tumourigenic pathways. J Pathol, 210, 67-74.

OJESINA, A. I., LICHTENSTEIN, L., FREEMAN, S. S., PEDAMALLU, C. S., IMAZ-ROSSHANDLER, I., PUGH, T. J., et al. 2014. Landscape of genomic alterations in cervical carcinomas. Nature, 506, 371-5.

OKADA, Y. 2010. Relationships of cervical lymph node metastasis to histopathological malignancy grade, tumor angiogenesis, and lymphatic invasion in tongue cancer. Odontology, 98, 153-9.

OLSEN, K. D., CARUSO, M., FOOTE, R. L., STANLEY, R. J., LEWIS, J. E., BUSKIRK, S. J., et al. 1994. Primary head and neck cancer. Histopathologic predictors of recurrence after neck dissection in patients with lymph node involvement. Arch Otolaryngol Head Neck Surg, 120, 1370-4.

OPITZ, O. G., SULIMAN, Y., HAHN, W. C., HARADA, H., BLUM, H. E. & RUSTGI, A. K. 2001. Cyclin D1 overexpression and p53 inactivation immortalize primary oral keratinocytes by a telomerase-independent mechanism. J Clin Invest, 108, 725-32.

PALMER, C., DIEHN, M., ALIZADEH, A. A. & BROWN, P. O. 2006. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics, 7, 115.

PAPAGIANNAKOPOULOS, T., SHAPIRO, A. & KOSIK, K. S. 2008. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. Cancer Res, 68, 8164-72.

PARFENOV, M., PEDAMALLU, C. S., GEHLENBORG, N., FREEMAN, S. S., DANILOVA, L., BRISTOW, C. A., et al. 2014. Characterization of HPV and host genome interactions in primary head and neck cancers. Proc Natl Acad Sci U S A, 111, 15544-9.

PARK, N. J., ZHOU, H., ELASHOFF, D., HENSON, B. S., KASTRATOVIC, D. A., ABEMAYOR, E., et al. 2009. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clin Cancer Res, 15, 5473-7.

PARK, R. W., KIM, T. M., KASIF, S. & PARK, P. J. 2015. Identification of rare germline copy number variations over-represented in five human cancer types. Mol Cancer, 14, 25.

PARSONS, D. W., JONES, S., ZHANG, X., LIN, J. C., LEARY, R. J., ANGENENDT, P., et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. Science, 321, 1807-12.

PASQUINELLI, A. E., REINHART, B. J., SLACK, F., MARTINDALE, M. Q., KURODA, M. I., MALLER, B., et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature, 408, 86-9.

PATMORE, H. S., ASHMAN, J. N., CAWKWELL, L., MACDONALD, A., STAFFORD, N. D. & GREENMAN, J. 2004. Can a genetic signature for metastatic head and neck squamous cell carcinoma be characterised by comparative genomic hybridisation? Br J Cancer, 90, 1976-82.

PATMORE, H. S., CAWKWELL, L., STAFFORD, N. D. & GREENMAN, J. 2005. Unraveling the chromosomal aberrations of head and neck squamous cell carcinoma: a review. Ann Surg Oncol, 12, 831-42.

PEKARSKY, Y., PALAMARCHUK, A., HUEBNER, K. & CROCE, C. M. 2002. FHIT as tumor suppressor: mechanisms and therapeutic opportunities. Cancer Biol Ther, 1, 232-6.

PELLEGRINO, L., STEBBING, J., BRAGA, V. M., FRAMPTON, A. E., JACOB, J., BULUWELA, L., et al. 2013. miR-23b regulates cytoskeletal remodeling, motility and metastasis by directly targeting multiple transcripts. Nucleic Acids Res, 41, 5400-12.

PETERS, L. J., GOEPFERT, H., ANG, K. K., BYERS, R. M., MAOR, M. H., GUILLAMONDEGUI, O., et al. 1993. Evaluation of the dose for postoperative radiation therapy of head and neck cancer: first report of a prospective randomized trial. Int J Radiat Oncol Biol Phys, 26, 3-11.

PICHLER, M. & CALIN, G. A. 2015. MicroRNAs in cancer: from developmental genes in worms to their clinical application in patients. Br J Cancer.

PICKERING, C. R., ZHANG, J., NESKEY, D. M., ZHAO, M., JASSER, S. A., WANG, J., et al. 2014. Squamous Cell Carcinoma of the Oral Tongue in Young Non-Smokers Is Genomically Similar to Tumors in Older Smokers. Clin Cancer Res.

PIENAAR, E., THERON, M., NELSON, M. & VILJOEN, H. J. 2006. A quantitative model of error accumulation during PCR amplification. Comput Biol Chem, 30, 102-11.

PINATEL, E. M., ORSO, F., PENNA, E., CIMINO, D., ELIA, A. R., CIRCOSTA, P., et al. 2014. miR-223 is a coordinator of breast cancer progression as revealed by bioinformatics predictions. PLoS One, 9, e84859.

PIXLEY, F. J. & STANLEY, E. R. 2004. CSF-1 regulation of the wandering macrophage: complexity in action. Trends Cell Biol, 14, 628-38.

POETA, M. L., MANOLA, J., GOLDWASSER, M. A., FORASTIERE, A., BENOIT, N., CALIFANO, J. A., et al. 2007. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. N Engl J Med, 357, 2552-61.

PRICE G, R. M., CROWTHER R, WIGHT R 2009. Profile of Head and Neck Cancers in England: Incidence, Mortality and Survival. In: UNIT, N. C. I. (ed.). Oxford: National Cancer Intelligence Network.

QIU, W., SCHONLEBEN, F., LI, X., HO, D. J., CLOSE, L. G., MANOLIDIS, S., et al. 2006. PIK3CA mutations in head and neck squamous cell carcinoma. Clin Cancer Res, 12, 1441-6.

QUAN, C. & GUO, X. 1998. [Analysis of Myc gene family in laryngeal cancer]. Zhonghua Er Bi Yan Hou Ke Za Zhi, 33, 273-5.

RAMDAS, L., GIRI, U., ASHORN, C. L., COOMBES, K. R., EL-NAGGAR, A., ANG, K. K., et al. 2009. miRNA expression profiles in head and neck squamous cell carcinoma and adjacent normal tissue. Head Neck, 31, 642-54.

RATHER, M. I., NAGASHRI, M. N., SWAMY, S. S., GOPINATH, K. S. & KUMAR, A. 2013. Oncogenic microRNA-155 down-regulates tumor suppressor CDC73 and promotes oral squamous cell carcinoma cell proliferation: implications for cancer therapeutics. J Biol Chem, 288, 608-18.

REIS, P. P., TOMENSON, M., CERVIGNE, N. K., MACHADO, J., JURISICA, I., PINTILIE, M., et al. 2010. Programmed cell death 4 loss increases tumor cell invasion and is regulated by miR-21 in oral squamous cell carcinoma. Mol Cancer, 9, 238.

RENTOFT, M., FAHLEN, J., COATES, P. J., LAURELL, G., SJOSTROM, B., RYDEN, P., et al. 2011. miRNA analysis of formalin-fixed squamous cell carcinomas of the tongue is affected by age of the samples. Int J Oncol, 38, 61-9.

RHEINWALD, J. G., HAHN, W. C., RAMSEY, M. R., WU, J. Y., GUO, Z., TSAO, H., et al. 2002. A two-stage, p16(INK4A)- and p53-dependent keratinocyte senescence mechanism that limits replicative potential independent of telomere status. Mol Cell Biol, 22, 5157-72.

RICHARD, J. M., SANCHO-GARNIER, H., MICHEAU, C., SARAVANE, D. & CACHIN, Y. 1987. Prognostic factors in cervical lymph node metastasis in upper respiratory and digestive tract carcinomas: study of 1,713 cases during a 15-year period. Laryngoscope, 97, 97-101.

RISCHIN, D., YOUNG, R. J., FISHER, R., FOX, S. B., LE, Q. T., PETERS, L. J., et al. 2010. Prognostic significance of p16INK4A and human

papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. J Clin Oncol, 28, 4142-8.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26, 139-40.

ROBSON, A. 2001. The management of the neck in squamous head and neck cancer. Clin Otolaryngol Allied Sci, 26, 157-61.

RODRIGO, J. P., GARCIA-CARRACEDO, D., GARCIA, L. A., MENENDEZ, S., ALLONCA, E., GONZALEZ, M. V., et al. 2009. Distinctive clinicopathological associations of amplification of the cortactin gene at 11q13 in head and neck squamous cell carcinomas. J Pathol, 217, 516-23.

RODRIGUEZ-CARUNCHIO, L., SOVERAL, I., STEENBERGEN, R. D., TORNE, A., MARTINEZ, S., FUSTE, P., et al. 2015. HPV-negative carcinoma of the uterine cervix: a distinct type of cervical cancer with poor prognosis. BJOG, 122, 119-27.

ROEPMAN, P., WESSELS, L. F., KETTELARIJ, N., KEMMEREN, P., MILES, A. J., LIJNZAAD, P., et al. 2005. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. Nat Genet, 37, 182-6.

ROLAND, N. J., CASLIN, A. W., NASH, J. & STELL, P. M. 1992. Value of grading squamous cell carcinoma of the head and neck. Head Neck, 14, 224-9.

ROTHENBERG, S. M. & ELLISEN, L. W. 2012. The molecular pathogenesis of head and neck squamous cell carcinoma. J Clin Invest, 122, 1951-7.

RUAS, M., BROOKES, S., MCDONALD, N. Q. & PETERS, G. 1999. Functional evaluation of tumour-specific variants of p16INK4a/CDKN2A: correlation with protein structure information. Oncogene, 18, 5423-34.

SAITO, M., SCHETTER, A. J., MOLLERUP, S., KOHNO, T., SKAUG, V., BOWMAN, E. D., et al. 2011. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung

adenocarcinoma: a retrospective analysis of three cohorts. Clin Cancer Res, 17, 1875-82.

SAMMAN, M., WOOD, H. M., CONWAY, C., STEAD, L., DALY, C., CHALKLEY, R., et al. 2015. A novel genomic signature reclassifies an oral cancer subtype. Int J Cancer.

SAMUELS, Y., WANG, Z., BARDELLI, A., SILLIMAN, N., PTAK, J., SZABO, S., et al. 2004. High frequency of mutations of the PIK3CA gene in human cancers. Science, 304, 554.

SANKUNNY, M., PARIKH, R. A., LEWIS, D. W., GOODING, W. E., SAUNDERS, W. S. & GOLLIN, S. M. 2014. Targeted inhibition of ATR or CHEK1 reverses radioresistance in oral squamous cell carcinoma cells with distal chromosome arm 11q loss. Genes Chromosomes Cancer, 53, 129-43.

SATHYAN, K. M., SAILASREE, R., JAYASURYA, R., LAKSHMINARAYANAN, K., ABRAHAM, T., NALINAKUMARI, K. R., et al. 2006. Carcinoma of tongue and the buccal mucosa represent different biological subentities of the oral carcinoma. J Cancer Res Clin Oncol, 132, 601-9.

SAWEY, E. T., CHANRION, M., CAI, C., WU, G., ZHANG, J., ZENDER, L., et al. 2011. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell, 19, 347-58.

SAYAN, A. E., GRIFFITHS, T. R., PAL, R., BROWNE, G. J., RUDDICK, A., YAGCI, T., et al. 2009. SIP1 protein protects cells from DNA damage-induced apoptosis and has independent prognostic value in bladder cancer. Proc Natl Acad Sci U S A, 106, 14884-9.

SCAPOLI, L., PALMIERI, A., LO MUZIO, L., PEZZETTI, F., RUBINI, C., GIRARDI, A., et al. 2010. MicroRNA expression profiling of oral carcinoma identifies new markers of tumor progression. Int J Immunopathol Pharmacol, 23, 1229-34.

SCHAEFER, A., JUNG, M., MOLLENKOPF, H. J., WAGNER, I., STEPHAN, C., JENTZMIK, F., et al. 2010a. Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. Int J Cancer, 126, 1166-76.

SCHAEFER, A., STEPHAN, C., BUSCH, J., YOUSEF, G. M. & JUNG, K. 2010. Diagnostic, prognostic and therapeutic implications of microRNAs in urologic tumors. Nat Rev Urol, 7, 286-97.

SCHAEFER, A., STEPHAN, C., BUSCH, J., YOUSEF, G. M. & JUNG, K. 2010b. Diagnostic, prognostic and therapeutic implications of microRNAs in urologic tumors. Nat Rev Urol, 7, 286-97.

SCHLECHT, N. F., BURK, R. D., ADRIEN, L., DUNNE, A., KAWACHI, N., SARTA, C., et al. 2007. Gene expression profiles in HPV-infected head and neck cancer. J Pathol, 213, 283-93.

SCHROEDER, M. P., RUBIO-PEREZ, C., TAMBORERO, D., GONZALEZ-PEREZ, A. & LOPEZ-BIGAS, N. 2014. OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. Bioinformatics, 30, i549-55.

SCHULLER, D. E., MCGUIRT, W. F., MCCABE, B. F. & YOUNG, D. 1980. The prognostic significance of metastatic cervical lymph nodes. Laryngoscope, 90, 557-70.

SCHUTTE, M., HRUBAN, R. H., HEDRICK, L., CHO, K. R., NADASDY, G. M., WEINSTEIN, C. L., et al. 1996. DPC4 gene in various tumor types. Cancer Res, 56, 2527-30.

SCHWEIGER, M. R., KERICK, M., TIMMERMANN, B., ALBRECHT, M. W., BORODINA, T., PARKHOMCHUK, D., et al. 2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS One, 4, e5548.

SCOTT, G. K., GOGA, A., BHAUMIK, D., BERGER, C. E., SULLIVAN, C. S. & BENZ, C. C. 2007. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. J Biol Chem, 282, 1479-86.

SEIWERT, T. Y., ZUO, Z., KECK, M. K., KHATTRI, A., PEDAMALLU, C. S., STRICKER, T., et al. 2015. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. Clin Cancer Res, 21, 632-41.

SESSIONS, D. G. 1976. Surgical pathology of cancer of the larynx and hypopharynx. Laryngoscope, 86, 814-39.

SETHI, N., WRIGHT, A., WOOD, H. & RABBITTS, P. 2014. MicroRNAs and head and neck cancer: reviewing the first decade of research. Eur J Cancer, 50, 2619-35.

SHAH, J. P. 1990. Patterns of cervical lymph node metastasis from squamous carcinomas of the upper aerodigestive tract. Am J Surg, 160, 405-9.

SHAH, J. P., CANDELA, F. C. & PODDAR, A. K. 1990. The patterns of cervical lymph node metastases from squamous carcinoma of the oral cavity. Cancer, 66, 109-13.

SHAH, J. P., CENDON, R. A., FARR, H. W. & STRONG, E. W. 1976. Carcinoma of the oral cavity. factors affecting treatment failure at the primary site and neck. Am J Surg, 132, 504-7.

SHAIN, A. H., YEH, I., KOVALYSHYN, I., SRIHARAN, A., TALEVICH, E., GAGNON, A., et al. 2015. The Genetic Evolution of Melanoma from Precursor Lesions. N Engl J Med, 373, 1926-36.

SHANKARAN, V., IKEDA, H., BRUCE, A. T., WHITE, J. M., SWANSON, P. E., OLD, L. J., et al. 2001. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature, 410, 1107-11.

SHARAN, R. N., MEHROTRA, R., CHOUDHURY, Y. & ASOTRA, K. 2012. Association of betel nut with carcinogenesis: revisit with a clinical perspective. PLoS One, 7, e42759.

SHARMA, H., SEN, S., LO MUZIO, L., MARIGGIO, A. & SINGH, N. 2005. Antisense-mediated downregulation of anti-apoptotic proteins induces apoptosis and sensitizes head and neck squamous cell carcinoma cells to chemotherapy. Cancer Biol Ther, 4, 720-7.

SHAW, R. J., LOWE, D., WOOLGAR, J. A., BROWN, J. S., VAUGHAN, E. D., EVANS, C., et al. 2010. Extracapsular spread in oral squamous cell carcinoma. Head Neck, 32, 714-22.

SHAW, R. J., MCGLASHAN, G., WOOLGAR, J. A., LOWE, D., BROWN, J. S., VAUGHAN, E. D., et al. 2009. Prognostic importance of site in squamous cell carcinoma of the buccal mucosa. Br J Oral Maxillofac Surg, 47, 356-9.

SHEU, J. J., HUA, C. H., WAN, L., LIN, Y. J., LAI, M. T., TSENG, H. C., et al. 2009. Functional genomic analysis identified epidermal growth factor receptor activation as the most common genetic event in oral squamous cell carcinoma. Cancer Res, 69, 2568-76.

SHIIBA, M., SHINOZUKA, K., SAITO, K., FUSHIMI, K., KASAMATSU, A., OGAWARA, K., et al. 2013. MicroRNA-125b regulates proliferation and radioresistance of oral squamous cell carcinoma. Br J Cancer, 108, 1817-21.

SILVA, S. D., ALAOUI-JAMALI, M. A., HIER, M., SOARES, F. A., GRANER, E. & KOWALSKI, L. P. 2014. Cooverexpression of ERBB1 and ERBB4 receptors predicts poor clinical outcome in pN+ oral squamous cell carcinoma with extranodal spread. Clin Exp Metastasis, 31, 307-16.

SIMS, D., SUDBERY, I., ILOTT, N. E., HEGER, A. & PONTING, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet, 15, 121-32.

SINGH, B., GOGINENI, S. K., SACKS, P. G., SHAHA, A. R., SHAH, J. P., STOFFEL, A., et al. 2001. Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. Cancer Res, 61, 4506-13.

SINGHI, A. D. & WESTRA, W. H. 2010. Comparison of human papillomavirus in situ hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. Cancer, 116, 2166-73.

SINHA, P., LEWIS, J. S., JR., PICCIRILLO, J. F., KALLOGJERI, D. & HAUGHEY, B. H. 2012. Extracapsular spread and adjuvant therapy in human papillomavirus-related, p16-positive oropharyngeal carcinoma. Cancer, 118, 3519-30.

SIOMI, H. & SIOMI, M. C. 2010. Posttranscriptional regulation of microRNA biogenesis in animals. Mol Cell, 38, 323-32.

SLAUGHTER, D. P., SOUTHWICK, H. W. & SMEJKAL, W. 1953. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer, 6, 963-8.

SLEBOS, R. J., YI, Y., ELY, K., CARTER, J., EVJEN, A., ZHANG, X., et al. 2006. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. Clin Cancer Res, 12, 701-9.

SMEETS, S. J., BRAAKHUIS, B. J., ABBAS, S., SNIJDERS, P. J., YLSTRA, B., VAN DE WIEL, M. A., et al. 2006. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. Oncogene, 25, 2558-64.

SMEETS, S. J., HESSELINK, A. T., SPEEL, E. J., HAESEVOETS, A., SNIJDERS, P. J., PAWLITA, M., et al. 2007. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. Int J Cancer, 121, 2465-72.

SMEETS, S. J., VAN DER PLAS, M., SCHAAIJ-VISSER, T. B., VAN VEEN, E. A., VAN MEERLOO, J., BRAAKHUIS, B. J., et al. 2011. Immortalization of oral keratinocytes by functional inactivation of the p53 and pRb pathways. Int J Cancer, 128, 1596-605.

SNOW, G. B., ANNYAS, A. A., VAN SLOOTEN, E. A., BARTELINK, H. & HART, A. A. 1982. Prognostic factors of neck node metastasis. Clin Otolaryngol Allied Sci, 7, 185-92.

SOBIN LH, G. M., WITTEKIND C. (ed.) 2009. TNM Claissification of Malignant Tumours: Wiley-Blackwell.

SONG, G., ZHANG, Y. & WANG, L. 2009. MicroRNA-206 targets notch3, activates apoptosis, and inhibits tumor cell migration and focus formation. J Biol Chem, 284, 31921-7.

SOO, K. C., CARTER, R. L., O'BRIEN, C. J., BARR, L., BLISS, J. M. & SHAW, H. J. 1986. Prognostic implications of perineural spread in squamous carcinomas of the head and neck. Laryngoscope, 96, 1145-8.

SOUTER, M. A., ALLISON, R. S., CLARKSON, J. H., COWAN, I. A., COATES, M. H. & WELLS, J. E. 2009. Sensitivity and specificity of computed tomography for detection of extranodal spread from metastatic head and neck squamous cell carcinoma. J Laryngol Otol, 123, 778-82.

SPIRO, R. H., ALFONSO, A. E., FARR, H. W. & STRONG, E. W. 1974. Cervical node metastasis from epidermoid carcinoma of the oral cavity and oropharynx. A critical assessment of current staging. Am J Surg, 128, 562-7.

SQUARIZE, C. H., CASTILHO, R. M., ABRAHAO, A. C., MOLINOLO, A., LINGEN, M. W. & GUTKIND, J. S. 2013. PTEN deficiency contributes to the development and progression of head and neck cancer. Neoplasia, 15, 461-71.

STANKIEWICZ, P. & BEAUDET, A. L. 2007. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev, 17, 182-92.

STANTA, G. & SCHNEIDER, C. 1991. RNA extracted from paraffin-embedded human tissues is amenable to analysis by PCR amplification. Biotechniques, 11, 304, 306, 308.

STATISTICS, O. F. N. 2012. Cancer Statistics Registrations, England (Series MB1) - No. 43. In: STATISTICS, O. F. N. (ed.).

STICHT, C., HOFELE, C., FLECHTENMACHER, C., BOSCH, F. X., FREIER, K., LICHTER, P., et al. 2005. Amplification of Cyclin L1 is associated with lymph node metastases in head and neck squamous cell carcinoma (HNSCC). Br J Cancer, 92, 770-4.

STOCKWELL, H. G. & LYMAN, G. H. 1986. Impact of smoking and smokeless tobacco on the risk of cancer of the head and neck. Head Neck Surg, 9, 104-10.

STOEHR, R., WILD, P. & HARTMANN, A. 2003. Lasermicrodissection--an important prerequisite for the molecular-genetic analysis of bladder cancer. Pathol Res Pract, 199, 355-62.

STRACHAN T, R. A. 1999. Human Molecular Genetics, John Wiley & Sons.

STRANGER, B. E., FORREST, M. S., DUNNING, M., INGLE, C. E., BEAZLEY, C., THORNE, N., et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science, 315, 848-53.

STRANSKY, N., EGLOFF, A. M., TWARD, A. D., KOSTIC, A. D., CIBULSKIS, K., SIVACHENKO, A., et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. Science, 333, 1157-60.

STRUSKI, S., DOCO-FENZY, M. & CORNILLET-LEFEBVRE, P. 2002. Compilation of published comparative genomic hybridization studies. Cancer Genet Cytogenet, 135, 63-90.

SUGAHARA, K., MICHIKAWA, Y., ISHIKAWA, K., SHOJI, Y., IWAKAWA, M., SHIBAHARA, T., et al. 2011. Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma. Int J Oncol, 39, 761-9.

SUH, S. S., YOO, J. Y., CUI, R., KAUR, B., HUEBNER, K., LEE, T. K., et al. 2014. FHIT suppresses epithelial-mesenchymal transition (EMT) and metastasis in lung cancer through modulation of microRNAs. PLoS Genet, 10, e1004652.

SUN, D., LEE, Y. S., MALHOTRA, A., KIM, H. K., MATECIC, M., EVANS, C., et al. 2011. miR-99 family of MicroRNAs suppresses the expression of prostate-specific antigen and prostate cancer cell proliferation. Cancer Res, 71, 1313-24.

SUN, W., GAYKALOVA, D. A., OCHS, M. F., MAMBO, E., ARNAOUTAKIS, D., LIU, Y., et al. 2014a. Activation of the NOTCH pathway in head and neck cancer. Cancer Res, 74, 1091-104.

SUN, Z., EVANS, J., BHAGWATE, A., MIDDHA, S., BOCKOL, M., YAN, H., et al. 2014. CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. BMC Genomics, 15, 423.

SUN, Z., EVANS, J., BHAGWATE, A., MIDDHA, S., BOCKOL, M., YAN, H., et al. 2014b. CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. BMC Genomics, 15, 423.

SUNG, C. O., CHOI, C. H., KO, Y. H., JU, H., CHOI, Y. L., KIM, N., et al. 2013. Integrative analysis of copy number alteration and gene expression profiling in ovarian clear cell adenocarcinoma. Cancer Genet, 206, 145-53.

SZAFRANSKA, A. E., DAVISON, T. S., SHINGARA, J., DOLESHAL, M., RIGGENBACH, J. A., MORRISON, C. D., et al. 2008. Accurate molecular characterization of formalin-fixed, paraffin-embedded tissues by microRNA expression profiling. J Mol Diagn, 10, 415-23.

TAI, S. K., LEE, J. I., ANG, K. K., EL-NAGGAR, A. K., HASSAN, K. A., LIU, D., et al. 2004. Loss of Fhit expression in head and neck squamous cell carcinoma and its potential clinical implication. Clin Cancer Res, 10, 5554-7.

TALMADGE, J. E. 2007. Clonal selection of metastasis within the life history of a tumor. Cancer Res, 67, 11471-5.

TAM, S., DE BORJA, R., TSAO, M. S. & MCPHERSON, J. D. 2014. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest, 94, 350-8.

TAMBORERO, D., GONZALEZ-PEREZ, A., PEREZ-LLAMAS, C., DEU-PONS, J., KANDOTH, C., REIMAND, J., et al. 2013. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep, 3, 2650.

TANG, F., ZHANG, R., HE, Y., ZOU, M., GUO, L. & XI, T. 2012. MicroRNA-125b induces metastasis by targeting STARD13 in MCF-7 and MDA-MB-231 breast cancer cells. PLoS One, 7, e35435.

TANIMOTO, K., HAYASHI, S., TSUCHIYA, E., TOKUCHI, Y., KOBAYASHI, Y., YOSHIGA, K., et al. 2000. Abnormalities of the FHIT gene in human oral carcinogenesis. Br J Cancer, 82, 838-43.

TAYLOR, B. S., BARRETINA, J., SOCCI, N. D., DECAROLIS, P., LADANYI, M., MEYERSON, M., et al. 2008. Functional copy-number alterations in cancer. PLoS One, 3, e3179.

TEAM, R. C. 2015. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.

THIBODEAU, B. J., GEDDES, T. J., FORTIER, L. E., AHMED, S., PRUETZ, B. L., WOBB, J., et al. 2014. Gene Expression Characterization of HPV Positive Head and Neck Cancer to Predict Response to Chemoradiation. Head Neck Pathol.

THOMAS, G. R., NADIMINTI, H. & REGALADO, J. 2005. Molecular predictors of clinical outcome in patients with head and neck squamous cell carcinoma. Int J Exp Pathol, 86, 347-63.

TOKER, C. 1963. Some observations on the deposition of metastatic carcinoma within cervical lymph nodes. Cancer, 16, 364-74.

TOMASETTI, C., MARCHIONNI, L., NOWAK, M. A., PARMIGIANI, G. & VOGELSTEIN, B. 2015. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci U S A, 112, 118-23.

TRASK, D. K., WOLF, G. T., BRADFORD, C. R., FISHER, S. G., DEVANEY, K., JOHNSON, M., et al. 2002. Expression of Bcl-2 family proteins in advanced laryngeal squamous cell carcinoma: correlation with response to chemotherapy and organ preservation. Laryngoscope, 112, 638-44.

TSUKAMOTO, Y., NAKADA, C., NOGUCHI, T., TANIGAWA, M., NGUYEN, L. T., UCHIDA, T., et al. 2010. MicroRNA-375 is downregulated in gastric carcinomas and regulates cell survival by targeting PDK1 and 14-3-3zeta. Cancer Res, 70, 2339-49.

UPILE, N. S., SHAW, R. J., JONES, T. M., GOODYEAR, P., LILOGLOU, T., RISK, J. M., et al. 2014. Squamous cell carcinoma of the head and neck outside the oropharynx is rarely human papillomavirus related. Laryngoscope.

VALASTYAN, S., CHANG, A., BENAICH, N., REINHARDT, F. & WEINBERG, R. A. 2011. Activation of miR-31 function in already-established metastases elicits metastatic regression. Genes Dev, 25, 646-59.

VALINEZHAD ORANG, A., SAFARALIZADEH, R. & KAZEMZADEH-BAVILI, M. 2014. Mechanisms of miRNA-Mediated Gene Regulation from Common

Downregulation to mRNA-Specific Upregulation. Int J Genomics, 2014, 970607.

VAN DEN BREKEL, M. W., CASTELIJNS, J. A., STEL, H. V., GOLDING, R. P., MEYER, C. J. & SNOW, G. B. 1993. Modern imaging techniques and ultrasound-guided aspiration cytology for the assessment of neck node metastases: a prospective comparative study. Eur Arch Otorhinolaryngol, 250, 11-7.

VAN DEN BREKEL, M. W., LODDER, W. L., STEL, H. V., BLOEMENA, E., LEEMANS, C. R. & VAN DER WAAL, I. 2012. Observer variation in the histopathologic assessment of extranodal tumor spread in lymph node metastases in the neck. Head Neck, 34, 840-5.

VASSILAKOPOULOU, M., PSYRRI, A. & ARGIRIS, A. 2015. Targeting angiogenesis in head and neck cancer. Oral Oncol, 51, 409-15.

VEIGAARD, C. & KJELDSEN, E. 2014. Exploring the genome-wide relation between copy number status and microRNA expression. Genomics, 104, 271-278.

VENKATRAMAN, E. S. & OLSHEN, A. B. 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics, 23, 657-63.

VENUTI, A. & PAOLINI, F. 2012. HPV detection methods in head and neck cancer. Head Neck Pathol, 6 Suppl 1, S63-74.

VERMEULEN, L., DE SOUSA, E. M. F., VAN DER HEIJDEN, M., CAMERON, K., DE JONG, J. H., BOROVSKI, T., et al. 2010. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol, 12, 468-76.

VICKERS, M. M., BAR, J., GORN-HONDERMANN, I., YAROM, N., DANESHMAND, M., HANSON, J. E., et al. 2012. Stage-dependent differential expression of microRNAs in colorectal cancer: potential role as markers of metastatic disease. Clin Exp Metastasis, 29, 123-32.

VIDAL, L. & GILLISON, M. L. 2008. Human papillomavirus in HNSCC: recognition of a distinct disease type. Hematol Oncol Clin North Am, 22, 1125-42, vii.

VIGNOT, S., FRAMPTON, G. M., SORIA, J. C., YELENSKY, R., COMMO, F., BRAMBILLA, C., et al. 2013. Next-generation sequencing reveals high concordance of recurrent somatic alterations between primary tumor and metastases from patients with non-small-cell lung cancer. J Clin Oncol, 31, 2167-72.

VIZCAINO, A. P., MORENO, V., BOSCH, F. X., MUNOZ, N., BARROS-DIOS, X. M., BORRAS, J., et al. 2000. International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. Int J Cancer, 86, 429-35.

VLACHOS, I. S., KOSTOULAS, N., VERGOULIS, T., GEORGAKILAS, G., RECZKO, M., MARAGKAKIS, M., et al. 2012. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. Nucleic Acids Res, 40, W498-504.

VLACHOS, I. S., PARASKEVOPOULOU, M. D., KARAGKOUNI, D., GEORGAKILAS, G., VERGOULIS, T., KANELLOS, I., et al. 2015. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids Res, 43, D153-9.

VOGELSTEIN, B. & KINZLER, K. W. 2004. Cancer genes and the pathways they control. Nat Med, 10, 789-99.

VON AHLFEN, S., MISSEL, A., BENDRAT, K. & SCHLUMPBERGER, M. 2007. Determinants of RNA quality from FFPE samples. PLoS One, 2, e1261.

VOORNEVELD, P. W., KODACH, L. L., JACOBS, R. J., LIV, N., ZONNEVYLLE, A. C., HOOGENBOOM, J. P., et al. 2014. Loss of SMAD4 alters BMP signaling to promote colorectal cancer cell metastasis via activation of Rho and ROCK. Gastroenterology, 147, 196-208 e13.

WAGHORNE, C., THOMAS, M., LAGARDE, A., KERBEL, R. S. & BREITMAN, M. L. 1988. Genetic evidence for progressive selection and overgrowth of primary tumors by metastatic cell subpopulations. Cancer Res, 48, 6109-14.

WALTER, V., YIN, X., WILKERSON, M. D., CABANSKI, C. R., ZHAO, N., DU, Y., et al. 2013. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS One, 8, e56823.

WANG, F., ZHENG, Z., GUO, J. & DING, X. 2010. Correlation and quantitation of microRNA aberrant expression in tissues and sera from patients with breast tumor. Gynecol Oncol, 119, 586-93.

WANG, L. G., NI, Y., SU, B. H., MU, X. R., SHEN, H. C. & DU, J. J. 2013. MicroRNA-34b functions as a tumor suppressor and acts as a nodal point in the feedback loop with Met. Int J Oncol, 42, 957-62.

WANG, L., CHANG, L., LI, Z., GAO, Q., CAI, D., TIAN, Y., et al. 2014. miR-99a and -99b inhibit cervical cancer cell proliferation and invasion by targeting mTOR signaling pathway. Med Oncol, 31, 934.

WANG, W., LIM, W. K., LEONG, H. S., CHONG, F. T., LIM, T. K., TAN, D. S., et al. 2015. An eleven gene molecular signature for extra-capsular spread in oral squamous cell carcinoma serves as a prognosticator of outcome in patients without nodal metastases. Oral Oncol, 51, 355-62.

WANG, W., SONGLIN, P., SUN, Y., ZHANG, B. & JINHUI, W. 2012. miR-21 inhibitor sensitizes human OSCC cells to cisplatin. Mol Biol Rep, 39, 5481-5.

WARD, M. J., THIRDBOROUGH, S. M., MELLOWS, T., RILEY, C., HARRIS, S., SUCHAK, K., et al. 2014. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. Br J Cancer, 110, 489-500.

WARNAKULASURIYA, S. 2009. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol, 45, 309-16.

WATANABE, J., NISHIZAKI, R., JOBO, T., KAMATA, Y., HATA, H., NISHIMURA, Y., et al. 2004. Expression of tumor suppressor gene product p14ARF in endometrioid adenocarcinoma of the uterine corpus. International Journal of Gynecological Pathology, 23, 234-240.

WAY, T. D., HUANG, J. T., CHOU, C. H., HUANG, C. H., YANG, M. H. & HO, C. T. 2014. Emodin represses TWIST1-induced epithelial-mesenchymal

transitions in head and neck squamous cell carcinoma cells by inhibiting the beta-catenin and Akt pathways. Eur J Cancer, 50, 366-78.

WEBSTER, R. J., GILES, K. M., PRICE, K. J., ZHANG, P. M., MATTICK, J. S. & LEEDMAN, P. J. 2009. Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7. J Biol Chem, 284, 5731-41.

WEINBERGER, P. M., YU, Z., HAFFTY, B. G., KOWALSKI, D., HARIGOPAL, M., BRANDSMA, J., et al. 2006. Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. J Clin Oncol, 24, 736-47.

WEIR, B. A., WOO, M. S., GETZ, G., PERNER, S., DING, L., BEROUKHIM, R., et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. Nature, 450, 893-8.

WEISS, M. H., HARRISON, L. B. & ISAACS, R. S. 1994. Use of decision analysis in planning a management strategy for the stage N0 neck. Arch Otolaryngol Head Neck Surg, 120, 699-702.

WELKOBORSKY, H. J., BERNAUER, H. S., RIAZIMAND, H. S., JACOB, R., MANN, W. J. & HINNI, M. L. 2000. Patterns of chromosomal aberrations in metastasizing and nonmetastasizing squamous cell carcinomas of the oropharynx and hypopharynx. Ann Otol Rhinol Laryngol, 109, 401-10.

WENG, L., WU, X., GAO, H., MU, B., LI, X., WANG, J. H., et al. 2010. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. J Pathol, 222, 41-51.

WERNER, M., CHOTT, A., FABIANO, A. & BATTIFORA, H. 2000. Effect of formalin tissue fixation and processing on immunohistochemistry. Am J Surg Pathol, 24, 1016-9.

WHITE, R. A., MALKOSKI, S. P. & WANG, X. J. 2010. TGFbeta signaling in head and neck squamous cell carcinoma. Oncogene, 29, 5437-46.

WHITNEY, A. R., DIEHN, M., POPPER, S. J., ALIZADEH, A. A., BOLDRICK, J. C., RELMAN, D. A., et al. 2003. Individuality and variation in

gene expression patterns in human blood. Proc Natl Acad Sci U S A, 100, 1896-901.

WHO. 2012. http://www.globocon.iarc.fr [Online]. [Accessed 08/09/2015.

WIGHTMAN, B., HA, I. & RUVKUN, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell, 75, 855-62.

WILLIS, R. A. 1930. Epidermoid carcinoma of the head and neck with special reference to metastasis. The Journal of Pathology and Bacteriology, 33, 501-26.

WILTING, S. M., DE WILDE, J., MEIJER, C. J., BERKHOF, J., YI, Y., VAN WIERINGEN, W. N., et al. 2008. Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. Genes Chromosomes Cancer, 47, 890-905.

WONG, Q. W., LUNG, R. W., LAW, P. T., LAI, P. B., CHAN, K. Y., TO, K. F., et al. 2008a. MicroRNA-223 is commonly repressed in hepatocellular carcinoma and potentiates expression of Stathmin1. Gastroenterology, 135, 257-69.

WONG, T. S., LIU, X. B., WONG, B. Y., NG, R. W., YUEN, A. P. & WEI, W. I. 2008b. Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue. Clin Cancer Res, 14, 2588-92.

WOOD, H. 2013. Applications of very low-coverage sequencing in cancer genomics: copy number, virus detection and survival. In: WEI WU, H. C. (ed.) Next Generation Sequencing in Cancer Research. Volume 1: Decoding the Cancer Genome. New York: Springer.

WOOD, H. M., BELVEDERE, O., CONWAY, C., DALY, C., CHALKLEY, R., BICKERDIKE, M., et al. 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. Nucleic Acids Res, 38, e151.

WOOD, H. M., BOLT, R. & HUNTER, K. D. 2012. Is next-generation sequencing an important tool in HPV subtype diagnosis? Expert Rev Mol Diagn, 12, 663-5.

WOOD, H. M., CONWAY, C., DALY, C., CHALKLEY, R., BERRI, S., SENGUVEN, B., et al. 2015. The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing. J Pathol.

WOOLGAR, J. A., ROGERS, S. N., LOWE, D., BROWN, J. S. & VAUGHAN, E. D. 2003. Cervical lymph node metastasis in oral cancer: the importance of even microscopic extracapsular spread. Oral Oncol, 39, 130-7.

WU, K. J. & YANG, M. H. 2011. Epithelial-mesenchymal transition and cancer stemness: the Twist1-Bmi1 connection. Biosci Rep, 31, 449-55.

WU, K. J. 2011. Direct activation of Bmi1 by Twist1: implications in cancer stemness, epithelial-mesenchymal transition, and clinical significance. Chang Gung Med J, 34, 229-38.

WU, X., WENG, L., LI, X., GUO, C., PAL, S. K., JIN, J. M., et al. 2012a. Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. PLoS One, 7, e35661.

WU, X., ZHANG, D. & LI, G. 2012b. Insights into the regulation of human CNV-miRNAs from the view of their target genes. BMC Genomics, 13, 707.

XIA, J., CHEN, Q., LI, B. & ZENG, X. 2007. Amplifications of TAOS1 and EMS1 genes in oral carcinogenesis: association with clinicopathological features. Oral Oncol, 43, 508-14.

XIAO, W., BAO, Z. X., ZHANG, C. Y., ZHANG, X. Y., SHI, L. J., ZHOU, Z. T., et al. 2012. Upregulation of miR-31* is negatively associated with recurrent/newly formed oral leukoplakia. PLoS One, 7, e38648.

XIE, T., G, D. A., LAMB, J. R., MARTIN, E., WANG, K., TEJPAR, S., et al. 2012. A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations. PLoS One, 7, e42001.

XU, C., LIU, Y., WANG, P., FAN, W., RUE, T. C., UPTON, M. P., et al. 2010. Integrative analysis of DNA copy number and gene expression in metastatic

oral squamous cell carcinoma identifies genes associated with poor survival. Mol Cancer, 9, 143.

XU, C., WANG, P., LIU, Y., ZHANG, Y., FAN, W., UPTON, M. P., et al. 2013a. Integrative genomics in combination with RNA interference identifies prognostic and functionally relevant gene targets for oral squamous cell carcinoma. PLoS Genet, 9, e1003169.

XU, J., ESCAMILLA, J., MOK, S., DAVID, J., PRICEMAN, S., WEST, B., et al. 2013b. CSF1R signaling blockade stanches tumor-infiltrating myeloid cells and improves the efficacy of radiotherapy in prostate cancer. Cancer Res, 73, 2782-94.

XU, Y. & FISHER, G. J. 2013. Role of met axis in head and neck cancer. Cancers (Basel), 5, 1601-18.

YAMASHITA, S. I., MASUDA, Y., YOSHIDA, N., MATSUZAKI, H., KURIZAKI, T., HAGA, Y., et al. 2008. p53AIP1 expression can be a prognostic marker in non-small cell lung cancer. Clin Oncol (R Coll Radiol), 20, 148-51.

YAN, B., FU, Q., LAI, L., TAO, X., FEI, Y., SHEN, J., et al. 2012. Downregulation of microRNA 99a in oral squamous cell carcinomas contributes to the growth and survival of oral cancer cells. Mol Med Rep, 6, 675-81.

YAN, D., DONG XDA, E., CHEN, X., WANG, L., LU, C., WANG, J., et al. 2009. MicroRNA-1/206 targets c-Met and inhibits rhabdomyosarcoma development. J Biol Chem, 284, 29596-604.

YAN, L. X., HUANG, X. F., SHAO, Q., HUANG, M. Y., DENG, L., WU, Q. L., et al. 2008. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA, 14, 2348-60.

YANG, S., JEUNG, H. C., JEONG, H. J., CHOI, Y. H., KIM, J. E., JUNG, J. J., et al. 2007. Identification of genes with correlated patterns of variations in DNA copy number and gene expression level in gastric cancer. Genomics, 89, 451-9.

YANG, W. H., LAN, H. Y., TAI, S. K. & YANG, M. H. 2013. Repression of bone morphogenetic protein 4 by let-7i attenuates mesenchymal migration of head and neck cancer cells. Biochem Biophys Res Commun, 433, 24-30.

YANG, Z., HAN, Y., CHENG, K., ZHANG, G. & WANG, X. 2014. miR-99a directly targets the mTOR signalling pathway in breast cancer side population cells. Cell Prolif, 47, 587-95.

YI, R., PASOLLI, H. A., LANDTHALER, M., HAFNER, M., OJO, T., SHERIDAN, R., et al. 2009. DGCR8-dependent microRNA biogenesis is essential for skin development. Proc Natl Acad Sci U S A, 106, 498-502.

YI, R., POY, M. N., STOFFEL, M. & FUCHS, E. 2008. A skin microRNA promotes differentiation by repressing 'stemness'. Nature, 452, 225-9.

YILMAZ, T., HOSAL, A. S., GEDIKOGLU, G., ONERCI, M. & GURSEL, B. 1998. Prognostic significance of vascular and perineural invasion in cancer of the larynx. Am J Otolaryngol, 19, 83-8.

YONESAKA, K., ZEJNULLAHU, K., OKAMOTO, I., SATOH, T., CAPPUZZO, F., SOUGLAKOS, J., et al. 2011. Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. Sci Transl Med, 3, 99ra86.

YONG, Z. W., ZAINI, Z. M., KALLARAKKAL, T. G., KAREN-NG, L. P., RAHMAN, Z. A., ISMAIL, S. M., et al. 2014. Genetic alterations of chromosome 8 genes in oral cancer. Sci Rep, 4, 6073.

YOSHIOKA, S., TSUKAMOTO, Y., HIJIYA, N., NAKADA, C., UCHIDA, T., MATSUURA, K., et al. 2013. Genomic profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization. PLoS One, 8, e56165.

YOUSEM, D. M., SOM, P. M., HACKNEY, D. B., SCHWAIBOLD, F. & HENDRIX, R. A. 1992. Central nodal necrosis and extracapsular neoplastic spread in cervical lymph nodes: MR imaging versus CT. Radiology, 182, 753-9.

YOUSIF, N. G., SADIQ, A. M., YOUSIF, M. G., AL-MUDHAFAR, R. H., AL-BAGHDADI, J. J. & HADI, N. 2015. Notch1 ligand signaling pathway

activated in cervical cancer: poor prognosis with high-level JAG1/Notch1. Arch Gynecol Obstet.

YU, C. C., CHEN, Y. W., CHIOU, G. Y., TSAI, L. L., HUANG, P. I., CHANG, C. Y., et al. 2011. MicroRNA let-7a represses chemoresistance and tumourigenicity in head and neck cancer via stem-like properties ablation. Oral Oncol, 47, 202-10.

YU, Z. W., ZHONG, L. P., JI, T., ZHANG, P., CHEN, W. T. & ZHANG, C. P. 2010. MicroRNAs contribute to the chemoresistance of cisplatin in tongue squamous cell carcinoma lines. Oral Oncol, 46, 317-22.

YUSUF, H. & YONG, S. L. 2002. Oral submucous fibrosis in a 12-year-old Bangladeshi boy: a case report and review of literature. Int J Paediatr Dent, 12, 271-6.

ZACK, T. I., SCHUMACHER, S. E., CARTER, S. L., CHERNIACK, A. D., SAKSENA, G., TABAK, B., et al. 2013. Pan-cancer patterns of somatic copy number alteration. Nat Genet, 45, 1134-1140.

ZENDER, L., SPECTOR, M. S., XUE, W., FLEMMING, P., CORDON-CARDO, C., SILKE, J., et al. 2006. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell, 125, 1253-67.

ZENG, W. & MORTAZAVI, A. 2012. Technical considerations for functional sequencing assays. Nat Immunol, 13, 802-7.

ZHANG, J. G., WANG, J. J., ZHAO, F., LIU, Q., JIANG, K. & YANG, G. H. 2010. MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). Clin Chim Acta, 411, 846-52.

ZHANG, L., HUANG, J., YANG, N., GRESHOCK, J., MEGRAW, M. S., GIANNAKAKIS, A., et al. 2006. microRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci U S A, 103, 9136-41.

ZHANG, L., XIA, L., ZHAO, L., CHEN, Z., SHANG, X., XIN, J., et al. 2015. Activation of PAX3-MET pathways due to miR-206 loss promotes gastric cancer metastasis. Carcinogenesis, 36, 390-9.

ZHANG, T., HAN, G., WANG, Y., CHEN, K. & SUN, Y. 2014. MicroRNA expression profiles in supraglottic carcinoma. Oncol Rep, 31, 2029-34.

ZHANG, X. C., XU, C., MITCHELL, R. M., ZHANG, B., ZHAO, D., LI, Y., et al. 2013. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. Neoplasia, 15, 1371-8.

ZHANG, Z. D. & GERSTEIN, M. B. 2010. Detection of copy number variation from array intensity and sequencing read depth using a stepwise Bayesian model. BMC Bioinformatics, 11, 539.

ZHAO, X. D., ZHANG, W., LIANG, H. J. & JI, W. Y. 2013. Overexpression of miR -155 promotes proliferation and invasion of human laryngeal squamous cell carcinoma via targeting SOCS1 and STAT3. PLoS One, 8, e56395.

ZHOU, T., CHOU, J. W., SIMPSON, D. A., ZHOU, Y., MULLEN, T. E., MEDEIROS, M., et al. 2006. Profiles of global gene expression in ionizing-radiation-damaged human diploid fibroblasts reveal synchronization behind the G1 checkpoint in a G0-like state of quiescence. Environ Health Perspect, 114, 553-9.

ZHOU, X., ZHANG, Z., YANG, X., CHEN, W. & ZHANG, P. 2009. Inhibition of cyclin D1 expression by cyclin D1 shRNAs in human oral squamous cell carcinoma cells is associated with increased cisplatin chemosensitivity. Int J Cancer, 124, 483-9.

ZHU, J., SANBORN, J. Z., BENZ, S., SZETO, C., HSU, F., KUHN, R. M., et al. 2009. The UCSC Cancer Genomics Browser. Nat Methods, 6, 239-40.

ZIMMERMANN, M., ZOUHAIR, A., AZRIA, D. & OZSAHIN, M. 2006. The epidermal growth factor receptor (EGFR) in head and neck cancer: its role and treatment implications. Radiat Oncol, 1, 11.

ZOUMALAN, R. A., KLEINBERGER, A. J., MORRIS, L. G., RANADE, A., YEE, H., DELACURE, M. D., et al. 2010. Lymph node central necrosis on computed tomography as predictor of extracapsular spread in metastatic head and neck squamous cell carcinoma: pilot study. J Laryngol Otol, 124, 1284-8.