The genomic landscape of oral verrucous carcinoma

Manar MohammedAbubakr I Samman

BSc, MSc

Submitted in accordance with the requirements for the Degree of Doctor of Philosophy

University of Leeds

School of Medicine

November 2014

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.

Chapter five is based on work from jointly authored publication (Next-Generation Sequencing Analysis for Detecting Human Papillomavirus in Oral Verrucous Carcinoma. Manar Samman; Henry M Wood; Caroline A Conway; Stefano Berri; Monica Pentenero; Sergio Gandolfo; Adele Cassenti; Paola Cassoni; Abdulaziz Al Ajlan; William A Barrett; Preetha Chengot; Pamela Rabbitts; Alec S High; Kenneth MacLennan. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2014 Jul;118 (1):117-125.

All the work in the above paper is part of this PhD project. I did: samples collection, all the required lab work, data analysis and paper writing. Henry M Wood and Stefano Berri did the bioinformatics. Caroline Conway provided technical training. Monica Pentenero, Sergio Gandolfo, Adele Cassenti, Paola Cassoni, Abdulaziz Al Ajlan and William A Barrett provided patients samples. Preetha Chengot, Alec S High and Kenneth MacLennan provided pathology diagnosis. Pamela Rabbitts supervised the project.

Acknowledgments

First and foremost, I am most grateful to almighty Allah (God) for providing me with blessings and giving me the strength to accomplish this work “Praise be to Allah under all conditions. Praise be to Allah through whose mercy (and favours) all good things are accomplished”. I would also like to ask Him that this project will prove to be another helpful tool in the treatment of oral cancer patients.

I would like to express my heartfelt gratitude to Professor Pamela Rabbitts and Dr Alec High, both of whom are not only advisors and supervisors but have also proven themselves to be my dear family in the UK. I could not have asked for better role models as scientists, teachers and mentors, each supportive, patient and inspirational. I could not be prouder of my scientific research roots, and it is my hope that I can, in turn, pass on the values and knowledge they have given to me. Thank you for helping to guide and shape the direction of this project with your instructive and careful comments, from the beginning until the end.

I would also like to thank Dr. Henry Wood (Senior Research Fellow), who provided the bioinformatics data analysis. It was not an easy task, and I am grateful for his insightful discussions and suggestions about this research, for his time and endless help and support.

It is a pleasure to thank those who made this project possible, such as Dr. Caroline Conway (Postdoctoral Scientist) for being a supportive mentor through all the stages of my PhD project, and also Dr. Lucy Stead (Postdoctoral Scientist), who gave me her valuable time to analyse my RNAseq data.

There is no way to express how much it has meant to me to be a member of the Pre-Cancer genomics group. I have been surrounded by brilliant friends and wonderful colleagues, all of whom have provided a rich and productive environment to study and investigate new ideas. I would like to thank the lab technologists—Philip Egan, Catherine Daly and Rebecca Chalky—for being extremely supportive and informative in teaching me the various lab techniques whilst pursuing my PhD studies. In their various ways, they have helped to make my time in the field both efficient and successful.

Undertaking this PhD project has been an amazing journey, which has taken me across three countries for the collection of my samples. Most of the results described in this thesis would not have been obtained without close collaboration with a number of laboratories. I am most grateful to the collaborators for lending me their patients’ samples—namely Dr Monica Pentenero (Department of Oncology, Oral Medicine and Oral Oncology Unit,

University of Torino, Turin, Italy), Dr William Barratt (Department of Histopathology, Queen Victoria Hospital, East Grinstead, UK), Dr Abdulaziz Al Ajlan and Dr Alaa Samkari (National Guard Health Affairs, Saudi Arabia). I am sincerely thankful for their input into this project.

I take this opportunity to sincerely acknowledge King Fahad Medical City for awarding me the PhD degree scholarship. My research would not have been possible without their financial support and help. Separately, I must thank Dr Mahmoud Al Yamany, Dr Saleh Altamimi, Dr Ameen Softah and Dr Abdulwahed Al-Dehaimi (King Fahad Medical City, Saudi Arabia) for their outstanding support throughout this scholarship. I am very grateful to them for all of their help.

I would also like to thank two other people who have had an important academic influence on me: Prof Mohammed Hannan, who always believed in my abilities and whose advice has been enormously helpful. Mr Khelad Al-Saidi, who introduced me into the cytogenetics field in 2005, has been a role model for me since I first started to work at King Fahad medical city.

I would not have contemplated this road and made it this far if not for my parents, my father Mohammed Samman and my mother Mona Al-Zankawi, who inspired me with faith, and a love for the creative pursuits and science, all of which finds a place in this PhD thesis. They sacrificed their lives for me and provided absolute love and generous care. I still remember when I was 11 years old, the first chromosomes picture my father gave to me and which I first played with, and tried to pair, when he was doing his own PhD research on oral cancer. That was the time when I knew that I wanted to grow up to be a scientist like him. My mother always stressed the importance of education, and I know that this respect for education has, in some insentient way, shaped my values and made me the person I am today. Despite the long distance between us, she has been always a source of constant and unconditional love. Words are short to express when striving to communicate my deep sense of gratitude towards my parents.

I have been blessed with a very loving and supportive family, comprising my two brothers, Mouath and Majed, and my sister Maram, all of whom have always encouraged me with their best prayers and wishes.

Being a mother, without any doubts, has been my greatest achievement and source of inspiration. Motherhood has taught me how to be a better person and how to provide unconditional love. To my children, my greatest blessings, King Faisal, Prince Talal, Mister Faris and Princess Rama, your cheerful little smiles always brightened my days. You all have demonstrated amazing patience throughout my long working hours over the last four years. You have been told, “Not right now, Mum is working, I’m busy” often—very often, in fact. I am

honoured to have you in my life and to raise you. Thank you for inspiring me to keep looking for excellence: you helped me to get through this tough period in the most positive way. You are just about the best children a mother could ever have and hope for: loving, happy, and fun to be with. Thank you for everything that you are, and everything you will become. There are no words to convey how much I love you all.

I also owe a sincere thank you, with love, to my friends near and far (too many to list here but you know who you are!) for encouraging me to pursue this project in the first place and for their constant friendship. I cannot and will not forget friends who have been through good and hard times together, with me, always cheering me on, and supporting and celebrating my accomplishments.

I am also grateful to Ascia Saikon for taking care of my children whenever I couldn't be there for them. Thank you, Ascia, for your patience. I would like to offer my blessings and regards also to those who stood by me in any respect during the completion of this project.

Last, but never least, I must thank my astonishingly supportive husband, Omar Kurdi. I find it difficult to express my deep appreciation as it is so boundless. You are my most enthusiastic supporter, my soulmate, my best friend, and the amazing husband and father I always dreamt of. This thesis would have taken even longer to complete without your bright optimism, and I would be lost without your love and full support. I am indebted to you not only because you have given up a lot to make my future and career a priority in our lives, but also because you have understood the ups and downs involved in the completion of an entire PhD process. You shared this miraculous journey with me; therefore, it seems only right that I dedicate this thesis to you. I will be forever thankful to you.

Abstract

Oral verrucous carcinoma (OVC) is categorised as a low-grade variant of oral squamous cell carcinoma (OSCC). The aetiology of OVC is unknown, and the suggested role of human papillomavirus (HPV) as a causative factor remains contentious. Distinguishing OVC from OSCC is problematic. The rarity of these lesions also makes them difficult to investigate, so most previous studies have been made on small numbers of cases.

The aim of this study is to use next generation (NG) copy number (CN) sequencing to identify OVC and oral verrucous hyperplasia (OVH) genomic features and determine if CN analysis could distinguish between the genomic damage pattern in OVH, OVC, and OSCC lesions. Additionally, this project aims to investigate the transcriptional and genomic changes that occur in OVC and compare them with the alterations that occur in OSCC using NG RNA-seq and whole-exome sequencing. Also, and since a verrucous appearance is suggestive of viral aetiology, this study aims to analyse OVC and OVH lesions for the presence of HPV.

A total of 57 OVC and 16 OVH FFPE cases were identified for CN analysis and were analysed for the presence of HPV subtypes and for all known human viruses. The CN karyograms of those cases were compared with 45 OSCC karyograms. Transcriptome and exome sequencing were performed on a subset of OVC cases and the results were compared with OSCC sequencing data (all OSCC data belongs to Pre-cancer Genomic Group).

CN results showed that the OVC lacked any of the classical OSCC genomic abnormalities such as gain of 3q and loss of 3p and demonstrated considerably less genomic instability than the OSCC cohort. OVC and OSCC profiles could be clearly distinguished. An HPV-16 sequence was identified in one OVC and one OVH, and an HPV-2 sequence was identified in one OVC out of the 73 cases but with low viral loads. Transcriptome sequencing also identified genes that are differentially expressed between the groups. Exome sequencing showed that OVC patients lacked mutations in any of the genes commonly associated with OSCC (*TP53*, *CDKN2A*, *NOTCH2* etc.).

Taken together, these results lead to the conclusion that no association between HPV infections and oral verrucous lesions. OVC is not a subtype of OSCC, but should be classified as distinct entity. The distinguishing features presented in this project should be of value in diagnosis.

Publications

Publications from the thesis

1. P066: No association between human papillomavirus infection and oral verrucous lesions. **Manar Samman**, Henry Wood, Stefano Berri, Monica Pentenero, Alec High, Pamela Rabbitts. Oral Oncology **49**. May 2013. PS116.

2. Next-Generation Sequencing Analysis for Detecting Human Papillomavirus in Oral Verrucous Carcinoma. **Manar Samman**; Henry M Wood; Caroline A Conway; Stefano Berri; Monica Pentenero; Sergio Gandolfo; Adele Cassenti; Paola Cassoni; Abdulaziz Al Ajlan; William A Barrett; Preetha Chengot; Pamela Rabbitts; Alec S High; Kenneth MacLennan. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2014 Jul;118 (1):117-125.

Submitted publications

1. **Manar Samman** and Neeraj Sethi."Using next generation sequencing to reveal patterns of chromosomal alterations in oral verrucous lesions." *Next Generation Sequencing in Cancer Research*. Volume 2. Springer New York, 2014. Submitted.

2. **Manar Samman** and Neeraj Sethi."Oral verrucous premalignant lesions and HPV." Clinical Otolaryngology, 2015. Submitted.

3. Stratifying tumour subtypes based on copy number alteration profiles using next-generation sequence data. Arief Gusnanto, Peter Tcherveniakov, Farag Shuweihdi, **Manar Samman**, Pamela Rabbitts, and Henry M. Wood. Submitted to Bioinformatics.

4. Multiple genomic and transcriptomic analyses redefine an oral cancer subtype. **Manar Samman**, Henry M Wood, Caroline Conway, Lucy Stead, Catherine Daly, Rebecca Chalkley, Burcu Senguven, Lisa Ross, Philip Egan, Preetha Chengot, Thian K Ong, Monica Pentenero, Sergio Gandolfo, Adele Cassenti, Paola Cassoni, Abdulaziz Al Ajlan, Alaa Samkari, William Barrett, Kenneth MacLennan, Alec High, Pamela Rabbitts. Submitted to International Journal of Cancer.

Poster presentations

1. Leeds dental institute research day, University of Leeds, 11th of July 2012: NGS analysis of DNA from tumour biopsies to reveal differences between genomes of pre-cancerous lesions and malignant tumours. **Manar M Samman**, Henry M Wood, Stefano Berri, Catherine L Daly, Alec S High , Pamela H Rabbitts.

2. 6th Saudi scientific international conference, London, UK. 11-14 October 2012: NGS analysis of DNA from tumour biopsies to reveal differences between genomes of pre-cancerous lesions and malignant tumours. **Manar M Samman**, Henry M Wood, Stefano Berri, Catherine L Daly, Alec S High , Pamela H Rabbitts.

3. Leeds Institutes of Molecular Medicine (LIMM) Postgraduate Research Symposium, university of Leeds, 15th April 2013: Next generation sequencing: a powerful tool and a new insight toward understanding the genomic landscape within FFPE oral verrucous carcinomas. **Manar Samman**, Henry Wood, Alec High, Pamela Rabbitts.

4. 4th world oral oncology congress, Greece, 15-18 May 2013: No association between human papillomavirus infection and oral verrucous lesions. **Manar Samman**, Henry Wood, Stefano Berri, Monica Pentenero, Alec High, Pamela Rabbitts.

5. The European Cancer Congress 2013, Amsterdam, 27 September-1 October 2013: Next Generation Sequencing: a powerful tool providing new insights into the genomic landscape of Oral Verrucous Carcinomas**. Manar Samman**, Henry Wood, Alec High, Pamela Rabbitts.

6. NCRI cancer conference 2013, Liverpool, 3-6 November 2013: Next-Generation Sequencing Analysis for Detecting Human Papillomavirus in oral verrucous lesions. **Manar Samman**, Henry M Wood, Monica Pentenero, William A Barrett, Alec S High, Pamela Rabbitts.

7. 7th Saudi scientific international conference, Edinburgh, UK, 1-2 February, 2014: Next Generation Sequencing (NGS): a powerful tool providing new insights into the genomic landscape of Oral Verrucous Carcinomas. **Manar Samman**, Henry Wood, Lucy Stead, Alec High, Pamela Rabbitts.

8. The Biology of Genomes 2014 Cold Spring Harbor Laboratory, USA, 6th to 10th of May 2014: Next Generation Sequencing (NGS) Copy Number analysis re-defines the classification of Oral Verrucous Carcinoma. **Manar Samman**, Henry Wood, Alec High, Pamela Rabbitts.

Oral presentations

1. Leeds dental institute research day, University of Leeds, 10th of July 2013: Next generation sequencing: a powerful tool providing new insights into the genomic landscape of oral verrucous carcinomas.

2. Leeds Institutes of Molecular Medicine (LIMM) Postgraduate Research Symposium, university of Leeds, 29th April, 2014: Next Generation Sequencing Copy Number analysis re-defines the classification of Oral Verrucous Carcinoma

Awards

1. 1st Prize – Poster Presentation: Next Generation Sequencing (NGS): a powerful tool providing new insights into the genomic landscape of Oral Verrucous Carcinomas. 7th Saudi scientific international conference, Edinburgh, UK. February 2014.

2. The Royal Embassy of Saudi Arabia in London held a ceremony to honour 100 Saudi students for their distinct discoveries and successes in British universities. During the tribute, I was identified by Prince Mohammed bin Nawaf, as one of the top ten students who had excellent scientific achievements and inventions. 23rd of September 2014.

3. King Fahad Medical City outstanding achievements award at KFMC 10th year anniversary ceremony, Riyadh, Saudi Arabia, 2015.

4. Best paper award at the 8th Saudi scientific international conference, London, UK. 31st January – 1st February 2015: Multiple genomic and transcriptomic analyses reclassify an oral cancer subtype.

5. 1st Prize – Oral Presentation: Multiple genomic and transcriptomic analyses reclassify an oral cancer subtype. 8th Saudi scientific international conference, London, UK. 31st January – 1st February 2015.

Table of contents

List of Figures

Chapter one:

Chapter two:

Chapter three:

Chapter four:

Chapter six:

Chapter seven:

List of Tables

Chapter one:

Chapter two:

Chapter three:

Chapter four:

Chapter five:

Chapter six:

Chapter seven:

List of abbreviations

°C (suffix)	degrees Celsius
aCGH	Array Comparative Genomic Hybridisation
cDNA	copy DNA
CGH	Comparative Genomic Hybridisation
Chr	Chromosome
CN	Copy number
CNA	Copy number alterations
CNV	Copy number variations
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DEGs	Differentially Expressed Genes
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
FFPE	Formalin-fixed paraffin-embedded
FPKM	Fragments per Kilobase per million Mapped
GISTIC	Genomic Identification of Significant Targets in Cancer
GO	Gene ontology
H&E	Hematoxylin and Eosin
H&N	Head and neck
HNC	Head and neck cancer
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
IHC	Immunohistochemistry
ISH	*In-situ* hybridization
KEGG	Kyoto Encyclopedia of Genes and Genomes
KSHV	Kaposi's sarcoma-associated herpesvirus
lincRNA	Large intergenic non-coding RNA
LOH	Loss of heterozygosity
LR	Logistic regression

ml	milliliter
mRNA	messenger RNA
ng	nanogram
NGS	Next Generation Sequencing
OLP	Oral lichen planus
OP	oropharyngeal
OPL	Oral premalignant lesion
OSCC	Oral squamous cell carcinoma
OVC	Oral verrucous carcinoma
OVH	Oral verrucous hyperplasia
PCR	Polymerase chain reaction
PI3K	phosphoinositide 3-kinase
PVL	Proliferative verrucous leukoplakia
RNA	ribonucleic acid
RNA Seq	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
WES	Whole exome sequencing
WHO	World Health Organization

Chapter 1 Introduction

1.1 Oral cancer

1.1.1 Incidence

Oral cancer can be defined as a type of head and neck malignancy that includes all neoplasms in any part of the mouth, and involves uncontrollable growth of cells that attack and harm the surrounding tissue (Silverman, 1999). Oral cancer is a growing serious problem in many parts of the world. Pharyngeal and oral cancers are the sixth most common type of cancer worldwide (Warnakulasuriya, 2009). The estimated annual incidence of oral cancer is about 275,000, with two-thirds of these cases occurring in developing countries (Figure 1.1) (Warnakulasuriya, 2009). In addition, oral cancer is more common in men than in women in most countries, and the recorded sex variations are attributable to more indulgence in risk behaviors by men (Warnakulasuriya, 2009). Oral cancer is still considered a fatal disease for more than 50% of annually diagnosed cases (Warnakulasuriya, 2009). This is basically due to the fact that most diagnosed cases are in advanced stages at presentation (Warnakulasuriya, 2009). Several studies have revealed a lack of awareness regarding oral cancer causes and symptoms, which underscores the urgent need for public education that specially targets high-risk groups (Warnakulasuriya *et al*., 1999, (Swerdlow *et al.*, 1995). Eighty-five percent of oral cancer arises in the lips, tongue, and floor of the mouth (Silverman, 1999), however, the tongue is the most common site, while oral squamous cell carcinoma (OSCC) forms the most common histological subtype, and involves about 9 out of 10 cases. Besides being the most common oral cancer type, squamous cell carcinoma (SCC) tends to spread quickly (Silverman, 1999). In some geographic areas, the lip is involved more frequently (Haya-Fernandez *et al.*, 2004).

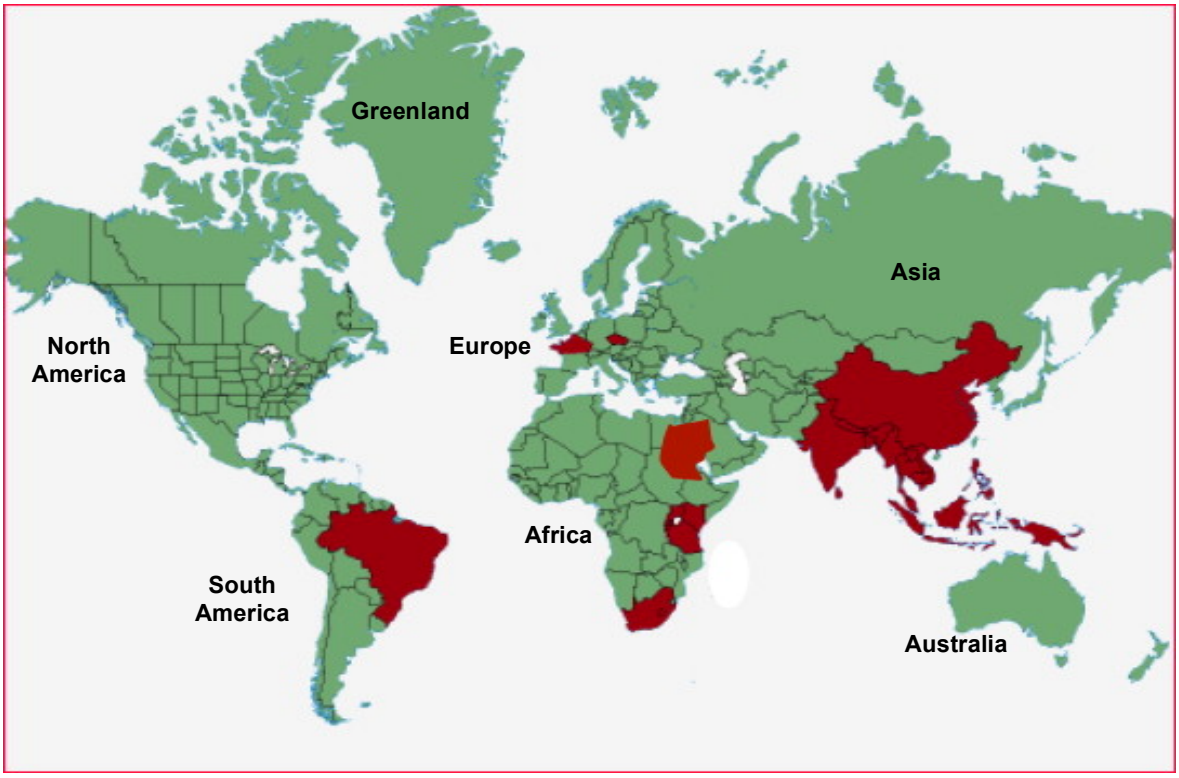


Figure 1.1 Countries with high oral cancer incidence (in red colour).

Figure adapted from (Warnakulasuriya, 2009).

Depending upon various factors, the survival rates for oral cancer patients vary from one case to another. These factors include: the lesion stage (i.e. lesion size, distant metastasis, lymph node involvement, and local extension); the primary tumour site (generally, posterior tumour site leads to negative prognosis); initial treatment adequacy; patient's ability to cope, and the histological differentiation of the tumour (Silverman, 1999). Patient's general health and life style (particularly alcohol and tobacco use) are other main secondary factors affecting their survival (Silverman, 1999). Any delay in the diagnosis allows deep tumours to penetrate into the local structures that can extend to reach the neck regional lymph nodes causing high mortality. Regular oral examinations play an essential role in detecting and controlling oral tumour (Coleman, 2002). Routine oral assessments can identify any mucosal changes that could be premalignant or malignant alterations, and as a result, speeding up the diagnosis and early treatment (Daniel *et al.*, 2004).

1.1.2 Development of oral cancer

Oral carcinogenesis is believed to be a complex process resulting from an accumulation of different cellular changes that are induced by some carcinogens. Alterations in genes regulating DNA synthesis and repair, cell cycle progression, and cell division are fundamental to this process (Thomson *et al.*, 2002), in addition to the mutational alteration of proto-oncogenes. The consequences of these genetic and molecular alterations are changes in the epithelial tissue phenotype, which could be histologically recognized as epithelial dysplasia, and eventually representing cell proliferation and differentiation dysregulation (Kushner *et al.*, 1997), (Gonzalez-Moles *et al.*, 2000). It is believed that cells with high proliferative activity are most likely to be associated with premalignant changes during carcinogenesis, and that increasing cell proliferation deregulation plays an important role in tumorogenesis (Jordan *et al.*, 1998). However, precise, predictive assessments of individual oral cancers and precancerous lesions and their clinical behavior and progression still remain undefinable in medical practice (Thomson *et al.*, 2002). Some oral squamous cell carcinomas arise in obviously normal mucosa, but others are preceded by clinically apparent premalignant lesions, mainly leukoplakia (white patch), erythroplakia (red patch), or speckled leukoplakia (white and red patches) (Cabay *et al.*, 2007).

1.1.3 Oral cancer arising from mutations in keratinocytes

Oral keratinocytes are assumed to be the cell of origin of OSCC (Scully and Bagan, 2009a). As for any cancer, OSCC is often caused by spontaneous DNA mutation but increased by prolonged exposure to any type of mutagen: microbial, physical, or chemical. Mutations in the DNA can promote development of a normal keratinocyte into a potentially malignant or pre-malignant keratinocyte that is described by the capability to proliferate in a less-controlled manner than normal (Scully and Bagan, 2009a). The cells become autonomous and a fully malignant cancer results by the invasion through the epithelial basement membrane and, eventually, metastasises to lymph nodes, bone and other sites (Scully and Bagan, 2009a).

1.1.4 Genetic changes in OSCC

A cancer arises in a multi-step way, with the involvement of a diversity of genetic changes that occur at each level, such as tumour suppressor gene inactivation (genes that prevent uncontrolled cellular proliferation), and oncogene activation (genes encoding proteins that can prompt cancer) (Hanahan and Weinberg, 2011). Such genetic changes are what push normal cells toward developing the main cancer hallmarks such as growth signal autonomy, inducing angiogenesis, uncontrolled and continuous proliferation, avoiding apoptosis, insensitivity to anti-growth signals, and metastatic and invasion capabilities (Hanahan and Weinberg, 2011). Passengers' genetic change contributes towards obtaining one or more of these cancer-associated phenotypes (Haddad and Shin, 2008). Wide-ranging research studies have been carried out or are still ongoing to discover and locate all the contributory genetic changes and their indirect or direct roles in head and neck cancer (HNC) pathogenesis, in order to develop effective cancer therapies, or at least to identify more reliable biomarkers (Haddad and Shin, 2008) that would help in early diagnosis and/or in monitoring therapeutic response.

Carcinogen exposure is known to induce a number of genetic defects identifiable in the aero-digestive tract epithelium, which in turn causes the epithelium to be exposed to high risk concerning the occurrence of premalignant lesions, which can arise at various phases of carcinogenesis (Thomas *et al.*, 2003). The issue of 'field cancerisation'—which is recognised as being a characteristic of cancer affecting the head and neck (H&N)—was first presented in the early 1950s (Slaughter *et al.*, 1953), the introduction of which centred on the hypothesis that long-term exposure to carcinogens can result in the independent progression of epithelial cells at a number of different locations in the adjacent mucosa. The large-scale accumulation of such genomic alterations throughout this developmental process is recognised as arising across a large population of cells—a heterogeneous 'field of genetically altered cells'—which could potentially induce a visible precursor lesion (da Silva *et al.*, 2011). Importantly, this idea seeks to describe the overall high frequency of local recurrences, in addition to the emergence of second primary tumours amongst OSCC patients (da Silva *et al.*, 2011). Additionally, conducting a

number of researches through the use of genetic markers has validated this theory. At the present time, data supports the view that wide fields of cells, comprising cancer-related genetic alterations, surround an estimated 30% of all oropharyngeal and oral cancer cases, which suggests a clonal association with the invasive carcinoma (Tabor *et al.*, 2001). These fields of cells regularly maintain their position upon the removal of tumours, thus, causing the occurrence of secondary tumours, which are clinically assigned as local recurrences and second primary tumours, although this ultimately depends on the time and distance related to the index tumour (Bedi *et al.*, 1996), (Califano *et al.*, 1996), (Califano *et al.*, 2000), (Tabor *et al.*, 2002), (Tabor *et al.*, 2004), (Perez-Ordonez *et al.*, 2006).

Metastasis is a complicated process that involves the progression of tumour cells through a number of different phases. It is controlled by consecutive changes in the expression of particular genes, or gene structure alterations and changes in encoded products (da Silva *et al.*, 2011). Cell disassociation is the preliminary stage within OSCC primary tumour that commonly induces metastasis in the regional (cervical) lymph nodes (Amaral *et al.*, 2004), (Vartanian *et al.*, 2004), (Kowalski *et al.*, 2005). Determining the biological parameters linked with regional metastasis could help to deliver additional information on tumours' metastatic behaviours, and could even prove valuable when making decisions concerning the clinical treatment of the neck (Takes *et al.*, 2008). When taking into account the overall intricacy apparent in the metastatic process, high-throughput approaches are acknowledged as important tools for predicting oral cancer's regional metastasis (Roepman *et al.*, 2005). Identifying variations in gene expression could even aid in identifying the critical genes involved in the metastasis process. In such a way, the gene signatures linked with nodal metastasis can then be established (Roepman *et al.*, 2005), (Takes *et al.*, 2008).

1.1.4.1 Oncogenes and tumour suppressor genes in OSCC

The abnormal activation of *ErbB* family members is a process that has been widely recognised as playing a role in a number of human cancers (Casalini *et al.*, 2004), such as those affecting the H&N (O-charoenrat *et al.*, 2002), which

has been associated with the increased development and metastasis of tumours (Casalini *et al.*, 2004), (Wei *et al.*, 2008), (Silva *et al.*, 2010). Throughout the progression of oral carcinogenesis, growth signalling is known to be deregulated through increased growth factor receptors levels, as well as their ligands, which are known to induce autocrine stimulation (Wei *et al.*, 2008), (Marcu and Yeoh, 2009). *EGFR* (Epidermal Growth Factor Receptor) is an oncogene in H&N tumours that has received much attention and analysis. Owing to its position as an oncogene, it needs to be activated, either through amplification or otherwise via mutation. In the case of the latter, these are not common, whereas amplification, on the other hand, is recognised in as many as 30% of all cases (Sheu *et al.*, 2009). In OSCC, the pathways most widely activated are mitogen-activated protein kinase, PI3K/AKT/mTOR and WNT (Nelson and Nusse, 2004), (Barker, 2008) (Agarwal *et al.*, 2010), (Courtney *et al.*, 2010), (Wong *et al.*, 2010). Importantly, in the case of *MYC* and *RAS* gene families, overexpression plays a fundamental role in the development of HNSCC tumours, and is known to be linked with poor prognoses (Bhattacharya *et al.*, 2009). In a similar vein, the large-scale expression of cyclins—particularly cyclin D1 (*CCND1*)—is a common (36–66%) characteristic in pre-malignant lesions and OSCC (Miyamoto *et al.*, 2003), (Myo *et al.*, 2005). It is important to highlight that *CCND1* is a key regulator, which has the propensity to induce the G1–S transition in the cell cycle regulated by *CDKs* (cyclin-dependent kinase) (Tsantoulis *et al.*, 2007). The amplification of the *CCND1* gene could potentially be responsible for a poorer prognosis, as well as higher risk of occult cervical lymph nodes metastasis in the case of H&N tumours (Miyamoto *et al.*, 2003), (Myo *et al.*, 2005).

Another point to highlight is the fact that tumour suppressor genes are able to stop cells from gathering malignant features, and commonly act in the regulation of discrete checkpoints throughout cell cycles progression, monitoring mitosis and DNA replication (Tsantoulis *et al.*, 2007). Tumour suppressor genes being inactivated may arise through genetic or epigenetic mechanisms. This selection of gene inactivation route throughout tumourigenesis has not yet come to be understood, although, it was suggested that the chemical carcinogens in tobacco smoke could act as a contributory

factors playing a role in *TP53* genetic mutations (Tsantoulis *et al.*, 2007). *TP53* tumour suppressor signalling pathway inactivation is usually identifiable in the cases of human cancers, such as that of OSCC (Hrstka *et al.*, 2009). Furthermore, abnormal *p53* protein activity could also be induced as a result of *TP53* sequence mutations, causing shortened sequence or inactive mutant proteins, or otherwise through the abnormal production of other proteins known to regulate the activity of *p53* (such as viral proteins or *MDM2* gene amplification) (da Silva *et al.*, 2011). Studies conducted recently proposed that inherited genetic polymorphisms in the *p53* pathway have an impact on tumour development and response to therapy (Hrstka *et al.*, 2009). Comparably, $p16^{INK4A}$ protein expression—encoded by the *CDKN2A* suppressor gene—is either low or completely negative in up to 60% of pre-malignant lesions and up to 83% of OSCC (Das and Nagpal, 2002). A number of different studies emphasise the frequent loss of gene expression or the presence of frequent *CDKN2A* gene mutations in the case of oral lesions, thus implying that this is an initial stage in oral carcinogenesis (Das and Nagpal, 2002).

In addition, several studies and reviews have been published on the identification of HNSCC candidate genes (Leemans *et al.*, 2011a). Table 1.1 shows a list of frequent genetic mutations and the cancer genes at these loci. These genes have been reported in several studies, where they have revealed mutations and/or deletions and/or they have shown a role in oncogenesis in frequently altered (>50%) chromosome locations. Several chromosome regions have been reported too, including allelic losses at 2q, 4p, 4q, 5q, 6p, 9q, 10q, 11q, 13q, 14q, 15q and 19q (Beder *et al.*, 2003), numerical gains at 5p, 8p, 9q, 17q, 19 and 20, and numerical losses at 1p, 4, 5q, 6q, 11q and 21 (Bockmuhl *et al.*, 1997), (Smeets *et al.*, 2006).

Table 1.1 Frequent genetic mutations in HNSCC candidate gene loci.

The table was adapted from (Leemans *et al.*, 2011a).

Chromosomal location	Gene	References
Tumour suppressor genes		
3p14	*FHIT*	(Virgilio *et al.*, 1996)
3p21	*RASSF1A*	(Hogg *et al.*, 2002)
8p23	*CSMD1*	(Sun *et al.*, 2001)
9p21	*CDKN2A*	(Smeets *et al.*, 2011),
9p23	*PTPRD*	(Reed *et al.*, 1996b)
10q23	*PTEN*	(Okami *et al.*, 1998)
17p13	*TP53*	(Somers *et al.*, 1992), (Brennan *et al.*, 1995b) (Smeets *et al.*, 2011)
18q21	*SMAD4*	(Bornstein *et al.*, 2009)
Oncogenes		
3q25	*CCNL1*	(Redon *et al.*, 2001)
3q25	*PARP1*	(Katoh and Katoh, 2003)
3q26	*PIK3CA*	(Okami *et al.*, 1998), (Murugan *et al.*, 2008)
3q26	*TP63*	(Rocco *et al.*, 2006), (Murugan *et al.*, 2008)
3q26	*DCUN1D1*	(Sarkaria *et al.*, 2006)
7p11	*EGFR*	(Sheu *et al.*, 2009)
7q31	*MET*	(Seiwert *et al.*, 2009)
8q24	*MYC*	(Rodrigo *et al.*, 1996), (Bockmuhl *et al.*, 1997), (Goessel *et al.*, 2005)
8q24	*PTK2*	(Agochiya *et al.*, 1999)
11q13	*CCND1*	(Inaba *et al.*, 1992), (Schuuring *et al.*, 1992), (Rodrigo *et al.*, 1996), (Rheinwald *et al.*, 2002), (Smeets *et al.*, 2011)
11q13	*CTTN*	(Schuuring *et al.*, 1992)
11q13	*FADD*	(Gibcus *et al.*, 2007)

1.1.5 Risk factors for oral cancer

Numerous risk factors for OSCC act mainly by enhancing and increasing the overall rate of mutations (da Silva *et al.*, 2011). These factors might include dietary influences, lifestyle habits (alcohol consumption and tobacco exposure), exposure to external agents, poor oral hygiene, genetic susceptibility, occupational activity and socioeconomic status (Das and Nagpal, 2002), (Hashibe *et al.*, 2009), (Hennessey *et al.*, 2009), (da Silva *et al.*, 2011). Moreover, there is also the likelihood that there are a number of other causative elements, which are yet to be established. Although alcohol drinking and tobacco smoking are the major etiological factors, oral cancer occurs also in patients who do not adopt these habits (Schmidt *et al.*, 2004).

1.1.5.1 Alcohol and tobacco

The harmful, negative impacts associated with excessive alcohol consumption and tobacco use have been well documented (Hashibe *et al.*, 2009), with the occurrence of OSCC amongst those who smoke as much as four to seven times greater than in those who do not smoke (Nozad-Mojaver *et al.*, 2009), (Poveda-Roda *et al.*, 2010). Despite the fact that both alcohol and tobacco are recognised as being independent risk elements, they have a synergistic impact with dose-dependent link between exposure frequency and duration, and the progression of the tumour (Tsantoulis *et al.*, 2007), (McCullough and Farah, 2008). Mucosa permeability is seen to increase as a result of ethanol, which facilitates the action of acetaldehyde, hydrocarbons and nitrosamines. Importantly, it is possible that carcinogenic agents can induce DNA mutations whilst also suppress the function of DNA repair enzymes (the most fundamental element against human cancer) (da Silva *et al.*, 2011). Brennan *et al*., for example, have provided support for the correlation between smoking and particular mutations of the tumour suppressor gene p53 (Brennan *et al.*, 1995a), although the precise mechanism of carcinogenesis might not be completely apparent (Tsantoulis *et al.*, 2007), (McCullough and Farah, 2008), (Hashibe *et al.*, 2009).

1.1.5.2 Viruses

HPV (Human Papillomavirus Virus) is also recognised as having a potential association with lifestyle, and is known to show a clear link to oropharyngeal cancer development. HPV, in the context of OSCC development, plays a role that is not well understood, but which could only involve a small portion of all cases, estimated at 5% (Braakhuis *et al.*, 2004), (Hennessey *et al.*, 2009). Unfortunately, however, some authors recognise cancers arising from the oral cavity and oropharynx as being combined to fall under the heading of 'oral cancer', as can be seen in the differences in reported incidence figures (da Silva *et al.*, 2011). OSCC genetic susceptibility (predisposition) is a critical consideration, particularly in the case of young individuals, and in line with inherited differences in a person's capacity to metabolise carcinogens, repair DNA, and control cell cycles, whether alone or in combination (Cloos *et al.*, 1996).

1.1.5.3 Betel

Moreover, a link can also be seen in the case of OSCC with the chewing of betel quid (areca nut being main component), which is known to be a habit of as many as 1 in 5 people across the globe (Cogliano *et al.*, 2004). Gene expression can be changed, and arecoline—one of the main alkaloids in the areca nut (the main component of betel quid)—could, through hypermethylation—cause tumour suppressor genes (TSGs), *p14*, *p15* and *p16* to be blocked, thus restricting the *p53* TSG, causing the repair of DNA to be restricted, and further inducing DNA damage responses in human epithelial cells to be triggered (Chen *et al.*, 2008b), (Takeshima *et al.*, 2008), (Tsai *et al.*, 2008). Other comparable chewing habits, such as khat use, for example, could similarly be identified in relation to OSCC in some communities (Fasanmade *et al.*, 2007), (Sawair *et al.*, 2007). A carcinogenic product, such as marijuana, is more debated in relation to OSCC aetiopathogenesis (Hashibe *et al.*, 2005).

1.1.5.4 Chewing tobacco and Shamma

Smokeless tobacco is one of the main risk factors associated with the high prevalence of oral potentially malignant diseases and head and neck cancers in South Asia (Gupta and Ray, 2003). Chewing tobacco is one of the forms of

smokeless tobacco. In a majority of Indian oral cancer patients, the etiology of oropharyngeal and oral cancers is chewing tobacco along with alcohol drinking and smoking (Kulkarni and Saranath, 2004). Shamma is a preparation mixture of powdered smokeless tobacco, ash, carbonate of lime, oil, black pepper and other flavouring additives, and is retained in the mouth as a quid (Yousef and Hashash, 1983), (Amer *et al.*, 1985). Oral cancer accounts for up to 20% of all diagnosed cancers in the south-western region of Saudi Arabia, as the use of white shamma is very common in that area (Ibrahim *et al.*, 1986), (Samman *et al.*, 1998). Also, a significant association has been found between the daily duration of shammah application with a specific dose-dependent manner in Yemeni users and the prevalence of oral leukoplakia (Scheifele *et al.*, 2007).

1.1.5.5 Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) can be identified as genome areas that have induced changes to the sequences of DNA, which might then not result in the alteration of amino acids, or otherwise sequences of DNA that do not induce negative effects in the case of healthy individuals, but; could be indicators in the tendency of diseases to occur, or could be used to genetically identify patients since SNPs have a tendency to cluster in relation to ethnic background (Scully and Bagan, 2009b). With specific consideration to TSGs, SNPs could also play a role in the development and progression of cancer (Drummond *et al.*, 2002), (Izzo *et al.*, 2003): for instance, and in cell-cycle, SNPs control pathway genes, namely *CCND1* splice variant P241P, and could play a role in the risk of potentially malignant lesions (Ye *et al.*, 2008b).

Oral cancer familial aggregation, potentially with the autosomal dominant mode of inheritance, has been recognised across a small number of individuals, although the genes responsible for this are, as yet, unidentified (Ankathil *et al.*, 1996). *p16* germline mutation was segregated with cancer predisposition in a single family with increased H&N cancer risk. Accordingly, there is the probability that the mutant p16 may be responsible in the case of HNSCC (Head and Neck Squamous Cell Carcinoma) tumourigenesis (Yu *et al.*, 2002).

1.2 Oral cancer and precancerous lesions: Potentially malignant oral disorders

The World Health Organization (WHO) has recommended using the term 'potentially malignant disorders' instead of 'potentially malignant conditions' and 'potentially malignant lesions' (Scully and Bagan, 2009a). Erythroplakia, leukoplakia, and lichenoid lesions are considered the most important potentially malignant lesions (Scully and Bagan, 2009a).

1.2.1 Leukoplakia

Leukoplakia is defined by the WHO as a white plaque or patch that is not attributable to a specific disease or cause, and requires taking a biopsy for histological examination (Pindborg JJ, 1997). Leukoplakia prevalence in the general population is less than 1% to 5%, and it is usually found in middle-aged and older men (Petti, 2003). Oral leukoplakia clinical phenotypes can range from thin homogeneous well-defined bordered white plaques to thick verrucous lesions (Kademani, 2007). It is the most frequently diagnosed oral premalignant lesion. Leukoplakia is also the most associated lesion with OSCC development (Mithani *et al.*, 2007). 2.9% to 17.9% of oral leukoplakia lesions have the potential for malignant transformation (Schepman *et al.*, 1998), (Liu *et al.*, 2010). Previous studies have shown up to 60% co-incidence of leukoplakia at the same time of OSCC diagnosis (Gupta *et al.*, 1980), (Bouquot *et al.*, 1988), (Reibel, 2003). Leukoplakia is further categorised based upon lesion heterogeneity, the more is the heterogeneity of a Leukoplakia lesion; the more is the possibility of malignant transformation (Schepman *et al.*, 1998).

Early studies on cancers of the upper aerodigestive tract defined loss of heterozygosity (LOH) as a process by which tumour suppressor genes in the genetic loci are eliminated. Increased LOH in the upper aerodigestive tract was correlated with histopathological progression of the disease (Califano *et al.*, 1996). Increased malignant potential in oral leukoplakia was associated with LOH of the 3p and 9p arms (Emilion *et al.*, 1996), (Mao, 1997), (Zhang and Rosin, 2001). Allelic loss of either the 3p or 9p arms have been found in 50% of leukoplakia lesions (Vanderriet *et al.*, 1994), (Rosin *et al.*, 2000). LOH at these loci alone are associated with 3.8-fold increased possibility of malignant

transformation, while further LOH at the 4q, 8p, 11q, 13q, and 17p arms are linked with 33-fold increased possibility of malignant transformation (Rosin *et al.*, 2000). In leukoplakia, LOH patterns are reflected in early carcinoma foci that are located within these lesions (Jiang *et al.*, 2001). *In situ* hybridization (ISH) has been used in oral leukoplakia to detect chromosomal replication. These studies show that most leukoplakia lesions have abnormal number of chromosomes 7 and 17 (Lee *et al.*, 1993), and lesions with trisomy 9 in more than 3% of cells had a significantly greater likelihood of cancer progression (Lee *et al.*, 2000). Furthermore, p53 expression was absent in normal oral mucosa by immunohistochemistry (IHC) while it has been identified in 90% of oral leukoplakia (Lippman *et al.*, 1995), though, it is unclear if this expression represents stabilized wild type p53 or the mutant protein.

1.2.1.1 Proliferative verrucous leukoplakia

Proliferative verrucous leukoplakia (PVL) was first reported in 1985 by Hansen *et al* (Hansen *et al.*, 1985). PVL is another aggressive, distinct form of oral leukoplakia with high rates of morbidity and mortality, and around 80% of the affected patients are women (Kademani, 2007). 56.2% of PVL lesions can transform to squamous cell carcinoma or verrucous carcinoma (Pentenero *et al.*, 2014). PVL has a high recurrence rate and may gradually spread to various oral sites (Bagan *et al.*, 2011). Histologically, early lesions exhibit only hyperkeratosis, then over time they may progress to become verrucous and commonly show variable degrees of epithelial dysplasia and a sudden change from hyperparakeratosis to hyperorthokeratosis, associated with verruciform or ridged surfaces (Radhakrishnan, 2011). PVL aetiology is still unknown, since tobacco does not seem to have a major role, nor do Candida species (Radhakrishnan, 2011). Furthermore, four previous studies have investigated the presence of HPV infection in case series (Gopalakrishnan *et al.*, 1997), (Fettig *et al.*, 2000), (Campisi *et al.*, 2004), (Bagan *et al.*, 2007). Neither Bagan *et al* nor Fettig *et al* identified any positive cases, while Gopalakrishnan *et a*l and Campisi *et al* reported an incidence of 20% to 24%, producing a pooled rate of 18% from a total of 91 PVL patients. PVL lesions are more common in older women. They most frequently appear in the gingiva (62·7%),

then the buccal mucosa (59·8%), followed by the tongue (49·1%) (Pentenero *et al.*, 2014).

1.2.1.2 Oral verroucous hyperplasia (OVH)

Oral verrucous hyperplasia (OVH) is a histological entity and precursor of oral verrucous carcinoma (OVC) that was first described by Shear and Pindborg (Shear and Pindborg, 1980b), although it was employed more than 20 years before by Ackerman and McGavran to describe a precursor to VC (Ackerman and Mc, 1958). OVH may transform into either an OVC or an OSCC (Wang *et al.*, 2009a). Shear and Pindborg, they described that 29% of OVH lesions also showed histological features of OVC (described in section 1.3.2). However, they separated these lesions according to the absence of invasive growth in OVH that is completely superficial and adjacent to normal epithelium. Very few studies have been published on OVH and the malignant transformation potential of verrucous hyperplasia lesions has not been inspected in detail (Shear and Pindborg, 1980b).

1.2.1.2.1 Incidence and risk factors for OVH

Until now, there have been only two clinicopathological analysis studies of OVH described by Wang *et al* and Zhu *et al*. OVH lesions are more common in 4th to 5th decade male patients; they occur mostly on the buccal mucosa and the tongue, and are usually highly associated with cigarette smoking, alcohol drinking, and the areca quid chewing habits (Wang *et al.*, 2009a), (Zhu *et al.*, 2012). However, Further studies are needed for oral verrucous lesions to fully evaluate the roles of the potential risk factors.

1.2.1.2.2 Diagnosis of OVH

OVH is a whitish or pink mass or an oral mucosal plaque with a papillary or verrucous surface (Figure 1.2) (Wang *et al.*, 2009a). Both OVH and PVL lesions can be histopathologically diagnosed as OVH since these two lesions demonstrate common epithelial hyperplasia features with verrucous surfaces (Figure 1.3) (Wang *et al.*, 2009a). Furthermore, OVH resemble OVC both histologically and clinically. Routine histological examination of hematoxylin and eosin (H&E) stained sections is currently the most reliable method to distinguish

between these entities; which is based on determining the endophytic and invasive growth pattern of OVC, from the exophytic growth pattern associated with OVH (Klieb and Raphael, 2007a). In1980, Shear and Pindborg described the histopathological key point features of oral verrucous lesions in 68 patients and indicated that OVH lesions are characterised by the superficial hyperplastic epithelium adjacent to normal epithelium, while OVC lesions are characterised by the hyperplastic epithelium with pushing-border invasion toward the underneath connective tissue with an intact basement membrane (Shear and Pindborg, 1980a).

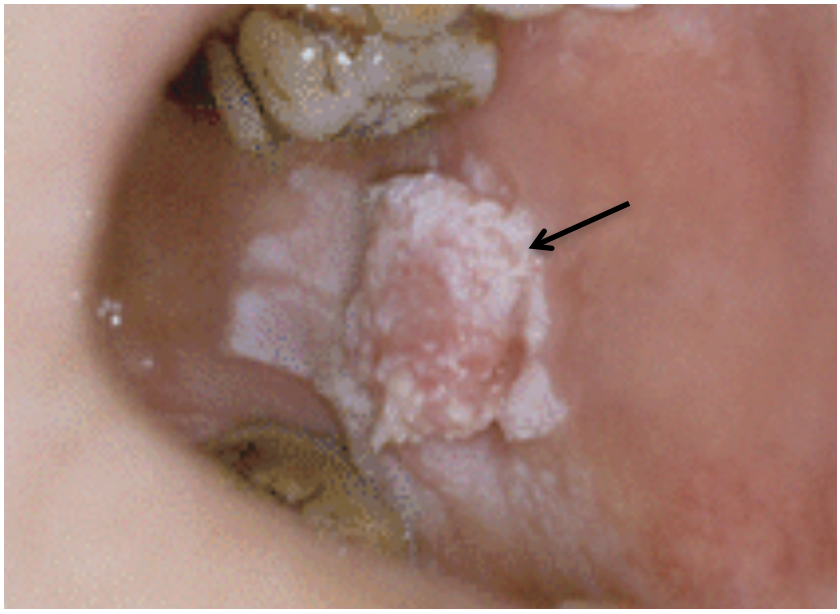


Figure 1.2 A clinical photograph of OVH lesion.

Note the whitish pink colour of the lesion. This figure was modified from (Wang *et al.*, 2009a)

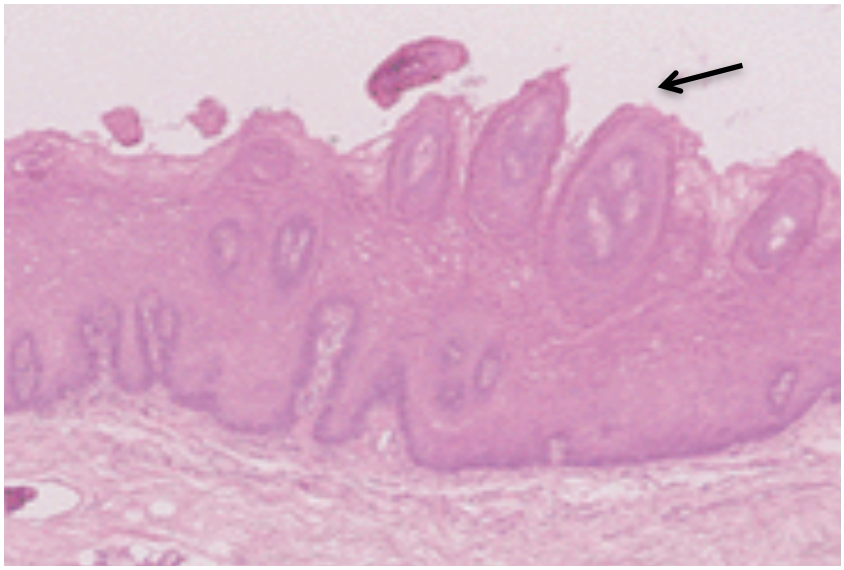


Figure 1.3 A histological photograph of OVH lesion.

Note the verrucous surface of the hyperplastic epithelium lesion. This figure was adapted from (Wang *et al.*, 2009a)

1.2.1.2.3 Treatment of OVH

OVH traditional treatment is total surgical removal of the affected lesion, however, this always causes a scar formation, especially for large OVH removed lesion (Chang and Yu, 2014). Cryotherapy was demonstrated previously as a treatment modality that could be used for OVH lesions (Yeh, 2000). The successful use of topical 5-aminolevulinic acid-mediated photodynamic therapy has been also shown before for the treatment of OVH (Lin *et al.*, 2010). Beside the relative lack of pain and scarring; being non-invasiveness and the low secondary infection incidence are all considered advantages of the two treatment modalities (Yeh, 2000) and (Dolmans *et al.*, 2003). Thus, a study in 2014 suggested using a combined treatment protocol of topical 5-aminolevulinic acid-mediated photodynamic therapy and cryotherapy since they has been shown to be effective in treating OVH lesions (Chang and Yu, 2014).

1.2.1.2.4 Previous molecular studies on OVH

A properly oriented histological H&E tissue sections is the gold standard way to distinguish between OVH and OVC lesions, however, the differentiation of these lesions is often difficult with poorly orientated specimens, small biopsies, and particularly, with biopsies that fail to show the margin of the lesion (Klieb and Raphael, 2007a). However, and even when a tissue section is available, the histological diagnosis of many OVH or OVC lesions is still challenging. An earlier study by Slootweg and Muller described that 25% of the lesions they examined could not easily be microscopically diagnosed as either VC or VH (Slootweg and Muller, 1983).

Therefore, and to distinguish OVH from OVC, a more pressing differential diagnosis is needed. An earlier study revealed a higher expression of glutathione S-transferase pi (GSTp)—a cytosolic acidic form of glutathione S-transferase that has a detoxification function and is a potential malignant transformation marker for some human cancers—and interleukin-1β in OSCCs and OVH lesions than in normal oral mucosa specimens. This study suggested a significant role of these two proteins in OVH transformation to OSCC (Chen and Lin, 1995), (Tsai *et al.*, 1999). Also, another study examined the expression of c-erbB-3 in 31 OVC and 18 OVH samples and suggested that c-erbB-3 expression was a malignancy index through progression from OVH to OVC and OSCC arising from VC (Sakurai *et al.*, 2000). Additionally, in 2001, a microsatellite analysis study was conducted to determine whether chromosomal regions frequently lost in SCC are lost as well in the VH/VC variant (Poh *et al.*, 2001). The results of this study showed high frequency of allelic loss in OVH lesions on seven chromosome arms (3p, 8p, 9p, 17p, 4q, 11q, 13q). They proposed that these findings might partly explain the potential malignant transformation of VH lesions (Poh *et al.*, 2001).

In further studies of OVH lesions, a high expression was found of both mRNA (60%), and inducible nitric oxide synthase (iNOS) protein (65–80%) (Chen *et al.*, 2002c), (Chen *et al.*, 2002a). Given that no mRNA or iNOS protein was found in normal oral mucosa samples, the authors suggested that the malignant transformation of OVH lesions may involve an iNOS-dependent mechanism.

Some previous studies have also been able to identify differences in Ki67 and P53 expression between OVC and both OSCC and OVH; less difference were noticed in the expression of retinoblastoma gene product (RBGP), p21, and p16 (Adegboyega *et al.*, 2005b), (Klieb and Raphael, 2007a). However, later studies failed to generate the same results (de Spindula *et al.*, 2011), (Lin *et al.*, 2011). A recent study conducted in 2014 investigated expression of p53, Ki67 and HuR in 17 OVC and six OVH samples using IHC. They detected an increase in Ki67 and p53 signals in OVC, while OVH presented rare positive signals. HuR diffuse staining pattern and epithelium expression were also observed (Habiba *et al.*, 2014).

1.2.2 Erythroplakia

Erythroplakia is described as an oral cavity red lesion that is not attributable to a specific disease, cannot be removed, and requires a biopsy for histological examination (Pindborg, 1997). This red plaque or patch could also have white areas surrounding or within the lesion and accordingly is termed as erythroleukoplakia (Kademani, 2007). Erythroplakia has a relatively rare incidence that ranges between 0.2%-0.8%, and occurs mostly in middle-aged men (Lumerman *et al.*, 1995). Clinically, erythroplakia is much more worrying than leukoplakia as lesions frequently shows a certain degree of dysplasia (Kademani, 2007). Alcohol intake and tobacco chewing are high risk factors for oral erythroplakia development (Hashibe *et al.*, 2000). Previous reported malignant transformation rate to OSCC ranged from 14.3% to 50.0% (Reichart and Philipsen, 2005). Another indication of the prognostic importance of oral erythroplakia is that earlier studies have found 85% to 90% of asymptomatic OSCC to be primarily reported as erythroplakia (Mashberg and Meyers, 1976), (Regezi JA, 1993).

Due to the histological similarities between oral leukoplakia and erythroplakia, it is expected to see subcellular changes that arise in oral leukoplakia also occur in erythroplakia, including aneuploidy, such as chromosomes 7 and 17 polysomy (Hittelman *et al.*, 1993). In addition, *p53* was frequently mutated in erythroplakia. A study in 1999 by Qin *et al* revealed a total of twelve *p53* mutations in 11 of 24 oral erythroplakia lesions, and these mutations have

altered p53 protein sequence (Qin *et al.*, 1999). However, the histological examination of the same samples did not show a significant difference between the frequency of *p53* mutations and the histological grade, which suggests an early *p53* mutation in OSCC development when the early lesion appeared as erythroplakia (Qin *et al.*, 1999).

1.3 Malignant lesions of the oral cavity

1.3.1 Squamous cell carcinoma

OSCC represents more than 90% of oral cancers (Johnson *et al.*, 2011). Despite the improvements in cancer diagnostic technologies over the past three decades, OSCC survival rates have not changed (Warnakulasuriya, 2009). Early detection of OSCCs can have an important impact on the disease therapy and management. Early OSCC usually presents as a pre-malignant lesion, which could be a red patch (erythroplakia), white patch (leukoplakia), or a combination of both white and red lesions (erythroleukoplakia). Over time, mucosal surface superficial ulceration may possibly develop (Neville and Day, 2002). OSCC lesions could range from millimeters to several centimeters in advanced cases. Primary lesions are generally asymptomatic since they are small (Bagan *et al.*, 2010a). The clinical appearance of these primary malignant lesions usually takes the form of an erytholeukoplastic lesion (Mashberg *et al.*, 1989). It contains red or white and red areas with a well-defined slight roughness and a change in the soft tissue elasticity to a harder sensation on palpation (Bagan *et al.*, 2010a). The typical features of oral carcinomas include nodularity, fixation to underlying tissues and ulceration (Neville and Day, 2002), (Scully and Bagan, 2009b). Histological features of all head and neck well-differentiated SCC following H&E staining are characterised by a pink cytoplasm with nests of squamous cells with variable degrees of squamous differentiation and marked cyto-nuclear atypia, intercellular bridges and keratin pearl formation, set in a background of stromal fibrosis (Pai and Westra, 2009).

The clinical features of OSCC are easily recognised in the advanced stages. However, a biopsy must always be taken to confirm the clinical diagnosis (Silverman, 1988). In the first disease stages, early diagnosis can improve patient survival rates by up to 80–90%, besides minimising surgery (Bagan *et*

al., 2010a). In general, it has been recommended to remove or destroy leukoplakias that reveal moderate epithelial dysplasia (Scully and Bagan, 2009b). The management of mild dysplasia lesions depends on the location, size, and the most probable cause of the lesion. Early dysplastic lesions can be occasionally reversed if the irritation source (e.g. heavy smoking) is eliminated (Scully and Bagan, 2009b). Treatment choices are variable and depend on the location and the size of the primary tumour, presence or absence of distant metastases, lymph node status, the patient's capability to tolerate treatment, and the patient's wishes (Scully and Bagan, 2009b). Surgery and/ or radiation therapy still remains the standard treatment modality for lip and oral cavity cancers (Neville and Day, 2002).

The histological progression and development from a simple squamous hyperplasia lesion throughout squamous dysplasia to invasive OSCC is driven by the progressive accumulation of genetic changes. Some alterations arise earlier than others, such as LOH at chromosomal loci 3p and 9p (Pai and Westra, 2009). A very recent systematic review aimed to identify the frequency of common genomic CN alterations (CNAs) in OSCC by involving 12 previous array comparative genomic hybridization (aCGH) studies. Furthermore, they revised the literature dealing with CNAs that are behind the development of oral premalignant lesion (OPL) to OSCC (Salahshourifar *et al.*, 2014). With the progression of the disease, the sequential accumulation of genomic alterations from OPL to OSCC increases in terms of type, size, and frequency of the abnormalities. Losses in 3p (37%), 8p (18%), 18q (11%), 9p (10%) and gains in 8q (47%), 11q (45%), 3q (36.5%), 20q (31%), 5p (23%), and 7p (21%) were the most common observations from the conducted systematic review (Salahshourifar *et al.*, 2014).

1.3.2 Verrucous carcinoma (VC)

In 1948, Ackerman defined VC; it is also known as Ackerman's tumour or verrucous carcinoma of Ackerman (Ackerman, 1948). VC is currently classified as a low grade, slow growing, non- metastasizing, rare variant of SCC (Barnes L, 2005) and affects both skin & mucosal sites.

1.3.2.1 Incidence of Oral Verrucous Carcinoma [OVC]

OVC accounts for 2-10% of all OSCC cases (Pentenero *et al.*, 2011). In terms of epidemiology, OVC is mostly seen in over sixth decade males (Oliveira *et al.*, 2006), (Walvekar *et al.*, 2009), (Alkan *et al.*, 2010b), (Rekha and Angadi, 2010), (Ray *et al.*, 2011b), (Zhu *et al.*, 2012). Furthermore, previous studies reported that the most common site for OVC is the buccal mucosa (Yeh, 2003), (Walvekar *et al.*, 2009), (Rekha and Angadi, 2010). However, the most affected areas in Alkan *et al's* study were the mandibular area followed by buccal mucosa (Alkan *et al.*, 2010b), and the predominant site of OVC lesions in Zhu *et al's* study was lower lip (Zhu *et al.*, 2012). These site variations were perhaps due to differences in the geographic locations and the ethnic populations between studies. For example, and based on the anatomical location, the lower lip is a sun-exposed area, which suggest the predominance of this oral affected site in Zhu *et al* OVC study. Also, this implies a possible aetiology of ultraviolet radiation for OVC (Zhu *et al.*, 2012).

1.3.2.2 Risk factors for VC

The aetiology of VC is not well known (Ray *et al.*, 2011b), though, it has been suggested that OVC develop from premalignant lesion (Shear and Pindborg, 1980a), (Bagan *et al.*, 2010b). Smoking appears to be related with the development of H&N mucosal VC (Alkan *et al.*, 2010a). The presence of leukoplakic lesions, and poor oral hygiene may also act as predisposing factors (Alkan *et al.*, 2010a). In Asia, bidis and cigarettes smoking is known to be associated with leukoplakia, areca quid, paan and miang chewing habits have also been found (Chung *et al.*, 2005). Chung *et al* reported that 55.6% (five out of 9 patients) of the patients with verrucous lesions were areca quid chewers, and they suggested that areca quid chewing could be a major causative factor for these lesions in Taiwan (Chung *et al.*, 2005). Alkan *et al* reported that 50% of the patients smoked tobacco (six out of 12 patients), and hence, they suggested that cigarette smoking appeared to be the major risk factor in their OVC patients (Alkan *et al.*, 2010a). However, in studies by Walvekar *et al* and Zhu *et al*, alcohol consumption and tobacco smoking and chewing were not identified to be statistically significant in OVC patients (Walvekar *et al.*, 2009), (Zhu *et al.*, 2012).

Since a verrucous appearance is suggestive of viral aetiology, this has prompted a number of investigations to study the putative association between HPV and those lesions (Stokes *et al.*, 2012). HPV has been cited as a probable aetiology in VC pathogenesis by various authors. However, the majority of the studies that investigated 'HPV presence' in verrucous lesions (as listed in Table 1.2) relied on polymerase chain reaction (PCR), and *in-situ* hybridization (ISH), for detection and did not identify HPV transcriptional activity markers or quantitate HPV viral load (Miller and Johnstone, 2001). The high sensitivity of the PCR technique can amplify very small quantities of HPV DNA and this can lead to detection of non-pathologic HPV infections or false-positives if sample contamination occurs (Ha *et al.*, 2002), (Kreimer *et al.*, 2005). Prior to and until 1997, 15 published studies investigated the presence of HPV DNA in OVC (Kari J. Syrjänen, 2000). Among the 159 samples analysed, HPV DNA was identified in 37.7% of the cases, and HPV subtypes 6 and 11 were the most predominant identified HPV infections (47%) (Kari J. Syrjänen, 2000). The possible role of HPV in VC pathogenesis is suggested by HPV incidence in VC cases, which varies from 30% to 100% (refer to Table 1.2 below). This range indicates that HPV prevalence in oral verrucous lesions and its actual role in cancer pathogenesis is controversial and inconclusive. This variation can be attributed to the deficiency of standardized detection procedures and the difficulty in defining complete histological criteria for OVH and OVC cases. Furthermore, the rarity of these types of lesions makes it difficult to study them in details, while most previous studies or case reports were based on small number of cases.

Table 1.2 Previous studies of HPV detection in OVC.

Diagnosis	Number of cases	Number of HPV-positive lesions	HPV detection method	Identified HPV genotypes	Reference
VC	9	3/9 (33.3%)	ISH	HPV2	(Adlerstorthz et al., 1986)
VC	25	12/25 (48%)	PCR	HPV6, 11,16,18	(Nobletopham et al., 1993)
VC	17	7/17 (41.2%)	PCR, ISH	HPV6, 11,16,18,31, 33, 35	(Shroyer et al., 1993)
VC	15	10/15 (66.7%)	PCR	HPV6, 11,16,18	(Balaram et al., 1995)
VC	4	4/4 (100%)	PCR	HPV2, 18, 20, 27, 57, 62, and partial sequences of an unknown HPV type	(Mitsuishi et al., 2005)
VC	23	11/23 (47.8%)	PCR, ISH (They used 10 non-neoplastic control lesion)	HPV6, 11, 18, 33, 74	(Mitsuishi et al., 2005)
VC Dysplastic verrusous lesions	7 13 (N verrucous lesions: 20)	1/7 5/13 (N +ve verrucous lesions: 6/20) (30%)	ISH, PCR	HPV16	(Stokes et al., 2012)

VC: verrucous carcinoma; VH: verrucous hyperplasia; PCR: polymerase chain reaction; ISH: In situ hybridization; N: number.

1.3.2.3 Diagnosis of OVC

Generally speaking, OVC clinico-histopathological diagnosis is usually exclusionary and extremely difficult, though, it has a better prognosis compared to other carcinomas (Ray *et al.*, 2011b). 'Verrucous' terminology is applied to lesions that show exophytic keratotic surfaces made of blunt or sharp epithelial projections, filled with keratin invaginations, but without clear fibrovascular cores (Ray *et al.*, 2011b). Histologically, OVC consists of thickened, club-shaped papillae and blunt stromal invaginations of well-differentiated squamous epithelium with marked keratinization (Figure 1.4), with the squamous epithelium lacking cytological criteria of malignancy. OVC invades underlying stroma with a pushing, rather than infiltrating front (Barnes L, 2005).

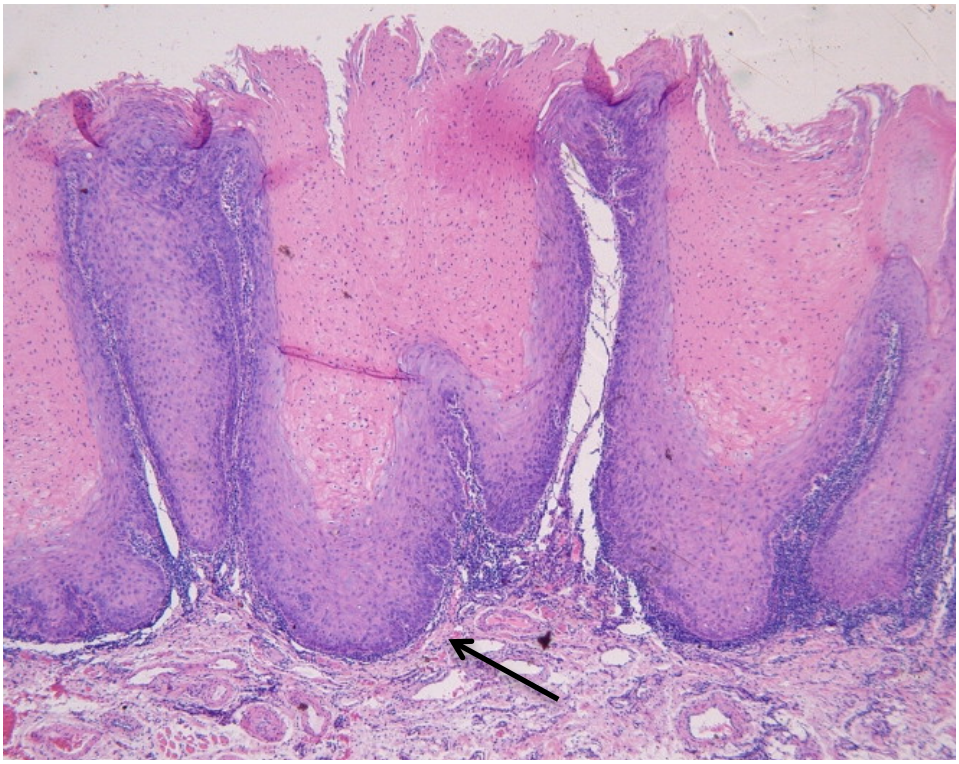


Figure 1.4 A histological photograph of OVC lesion.

Note the ‘elephant feet-‘ like down-growth with abundant parakeratin production. This figure was modified from (van Heerden and van Zyl, 2009)

Clinically, VC appears as an exophytic mass with—as the name suggests—verrucous surface (Rekha and Angadi, 2010). In general, OVC histopathological diagnosis is ambiguous and difficult, particularly, when dealing with poorly orientated specimens and small superficial biopsies. As they show an exophytic ‘warty’ growth pattern, it is not uncommon to sample only the most superficial area of the tumours for the biopsy specimens; and hence, the histopathological diagnosis by light microscopy will be difficult at the time of initial biopsy examination (Devaney *et al.*, 2011b). Not all H&N, white verrucous lesions will prove to be VC; VH, squamous papilloma, and viral verruca all enter into a differential diagnosis (Devaney *et al.*, 2011b). Additionally, conventional SCC might be also surmounted by a verrucous surface, and hence, will show greater degree of mitotic activity and cytologic atypia and will behave more aggressively (Rekha and Angadi, 2010). The distinction of OVC from classical OSCC is a common problem for both pathologists and clinicians (Ray *et al.*, 2011b).

1.3.2.4 Previous molecular studies of verrucous carcinoma

OVC histological features can lead to misclassification during histopathological examination (Devaney *et al.*, 2011b). For this reason, new approaches are needed to evaluate biological biomarkers that are required for better understanding of the development and progression of OVCs. Furthermore, and because of the different features such as metastatic potential, morbidity and mortality rates, and treatments response, OVC and OSCC lesions must be correctly distinguished by the pathologist (Saito *et al.*, 1999b), (Yoshimura *et al.*, 2001) (Pereira *et al.*, 2007).

Genomic changes such as CN alterations and point mutations, gene expression changes, as well as epigenetic changes have been discovered previously in OSCC, which could help in the development of biomarkers and assist in clinical choices and decisions (Tuch *et al.*, 2010), (Gibb *et al.*, 2011a), (Zhang *et al.*, 2013). However, previous VC molecular studies have yielded mixed results (Devaney *et al.*, 2011b). GimenezConti *et al* (1996) investigated by IHC the expression of cyclin D1, p53 and Rb in 29 OVC cases. They reported p53 accumulation in 15 OVC samples and suggested possible gene mutations. They also reported cyclin D1 overexpression (61%) and no Rb staining alterations, and suggested that Rb function may be inactivated by HPV infection or overexpression of cyclin D1. In the case of p53, immunohistochemical protein expression does not always mean the presence of a mutation. Nevertheless, Angadi and Krishnapillai (2007) reported no statistical significance in the overexpression of cyclin D1 in 29 out of 41 well-differentiated OSCC cases and in 19 out of 30 OVC cases. In 1999, Saito *et al.* examined the expression of the cell cycle-associated proteins p53, p16, p27, pRb and Ki-67 using IHC in 15 OVC and 44 OSCC samples (Saito *et al.*, 1999a). They reported differences in expression levels of p53, p27 and Ki-67 that increased in OSCC when compare to OVC and suggested that this might reflect differences in cell cycle-associated proliferative activities between both tumours. However, they noted high expression levels of both pRb and p16 in OVC when compared to OSCC samples and proposed a possible relationship between HPV infection and the formation of OVC lesions (Saito *et al.*, 1999a). Similarly, in studies by Mohatasham *et al* study, p53 expression was found to have increased gradually

from the normal oral mucosa control group toward OVC and then to OSCC (Mohtasham *et al.*, 2013). However, and unlike Saito *et al's* study, Zargaran *et al* (2012) examined the immunohistochemical protein expression of Ki67 in 15 cases of OVC versus 15 cases of OSCC and concluded that it was not a reliable marker to evaluate invasion level and differentiate OVC from OSCC. Ki67 was also not found to be a good marker in 20 cases of OVC compared with 42 cases of OSCC confirming the earlier findings of Mohatasham *et al* (Mohtasham *et al.*, 2013).

Impola *et al* (2004) investigated the expression of MMP-7 and MMP-9 in 15 OSCC and 15 OVC ceases and attributed the pattern of MMPs expression, which increased from OVC toward OSCC, to the less invasive nature of OVC. This finding was consistent with that of Mohatasham *et al*, as they reported the same pattern of differences between OVC and OSCC expressing MMP-2 and MMP-9 (Mohtasham *et al.*, 2013). Though, and unlike the observation of Mohatasham *et al*, Tang *et al* (2005) found that MMP-2 immunohistochemical protein expression was more frequent in 10 OVC cases compared to 15 OSCC cases. In addition, Oliverira *et al* studied the protein expression of cytokeratins 10, 13, 14 and 16 in eight OVC samples using IHC and concluded that OVC cytokeratins profile was similar to OSCC cytokeratins profile reported in the literature (Oliveira *et al.*, 2005).

In a study in 2010, immunohistochemical protein expression in some basement membrane elements (in particular, laminin, laminin-5, fibronectin, and collagen IV) showed quantitative differences between 20 OVC cases, on the one hand, and ten OSCC cases, on the other hand (Arduino *et al.*, 2010). The study reported less intensive laminin staining pattern in SCC when compared with VC. However, these differences are not precise enough to be used in differential diagnosis (Arduino *et al.*, 2010). In 2011, a study was conducted to investigate the differences in chromosomal instability (CIN) biomarkers between nine OVC cases and 25 OSCC lesions (Pentenero *et al.*, 2011) The study suggested that OVCs are characterised by a lower CIN and tumour heterogeneity degree than OSCCs (Pentenero *et al.*, 2011).

Since loss of basement membrane can be linked with stromal invasion and metastatic progression, cell migration and destruction of basement membrane were evaluated by Zargaran *et al* using IHC in 15 cases of OVC and 15 cases of OSCC (Zargaran *et al.*, 2011b). There was no significant difference in type IV collagen staining signals between OVC and OSCC groups. However, the study suggested that the expression of Ln-332 $_{Y}$2 chain in over 5% of the cells supports OSCC diagnosis and it may provide further predictive and diagnostic information for clinicians and pathologists (Zargaran *et al.*, 2011b). In addition, Quan *et al* (2012) investigated the anti-apoptotic function of αB-crystallin in 17 OVC samples using 15 OSCC samples as a control group (Quan *et al.*, 2012). When comparing the OSCC cases with OVC cases, immunohistochemical staining of αB-crystallin was lower in OVC compared to OSCC, and hence, they suggested that this might partially illustrate the less aggressiveness behaviour of OVC when compared with OSCC.

The possible link between the biological behaviour and the structural and functional features with the clinical outcome in OVCs is still not clear and requires further investigation. The rarity of OVC lesions also makes them difficult to investigate. Using different samples, sample numbers, variations in clinical diagnosis, difficulties in defining 'gold-standard' histological criteria for diagnosing verrucous lesions, and different staining procedures and analysis methods may explain the lack of concordance between these studies.

1.3.2.5 Treatment of OVC

The traditional treatment of OVC is complete surgical excision of the lesion (Ferlito and Recher, 1980), (McCoy and Waldron, 1981), (McDonald *et al.*, 1982), (Kang *et al.*, 2003), with radiation treatment applied in certain conditions of either poor surgical candidates or for extensive disease cases (Kolokythas *et al.*, 2010). However, ideal treatment of OVC is still debatable, and there is no worldwide agreement (Karagozoglu *et al.*, 2012). Anaplastic transformation after radiotherapy has been reported in OVC cases (Karagozoglu *et al.*, 2012). However, anaplastic changes have been reported as well in an untreated OVC and after resection; which may be a consequence of a misleading histopathological diagnosis (Karagozoglu *et al.*, 2012). Patients with a

verrucous lesion coexistent with a conventional SCC must be treated as if they had invasive SCC (Sheen *et al.*, 2004). The efficiency of chemotherapy treatment for OVC lesions is still not well defined, and only a small series and few case studies have been reported (Dame *et al.*, 1974) (Sheen *et al.*, 2004), (Wu *et al.*, 2008). Karagozoglu *et al* described the use of chemotherapy treatment with methotrexate, which was given to 12 OVC patients who had not been conventionally treated because of their poor general condition and the extent of their oral verrucous lesions. Treatment with methotrexate was useful in 11 out of 12 patients and only one patient failed to respond. They suggested that the use methotrexate chemotherapy treatment alone might improve quality of life and minimise morbidity, especially among old patients (Karagozoglu *et al.*, 2012).

1.3.3 Hybrid verrucous carcinoma

VC has two types: classic and hybrid. The classic variant is the more common type (Devaney *et al.*, 2011b). A study by Medina *et al* (1984) reported the coexistence of SCC foci within OVCs in a review of 104 cases (Medina *et al.*, 1984). This supports the importance of taking an adequate depth for the initial biopsy in order to evaluate the verrucous tissue for the presence of SCC (Kolokythas *et al.*, 2010). Hybrid verrucous carcinoma is a mixed tumour that contains verrucous and classical squamous cell carcinoma and has the capability to metastasise (Devaney *et al.*, 2011b). About 20% of all OVC and 10% of all laryngeal VC are of the hybrid type (Medina *et al.*, 1984), (Orvidas *et al.*, 1998), (Kolokythas *et al.*, 2010). Hence, All VCs should be carefully assessed to exclude any potential hybrid variant, and clinicians as well as pathologists must be aware of it (Devaney *et al.*, 2011b). The hybrid tumours behave like conventional SCC, and accordingly, treatment protocols of these tumours should be also as such (Thomas and Barrett, 2009).

1.4 Next generation sequencing in cancer biology

It is commonly believed that cancers result from the accumulation of genetic mutations. During the past half-decade, the development of next-generation sequencing (NGS) technologies has enabled high sensitivity and resolution studies of cancer genomes through whole-exome and whole-genome

sequencing approaches (Haimovich, 2011). Analytical methods are now applied to identify numerous somatic genome alterations, including insertions/ deletions, nucleotide substitutions, chromosomal rearrangements, and copy number variations (Haimovich, 2011). These current sequencing methods have succeeded Sanger sequencing (Cabelguenne *et al.*, 2000), while substituting microarray analysis as the genotyping and discovery platform (Alkan *et al.*, 2011). Since NGS technologies are improving, sequencing costs will continue to decrease, allowing this technology to be more available for the use in cancer biology studies (Kozarewa *et al.*, 2012). Ultimately, this will allow researchers to understand and study a wide panel of gene mutations within malignant tumours that will lead further towards the personalised medicine era in the near future.

Capillary-based dideoxy-terminator methods were used in conventional Sanger sequencing. NGS refers to approaches developed following to Sanger automated method, which rely on the preparation of the DNA template, sample sequencing, imaging, and genome alignment and data assembly (Feldman, 1973). First, DNA library is constructed by fragmenting the genomic DNA into shorter segments. These are sequenced afterward by detecting emitted signals from each fragment, while they are re-synthesised from the template DNA strand. This is performed in parallel reactions millions of times and the generated reads are aligned next using the reference known genome (Feldman, 1973). The fact that NGS is relatively a new technology is one of its limitations. The produced errors are not well known; it is also a costly technique from sample preparation to the analysis of the results (Wood *et al.*, 2010).

1.4.1 Cancer samples-specific considerations

Cancer samples have some characteristics that are different from other tissue samples and from germ line inherited genomic sequences, which require specific consideration in NGS analyses (Meyerson *et al.*, 2010). Additionally, cancer nucleic acids are usually of lower quality than blood purified nucleic acids. One reason for this is biological: tumour samples often include apoptotic or necrotic cell fractions that lowers the average quality of the nucleic acid. A second reason for this variance in the quality of the nucleic acid is technical: most tumour tissue samples are FFPE materials from the histopathological

microscopic examination. Nucleic acids extracted from FFPE materials may be degraded and have probably went through crosslinking (Gilbert *et al.*, 2007). NGS analysis of nucleic acids extracted from FFPE specimens may require certain experimental (Wood *et al.*, 2010) and computational procedures to overcome the increased mutational background rate (Marchetti *et al.*, 2006), (Ruiz *et al.*, 2007). Thus, experimental procedures must be adapted too to account for this (Meyerson *et al.*, 2010). Notably, the many-fold coverage by NGS can allow the production of high-quality data from lower quality tumour materials (Thomas *et al.*, 2006). Furthermore, cancer samples have a mixture of non-malignant and malignant cells and, consequently, a mixture of normal and cancer genomes (Meyerson *et al.*, 2010). Also, the tumours themselves can be very heterogeneous and composed of a diversity of clones resulting from their genomic instability leading to different genomes (Navin *et al.*, 2010). Hence, NGS analytical models for cancer genomes must take these two heterogeneity types into account in their genome alterations prediction analysis (the within-cancer heterogeneity and cancer versus normal heterogeneity) (Meyerson *et al.*, 2010).

1.4.2 NGS experimental approaches

Next-generation sequencing can be used with cancer samples in many ways (Meyerson *et al.*, 2010). These vary according to the type of the input material (e.g. DNA or RNA), the proportion of the targeted genome (a subset of genes, the whole genome, or transcriptome) and the type of the studied variation (gene expression, point mutation, or structural change) (Meyerson *et al.*, 2010). The use of these approaches with large sample numbers may lead to the discovery of the main recurrent cancer translocations. The rearrangements that can be identified by next-generation sequencing include structural chromosomal rearrangements (e.g. inversions, deletions, duplications, and reciprocal and non-reciprocal chromosomal rearrangements), non-endogenous sequences insertions such as viral sequences and complex rearrangements involving combinations of several events (Meyerson *et al.*, 2010).

1.4.2.1 Whole-genome sequencing

In 2008, the first cancer whole genome nucleotide sequence was described

from an acute myeloid leukaemia compared with normal skin DNA extracted from the same patient (Ley *et al.*, 2008). Since that time, more complete cancer genomes sequences along with the matched normal genomes have been described and this number is growing rapidly (Mardis *et al.*, 2009), (Pleasance *et al.*, 2010a), (Pleasance *et al.*, 2010b), (Ding *et al.*, 2010a), (Puente *et al.*, 2011), (Link *et al.*, 2011), (Fujimoto *et al.*, 2012), (Ding *et al.*, 2012), (Dulak *et al.*, 2013), (Wang *et al.*, 2014a), (Morrison *et al.*, 2014). Complete genome sequencing of DNA from cancer tissue through utilising germline DNA sequence taken from the same patient for comparison, will identify the complete series of genomic alterations (including copy number alterations, structural rearrangements, and nucleotide substitutions) in a single approach (Ley *et al.*, 2008), (Mardis *et al.*, 2009), (Pleasance *et al.*, 2010a), (Pleasance *et al.*, 2010b), (Ding *et al.*, 2010a), (Agrawal *et al.*, 2011), (Ross *et al.*, 2014). Thus, whole-genome sequencing provides the most inclusive characterisation of the cancer genome (Meyerson *et al.*, 2010). However, since it involves the highest sequencing amount, it is also the most costly. The detection of structural chromosomal rearrangements is one of the major potentials of cancer whole-genome sequencing as compared to exome sequencing (Meyerson *et al.*, 2010).

1.4.2.2 Whole-exome sequencing

NGS approaches such as whole exome sequencing (WES) have greatly clarified the genetic alteration landscape in several tumour types and provided biological understandings relevant to clinical contexts (Garraway and Lander, 2013). The high throughput, decreased cost, and increased practical accessibility of tumour genomic profiling in clinical oncology has offered opportunities to test the hypothesis: ‘precision medicine’ (Garraway, 2013). Furthermore, The affordability of achieving higher coverage for large sample numbers makes WES highly suitable for mutation detection in mixed purity cancer samples (Meyerson *et al.*, 2010). In general, knowledge of changes in the coding exons regions of all genes can suggest treatment selections and further therapeutic options (Garraway and Janne, 2012). Several challenges remain to widespread clinical application of exome sequencing (Van Allen *et al.*, 2014). One challenge includes generating WES high throughput data from

FFPE archival tumour tissue material (Goetz *et al.*, 2013). A second one involves interpreting clinical exome sequencing data to be used for clinical and biological investigation (Van Allen *et al.*, 2014). A third includes developing a method to question reasonably actionable variants of undefined significance. Overcoming the previously mentioned challenges of WES would inform selected experimental follow-up and guide clinical decision-making (Van Allen *et al.*, 2014).

1.4.2.3 Transcriptome sequencing

NGS of the transcriptome—as cDNA derived from total RNA, mRNA or other RNAs as microRNA—is a powerful technique for understanding molecular changes in cancer (Meyerson *et al.*, 2010). This has improved the possibility of characterising different tumours at the molecular levels across the whole genome. Transcriptome sequencing or RNA-Seq is an efficient and sensitive methodology to identify intragenic fusions that lead to oncogene activation (Maher *et al.*, 2009), (Berger *et al.*, 2010). RNA-Seq can also be used to identify somatic mutations, however, finding a normal matched sample for assessment and comparison is a challenge, since it is unlikely that normal tissue will express the same genes exactly as the tumour sample (Meyerson *et al.*, 2010). Also, mutation detection is hampered in genes expressed at low levels due to the lack of statistical power. Additionally, the possibilities of RNA editing and reverse transcriptase errors require consideration (Shah *et al.*, 2009).

Remarkable recent advances in NGS technology provided massive data volumes that can identify genetic variations in individuals' genomes and even through using FFPE samples (Sinicropi *et al.*, 2012), (Zhang *et al.*, 2013). It has been demonstrated from the application of older methods, such as RT-PCR and DNA microarray that gene expression profiles (RNA transcripts levels) can categories patients and predicts their outcomes in a range of different diseases, providing new insights for many significant clinical tests (Mehra *et al.*, 2007), (Mehra *et al.*, 2008), (Chudova *et al.*, 2010). Although significant gene expression differences have been previously identified using microarray analysis, (Ye *et al.*, 2008a), (Estilo *et al.*, 2009), (Han *et al.*, 2009), this

technique has a limited sensitivity in analysing the transcriptome (Tuch *et al.*, 2010). As an alternative, transcriptome sequencing is a deep sequencing technology, which is widely used for transcriptomic profiling now because of its reasonable costs (Zhang *et al.*, 2013). When compared with microarray, RNA-Seq provide much more accurate measurement of gene expression levels and more advanced categorisation of transcript isoforms (Mortazavi *et al.*, 2008), (Wang *et al.*, 2009c).

1.4.2.4 Copy number Analysis

Array-based techniques have been a powerful approach to identify copy number alterations pattern in cancer, from focal deletions and amplifications to whole chromosome or chromosome arms gain or loss that could range in size from ten kilobases to ten megabases (Bignell *et al.*, 2004), (Zhao *et al.*, 2004), (Beroukhim *et al.*, 2007), (Beroukhim *et al.*, 2010), (Bignell *et al.*, 2010). One of the applications of NGS technology is to identify DNA copy number (CN) changes by sequencing at relatively low coverage but with a higher resolution compared to array-based comparative genomic hybridisation (aCGH) (Wood *et al.*, 2010). Unlike aCGH, NGS CN analysis can be used with nanogram DNA quantities extracted from challenging tissue samples such as FFPE materials. Also, the obtained signals from the DNA molecules are directly read out rather than inferred from hybridisation (Wood *et al.*, 2010). NGS techniques offer considerable benefits for copy number analysis (CNA), including precise delineation of the CN breakpoints, and higher resolution (can detect single-base insertions or deletions) of CN changes (Meyerson *et al.*, 2010). NGS enables the estimation of tumour-to-normal CN ratio at a genomic locus through counting the number of reads at this locus in normal and tumour samples (Meyerson *et al.*, 2010). Copy number variation (CNV) data are produced by analysing aligned read distributions to a reference genome (Ding *et al.*, 2010b). Identifying Copy Number Alterations in cancer cells is an essential step toward determining chromosomal regions with breakpoints and to access chromosomal rearrangements severity. Additionally, comparison of CN genomic profiles between tumours from different patients can define common lost or duplicated regions to highlight the positions of oncogenes or tumour suppressor genes (Hartwell and Kastan, 1994a).

1.4.3 Genomics of HNSCC

Next-generation sequencing research studies of HNSCC biology have improved the understanding and provided clearer insight of the etiological and molecular aspects of HNSCC. Four key papers have now been published on HNSCC genomics. The first next generation whole exome sequencing was performed previously to identify the mutational landscape and events in key cell cycle components of HNSCC patient tumours (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). Further discoveries involved cell differentiation pathways mutations, mainly, in *FXBW7* and *NOTCH1* (Agrawal *et al.*, 2011). These studies indicated that up to 20% of tumours have *NOTCH1* loss-of-function mutations and >80% contains *TP53* mutations. Additionally, WES data of HNSCC collected tumours in Lui *et al's* study showed that mutational events in PI3K pathway are the most frequent among all mutated oncogenic pathways in HNSCC (30% of HNSCCs have genomic mutations in this pathway) (Lui *et al.*, 2013). Furthermore, Pickering *et al* performed integrated genomic analysis of copy number, gene expression, point mutations, and methylation in OSCC (Pickering *et al.*, 2013). Comprehensive genomic analysis identified four main driver pathways (Notch, TP53, cell cycle, and mitogenic signalling) and two other important genes (*CASP8* and *FAT1*) (Pickering *et al.*, 2013). Loss of *CDKN2A* and amplification of *CCND1* are two of the most common genomic changes in OSCC that were found in 94% of the tumours in Pickering *et al* study. Alterations in the Notch pathway were detected in 66% of patients, and in follow-up mechanism studies, NOTCH1 functional signalling inhibited OSCC cell lines proliferation. *FAT1* was also frequently mutated in 30% of OSCC patients. They discovered as well a new OSCC molecular subtype associated with frequent mutation *of HRAS* and *CASP8* and with fewer total CN changes (10% of OSCC tumours harboured *CASP8* mutation. mutations *in CASP8* were strongly associated with mutations in *HRAS*) (Pickering *et al.*, 2013).

Selection of efficient cancer therapy requires knowledge of tumour biomarkers and their roles in the biological processes leading to cancer. It is well known that the accumulation of several genetic changes in molecular pathways and in different genes is the main cause of OSCC development (Gibb *et al.*, 2011b). A recent conducted research has used RNA-Seq data of three paired tumour and

matched normal tissues to analyse somatic mutations in OSCC (Zhang *et al.*, 2013). They identified 156 tumour-specific disruptive genes and 515 significantly mutated genes, with six genes in both groups, including *GTF2H5, ANKRA2*, *TAF1L, NUP37, PPP1R26, and STOML1*. Their pathway and gene ontology analysis also suggested that significantly mutated genes were enriched in cell adhesion, which is an indicative of cancer development (Zhang *et al.*, 2013). The application of NGS approaches will allow researchers to understand and study gene mutations within malignant tumours that will eventually also lead towards personalised medicine and

1.5 Aims of the present studies

It is known that molecular changes drive the cellular phenotype of any tumour. Until now, most the previously reported molecular studies of oral verrucous lesions (including OVH and OVC) have inspected candidate genes rather than taking a complete genome wide approach. The study of HNSCC biology using NGS techniques has guided to a clearer understanding of the etiological and molecular aspects of HNSCC (Rizzo *et al.*, 2014). However, NGS data (including copy number analysis, exome sequencing, or transcriptome sequencing) have not previously been reported for oral verrucous tumours. Nonetheless, distinguishing OVC from OVH lesions is often difficult. In addition, distinguishing OVC from classical OSCC is a common problem for pathologists due to the poorly defined diagnostic criteria. Also, the aetiology of OVC is not well known, and the suggested role of human papillomavirus HPV as a causative factor remains contentious. The rarity of these lesions makes them difficult to investigate, so most earlier studies have been made on small numbers of cases.

In view of the above considerations, the specific aims of the present thesis studies were designed as follow:

1. Describe the clinical and histopathological features of OVH and OVCs for all the samples collected and used in this PhD project.

2. To use next generation sequencing copy number analysis at low coverage to identify copy number variations as a reflection of the molecular changes in oral verrucous lesions and to identify OVC and OVH genomic characteristic features.

3. To determine if next generation sequencing copy number analysis could distinguish between the genomic damage pattern in OVH, OVC, and OSCC lesions.

4. To analyse a subset of oral verrucous lesions (including VC, and VH cases) for the presence of HPV subtypes and all characterized human viral genomes.

5. To investigate transcriptional changes in OVC and compare them with changes in OSCC using next generation RNA sequencing approach.

6. To use next generation whole-exome sequencing to further investigate the contribution of somatic genomic alteration in the pathogenesis of OVC and gain a comprehensive view of the genetic mutations underlying these lesions and compare them with the genomic mutations underlying OSCCs.

Chapter 2 Materials and Methods

Methods used throughout this project are detailed here. Room temperature (RT) is taken to be 21-24˚C. Invitrogen, Sigma-Aldrich, Gibco or BDH supplied all materials, unless elsewhere mentioned. Supplier addresses and E-mail addresses are presented in Appendix 2.1. Figure 2.1 below demonstrates the study design of this project. All OSCC data used in this project belong to the Pre-cancer Genomics Group.

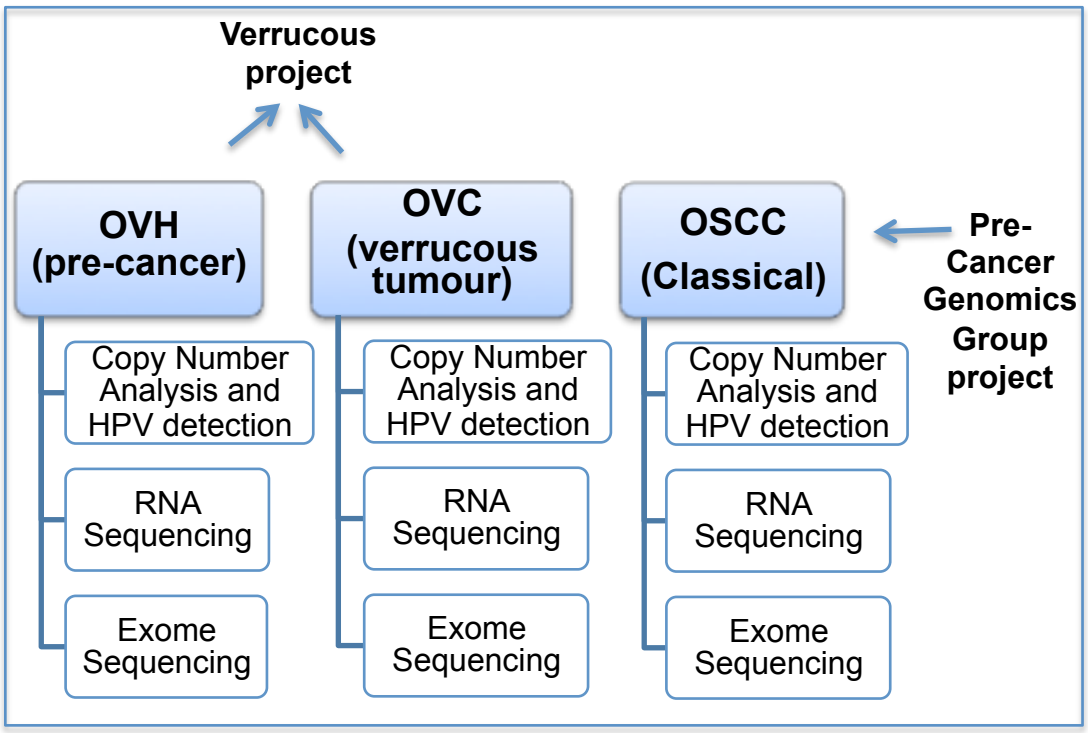


Figure 2.1 Overview of the study design.

2.1 Sample selection

All pathological materials used for this PhD project (including copy number analysis, RNA sequencing and Exome sequencing) from each case were available in the form of Archival Formalin-Fixed Paraffin-Embedded (FFPE) tumour blocks. Samples from Turin, Italy, were taken as sections on glass slides (10μm sections onto 10 plain glass slides from each block). Fifteen OVC, and 13 OVH FFPE blocks were retrieved from the Pathology Department, Bexley Wing, St James's University Hospital. Fifteen OVC samples, and one OVH FFPE block were provided from the Pathology Division, Queen Victoria Hospital, West Sussex, UK. Forty OVC samples and one OVH FFPE block were provided from the Pathology Division, University of Torino, Italy. Seven OVC FFPE blocks were provided from the Department of Pathology, National Guard Hospital, Saudi Arabia. Written informed consent and approval was obtained for all patients for the use of their tissue in this research. (Local ethics committee REC reference: 07/Q1206/30) refer to appendix 2.2 for ethical approval document. In total, 92 oral verrucous samples, malignant (OVC) and pre-malignant (OVH) regions were identified and the original diagnoses were confirmed by Dr. Alec High (reference pathologist). The clinicopathological characteristics of OVH and OVC samples are discussed in chapter three.

2.2 Sectioning FFPE samples

FFPE blocks were placed onto ice for 1-2 hours prior to sectioning. The water bath was set at 45^{o}C. A manual rotary microtome was used to cut one section at 5μm thickness and these slides were then placed on the hot plate for 2 hours for haematoxylin and eosin (H+E) staining, so OVC and OVH areas to be dissected are microscopically identified and marked. Slides were labelled with the study ID, section number and thickness. Then, seven 10μm sections were cut from each block onto plain glass slides for DNA extraction step used for copy number analysis. Ten 10μm sections were cut from each block onto plain glass slides for DNA and RNA dual extraction step used for exome and RNA sequencing. Sections on slides were air dried in a rack overnight and stored later at 4^{o}C until macro-dissection (maximum 7days). Slides were placed on a hot plate for 3min before the macro-dissection steps.

2.3 Haematoxylin and eosin (H+E) staining

In a fume hood, the slides were dipped for de-waxing into four separate xylene (Sigma-Aldrich, USA) jars for 3min in each. Next, the slides were rehydrated in graded ethanols (Sigma-Aldrich, USA): four separate 100% ethanol jars for 3min in each, 70% ethanol jar for 3min, 50% ethanol jar for 3min and 25% ethanol jar for 3min. The slides were then rinsed under running tap water for 2min. Afterward; the slides were immersed in Mayer's haematoxylin stain (Sigma) for 2.5min. The slides were then rinsed under running tap water for 1min. The slides were immersed in Scott's tap water substitute (Sigma) for 2min and were then washed with running tap water for 1min. Next, the slides were immersed in eosin (BDH) for 2min then washed under running tap water for 1min. Afterward, The slides were dipped for dehydration in four separate 100% ethanol jars for 15sec, 1min and 5min in each of the last two jars. The slides were then dipped into three separate xylene jars for 3min in each. Finally, slides were covered by coverslips using DPX mounting medium (Solmedia, UK).

2.4 FFPE tissue Macro-dissection

This step was performed using the marked H+E slide as a guide to obtain at least 70% tumour cell content for DNA or RNA extraction steps. The sections were initially de-paraffinised by xylene and graded ethanol washes: five glass solvent Coplin jars were placed in the fume hood and filled serially with xylene, jar two with absolute ethanol, jar three with 90% ethanol solution, jar four with 70% ethanol solution and jar five with Millipore diH_2O. All sections from a particular block were placed into a wire rack and submerged for de-waxing into the first xylene jar for 5min. Next, the slide rack was lifted from xylene and allowed to drain well then dipped in 100% ethanol jar for 3min. The rack was lifted again and immersed in 90% ethanol jar for 3min, and then dipped into 70% ethanol jar for 3min as well. For DNA/RNA dual extraction, the slides rack was kept into the 70% ethanol jar until I start the macro-dissection, then I takeout slide by slide. For DNA extraction, the slides rack was placed into Millipore diH_2O jar for few minutes and kept in water until macro-dissection, slide by slide. With the marked H+E slide as a guide, the desired tissue was

dissected off using a size 11 disposable sterile scalpel blade, and the tissue was collected and placed into a closed 1.5ml centrifuge tube labelled for that sample. Slide number seven or ten from each block was saved after macro-dissection for H+E post sampling staining. DNA or RNA extraction steps proceeded immediately according to the protocol, suitable for the macro-dissected sample area.

2.5 DNA extraction

This step was made to isolate and extract DNA from de-waxed and macro-dissected FFPE tissue using Qiagen DNA extraction kits (Qiagen, Sussex, UK) as instructed by the manufacturer's. According to the size of the area being sampled for extraction, one of two Qiagen DNA extraction kits were used:

2.5.1 Tissue area sampled per slide = <5 mm^2

The Qiagen QIAamp DNA micro kit reagents and columns (Qiagen, Sussex, UK) were used for these samples. After transferring scraped FFPE tissue samples in a 1.5 ml centrifuge tube, 30µl of ATL buffer, and 10µl of Proteinase K were immediately added to the tubes and mixed by pulse-vortexting for 15sec then incubated at 56°C in a water bath for 72hr until the samples were completely lysed with occasional agitation (with new addition of 10µl Proteinase K every 24 hours if tissue fragments remained). Following this, the tubes were incubated for 1hr at 90°C on a heating block for DNA fragmentation, and to ensure efficient lysis, 50µl of AL buffer and 10µl of ATL buffer was added and thoroughly mixed by pulse-vortexting for 15sec. 50µl of 100% ethanol was added to the tubes and mixed thoroughly by pulse-vortexing for 15sec and Incubated for 5min at room temperature (15–25°C). The entire lysate was transferred to a labelled QIAamp MinElute Column and centrifuged at 8000rpm for 1min. The column was then placed into a clean new 2ml collection tube, and the flow-through tube was discarded. 500µl of AW1 buffer was added to the tubes and centrifuged at 8000 rpm for 1min. Again, the column was then placed into clean new 2ml collection tube, and the flow-through tube was discarded. 500µl of AW2 buffer was added to the tubes and centrifuged at 14,000 rpm for 3min. The column was placed next into clean 2ml collection tube for a dry spin

at 14,000 rpm for 1min. The columns were then placed into new clean 1.5ml labelled microcentrifuge tubes, and the flow through collection tubes were discarded. Next, 30 µl of AE elution buffer was added to the centre of the column membrane to ensure that the entire bound DNA is eluted. After 5min of incubation at room temperature (15–25°C), tubes were centrifuged at 14,000 rpm for 1min. DNA was stored at 4°C overnight or at -20°C for long-term storage.

2.5.2 Tissue area sampled per slide = 5-10 mm2 and Tissue area sampled per slide = >10 mm2

The Qiagen QIAamp DNA mini kit reagents and columns (Qiagen, Sussex, UK) were used for these samples. The protocols for the two tissue areas were very similar except that with the 5-10 mm^2 samples, QIAamp MinElute Columns are used and the DNA is eluted in 50µl AE elution buffer, while the QIAamp Mini spin columns were used with the = >10 mm^2 samples and the DNA was eluted in 100µl AE elution buffer. For these dissected tissue samples 180µl Buffer ATL and 20µl Proteinase K were immediately added to the tubes and mixed by pulse-vortexting for 15sec then incubated at 56°C water bath for 72 hours until the samples were completely lysed with occasional agitation (with new addition of 20µl Proteinase K every 24 hours if tissue fragments remained). 200µl of 100% ethanol was added to the tubes and mixed thoroughly by pulse-vortexing for 15sec and Incubated for 5min at room temperature (15–25°C). The entire lysate was transferred to labelled QIAamp MinElute Column or QIAamp Mini spin columns according to the sampled tissue area size (as discussed above in this section) and centrifuged at 8000 rpm for 1min. The column was then placed into a clean new 2ml collection tube, and the flow-through tube was discarded. 500µl of AW1 buffer was added to the tubes and centrifuged at 8000 rpm for 1min. Again, the column was then placed into clean new 2ml collection tube, and the flow-through tube was discarded. 500µl of AW2 buffer was added to the tubes and centrifuged at 14,000 rpm for 3min. The column was placed next into clean 2ml collection tube for a dry spin at 14,000 rpm for 1min. The columns were then placed into new clean 1.5ml labelled microcentrifuge tubes, and the flow through collection tubes were discarded. Next, 50-100 µl of AE elution

buffer was added to the centre of the column membrane (according to the sampled tissue area size as discussed above in this section) to ensure that the entire bound DNA is eluted. After 5min of incubation at room temperature (15–25°C), tubes were centrifuged at 14,000 rpm for 1min. DNA was stored at 4°C overnight or at -20°C for long-term storage.

2.6 Dual DNA/RNA extraction

The Qiagen AllPrep DNA/RNA FFPE kit was used for co-extraction of genomic DNA and total RNA from FFPE tissue sections. The sections were initially de-paraffinised and macro-dissected as described in section 2.4. The tissue was collected and placed into a labelled 1.5ml centrifuge tube and incubated for ten min to dry the pellet from the residual ethanol at 37°C. Next, The pellet was re-suspended and subjected to lyses and digestion by adding 150µl of PKD buffer and 10µl of proteinase K with tapping the tube to loosen the pellet and the suspension was mixed then by vortexing and incubated at 56°C for 40min. For efficient precipitation, the sample was incubated for complete cooling on ice for 3min. Samples were centrifuged at 14,000 rpm for 15min to separate the RNA-containing supernatant from DNA-containing pellet. After this, the supernatant was carefully transferred to a new 2ml centrifuge tube without disturbing the pellet for RNA extraction. At this point, both nucleic acids are separated and the RNA purification protocol is followed, while the DNA pellets were stored at -20°C for extraction at a later date.

2.6.1 Total RNA extraction

The RNA purification protocol involved incubating the RNA supernatant for 15min at 80°C to start the removal of formaldehyde crosslinks, and followed by addition of 320µl RLT buffer to adjust the binding conditions then mixed by vortexing. Next, 1120µl of 100% ethanol was added and the tube was mixed afterwards by vortexing. The entire sample was transferred to an RNeasy MinElute spin column placed in a 2ml collection tube and centrifuged at ≥10.000 rpm for 15sec. The flow-through was discarded after the centrifugation and 350µl of FRN buffer was added to the spin column and centrifuged again at ≥10.000 rpm for 15sec. The flow-through was discarded after the centrifugation

and the sample was treated with 10µl DNase I stock solution that was added to 70µl of RDD buffer and mixed gently by inverting the tube, and the DNase I digestion mixture (80µl) was added directly to the RNeasy MinElute spin column membrane and incubated for 15min at room temperature. Total RNA washing started then by adding 500µl of FRN buffer to the spin column that was centrifuged at ≥10.000 rpm for 15sec. The flow-through was not discarded this time but was re-applied instead to the spin column that was placed in a new 2ml collection tube and centrifuged at ≥10.000 rpm for 15sec. The flow-through was discarded after the centrifugation and 500µl of RPE buffer was added to the spin column and centrifuged at ≥10.000 rpm for 15sec. Again, after the centrifugation, the flow-through was discarded and 500µl of RPE buffer was added to the spin column and centrifuged at ≥10.000 rpm for 15sec. Next, the collection tube was discarded along with the flow-through and the spin column was placed into a new 2ml collection tube for dry centrifugation at 14,000 rpm for 5min to ensure that no residual ethanol is carried out through RNA elution step. Finally, the RNeasy MinElute spin column was placed into a new 1.5ml collection tube and 25µl of RNase-free water was added directly to the spin column membrane to elute the RNA and was incubated for 1min at room temperature then centrifuged at 14,000 rpm for 1min. The purified total RNA was then stored at -80°C.

2.6.2 DNA extraction

The DNA purification protocol involved lysing the DNA pellet through re-suspension in 40µl of proteinase K and 180µl of ATL buffer that was followed by vortexing and incubation for one hour at 56°C then incubation for two hours at 90°C to start the removal of formaldehyde crosslinks. Next, 200µl of AL buffer and 200µl of 100% ethanol were added to the sample and mixed thoroughly by vortexing. After that, the entire sample was transferred to a QIAamp MinElute spin column placed in a 2ml collection tube and centrifuged at ≥10.000 rpm for 1min. The collection tube was discarded along with the flow-through and the spin column was placed into a new 2ml collection tube and 700µl of AW1 washing buffer was added to the spin column that was centrifuged after that at ≥10.000 rpm for 15sec. Again, after the centrifugation, the flow-through was

discarded and 700µl of AW2 washing buffer was added to the spin column that was then centrifuged at ≥10.000 rpm for 15sec. The flow-through was discarded after the centrifugation and 700µl of 100% ethanol was added to the spin column that was then centrifuged at ≥10.000 rpm for 15sec. Next, the collection tube was discarded along with the flow-through and the spin column was placed into a new 2ml collection tube for dry centrifugation at 14,000 rpm for 5min to ensure that no residual ethanol is carried out through DNA elution step. Finally, the QIAamp MinElute spin column was placed into a new 1.5ml collection tube and 45µl of ATE elution buffer was added directly to the spin column membrane to elute the DNA and was incubated for 5min at room temperature then centrifuged at 14,000 rpm for 1min. The purified total DNA was then stored at -20°C.

2.7 Nucleic acids quantification

After DNA and RNA extractions, both nucleic acids were then quantified using two methods:

2.7.1 Spectrophotometry

The Nanodrop ultra-violet (UV) spectrophotometer: Nanodrop-8000 (Thermo Scientific, UK) was used to determine DNA and RNA concentration and purity within the sample. Reading of DNA concentrations were taken against the elution buffer in which the DNA was dissolved. Reading of RNA concentrations were taken against RNase-free water in which the RNA was dissolved. Beside the concentration, $A_{260:\,230}$ and $A_{260:\,280}$ ratios were obtained which are used as a useful indicators of the purity of the samples and gives an indication of any chemical or protein contaminations using the UV radiation absorption. A copy was stored for each absorbance against wavelength plot and the values for concentration, A_{260}, A_{280}, and $A_{260:280}$ and $A_{260:230}$ ratios were recoded.

2.7.2 Fluorometry

2.7.2.1 Measuring DNA concentration

DNA concentration was specifically quantified using the Quant-iT PicoGreen dsDNA BR assay kit (Invitrogen, UK). Quant-iT Working solution was made by diluting the Quant-iT reagent 1:200 in Quant-iT buffer. Assay tubes were prepared according to table 2.1 below. After incubating the tubes at room temperature for 2min, concentration readings were taken in Qubit Fluorometer following the on-screen directions, and starting with the standards. DNA concentrations were then recorded to be used next for sequencing library preparation steps.

Table 2.1 DNA quantification using Quant-iT dsDNA BR kit

	Standard (1 and 2) Assay Tubes	**User Sample Assay Tubes**
Volume of working solution to add	190µl	199µl
Volume of Standard (from kit) to add	10µl	__
Volume of User Sample to add	__	1µl
Total Volume in each Assay Tube	200µl	200µl

2.7.2.2 Measuring RNA concentration

RNA concentration was specifically quantified using the Quant-iT RiboGreen RNA Assay Kit (Invitrogen, UK). Quant-iT Working solution was made by diluting the Quant-iT RiboGreen reagent 1:200 in Quant-iT buffer. Assay tubes were prepared according to table 2.2 below. After incubating the tubes at room temperature for 2min, concentration readings were taken in Qubit Fluorometer following the on-screen directions, and starting with the standard. RNA concentrations were then recorded to be used next for sequencing library preparation steps.

Table 2.2 RNA quantification using Quant-iT RiboGreen RNA Assay Kit

	Standard (1 and 2) Assay Tubes	**User Sample Assay Tubes**
Volume of working solution to add	190µl	199µl
Volume of Standard (from kit) to add	10µl	__
Volume of User Sample to add	__	1µl
Total Volume in each Assay Tube	200µl	200µl

2.8 Copy number analysis library preparation and sequencing

DNA libraries were prepared following two protocols, a protocol for the Illumina Genome analyser GAIIx sequencer (before the upgrade to HiSeq 2500), and a protocol after the upgrade for Illumina HiSeq 2500.

2.8.1 Library preparation for Illumina Genome analyser sequencing

By following standard Illumina protocols, DNA samples were used to make DNA libraries for sequencing.

2.8.1.1 Shearing

Between 0.5µg to 1µg of DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA).

2.8.1.1.1 Sample preparation and shearing

To prepare the added amounts of each DNA sample, the following calculations were made:

Concentration of Starting Material: (Picco Green result, ng/µl)

Volume of DNA Used (µl): (1000 / Concentration of Starting Material).

Volume of TE Buffer used (µl): 250 - Volume of DNA Used (µl)

Total Amount of DNA Used (ug): (Concentration of Starting Material x Volume of DNA Used (µl)) / 1000

1x TE buffer (made up from 100x TE buffer and nuclease-free water) was added to each DNA shearing sample tube according to the above calculations. DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA). The Covaris S2 Focused-ultrasonicator was used for each sample to shear the DNA at 19^{o}C in batches of 25 cycles using settings shown in Table 2.3. After shearing, samples were processed through a MinElute column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.3 Covaris S2 batch settings

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

2.8.1.1.2 Clean-up with a MinElute column

Five volumes (more than the total samples volumes) of PB binding buffer (Qiagen, Sussex, UK) and DNA samples from the previous step were added to new MinElute columns (Qiagen, Sussex, UK) and centrifuged at 13000rpm for 1min. The flow-through was discarded and 750 µl of PE washing buffer (Qiagen, Sussex, UK) was added to each column and centrifuged at 13000rpm for 1min. The flow-through was discarded again and the columns were centrifuged at 13000rpm for 1min (dry spin). Finally, columns were placed in clean labelled 1.5 ml centrifuge tubes, and 10 µl of EB elution buffer was added to the centre of each column membrane then centrifuged at 13000 rpm for 1min.

2.8.1.2 Agilent Bioanalyser checkpoint

DNA sheared samples were checked for appropriate size distribution according to the manufacturer's instructions on an Agilent Bioanalyser DNA 1000 LabChip (Agilent technologies, Inc., USA). To prepare the Chip, Gel-dye mix was made and added to the corresponding wells in the Chip. 1µl of the DNA ladder was added in the well marked with the ladder symbol. 5µl of DNA marker was added

into the DNA ladder well, as well as all samples wells. Finally, 1µl of each DNA sheared sample was added to the wells and the Chip was inserted into the Agilent Bioanalayser to run the programme.

2.8.1.3 End-repair

The End repair step was performed using End-It DNA End Repair Kit (Epicentre Biotechnologies, USA) according to the manufacturer's instructions. The End-It DNA End-Repair Kit was used to convert DNA with damaged ends to blunt-ended, 5´-phosphorylated DNA. Refer to Table 2.4 below for reaction components. In a PCR plate, 41µl of the reaction master mix was added to each sample in the plate and incubated at room temperature for 45min in the PCR machine. Next, samples were processed through a QiaQuick column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.4 End-repair reaction components.

Component	Volume x1
DNA	9µl
End-It Buffer	5µl
dNTP mix	5µl
ATP	5µl
End-It Enzyme mix	1µl
dH20 (nuclease free water)	25µl

2.8.1.3.1 Clean up with a QiaQuick column

Five volumes (more than the total samples volumes) of PB buffer (Qiagen, Sussex, UK) and DNA samples from the previous step were added to new QiaQuick columns and centrifuged at 13000 rpm for 1min. The flow-through was discarded and 750 µl of PE washing buffer (Qiagen, Sussex, UK) was added to each column and centrifuged at 13000 rpm for 1min. The flow-through was discarded again and the columns were centrifuged at 13000 rpm for 1min (dry spin). Finally, columns were placed in clean labelled 1.5ml centrifuge tubes,

and 34.5µl of EB elution buffer (Qiagen, Sussex, UK) was added to the centre of each column membrane then centrifuged at 13000 rpm for 1min.

2.8.1.4 A-Addition

Klenow DNA polymerase was used to add an A base to each blunt-ended DNA fragment so that adapters could be ligated. A-addition step was performed by adding Klenow DNA polymerase using Kit. Refer to Table 2.5 below for reaction components. In a PCR plate, 15.5µl of the reaction master mix was added to each sample in the plate and incubated at 37°C for 30min in the PCR machine. Next, samples were processed through a MinElute column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.5 A-addition reaction components

Component	Volume x1
DNA	34.5 from last step µl
DNA Pol (Klenow) buffer	5µl
1mM dATP	10µl
Klenow (3'>5' exo minus)(9u/UL, Promega)	0.6µl

2.8.1.4.1 Clean-up with a MinElute column

Protocol is described above in section 2.8.1.3.1, final elution in 10µl.

2.8.1.5 Ligation

Six bp of unique oligonucleotide tag sequence (adapter) was ligated to the ends of the DNA fragments. Tags were chosen in this way to avoid over-representation of any base at all positions that could interfere with cluster identification. Refer to Table 2.6 below for reaction components. In a PCR plate, 19.5 µl of the reaction master mix and 0.5 µl of each individual adaptor were added DNA samples in the plate and incubated at room temperature for 15min in the PCR machine, then incubated at 65°C for 20min to inactivate the enzyme. Next, samples were processed through a QiaQuick column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.6 Ligation reaction components

Component	Volume x1
DNA	10µl
Liga-fast reaction buffer	15µl
dH2O	2µl
Adaptor (tag): different tag With each sample	0.5µl
T4 DNA Ligase	3µl

2.8.1.5.1 Clean up with a QiaQuick column

Protocol is described above in section 2.8.1.3.1, final elution in 30µl.

2.8.1.6 Size selection

DNA fragments were size selected to 200bp using magnetic beads size selection method that separates the DNA from the rest of the sample. In this step, 27µl of AMPure beads (Agencourt Bioscience, Beverly, MA), were added to 30µl ligation reaction (DNA sample from the last step), mixed thoroughly and incubated at room temperature for 5min. The reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads. Once the separation occurred, the cleared supernatant was aspirated and discarded and 200µl of 70% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, then magnetic beads were kept to dry at room temperature for 10-20min. Finally, 40µl of elution buffer (EB from the Qiagen kit) was added and mixed with sample to elute the DNA from the magnetic beads, and the eluted samples were transferred next into new tubes.

2.8.1.7 Enrichment

DNA samples were enriched using a 12- enrichment PCR cycle. Refer to Table 2.7 below for reaction components. In a PCR plate, 35µl of the reaction master

mix was added to15µl of each DNA sample (from previous step). Refer to Table 2.8 below for PCR reaction conditions. Next, samples were processed through a QiaQuick column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.7 Enrichment reaction components.

Component	Volume x1
DNA	15µl
Phusion HF Master Mix	25µl
PE-PTO-F primer (25uM) 1 in 4 dilution	1µl
PE-PTO-R primer (25uM) 1 in 4 dilution	1µl
dH20	13µl

Table 2.8 Enrichment PCR program

Time	Temperature
30sec	98°C
12 cycles of:	
10sec	98°C
30sec	65°C
30sec	72°C
5min	72°C
Hold @ 4°C	

2.8.1.7.1 Clean up with a QiaQuick column

Protocol is described above in section 2.8.1.3.1, final elution in 30µl.

2.8.1.8 Library quality control

Libraries were then examined using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit and Agilent Bioanalyser DNA 1000 LabChip (Agilent technologies, Inc., USA) to assess DNA concentration and quality, respectively. Protocols are described above sections 2.7.2.1 and 2.8.1.2.

2.8.1.9 Sample pooling and sequencing

Details of the prepared libraries were uploaded to the group server for storage. Equimolar amounts of each DNA library were pooled for cluster amplification and multiplexed up to 20 samples per lane for 76bp Illumina single end sequencing, where each read includes 6bp of tagged adapter and 70bp of genomic DNA sequence. The pooled library samples submitted to the sequencing team for running on the sequencer and were run on the Illumina Genome analyser GAIIx sequencer. This offers 20-30 million reads per lane. The Illumina Genome analyser GAIIx sequencer produces a fast Q file. This contains information about each read (including quality of read, location and size) besides the raw nucleotide sequence.

2.8.2 Library preparation for Illumina HiSeq 2500

For Illumina HiSeq 2500 sequencing, DNA libraries were prepared using the NEBNext DNA Library Prep Master Mix Set and NEBNext Singleplex Oligos as described for Illumina with some modifications (New England BioLabs inc., UK).

2.8.2.1 Shearing

200ng of DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA)

2.8.2.1.1 Sample preparation and shearing

To prepare the added amounts of each DNA sample, the following calculations were made:

Concentration of Starting Material: (Picco Green result, ng/µl)

Volume of DNA Used (µl): (200 / Concentration of Starting Material).

Volume of TE Buffer used (µl): 250 - Volume of DNA Used (µl)

Total Amount of DNA Used (ug): (Concentration of Starting Material x Volume of DNA Used (µl)) / 1000

1x TE buffer (made up from 100x TE buffer and nuclease-free water) was added to each DNA shearing sample tube according to the above calculations. DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA). The Covaris S2 Focused-ultrasonicator was used for each sample to shear the DNA at 19^{o}C in batches of 25 cycles using the below settings shown in Table 2.9. After shearing, samples were processed through a MinElute column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.9 Covaris S2 batch settings

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

2.8.2.1.2 Clean-up with a MinElute column

Five volumes (more than the total samples volumes) of PB binding buffer (Qiagen, Sussex, UK) and DNA samples from the previous step were added to new MinElute columns (Qiagen, Sussex, UK) and centrifuged at 13000 rpm for 1min. The flow-through was discarded and 750 µl of PE washing buffer (Qiagen, Sussex, UK) was added to each column and centrifuged at 13000 rpm for 1min. The flow-through was discarded again and the columns were centrifuged at 13000rpm for 1min (dry spin). Finally, columns were placed in clean labelled 1.5ml centrifuge tubes, and 10µl of EB elution buffer (Qiagen, Sussex, UK)was added to the centre of each column membrane then centrifuged at 13000 rpm for 1min.

2.8.2.2 Agilent 2200 Tapestation checkpoint

DNA sheared samples were checked for appropriate size distribution according to the manufacturer's instructions on an Agilent 2200 TapeStation D1K High Sensitivity Screentape (Agilent technologies, Inc., USA). To prepare the TapeStation, Agilent 2200 TapeStation software was switched-on. The tips as well as the High Sensitivity D1K ScreenTape were loaded into the 2200 TapeStation. To prepare the samples, 3µl of High Sensitivity D1K Ladder were

added into the first tube, and 2µl of each DNA sample were mixed with 2 µl of High Sensitivity D1K Sample Buffer by vortex for 5sec. after a Quick spin down of the samples to position them at the bottom of the tubes; sample tubes were loaded into the TapeStation. Finally, samples were selected and named on the controller software to run the programme.

2.8.2.3 End-repair

Following purification, DNA fragments were end repaired to convert DNA containing damaged ends to blunt-ended, 5´-phosphorylated DNA according to the manufacturer instructions. Refer to Table 2.10 below for reaction components. In a PCR plate, 41µl of the reaction master mix was added to each 9µl DNA sample in the plate and incubated at room temperature for 30min in the PCR machine. Next, samples were processed through a QiaQuick column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.10 End-repair reaction components.

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

2.8.2.3.1 Clean up with a QiaQuick column

Five volumes (more than the total samples volumes) of PB binding buffer (Qiagen, Sussex, UK) and DNA samples from the previous step were added to new QiaQuick columns (Qiagen, Sussex, UK) and centrifuged at 13000 rpm for 1min. The flow-through was discarded and 750 µl of PE washing buffer (Qiagen, Sussex, UK) was added to each column and centrifuged at 13000 rpm for 1min. The flow-through was discarded again and the columns were centrifuged at 13000 rpm for 1min (dry spin). Finally, columns were placed in clean labelled 1.5 ml centrifuge tubes, and 21 µl of EB elution buffer (Qiagen,

Sussex, UK) was added to the centre of each column membrane then centrifuged at 13000 rpm for 1min.

2.8.2.4 dA-Tailing of end-repaired DNA

End-repaired DNA fragments were dA-tailed so that adapters could be ligated. Refer to Table 2.11 below for reaction components. In a PCR plate, 4 µl of the reaction master mix was added to each 21 µl DNA sample in the plate and incubated at 37°C for 30min in the PCR machine. Next, samples were processed through a MinElute column according to the Qiagen protocol (Qiagen, UK) for cleaning-up.

Table 2.11 dA-Tailing reaction components.

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

2.8.2.4.1 Clean up with a QiaQuick column

Protocol is described above in section 2.8.2.3.1, final elution in 12.5 µl.

2.8.2.5 Ligation

dA-tailed DNA fragments were ligated to the ends of the DNA with NEBNext adaptor. Refer to Table 2.12 below for reaction components. In a PCR plate, 12.5 µl of the reaction master mix was added to the dA-tailed DNA in the plate and incubated at room temperature for 15min in the PCR machine, then 3µl of USER enzyme mix was added to the sample and incubated at 37°C for 15min. Next, samples were cleaned and smaller fragments of DNA were removed using AMPure Solid-Phase Reversible Immobilisation beads (AMPure SPRI beads) (Agencourt Bioscience, Beverly, MA).

Table 2.12 Ligation reaction components.

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

2.8.2.6 Size-selection

Adaptor ligated DNA fragments were size selected to 200 bp using magnetic beads (AMPure SPRI beads) (Agencourt Bioscience, Beverly, MA). In this step, 40µl (0.8 concentration) SPRI beads were added to 12.5µl-ligation reaction DNA (sample from the last step), mixed thoroughly and incubated at room temperature for five min to bind larger fragments of DNA. The reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the supernatant was transferred to a new tube and 10µl (0.2 concentration) SPRI beads were added to the tube (supernatant DNA targets, approximately 200bp in length) and placed again into Agencourt magnetic rack. The bound DNA was washed twice with 200µl of 80% ethanol and incubated for 30sec at room temperature to remove any proteins or chemicals. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and the magnetic beads were kept to dry at room temperature for 10-20min. Finally, 20µl of elution buffer (EB from the Qiagen kit) was added and mixed with sample to elute the DNA from the magnetic beads, and the eluted samples were transferred after into new tubes. Ten µl of each sample was aliquoted and labelled in a separate tube and stored at -20^{o}C (Pre-PCR).

2.8.2.7 Enrichment

Adaptor ligated DNA fragments were tagged to 96 different indexed primers custom-designed by Pre-Cancer Genomics Group (Integrated DNA technologies, Inc., UK), each containing a unique identifying 6bp tag, and was then subjected to PCR enrichment using 15- enrichment PCR cycles. Refer to

Table 2.13 below for reaction components. In a PCR plate, 13.75µl of the master mix was added to 10µl of each DNA sample and 1.25µl of a separate indexed primer. Refer to Table 2.14 for PCR reaction conditions.

Table 2.13 Enrichment reaction components.

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

Table 2.14 Enrichment PCR program

Time	Temperature
30sec	98°C
15 cycles of:	
10sec	98°C
30sec	65°C
30sec	72°C
5min	72°C
Hold 4°C	

2.8.2.7.1 Clean up with AMPure SPRI beads

Next, The post-PCR samples were processed for cleaning-up using AMPure SPRI beads (Agencourt Bioscience, Beverly, MA). In this step, 2.5x concentration of SPRI beads was added to the enriched DNA targets (DNA sample from the last step), mixed thoroughly and incubated at room temperature for 5min. The reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 200µl of 70% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was

repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10-20min. Finally, 40µl of elution buffer (EB from the Qiagen kit) was added and mixed with sample to elute the DNA from the magnetic beads, and the eluted samples were transferred next into new tubes.

2.8.2.8 Library quality control

Libraries were examined using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit and Agilent 2200 TapeStation D1K High Sensitivity ScreenTape to assess DNA concentration and quality, respectively. Protocols are described in sections 2.7.2.1 and 2.8.2.2.

The TapeStation also reveals any excess adaptor oligonucleotide that has not been removed from the sample. If there was an adaptor contaminating of a total greater than 10% of the final library concentration then the library DNA sample is re-cleaned using 2.5x concentration of AMPure SPRI beads. If there was low DNA amounts in the prepared library to allow this then the Pre-PCR sample is used for further bead-cleaning step using a 1.8x concentration of AMPure SPRI beads before the enrichment step and with 18 cycles of PCR target enrichments. Then, the sample is cleaned using a 2.5x concentration of AMPure SPRI beads as described in section 2.8.2.7.1.

2.8.2.9 Sample pooling and sequencing

Details of the prepared libraries were uploaded to the group server for storage. Equimolar amounts (typically 20ng) of each DNA library were pooled before being sent to the sequencing team for cluster amplification and multiplexed up to 40 samples per lane and paired-end sequenced (2X100bp) on an Illumina HiSeq 2500. This generates 200 million reads per lane. Illumina HiSeq 2500 sequencer produces a fast Q file that provides information about each read (including quality of read, location and size) besides the raw nucleotide sequence.

2.8.3 Alignment and data analysis

Alignment and data analysis steps are the same for both Illumina sequencers (Illumina Genome analyser GAIIx and Illumina HiSeq 2500). The methodology described in this section was performed by Dr Henry Wood, a [Bioinformatics, Pre-cancer Genomics Group]. Briefly:

After passing quality control examination and sample pooling, samples were sent for Illumina sequencing. The bioinformatics group, clinical science building, St James's University Hospital, performed the following steps: Briefly, Reads were split into separate files according to tag and Cutadapt software was used to trim the adaptor sequences. The remaining reads were then aligned to the human reference genome using the Burrows Wheeler Aligner program (BWA) (University of California Santa Cruz version GRCh37/hg19, http://genome.ucsc.edu) and to all known viral genomes, including HPV subtypes, downloaded from the National Center for Biotechnology information:

(http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid%20=%201023 9). Digital karyograms were constructed from these data using the CNAnorm program in order to be analysed for copy number variation.

2.8.3.1 Human genomic copy number analysis

DNA was sequenced at 0.033X - 0.33X coverage**.** Sample reads were arranged and organized by chromosome and position. The ratio of test to control reads was calculated across the genome in equally sized windows, averaging 200 reads in some cases and 400 reads in others. A control sample was pooled from a group of 20 data sets for normal individuals downloaded from the 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp). Digital karyograms were constructed from these data using the CNAnorm program in which CNV can be analysed.

2.8.3.1.1 visual assessment of CN karyogram components

In each CN karyogram, chromosomal position is on the x-axis and tumor:normal ratio is on the y-axis. The black lines are regions of common copy number

between breakpoints. Windows of gain and loss are red and blue respectively. During the visual examination of the CN abnormalities within the karyograms, red windows associated with black lines above the centre line in each genomic position along the genome were considered regions of chromosomal gain. In addition, blue windows associated with black lines below the centre line in each genomic position along the genome were considered regions of chromosomal loss. Technical artefacts regions such as in chromosome 19 and in the centromere and telomere chromosomal regions were excluded from the visual examination of the karyograms.

2.8.3.1.2 Frequency karyograms

The methodology described in this section was performed by Dr Henry Wood. Briefly, frequency karyograms were produced for OVH, OVC, and OSCC cohorts using a program that takes all BED files from the copy number analysed samples lists. The selected CN threshold was of 0.05, above or below was considered a gain or loss. (Primary OSCC data belong to the Pre-cancer Genomics Group). Known copy number artefacts chromosomal regions (i.e. chromosome 19, telomeres and centromeres) were excluded from the analysis (these regions were considered technical artefacts as they appear in all and different tumours karyograms.

2.8.3.1.3 Logistic regression technique

The ability to distinguish OVC and OSCC samples using their copy number profiles was also tested using a novel logistic regression computational technique (Gusnanto, Wood *et al.* accepted pending corrections, Bioinformatics). Each sample was removed from the total data set in turn. The remaining samples were then used to build a predictive model. Each genomic window was given a score based on its ability to distinguish the two groups. The model was then applied to the test sample and a subtype prediction was made.

2.8.3.1.4 Hierarchical clustering

Chromosomal region alterations in OVC, OVH, and OSCC were analysed using NGS copy number analysis. A complete linkage unsupervised hierarchical clustering was performed using R package HCLUST software for all the three groups based on DNA CN changes.

2.8.3.1.5 Genomic Identification of Significant Targets in Cancer (GISTIC2.0) (computational approach)

The GISTIC algorithm identifies likely somatic driver CN alterations through evaluating the amplitude and frequency of amplified or deleted observed events (Mermel *et al.*, 2011). It identifies copy number regions with a statistically high frequency of aberrations over the background aberrations. This computational tool evaluates both significance and frequency to detect regions of interest. Within each identified statistically significant CN region, a peak region is determined, and accordingly, the region with a minimal q-value and maximal G-score is most likely to include affected genes. Additionally, regions that pass both the Q-bound and G-Score threshold cut-offs will be only shown. Here, segmentation files (-seg) for OVH, OVC and OSCC cohorts were uploaded at the GISTIC analysis tool. The segmentation file includes samples segmented data identified by the segmentation algorithm from the CN analysis pipeline designed by the bioinformatics team in Pre-cancer Genomics Group.

The copy number profiles of OVH, OVC and OSCC were characterised by several approaches using the GISTIC2.0 algorithm including: amplification and deletion plots of copy number alterations, the identification of amplification and deletion genes within copy number altered regions, and segmented copy number heat maps. All the parameters used in the GISTIC analysis are in appendix 2.3. (Primary OSCC data belong to the Pre-cancer Genomics Group).

2.8.3.1.6 Assessment of the list of genes with copy number alterations in OVH, OVC and OSCC cohorts

Genes contained in the regions of highest significant gain and loss regions identified by GISTIC method (section 2.8.3.1.5) were further analysed by

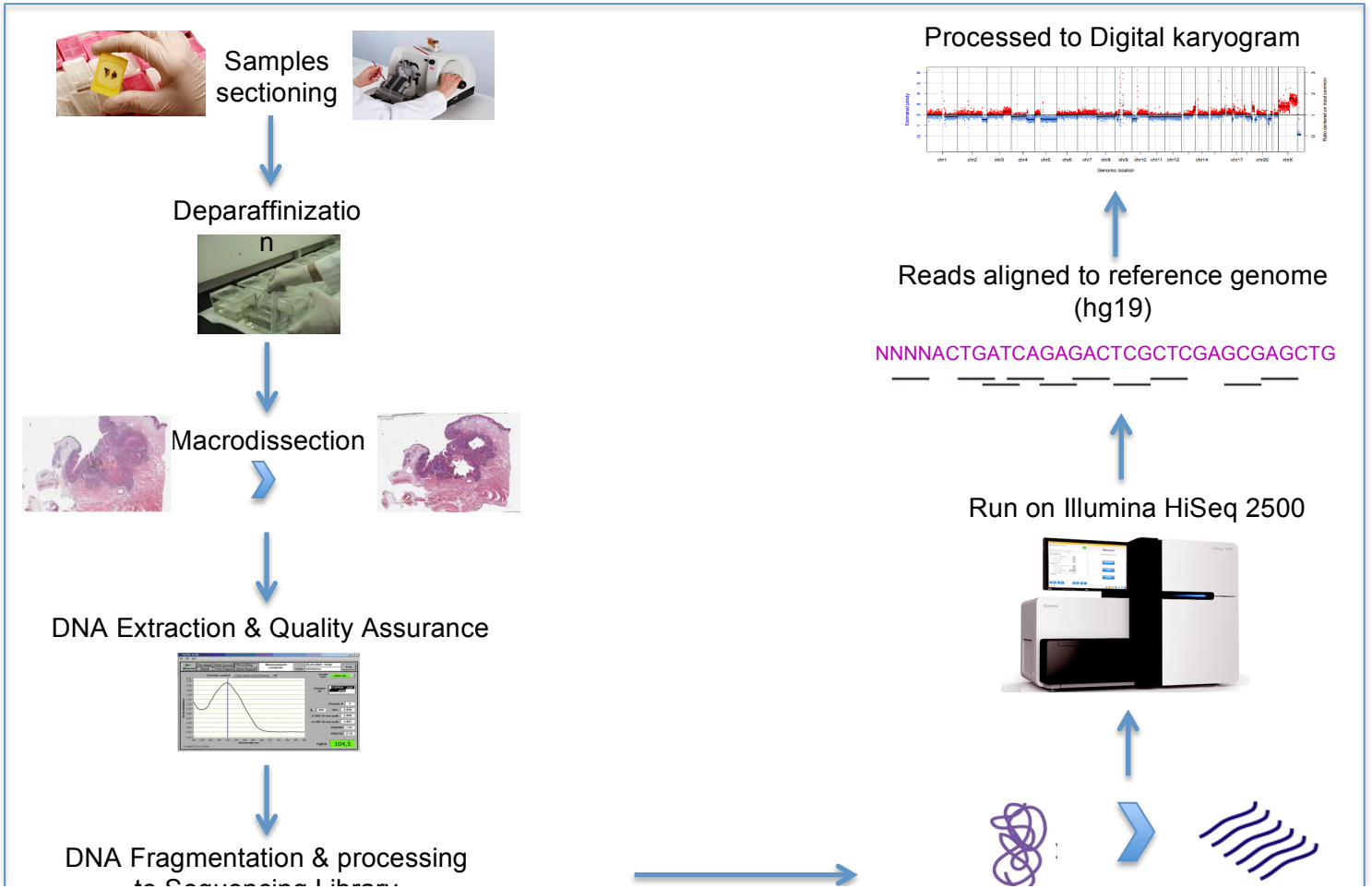
running them against 13 KEGG pathways (Kyoto Encyclopaedia of Gene and Genomes), which are more related to head and neck cancers (a large database project for metabolic pathways), as well as cancer gene census and Stransky mutation list (76 previously identified genes in HNSCCs harbouring high statistically significant mutations) (Stransky *et al.*, 2011). The 13 KEGG pathways include: P13K, WNT signalling pathway, cell cycle, calcium signalling pathway, VEGF signalling pathway, MAPK pathway, DNA replication pathway, Phosphatidylinositol signalling system, P53 signalling pathway, NOTCH signalling pathway, JAK STAT signalling pathway, ERBB signalling and hedgehog.

2.8.3.2 Viral genomes and HPV subtypes Detection and Load Measurements by Sequencing

The methodology described in this section was performed by Dr Henry Wood. Briefly, viral load was measured as described in (Conway *et al.*, 2012). The number of reads aligning to the human genome was used to calculate read depth in terms of reads per kilobase. Next, the number of reads uniquely aligning to viral genomes was counted. This was equated to a certain number of Kb viral sequence per human genome, and hence the number of viral genomes per human genome. Given a certain number of human reads, the possible viral load that could be detected with 95% confidence is:

(0-log(1-0.95) x 6Mb diploid human genome)/(7.9Kb viral genome x number of human reads)

HPV sequencing data from a previous study published by Pre-Cancer Genomics group were used to provide positive and negative controls. Oral verrucous samples were matched with 16 oral and oropharyngeal (OP) cases (Conway *et al.*, 2012).


Samples sectioning
Deparaffinizatio n
Macrodissection
DNA Extraction & Quality Assurance
DNA Fragmentation & processing
Processed to Digital karyogram
Reads aligned to reference genome (hg19)
NNNNACTGATCAGAGACTCGCTCGAGCGAGCTG
Run on Illumina HiSeq 2500

2.9 RNAseq library preparation and sequencing

Strand-directional whole transcriptome sequencing libraries were prepared from total RNA following ScriptSeq complete kit (Human/Mouse/Rat–Low Input) protocol manufacturer's instructions (Epicenter Biotechnologies, USA) for Illumina HiSeq 2500 sequencing.

2.9.1 Ribosomal RNA (rRNA) Depletion

rRNA depletion step was carried out to remove ~99% of cytoplasmic rRNA, which accounts for ~80% of total RNA and can consequently obstruct enough coverage of mRNA, the major focus of RNAseq studies (Morlan *et al.*, 2012). Ribo-Zero Magnetic Gold kit (Human/Mouse/Rat) (Epicenter Biotechnologies, USA) was used to remove rRNA from 80ng to 1.5µg of the total isolated RNA according to the manufacturer's instructions:

2.9.1.1 Preparation of the magnetic beads

The first step in rRNA removal was beads washing. In this step, 90µl of the supplied beads (quantity for each sample) were added to 1.5ml empty tubes and placed on the magnetic rack for 1min. Next, the clear supernatants were discarded and the tubes were removed from the magnetic stand. 90 µl of RNase-free water was added to the tubes contacting the beads and mixed by repeated pipetting. The tubes were placed again on the magnetic rack for 1min and the clear supernatants were discarded. After that, the tubes were removed from the magnetic stand and 32.5µl of the Magbead re-suspension solution was added to the tubes contacting the beads and mixed by repeated pipetting. Then, 35ml was taken from each tube contacting the beads into new 1.5 ml RNase-free tubes (for each sample). Finally, 0.5 µl of RiboGuard (protect RNA from degrading) was added to each tube, mixed briefly by vortexing, and stored at RT for the next part of rRNA depletion.

2.9.1.2 rRNA removal

The reaction components in Table 2.15 were added into a 0.5ml tube in the same order, gently mixed by pipetting and incubated at 68°C for10min then at

room temperature for 5min to treat each RNA sample with rRNA removal solution. Next, treated RNA samples were added to the tubes contacting washed beads and mixed by pipetting at least ten times then vortexed at medium speed for 10sec and placed at room temperature for 5min. After that, the tubes were incubated at 50°C for 5min then immediately placed on the magnetic stand for 1min. Following this, the clear supernatants (50-52µl) were transferred to new 1.5ml tubes and placed on ice for the next step (sample purification).

Table 2.15 rRNA removal reaction component.

Components	Volume
RNase-free water	To make a final volume of 20 µl
Ribo-Zero reaction buffer	2 µl
Total RNA sample	According to RNA amount of each sample (80 ng to 1.5µg)
Ribo-Zero removal solution	4 µl
Total volume	20 µl

2.9.1.3 Purification of rRNA-depleted samples

The RNeasy MinElute Cleanup Kit (Qiagen, Sussex, UK) was used to enable clean-up and concentration of RNA after rRNA-depletion. In this step, rRNA-depleted samples were adjusted to 100µl with adding RNase-free water (50-48µl). Then, 350µl of RLT lysis buffer and 550µl of 100% ethanol were added to each sample and mixed by pipetting to create conditions that encourage selective binding of rRNA-depleted samples to the membrane of the RNeasy MinElute spin column. The samples were then transferred to MinElute columns, centrifuged for 15sec at 8000 rpm, and the flow-through with the collection tubes were discarded. Next, 500µl of RPE washing buffer were added to each MinElute spin column, centrifuged for 15sec at 8000rpm, and the flow-through were discarded. Following this, 500µl of 80% ethanol were added to each MinElute spin column for efficient washing away of any contaminants, centrifuged for 2min at 8000rpm, and the flow-through with the collection tubes

were discarded. MinElute spin column were placed then into a new clean 2ml collection tube for a dry spin at 14,000 rpm for 5min. After this, the columns were placed into new clean 1.5ml labelled microcentrifuge tubes, and the flow through collection tubes were discarded. Next, 12µl of RNase-free water were added to the center of each column membrane to ensure that the entire bound RNA is eluted and the tubes were centrifuged at 14,000 rpm for 1min. Then, rRNA-depleted and purified samples were quantified using High Sensitivity R6K ScreenTape on Agilent 2200 TapeStation system (Agilent technologies, Inc., USA).

2.9.1.4 Agilent 2200 TapeStation for rRNA-depleted samples quantification

rRNA-depleted and purified samples were quantified using High Sensitivity R6K ScreenTape (Agilent technologies, Inc., USA) according to the manufacturer's instructions on an Agilent 2200 TapeStation system (Agilent technologies, Inc., USA). To prepare the samples, 2µl of each RNA sample wre mixed with 1µl of High Sensitivity R6K Sample Buffer and heated at 72°C for 3min then placed on ice for 2min, to allow samples denaturation. samples were then centrifuged briefly to position them at the bottom of the tubes and loaded into the TapeStation. Finally, samples were selected and named on the controller software to run the programme. Note: a software ladder is added in the Agilent 2200 TapeStation system to be used for the analysis.

2.9.2 cDNA synthesis and terminal tagging

The remaining RNA after depletion of total RNA was used as input to prepare barcoded strand-specific cDNA libraries.

2.9.2.1 RNA fragmentation

500 pg to 50 ng of rRNA-depleted RNA is needed per reaction for chemical fragmentation. The reaction components in Table 2.16 were added into a 0.2 ml tube and incubated at 65°C for five min then placed on ice. This step was performed to anneal the primer to RNA samples (hot start). A reaction template control was also included starting from this step.

Table 2.16 RNA fragmentation reaction component.

Components	Volume (samples)	Volume (control)
Nuclease-Free water	To make a final volume of 11µl	9µl
Ribo-Zero treated RNA	500 pg to 50 ng	-
cDNA Synthesis Primer	2µl	2µl
Total volume	11µl	11µl

2.9.2.2 cDNA synthesis

RNA fragmentation and single strand cDNA transcription steps were achieved by adding the reaction components in Table 2.17 into sample tubes from the previous step (primer annealing step). The components were mixed gently and thoroughly, then incubated at 25°C for 5min, 42°C for 20min, paused at 37°C to add 1µl of the Finishing Solution to each tube, and mixed by pipetting. Following this, the tubes were incubated at 37°C for 10min, 95°C for 3min and paused at 25°C. Note: all reaction components were added mixed on ice.

Table 2.17 cDNA synthesis reaction components

	Volumes add for each sample	Volumes add for control
RNA Fragmentation Solution	1.0µl	1.0µl
cDNA synthesis PreMix	3µl	3µl
100 mM DTT	0.5µl	0.5µl
StarScript Reverse Transcriptase	0.5µl	0.5µl
Total volume	5µl	5µl
	Incubate at 25°C for 5min	
	Incubate at 42°C for 20min	
Finishing Solution	Pause at 37°C, add 1.0µl of Finishing Solution to each reaction	
	Incubate at 37°C for 10min	
	Incubate at 95°C for 3min	
	Pause at 25°C	

2.9.2.3 cDNA Terminal-Tagging

During the 95°C incubation (last step), the reaction components in Table 2.18 were prepared on ice for terminal tagging to produce di-tagged, single-stranded cDNA. The components were mixed gently and thoroughly by pipetting, added into sample tubes from the previous step (cDNA synthesis) at the 25°C pausing step, then incubated at 25°C for 15min, 95°C for 3min and kept hold at 4°C.

Table 2.18 cDNA Terminal-Tagging reaction components

	Volumes add for each sample	**Volumes add for control**
Terminal Tagging Premix	7.5µl	7.5µl
DNA Polymerase	0.5µl	0.5µl
Total volume	8µl	8µl
	Incubate at 25°C for 15min	
	Incubate at 95°C for 3min	
	Hold at 4°C	

2.9.2.4 cDNA purification

Purification of cDNA samples was carried out using Agencourt Ampure XP beads (BeckmanCoulter, Agencourt Bioscience, Beverly, MA). In this step, 1.8x concentration (45µl) of SPRI beads was added to each cDNA sample (samples from the last step), mixed thoroughly and incubated at room temperature for 15min. The reaction tubes were then placed in an Agencourt magnetic rack for 5min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 200µl of 80% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10-15min. Finally, 22.5µl of RNase-free water was added and mixed with each sample to elute the cDNA from the

magnetic beads, incubated for 2min, and the eluted samples were transferred next into new tubes.

2.9.3 Library amplification and Indexing

The di-tagged cDNA was amplified by 15-cycle PCR and indexed using ScriptSeq index PCR primer (Epicenter Biotechnologies, USA).

2.9.3.1 PCR

The reaction components in Table 2.19 were prepared on ice, mixed gently and thoroughly by pipetting, and added into sample tubes from the previous step (di-tagged cDNA purification: 22.5µl). In 0.2 ml tubes, 26.5µl of the master mix was added to 22.5 µl of each purified cDNA sample and 1µl of a separate indexed primer. Refer to Table 2.20 for PCR reaction conditions.

Table 2.19 Library amplification reaction components.

	Volumes add for each sample	Volumes add for control
di-tagged cDNA	22.5µl	22.5µl from previous step control
FailSafe PCR Premix E	25µl	25µl
Forward PCR Primer	1µl	1µl
Reverse PCR Primer	1µl	1µl
FailSafe PCR Enzyme	0.5µl	0.5µl
Total volume	50µl	50µl

Table 2.20 PCR program.

Time	Temperature
1 minute	95°C
15 cycles of:	
30sec	95°C
30sec	55°C
3min	68°C
Incubate after the final cycle for:	
7min	68°C

2.9.3.2 Library purification

Before starting the purification step, 1µl of Exonuclease I enzyme (New England BioLabs inc., UK) was added to each reaction tube then incubated at 37°C for 15min to ensure degradation of any excess single-stranded primer oligonucleotide within the reaction mixture. Next, the di-tagged amplified cDNA libraries were purified using Agencourt Ampure XP beads (BeckmanCoulter, Agencourt Bioscience, Beverly, MA). In this step, 1 x concentration (51 µl) of SPRI beads was added to each di-tagged amplified cDNA sample (samples from the last step), mixed thoroughly and incubated at room temperature for 15min. The reaction tubes were then placed in an Agencourt magnetic rack for 5min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 200µl of 80% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10-15min. Finally, 20µl of RNase-free water was added and mixed with each sample to elute cDNA libraries from the magnetic beads, incubated for 2min, and the eluted samples were transferred next into new tubes.

2.9.4 Library quality control

Libraries were examined using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit and Agilent 2200 TapeStation D1K High Sensitivity ScreenTape to assess DNA concentration and quality, respectively. Protocols are described in sections 2.7.2.1 and 2.8.2.2. The TapeStation also can reveal any excess adaptor oligonucleotides that has not been removed from the sample. If there was an adaptor contaminating of a total greater than 10% of the final library concentration then the library sample is subjected to re-cleaning using the Ampure XP system.

2.9.5 Sample pooling and sequencing

Details of the prepared libraries were uploaded to the group server for storage. Equimolar amounts (typically 5ng) of each cDNA library were pooled before being sent to the sequencing team for cluster amplification and multiplexed up to six samples per lane and paired-end sequenced (2X100bp) on an Illumina HiSeq 2500. Illumina HiSeq 2500 sequencer produces a fast Q file that provides information about each read (including quality of read, location and size) besides the raw nucleotide sequence.

2.9.6 Alignment and data analysis

The methodology described in this section was constructed and applied by Dr Lucy Stead, [bioinformatics, Pre-cancer Genomics Group]. Briefly:

Libraries were sequenced to an equivalence of 6 samples per lane of a HiSeq2500. Fastq files were processed using trim_galore to remove low quality bases, trim adaptors and fix paired-end reads. Processed reads were aligned to the human genome GRCh37.p11 by Tophat2.0.7, using the gencode.v17 genome annotation as a guide. Reads could align a maximum of 5 times, with 2 mismatches. Alignment statistics were ascertained using the samtools flagstat command and CollectRnaSeqMetrics programme in the Picard software suite, version 1.56. Expression quantification and differential expression analysis (FDR of 0.01) were performed using cuffdiff, version 2.1.1, with multireads assigned using the –u parameter.

2.9.6.1 Principal Component Analysis

Principal Component Analysis (PCA) was performed for all expressed protein-coding genes in OVC and OSCC samples, using the prcomp function in R.

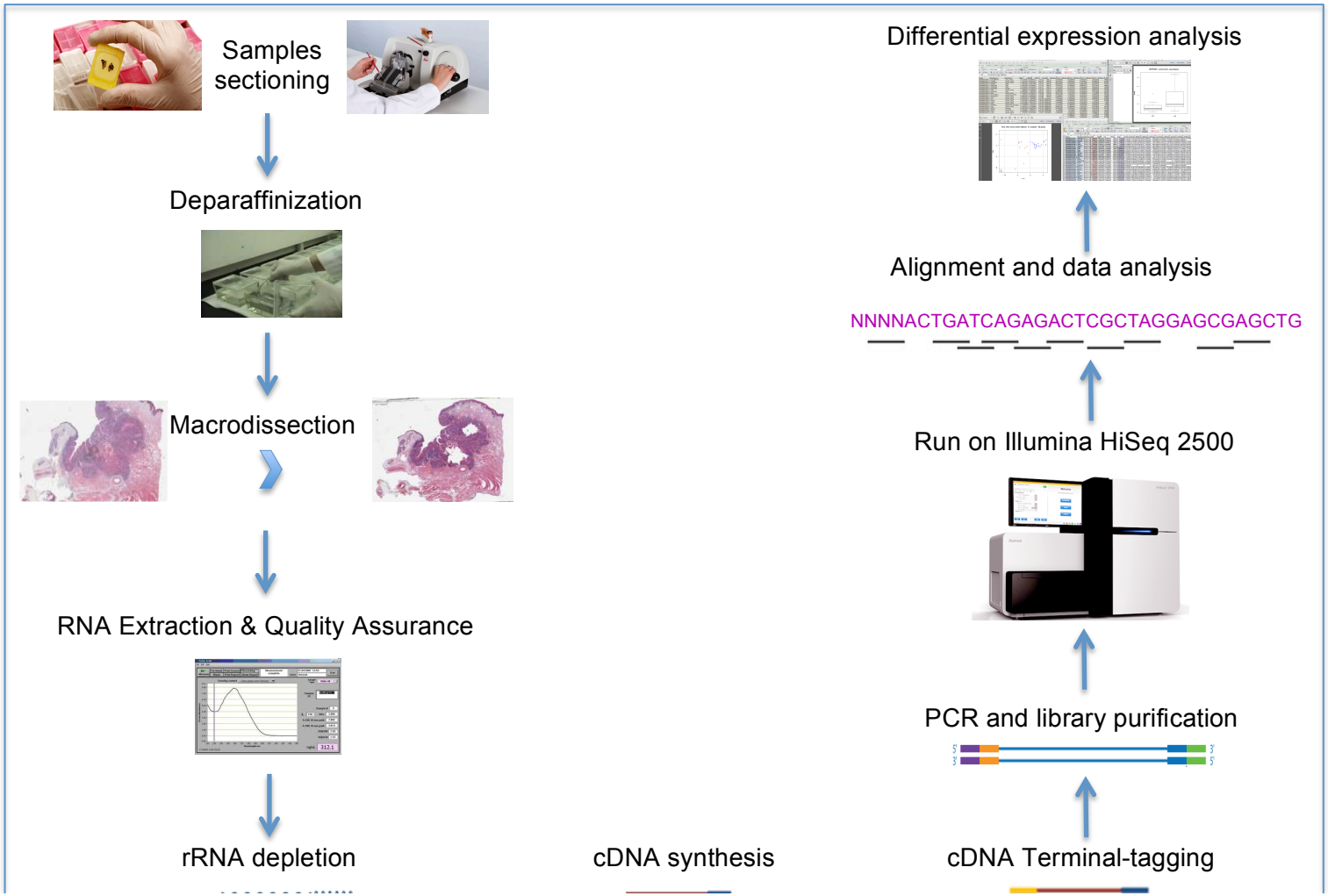
2.9.6.2 Functional Enrichment

Using a significance threshold of 1%, numbers of differentially expressed genes were obtained and detailed in chapter five. DAVID Bioinformatics Database 6.7 (Database for Annotation, Visualisation and Integrated Discovery) web server

was used to assess their functional enrichment (Huang *et al.*, 2008).These individual gene lists were then uploaded using their official gene symbol and the background from which to measure enrichment was the Homo sapiens list of all known genes. The inspection of the functional enrichment with the different biological interpretation is described in chapter five.

2.9.6.3 Integration of CN analysis and RNAseq gene expression data

To highlight novel genes of possible clinical and biological importance in OVC, the list of genes generated from the copy number analysis data (chapter four, Table 4.3) was integrated with genes from the significant differential expression lists for protein-coding genes in the matched normal versus verrucous samples identified by RNAseq. In this step, CN gene data from Table 4.3 were run against RNAseq significant differential expression gene lists in an Excel sheet and the matched genes in both lists were identified and recoded (chapter 6).


Samples sectioning
Deparaffinization
Macrodissection
RNA Extraction & Quality Assurance
rRNA depletion
cDNA synthesis
cDNA Terminal-tagging
PCR and library purification
Run on Illumina HiSeq 2500
Alignment and data analysis
NNNNACTGATCAGAGACTCGCTAGGAGCGAGCTG
Differential expression analysis

2.10 Exome sequencing library preparation and sequ

DNA sequencing libraries were prepared from extracted D NEBNext singleplex library preparation kit (New England BioL then, exome DNA was captured following SureSelect^{XT} targ System protocol (Agilent technologies, Inc., USA) for Illumina paired end sequencing.

2.10.1 DNA quality control PCR

2.10.1.1 PCR reaction

As the DNA used here was extracted from FFPE tissue, performance might be seen when using the same DNA starting a due to varied DNA degradation levels triggered by sub processing or DNA storage. For exome sequencing library prepa to know if the DNA sample is poor before being used, so that it omitted all together, or used in increased amounts. The metho section describe a multiplex PCR that allows DNA quality quantifi to use this information to inform how much DNA will be neec sequencing library preparation.

Here, DNA concentrations were measured using Invitrog Picogreen dsDNA BR assay kit as described in sections 2.7.2.1 PCR Plates, all DNA samples were diluted to 5ng/ul. A positive c quality DNA sample and a negative control of nuclease fre included. Next a PCR master mix was prepared as described below and loaded into each PCR well of the plate, and 2µl of eac was added to each PCR well. The PCR plate was placed into the after a brief vortex and centrifugation. Refer to Table 2.22 for conditions.

Table 2.21 PCR reaction component.

Component	Volume x1
Phusion High Fidelity Master Mix (New England BioLabs inc., UK).	12.5µl
95bp Forward primer (100uM)	0.5µl
95bp Reverse primer (100uM)	0.5µl
235bp Forward primer (10uM)	1.5µl
235bp Reverse primer (10uM)	1.5µl
Nuclease Free Water	6.5µl
Total	23µl

Table 2.22 PCR program.

Time	Temperature
10min	95°C
39 cycles of:	
20sec	95°C
20sec	60°C
20sec	72°C
Incubate after the final cycle for:	
5min	72°C
Hold @ 4°C	

2.10.1.2 Purification

All PCR products were transferred into 1.5 ml tubes. Next, the post-PCR samples were processed for cleaning-up using AMPure XP SPRI beads (Agencourt Bioscience, Beverly, MA). In this step, 2x concentration of SPRI beads was added to each DNA sample from the PCR step, mixed thoroughly and incubated at room temperature for 5min. The reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 200µl of 70% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then

this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10-20min. Finally, 25µl of elution buffer (EB from the Qiagen kit) was added and mixed with sample to elute the DNA from the magnetic beads, and the eluted samples were transferred next into new tubes.

2.10.1.3 DNA quality and concentration assessment

DNA samples were checked on an Agilent 2200 TapeStation D1K High Sensitivity Screentape (Agilent technologies, Inc., USA) as described in section 2.8.2.2. For each DNA sample, the concentration ratios of the control DNA to the test DNA was calculated using the following formula:

$Conc^n$ Ratio 95bp =

$Conc^n$ 95bp (test sample)/$Conc^n$ 95bp (control sample)

$Conc^n$ Ratio 235bp =

$Conc^n$ 235bp (test sample)/$Conc^n$ 235bp (control sample)

Average $Conc^n$ Ratio =

($Conc^n$ Ratio 95bp + $Conc^n$ Ratio 235bp)/2 = QC Score

The QC Score of each sample can be used then to determine the amount of DNA that should be used for DNA sequencing libraries the guidelines in Table 2.23:

Table 2.23 QC score and DNA input guidelines.

QC Score	Adjusted DNA input
1+	200-500ng
0.5-1	500-1000ng
0.01-0.5	1000-2000ng

If less DNA amounts than recommended by the QC score were available, PCR cycles were increased from 12 to 14 in the ‘adaptor-ligated library amplification’ step, section 2.10.3.3.

2.10.2 Pre-capture library preparation for SureSelect^{XT} target enrichment paired end sequencing

Libraries preparation protocol for the SureSelect^{XT} target enrichment system for Illumina HiSeq 2500 paired end sequencing recommends a starting high quality DNA amount of 3µg. This is usually not obtainable when working with FFPE degraded DNA, or using DNA from a low stock. The modified protocol described in this section uses less concentrating and clean-up steps to reduce DNA wastage. The resulting prepped DNA library should then contain the required 750ng recommended for hybridisation in the SureSelect^{XT} protocol.

2.10.2.1 DNA preparation

DNA quality was assessed as described in section 2.10.1 above and the amount of used DNA starting material was decided according to the guidelines in Table 2.22 above. The DNA concentration was measured and confirmed on the day of library prep, using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit as described in sections 2.7.2.1.

2.10.2.2 Shearing

750ng of DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA)

2.10.2.2.1 Sample preparation and shearing

To prepare the added amounts of each DNA sample, the following calculations were made:

Concentration of Stating Material: (Picco Green result, ng/µl)

Volume of DNA Used (µl): (750 / Concentration of Stating Material)

Volume of EB Buffer used (µl): 50 - Volume of DNA Used (µl)

Total Amount of DNA Used (ug): (Concentration of Stating Material x Volume of DNA Used (µl)) / 1000

DNA was sheared on a Covaris S2 Sample Preparation System (Covaris Inc., USA). The Covaris S2 Focused-ultrasonicator was used for each sample to shear the DNA at 4^{o}C in run mode using the below settings shown in Table 2.24.

Table 2.24 Covaris S2 run settings

Temperature of water bath	4^{o}C
Water level	12
Duty Cycle	20%
Intensity	5
Cycle Burst	200
Time	140sec

2.10.2.2.2 Agilent 2200 Tapestation checkpoint

After shearing, 2µl of each DNA sample were checked for appropriate size distribution on an Agilent 2200 TapeStation D1K High Sensitivity Screentape (Agilent technologies, Inc., USA) as described in section 2.8.2.2.

2.10.2.3 End-repair

DNA fragments were end repaired to convert DNA containing damaged ends to blunt-ended, 5´-phosphorylated DNA using reagents from NEBNext ultra DNA prep kit (New England BioLabs inc., UK). Reaction components in Table 2.25 were mixed by pipetting, quickly centrifuged to collect all the liquid from the side of the tube, added to sheared DNA samples in a PCR plate, and incubated for 30min at 20^{o}C, then for 30min at 65^{o}C.

Table 2.25 End-repair reaction components.

Component	**Volume x1**
Sheared DNA	48µl
End Repair reaction buffer	6.5µl
EB buffer	7.5µl
End-It Enzyme mix	3µl

2.10.2.4 Adaptor ligation

End-repaired DNA fragments were dA-tailed so that adapters could be ligated to the ends of the DNA with SureSelect Adaptor. Refer to Table 2.26 below for reaction components. In a PCR plate, 18.5µl of the reaction master mix was added to the end-repaired DNA samples: 65µl, and incubated at room temperature for 15min. Next, samples were cleaned and smaller fragments of DNA were removed using AMPure XP beads (Agencourt Bioscience, Beverly, MA).

Table 2.26 Adaptor ligation reaction components.

Component	**Volume per DNA sample/**µl
Blunt/TA Ligase Master Mix	15µl
SureSelect Adaptor Oligo Mix	2.5µl
Ligation Enhancer	1.0µl
Total volume	18.5µl

2.10.2.5 Clean up with AMPure XP beads

All DNA samples were transferred into 1.5ml tubes. Next, samples were processed for cleaning-up using AMPure XP beads (Agencourt Bioscience, Beverly, MA). In this step, 1x concentration of AMPure XP beads (83.5µl) was added to each DNA sample from the previous step, mixed thoroughly and incubated at room temperature for 5min. The reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 200µl of 70% ethanol was added and incubated for 30sec at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10-20min. Finally, 23µl of 0.1x TE

buffer was added and mixed with sample to elute the DNA from the magnetic beads, and the eluted samples were transferred next into new tubes.

2.10.2.6 PCR amplification

Adaptor ligated DNA fragments were subjected to PCR amplification using 12-enrichment PCR cycles. Refer to Table 2.27 below for reaction components. In a PCR plate, 27µl of the master mix was added to 23 µl of each DNA sample (DNA from previous step) and placed into the thermal cycler. Refer to Table 2.28 for PCR reaction conditions.

Table 2.27 PCR amplification reaction components.

Components	Volume per DNA sample/µl
NEB High Fidelity 2x PCR master mix	25µl
Sure Select Primer	1.0µl
SureSelect ILM Indexing Pre Capture PCR Reverse Primer	1.0µl

Table 2.28 PCR program

Time	Temperature
30sec	98°C
15 cycles of:	
10sec	98°C
30sec	65°C
30sec	72°C
5min	72°C
Hold 4°C	

2.10.2.7 Clean up of PCR product with AMPure XP beads

All PCR products were transferred into 1.5 ml tubes. Next, the post-PCR DNA samples were processed for cleaning-up using AMPure XP beads (Agencourt Bioscience, Beverly, MA). In this step, 1x concentration of AMPure

XP beads (50µl) was added to each DNA sample from the PCR step. The rest of the cleaning steps were exactly as described in section 2.10.2.5 above, except for the final elution that was made using 25µl of nuclease free water.

2.10.2.8 Quality and quantity assessment of pre-capture library

Libraries were examined using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit and Agilent 2200 TapeStation D1K High Sensitivity ScreenTape to assess DNA concentration and quality, respectively. Protocols are described in sections 2.7.2.1 and 2.8.2.2. Once the concentration is determined, 750ng of each library was calculated and transferred to a new, labelled 1.5ml tube and proceed to hybridisation (SureSelect^{xt} target enrichment system for illumina paired-end sequencing protocol).

2.10.3 SureSelect^{xt} target enrichment system for illumina HiSeq 2500 paired-end sequencing

DNA sequencing libraries were previously prepared as described in section 2.10.2 from extracted DNA using the NEBNext singleplex library preparation kit (New England BioLabs, UK). Next, exome DNA was captured following SureSelect^{XT} target enrichment System protocol (Agilent technologies, Inc., USA) for Illumina HiSeq 2500 paired end sequencing.

2.10.3.1 Library hybridisation

In this step, DNA libraries were hybridized for target enrichment. Each hybridization reaction requires 3.4µl of 750ng of the prepped library DNA. If the prepped library concentration was below 750ng, a concentrator was used to concentrate the sample at 30°C for 15min in which the entire prepped library is added to an eppendorf tube. Then, holes were made into the lid with a narrow needle and the samples were transferred to the vacuum concentrator to dehydrate. Once the run has completely finished and no liquids can be seen in the tubes, the samples were re-suspended by adding 3.4 µl of nuclease-free water and mixed on a vortex.

2.10.3.1.1 Hybridisation buffer

To prepare the hybridisation buffer, the components in Table 2.29 were mixed by pipetting, centrifuged briefly, incubated at 65°C for 5min and stored at room temperature.

Table 2.29 Hybridisation buffer mix

Components	Volume for 1 capture (µl)
SureSelect Hyb buffer #1 (orange cap)	25µl
SureSelect Hyb buffer #2 (red cap)	1µl
SureSelect Hyb buffer #3 (yellow cap)	10µl
SureSelect Hyb buffer #4 (black cap)	13µl
Total	49µl

2.10.3.1.2 SureSelect Capture Library

The SureSelect capture library (RNase Block) component mix was prepared in a PCR plate, mixed by pipetting, centrifuged briefly and stored in ice for target enrichment reaction as described in Table 2.30 below. (Library capture size < 3.0 Mb).

Table 2.30 RNase Block mix

Components	Volume for 1 capture (µl)
RNase Block (purple cap) for 1 sample	1µl
Nuclease-free water for 1 sample	3µl
2 µl is needed for each sample	

2.10.3.1.3 SureSelect Block

The SureSelect Block component mix was prepared as shown in Table 2.31, mixed to make the correct amount for the number of samples used and stored in ice.

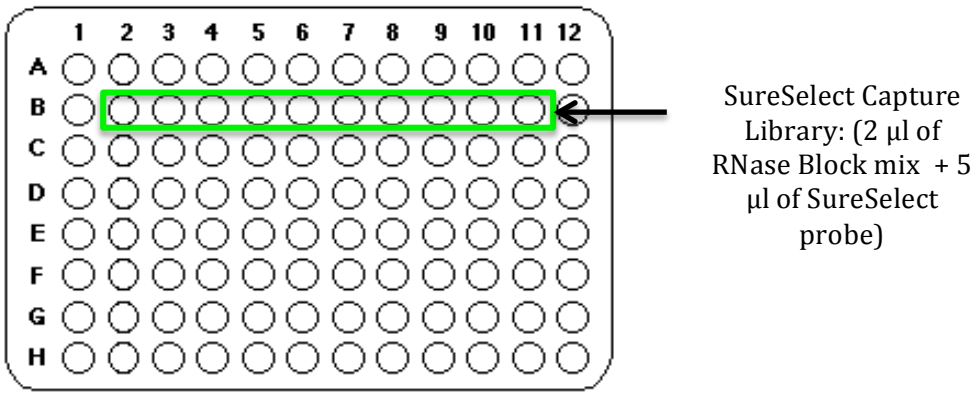
Table 2.31 SureSelect Block mix

Components	Volume for 1 capture (µl)
SureSelect Indexing Block #1 (green cap)	2.5µl
SureSelect Block #2 (blue cap)	2.5µl
SureSelect Indexing Block #3 (brown cap)	0.6µl
Total	5.6µl

2.10.3.1.4 Preparing PCR plates

Two PCR palates were used for library hybridisation reaction. In PCR plate No.1 (shown below), 2µl per sample of the RNase Block mix that was previously prepared in section 2.10.3.1.2 were added along with 5 µl / per sample of SureSelect Human all Exone V5 probe into the wells in row B, but leaving the first and last wells that are more exposed to heat and evaporation inside the thermal cycler and the PCR plate was then stored in ice.

PCR plate (1)



In PCR plate No.2 (shown below), 3.4µl of the previously prepared library DNA samples were added along with 5.6µl of SureSelect Block mix that was prepared before in section 2.10.3.1.3 into the wells in row B, but leaving the first and last wells that are more exposed to heat and evaporation inside the thermal cycler. Next, PCR plate No.2 was covered by a heated lid and placed into the thermal cycler for the indicated program below in Table 2.32.

PCR plate (2)

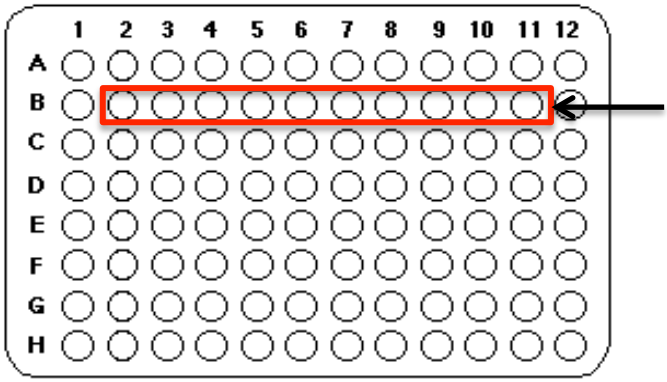
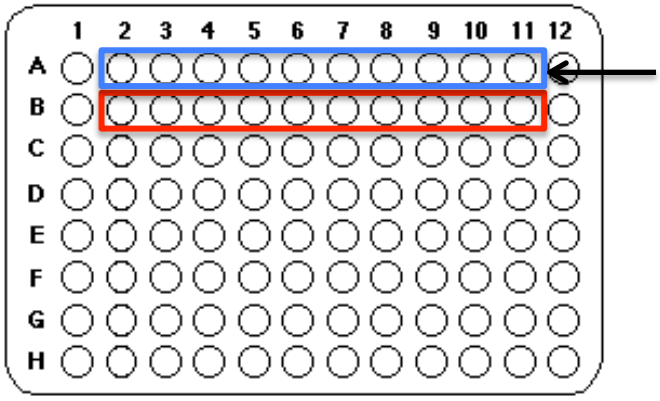


Table 2.32 PCR program

Time	Temperature
5min	95°C
Hold @ 65°C	

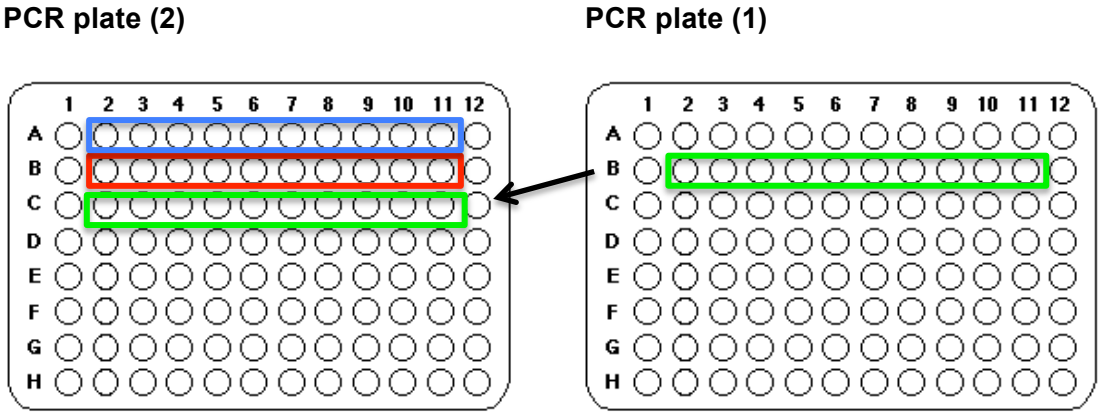
After 5min, the sealing heated lid was removed from PCR plate No.2 and 40µl of the Hybridisation buffer mix that was prepared already in section 2.10.3.1.1 was added into the wells in row A in PCR plate No.2, but leaving the first and last wells that are more exposed to heat and evaporation inside the thermal cycler. Then, PCR plate No.2 was covered by a heated lid and placed into the thermal cycler and incubated at 65°C for 5min.

PCR plate (2)

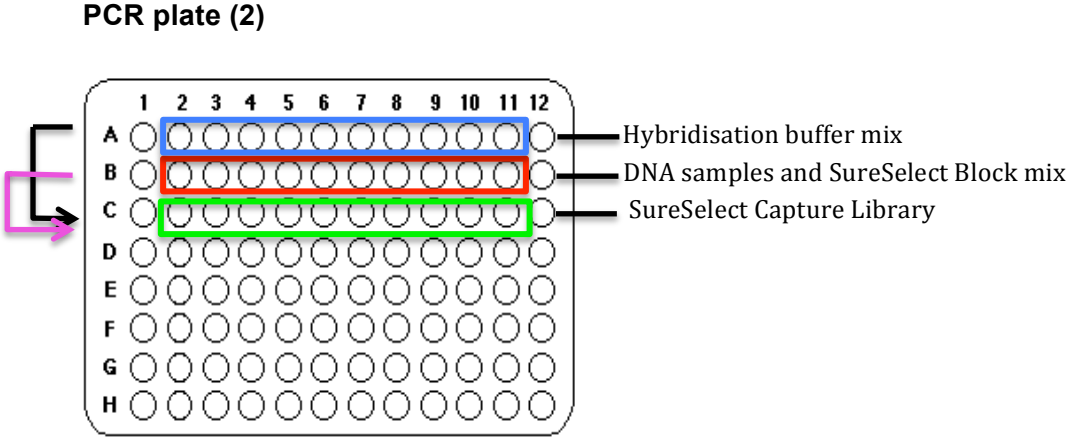


After 5min, the sealing heated lid was removed from PCR plate No.2 and the 7 µl in row B, in PCR plate No.1, was transferred into row A in PCR plate No.2.

After that, PCR plate No.2 was covered by a heated lid and placed into the thermal cycler and incubated at 65°C for 2min.



After 2min, the sealing heated lid was removed from PCR plate No.2 and 13µl from row A in the same plate were transferred to row C. Additionally, 9µl from row B in the same plate were transferred to row C and mixed well by slow pipetting. Next, PCR plate No.2 was covered by a heated lid and placed into the thermal cycler and incubated at 65°C for 24 hours. All the transferring steps were made using a multi-channel pipette and the PCR plate was maintained at 65°C.



2.10.3.2 magnetic beads preparation

The SureSelect Wash buffer #2 from the SureSelect Target Enrichment Kit (Agilent technologies, Inc., USA) was pre-warmed at 65°C. Then, 50µl of the Dynabeads MyOne Streptavidin (Invitrogen, UK) was added for each sample to a separate 1.5ml tube along with 200µl of SureSelect binding buffer. After a

brief vortexing, the reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded. The previous washing step was repeated three times followed by a re-suspension in 200µl of SureSelect Binding Buffer.

2.10.3.3 Select hybrid capture with SureSelect

While PCR plate No.2 is still at 65°C in the PCR machine, the hybridization mixture was added directly from the thermal cycler to the Dynabeads solution prepared earlier in section 2.10.3.2 and the tubes were then inverted 3 to 5 times to mix the components. Next, the tubes were incubated at room temperature for 30min on a 3D gyratory rocker (Stuart, UK). Following this, the tubes were briefly centrifuged and the reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded. The tubes were then removed from the magnetic rack and 500µl of SureSelect Wash #1 was added to each sample, vortexed and incubated at room temperature for 15min. After a brief centrifugation, the reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution and the cleared supernatant was aspirated and discarded. Next, 500µl of the pre-warmed SureSelect Wash buffer #2 was added to the tubes, vortexed and incubated at 65°C for 10min. After a brief centrifugation, the reaction tubes were then placed in an Agencourt magnetic rack for 2min to separate the beads from solution and the cleared supernatant was aspirated and discarded. This previous washing step with SureSelect Wash buffer #2 was repeated 3 times followed by a re-suspension in 30µl of nuclease-free water.

2.10.3.4 Addition of index tags by post-hybridization amplification

One amplification reaction was prepared for each hybrid capture and a negative no-template control was included. In a PCR strip tubes, the reaction mix described in Table 2.33 below was prepared on ice and mixed well on a vortex mixer. Then, 35µl of the reaction mix was added to each tube, and 1µl of the

appropriate index PCR Primer (selected from index 1 through index 16) from the SureSelect Library Prep Kit, was added to each tube and mixed by pipetting. A different index primer was used for each sample in order to be sequenced in the same lane. The PCR tubes were then placed into the thermal cycler. Refer to Table 2.34 for PCR reaction conditions.

Table 2.33 PCR amplification and index tags addition reaction components.

Components	Volume per DNA sample/µl
5× Herculase II Rxn Buffer	10µl
100 mM dNTP Mix	0.5µl
Herculase II Fusion DNA Polymerase	1.0µl
Indexing Post Capture Forward PCR Primer	1.0µl
Nuclease-free water.	22.5µl
Total 35µl	
PCR Primer Index, selected from index 1 through index 16	1.0µl
Captured on-bead DNA sample	14µl

Table 2.34 PCR program

Time	Temperature
2min	98°C
10 cycles of:	
30sec	98°C
30sec	57°C
I minute	72°C
Incubate after the final cycle for:	
10min	72°C
Hold 4°C	

2.10.3.5 Samples purification

Purification of amplified DNA libraries was carried out using Agencourt Ampure XP beads (Agencourt Bioscience, Beverly, MA). In this step, 1.8x concentration (90µl) of SPRI beads was added to each DNA sample (samples from the last

step), mixed thoroughly and incubated at room temperature for 15min. The reaction tubes were then placed in an Agencourt magnetic rack for 5min to separate the beads from solution. Once the separation occurred, the cleared supernatant was aspirated and discarded, and 500µl of 70% ethanol was added and incubated for 1 minute at room temperature. Ethanol was aspirated and discarded afterward, and then this previous step was repeated for a total of two washes. Next, all of the ethanol from the bottom of the tubes was aspirated as it may contain residual contaminants, and then magnetic beads were kept to dry at room temperature for 10min. Finally, 30µl of nuclease-free water was added and mixed with each sample to elute the DNA from the magnetic beads, incubated for 2min, and the eluted samples were transferred next into new-labelled tubes.

2.10.3.6 Library quality control

Libraries were examined using Invitrogen's Quant-iT Picogreen dsDNA BR assay kit and Agilent 2200 TapeStation D1K High Sensitivity ScreenTape to assess DNA concentration and quality, respectively. Protocols are described in sections 2.7.2.1 and 2.8.2.2.

2.10.3.7 Sample pooling and sequencing

Details of the prepared libraries were uploaded to the group server for storage. Equimolar amounts (typically 10ng) of each exome captured DNA library were pooled before being sent to the sequencing team for cluster amplification and multiplexed up to 4 samples per lane and paired-end sequenced (2X100bp) on an Illumina HiSeq 2500 to an average of 90X coverage. Illumina HiSeq 2500 sequencer produces a fast Q file that provides information about each read (including quality of read, location and size) besides the row nucleotide sequence.

2.10.3.8 Alignment and data analysis

Sequencing was performed to an average of 90X coverage. Reads were trimmed using cutadapt (Martin, 2011) to aligned to the human genome (hg19)

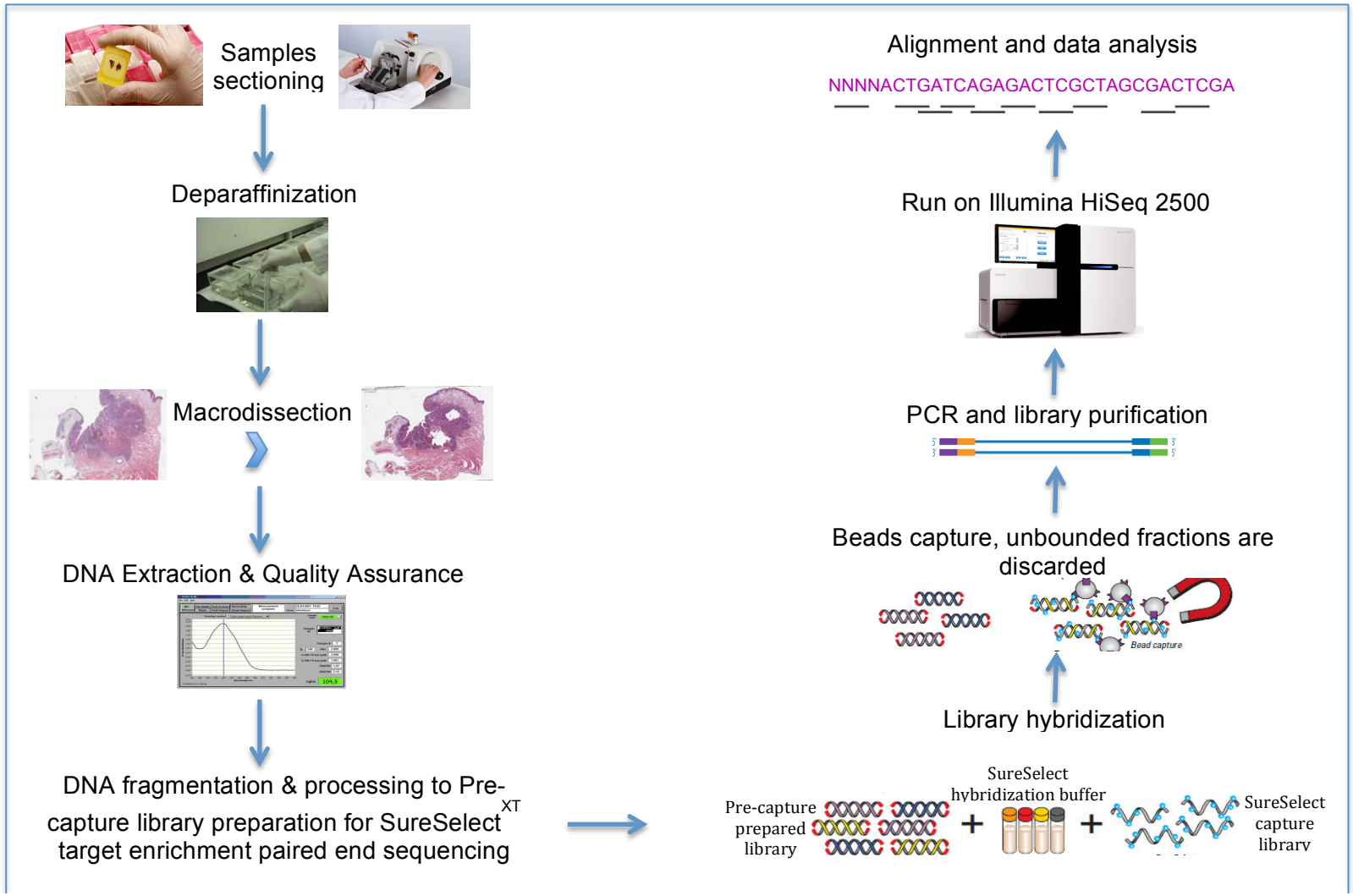
using BWA (Li and Durbin, 2009). PCR duplicates were removed using Picard (http://picard.sourceforge.net), and indel realignment and quality score calibration performed using GATK (McKenna *et al.*, 2010). Variant calling was performed by Varscan2 in somatic mode (Koboldt *et al.*, 2012), and variant consequences were then predicted using Chasm (Wong *et al.*, 2011) and Variant Effect Predictor (McLaren *et al.*, 2010). Variants were filtered using a number of criteria to enrich the mutation list for functionally important genes: Mutations had to pass a Varscan2 Phred somatic score of threshold of 15 (p-value of less than 0.05); mutations had to be present in over 50% of tumour cells, calculated taking into account tumour cell percentages and local copy number, as well as absent in the matched normal sample; mutations that have a predicted possible consequence on protein function (deleterious, splice variant, probably deleterious, possibly deleterious, stop gained).

2.10.3.9 Functional analysis of mutated genes in OVC. (DAVID gene set enrichment analysis)

I used DAVID Bioinformatics Database 6.7 to assess the functional enrichment of the significant mutated genes in OVC (Huang *et al.*, 2008).These genes were uploaded using their official gene symbol and the background from which to measure enrichment was the Homo sapiens list of all known genes. The inspection of the functional enrichment with the different biological interpretation is described in chapter 7.

2.10.3.10 Integration of exome sequencing data and RNAseq gene expression data

To highlight genes of possible clinical and biological importance in OVC, the list of significant mutated genes generated from exome sequencing analysis was integrated with genes from the significant differential expression lists for protein coding genes in OVC versus its matched normal, and OVC versus OSCCs, that was generated from RNAseq data analysis (chapter 6).


Samples sectioning
Deparaffinization
Macrodissection
DNA Extraction & Quality Assurance
DNA fragmentation & processing to Pre-capture library preparation for SureSelect XT target enrichment paired end sequencing
Alignment and data analysis
NNNNACTGATCAGAGACTCGCTAGCGACTCGA
Run on Illumina HiSeq 2500
PCR and library purification
Beads capture, unbounded fractions are discarded
Bead capture
Library hybridization
Pre-capture prepared library
SureSelect hybridization buffer
SureSelect capture library

Chapter 3 Clinicopathological study of OVH and OVC samples

3.1 Introduction

In 1948, Ackermann defined verrucous carcinoma; it is also known as Ackermann's tumour or verrucous carcinoma of Ackermann (Ackerman, 1948). VC is currently classified as a low grade, slow growing, non- metastasizing, rare variant of SCC (Barnes L, 2005) and affects both skin & mucosal sites. The most common site of verrucous carcinoma occurrence is the oral cavity (Walvekar *et al.*, 2009). Additionally, it is known to occur in the nasal cavity and paranasal sinuses, larynx, esophagus, leg, vulva, vagina, uterine cervix, scrotum, perineum, penis, and the skin (Spiro, 1998), (Walvekar *et al.*, 2009). Oral verrucous carcinoma (OVC) is a low grade, rare variant of OSCC, and it accounts for 2-10% of all OSCC cases (Pentenero *et al.*, 2011). 'Verrucous' terminology is applied for lesions that show exophytic, keratotic surfaces, made of blunt or sharp epithelial projections, filled with keratin invaginations, but without clear fibrovascular cores (Ray *et al.*, 2011a). Generally speaking, the accurate histological classification of squamous mucosal lesions with an exophytic growth pattern is often difficult and requires experience (Santoro *et al.*, 2011). Hence, OVC clinico-histopathological diagnosis is usually exclusionary and extremely problematic (Rekha and Angadi, 2010). The aetiology of OVC is not well known, and in terms of epidemiology, it is mostly seen in over sixth decade males (Ray *et al.*, 2011a), and commonly arises in buccal mucosa (Walvekar *et al.*, 2009). Nevertheless, VC has a better prognosis compared to other carcinomas (Ray *et al.*, 2011a).

Oral verrucous hyperplasia (OVH) is a histological precursor of oral verrucous carcinoma; and may transform into either an OVC or an OSCC and is believed to have the same biological features as of OVC (Wang *et al.*, 2009b). Although OVH and OVC share similar clinical and histopathological morphology, they are two distinctive oral verrucous lesions (Shear and Pindborg, 1980a), (Klieb and Raphael, 2007b), (Zhu *et al.*, 2012). Furthermore, both lesions may present clinically as a thick, extensive, white mass or plaque with an exophytic,

verrucous appearance; and therefore, there are no precise distinguishing characteristic features to differentiate between OVH and OVC (Zhu *et al.*, 2012). In 1980, sixty-eight patients with oral verrucous lesions were described by Shear and Pindborg and they indicated histopathological key points for both lesions; VH was categorised by the presence of hyperplastic epithelium adjacent to superficial normal epithelium, while VC was categorised by the presence of a hyperplastic pushing-border invasion of the epithelium into the underlying connective tissue, whereas the basement membrane is still intact (Shear and Pindborg, 1980a). To date, there have been few studies focusing on the clinicopathological features of VH and VC (Oliveira *et al.*, 2006), (Walvekar *et al.*, 2009), (Rekha and Angadi, 2010), and hardly any reports have been found on comparative clinicopathological analysis of both lesions (Zhu *et al.*, 2012).

3.2 Aim

Distinguishing OVC from OVH lesions is often difficult. Furthermore, distinguishing OVC from classical OSCC is a common problem for pathologists due to poorly-defined diagnostic criteria. The rarity of these lesions also makes them difficult to investigate, so most previous studies have been made on small numbers of cases. The objective of this study is to investigate and describe the clinical and histopathological features of OVH and OVC lesions for all the samples used in this PhD project.

3.3 Results

3.3.1 Sample selection

The rarity of OVC and OVH lesions was one of the obstacles in this project; and therefore, I looked for different sources to get the tissue materials needed for this study. An agreed collaboration have was reached after a visit to the Pathology Division, University of Torino, in Italy with Dr. Monica Pentenero and her group whom had an interest in studying OVC and had published a recent article in the field (Pentenero *et al.*, 2011). An agreed collaboration has been reached as well with Dr. Abdulaziz Al-Ajlan, the Chairman of the Department of Pathology & Laboratory Medicine, at National Guard Hospital, Saudi Arabia, and Dr. William Barratt, at the Department of Histopathology, Queen Victoria Hospital, East Grinstead, UK to use their archival OVC blocks. All pathological materials used for this PhD project from each case were available in the form of Archival Formalin-Fixed Paraffin-Embedded (FFPE) tumour blocks. Samples from Turin, Italy, were taken as sections on glass slides (10μm sections onto 10 plain glass slides from each block).

Fifteen OVC, and 13 OVH FFPE blocks were retrieved from the Pathology Department, Bexley Wing, St James's University Hospital. Fifteen OVC samples, and one OVH FFPE block were provided from the Pathology Division, Queen Victoria Hospital, West Sussex, UK. Forty OVC samples and one OVH FFPE block were provided from the Pathology Division, University of Turin, Italy. Seven OVC FFPE blocks were provided from the Department of Pathology, National Guard Hospital, Saudi Arabia. Written informed consent and approval was obtained for all patients for the use of their tissue in this research. (Local ethics committee REC reference: 07/Q1206/30) refer to appendix 2.2 for ethical approval document.

Five cases had an initial diagnosis of OVC, and following the examination of the copy number karyograms for those samples (discussed in details in chapter 4 section 4.4); they had shown OSCC chromosomal signature features. Dr Alec High and Prof. Kenneth MacLennan have performed careful histological revision, and the final diagnosis was OSCC with verrucous appearance, and

these cases were excluded from the study (explained in details in chapter 4, section 4.4.

In total, 92 oral verrucous samples, malignant (OVC) and pre-malignant (OVH) regions were identified and the original diagnoses were confirmed by Dr. Alec High (reference pathologist). World Health Organisation (WHO) definitions and criteria were used for the histological diagnosis of OVH and OVC lesions (Barnes L, 2005). Verrucous appearing, but clearly 'invasive' squamous lesions were classified as verrucous squamous cell carcinoma and excluded (Barnes L, 2005). Because of different reasons such as low yields of the extracted DNA or RNA and failed library preparations, 73 cases out of 92 were suitable for NGS analysis, including copy number analysis, RNA sequencing and Exome sequencing.

3.3.2 Clinical data and characteristics of OVH and OVC samples

Seventy-three cases out of 92 were suitable for NGS analysis in this project and clinical data were obtained for those patients only. A total of 16 OVH patients were identified, ranging from 52–80 years with an average age of 66.4 years at the time of diagnosis. There were ten females and six males. The buccal mucosa was affected in 39% of patients with OVH, followed by the tongue (22%), followed by the palate (17%) (Figure 3.1). Smoking and alcohol intake data were not available for 11 OVH patients out of 16 (refer to table 3.1).

A total of 57 OVC patients were identified, ranging from 46–96 years with an average age of 68.7 years at the time of diagnosis. There were 26 females and 31 males. The palate was affected in 26% of patients with OVC, followed by the buccal mucosa (20%), followed by the tongue (18%) (Figure 3.2). Smoking and alcohol intake data were not available for 39 OVC patients out of 57 (refer to table 3.1).

For all the 73 verrucous patients (OVH and OVC), the gender ratio was 37 males to 36 females, mean age: 68.2 years (range 46-96). The buccal mucosa was affected in 33%, followed by the tongue (19%), followed by the palate (18%). The verrucous surface is the most characteristic feature of verrucous lesions (OVH and OVC lesions). Clinically, distinguishing OVH from OVC lesions is often difficult (Zhu *et al.*, 2012).

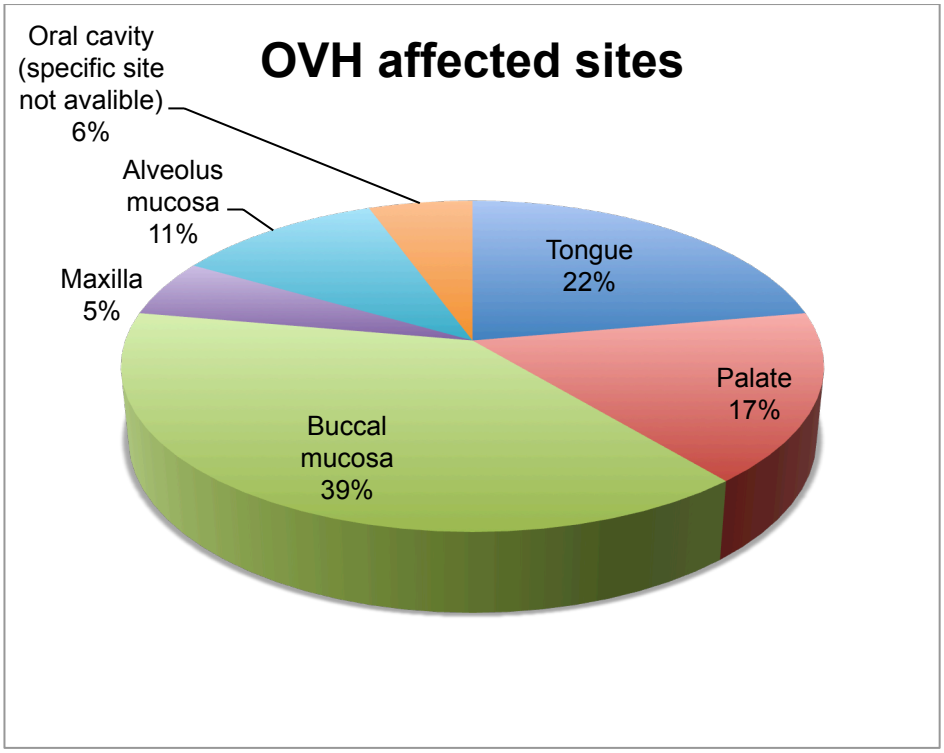


Figure 3.1 The percentage distribution of affected sites in OVH.

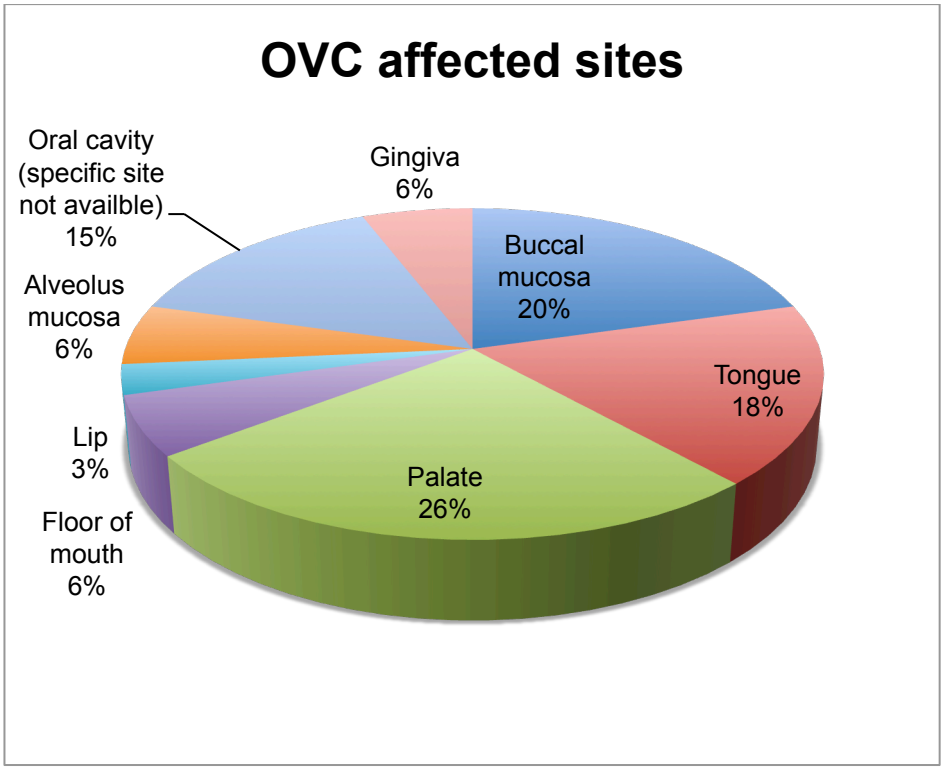


Figure 3.2 The percentage distribution of affected sites in OVC.

Table 3.1 Clinical data of 73 patients (OVH:16 cases and OVC:57 cases).

Study ID	Diagnosis	Age at diagnosis	Sex	Primary Site	Smoking (current, no, previous)	Form (eg. Pipe, snuff, cigarettes)	Drinking (units/wk)	Comments/additional data/ any history of illness	CNA	RNA seq	Exome seq
V-04-01-E9	OVC	77	M	Mandible and adjacent buccal and lingual mucosa	No		Little	Had skin cancer (scalp), father had a lung cancer	Yes	Yes	Yes
V-07-01-A	OVC	86	F	Lower lip	-	-	-		Yes		
V-10-01-1 e1	OVC	62	M	Left cheek	Yes	Chew tobacco	-		Yes		
V-014-01-5	OVC	62	M	Left buccal mucosa	Yes	Chew tobacco	-		Yes	Yes	Yes
V-019-01-I4	OVC	63	F	Buccal mucosa	-	-	-		Yes		
V-20-3	OVC	55	M	Right soft palate	-	-	-	Had multiples oral and oropharyngeal region of SCCs since 2001, lung metastases	Yes		
V-026-01-A4	OVC	72	M	Left buccal mucosa	Gave up 20 years ago	-	Yes	Hypertension	Yes		
V-060-01	OVC	65	M	Tongue	-	-	-		Yes		
V-61-1-4	OVC	96	F	Dorsum of the tongue	-	-	-		Yes		
V-62-1-B	OVC	73	F	Ventral tongue	-	-	-		Yes		
V-63-1-2	OVC	66	F	Buccsal mucosa	-	-	-		Yes		
V-65-1-D1	OVC	78	M	Upper and lower lip	-	-	-		Yes		
V-66-1-B	OVC	73	M	Buccal	Ex.	Cigarette	-		Yes		
V-67-1-B1	OVC	61	F	Floor of mouth, left.	-	-	-		Yes		
V-68-1	OVC	61	M	Buccal	-	-	-		Yes		

V-74-1-A	OVC	88	F	Hard palate	No		-		Yes		
V-75-1	OVC	49	F	Oral cavity	-	-	-		Yes		
V-77-1	OVC	81	M	Oral cavity	-	-	-		Yes		
V-78-1-A	OVC	74	M	Buccal	-	-	-		Yes		
V-79-1-B	OVC	77	F	Palate	-	-	-		Yes		
V-80-1	OVC	63	F	Alveolar ridge	-	-	-		Yes		
V-83-1-C	OVC	72	F	Gingiva/mucobuccal fold/buccal	-	-	-		Yes		
V-84-1	OVC	61	F	Upper lip	-	-	-		Yes		
V-85-1	OVC	73	M	Oral cavity	-	-	-		Yes		
V-86-1	OVC	70	M	Oral cavity	-	-	-		Yes		
V-88-1	OVC	74	M	Buccal mucosa	-	-	-		Yes		
V-89-1-A	OVC	53	M	Inside the right cheek	-	-	-		Yes		
V-90-1	OVC	61	M	Oral cavity	-	-	-		Yes		
V-91-1	OVC	60	F	Palate	-	-	-		Yes		
V-92-1	OVC	57	M	Oral cavity	-	-	-		Yes		
V-93-1	OVC	77	M	Oral cavity	-	-	-		Yes		
V-94-1	OVC	46	M	Buccal	Yes	Cigarette (10/day)	At least 20 units/week		Yes		
V-98-1-D	OVC	85	F	Tongue	-	-	-	Previous diagnosis of oral lichen planus follow-up ends on 12-01-2007	Yes		Yes
V-99-1-4	OVC	81	F	Tongue	-	-	-	Previous dysplasia same site on 1996; follow-up ends on 03/12/2007	Yes		
V-100-1-B	OVC	69	M	Buccal	Yes	Cigarette (7-8/day)	-		Yes		
V-101-1-F	OVC	80	F	Palate	-	-	-		Yes		
V-102-1	OVC	55	F	Gingiva/mucobuccal fold	-	-	-		Yes		
V-105-01	OVC	47	M	Tongue, right, lateral border	Yes	20 cigarettes/day for 30 years	-	candidal infection	Yes		
V-108-01	OVC	78	F	Tongue, left, dorsum	-	-	-	Long term lichen	Yes	Yes	Yes

								planus			
V-109-01	OVC	82	F	Hard palate, left	-	-	-	Type II diabetes, heart failure, rheumatic fever	Yes	Yes	Yes
V-110-01	OVC	82	F	Cheek mucosa, left	-	-	-		Yes		
V-112-01	OVC	79	F	Hard palate, left	-	-	-		Yes	Yes	Yes
V-113-01	OVC	64	M	Alveolus, left, maxilla	No		20 units/week	Laryngeal SCC 1997	Yes		
V-114-01	OVC	53	M	Tongue, left, anterior	No		3-4 units/week	Non-dysplastic keratosis, at the same site, 2000	Yes	Yes	
V-116-01	OVC	60	M	Hard palate, right, upper molar area	-	-	-		Yes	Yes	Yes
V-117-01	OVC	58	M	Floor of mouth, left. Invading mandible	Yes	25 roll-ups/day	None for 2 years (in jail)		Yes	Yes	
V-118-01	OVC	58	M	Cheek mucosa, left, extending to commissure	Yes	10 cigarettes/ day for 50 years	4 pints beer/day (14 units alcohol per day, 100/week)	Chronic candidosis	Yes		
V-119-01	OVC	67	F	Hard palate, left	Gave up in 2002	15-20 cigarettes/ day	No		Yes	Yes	Yes
V-122-01	OVC	76	M	Floor of mouth, right	Gave up 14 years ago	20 cigarettes/ day	6 units/day		Yes		
V-123-01	OVC	60	M	Hard palate, right, upper molar area	No		-		Yes	Yes	Yes
V-124-01	OVC	66	M	Soft palate, left	Gave up 1991	Unknown	4 units/day		Yes		Yes
V-125-01	OVC	79	M	Tongue, left	Yes	20 cigarettes/ day	-		Yes	Yes	Yes
V-95-1-B	OVH	53	M	Oral cavity	-	-	-		Yes		
V-02-02-C1	OVH	60	F	Left lower alveolus and inferior buccal mucosa	-	-	-		Yes		
V-03-01-A1	OVH	62	F	Left buccal sulcus and alveolus	-	-	-		Yes		

V-015-01-A3	OVH	75	F	Left hard palate, buccal	Gave up in 2007		No		Yes		
V-022-01-4	OVH	67	F	Right buccal mucosa	-	-	-	Type II diabetes, iron deficiency anaemia	Yes		
V-025-01-D2	OVH	80	F	Tongue, left	No		No	Lichen planus	Yes		
V-029-02-A3	OVH	67	F	Hard palate	-	-	-		Yes	Yes	Yes
V-30-1	OVH	80	F	Right buccal mucosa	Snuff only	Snuff	-	Hypertension and osteoarthritis	Yes		
V-031-01-B1	OVH	72	M	Right buccal mucosa	Gave up 20 years previously	-	Yes	Hypertension	Yes		
V-033-01-4	OVH	72	M	Left buccal mucosa	-	-	-		Yes		
V-40-1	OVH	63	F	Upper alveolus	-	-	-		Yes		
V-041-01-G	OVH	67	M	Right maxilla	-	-	-		Yes		
V-42-1-G	OVH	70	F	Tongue	No		No		Yes		
V-44-1-3J	OVH	54	M	Tongue	-	-	-		Yes		
V-46-1-2	OVH	60	M	Palate	-	-	-	Bladder cancer	Yes		
V-106-01	OVH	76	F	Tongue, left, lateral border	-	-	-		Yes		

3.3.3 Histological characteristics of OVC and OVH samples

Pathological diagnoses of OVH and OVC were made based on criteria recommended by WHO (Barnes L, 2005). The features of the diagnoses of OVH and OVC are listed in Table 3.2. Histologically, OVC consists of thickened, club-shaped papillae and blunt stromal invaginations of well-differentiated squamous epithelium with marked keratinisation (Figure 3.3 a and b), with the squamous epithelium lacking cytological criteria of malignancy. OVC invades underlying stroma with a pushing, rather than infiltrating front (Barnes L, 2005). On the other hand, the histological diagnosis features for OVH included (Figure 3.4 a and b): verrucous surface, epithelial hyperplasia with hyperkeratosis or parakeratosis, and as compared with normal adjacent mucosal epithelium, the hyperplastic epithelium does not invade the lamina propria (Barnes L, 2005), (Wang *et al.*, 2009b).

Table 3.2 histopathological diagnosis features for OVH and OVC

Histological feature	OVH	OVC
Surface	Verrucous surface	Verrucous surface
Epithelia	Epithelial hyperplasia with hyperkeratosis or parakeratosis	Thickened, club-shaped papillae and blunt stromal invaginations of well-differentiated squamous epithelium with marked keratinization and minimal cellular atypia
Invasion	No invasion into the lamina propria	Epithelial overgrowth with elongated and wide rete ridges showing a pushing-border invasion into the beneath connective tissue with intact basement membrane

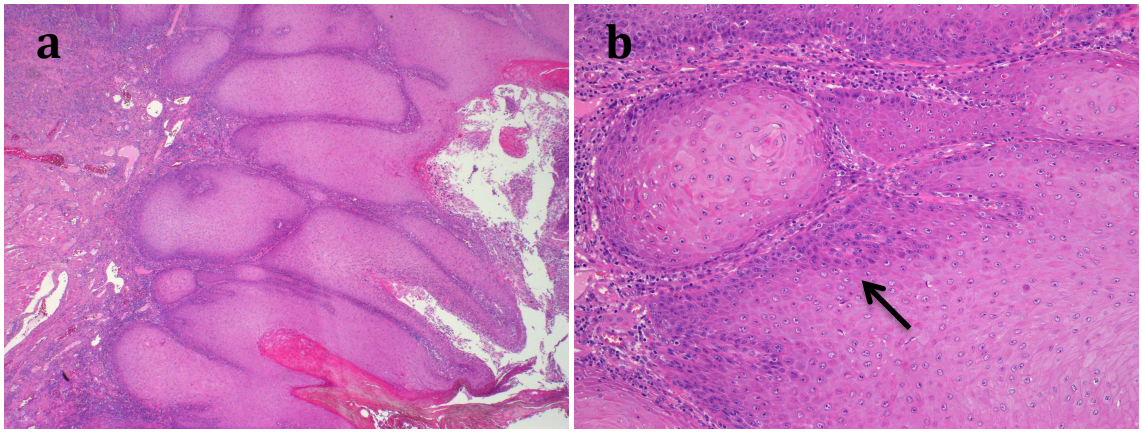


Figure 3.3 Photomicrographs of OVC.

Figure 3.3 a: showing 'extension' into underlying mucous salivary glands. It 'retains' the bulbous rete seen in verrucous hyperplasia, but clearly has a 'pushing' advancing front that now extends considerably deeper than adjacent normal epithelium. (Haematoxylin and Eosin stain. Magnification: x100 approximately). Figure 3.3 b. A higher power view of the deep margin from Figure 3a. Note the lack of nuclear pleomorphism (black arrow). (H&E stain. Magnification: x400 approximately). Figures adapted from (Samman *et al.*, 2014).

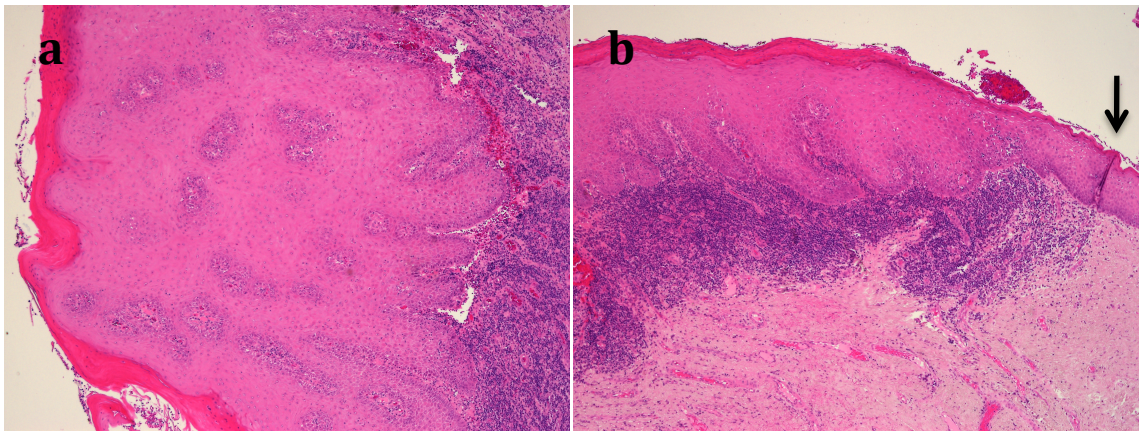


Figure 3.4 Photomicrographs of OVH.

Figure 3.4 (a & b) showing markedly acanthotic epithelium, with bulbous, 'club-shaped' papillae lacking significant cellular atypia (3.4 a), lying adjacent to more normal epithelium (black arrow) (3.4 b). (both H&E stain. Magnification: x250 approximately). In comparison to Figure 3.3, the affected epithelium does not show any significant 'deep extension'. Figures adapted from (Samman *et al.*, 2014).

3.4 Discussion

OVC is an exophytic, slow-growing, hyperkeratotic, rare variant of OSCC, that typically presents as a white, extensive, warty lesion (Jordan, 1995). OVH and OVC are clinically indistinguishable (Alkan *et al.*, 2010a) (Devaney *et al.*, 2011b). The aetiology of verrucous lesions (OVH and OVC) is still not well known (Alkan *et al.*, 2010a). Smoking, betel quid and tobacco chewing, presence of oral lichen planus, and poor oral hygiene may act as predisposing factors associated with the development of head and neck VC (Alkan *et al.*, 2010a), (Devaney *et al.*, 2011b). It is often difficult to distinguish between OVC and OVH, and differential diagnosis is usually made histopathologically. However, a good biopsy specimen is needed for correct diagnosis (Alkan *et al.*, 2010a). This chapter attempts to elucidate the clinical features and histopathological characteristics of all oral verrucous cases included in this project. Sixteen OVH and 57 OVC cases that fulfilled the histopathological criteria described above in section 3.3.3 were identified for NGS analysis (CNA, RNA-Seq and exome sequencing).

Given that the samples used in this project were obtained from different places across UK or even around the world (i.e. Leeds, west Sussex, Saudi Arabia and Italy); no noticeable differences were detected through processing them in the lab or after sequencing, accept for two cases (V-112-01 and V-119-01) that were from the same region, and their exome sequencing data showed high levels of FFPE damage. Hence, I decided to exclude those samples from my analysis (explained in detail in chapter 7, section 7.3.2).

Until now, there have been few clinicopathological studies in the literature on OVH and OVC (Wang *et al.*, 2009b), (Walvekar *et al.*, 2009), (Rekha and Angadi, 2010), (Zhu *et al.*, 2012). In general, OVH is superficial, adjacent to normal epithelium and does not extend into deeper tissues, whereas OVC spreads more deeply (Barnes L, 2005). All OVH lesions included in this project were similar in clinical behaviour and characteristic. Likewise, the histopathological appearance of all the included OVC cases in this study was concurrent with the description mentioned above in in section 3.3.3.

The current study showed that 39% of OVH lesions occurred on the buccal

mucosa. It was suggested that this site might explain the association of OVH lesions with cigarette smoking habits and tobacco or betel chewing (Wang *et al.*, 2009b). Though, (Zhu *et al.*, 2012) study showed that the most common affected site for OVH was the tongue followed by the buccal mucosa. The mean age observed at diagnosis in OVH cases here is 66.4 years, which was higher than the mean age (52 years) in (Wang *et al.*, 2009b) study and the mean age (58.5) in (Zhu *et al.*, 2012) study. In addition, OVH cohort in this project included six males and ten females; and this was different as well from (Wang *et al.*, 2009b) study, in which they had more male to female ratio. These differences are probably due to variations in the geographical areas and in study populations.

OVC is a slow-growing tumour, mainly seen in men over 60 years (Devaney *et al.*, 2011b); while here, the observed mean age at diagnosis in OVC cases is 68.7 years, which was slightly higher than the mean age (64.3 years) observed by (Zhu *et al.*, 2012) study. Furthermore, OVC cohort in this project included 31 males and 26 females; and this was very close to the gender ratio for OVC cohort (56 patients) in (Zhu *et al.*, 2012) study, in which they included 26 females and 30 males. However, their study showed that the lower lip was the predominant site of OVC lesions followed by buccal mucosa, while in this study, the palate was affected in 26% of patients with OVC, followed by the buccal mucosa (20%), followed by the tongue (18%). In contrast, previous studies reported the main site of OVC lesions was the buccal mucosa (Walvekar *et al.*, 2009), (Rekha and Angadi, 2010), and as discussed earlier, this site might explain the association of verrucous lesions with cigarette smoking and betel chewing, which were probably more in the cohorts of the previous studies due to geographic difference and ethnic population variations. It was suggested previously that oral verrucous lesions incidence are associated with alcohol drinking, smoking and tobacco or betel chawing habits (Zhu *et al.*, 2012). Whereas the proportions of alcohol users and smoking were not high in this study, and this was mainly due to lack of clinical data information for 50 oral verrucous patients (11 OVH and 39 OVC) out of 73 patients.

There are two histological variants of OVCs: the most common type is the

classical type or pure OVC tumour, and the hybrid type (Devaney *et al.*, 2011b). The classical type (as previously described in section 3.3.3) is a histologically uniform verrucous tumour that does not metastasise, and can be locally aggressive, and in contrast, the hybrid form is a mixed tumour that contains both conventional SCC and verrucous carcinoma; which hence, can metastasise (Devaney *et al.*, 2011b). Nevertheless, and as reported by (Kolokythas *et al.*, 2010), the hybrid type is not infrequent, it accounts for 20% of VC cases. Accordingly, all OVC cases should be evaluated carefully to role out a possible hybrid variant. Here, five cases had an initial diagnosis of OVC, and following the examination of the CN karyograms for those samples; they had shown OSCC chromosomal signature features described in chapter 4, section 4.4. Careful histological revision indicated the final diagnosis of OSCC with verrucous appearance (hybrid verrucous carcinoma), and these cases were excluded from the study.

In summary, this chapter elucidated the clinicopathological features of all patients with oral verrucous lesions. OVH lesions appear to occur more in patients between the 5th and 7th decade and most often on the buccal mucosa. OVC risk was found to be found higher male gender and in the elderly patients (over 60 years). Smoking and alcohol intake data were not available for 50 out of 73 oral verrucous patients; and therefore, further studies are needed to evaluate the potential risk factors roles for oral verrucous lesions.

Chapter 4 Next generation sequencing copy number analysis to identify OVC genomic characteristic features, and determine if the CNA could distinguish between the genomic damage pattern in OVH, OVC, and OSCC lesions

4.1 Introduction

OVC clinico-histo-pathological diagnosis is usually difficult and exclusionary (Ray *et al.*, 2011b).The accurate histological classification of mucosal lesions with an exophytic growth pattern encountered experience and is often challenging (Santoro *et al.*, 2011). OVC is a near-diploid aneuploid lesion by flow cytometry; on the other hand, conventional OSCC lesions show higher degree of genomic instability and aneuploidy (Pentenero *et al.*, 2011). Previous studies suggested putative tumor suppressor genes and oncogenes associated with OSCC where losses mapped on 3p, 4q, 9p, and 18q, and gains mapped on chromosomal arms 3q, 6q, 8q, 9p, 9q, 11p, 11q, 14q, 17q and 20q (Snijders *et al.*, 2005), (Baldwin *et al.*, 2005), (Liu *et al.*, 2006), (Nakamura *et al.*, 2008), (Freier *et al.*, 2010), (Jarvinen *et al.*, 2008).

OVH is a precursor of OVC that was first described by Shear and Pindborg (Shear and Pindborg, 1980b), and may transform into either an OVC or an OSCC. OVH resemble OVC both histologically and clinically, and the distinction of OVC from OVH cannot rely on the cytological features (Woolgar and Triantafyllou, 2009), (Eversole and Papanicolaou, 1983). The possible link between the biological behavior and the structural and functional features with the clinical outcome in OVCs is still not clear and requires further investigation. Furthermore, previous immunohistochemical studies to distinguish between oral verrucous lesions and OSCCs have yielded mixed results (Devaney *et al.*, 2011a). For this reason, new experimental approaches are needed to improve the pathological diagnostic criteria for OVH and OVC, and for better understanding of the development and progression of those lesions.

Cancer cells usually show several karyotypic changes: structural rearrangements such as deletions, amplifications and translocations, and whole chromosome gain or loss that result in extensive aneuploidy (Hartwell and Kastan, 1994b). Identifying Copy Number Alterations (CNA) in cancer cells is an essential step toward determining chromosomal regions with breakpoints and to access chromosomal rearrangements severity. Additionally, comparison of copy number (CN) genomic profiles between tumours from different patients could define common lost or duplicated regions to highlight the positions of oncogenes or tumor suppressor genes (Hartwell and Kastan, 1994a).

Copy Number Alterations can be detected using several methodologies, single nucleotide polymorphism array (SNP arrays) (Bignell *et al.*, 2004), Comparative Genomic Hybridisation (CGH) (Kallioniemi *et al.*, 1992), array Comparative Genomic Hybridisation CGH (aCGH) (Pinkel *et al.*, 1998), and lately, Next Generation Sequencing (NGS) platforms (Illumina MiSeq, GAII, HiSeq, Roche 454, Ion Torrent PGM, ABI SOLiD). In addition, and whilst sequencing technologies are currently becoming more widespread, accurate and affordable low coverage sequencing CN analysis will become more informative and expedient (Wood *et al.*, 2010) (Gusnanto *et al.*, 2012). Next-generation sequencing techniques offer considerable benefits for CN analysis, including precise delineation of the CN breakpoints, and higher resolution (can detect single-base insertions or deletions) of CN changes (Meyerson *et al.*, 2010). It enables the estimation of tumour-to-normal CN ratio at a genomic locus through counting the number of reads at this locus in normal and tumour samples (Meyerson *et al.*, 2010). Nevertheless, sequencing data can be produced even with nanogram amounts of DNA extracted from formalin-fixed paraffin-embedded (FFPE) materials (Wood *et al.*, 2010).

4.2 Aims

The cancer genome of any tumour involves a number of genetic abnormalities, some are tumour type-specific, and many are idiopathic. The aims of the work described in this chapter are:

1. To use next generation sequencing copy number analysis to identify

OVC and OVH genomic characteristic features.

2. To use next generation sequencing copy number analysis data to distinguish between the genomic damage pattern in OVH, OVC, and OSCC lesions.

4.2.1 Aim 1: To use next generation sequencing copy number analysis to identify OVC and OVH genomic characteristic features

4.2.1.1 Results

4.2.1.1.1 Characteristics of the study cohort

From the verrucous cohort study described in Chapter 3, pathological materials from each case were available in the form of FFPE tumour blocks. Written informed consent and approval was obtained for all patients for the use of their tissue in this research. (Local ethics committee REC reference: 07/Q1206/30). In total, 92 oral verrucous samples, malignant (OVC) and pre-malignant (OVH) regions were identified and Dr Alec High assessed the original diagnoses. Samples from 73 patients out of 92 were selected for CN analysis and 19 samples failed (eight yielded low DNA amounts after extraction, and 11 failed library preparations) refer to appendix 4.1 for failed samples. Neither the age of the FFPE block, nor the amount of tissue sampled, correlated with low yields of DNA. A database was created to record the patho-clinical characteristics for the 73 patients (37 males and 36 females, mean age 68.2 years, range 46-96), refer to Table 4.1 below and Table 3.1 in chapter 3. The site distributions of primary verrucous lesions (OVC and OVH) are illustrated in figure 4.1.

Table 4.1 Overview of analysed verrucous lesions clinicopathological data.

Type of lesion	N	Age	Male	Female
OVC	57	46-96 (Mean=68.7)	31	26
OVH	16	52-80 (Mean=66.4)	6	10

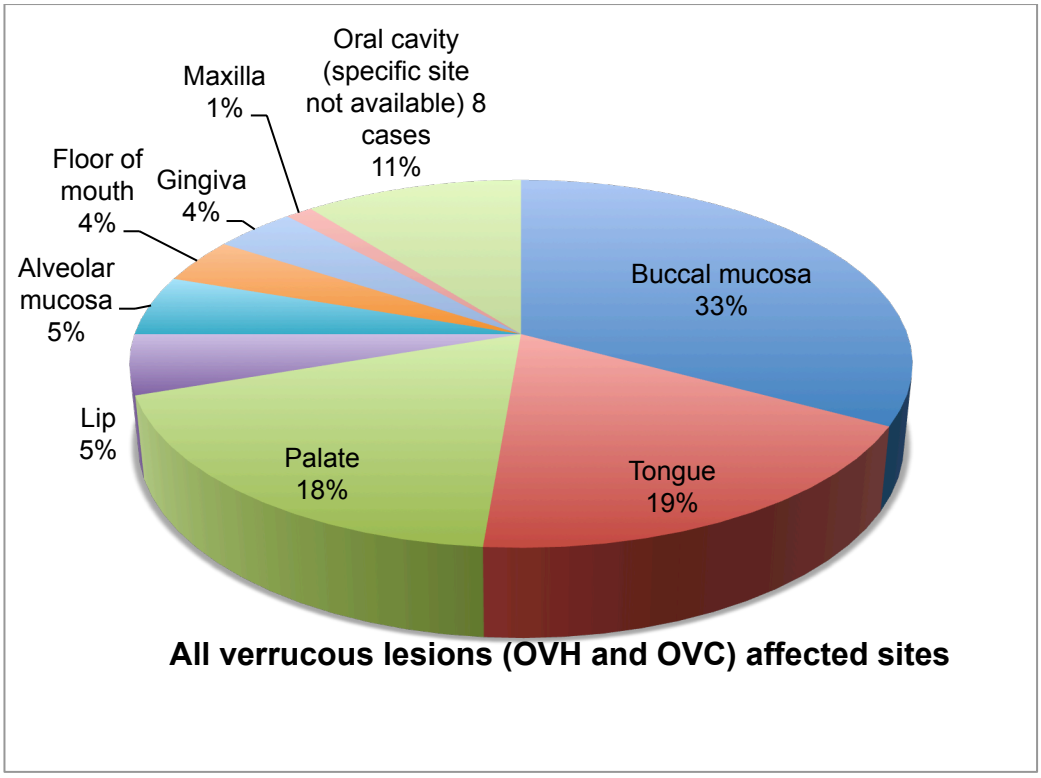


Figure 4.1 Site distributions of primary verrucous lesions (OVC and OVH).

Buccal mucosa 25 cases, tongue 14 cases, hard palate 12 cases, lip 4 cases, alveolar mucosa 4 cases, floor of mouth 3 cases, gingiva 3 cases, soft palate 2 cases, maxilla 1 case, oral cavity (specific site not available) 8 cases.

4.2.1.1.2 Genomic CN analysis of OVC and OVH samples

High-resolution mapping of CNV involves sequenced reads aligned to a reference genome (Xie and Tammi, 2009). Then, the aligned reads distribution is analysed on a genomic segmental window-by-window basis to define alterations in read-depth between the reference genomes and tests (Hayes *et al.*, 2013). When compared to the control sample, a reduction in sample read-depth across a window suggests a loss in genomic component; an increase in read-depth represents a gain (Chiang *et al.*, 2009).

Here, Between one and ten million reads per sample were generated for copy number analysis, equating to one read to every 300bp-3Kb, or 0.033X - 0.33X coverage, and the CN was calculated and analysed as described in Chapter 2 section 2.3.

In general, sequence variations are usually not distributed uniformly within genomes (Nguyen *et al.*, 2006). Nonetheless, CNVs that are enriched in simple tandem repeats occur more often towards centromeres and telomeres, and are not elevated in G + C content or SNPs (Nguyen *et al.*, 2006). Since these regions are not overrepresented in CNVs (Sharp *et al.*, 2005), I excluded CNAs in all centromere and telomere chromosomal regions throughout the different CN data analysis approaches.

Visual inspection of the 73 patient (OVHs and OVCs) genomic CN karyograms demonstrated regions of gain and loss along the whole genome in OVC cases (refer to chapter 2, section 2.8.3.1.1, for the visual examination criteria of the CN karyograms). In general, OVC karyograms showed various copy number patterns, ranged from whole chromosome gain to amplified or lost chromosome arms and regions. Figure 4.2 below (from case V-109-N) shows a CN karyogram for a histologically normal oral epithelium tissue, with no chromosomal gain or loss.

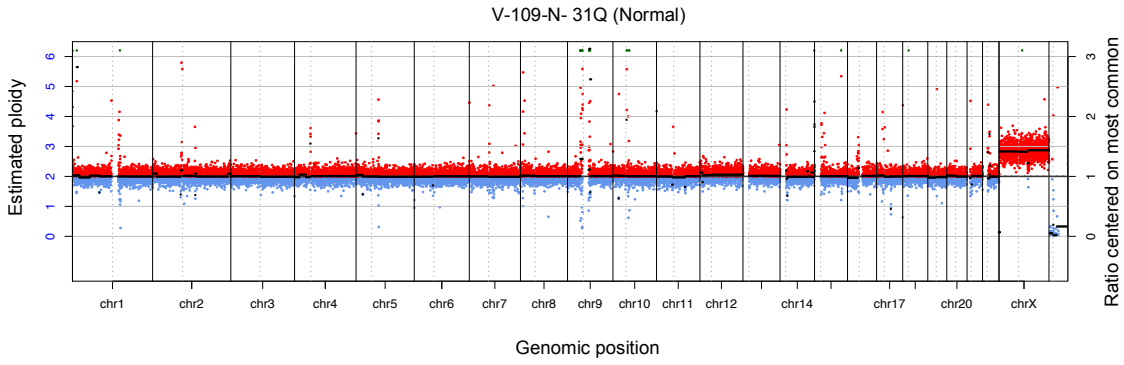


Figure 4.2 An example of the genomic profile of a histologically normal oral epithelium by NGS CN analysis.

Each data point represents one window of approximately 200 reads. Genomic position is on the x-axis and tumor:normal ratio is on the y-axis. The black lines are regions of common copy number between breakpoints. Windows of gain and loss are red and blue respectively.

4.2.1.1.2.1 Genomic profiling of OVH by NGS CN analysis

In general, gain and loss features were minimally found in OVH cases. Visual examination of the 16 OVH copy number traces revealed a very lower level of the genomic damage compared to oral verrucous tumours (N: 57), indicating that the genomic profile of these cases has minimal chromosomal abnormalities and is more similar to normal. (Refer to appendix 4.2 for all OVH and appendix 4.3 for all OVC karyograms). Whole genomic profile of a representative OVH sample (figure 4.3) from case V-029-02-A3 is shown below. The blue arrow points to gain in Chr7. Horizontal lines above the centre demonstrate regions of gain, and those below the centre demonstrate regions of loss.

These findings are surprising since it has been well-known that OVH shares similar clinical and histological morphology to OVC, and the clinical differentiation of the verrucous hyperplastic lesions from OVC is often difficult (Shear and Pindborg, 1980b), (Poh *et al.*, 2001), (Zhu *et al.*, 2012). From what has been found here, and despite the similar clinical and histological features that OVH and OVC share, the analysis of OVH individual CN karyograms showed that these lesions has different genomic profile from OVC with very low, narrow levels of DNA aneuploidy.

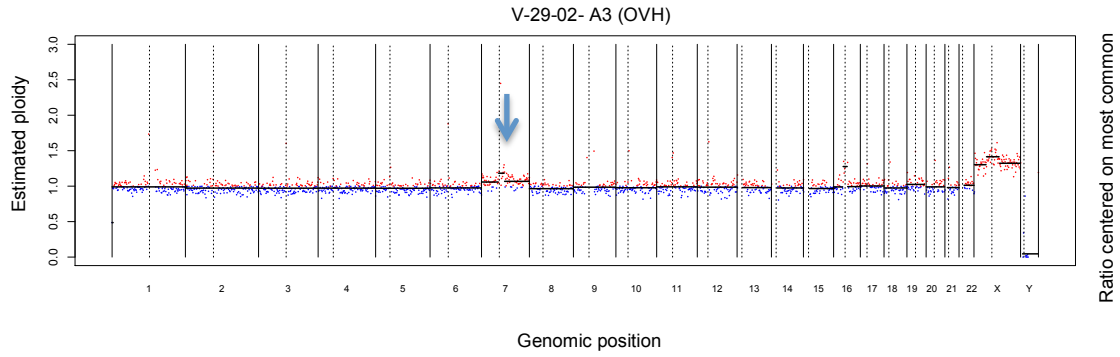


Figure 4.3 An example of the genomic profile of an OVH sample by NGS CN analysis.

Each data point represents one window of approximately 200 reads. Genomic position is on the x-axis and tumor:normal ratio is on the y-axis. The black lines are regions of common copy number between breakpoints. Windows of gain and loss are red and blue respectively. Blue arrow points to gain in chromosome 7.

4.2.1.1.2.2 Genomic profiling of OVC by NGS CN analysis

Visual examination of OVC (N: 57) genomic CN karyograms revealed a higher level of CN alterations compared to OVH. OVC karyograms appear to be in an early stage of DNA near-diploid aneuploidy (Refer to appendix 4.3 for all OVC karyograms). In addition, and as shown in figure 4.4, gains at 7q, 16q and 17q (represented by red with black lines) were detected frequently in the OVC cohort, suggesting that these CN alterations may be involved in the development of OVC. Notably, deletion trends were minimally found in OVC's, suggesting that overexpression of oncogenes is most likely to be involved in the development of OVC. Whole genomic profiling of a representative OVC sample (figure 4.4) from case V-78-01-A is shown below. Blue arrows point at gains in Chr2, Chr7, Chr10, Chr16 and Chr17. Horizontal lines above the center demonstrate regions of gain, and those below the center demonstrate regions of loss.

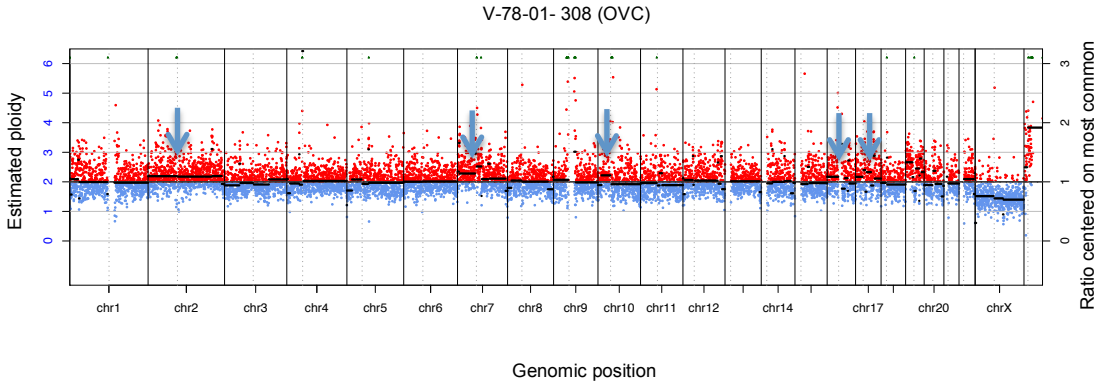


Figure 4.4 Representative genomic profiling of OVC by NGS CN analysis.

Each data point represents one window of approximately 200 reads. Genomic position is on the x-axis and tumor:normal ratio is on the y-axis. The black lines are regions of common copy number between breakpoints. Windows of gain and loss are red and blue respectively. Blue arrows point to regions of chromosomal gain.

4.2.1.1.2.3 Comparison of the genome-wide frequency karyograms of CNAs in OVH and OVC

In order to compare OVH and OVC as groups; frequency accumulative karyograms were produced using a program that takes all BED files from the CN analysed samples lists. The selected CN threshold of 0.05 above or below was considered a gain or loss. In general, visual examination of OVH (N: 16) genomic CN frequency karyogram (figure 4.5 a) noticeably illustrates the very low level of CN alterations in OVHs in compare to OVCs, indicating that the genomic profile of these cases has minimal chromosomal abnormalities and is most similar to normal. The genomic CN profile of each chromosome from OVH frequency karyogram is shown below in figure 4.6. Visual examination was carried out on chromosome plots in order to investigate the genomic locations of chromosomal segments with altered CN in OVHs. Gains mapped at chromosome 7q11.2 and 7q22 (represented by red colour) were noticed in OVHs at a frequency of ~50%, suggesting that this CN alteration may be related to the development of OVH. These results are different from a study in 2001, which reported a high frequency of allelic loss in 20/25 OVH cases at loci on 3p, 9p, 4q, 8p, 11q, 13q and 17p chromosome arms, and suggested that LOH on these arms may explain the malignant potential of OVH lesions (Poh *et al.*, 2001). Allelic loss without CN loss is possible, however, it is unlikely not to identify any in OVH cohort here at all these loci. Nonetheless, it is important to keep in mind the inability of the LOH techniques to identify chromosomal gains, which differ from aCGH or NGS CN analysis capabilities in detecting both, chromosomal losses and gains (Mohapatra *et al.*, 2006). Interestingly, the CN gain in OVH group at chromosome 7q arm, with a frequency of ~50%, was present as well in OVCs, suggesting that this region might harbour the first CN alteration involved in the development of oral verrucous lesions.

Visual examination of OVC (N: 57) genomic CN frequency karyogram (figure 4.5 b) revealed a higher level of CNA compared to OVH. Locus-specific differences in CN can also be seen by comparing chromosome plots frequency diagrams of the two groups (figure 4.6, figure 4.7). In OVCs, there is no loss at chromosome 3p or gain at 3q arms; which are the main chromosomal

abnormality features in OSCCs. Furthermore, gains mapped at chromosome 7p22, 7q11.2 and 7q22 (represented by red colour) were observed in OVCs at a frequency of ~50%, in addition to gains mapped at chromosomes 3p21 (at a frequency of ~30%), 15q15 (at a frequency of ~30%), 16q22 (at a frequency of ~25%) and 17q23 (at a frequency of ~25%), as well as losses on chromosomes 6p21 (at a frequency of ~25%) and 17q12 (at a frequency of ~50%) represented by green colour, suggesting that these CN alterations may be involved in the development of OVC.

Gains at 7q, 16q and 17q were detected in OVCs at a frequency of 50% and has not been previously identified as a common CN altered chromosome lesions in oral cancer. Deletion trends were also minimally found in OVC's frequency karyogram. In 2001, a study was conducted to investigate the frequency of allelic loss in oral verrucous lesions, including 17 OVC samples (Poh *et al.*, 2001). They reported high frequency of allelic loss at loci on 3p, 9p, 4q, 8p, 11q, 13q and 17p chromosome arms and suggested that LOH on these arms may explain the malignant potential of OVCs (Poh *et al.*, 2001). From their findings, two chromosomal regions were comparable here to the CN aberrations outcome in OVC cohort (losses in chromosomes 8p23.3 and 9p21). Though, loss of chromosome 8p23.3 (which is a telomere region as well) was at a frequency of ~10%, and loss of chromosome 9p21 was at a frequency of ~5% in the OVC study here. In addition, the previous report included 17 OVC samples, while in this project, 57 OVC samples were included. Again, it is important to keep in mind the inability of LOH techniques to identify chromosomal gains, which differ from aCGH or NGS CN analysis capabilities in detecting chromosomal losses and gains (Mohapatra *et al.*, 2006).

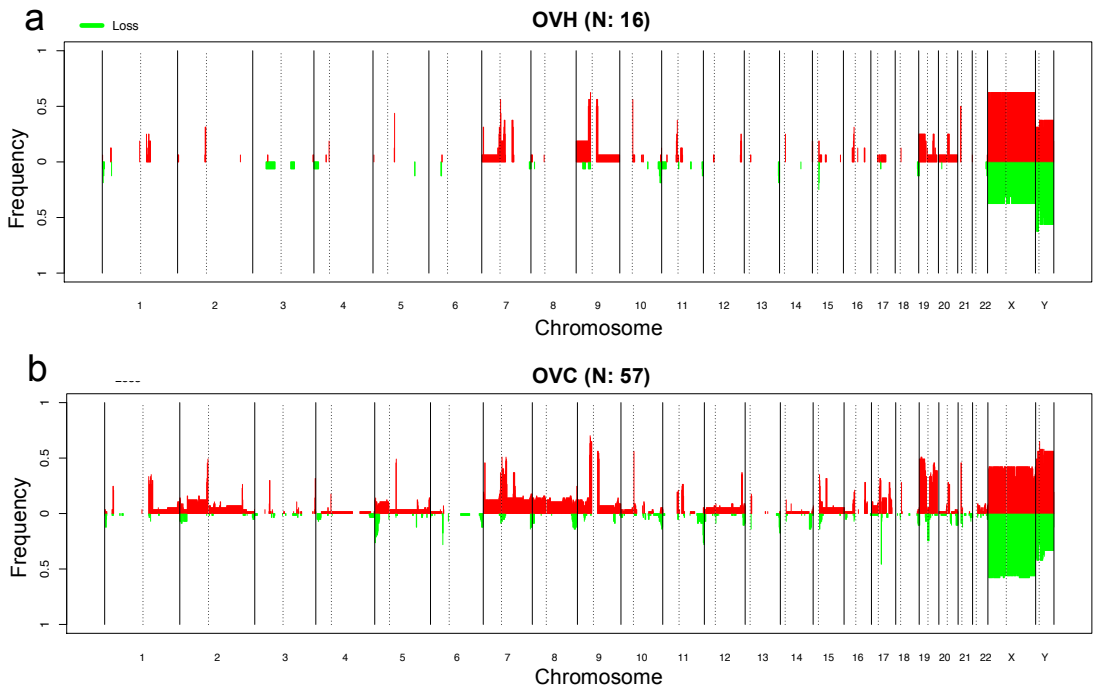
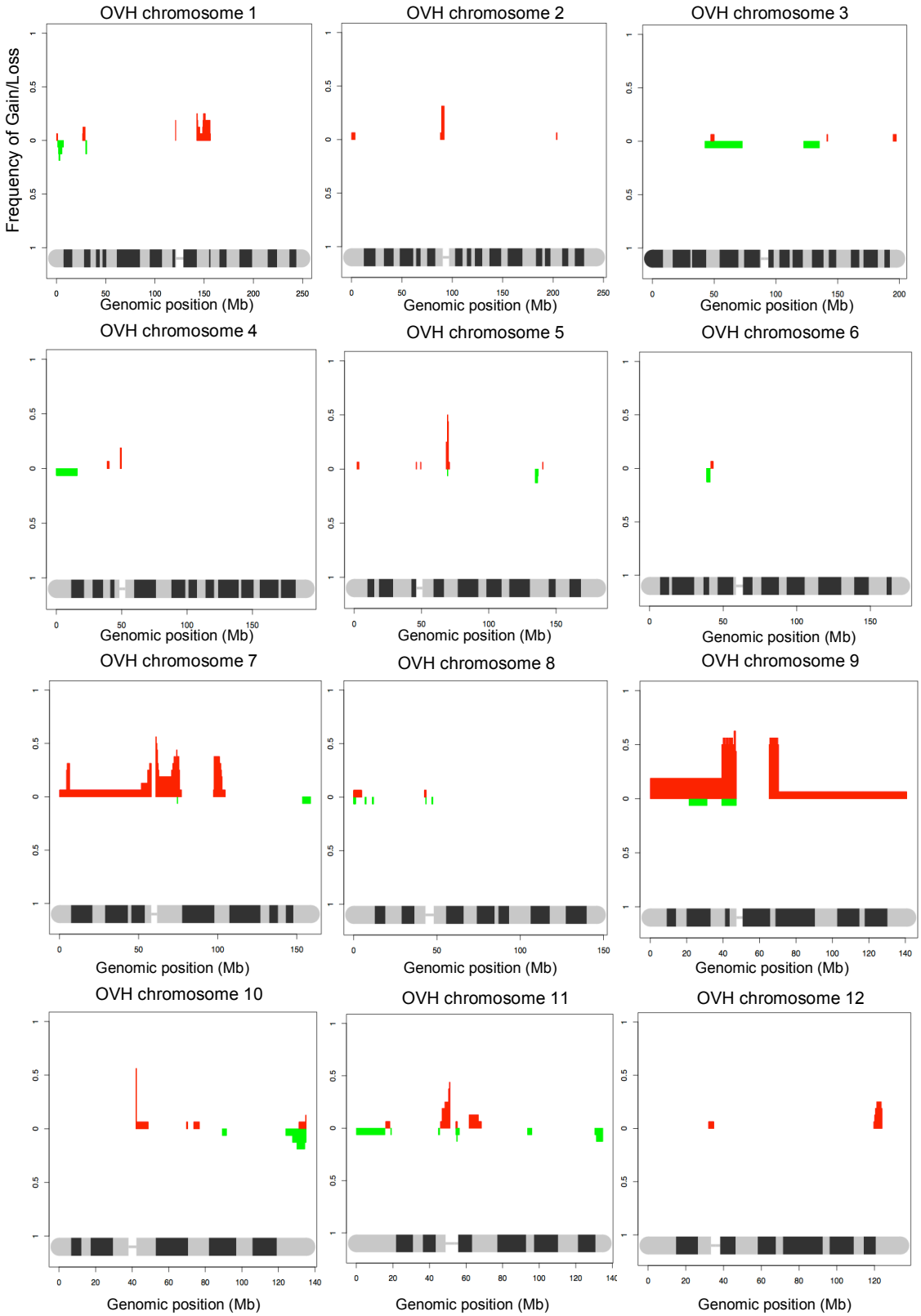


Figure 4.5 Frequency of genomic gain and loss for OVH (a) and OVC (b).

Genomic position is on the x-axis, frequency (%) of gains (red) and losses (green) are shown on the y-axis.


OVH chromosome 1
OVH chromosome 2
OVH chromosome 3
OVH chromosome 4
OVH chromosome 5
OVH chromosome 6
OVH chromosome 7
OVH chromosome 8
OVH chromosome 9
OVH chromosome 10
OVH chromosome 11
OVH chromosome 12
Frequency of Gain/Loss
Genomic position (Mb)

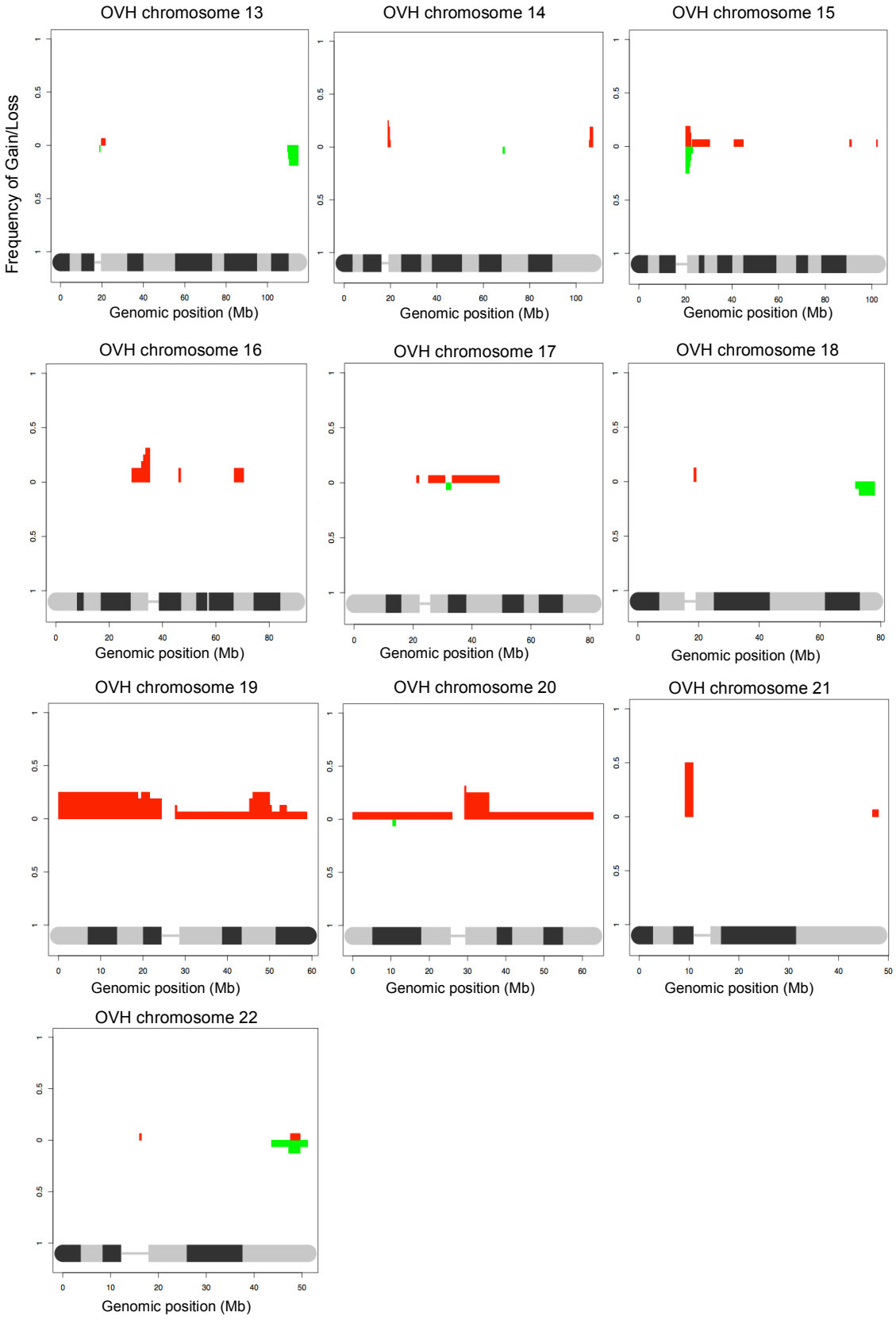
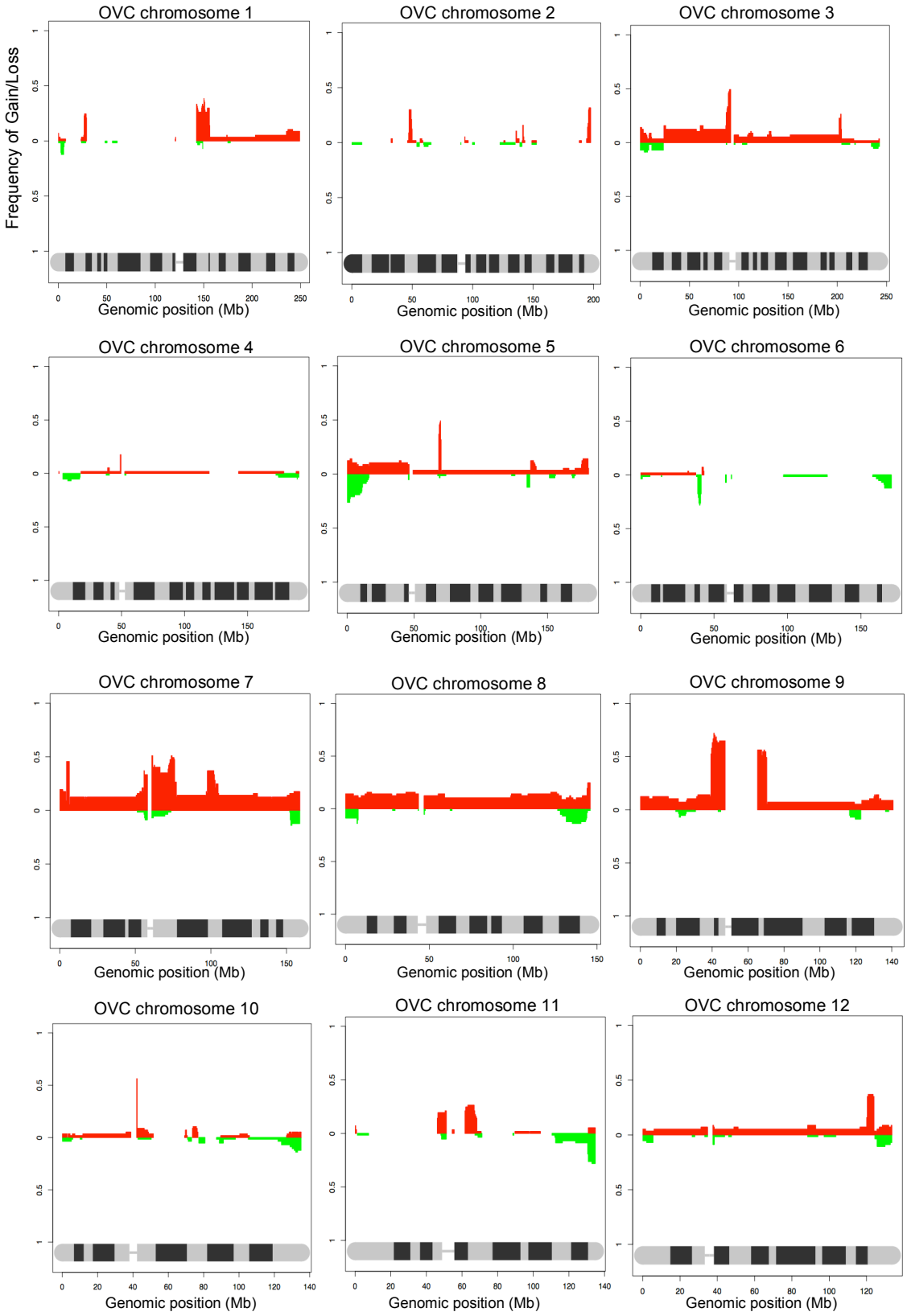


Figure 4.6 Genomic locations of chromosomal segments with altered CN in OVH.

Genomic position is on the x-axis, frequency (%) of gains (red) and losses (green) are shown on the y-axis.


OVC chromosome 1
OVC chromosome 2
OVC chromosome 3
OVC chromosome 4
OVC chromosome 5
OVC chromosome 6
OVC chromosome 7
OVC chromosome 8
OVC chromosome 9
OVC chromosome 10
OVC chromosome 11
OVC chromosome 12
Frequency of Gain/Loss
Genomic position (Mb)

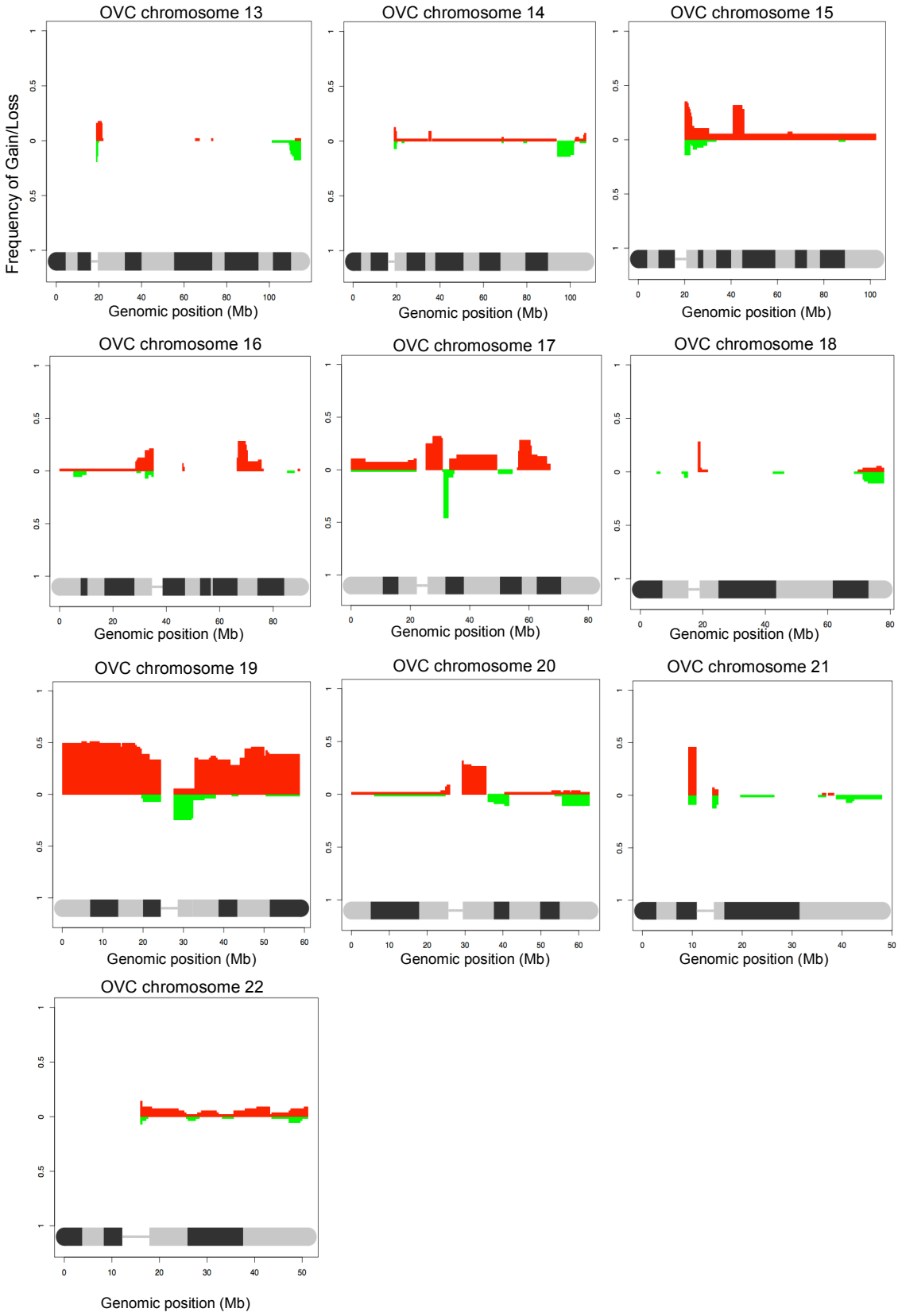


Figure 4.7 Genomic locations of chromosomal segments with altered CN in OVC.

Genomic position is on the x-axis, frequency (%) of gains (red) and losses (green) are shown on the y-axis.

4.2.1.1.3 Genomic Identification of Significant Targets in Cancer (GISTIC2.0) (computational approach)

The GISTIC algorithm identifies likely somatic driver CN alterations through evaluating the amplitude and frequency of amplified or deleted observed events (Mermel *et al.*, 2011). GISTIC has been used and applied to many cancer types, including lung and esophageal squamous carcinoma (Bass *et al.*, 2009), colorectal carcinoma (Firestein *et al.*, 2008), melanoma (Lin *et al.*, 2008), and ovarian carcinoma (Etemadmoghadam *et al.*, 2009), and has facilitated the identification of several new amplification targets, including: *SOX2 (Bass et al., 2009)*, *CDK8 (Firestein et al., 2008)*, *NKX2-1 (Weir et al., 2007)*, and *VEGFA (Chiang et al., 2008)*, besides deletions in: *EHMT1* (Northcott et al., 2009).

Here, and in order to compare the CN profiles of OVH and OVC; they were additionally characterised by several approaches using the GISTIC2.0 algorithm including: amplification and deletion plots of CNAs, the identification of amplification and deletion genes within CN altered regions, and segmented CN heat maps. All the parameters used in the GISTIC analysis are in appendix 2.3. A number of regions of recurrent CN gains and losses were evident in the GISTIC analysis in OVH and OVC cohorts, and matched the generated frequency karyograms CN aberrations in section 4.3.1.2.3. Genomic positions of the most significant amplification and deletion peaks (from the GISTIC analysis) including the list of genes contained in them for OVH and OVC samples were identified tables 4.2 and 4.3. The results were then further analysed by running the gene lists against cancer gene census and Stransky mutation list (76 previously identified genes in HNSCCs harbouring high statistically significant mutations) (Stransky *et al.*, 2011), as well as 13 KEGG pathways, which are more related to head and neck cancers (explained in details in sections 4.2.1.1.3.5 and 4.2.1.1.3.6 below).

4.2.1.1.3.1 Genome-wide amplification and deletion plots of CNAs in OVH

Regions of significant gains or losses were identified using the GISTIC algorithm. Two chromosomal regions (deletions) from the CNAs identified by

GISTIC analysis were significantly altered in OVH patients' genomes according to this analysis (orange highlighted chromosomal positions in Figure 4.8 b). These two deletion regions that surpass the significance threshold are in chr 5q31.1, and 17q12 with a frequency less than 20%. Surprisingly, no significant amplification regions were detected by GISTIC analysis, and the chromosomal regions shown in the amplification plot below are centromeres (e.g. Chr 7q11.1). In general, visual examination of OVH genomic CN plots noticeably illustrates the very low level of CN alterations in OVHs in compare to OVC genomic CN plots.

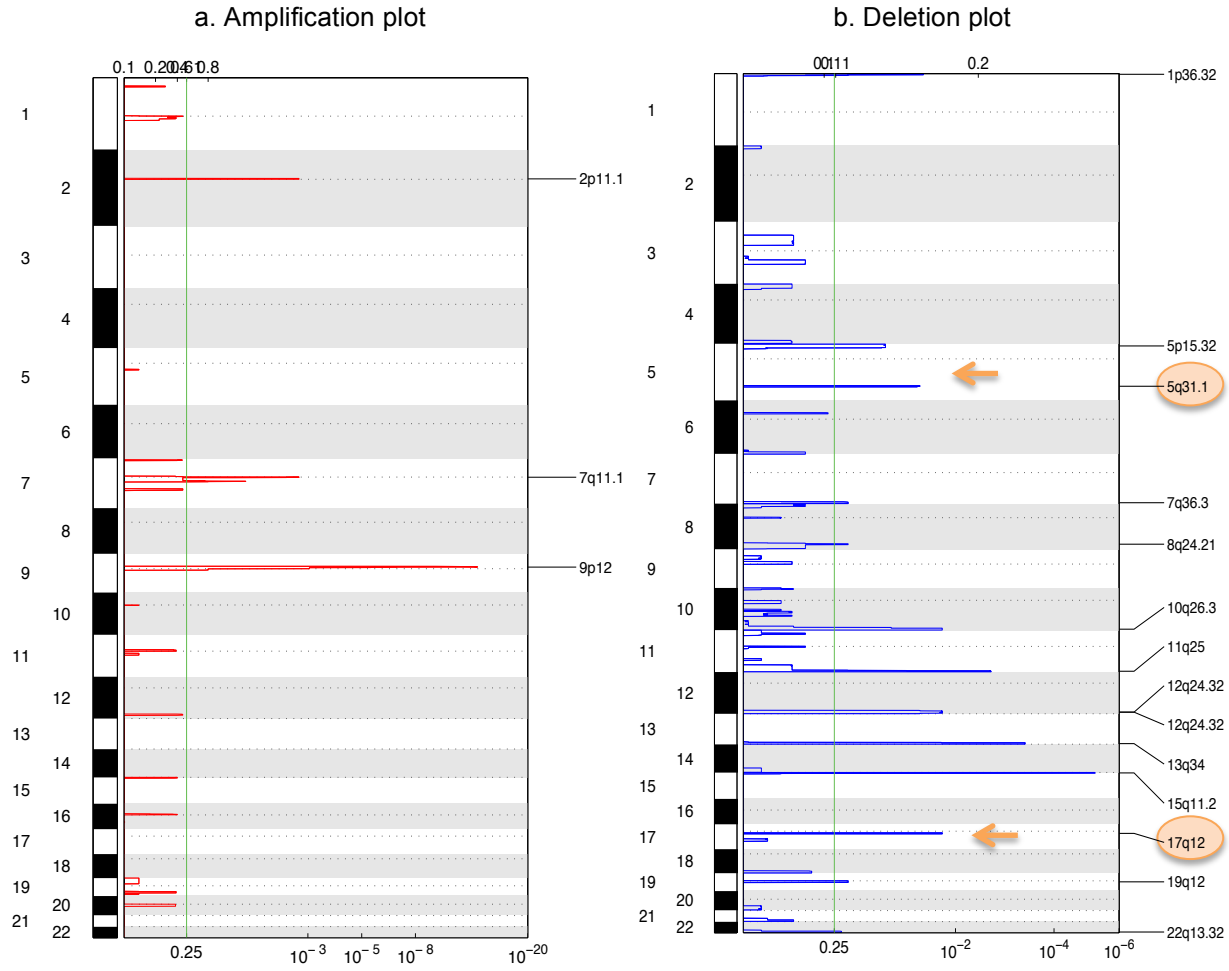


Figure 4.8 Genome-wide amplification and deletion plots of CNAs in OVH.

Genomic positions are indicated along the y axis with centromere locations showed by dotted lines. Amplification (red) and deletion (blue) GISTIC plots show q- values (bottom on the x axis), the G-scores that considers the frequency the aberration occurrence as well as its amplitude across samples (top) and the significance threshold is indicated by the green line at 0.25, with respect to amplifications and deletions for all markers over the entire analysed region. Orang arrows and circles point to regions with significant loss

4.2.1.1.3.2 Genome-wide amplification and deletion plots of CNAs in OVC

Regions of significant gains or losses were identified in OVC samples using GISTIC algorithm. Ten chromosomal regions (seven amplifications and three deletions) from the CNAs identified by GISTIC analysis were significantly altered in OVC genomes (Figure 4.9 a, and b). The seven most significant amplifications from GISTIC peaks (Figure 4.5 a, blue circles) that also surpass the significance threshold include chromosomes 3p21.31, 7p22.2, 7q11.23, 7q22.1, 15q15.2, 16q22.1 and 17q23.2. Gains mapped at chromosome 3p21 (at a frequency of ~50%), 7p22 (at a frequency of ~75%), 7q11.2 (at a frequency of ~70%), 7q22 (at a frequency of ~35%), 15q15 (at a frequency of ~40%), 16q22 (at a frequency of ~40%) and 17q23 (at a frequency of ~45%) were observed as well in OVC frequency karyograms although with different frequencies (refer to section 4.3.1.2.3). The variation in the frequencies between GISTIC analyses CN plots and the frequency karyograms generated from OVC individual CN karyograms can be attributed to the non specific visual examination method and human eye errors, as there was no algorithm to give the exact frequency percentage at the time of my analysis.

The three most significant deletions from GISTIC peaks (Figure 4.9 b, orange circles) that also surpass the significance threshold include chromosomes 5q31.1 (at a frequency of ~15%), 6p21.2 (at a frequency of ~25%) and 17q12 (at a frequency of ~15%). losses on chromosomes 6p21 and 17q12 were also observed in OVC frequency karyograms but with different frequencies (refer to section 4.3.1.2.3). Again, The variation in the frequencies between GISTIC analyses CN plots and the frequency karyograms generated from OVC individual CN karyograms can be attributed to the inaccurate visual estimation and human eye errors, as there was no algorithm to give the exact frequency percentage at the time of my analysis. Notably, losses on chromosomes 5q31.1 and 17q12 (at a frequency of ~15%) were also observed in OVH (Figure 4.8) and with similar frequencies, which suggests a possible role for those regions in the development of oral verrucous lesions.

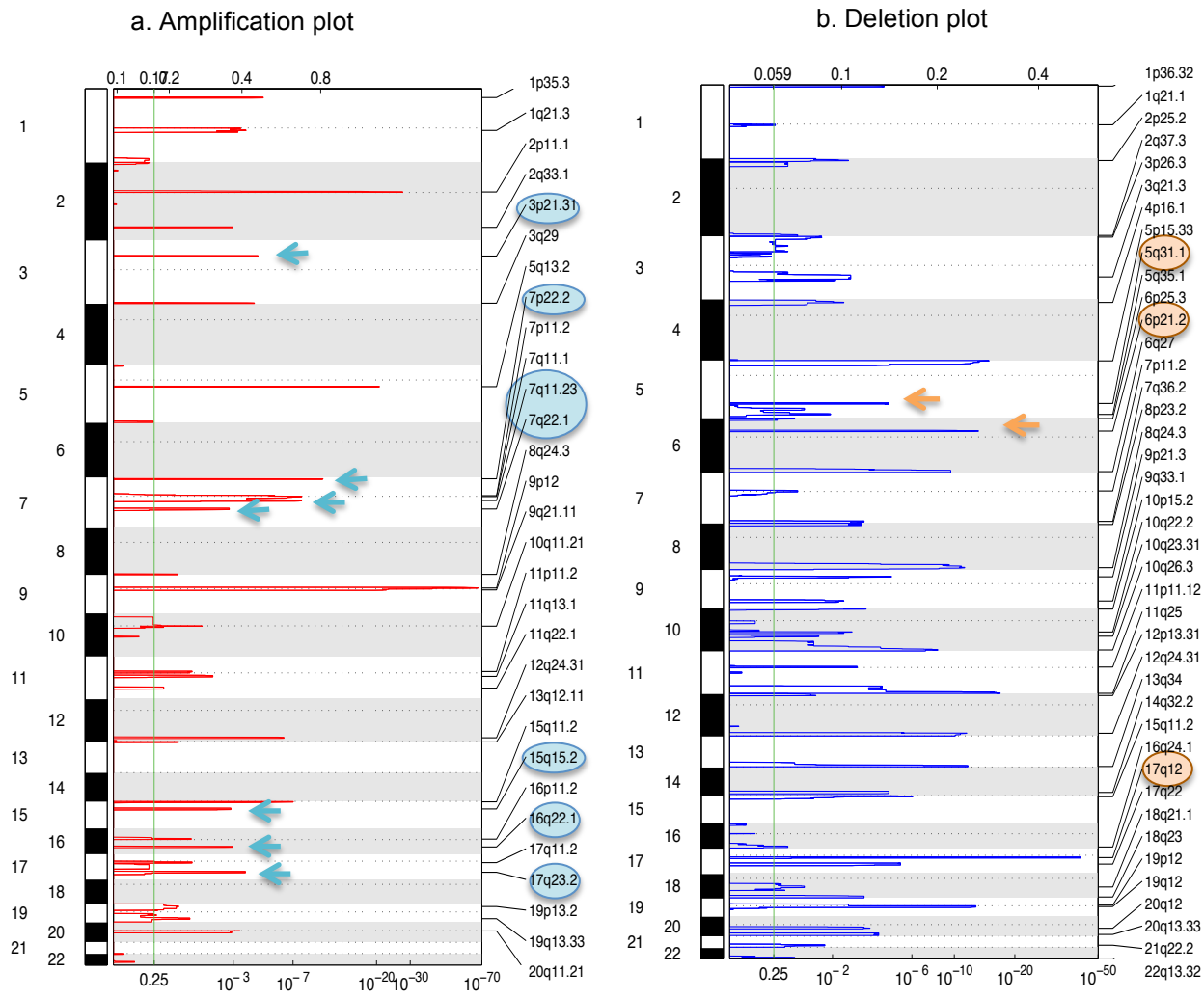


Figure 4.9 Genome-wide amplification and deletion plots of CNAs in OVC.

Genomic positions are indicated along the y axis with centromere locations showed by dotted lines. Amplification (red) and deletion (blue) GISTIC plots show q- values (bottom on the x axis), the G-scores that considers the frequency the aberration occurrence as well as its amplitude across samples (top) and the significance threshold is indicated by the green line at 0.25, with respect to amplifications and deletions for all markers over the entire analysed region. Blue arrows and circles point to regions with significant gain, and orang arrows and circles point to regions with significant loss.

4.2.1.1.3.3 Regions of focal copy number change and genes within copy number–altered regions in OVH

A number of regions of recurrent copy number gain and loss were evident in the GISTIC analysis (figure 4.8). Genomic positions of amplification and deletion peaks (identified in the GISTIC analysis) are listed below (table 4.2) in order to explain the next step, including the list of genes contained in them. Focal event regions selected from the highlighted deletion circles in figure 4.8. The threshold for q-values—the calculated false discovery rate for the abnormal regions—is 0.25; regions with q-values lower than this number were considered significant and genes within those regions will be further investigated in section 4.2.1.1.3.5.

4.2.1.1.3.4 Regions of focal copy number change and genes within copy number–altered regions in OVC

A number of regions of recurrent copy number gain and loss were evident in the GISTIC analysis (figure 4.9). Genomic positions of amplification and deletion peaks (identified in the GISTIC analysis) are listed below (table 4.3) in order to explain the next step, including the list of genes contained in them. Focal event's regions where selected from the highlighted amplification and deletion circles in figure 4.9. The threshold for q-values is 0.25; regions with q-values lower than this number were considered significant and genes within those regions will be further investigated in section 4.2.1.1.3.6.

Table 4.2 Lists of genes located in the most common regions of recurrent DNA copy number change in OVH.

Focal events	Genomic position	Genes mapping within region
Focal deletion (q value: 0.032774)	5q31.1, wide peak boundaries: chr5:133200002-137600000	CAMLG, IL9, LECT2, SMAD5, NEUROG1, NPY6R, PITX1, PPP2CA, SKP1, SPOCK1, TCF7, TGFBI, UBE2B, VDAC1, WNT8A, NME5, CDC23, MYOT, CXCL14, H2AFY, SMAD5-AS1, DDX46, KIF20A, SEC24A, BRD8, HNRNPA0, PHF15, FBXL21, KLHL3, PKD2L2, SAR1B, CDKL3, FAM13B, C5orf15, TRPC7, TXNDC15, PCBD2, CDKN2AIPNL, C5orf24, C5orf20, SLC25A48, LOC340073, LOC340074, CATSPER3, LOC389332, TIFAB, VTRNA2-1, MIR874, MIR3661, MIR4461
Focal deletion (q value: 0.015741)	17q11.2, wide peak boundaries: chr17:30000002-33600000	hsa-mir-632, ACCN1, LIG3, MYO1D, PSMD11, RAD51D, CCL1, CCL2, CCL7, CCL8, CCL11, CCL13, SH3GL1P1, ZNF207, CDK5R1, CCT6B, SUZ12, TMEM98, NLE1, FNDC8, RHOT1, C17orf79, UTP6, C17orf75, ZNF830, LRRC37B, RFFL, TMEM132E, SPACA3, SLC35G3, UNC45B, SLFN5, RHBDL3, C17orf102, ARGFXP2, MIR632, AA06, RAD51L3-RFFL

Table 4.3 Lists of genes located in the most common regions of recurrent DNA copy number change in OVC.

Focal events	Genomic position	Genes mapping within region
Focal amplification (q value: 4.68E-05)	3p21.31, wide peak boundaries: chr3:46400002-50000000	AMT, APEH, RHOA, SLC25A20, CAMP, CDC25A, CCR5, COL7A1, DAG1, CELSR3, GPX1, IMPDH2, LAMB2, LTF, MAP4, MST1, MST1R, MYL3, PFKFB4, PLXNB1, PRKAR2A, PTH1R, QARS, SMARCC1, TCTA, TDGF1, UBA7, USP4, UQCRC1, BSN, CCRL2, IP6K1, RBM6, NME6, TRAIP, ARIH2, CSPG5, USP19, WDR6, TREX1, DHX30, SCAP, LAMB2P1, NBEAL2, KLHL18, NDUFAF3, PTPN23, SETD2, PRSS50, GMPPB, SHISA5, CCDC72, ZNF589, IP6K2, NCKIPSD, P4HTM, C3orf75, QRICH1, DALRD3, RNF123, KIF9, CCDC71, SLC26A6, CAMKV, LRRC2, CCDC51, RTP3, ATRIP, NICN1, MON1A, UCN2, CCDC12, KLHDC8B, ALS2CL, TMIE, FBXW12, FLJ39534, CCDC36, PRSS42, C3orf62, PRSS45, AMIGO3, CDHR4, FAM212A, MIR191, TMEM89, MIR425, SPINK8, C3orf71, LOC646498, CCR2, NRADDP, LOC100132146, BSN-AS2, PRSS46, MIR1226, MIR711, MIR4271, MIR4793, MIR4443
Focal amplification (q	7q11.23, wide peak	CLDN4, CLDN3, ELN, GTF2I, GTF2IP1, HIP1, HSPB1, LIMK1, MDH2, PMS2L2, PMS2P5, PMS2P3, POR

		NCF1C, SNORA14A, MIR590, GATSL2, LOC100093631, POM121C, LOC100133091, MIR4284, MIR4651
Focal amplification (q value: 4.93E-11)	7p22.2, wide peak boundaries: chr7:3600002-7200000	ACTB, PMS2, RAC1, FSCN1, ZNF12, AIMP2, CYTH3, KIAA0415, KDELR2, WIPI2, EIF2AK1, CCZ1, RNF216, ZNF853, ZDHHC4, RADIL, PAPOLB, RBAK, C7orf26, FBXL18, USP42, TNRC18, C7orf70, SDK1, FOXK1, MMD2, DAGLB, CCZ1B, SLC29A4, RSPH10B, LOC389458, GRID2IP, ZNF815, RNF216P1, PMS2CL, ZNF890P, OCM, MIR58, RSPH10B2, LOC100131257, RBAK-LOC389458, MIR4656
Focal amplification (q value: 0.0015106)	7q21.3, wide peak boundaries: chr7:96800002-102800000	ACHE, ASNS, AZGP1, AP1S1, CUX1, CYP3A7, CYP3A4, CYP3A5, EPHB4, EPO, GNB2, AGFG2, LRCH4, MCM7, NPTX2, OCM2, SERPINE1, PCOLCE, PMS2P1, POLR2J, TAC1, TAF6, TFR2, TRIP6, VGF, ZAN, ZNF3, ZKSCAN1, ZSCAN21, TRRAP, BUD31, PLOD3, AP4M1, ATP5J2, MUC12, ARPC1B, RASA4, LRRC17, POP7, ZNHIT1, ARPC1A, SH2B2, STAG3, CPSF4, COPS6, PDAP1, LMTK2, ZKSCAN5, CLDN15, BRI3, TECPR1, PTCD1, FBXO24, PILRB, PILRA, FIS1, ACTL6B, SRRT, ALKBH4, ZCWPW1, C7orf43, BAIAP2L1, MEPCE, SLC12A9, ACN9, SMURF1, MOSPD3, GIGYF1, RABL5, CYP3A43, ZNF655, PVRIG, GAL3ST4, PRKRIP1, ORAI2, OR2AE1, TSC22D4, TRIM56, ARMC10, ZNF394, MYH16, TRIM4, MYL10, EMID2, MUC17, BHLHA15, ZNF498, FAM200A, PPP1R35, GPC2, LRWD1, FAM185A, FBXL13, NAPEPLD, TMEM130, NYAP1, CNPY4, POLR2J2, MBLAC1, ZNF789, MOGAT3, GJC3, GATS, NAT16, MGC72080, C7orf59, KPNA7, C7orf61, UFSP1, MIR106B, MIR25, MIR93, SPDYE3, SPDYE2, POLR2J3, AZGP1P1, SPDYE6, RPL19P12, LOC100129845, UPK3BL, LOC100289187, LOC100289561, SPDYE2L, SAP25, MIR4285, MIR3609, ATP5J2-PTCD1, MIR4653, MIR4467, MIR4658, LOC100630923, CYP3A7-CYP3AP1
Focal amplification (q value: 0.0012753)	15q15.1, wide peak boundaries: chr15:40000002-45600000	B2M, BUB1B, CAPN3, CKMT1B, EPB42, GANC, GCHFR, PDIA3, ITPKA, IVD, LTK, MAP1A, MFAP1, PLCB2, RAD51, SORD, SPINT1, SRP14, TP53BP1, TYRO3, EIF3J, JMJD7-PLA2G4B, SNAP23, SLC28A2, TGM5, PPIP5K1, LCMT2, SERF2, GPR176, CHP, OIP5, BAHD1, MAPKBP1, RTF1, MGA, VPS39, CCNDBP1, C15orf63, TMEM87A, RPAP1, RPUSD2, TUBGCP4, EHD4, DUOX2, NDUFAF1, NUSAP1, SPTBN5, CTDSPL2, DUOX1, DLL4, INO80, PPP1R14D, HAUS2, FAM82A2, DNAJC17, PAK6, CASC5, STARD9, VPS18, ZFP106, CHAC1, WDR76, TMEM62, SPG11, ELL3, ZFYVE19, FRMD5, DISP2, C15orf57, C15orf23, BMF, SHF, DUOXA1, CHST14, CASC4, TGM7, CATSPER2, PLA2G4E, TRIM69, C15orf43, ZSCAN29, TTBK2, CDAN1, STRC, ADAL, EXD1, FSIP1, RHOV, UBR1, PATL2, PLA2G4F, LRRC57, PLA2G4D, MRPL42P5, C15orf52, DUOXA2, EIF2AK4, CATSPER2P1, CKMT1A, SERINC4, C15orf62, C15orf56, PHGR1, LOC645212, MIR626, MIR627, LOC728758, OIP5-AS1, LOC100131089, ANKRD63, JMJD7, PLA2G4B, MIR1282, MIR4310, LOC100505648, SERF2-C15ORF63
Focal amplification (q value: 0.0010611)	16q22.1, wide peak boundaries: chr16:66000002-70400000	AARS, AGRP, CA7, CBFB, CDH1, CDH3, CDH5, CDH16, CTRL, NQO1, DYNC1LI2, E2F4, HAS3, HSD11B2, HSF4, LCAT, NFATC3, PSKH1, PSMB10, RRAD, SLC9A5, SLC12A4, SNTB2, TERF2, TK2, TRADD, CES2, NAE1, NOL3, SLC7A6, ATP6V0D1, NUTF2, CTCF, NFAT5, WWP2, DDX19B, CES3, EDC4, PLA2G15, PLEKHG4, LRRC29, VPS4A, NOB1, TMEM208, FHOD1, ZDHHC1, PARD6A, CKLF, NIP7, FAM96B, TPPP3, PRMT7, DUS2L, CHTF8, PDPR, LRRC36, DDX19A, FBXL8, SMPD3, DDX28, TSNAXIP1, THAP11, PDP2, RANBP10, PDF, DPEP2, DPEP3, ACD, FAM65A, TMCO7, ELMO3, ESRP2, CENPT, C16orf70, CYB5B,

		GFOD2, C16orf48, SLC7A6OS, COG8, B3GNT9, CIRH1A, CMTM1, EXOSC6, NRN1L, CMTM3, ZFP90, RLTPR, KCTD19, CMTM4, CMTM2, BEAN1, TMED6, CCDC79, CES4A, EXOC3L1, PDXDC2P, CLEC18C, CLEC18A, C16orf86, MIR140, MIR328, KIAA0895L, LOC729513, MIR1538, MIR1972-1, MIR1972-2, LOC100505865, LOC100506083, CKLF-CMTM1
Focal amplification (q value: 0.00024633)	17q22, wide peak boundaries: chr17:56000002-61200000	CA4, CLTC, LPO, MPO, TRIM37, SEP4, RAD51C, RPS6KB1, SRSF1, SUPT4H1, TBX2, VEZF1, EPX, PPM1D, MTMR4, BZRAP1, TBX4, MRC2, MED13, APPBP2, TLK2, PPM1E, TANC2, OR4D1, RNFT1, TUBD1, PTRH2, BCAS3, RNF43, MKS1, SMG8, MSX2P1, PRR11, TEX14, INTS2, HEATR6, DHX40, VMP1, BRIP1, USP32, HSF5, OR4D2, C17orf64, DYNLL2, EFCAB3, MARCH10, C17orf47, GDPD1, METTL2A, NACA2, SKA2, YPEL2, C17orf82, MIR142, MIR21, MIR301A, TBC1D3P2, LOC645638, TBC1D3P1-DHX40P1, LOC653653, SCARNA20, MIR454, MIR548W, LOC100506779, MIR4729, MIR4737, MIR4736
Focal deletion (q value: 3.44E-05)	5q31.1, wide peak boundaries: chr5:133200002-137600000	CAMLG, IL9, LECT2, SMAD5, NEUROG1, NPY6R, PITX1, PPP2CA, SKP1, SPOCK1, TCF7, TGFBI, UBE2B, VDAC1, WNT8A, NME5, CDC23, MYOT, CXCL14, H2AFY, SMAD5-AS1, DDX46, KIF20A, SEC24A, BRD8, HNRNPA0, PHF15, FBXL21, KLHL3, PKD2L2, SAR1B, CDKL3, FAM13B, C5orf15, TRPC7, TXNDC15, PCBD2, CDKN2AIPNL, C5orf24, C5orf20, SLC25A48, LOC340073, LOC340074, CATSPER3, LOC389332, TIFAB, VTRNA2-1, MIR874, MIR3661, MIR4461
Focal deletion (q value: 5.70E-14)	6p21.2, wide peak boundaries: chr6:38400002-41200000	DNAH8, GLO1, GLP1R, MOCS1, NFYA, KCNK5, APOBEC2, DAAM2, TREM2, SAYSD1, LRFN2, TREML2, KCNK16, KCNK17, LOC221442, C6orf130, KIF6, TSPO2, UNC5CL, TREML1, TREML3, FLJ41649, TDRG1, LOC100131047
Focal deletion (q value: 0.015741)	17q11.2, wide peak boundaries: chr17:30000002-33600000	hsa-mir-632, ACCN1, LIG3, MYO1D, PSMD11, RAD51D, CCL1, CCL2, CCL7, CCL8, CCL11, CCL13, SH3GL1P1, ZNF207, CDK5R1, CCT6B, SUZ12, TMEM98, NLE1, FNDC8, RHOT1, C17orf79, UTP6, C17orf75, ZNF830, LRRC37B, RFFL, TMEM132E, SPACA3, SLC35G3, UNC45B, SLFN5, RHBDL3, C17orf102, ARGFXP2, MIR632, AA06, RAD51L3-RFFL
Focal deletion (q value: 6.26E-06)	17q21.33, wide peak boundaries: chr17:48400002-55200000	CHAD, COX11, HLF, NME1, NME2, TRIM25, COIL, AKAP1, DGKE, ABCC3, CACNA1G, SPAG9, NOG, TOM1L1, TOB1, MMD, UTP18, MRPL27, LUC7L3, MBTD1, LINC00483, EPN3, TMEM100, RSAD1, LRRC59, CA10, PCTP, SCPEP1, XYLT2, SPATA20, ACSF2, MYCBPAP, KIF2B, ANKRD40, WFIKKN2, EME1, ANKFN1, MTVR2, STXBP4, LOC253962, C17orf67, RNF126P1, LOC400604, NME1-NME2, MIR3614

4.2.1.1.3.5 Assessment of the list of genes with CNAs in OVH

The GISTIC method was used to identify the most significant amplifications and deletions as described previously. Two peaks were identified, and these regions had a large list of genes (Table 4.2). The results were then further analysed by running the gene lists against 13 KEGG pathways, which are more related to head and neck cancers, as well as cancer gene census and Stransky mutation list (76 previously identified genes in HNSCCs harbouring high statistically significant mutations) (Stransky *et al.*, 2011) (refer to chapter 2, section 2.8.3.1.6).

Out of eight key genes hits (refer to figure 4.10), four genes were involved in KEGG WNT signalling pathway (36% of the CN altered genes in OVH cohort were involved in this pathway). In addition, two genes were involved in KEGG cell cycle pathway (18% of the CN altered genes in OVH cohort were involved in this pathway), and one cancer gene was located as well among the eleven genes list (*SUZ12*). Table 4.4 lists all key genes founded to be in CN altered regions with the highest significance losses in OVH cohort. OVH illustrates very low level of CNAs, and as a result, low number of genes were identified.

Many WNTs are frequently overexpressed in HNCs (Barker and Clevers, 2007). However, in the OVH cohort, genes involved in WNT signalling pathway were in CN loss regions. The *SUZ12* gene is located at chromosome 17q11.2, which has been deleted at a frequency of ~15% in OVH deletion plot. The role of *SUZ12* has been investigated previously in epithelial ovarian cancer and revealed high significant expression levels when compared with either, fallopian tube epithelium or normal ovarian surface epithelium, as it inhibits apoptosis by stimulating the proliferation of epithelial ovarian cancer cells (Li *et al.*, 2012). However, no reports were found to illustrate the role of *SUZ12* when down regulated. Additionally, candidate cancer genes at regions of low CN loss levels might function differently than when at high CN loss levels, and this gene was located in a loss region at a frequency of ~15% only. Nonetheless, it must be remembered that the suggested cancer gene here, at best, can only report on the “likelihood” of progression of OVH and requires further investigations.

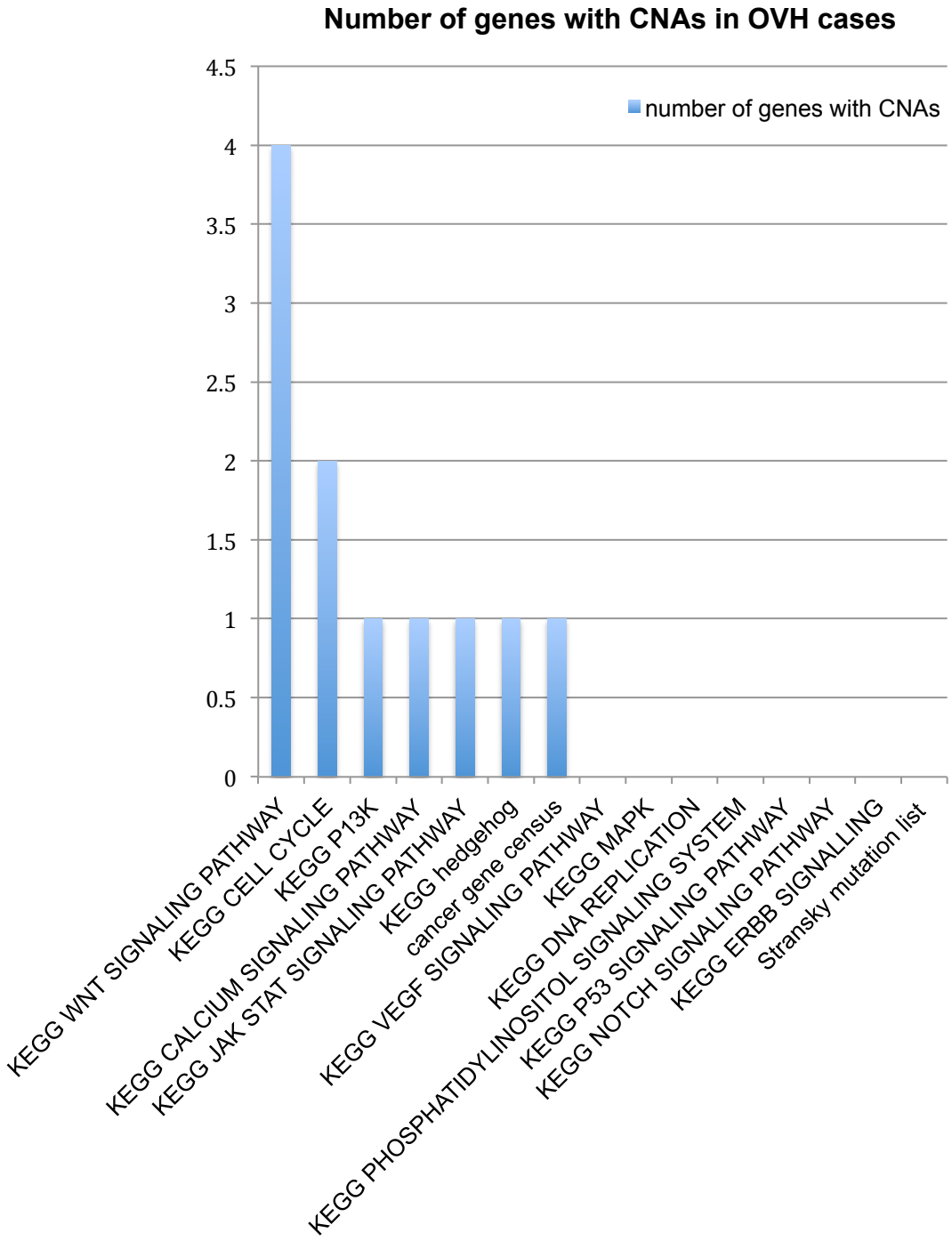


Figure 4.10 Number of genes with CNAs in OVH cases

A graphical representation of: 13 KEGG pathways, cancer gene census and Stransky mutation list (Stransky *et al.*, 2011) on the x-axis ranked by the number of genes with CNAs from OVH samples in each pathway and list on the y-axis.

4.2.1.1.3.6 Assessment of the list of genes with CNAs in OVC

The GISTIC method was used to identify the most significant amplifications and deletions as described previously. Twelve peaks were identified, and these regions had a large gene lists (Table 4.2). The results were then further analysed by running the gene lists against 13 KEGG pathways, which are more related to head and neck cancers, as well as cancer gene census and Stransky mutation list (Stransky *et al.*, 2011) (refer to chapter 2, section 2.8.3.1.6).

Out of 49 key genes hits (refer to figure 4.11), thirteen genes were cancer genes (17% of the CN altered genes in OVC cohort were found to be related with cancer). Furthermore, eleven genes were involved in KEGG WNT signalling pathway (15% of the CN altered genes were involved in this pathway), and eight genes in P13K pathway (10% of the CN altered genes were involved in this pathway). Seven genes were in KEGG cell cycle pathway, and seven genes as well in KEGG VEGF signalling pathway (9% of the CN altered genes are in these pathways). Similarly, six genes were in the KEGG MAPK pathway, and six genes involved KEGG calcium-signalling pathway (8% of the CN altered genes are in these pathways). Table 4.4 lists all key genes founded to be in CN altered regions with the highest significance gains and losses in OVC cohort.

As can be seen from table 4.4, all significant gene hits in OVH group were present in the OVC significant gene hit lists, which suggests that OVH is a (histological) precursor for OVCs. In the analysis of OVC gene hit lists, I focused on genes that had a known role in head and neck cancers. *CDH1* or epithelial -cadherin gene, located on chromosome 16q22.1, was in a gain chromosomal region and therefore, is likely to be overexpressed in the OVC cohort. The function of E-cadherins has been well-established in maintaining junctions, E-cadherin loss enables the disaggregation of malignant cells from one another and promotes metastasis (Berx *et al.*, 1998), (Onder *et al.*, 2008). In human cancers, reduction or loss of E-cadherin expression can be triggered by silencing of the CDH1 promoter, chromosomal deletions and somatic mutations (Berx *et al.*, 1998), (Onder *et al.*, 2008). However, and in light of the

possibility of overexpression of *CDH1* in the OVC group, I therefore suggest that this might be a reason behind the fact that OVCs do not metastasise, unlike OSCCs, where the CN data of this cohort showed deletion in chromosome 18q21.3 that harbour *CDH20* gene, which has been reported previously to be involved in tumour invasion regulation (Vermeulen *et al.*, 1996).

In addition, the *MCM7* gene located in chromosome 7q22.1, was in a chromosomal region showing gain at a frequency of ~50% according to the OVC frequency karyogram (section 4.2.1.1.2.3) and therefore, is likely to be overexpressed in the OVC cohort. It has been demonstrated in a previous study that *MCM7* gene is expressed in normal oral mucosa and variably overexpressed in dysplasias and OSCCs (Tamura *et al.*, 2010). Likewise, the *SERPINE1* gene also located on the chromosome 7q22.1 arm that presented a gain is likely to be overexpressed in OVC cohort. In 2005, a study revealed that the expression of the *SERPINE1* gene in primary head and neck tumours was up-regulated in comparison to normal mucosa by an expression ratio of: 6.22, using microarray (Chin *et al.*, 2005). *SERPINE1* overexpression was suggested to play a key role in chromosome 7q21.3 – 22 karyotypic changes and in oral oncogenesis (Chen *et al.*, 2004b).

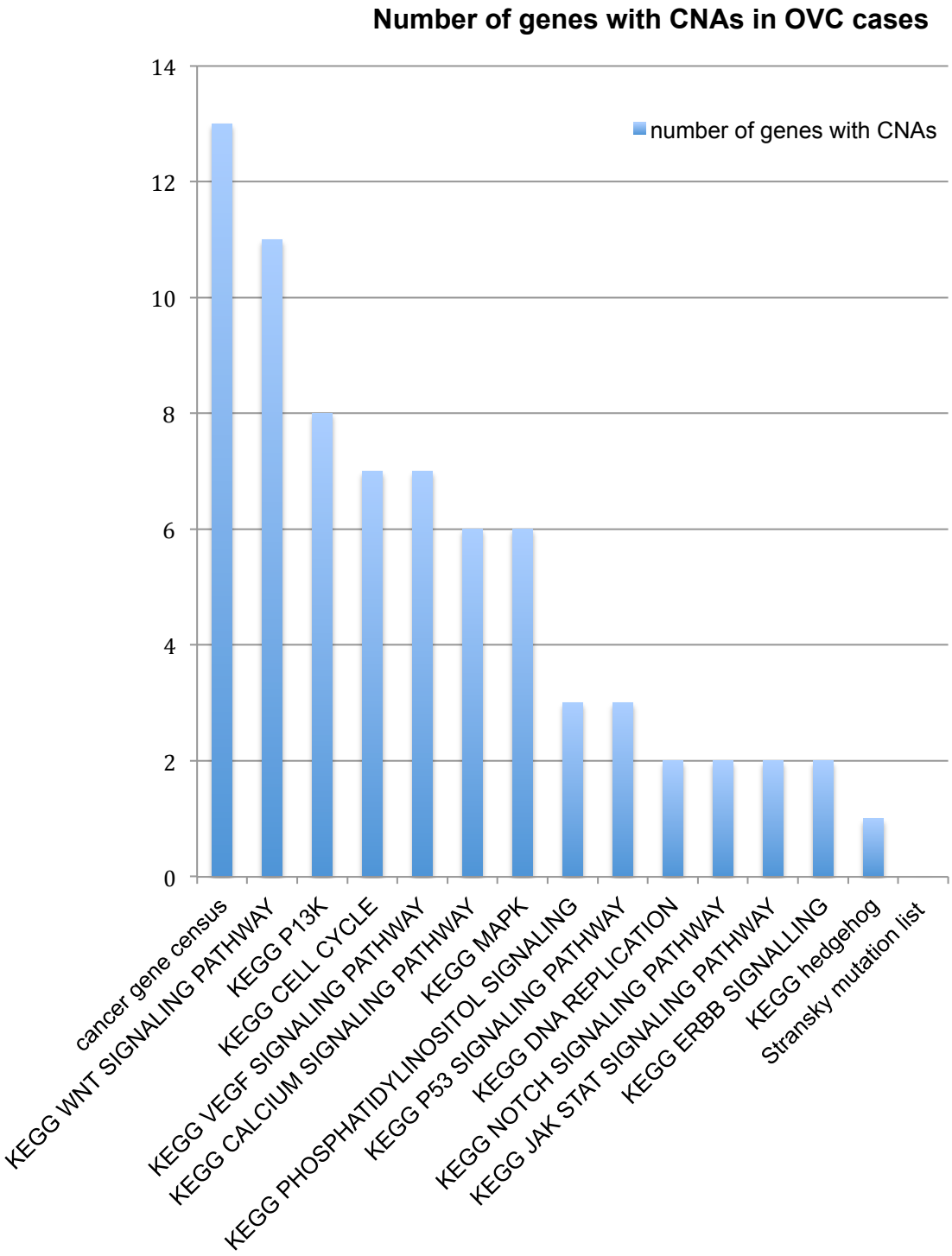


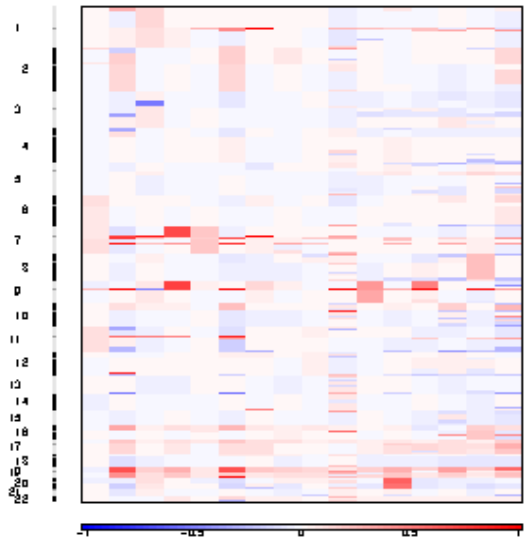
Figure 4.11 Number of genes with CNAs in OVC cases

Figure 0.11 A graphical representation of: 13 KEGG pathways, cancer gene census and Stransky mutation list (Stransky *et al.*, 2011) on the x-axis ranked by the number of genes with CNAs from OVC samples in each pathway and list on the y-axis.

4.2.1.1.3.7 Comparison of DNA copy-number profiles (GISTIC G-scores heat maps) between OVH and OVC samples.

Chromosomal alteration regions based on DNA CN changes in OVH and OVC groups are illustrated in the heat maps below generated from GISTIC G-scores analysis (figure 4.12). Visual examination of the OVH heat map (figure 4.12 a) and OVC's (figure 4.12 b) illustrates the very low level of CNAs in OVHs in compare to OVCs, indicating that the genomic profile of these cases has minimal chromosomal abnormalities and is most similar to normal. Nevertheless, gain at chromosome 7q (represented by a lineage red colour) was noticed in OVHs at a frequency of more than 50%. In addition, as shown in figure 4.12 b, gains at chromosome arms 7q, 16q and 17q (represented by a lineage red colour), and loss at chromosome 5p (represented by a lineage blue colour) were detected in OVCs at a frequency of 50%. Again, and as seen in the heat map below; deletions were minimally found in OVC, which could be a reason behind the minimal histological cytological atypia features found in this tumour.

a. OVH



b. OVC

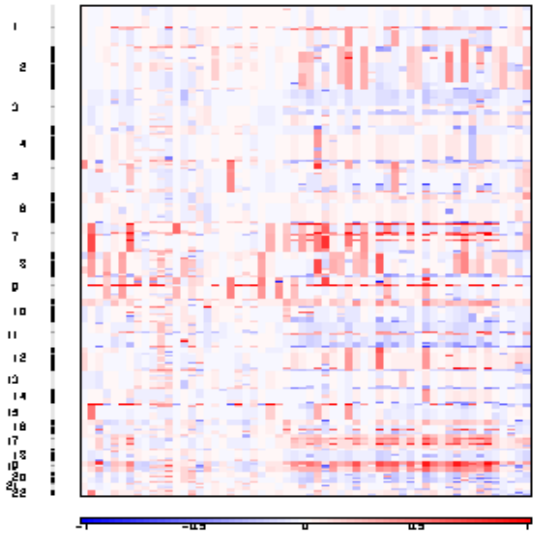


Figure 4.12 Genome-wide amplification and deletion plots of CNAs in OVH.

Genomic positions are indicated along the y axis with centromere locations showed by dotted lines. Amplification (red) and deletion (blue) GISTIC plots show q- values (bottom on the x axis), the G-scores that considers the frequency the aberration occurrence as well as its amplitude across samples (top) and the significance threshold is indicated by the green line at 0.25, with respect to amplifications and deletions for all markers over the entire analysed region. Blue arrows and circles point to regions with significant gain, and orang arrows and circles point to regions with significant loss

4.2.2 Aim 2: To use next generation sequencing copy number analysis data to distinguish between the genomic damage pattern in OVC and OSCC lesions.

Copy number sequencing data were generated for 45 OSCC samples (34 males and 11 females, mean age 61.4) as a part of another study conducted by Pre-cancer Genomic Group in this institute to investigate the copy number changes and clonal relationships that occur between the stages of normal epithelium, dysplasia and OSCC. Catherine Daly, Rebecca Chalkley, Rajni Bhardwaj and Henry Wood did the entire work on OSCCs, starting from sectioning, macro-dissection, DNA extraction, DNA quantification and DNA CN sequencing library preparation steps. I used OSCC CN analysis data and karyograms from the Pre-cancer Genomic Group to compare the genomic damage pattern that occur in them with the genomic damage pattern that occur in OVH and OVC. The designed study groups (including samples IDs) are shown in appendix 4.4.

4.2.2.1 Genomic profiling of OSCC by NGS CN analysis

In general, OSCC karyograms shows more whole chromosome and localised gain and loss and a higher degree of aneuploidy across the whole genome when compared to OVCs. (Refer to appendix 4.5 for eight examples of OSCC karyograms). Loss of chromosome 3p and gain at 3q were the copy number variations most frequently detected to be different between oral verrucous tumours and oral squamous tumours. No losses at chromosome 3p or gains at 3q were identified in OVCs. Losses mapped on 3p, 4q, 9p, and 18q, and gains mapped on chromosomal arms 3q, 4q, 5q, 8q, 9q and 20p are chromosomal signatures commonly linked with OSCC and frequently identified in OSCC cohort here. A representative classical OSCC CN karyogram from case PG123-T-3_2KB is shown below in figure 4.9 a to illustrate the differences between the three cohorts individual karyograms (data belong to the Pre-cancer Genomic Group). Blue arrows point at gains in Chr3q, Chr4p, Chr5p, Chr8q. Chr9q and Chr18p. Orange arrows point at losses in Chr3p, Chr4q, Chr8p, Chr9p and Chr18q. Horizontal lines above the center demonstrate regions of gain, and those below the center demonstrate regions of loss.

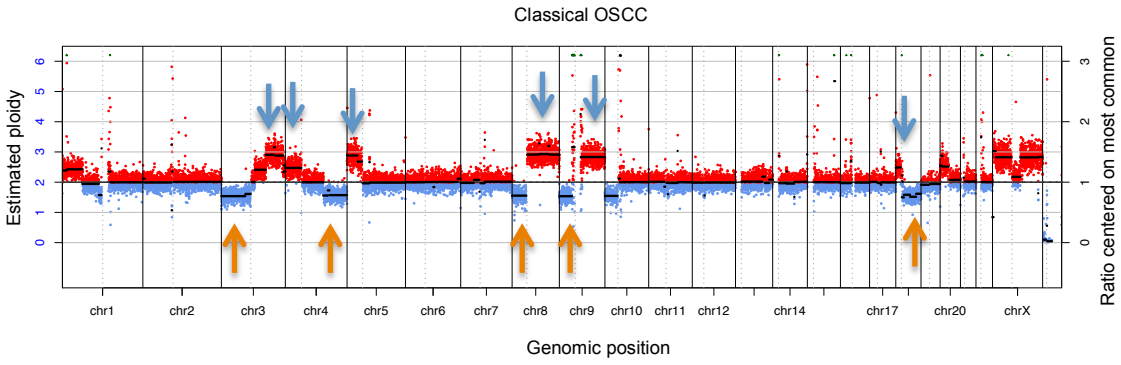


Figure 4.13 Representative genomic profiling of OSCC by NGS CN analysis.

Each data point represents one window of approximately 200 reads. Genomic position is on the x-axis and tumor:normal ratio is on the y-axis. The black lines are regions of common CN between breakpoints. Windows of gain and loss are red and blue respectively.

4.2.2.2 Results

4.2.2.2.1 Comparison of the genome-wide frequency karyograms of CNAs in OVC and OSCC

In order to compare OVC and OSCC as groups; frequency accumulative karyograms were produced for the two cohorts. Visual examination of OVC (N: 57) genomic CN frequency karyogram (figure 4.14 a) revealed a lower level of CNA compared to OSCCs (N: 45) (figure 4.14 b). This result suggests that OVC is characterised by a lower degree of chromosomal instability than OSCC. Similar results were obtained in 2011 by another study in which they have investigated differences in chromosomal instability between OVC and OSCC lesions using high-resolution DNA flow cytometry (Pentenero *et al.*, 2011). In this study, they reported a lower degree of tumour heterogeneity and chromosomal instability in OVCs when compared to OSCCs. Locus-specific differences in CN can also be seen by comparing chromosome plots frequency diagrams of the two groups (figure 4.7, figure 4.15). In addition, Loss of chromosome 3p and gain at 3q were the CN variations most frequently detected to be different between oral verrucous tumours and OSCCs. No losses at chromosome 3p or gains at 3q were identified in OVCs. Furthermore, gains mapped at chromosome 7p22, 7q11.2 and 7q22 (represented by red colour) were observed in OVCs at a frequency of ~50%, in addition to gains mapped at chromosomes 3p21 (at a frequency of ~30%), 15q15 (at a frequency of ~30%), 16q22 (at a frequency of ~25%) and 17q23 (at a frequency of ~25%), as well as losses on chromosomes 6p21 (at a frequency of ~25%) and 17q12 (at a frequency of ~50%) represented by green colour. Nevertheless, deletion trends were minimally found in OVC's frequency karyogram.

On the other hand, as can be seen in figure 4.14 b, and figure 4.15, Loss of chromosome 3p and gain at 3q (at a frequency of 50%) were the CN variation most frequently distinguishing OVC and OSCC (primary OSCC data belong to Pre-cancer Genomics Group). Gains mapped at chromosomes 3q (at a frequency of 50%), 5p (at a frequency of ~50%), 7p (at a frequency of ~25%), 8q (at a frequency of ~40%) and 20p (at a frequency of ~20%) represented by red colour, and losses mapped at chromosomes 3p (at a frequency of 50%), 4p

(at a frequency of ~25%), 8p23 (at a frequency of ~25%), 18q21 (at a frequency of ~30%) and 18q23 (at a frequency of ~30%) represented by green colour were detected in OSCCs. These chromosomal abnormalities and CN alterations may be involved in the development of OSCC. As has been observed, losses were detected more frequently in OSCCs than OVCs, suggesting that these CN alterations may be related to the aggressiveness behaviour of OSCC cells. Losses on chromosomal arms 3p, 4q and 18q, and gains on chromosomal arms 3q, 5q, 8q, and 20p are chromosomal signatures commonly linked with OSCCs.

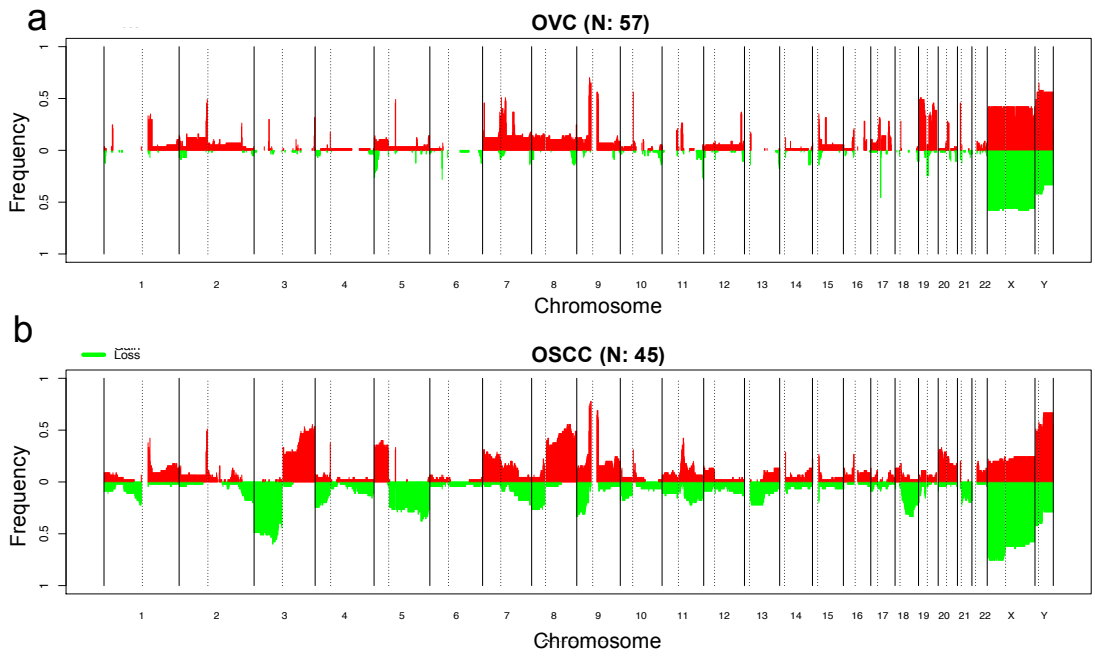
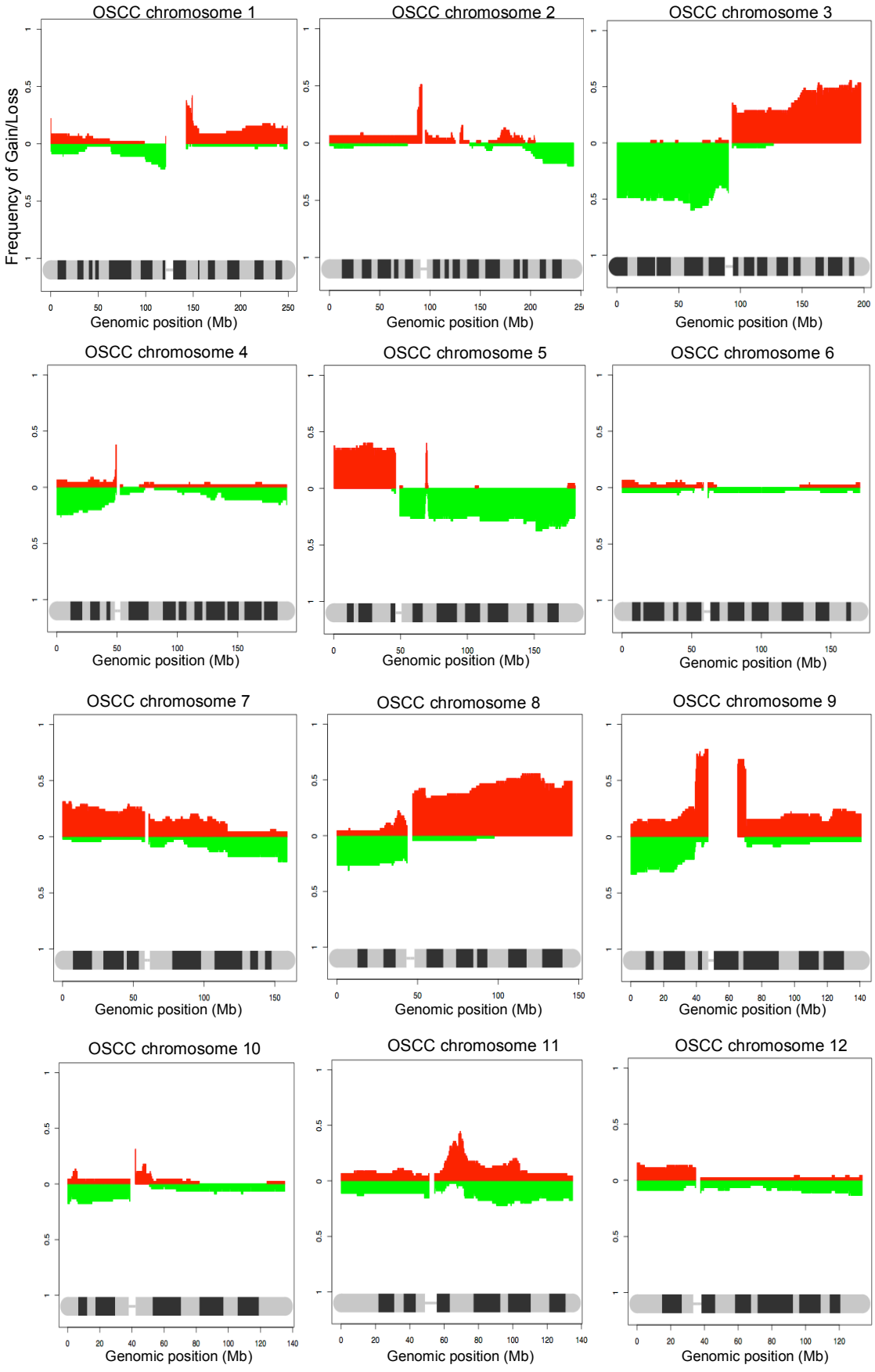


Figure 4.14 Frequency of genomic gain and loss for OVC (a) and OSCC (b).

Genomic position is on the x-axis, frequency (%) of gains (red) and losses (green) are shown on the y-axis.


OSCC chromosome 1
OSCC chromosome 2
OSCC chromosome 3
OSCC chromosome 4
OSCC chromosome 5
OSCC chromosome 6
OSCC chromosome 7
OSCC chromosome 8
OSCC chromosome 9
OSCC chromosome 10
OSCC chromosome 11
OSCC chromosome 12
Frequency of Gain/Loss
Genomic position (Mb)

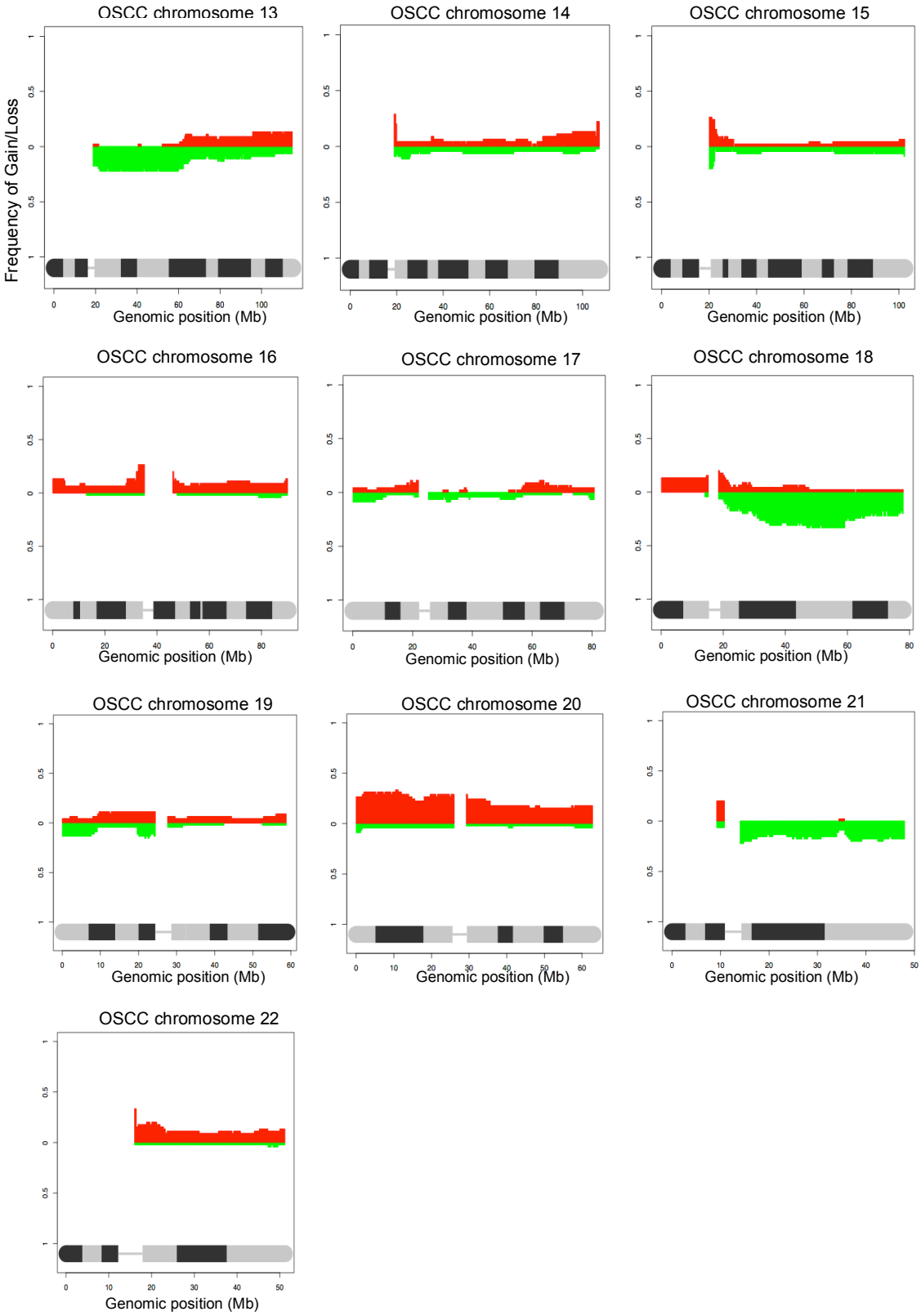


Figure 4.15 Genomic locations of chromosomal segments with altered CN in OSCC.

Genomic position is on the x-axis, frequency (%) of gains (red) and losses (green) are shown on the y-axis

4.2.2.2.2 Genomic Identification of Significant Targets in Cancer (GISTIC2.0) (computational approach)

Copy number profiles of OSCC were characterised by several approaches using the GISTIC2.0 algorithm in order to compare this cohort with the OVC cohort. These analysis forms include: amplification and deletion plots of CNAs, the identification of amplification and deletion genes within CN altered regions, and segmented copy number heat maps. All the parameters used in the GISTIC analysis are presented in appendix 2.3. (Primary data belong to the Pre-cancer Genomic Group).

4.2.2.2.2.1 Genome-wide amplification and deletion plots of CNAs in OSCC

Regions of significant gains or losses were identified using the GISTIC algorithm. Sixteen chromosomal regions (thirteen amplifications and three deletions) from the CNAs identified by GISTIC analysis were significantly altered in OSCC genomes with CNAs ranging from single copy gains and losses of broad chromosomal regions **(**Figure 4.16 a, and b). The three most significant amplifications from GISTIC peaks (Figure 4.16 a, blue circles) that also surpass the significance threshold include chromosomes 3q26.1, 3q28, 7q11.23, and 11q22.1.The thirteen most significant deletions from GISTIC peaks (Figure 4.16 b, orange circles) that also surpass the significance threshold include chromosomes 2q37.3, 3p26.2, 3p14.2, 4p16.3, 5q35.1, 7q36.2, 8p23.1, 9p24.3, 11q25, 13q12.13, 18q12.3, 18q21.2 and 18q23.

Previous CN studies of chromosomal signatures commonly linked with OSCC revealed losses mapped on 3p, 4q, 9p, and 18q, and gains mapped on chromosomal arms 3q, 4q, 5q, 8q, 9q and 20p (Baldwin *et al.*, 2005), (Snijders *et al.*, 2005), (Liu *et al.*, 2006), (Jarvinen *et al.*, 2008), (Nakamura *et al.*, 2008), (Freier *et al.*, 2010). These findings were frequently identified as well in OSCC cohort here, suggesting that putative tumour suppressor genes and oncogenes on those arms are associated with OSCC and may be involved in the development of this tumour.

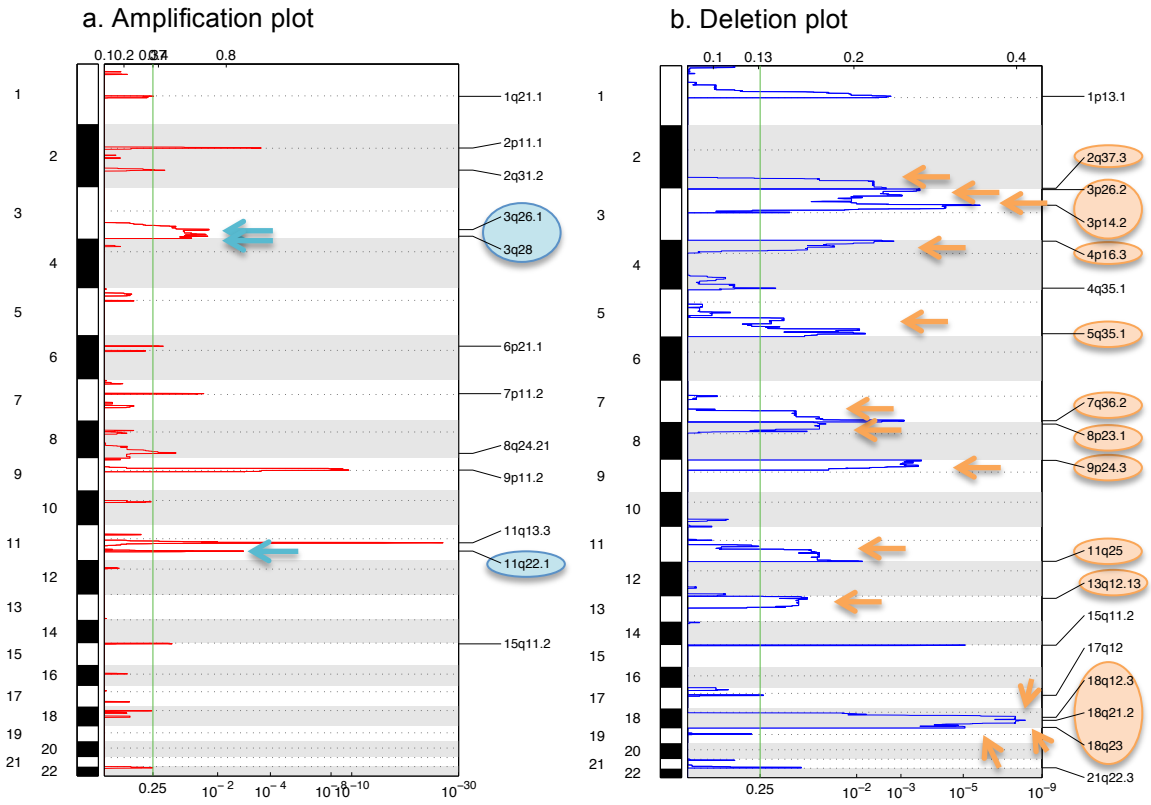


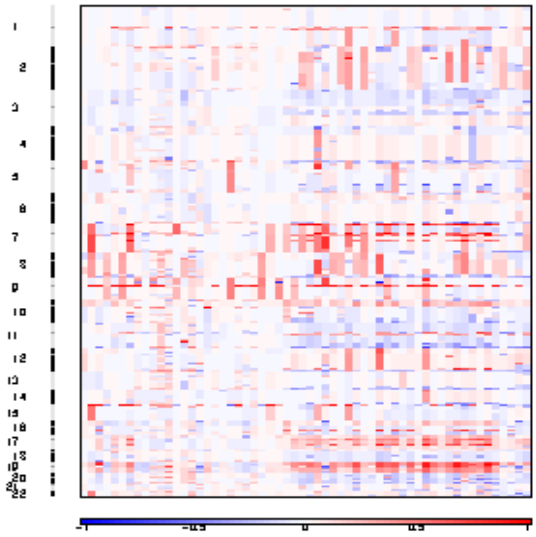
Figure 4.16 Genome-wide amplification and deletion plots of CNAs in OSCC.

Genomic positions are indicated along the y axis with centromere locations showed by dotted lines. Amplification (red) and deletion (blue) GISTIC plots show q- values (bottom on the x axis), the G-scores that considers the frequency the aberration occurrence as well as its amplitude across samples (top) and the significance threshold is indicated by the green line at 0.25, with respect to amplifications and deletions for all markers over the entire analysed region. Blue arrows and circles point to regions with significant gain, and orang arrows and circles point to regions with significant loss.

4.2.2.2.2.2 Comparison of DNA copy-number profiles (GISTIC G-scores heat maps) between OVC and OSCC samples.

Chromosomal alteration regions based on DNA CN changes in OVC and OSCC groups are illustrated in the heat maps below generated from GISTIC G-scores analysis (figure 4.17) (primary data belong to the Pre-cancer Genomic Group). Visual examination of OVC heat map and OSCC's (figure 4.17 a and b) as it is clearly show the lower degree of CN alterations in OVC cases as compared to OSCCs. In addition, as shown in figure 4.17 a, gains at chromosome arms 7q, 16q and 17q (represented by a lineage red colour), and loss at chromosome 5p (represented by a lineage blue colour) were detected in OVCs at a frequency of 50%. Moreover, deletion trends were minimally found in OVC's heat map. On the other hand, as can be seen in figure 4.17 b, gains at chromosome arms 3q, 5p and 8q (represented by a lineage red colour), and losses at chromosome arms 3p, 4p, 5q, 8p, 13q and 18q (represented by a lineage blue colour) were detected in OSCCs at a frequency of more than 50%. Remarkably, losses were detected more frequently in OSCC cohort than OVCs, suggesting that these CN alterations may be related to the aggressiveness behaviour of OSCC cells. In addition, the minimal or absent histological cytological atypia feature of OVC (Zargaran *et al.*, 2012) could be another reason behind the lower level of chromosomal instability in OVCs in compare to OSCCs.

a. OVC



b. OSCC

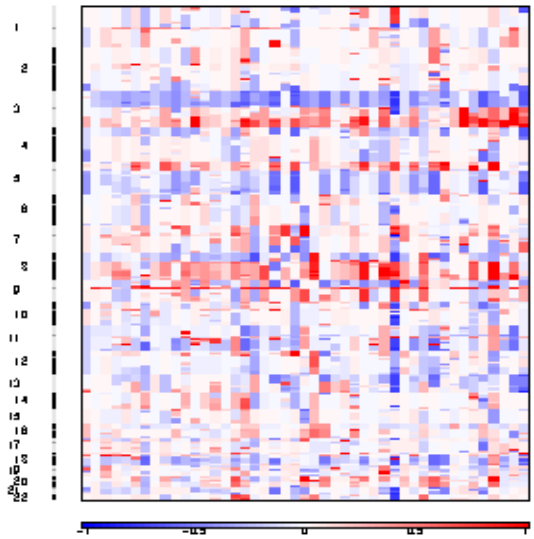


Figure 4.17 Heat map images of OVC and OSCC based on total segmented DNA copy number variation profiles.

Images were analysed using (GISTIC2.0). In each heat map, the samples are arranged from left to right, and chromosomes are arranged vertically from top to bottom. Red represents CN gain and blue represents CN loss.

4.2.2.2.3 OVC, OVH and OSCC hierarchical classification based on the genomic CNAs

Chromosomal region alterations in OVC, OVH, and OSCC were analysed using NGS CN analysis. A complete linkage unsupervised hierarchical clustering was performed for all the three groups based on DNA CN changes (figure 4.18) (primary OSCC data belong to the Pre-cancer Genomics Group). Samples from oral verrucous lesions group (OVC and OVH) tended to be located relatively close together in the dendrogram below (figure 4.18, clusters b, d and e). However, a few OSCC samples were included within these group clusters. This can be attributed to the similarities in some CN features, as in centromere and in telomere chromosomal regions that arise exactly in the same place in all groups' samples (centromere and telomere chromosomal regions were excluded from visual examination of the karyograms but not from the BED files (chromosomal break points files) of every case). In contrast, samples from the OSCC group tended to be located fairly close together in the dendrogram below (figure 4.18, clusters a and c). Nevertheless, few verrucous samples were included within these clusters. This can be attributed to the similarities in some CN features, as in centromere in telomere chromosomal regions that arise exactly in the same place in all groups' samples. In general, this methodology was able to segregate oral verrucous samples from OSCCs using the entire pattern of CN changes, including the shared ones. This finding confirms the previous result based on individual karyograms for the three cohorts, frequency karyograms and heat maps data, and prove that oral verrucous lesions are distinct entity according to its clinical appearance, histological features, and its genomic profiles and behaviour mode.

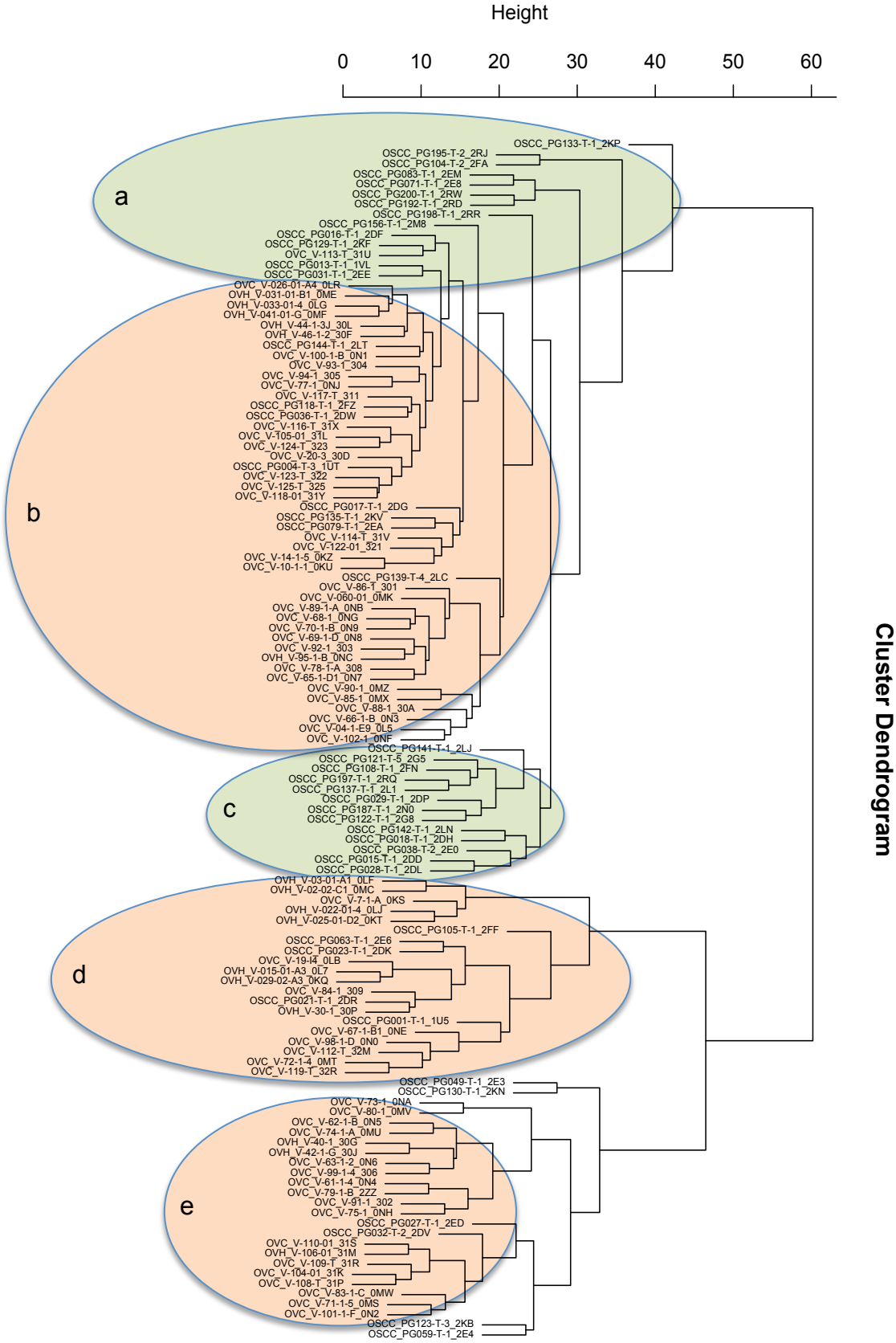


Figure 4.18 Hierarchical cluster dendrogram of the CNAs data from OVH (N: 16), OVC (N: 57) and OSCC (N: 45) tissues by unsupervised clustering.

The scale on the top bar indicates Manhattan distance.

4.2.2.2.4 Logistic regression (LR) analysis

A novel LR method that blindly predicts the subtype of an 'unknown' sample based on the copy number of a two groups of 'known' samples was applied on OVC and OSCC cohorts. Of the 45 OSCC samples, four were misclassified, whereas only one out of 57 OVC samples was incorrectly predicted. The logistic regression analysis successfully distinguished between the two patient groups. (Gusnanto, Wood *et al.* accepted pending corrections, Bioinformatics). Figure 4.19 below show the percentage of the error for OVC and OSCC using two analysis sets. In the "50" analysis (a), The samples were divided into 50% training set and 50% test, while for the "65" analysis (b), 65% of the samples went into the training set. The plots below illustrated that the more samples included in the training set (b), the less was the percentage error.

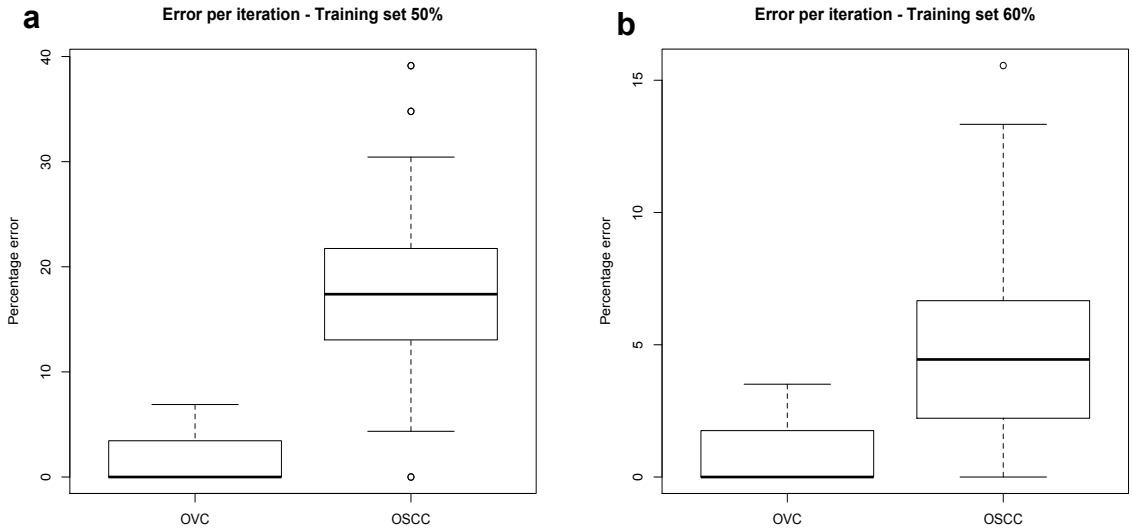


Figure 4.19 Percentage error for OVC and OSCC plots.

Figure 4.19 (a) is for the "50" analysis and (b) is for the "65" analysis.

4.2.2.2.5 Assessment of the list of genes with CNAs in OSCC

The GISTIC method was used to identify the most significant amplifications and deletions as described previously. Sixteen peaks were identified within the regions of highest significance gains and losses, and these peaks had a large gene lists that will be further investigated in this section (OSCC gene lists are not shown, too large data). The results were then further analysed by running the gene lists against 13 KEGG pathways, which are more related to H&N cancers, as well as cancer gene census and Stransky mutation list (76 previously identified genes in HNSCCs harbouring high statistically significant mutations) (Stransky *et al.*, 2011) (refer to chapter 2, section 2.8.3.1.6).

Out of 138 key genes hits (refer to figure 4.20), fifty genes were cancer genes (25% of the CN altered genes in OSCC cohort were found to be related with cancer). Additionally, 38 genes were involved in P13K pathway (19% of the CN altered genes are in this pathway), which considered crucial as aberrations in genes in this pathway were frequently found in OSCCs (Chang *et al.*, 2013). Twenty-two genes were in the KEGG JAK STAT signalling pathway (11% of the CN altered genes are in this pathway), and 17 genes were involved in the KEGG WNT signalling pathway (8% of the CN altered genes were in this pathway). Similarly, 15 genes were involved in KEGG calcium signalling pathway (7% of the CN altered genes were in this pathway). Table 4.4 lists all key genes founded to be in CN altered regions with the highest significance gains and losses in OVH, OVC and OSCC cohorts. In general, OSCCs illustrate more whole chromosome and localised gain and loss and a higher degree of CN alterations compared to OVCs, and accordingly, more genes were identified, noticeably, among cancer genes and in the P13K pathway. This agrees with the recent analysis report of whole-exome sequencing data for 151 HNSCC tumours, which revealed that the most frequently mutated oncogenic pathway, with significantly higher mutation rates in known cancer genes, was the PI3K pathway (30.5%) (Lui *et al.*, 2013). In contrast, OVC showed a lower degree of CN alterations compared to OSCCs, and consequently, fewer genes were identified table 4.4.

I then focused on genes that had a significant role in oral and head and neck cancers, which were in CN altered regions in the OSCC group but not in OVCs (primary OSCC data belong to Pre-cancer Genomics Group). The *SMAD4* gene located on chromosome 18q21.1 was in a deleted chromosomal region in the OSCC cohort. Loss of the *SMAD4* gene appears to play a key role in HNSCC tumours progression (Snijders *et al.*, 2005), and LOH at this locus has been described at ~50% of HNSCCs (Kim *et al.*, 1996), (Takebayashi *et al.*, 2000). Furthermore, expression loss of *SMAD4* protein had an impact on lymph node metastasis and tumour depth in patients with esophageal squamous cell carcinoma (Natsugoe *et al.*, 2002). Similarly here, *TP53AIP1* located on chromosome 11q24.3 was in a deleted chromosomal region in OSCC cohort. *TP53AIP1* plays an important role in the apoptotic signalling of tumour suppressor *p53*, and was found to be mutated in prostate cancer tissue (Wang *et al.*, 2006). In this study, *CDKN2A* gene in 9p21 was also in a deleted chromosomal region in OSCC cohort. *CDKN2A* gene is described as one of the early aberrations and could be considered as a significance prognostic for patients with oral cancer, or pre-malignant lesions (Rosin *et al.*, 2000), (Jiang *et al.*, 2001), (Leemans *et al.*, 2011b). One of the most vulnerable regions of the genome in OSCCs and HNSCCs is chromosome 3p and 3q arms (Leemans *et al.*, 2011b). *FHIT* gene located at 3p14.2 was in a deleted chromosomal region, and *TP63* gene (from *TP53* gene family), located at 3q28, was in a gain chromosomal region in OSCC cohort. *FHIT* and *TP63* genes are another HNSCC candidate cancer genes that have been identified some time ago (de Oliveira *et al.*, 2007), (Tsantoulis *et al.*, 2007), (Leemans *et al.*, 2011b). The above genes in CN altered regions in OSCC group have not been identified to be in CN altered regions in OVC cohort, which could again be another explanation for the aggressiveness behaviour of OSCC and the benign behaviour for OVC.

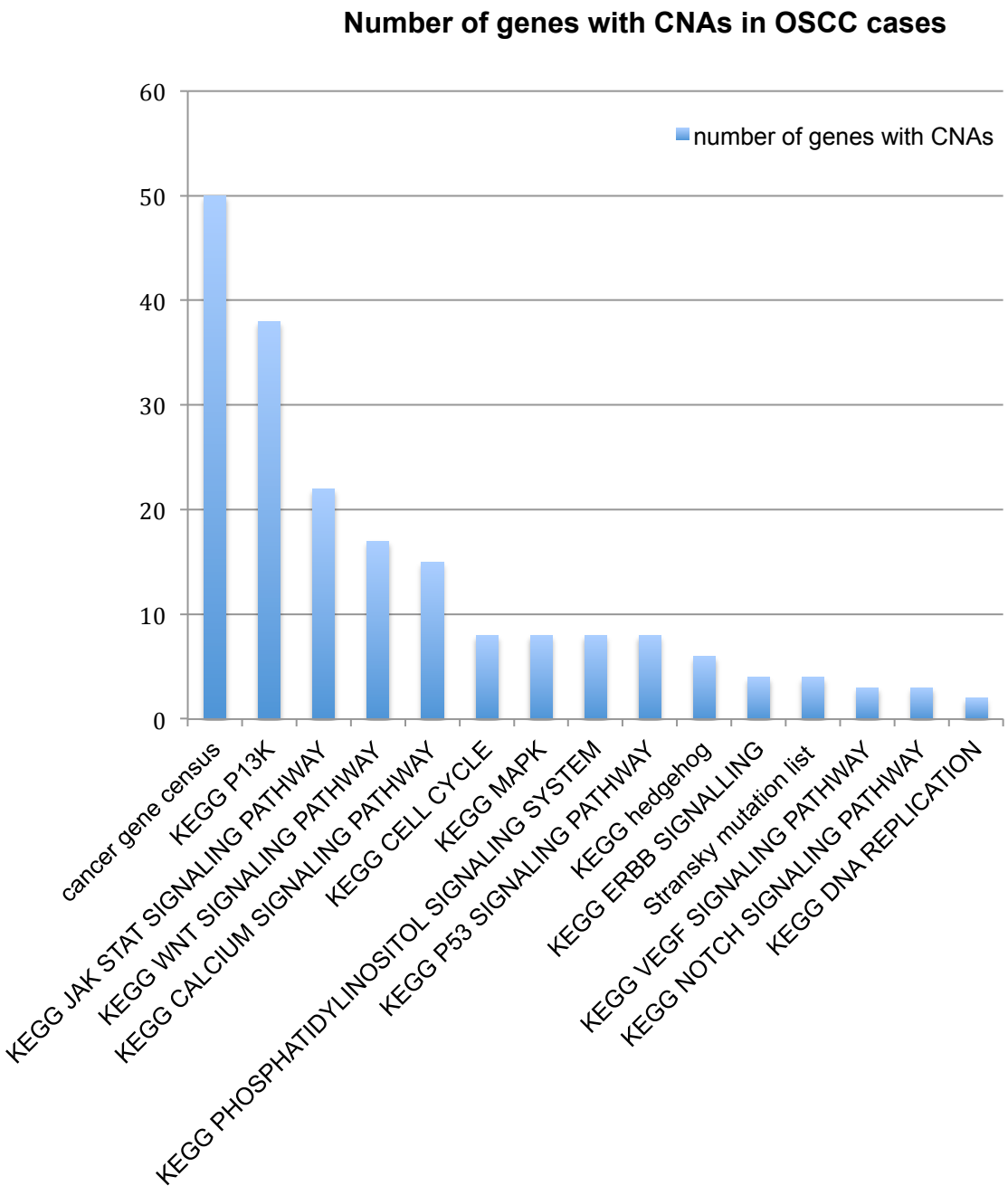


Figure 4.20 Number of genes with CNAs in OSCC cases

A graphical representation of: 13 KEGG pathways, cancer gene census and Stransky mutation list (Stransky *et al.*, 2011) on the x-axis ranked by the number of genes with CNAs from OSCC samples in each pathway and list on the y-axis

Table 4.4 Lists of All key genes founded to be CN altered within regions with the highest significance gains and losses in OVH, OVC and OSCC cohorts.

	Genes		
	OVH	**OVC**	**OSCC**
KEGG WNT SIGNALING PATHWAY	*PPP2CA, SKP1, TCF7, WNT8A*	*RAC1, RHOA, NFATC3, NFAT5, PLCB2, CHP, DAAM2, PPP2CA, SKP1, TCF7, WNT8A*	*PRICKLE2, CER1, CTBP1, WNT6, FZD5, WNT10A, SMAD4, MMP7, PPP2R1B, WNT7A, SMAD2, PPP2CB, PPP3CC, FZD3, NFATC1, CSNK1A1L, MMP7*
KEGG JAK STAT SIGNALING PATHWAY	*IL9*	*EPO, IL9*	*IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA7, IFNA8, IFNA10, IFNA13, IFNA14, IFNA16, IFNA17, IFNA21, IFNB1, IFNW1, JAK2, IFNK, IFNE, CBL, IL10RA, IL5RA, PIAS2*
KEGG MAPK		*RAC1, JMJD7-PLA2G4B, CHP, PLA2G4E, PLA2G4B, CACNA1G*	*FGFR3, HSPA8, RAF1, DUSP4, PPP3CC, FGF17, FGF20, FGF9*
KEGG DNA REPLICATION		*RFC2, MCM7*	*RFC3, RNASEH2B*
KEGG NOTCH SIGNALING PATHWAY		*DTX2, DLL4*	*CTBP1, MAML2, KAT2B*
KEGG PHOSPHATIDYLINOSITOL SIGNALING SYSTEM		*ITPKA, PLCB2, DGKE*	*DGKQ, INPP5D, DGKD, PLCD4, PIKFYVE, ITPR1, PIK3C3, DGKH*
KEGG P13K	*PPP2CA*	*RAC1, YWHAG, LAMB2, RPS6KB1, EPO, GNB2, CHAD, PPP2CA*	*IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA7, IFNA8, IFNA10, IFNA13, IFNA14, IFNA16, IFNA17, IFNA21, IFNB1, JAK2, RPS6, TEK, RHEB, FGFR3, PPP2R2C, COL4A3, COL4A4, COL6A3, CREB1, FN1, IRS1, EIF4E2, PPP2R1B, PDGFD,*

KEGG CALCIUM SIGNALING PATHWAY	*VDAC1*	*FZD9, ITPKA, PLCB2, CHP, CACNA1G, VDAC1*	*HTR5A, DRD5, ERBB4, HTR2B, PLCD4, GRM5, ATP2B2, HRH1, ITPR1, OXTR, ADRA1A, PTK2B, PPP3CC, HTR2A, CYSLTR2*
KEGG ERBB SIGNALLING		*RPS6KB1, PAK6*	*ERBB4, CBL, RAF1, NRG1*
KEGG CELL CYCLE	*SKP1, CDC23*	*YWHAG, CDC25A, E2F4, BUB1B, MCM7, SKP1, CDC23*	*CDKN2A, CDKN2B, SMAD4, ATM, CHEK1, SMAD2, RB1, CCNA1*
KEGG P53 SIGNALING PATHWAY		*SHISA5, PPM1D, SERPINE1*	*CDKN2A, PMAIP1, ATM, CHEK1, EI24, TP53AIP1, SESN3, TNFRSF10B*
KEGG VEGF SIGNALING PATHWAY		*RAC1, NFATC3, NFAT5, JMJD7-PLA2G4B, CHP, PLA2G4E, PLA2G4B*	*RAF1, PPP3CC, NFATC1*
Transky mutation list			*CDKN2A, SI, SLITRK3, TP63*
Cancer gene census	*SUZ12*	*PMS2, RAC1, ELN, HIP1, SETD2, CLTC, RNF43, BRIP1, CBFB, CDH1, BUB1B, HLF, SUZ12*	*FHIT, MITF, FOXP1, CDKN2A, JAK2, MLLT3, NFIB, CD274, MLL3, FGFR3, WHSC1, ATIC, CREB1, ACSL3, IDH1, PAX3, FEV, MALT1, BIRC3, ATM, DDX6, DDX10, CBL, FLI1, KCNJ5, MLL, PAFAH1B2, POU2AF1, SDHD, PCSK7, ARHGEF12, MAML2, FANCD2, PPARG, RAF1, VHL, XPC, SRGAP3, SETBP1, PCM1, WRN, BRCA2, CDX2, FLT3, LCP1, RB1, LHFP, TTL, LPP, BIRC3*

Genes in black are genes within amplification regions. Genes in red are genes within deletion regions.

4.3 Discussion

In this study, the NGS copy number analysis method was used quantities of DNA isolated from FFPE tissue, in the largest sample cohort described to date, with the aim of identifying C genomic characteristic features, and determine if next generati CNA could distinguish between the genomic damage pattern in C OSCC lesions. Until now, the literature focusing on oral verrucous and OVC) clinicopathological characteristics is not robust. The lesions also makes them difficult to investigate, and most previou been made on small numbers of cases. The possible link betwee behavior and the structural and functional features with the clini OVH and OVC is still not clear and requires further investigation previous studies that intended to distinguish between OVC an yielded mixed results (Devaney *et al.*, 2011a). The distinction of C is a frequent problem also for pathologists and cannot be base histological features as it appears benign cytologically, beside hyperchromatism of the focal basal cell layer (Eversole and 1983), (Woolgar and Triantafyllou, 2009). Hence, molecular app offer an important help.

Because sequencing technologies are currently becoming mor accurate and affordable, low coverage sequencing CN analysi more informative and expedient (Gusnanto *et al.*, 2012). sequencing data can be produced even with low DNA amounts formalin-fixed paraffin-embedded (FFPE) materials (Wood *et al* generation sequencing techniques offer considerable benefits fo including precise delineation of the CN breakpoints, and higher detect single-base insertions or deletions) of CN changes, ir (Meyerson *et al.*, 2010). It enables the estimation of tumour-to-n at a genomic locus through counting the number of reads at this l and tumour samples (Meyerson *et al.*, 2010).

Classical OSCC have the same distinctive pattern of amplification and deletion commonly seen in other tumours with squamous histology (Hoadley *et al.*, 2014) but this was completely absent from OVC. Furthermore, oral verrucous tumours had their own set of chromosomal abnormalities, although fewer in number as would be expected from their reduced aneuploidy (Pentenero *et al.*, 2011). This lower level of chromosomal instability could be linked to the minimal or absent histological cytological atypia found in OVC (Zargaran *et al.*, 2012). Interestingly, losses were detected frequently in OSCC genomes but rarely in OVCs. Losses on chromosomal arms 3p, 4q, 9p and 18q, and gains on 3q, 5q, 8q, and 20p are chromosomal signatures commonly linked with OSCCs (Baldwin *et al.*, 2005), (Snijders *et al.*, 2005), (Liu *et al.*, 2006), (Jarvinen *et al.*, 2008), (Freier *et al.*, 2010) and were frequently identified in this OSCC cohort but were absent in OVC, suggesting that these CN alterations may be related to the more aggressiveness behaviour of OSCC tumours. Gains of chromosomes 7q, 16q and 17q were detected in OVCs at a frequency of ~50%, suggesting that these CN alterations may be involved in the development of OVC. Significant CN gain in OVH group at 7q, with a frequency of ~50%, was present in OVCs, suggesting that this region might harbour the first CN alteration involved in oral verrucous lesions.

In many human cancers, loss of chromosome 3p is a common genetic variant, as this region include several putative tumour suppressor genes (Zabarovsky *et al.*, 2002), (Garnis *et al.*, 2004). Loss of 3p carries prognostic consequence in oral cancer, mainly for second primary tumours development, disease progression risk and presence of local recurrences (Rosin *et al.*, 2002), (Mao *et al.*, 2004). In addition, variations on chromosome 3p are thought to be crucial events in the development of oral premalignant lesions to aggressive disease (Tsui *et al.*, 2008). Previous reports have demonstrated alterations and LOH in specific segments on this chromosome arm in oral premalignant lesions and that these alteration were observed more with lesions progression to cancer (Califano *et al.*, 1996) (Mao *et al.*, 1996) (Rosin *et al.*, 2000). A study was conducted in 2008 to examine chromosome 3p arm throughout the multistep oral tumorigenesis for genetic pattern, frequency of variation and size of

alteration (Tsui *et al.*, 2008). Their data showed an increase in the genomic instability pattern at chromosome 3p as specific genetic events were associated with increased histopathological progression and disease stage (Tsui *et al.*, 2008). They reported whole 3p arm loss in OSCCs, high-grade dysplasias had segmental losses and most low-grade dysplasias had no losses (Tsui *et al.*, 2008). From the earlier reports described here, it is clear now that loss in 3p is a chromosomal signature usually associated with OSCCs, which was shown as well in the CN karograms of OSCC cohort (pre-cancer genomic data) and might explain the aggressiveness behaviour of this tumour. In contrast, OVC CN karyograms did not show any loss in 3p within the entire cohort, which might explain the indolent behaviour of this tumour. Additionally, and based on the genomic CN profiles of OVC and OSCC cohorts, the logistic regression analysis successfully distinguished between the two patient groups, correctly identifying 56 out of 57 OVC samples and 41 out of 45 OSCCs.

The defined break points in gain and loss regions in the current work has allowed naming specific gene candidates (refer to table 4.4). Some insight into the genes that might have a role in the development of OVC (as discussed in section 4.2.1.1.3.6) may be obtained by considering candidate tumour suppressors and oncogenes that have been previously suggested for other cancers based on finding that they are deleted or amplified in those tumours. Besides, it is well understood now that the more were the numbers of genomic variations can be an indication of cancer progression. However, It is important to keep in mind that candidate oncogenes at regions of low CN gain levels might function differently than when at high CN gain levels (Bhattacharya *et al.*, 2011). Nonetheless, it must be remembered as well that the suggested genes in this study, at best, can only report on the “likelihood” of progression of OVC and requires further investigations.

Studies on a number of tumour types (breast cancer, oral and oropharyngeal cancers, soft tissue sarcoma and prostate cancer) have identified that subtypes usually associated with better prognosis often lack CN instability (Fridlyand *et al.*, 2006), (Smeets *et al.*, 2009), (Barretina *et al.*, 2010), (Taylor *et al.*, 2010). In oral cancer, a study in 2011 distinguished two oral lesions subtypes that are

suggestive for the development of at least two oral cancer pathways, which differ, in chromosomal instability and metastasis risk (Bhattacharya *et al.*, 2011). They applied aCGH technique to determine CN alterations in OSCC and dysplasia groups. They revealed two subtypes that were distinguished by acquisition of one or more specific CN alterations at four genomic regions: gain in 3q, loss in 8p, gain in 8q, and gain in chromosome 20 (Bhattacharya *et al.*, 2011). Furthermore, these subtypes were significantly differing in their metastasis risk, and accordingly, they suggested that these variations in CN aberrations constitute a biomarker with clinical value in identifying patients' treatment on the basis of cervical lymph node metastasis risk as their results showed that neck metastasis was present in 22 of 48 (46%) of +3q-8p+8q+20 tumours and in only one of 15 (7%) of non +3q-8p+8q+20 tumours (Bhattacharya *et al.*, 2011). The non +3q-8p+8q+20 tumour subtype was clearly a member of the low genomic instability and the low metastasis risk group, and here, OVC copy number analysis study showed that this cohort is similar to non +3q-8p+8q+20 tumours, in which it lack these chromosomal abnormalities, characterised by low levels of genomic instability and known to be a non-metastasising carcinoma.

Considered together the distribution of CN aberrations in OVC and OSCC suggest that there are two different distinct routes to two different oral cancers, one associated with greater genomic rearrangements and acquisition of previously known OSCC chromosomal signatures, and the other lacking these CN aberrations and with lower level of chromosomal instability detectable by CN analysis. However, the driving potency for both tumours remains unclear. It was noted previously as well that differential methylation was linked with the cases with high levels of CN changes (Poage *et al.*, 2010) and (Bhattacharya *et al.*, 2011). The low level of chromosomal instability in OVC tumours may also suggests that the development of these tumours could be linked with other neutral CN mechanisms such as epigenetic alterations or microsatellite instability. Genomic changes in methylation patterns were observed in head and neck SCC (Poage *et al.*, 2010); however, microsatellite instability is not known to be common in oral cancer (Shaw *et al.*, 2008). Therefore, to investigate

whether OSCC and OVC differ in methylation patterns, I suggest a further analysis for both, CN and methylation measurements for the two cohorts. In addition, there is still an open possibility that there might be underlying balanced chromosomal rearrangements in OVC, or, an extrinsic factor, such as inflammation of neighbouring cells, which may affect the growth of oral epithelial cells (Arwert *et al.*, 2010). HPV infection is another candidate that has been previously reported with OVC (Shroyer *et al.*, 1993), (Mitsuishi *et al.*, 2005), (Fujita *et al.*, 2008) and will be discussed in the next chapter (chapter 5).

Tumor genomic profiles can be produced using NGS copy number analysis for DNA samples extracted from patients' FFPE materials. However, the tumour cells content in the extracted DNA can be affected by the presence of stroma cells in the macrodissected carcinoma tissues. To avoid this complication, laser microdissection can be used to isolate tumour epithelium tissue from the surrounding healthy and stroma materials. On the other hand, a previous report suggested that the concealing stroma effect on tumour epithelium appears to be minimal in macrodissected extracted tissue, and that therefore; macrodissection is the convenient preferable technique for nucleic acids extraction (de Bruin *et al.*, 2005). Given these previous points, the performed macrodissection method on OVC tissue here does not show an effect on the low-coverage CN produced genomic profiles. The tumour cells content were enough to reveal CN alterations in all macrodissected tissue samples.

It has been previously demonstrated that NGS can provide genomic CN gain and loss details in a cost-effective manner from DNA isolated from different sources, including FFPE tissue blocks stored after histopathological diagnosis, frozen tumour samples and cell lines (Wood *et al.*, 2010), (Hayes *et al.*, 2013). It has been also shown that the resolution of NGS copy number analysis method has a high degree of correlation and comparable with aCGH, but gave more information for less money when applied at low multiplexing levels, and it is extremely adjustable (Wood *et al.*, 2010). It is also important to keep in mind that aCGH technique has shown difficulty to use with DNA extracted from FFPE materials (Hostetter *et al.*, 2010). Additionally, aCGH requires microgram DNA quantities while NGS can produce CN genomic keryograms from nanogram

quantities of DNA (less than 100ng) (Wood *et al.*, 2010). When compared to PCR-based methods such as LOH analysis, NGS produces much more data when performed at high multiplexing levels (Wood *et al.*, 2010).

The CN analysis method applied in this study provided a digital readout of viral subtypes, loads, as well as tumour karyograms in a single test (discussed in further details in chapter 5). It has been also revealed here that good quality CN data can be attained when multiplexing 40 samples on one single lane of an Illumina HiSeq 2500. Multiplexing is an essential aspect in designing research studies according to the required selected resolution, available resources and accessible sample numbers. Another key point, copy number libraries can be used for several times after being aliquoted. Accordingly, further examination of previously prepared and low-resolution screened libraries can be obtained without the need of additional preparation steps, and hence, data from both screenings on the same sample can be compound to provide a double coverage (Wood *et al.*, 2010). Despite the proven utility of next generation sequencing copy number aberrations detection (Wood *et al.*, 2010), (Hayes *et al.*, 2013), it cannot detect neutral CN variations (genomic variations that do not cause changes in the amount of the genetic material), such as inversions and balanced translocations (Coughlin *et al.*, 2012). Balanced translocations and Inversions that occur in coding region breakpoints can result in a disease phenotype (Coughlin *et al.*, 2012). One of the limitations here was the lack of technical replicates. Though, pre-cancer genomics group previously validated the reproducibility of the same methodology I used here for CN analysis as they did technical replicates back when they first developed the technique (Wood *et al.*, 2010). They sequenced the same DNA libraries twice and made libraries from the same DNA twice, and all times, the produced CN karyograms were virtually identical (Wood *et al.*, 2010).

Another limitation in this study was that I did not check the effect of the fixation procedure on the produced CN genomic profiles by comparing the generated karygrams for DNA extracted from FFPE materials with CN karyograms for DNA extracted from fresh frozen tissue from the same OVC samples. Nevertheless, the rarity of oral verrucous lesions made it really hard to get any

fresh frozen OVC samples. DNA is susceptible to degradation in fixative solutions used for tissue preservation in histopathology labs (Ferrer *et al.*, 2007), and again, pre-cancer genomics group have previously investigated the effect of fixation on CN karyograms produced from DNA extracted from FFPE materials when they first developed the method. They compared the CN genomic profiles for DNA extracted from fresh frozen against FFPE materials from the same lung carcinomas (Wood *et al.*, 2010). They have shown that the corresponding fixed and fresh CN karyogarms for DNA extracted from the same samples were nearly identical (Wood *et al.*, 2010). The slight differences were in the magnitude of same CN variants and was attributed to macrodissection of non-cancerous cells in fixed samples, such as inflammatory and stroma cells (Wood *et al.*, 2010). Additionally, the lack of paired tumour and normal samples was again another limitation in the current study which if were available would reduce the noise usually associated with CN profiles produced from DNA extracted from FFPE materials (Bhattacharya *et al.*, 2011).

The conducted copy number variation analysis in this project was carried out following a previously published procedure by pre cancer genomics group. Generally speaking, sequence variations are usually not distributed uniformly within genomes (Nguyen *et al.*, 2006). Nonetheless, CNVs that are enriched in simple tandem repeats occur more often towards centromeres and telomeres, and are not elevated in G + C content or SNPs (Nguyen *et al.*, 2006). Since these regions are not overrepresented in CNVs (Sharp *et al.*, 2005), I excluded CNAs in all centromere and telomere chromosomal regions throughout the different CN data analysis approaches. For the CN obtained data here, I obtained between one million and ten million reads. One million reads is one read every 3 kb, which is 0.033 X coverage, and ten million reads is 0.33 X coverage; which in turn considered a relatively low coverage. Though, even with this low, genomic CN profiles generated clear karygrams with apparent chromosomal gain and loss features and enabled the differentiation between OVH, OVC, and OSCC cohorts. On the other hand, and to clarify the choice of the read window size used in OVC CN analysis, the selected 400 read window size is a compromise between the resolution and the noise. 200 read windows

gives better resolution but are noisier and 800 read windows are very clean, but small genomic events will be missed. 400 read window size provides the nicest looking karyogram pictures for human readability. Also, and to explain more on the obtained coverage, in next generation sequencing, the number of reads produced by the sequencer dictates what coverage of the genome is obtained. At high coverage, there are many hundreds of millions of reads which cover the whole or target area of the genome many times and can allow for observations to be inferred about point mutations. At low coverage, few reads per genome are obtained and these may be more spread out so that essentially the genome is sampled. This produces a relatively low-resolution genomic data compared to whole genome sequencing equivalent, but nonetheless, it is much cheaper and many more samples can be processed at once.

4.4 Copy number karyograms changed a misrepresented histological diagnosis

Five cases in this project had an initial suspected diagnosis of OVC. Following the visual examination of the CN karyograms for those samples; they revealed OSCC chromosomal signatures, mainly, loss in chromosome 3p arm and gain in 3q (refer to appendix 4.6). Careful, blind histological revision by two different pathologists indicated that the final diagnosis is OSCC with verrucous appearance and these cases were excluded from the study. CN karyograms changed the misrepresented pathological diagnosis in all five cases and hence; can be used for differential diagnosis of OVCs and OSCCs.

In summary, this is the first study of OVH and OVC profiling from FFPE tissue blocks up to date. The visual inspection of patient's CN karyograms revealed that genomic signatures usually associated with OSCCs were completely absent in oral verrucous lesions and losses were detected more frequently in OSCCs than OVCs. The current study has demonstrated that NGS CN analysis can be used for more specific assessment and evaluation of OVH and OVC heterogeneity based on the analysis of the whole genome CN karyograms. I demonstrate here that CN analysis could contribute to differential diagnosis of oral verrucous lesions and classical OSCCs using routine biopsy specimens.

Chapter 5 Next-Generation Sequencing Analysis for Detecting Human Papillomavirus in oral verrucous lesions.

5.1 Introduction

The association between human papillomavirus (HPV) and HNSCC is strongest among oropharyngeal squamous cell carcinomas, especially for cancers of the lingual and palatine tonsils (Schwartz *et al.*, 1998), (Gillison *et al.*, 2000), (Mork *et al.*, 2001), (Ernster *et al.*, 2007) Furthermore, the risk of developing oropharyngeal cancer when adjusted for tobacco and alcohol use is substantially increased with high-risk HPV oral infections (Hansson *et al.*, 2005). HPV has been identified in 45% to 95% of OPSCC (Fakhry and Gillison, 2006), (Hammarstedt *et al.*, 2006), (Nasman *et al.*, 2009) and is believed to be an aetiological agent (Marklund and Hammarstedt, 2011).

OVC is a low-grade, rare variant of OSCC. Verrucous in this context describe a lesion that morphologically mimics verruca vulgaris (a lesion caused by HPV). Since a verrucous appearance is suggestive of viral aetiology, this has prompted a number of investigations to study the putative association between HPV and those lesions (Stokes *et al.*, 2012). HPV has been cited as a probable aetiology in VC pathogenesis by various authors (Spiro, 1998). However, the majority of the studies that investigated 'HPV presence' in verrucous lesions relied on PCR, and *in-situ* hybridization (ISH), for detection and did not identify HPV transcriptional activity markers or quantitate HPV viral load (Miller and Johnstone, 2001). The high sensitivity of the PCR technique can amplify very small quantities of HPV DNA and this can lead to detection of non-pathologic HPV infections or false-positives if sample contamination occurs (Ha *et al.*, 2002), (Kreimer *et al.*, 2005). Prior to and until 1997, 15 published studies investigated the presence of HPV DNA in OVC. Among the 159 analysed samples, HPV DNA was identified in 37.7% of the cases, and HPV subtypes 6 and 11 were the most predominant identified HPV infections (47%) (Kari J. Syrjänen, 2000). At present, the possible role of HPV in VC pathogenesis is

proposed by HPV incidence in VC cases, which varies from 0% to 100% (Stokes *et al.*, 2012), (Altshuler *et al.*, 2010), (Kondi-Paphitis *et al.*, 1998). This range indicates that HPV prevalence in oral verrucous lesions and its actual role in cancer pathogenesis is controversial and inconclusive. This variation can be attributed to the deficiency of standardized detection procedures and the difficulty in defining complete histological criteria for OVH and OVC cases. Furthermore, the rarity of these types of lesions makes it difficult to study and investigate, and most previous studies or case reports have been made on small number of cases.

5.2 Aim

The aetiology of OVC is unknown, and the suggested role of human papillomavirus HPV as a causative factor remains contentious. The aim of this study was to analyse a subset of oral verrucous lesions (including VC, and VH cases) for the presence of HPV subtypes and all characterized human viral genomes.

5.3 Results

5.3.1 Characteristics of the study cohort

From the verrucous cohort study described in Chapter three and table 3.1, pathological materials from each case were available in the form of FFPE tumour blocks. In total, 73 oral verrucous samples, malignant (57 OVC) and pre-malignant (16 OVH) regions were identified and Dr Alec High assessed the original diagnoses. All the 73 samples were selected for copy number analysis and viral subtypes and loads detection (37 males and 36 females, mean age 68.2 years, range 46-96).

5.3.2 HPV Detection by NGS

Viral load was measured as described in chapter two, section 2.8.3.2. I used HPV sequencing data from a previous study published by pre-cancer genomics group to provide positive and negative controls. I matched my samples with 16 oral and oropharyngeal (OP) cases. Nine positive HPV cases were detected out of 16 successfully sequenced samples (Conway *et al.*, 2012). Sequencing libraries were prepared from all 73 samples as described in chapter two, section 2.8. Table 5.1 lists the range of human and viral reads. HPV-16 sequence was identified in one OVH and one OVC, and HPV-2 sequence was detected in one OVC out of 73 oral verrucous samples at 95% confidence level with [2.24, 8.16, and 0.33 viral genomes per cell] respectively. The standard deviation of reads taken from all patient verrucous samples was 4 257 556.901 [ranging from 296 655 to 115 682 098] human reads.

Table 5.1 **Viral load determined by next generation sequencing in human oral verrucous carcinomas (n = 57) and hyperplasias (n = 16)**

Patient ID	Species	Virus length	Human reads	Virus reads	Read density Kb per read	Kb virus sequence	Viral load	Detectable HPV load at 95% confidence	Probability of detecting 1 copy of HPV per cell
V-04-01-E9	None	0	589117	0	10.18	0	0	3.8621	0.5396
V-07-01-A	Herpesvirus 6A	159322	671946	8	8.93	71.43	0.45	3.3860	0.5872
V-07-01-A	Herpesvirus 1	152261	671946	1	8.93	8.93	0.06	3.3860	0.5872
V-10-01-1 e1	None	0	1556862	0	3.85	0	0	1.4614	0.8712
V-014-01-5	None	0	1288201	0	4.66	0	0	1.7662	0.8166
V-019-01-I4	None	0	1182923	0	5.07	0	0	1.9234	0.7893
V-20-3	None	0	4253014	0	1.41	0	0	0.5350	0.9963
V-026-01-A4	None	0	666770	0	8.99	0	0	3.4123	0.5844
V-060-01	None	0	289026	0	20.76	0	0	7.8721	0.3165
V-61-1-4	None	0	4901408	0	1.22	0	0	0.4642	0.9984
V-62-1-B	Herpesvirus 6A	159322	5978774	2	1	2.01	0.13	0.3806	0.9996
V-63-1-2	None	0	6794116	0	0.88	0	0	0.3349	0.9999
V-65-1-D1	None	0	7769042	0	0.77	0	0	0.2929	1.0000
V-66-1-B	None	0	4446770	0	1.35	0	0	0.5117	0.9971
V-67-1-B1	None	0	6100026	0	0.98	0	0	0.3730	0.9997
V-68-1	Herpesvirus 5	235646	2984540	4	2.01	8.04	0.03	0.7623	0.9803
V-69-1-D	Herpesvirus 5	235646	8960354	2	0.67	1.34	0.01	0.2539	1.0000
V-70-1-B	Herpesvirus 7	153080	6371664	2	0.94	1.88	0.01	0.3571	0.9998
V-71-1-5	None	0	7896410	0	0.76	0	0	0.2881	1.0000
V-72-1-4	None	0	6091470	0	0.99	0	0	0.3735	0.9997
V-73-1	None	0	7774162	0	0.77	0	0	0.2927	1.0000
V-74-1-A	None	0	8256130	0	0.73	0	0	0.2756	1.0000
V-75-1	None	0	1527565	0	3.93	0	0	1.4895	0.8662
V-77-1	None	0	7883782	0	0.76	0	0	0.2886	1.0000
V-78-1-A	Herpesvirus 5	235646	6029256	4	0.99	3.98	0.017	0.3774	0.9996
V-79-1-B	Herpesvirus 6B	162114	6483082	2	0.93	1.85	0.01	0.3510	0.9998

V-86-1	None	0	15107456	0	0.4	0	0	0.1506	1.0000
V-88-1	Herpesvirus 1	152261	6584298	2	0.91	1.82	0.01	0.3456	0.9998
V-89-1-A	None	0	6898174	0	0.87	0	0	0.3298	0.9999
V-90-1	None	0	4493076	0	1.34	0	0	0.5064	0.9973
V-91-1	Herpesvirus 1	152261	8733136	2	0.69	1.37	0.01	0.2605	1.0000
V-92-1	Herpesvirus 1	152261	10556564	4	0.57	2.27	0.02	0.2155	1.0000
V-93-1	None	0	1538774	0	3.90	0	0	1.4786	0.8682
V-94-1	HPV-2	7860	4614998	2	1.3	2.6	0.33	0.4930	0.9977
V-94-1	Herpesvirus 1	152261	4614998	4	1.3	5.2	0.03	0.4930	0.9977
V-98-1-D	None	0	7454606	0	0.81	0	0	0.3052	0.9999
V-99-1-4	None	0	6585616	0	0.91	0	0	0.3455	0.9998
V-100-1-B	None	0	4716241	0	1.27	0	0	0.4824	0.9980
V-101-1-F	None	0	4412444	0	1.36	0	0	0.5156	0.9970
V-102-1	None	0	6610388	0	0.91	0	0	0.3442	0.9998
V-105-01	Herpesvirus 1	152261	10847800	2	0.55	1.10	0.01	0.0109	1.0000
V-108-01	None	0	13893548	0	0.43	0	0	0.1638	1.0000
V-109-T	Herpesvirus 4	171823	14969258	6	0.40	2.41	0.01	0.006988318	1.0000
V-110-01	None	0	12720402	0	0.47	0	0	0.1789	1.0000
V-112-T	None	0	5062270	0	1.19	0	0	0.4495	0.9987
V-113-T	None	0	4086390	0	1.47	0	0	0.5568	0.9954
V-114-T	None	0	7797994	0	0.77	0	0	0.2918	1.0000
V-116-T	None	0	10659704	0	0.56	0	0	0.2134	1.0000
V-117-T	None	0	414682	0	14.47	0	0	5.4867	0.4207
V-118-01	None	0	15473750	0	0.39	0	0	0.1471	1.0000
V-119-T	None	0	3504536	0	1.71	0	0	0.6492	0.9901
V-122-01	None	0	12558750	0	0.48	0	0	0.1812	1.0000
V-123-T	None	0	11502764	0	0.52	0	0	0.1978	1.0000
V-124-T	Herpesvirus 6B	162114	9115790	142	0.66	93.46	0.58	0.0122	1.0000
V-125-T	HPV-16	7905	10041734	108	0.60	64.53	8.16	0.2264	1.0000
V-95-1-B	Herpesvirus 1	152261	11163878	4	0.54	2.15	0.01	0.2038	1.0000
V-95-1-B	Herpesvirus 6B	162114	11163878	4	0.54	2.15	0.01	0.2038	1.0000
V-02-02-C1	Herpesvirus 5	235646	296655	1	20.23	20.23	0.09	7.6696	0.3233
V-02-02-C1	Herpesvirus 1	152261	296655	2	20.23	40.45	0.27	7.6696	0.3233
V-03-01-A1	None	0	431502	0	13.9	0	0	5.2728	0.4334
V-015-01-A3	None	0	1554350	0	3.86	0	0	1.4638	0.8708
V-022-01-4	HPV-16	7905	2368058	7	2.53	17.74	2.24	0.9608	0.9558

V-025-01-D2	None	0	668260	0	8.98	0	0	3.4047	0.5852
V-029-02-A3	None	0	840256	0	7.14	0	0	2.7078	0.6692
V-30-1	Herpesvirus 1	152261	2943604	4	2.04	8.15	0.05	0.7729	0.9793
V-031-01-B1	None	0	583436	0	10.28	0	0	3.8997	0.5361
V-033-01-4	herpesvirus 6A	159322	1182195	18	5.08	91.36	0.57	1.9246	0.7891
V-033-01-4	Herpesvirus 6B	162114	1182195	1	5.08	5.08	0.03	1.9246	0.7891
V-40-1	None	0	7611248	0	0.79	0	0	0.2989	1.0000
V-041-01-G	None	0	2324407	0	2.58	0	0	0.9788	0.9531
V-42-1-G	Herpesvirus 1	152261	14884005	6	0.4	2.42	0.02	0.1529	1.0000
V-44-1-3J	Herpesvirus 1	152261	4793286	2	1.25	2.5	0.016	0.4747	0.9982
V-46-1-2	Herpesvirus 6B	162114	12047262	150	0.5	74.71	0.46	0.1889	1.0000
V-106-01	None	0	9112970	0	0.66	0	0	0.2497	1.0000

5.3.3 Presence of Herpes viruses in verrucous samp

Patient tumour DNA's were scanned for all characterized sequences. Human herpesvirus sequences were detected in 21 c DNA samples, seven OVHs, and 14 OVCs, with viral loads rangir 0.58 viral genomes per cell as shown in Table 5.1, and with 0 standard deviation. Eleven samples were positive for herpesvirus positive for herpesviruses 6B, four were positive for herpesviruse positive for herpesviruses 6A, one was positive for herpesvirus was positive for herpesviruses 4. Four cases had a double infections for: (herpesviruses 6A and herpesviruses 1), (herpes herpesviruses 6B), (herpesviruses 5 and herpesvirus (herpesviruses6A and herpesviruses 6B).

To investigate whether the prevalence of Herpes virus detected VC cohort in this study, sequencing data from 23 head anc samples from a previous study published by pre-cancer ge (Conway *et al.*, 2012) were scanned for all characterized sequences. Human herpesvirus sequences were detected in e cases (seven oral, and one pharyngeal), with viral loads ranging 0.0362 viral genomes per cell, and with 0.00702 viral loads stan From the eight positive samples, four were positive for herpesv were positive for herpesviruses 5, and two were positive for he One case had a double herpesviruses infection for herpesv herpesviruses 5.

5.4 Discussion

NGS was described here as a novel but validated, powerful, high-throughput method to investigate the presence of HPV and all characterised human viral genomes loads and subtypes in the largest oral verrucous sample cohort described to date, following careful histological definition for OVC and OVH lesions. Testing for all HPV subtypes here revealed HPV-16 positivity in only one OVH and one OVC sample and HPV-2 positivity in one OVC sample out of 73 oral verrucous lesions with 2.2, 8.1 and 0.33 viral genomes per cell respectively. Although it is difficult to accurately predict the exact viral load with only a very small number of aligning viral reads, viral loads obtained in this study were clearly much lower than the viral loads obtained in the previously published study of HNSCC by pre-cancer genomics group (Conway *et al.*, 2012), in which the standard deviation of the viral loads obtained was 37.75 suggesting that the virus was not contributing to disease aetiology.

While some studies on verrucous lesions have shown HPV DNA positivity in their investigations, others have failed to detect HPV DNA in their cases (Table 1). A study was conducted in 1998 to investigate the prevalence of HPV infection in three vulvar VCs and indicated the presence of HPV 6 and 11 in all three cases (Kondi-Paphitis *et al.*, 1998). Another study was performed in 1993 where 17 VC cases were tested using PCR followed by DNA slot-blot hybridization for the presence of HPV DNA and showed a positive results in all samples (Shroyer *et al.*, 1993). Mitsuishi. *et al*. in 2005, investigated the presence of HPV DNA in four VCs of the lip using PCR with sequence analysis (Mitsuishi *et al.*, 2005). They suggested that numerous cutaneous and mucosal HPVs of high-risk types might participate with lip VC pathogenesis, as all samples harbored HPV DNA (Mitsuishi *et al.*, 2005). However, there is a suggestion that PCR amplification of viral DNA can be too sensitive, and produces false positive results (Ha *et al.*, 2002), (Kreimer *et al.*, 2005). In addition, Fujita. *et al*. in 2008, identified HPV genotypes in 11 OVCs out of 23 cases by using short PCR fragment (SPF)-PCR assay and sequencing, *in-situ* hybridization (ISH), and immunohistochemistry (Fujita *et al.*, 2008). They proposed that low- and high-risk multiple HPVs infections maybe associated in the histogenesis of OVC during hyperkeratinization (Fujita *et al.*, 2008).

In contrast, in 2009, Stankiewicz. *et al*. performed PCR to investigate HPV infection in 13 penile VC and identified HPV DNA in three cases only. They suggested that HPV infection and the tumour suppressor genes and oncogenes usually altered by virus infection are unrelated to penile VC pathogenesis (Stankiewicz *et al.*, 2009). Additionally, in 2012 Stokes. *et al*. studied the role of HPV in malignant and dysplastic oral verrucous lesions and proposed that although high- risk HPV DNA was identified in one of seven carcinomas and five of thirteen dysplasias, the oncogenic process is not enhanced by HPV oncoproteins as p16 overexpression was lacking and concluded that further work is needed on a larger cohort to determine HPV's biological significance in the development of VC (Stokes *et al.*, 2012). In 2012, del Pino. *et al*. investigated the prevalence of HPV in a total of 18 verrucous lesions. By performing PCR, only one head and neck VC and one penile VH were positive for HPV infection, and hence, they concluded that VC development is unlikely to be related to HPV infection (del Pino *et al.*, 2012). Another recent article by Patel *et al*. used HPV RT-PCR (reverse transcription PCR). Though, they had a significant subgroup of cases in which insufficient extracted RNA from the FFPE tumour sections prevented HPV RT-PCR testing (Patel *et al.*, 2013). Nonetheless, they concluded, "Active HPV in (H&N) verrucous carcinomas is rare enough to likely be clinically inconsequential."

In comparison with all previously published papers, NGS was used in this study for the detection of HPV in verrucous samples on the largest cohort to date. Moreover, the histological diagnostic criteria were clearly defined and oral verrucous appearing lesions with a focal stromal invasion in the bulk of the tumour were classified as SCCs with verrucous architecture, and hence, were excluded from the study. Also, it is often difficult to distinguish between OVH and OVC, since there are no clear-cut criteria in the literature to differentiate them (Alkan *et al.*, 2010a). In the current study, these lesions were separated according to the absence of invasive growth in OVH, as it is completely superficial and usually adjacent to normal epithelium.

Nevertheless, and along with the tested samples here, HPV sequencing data from a previous study conducted by pre-cancer genomics group were used to

provide positive and negative controls for the presence of HPV subtypes and all characterized human viral genomes (Conway *et al.*, 2012). Also, the applied method in the current study was validated before (on the control sample set) by detecting HPV sequences using PCR, and by evaluating P16 expression as a marker for HPV infection. It has been shown from the assessment of HPV screening results of the three approaches that NGS method has a high specificity and sensitivity for HPV detection when compared to the two other techniques (Conway *et al.*, 2012). Furthermore, It has been previously suggested that PCR methods can be over sensitive (Smeets *et al.*, 2007), while the method used here can provide a better specificity, as demonstrated by the observation that all p16 positive samples were also positive for HPV-16 by sequencing (Conway *et al.*, 2012). Moreover, and from the same previous study, HPV-61 was detected in one oral tumour by sequencing and was not detected by any other method, which again shows the ability of this method in detecting all HPV subtypes and loads (Conway *et al.*, 2012).

Previous studies have relied mostly on PCR and ISH to investigate the presence of HPV subtypes in verrucous lesions without quantitating HPV viral loads (Miller and Johnstone, 2001). Furthermore, HPV DNA may degrade in paraffin-embedded tissues. Sequencing may be less affected by this than PCR. The standard PCR test for HPV requires a 120-bp fragment to be amplified. DNA libraries are size selected here to be around 200 bp to ensure that enough fragments of <100 bp are sequenced. If an HPV sequence is in one of these, it would be picked up by sequencing but not by PCR. Besides, PCR methods for viral detection are specific to certain subtypes per test. One of the main advantages of using NGS is the fact that sequencing is blind. All known viral subtypes can be quantified in a single test. The method provides a digital readout of viral subtypes and loads with high sensitivity and specificity (Conway *et al.*, 2012), and the same sequence data can also be re-analysed to produce tumour karyograms. These data are extremely cheap to produce compared to many NGS methods, and can be multiplexed to over 50 samples per lane of an Illumina HiSeq. The analysis of oral verrucous data set karyograms are presented and discussed in chapter four.

The power of this method was also shown through the detection of other viruses by screening all verrucous samples for all other known virus sequence genomes. Human herpes viruses were identified in (21/73) of the oral verrucous lesions (28.77%), although these results has not been confirmed using any other diagnostic test. In addition, the control samples were scanned for all human viruses sequences and eight positive cases were identified out of 23 head and neck samples (34.78%). In general, Herpes simplex viruses-related infections are among the highest widespread diseases, affecting approximately 60% to 95% of adult human population (Brady and Bernstein, 2004). The two human herpesviruses known to be associated with cancer are Kaposi's sarcoma-associated herpesvirus (KSHV), and Epstein-Barr virus (EBV) (Everly *et al.*, 2012), and these were not detected in oral verrucous samples here. Nonetheless, it is important to point out that the detection of virus DNA in patients' samples does not essentially indicate a viral pathogenic role in a disease. NGS tells nothing about transcriptional activity, so it is not possible to speculate further on the clinical significance of this finding. However, by infecting defense cells, many herpes viruses can persistently arise in different human tissues in the event of inflammation (Ferreira *et al.*, 2011), and accordingly, viral genomes accumulate till they become detectable in these infected cells (Brady and Bernstein, 2004). The finding that herpes sequence could be detected in 28.77% of oral verrucous lesions while those lesions did not harbour any HPV infection shows further the value of this method. Herpes infection may not be the cause of this disease, but future studies of a similar nature may reveal previously unsuspected oncoviruses to be common in a different tumour type. The fact that the read depth here is enough to detect 1 HPV copy per cell with 95% confidence in most of the samples, combined with the previous ease of detecting HPV in oropharangeal cancer using the same method (Conway *et al.*, 2012) confirms that the failure to detect HPV in these samples is not a technical error, but a real biological finding.

In conclusion, the results of this study suggest that oral verrucous lesions are not associated with HPV or any other human virus. The data and findings of this chapter were published in: Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology Journal (Samman *et al.*, 2014).

Chapter 6 Revealing the transcriptional events that occur in OVC compared to OSCC.

6.1 Introduction

Oral and oropharyngeal SCCs are very invasive and prone to metastasises at the later stages and considered a threat to an individual's life (Jemal *et al.*, 2009). Several molecular studies suggest a link between the accumulation of genetic changes at DNA and RNA levels in SCC initiation and development (Gibb *et al.*, 2011a), (Zhang *et al.*, 2013) . Genomic changes such as CN alterations and point mutations, gene expression changes, as well as epigenetic changes have been discovered previously in OSCC, which could help in the development of biomarkers and assist in clinical choices and decisions (Tuch *et al.*, 2010), (Gibb *et al.*, 2011a), (Zhang *et al.*, 2013). Furthermore, tumours in different locations inside the oral cavity could differ in their clinical presentations, outcomes, and consequently, in their expression profile (Rautava *et al.*, 2007), (Ye *et al.*, 2008a), (Sajnani *et al.*, 2012).

In comparison to OSCC, OVC grows slowly, has a better prognosis, behaves less aggressively, and despite being locally invasive, does not metastasise. Accordingly, OVC has more conservative treatment than OSCC, (Strojan *et al.*, 2006), (Walvekar *et al.*, 2009), and therefore, a correct diagnosis of OVC is essential. However, this is often difficult, and the rate of preliminary misdiagnosis is high, mainly from small biopsy samples (Orvidas *et al.*, 1998), (Odar *et al.*, 2012a). In addition, OVC molecular background is still not clear, and related molecular studies are limited (Odar *et al.*, 2012a) (Odar *et al.*, 2012b). Consequently, the identification of additional biomarkers to assist in the diagnosis of OVC is an important aim.

Tumours FFPE tissue blocks from surgical resections and biopsies with their clinical data records exist in the pathology archives. This material can be used for transcriptomic profiling analyses and consequently assist in identifying clinical biomarkers in statistically well-powered studies (Lewis *et al.*, 2001), (Cronin *et al.*, 2004), (Sinicropi *et al.*, 2012). Remarkable recent advances in

sequencing technology (NGS) is providing massive data volumes that could identify genetic variations in individuals' genomes Using FFPE samples (Sinicropi *et al.*, 2012), (Zhang *et al.*, 2013),it has been demonstrated from the application of older methods, such as RT-PCR and DNA microarray that gene expression profiles (RNA transcripts levels) can categories patients and predicts their outcomes in a range of different diseases, providing new insights for many significant clinical tests (Mehra *et al.*, 2007), (Mehra *et al.*, 2008), (Chudova *et al.*, 2010). Although significant gene expression differences have been previously identified in oral cancers using microarray analysis, (Ye *et al.*, 2008a), (Estilo *et al.*, 2009), (Han *et al.*, 2009), this technique has a limited sensitivity in analysing the transcriptome (Tuch *et al.*, 2010).

As an alternative, RNA-Seq is a recently developed, deep sequencing technology, which is now widely used for transcriptomic profiling because of its reasonable costs (Zhang *et al.*, 2013). When compared with other older technologies like microarray, RNA-Seq provide much more accurate measurements of gene expression levels and more advanced categorisation of transcript isoforms (Mortazavi *et al.*, 2008), (Wang *et al.*, 2009c). Additionally, RNA-Seq is a cost-effective and efficient method to study genomic variations, such as gene fusions (Cloonan *et al.*, 2008), (Gregg *et al.*, 2010b), (Gregg *et al.*, 2010a), or somatic mutations in transcribed regions (Morozova *et al.*, 2009), (Tuch *et al.*, 2010), (Cirulli *et al.*, 2010), (Kridel *et al.*, 2012). This has improved the possibility of characterising different tumours at a molecular level across the whole genome.

6.2 Aims

Changes in transcriptional events drive the cellular phenotype of any tumour. This is the first study to date that aims to investigate the transcriptional changes that occur in OVC and compare them with the transcriptional changes that occur in OSCC using next generation RNA sequencing (RNAseq) on FFPE extracted RNA, which could henceforth aid in the histopathological diagnosis and treatment choices of OVC.

6.3 Results

6.3.1 Characteristics of the study cohort

From the verrucous cohort described in chapter 3, pathological material from each case was available in the form of FFPE blocks. In total, 13 OVC samples were identified and selected for this study. Twelve OVC samples were successfully prepared for CN analysis and RNAseq, and one OVC sample failed in RNAseq library preparation. Patients' clinical characteristics are given in Table 3.1, chapter 3. FFPE blocks were used from patients with surgery between 2004 and 2013 (median age of blocks: 5 years).

6.3.2 High throughput transcriptome sequencing identifies differentially expressed genes (DEGs) in OVC

The high throughput RNA-Seq data from OVC and adjacent normal tissues transcriptome using Illumina HiSeq 2500 NGS technology has successfully identified a list of significant differentially expressed transcripts. Gene transcription profiles were generated for all twelve OVC, with an average of 51412094 reads (ranging from 25602577 to 569763697), and with a median of 82% mapped reads. Ribosomal RNA ranged between 0% - 0.1% of the total reads (Table 6.1). Gene expression was quantified as FPKM (Fragments per Kilobase per million Mapped) for each protein-coding and non-coding gene. The threshold of 0.1 FPKM was used to determine whether or not a gene was expressed.

Table 6.1 OVC RNA alignment statistics.

Sample	Mapped Reads	Unmapped Reads	Total Reads	% Mapped	Fraction Ribosomal Bases	Fraction Intronic Bases	Fraction Intergenic Bases	Fraction mRNA Bases	Fraction Correct Strand
V-04-1-N	62383324	10228920	72612244	85%	0.06%	45.70%	34.65%	19.59%	97.05%
V-04-1-T	84774453	14533767	99308220	85%	0.00%	19.36%	76.11%	4.53%	92.52%
V-104-N	47348390	10079920	57428310	82%	0.00%	54.08%	23.71%	22.20%	96.96%
V-104-T	105360467	18276356	123636823	85%	0.00%	41.33%	32.82%	25.86%	97.63%
V-109-N	31962699	11114310	43077009	74%	0.00%	53.75%	16.86%	29.39%	96.79%
V-109-T	40430525	7113360	47543885	85%	0.00%	62.46%	16.68%	20.86%	96.70%
V-112-N	38212782	13824021	52036803	73%	0.00%	41.79%	49.73%	8.48%	87.88%
V-112-T	38726544	9482215	48208759	80%	0.00%	60.86%	24.55%	14.58%	93.21%
V-114-N	21306577	16198190	37504767	56%	0.01%	45.58%	44.78%	9.64%	93.24%
V-114-T	24753101	8894992	33648093	73%	0.01%	53.90%	32.02%	14.08%	95.23%
V-116-N	50551731	12262156	62813887	80%	0.00%	49.52%	29.16%	21.32%	95.95%
V-116-T	47935836	10513329	58449165	82%	0.00%	46.71%	35.54%	17.75%	95.71%
V-117-N	20048100	24757102	44805202	44%	0.00%	14.07%	84.94%	0.99%	68.47%
V-117-T	16468284	23438878	39907162	41%	0.00%	17.52%	81.39%	1.09%	61.87%
V-119-N	40422411	10722092	51144503	79%	0.00%	24.33%	72.29%	3.38%	78.24%
V-119-T	33147501	9895621	43043122	77%	0.00%	49.92%	36.37%	13.71%	93.73%
V-123-N	29044393	4932284	33976677	85%	0.03%	53.86%	24.66%	21.45%	96.74%
V-123-T	77380459	11964620	89345079	86%	0.05%	49.47%	28.12%	22.36%	97.49%
V-125-N	100644390	10981560	111625950	90%	0.10%	38.71%	41.57%	19.62%	93.86%
V-125-T	496636528	73127169	569763697	87%	0.06%	23.03%	61.82%	15.10%	95.14%
V-14-1-N	22204734	3397843	25602577	86%	0.07%	50.30%	32.01%	17.62%	91.54%
V-14-1-T	54734939	11888207	66623146	82%	0.00%	41.05%	53.58%	5.37%	82.82%
V-29-2-N	44494205	7185480	51679685	86%	0.09%	47.67%	34.83%	17.41%	92.67%

From the bioinformatics analysis of the generated RNAseq data, generated for the DEGs between OVC and the adjacent normal t for all gene types (i.e. protein-coding, small nuclear RNA (intergenic non-coding RNAs (lincRNAs) & pseudogenes), and th protein-coding genes only. To indicate the significant differenti genes in both lists, I ranked them by the adjusted p-value (p.a genes were selected with p.adj ≤ 0.01. From the all gene type significant differentially expressed genes were identified based value (Table 6.2). Similarly, and based on the cut off value, four differentially expressed genes were identified from the protein-co (Table 6.3).

From what is shown in Tables 6.2 and 6.3 below, for the differen analysis on the 12-matched normal vs. OVC samples, the first tw for gene names and types. The third column is for the log2 (Log2FC) and the final column is for the p-adjusted values. The run once for all genes (counts per million per sample for each g the data used to ascertain the p-value and can be compared between samples). Then run again separately per gene type kilobase per million mapped reads expression values, and compared between both genes and samples).

As can be seen from the protein-coding genes list, three protein- were significantly overexpressed in OVC compared to its ad epithelium tissue and eleven were significantly down-regulated. hand, the all gene types list included one significant overexp coding gene and one significant overexpressed pseudogene i compared to it's adjacent normal tissue. Since ribosomal RNA (r is important for successful transcriptome profiling (Peano *e* therefore , all rRNA genes were excluded from my analysis. Add protein- coding genes, and one pseudogene have been sign regulated in OVC compared to it's adjacent normal tissue. functional enrichment of the differentially expressed genes is section 6.3.2.1 below.

Table 6.2 significant differential expression list for all gene types in 12-matched normal versus OVC.

Gene Name	Gene Type	Log2 FC	P.adj
C6orf141	Protein coding	1.79	9.44E-06
UGT1A10	Protein coding	-4.30	9.44E-06
TMPRSS11B	Protein coding	-1.63	5.12E-05
ALDH3A1	Protein coding	-1.78	5.74E-05
EEF1A1P3	Pseudogene	7.58	0.000301
PAX9	Protein coding	-1.51	0.000871
RNA5S16	rRNA	4.81	0.000871
TGM3	Protein coding	-1.61	0.001989
RP11-646E20.6	Pseudogene	-1.23	0.001989
CRNN	Protein coding	-2.19	0.006366
RNA5SP50	rRNA	2.83	0.006366
TMPRSS11BNL	Protein coding	-1.72	0.008187

(+ve Log fold change (FC) means overexpression in OVC when compared to its normal).

Table 6.3 significant differential expression list for protein-coding genes in 12-matched normal versus OVC.

Gene Name	Gene Type	Log2 FC	P.adj
TMPRSS11B	Protein coding	-1.63	2.57E-08
UGT1A10	Protein coding	-4.35	3.85E-08
ALDH3A1	Protein coding	-1.77	9.49E-07
PAX9	Protein coding	-1.50	9.49E-07
TGM3	Protein coding	-1.62	1.66E-06
C6orf141	Protein coding	1.82	1.66E-06
CRNN	Protein coding	-2.17	0.0002131
SCO2	Protein coding	1.38	0.0005494
ADH7	Protein coding	-1.44	0.0040428
RPTN	Protein coding	-1.86	0.0044138
SERPINE1	Protein coding	1.49	0.0072266
ABCA8	Protein coding	-1.11	0.0074045
PCDHGA2	Protein coding	-3.81	0.0102131
BOC	Protein coding	-1.20	0.0102131
CDH3	Protein coding	1.22	0.0109073

(+ve Log fold change (FC) means overexpression in OVC (blue highlighted rows) when compared to its normal).

6.3.2.1 Enrichment of genes associated with the formation and biological processes of OVC (DAVID functional analysis)

The DAVID Bioinformatics Database 6.7 web server was used to assess the functional enrichment for differentially expressed genes in OVC. In this analysis, significant genes were selected based on the cut off adjusted p-value of: p.adj ≤ 0.01 as discussed above. When compared to matching normal tissue, the up regulated uniquely identified genes in OVC were categorised as up-regulated, whereas the down-regulated uniquely identified genes in OVC were categorised as down-regulated for differential enrichment analysis (+ve Log fold change (FC) means overexpression in OVC when compared to its normal). (Tables 6.2 and 6.3).

Genes that are differentially expressed between normal versus OVC are those that are dysregulated in the pathological process as a trigger or consequence of the epithelial cells growing abnormally and forming the histological and clinical appearance of OVC. Here, this significant differentially expressed gene (DEG) subset consists of 22 genes (significant DEGs (all types) in 12-matched normal versus verrucous samples). DAVID functional analysis of significantly enriched genes revealed involvement of three plasma membrane genes in response to organic substance, response to ethanol and response drug. All (*UGT1A10* and *ADH7*) except *CDH3* were significantly down-regulated at the 1% threshold in OVC. Additionally, pathway analysis also highlighted drug and ethanol response networks as a part of KEGG drug metabolism and KEGG retinol metabolism pathways. The expression profile of these ethanol-response and drug-response genes may reflect the extent of ethanol and drug exposure of the individuals in OVC cohort; however, alcohol and drug intake data were not available for patients involved in this RNAseq study. Furthermore, adherens junctions were highlighted in four plasma membrane genes in OVC and found that all (*BOC, CRNN* and *PCDHGA2*) except *CDH3* were significantly down-regulated at the 1% threshold. This shows that expression of adherens junction components is decreased in verrucous tissue, indicating a role in the disruption in normal cellular morphology and formation, which results in progression of malignant lesion.

The functional analysis of significantly enriched genes also revealed involvement of the plasma membrane *SERPINE1* gene in response to oxidative stress and regulation of angiogenesis. In addition, pathway analysis also highlighted *SERPINE1* role in the inhibition of angiogenesis and metastasis as a part of KEGG p53 signalling pathway (pathway in appendix 6.1). Here, *SERPINE1* was significantly overexpressed at the 1% threshold. The elevated expression levels of this gene might explain the development of OVC carcinogenesis. Likewise, DAVID functional analysis also revealed involvement of the plasma membrane *TGM3* gene in epithelial cells differentiation and keratinocytes differentiation. In the current study, *TGM3* was significantly down-regulated at the 1% threshold, which has been reported as well in multiple instances in HNSCC (Mendez *et al.*, 2007). Moreover, The functional analysis of significantly enriched genes also revealed down-regulation of *PCDHGA2* gene (involved in proliferation and cell death) at the 1% threshold in OVC. A previous *in vitro* study revealed that *PCDHGA2* suppresses the growth of carcinoma cells in Wilms' tumour (nephroblastoma) (Dallosso *et al.*, 2009).

6.3.3 Differential expression analysis on OVC versus OSCC

High throughput RNA-Seq data were generated for 16 OSCC samples. Caroline Conway did the work for OSCC RNA sequencing library preparation. This work was part of another study to investigate the transcriptional changes that occur between the stages of normal epithelium, dysplasia and OSCC. I used the OSCC RNA-Seq data to compare the transcriptional changes that occur in them with the transcriptional changes that occur in OVC. The differential expression analysis of OVC and OSCC using Illumina HiSeq 2500 NGS technology has successfully identified a list of significant differentially expressed transcripts.

From the bioinformatics analysis of the generated RNA Seq data, a list of all gene types was produced for the differentially expressed genes between the two cohorts (i.e. protein-coding, small nuclear RNA (snRNA), large intergenic non-coding RNAs (lincRNAs) & pseudogenes). To indicate the significant differentially expressed genes, I ranked them by the adjusted p-value (p.adj), and significant genes were selected with p.adj ≤ 0.01. From the all gene types

list, 57 significant differentially expressed genes were identified based on the cut off value (Table 6.4), including: 42 protein coding genes, two lincRNAs, three pseudogenes, and four snRNAs. All four rRNA genes were excluded from my analysis. Similarly, the first two columns are for name and type. The third column is for the log2 fold changes (Log2FC) and the final column is for the p-adjusted values. The analysis was run once for all genes and then separately per gene type.

As can be seen from the protein-coding genes list in Table 6.5, 23 protein-coding genes were significantly overexpressed in OVC when compared to OSCCs (gray highlighted rows), and 19 protein-coding genes were significantly overexpressed in OSCCs versus OVC. The inspected functional enrichment of the differentially expressed protein coding genes is described in section 6.3.3.1 below.

Table 6.4 significant differential expression list for all gene types in oral OVC versus OSCCs.

Gene Name	Gene Type	Log2 FC	P.adj
UNC45B	Protein coding	5.73176541	1.02E-09
ANKRD30BL	Protein coding	5.068125637	3.00E-09
RNA5S9	rRNA	5.38205729	8.91E-08
RNA5SP338	rRNA	5.543475785	1.89E-07
RNA5SP429	rRNA	5.728052924	1.89E-07
AC004448.5	lincRNA	3.426597361	3.11E-07
KRT2	Protein coding	3.28887873	4.25E-07
RNA5SP225	rRNA	4.948108267	2.39E-06
DLG2	Protein coding	3.142861272	3.25E-06
RP11-181C3.1	Protein coding	2.589914427	4.46E-06
KRT76	Protein coding	4.393647354	4.37E-05
LOR	Protein coding	3.696791035	0.000182642
ELOVL4	Protein coding	1.913915467	0.000341653
HTRA3	Protein coding	-1.624779995	0.000341653
PDK4	Protein coding	-2.442699908	0.000497839
SERPINB11	pseudogene	1.713018979	0.000613446
FLG2	Protein coding	4.155257725	0.000682528
MT2A	Protein coding	-2.067002668	0.000737808
HPGD	Protein coding	1.75292207	0.00078544
TNFRSF12A	Protein coding	-1.923891955	0.000790342
U6	snRNA	2.776570725	0.000816463

RNA45S5	pseudogene	3.505939826	0.000896898
IGFBP6	Protein coding	-2.386076713	0.000966215
ATP10B	Protein coding	1.718836376	0.001069055
HBB	Protein coding	3.554687453	0.001114319
SLC11A1	Protein coding	-1.677099745	0.001188788
FDCSP	Protein coding	3.705508687	0.001313018
FSTL3	Protein coding	-1.641022383	0.001395803
PTGDR2	Protein coding	2.991133602	0.001395803
U6	snRNA	1.978492893	0.001522023
DSC1	Protein coding	2.515586707	0.001634268
TCAP	Protein coding	-4.437644385	0.001634268
DDIT4	Protein coding	-2.274816537	0.002274577
U6	snRNA	2.042741149	0.002274577
INPP5F	Protein coding	2.093068765	0.002500484
PID1	Protein coding	1.742685358	0.002934709
ENTPD3	Protein coding	1.037851739	0.002934709
RP11-229C3.2	lincRNA	1.320493545	0.002934709
PLA2G4D	Protein coding	2.117301287	0.003105351
CTC-236F12.4	Protein coding	1.109388981	0.003105351
PDLIM3	Protein coding	-2.03957167	0.003434804
RP11-274B21.4	pseudogene	-1.780626402	0.003720488
C10orf53	Protein coding	2.290929173	0.004033844
THBS1	Protein coding	-2.189567433	0.005454379
CXCL5	Protein coding	-2.570674984	0.006168571
SERPINH1	Protein coding	-1.43209521	0.007136242
MT1X	Protein coding	-1.54913047	0.007136242
WNT9A	Protein coding	-1.802040084	0.007747887
RP11-1023L17.2	pseudogene	-1.526231287	0.008337272
LCE1A	Protein coding	2.964477071	0.008599762
LAMC2	Protein coding	-2.28993045	0.008730485
CYC1	Protein coding	-1.369128527	0.008958146
SMR3B	Protein coding	-3.661364151	0.009270752
U6	snRNA	1.845558757	0.009270752
HLF	Protein coding	1.349932377	0.009352683
CTGF	Protein coding	-1.972202278	0.009409324
RNU6-21	snRNA	1.965015019	0.009438553

(+ve Log fold change (FC) means overexpression in OSCCs when compared to OVC).

Table 6.5 Significant differential expression list for protein-coding genes in OVC versus OSCCs.

Gene Name	Gene Type	Log2 FC	P.adj
UNC45B	Protein coding	5.73176541	1.02E-09
ANKRD30BL	Protein coding	5.068125637	3.00E-09
KRT2	Protein coding	3.28887873	4.25E-07
DLG2	Protein coding	3.142861272	3.25E-06
RP11-181C3.1	Protein coding	2.589914427	4.46E-06
KRT76	Protein coding	4.393647354	4.37E-05
LOR	Protein coding	3.696791035	0.000182642
ELOVL4	Protein coding	1.913915467	0.000341653
HTRA3	Protein coding	-1.624779995	0.000341653
PDK4	Protein coding	-2.442699908	0.000497839
FLG2	Protein coding	4.155257725	0.000682528
MT2A	Protein coding	-2.067002668	0.000737808
HPGD	Protein coding	1.75292207	0.00078544
TNFRSF12A	Protein coding	-1.923891955	0.000790342
IGFBP6	Protein coding	-2.386076713	0.000966215
ATP10B	Protein coding	1.718836376	0.001069055
HBB	Protein coding	3.554687453	0.001114319
SLC11A1	Protein coding	-1.677099745	0.001188788
FDCSP	Protein coding	3.705508687	0.001313018
FSTL3	Protein coding	-1.641022383	0.001395803
PTGDR2	Protein coding	2.991133602	0.001395803
DSC1	Protein coding	2.515586707	0.001634268
TCAP	Protein coding	-4.437644385	0.001634268
DDIT4	Protein coding	-2.274816537	0.002274577
INPP5F	Protein coding	2.093068765	0.002500484
PID1	Protein coding	1.742685358	0.002934709
ENTPD3	Protein coding	1.037851739	0.002934709
PLA2G4D	Protein coding	2.117301287	0.003105351
CTC-236F12.4	Protein coding	1.109388981	0.003105351
PDLIM3	Protein coding	-2.03957167	0.003434804
C10orf53	Protein coding	2.290929173	0.004033844
THBS1	Protein coding	-2.189567433	0.005454379
CXCL5	Protein coding	-2.570674984	0.006168571
SERPINH1	Protein coding	-1.43209521	0.007136242
MT1X	Protein coding	-1.54913047	0.007136242
WNT9A	Protein coding	-1.802040084	0.007747887
LCE1A	Protein coding	2.964477071	0.008599762

LAMC2	Protein coding	-2.28993045	0.008730485
CYC1	Protein coding	-1.369128527	0.008958146
SMR3B	Protein coding	-3.661364151	0.009270752
HLF	Protein coding	1.349932377	0.009352683
CTGF	Protein coding	-1.972202278	0.009409324

Gray highlighted rows are for genes overexpressed in OVC, while white rows are for genes overexpressed in OSCC.

6.3.3.1 Enrichment of significant DEGs that distinguish OVC and OSCC (DAVID functional analysis and literature search)

The second functional classification using DAVID functional analysis is for significant differentially expressed protein coding genes between OVC versus OSCCs (Table 6.6). In this analysis, I tested for enrichment of Gene Ontology (GO) categories within each of these genes with a threshold of: p.adj ≤ 0.01 as discussed above. The highlighted cancer genes that have been also reported by previous cancer studies will be discussed below on an individual basis.

Table 6.6 David functional analysis for differentially overexpressed genes in OVC and OSCC cohorts.

GO category	GO term associated with the gene list	No. of genes overlapping GO term	% of total genes tested overlapping GO term	DAVID P-Value
Up-regulated genes in OVC				
Molecular function	Keratinization	4	2.9	0.02
	Keratinocyte differentiation	4	2.9	0.05
Biological Process	Keratinization	3	2.2	6.38E-04
	Keratinocyte differentiation	3	2.2	0.002
	Epidermal cell differentiation	3	2.2	0.002
	Epithelial cell differentiation	3	2.2	0.006
	Epidermis development	3	2.2	0.01
	Ectoderm development	3	2.2	0.01
	Epithelium development	3	2.2	0.02
Up-regulated genes in OSCC				
Molecular function	Pattern binding	3	1.7	0.013
	Structural molecule activity	4	2.3	0.034
	Insulin-like growth factor binding	3	1.7	3.68E-04
	Glycosaminoglycan binding	3	1.7	0.011
	Heparin binding	3	1.7	0.006
	Growth factor binding	4	2.3	2.18E-04
	Polysaccharide binding	3	1.7	0.013
	Protein complex binding	3	1.7	0.021
Biological Process	Angiogenesis	3	1.7	0.02
	Regulation of cell growth	4	2.3	0.002
	Blood vessel development	3	1.7	0.04
	Vasculature development	3	1.7	0.04
	Negative regulation of signal transduction	3	1.7	0.03
	Negative regulation of cell communication	3	1.7	0.04
	Cell migration	3	1.7	0.05
	Regulation of growth	4	2.3	0.01
	Blood vessel morphogenesis	3	1.7	0.03

In this study, protein coding differentially expressed genes consists of 42 genes (significant DEGs between OVC and OSCC). DAVID functional analysis of significantly enriched genes revealed involvement of immune response genes: *CTGF*, *CXL5*, and *SLC11a1* in the development of OSCC. All three genes were significantly overexpressed at the 1% threshold in OSCC group versus OVCs. The elevated expression levels of these genes suggest that the immune system is acting to identify and eradicate OSCC malignant transformed cells. *CXCL5* is an immune function modulator that has been found to be markedly up-regulated as well using microarray hybridization data in head and neck cancer cell line (HN12) that was obtained from a nodal metastasis (Miyazaki *et al.*, 2006). SLC11A1 is another immune related protein that was overexpressed in OSCC cohort when compared to OVCs. In a South African coloured population, expression of *SLC11A1* was found to be s associated with oesophageal cancer (Zaahl *et al.*, 2005).

In addition, CTGF (connective tissue growth factor), an immune related protein, also named (CCN2), belongs to the multifunctional protein family (CCN family) that regulates cell adhesion, angiogenesis, differentiation and migration (Brigstock *et al.*, 2003). *CCN2* is overexpressed in breast cancer and oesophageal adenocarcinoma, and higher expression levels are associated with more progressive cancer stages (Xie *et al.*, 2001), (Koliopanos *et al.*, 2002). Moreover, a study in 2008 showed that CCN2 was localised to tumour, vascular endothelium cells and stromal fibroblasts in head and neck SCC using immunohistochemistry. They also reported overexpression of CCN2 mRNA using quantitative RT-PCR procedure (Mullis *et al.*, 2008). In contrast, other studies reported that the increase of *CCN2* expression levels are associated with better survival and less severe cancer stages for several other tumours, including colorectal cancer, lung adenocarcinoma, and oesophageal SCC (Lin *et al.*, 2005), (Chang *et al.*, 2004), (Koliopanos *et al.*, 2002). Another study in 2012 indicated that high expression level of *CTGF* was associated with a better outcome and lower clinical stage in OSCC samples (Yang *et al.*, 2012). Consequently, the mixed results of the previous mentioned studies suggest that the role of CTGF in different cancer types may significantly vary, depending on

the involved tissue. Nonetheless, the role of CTGF in tumour metastasis and underlying mechanisms behind this are not fully understood.

Furthermore, The functional analysis of significantly enriched genes also revealed involvement of three cellular membrane genes *(LAMC2*, *THBS1* and *CTGF*) in cell motility and cell-cell signalling process that facilitates information transfer from one cell to another. All the three genes were significantly overexpressed at the 1% threshold in OSCC group versus OVCs. The elevated expression levels of these genes suggest the disruption in normal cellular information processing, activities, features and morphology, which might be involved in the development of OSCC. Earlier studies reported overexpression of *LAMC2* (laminin γ_2) protein coding gene mainly in the tumour invasive front, of HNSCC (Patel *et al.*, 2002), (Lindberg *et al.*, 2006). In 2007, Pyeon *et al* also found high gene expressions of *COL1A1, LAMC2 and COL4A1* in 42 OSCC samples, compared with 14 normal controls (Pyeon *et al.*, 2007). Furthermore, it has been previously found that the expressions of two genes pairs, including: *COL1A1-PADI1* and *LAMC2-COL4A1*, in independent testing sets, were specifically effective in differentiating OSCC from normal oral tissue (Chen *et al.*, 2008a). It has been suggested too that *LAMC2* alone can identify OSCC patients with a poor survival (Mendez *et al.*, 2009). Likewise, *THBS1* gene, also named (TSP1), that plays a role in cell-to-cell signalling and cell motility was *one of five genes* involved in the process of invasion and metastasis *that were also* predictive of OSCC-specific mortality (Mendez *et al.*, 2009). *THBS1* plays a role in platelet aggregation and cancer metastasis by enabling the interactions between tumour cells and platelets which facilitate the metastatic process (Bornstein, 1995). In addition, *TSP1* expression has been correlated inversely with survival rate in colon, bladder, and thyroid carcinomas but not in other cancer types (de Fraipont *et al.*, 2001). KEGG analysis revealed enrichment of *THBS1* in P53 signalling pathway.

Additionally, KEGG pathway analysis also highlighted three genes (*THBS1*, *LAMC2*, and *DDIT4*) as a part of p13K signalling pathway. All three genes were significantly overexpressed at the 1% threshold in OSCC in comparison to OVCs. Aberrations in genes in the p13K signalling pathway were frequently

found in OSCCs (Chang *et al.*, 2013). DNA damage inducible transcript 4 (*DDIT4*) has been reported by a study in 2011 to be involved in the progression of hepatocellular carcinoma (Huang and He, 2011). KEGG pathway analysis also revealed enrichment of *WNT9A* in WNT and Hedgehog signalling pathways. *WNT9A* protein coding gene was overexpressed in OSCC cohort in this study when compared with OVCs. *WNT9A* was markedly down-regulated in metastases from mice modules with p53-/+ HNC (Ku *et al.*, 2007).

Moreover, overexpression of *MT2A* and *MT1X* genes (involved in ion binding and in transcription factor regulation) at the 1% threshold in OSCC group versus OVCs suggest the malignant transformation of oral epithelial cells; since metallothioneins, including: MT2A and MT1X, have been linked with the metastasis process, but the mechanism is still unclear (Weinlich *et al.*, 2006) (Pedersen *et al.*, 2009), (Szelachowska *et al.*, 2009). Expression of metallothionein has been also associated with cell invasion and proliferation in breast cancer (Jin *et al.*, 2002), and has been also related with poor survival in colorectal tumours (Janssen *et al.*, 2002). Metallothioneins potential role in oral cancers and melanoma metastasis has also been suggested (Szelachowska *et al.*, 2009), (Weinlich *et al.*, 2006). In Skubitz *et al* study in 2012, soft tissue sarcoma samples that were in the group with the highest metastases rate overexpressed several metallothioneins, including: MT2A and MT1X (Skubitz *et al.*, 2012).

In addition, the functional analysis of significantly enriched genes also revealed involvement of two cytoskeleton keratin genes (*KRT76* and *KRT2*) in maintaining cellular structural and functional integrity process. Both genes were significantly overexpressed at the 1% threshold in OVC group versus OSCCs. The elevated expression levels of these genes reflect the epithelial origin of OVC tumours. *KRT76* protein coding gene that encodes a filament protein, which is involved in epithelial cells structural integrity, was one of the significant DEGs overexpressed in OVC samples when compared to OSCCs, with 4.4 increases in fold change. A previous study in 2010 identified down-regulation of *KRT76* in five oral cancer samples taken from smoking patients as the most down-regulated gene using next generation sequencing (Illumina Genome

analyser) (Zain *et al.*, 2010). Additionally, a recent study used a hamster model of oral cancer and reported the down-regulation of *KRT76* in this model as well as in human OSCC and oral precancerous lesion using qRT-PCR and immunohistochemistry (Ambatipudi *et al.*, 2013). *KRT2* protein coding gene, which is involved in epidermal differentiation; was another significant differentially expressed keratinocyte-related gene, overexpressed in OVC cohort when compared to OSCCs, with 3.3 increases in fold change. In Zain *et al* study, *KRT2* was also one of the down-regulated genes in the five OSCC samples (Zain *et al.*, 2010). Nonetheless, and from the same study, *KRT76* and *KRT2* were both down-regulated but to a much lesser level along with some other genes (not discussed here) in the two proximal normal samples used in their analysis and they suggested a possible existence of 'field cancerization'' theory as a result of similar pattern of gene down-regulation in the cancerous and proximal normal samples (Zain *et al.*, 2010). Figures 6.1 and 6.2 below show the boxplots for *KRT76* and *KRT2* differentially expressed genes (at 1% threshold) and the overexpression in OVC samples when compared with OSCCs (based on actual FPKM levels).

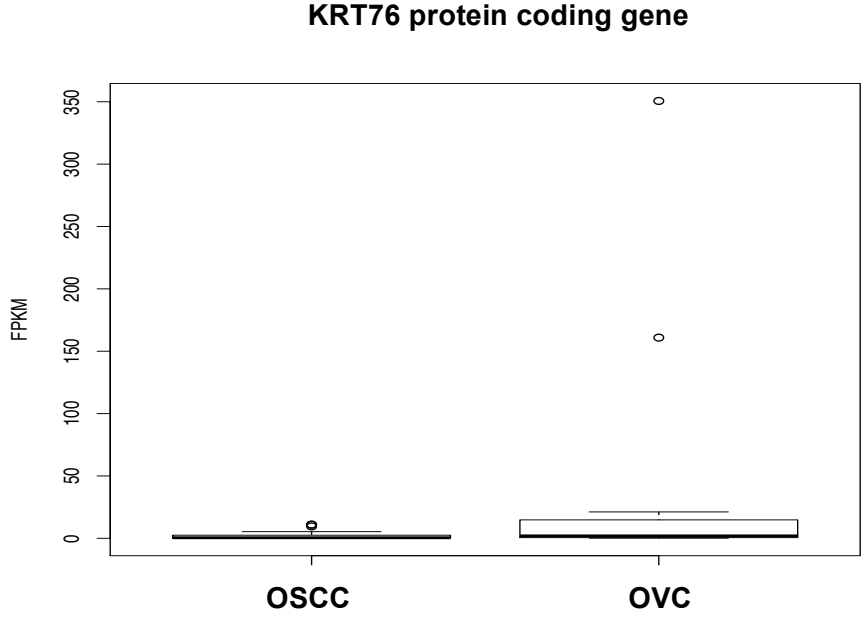


Figure 6.1 *KRT76* differential expression boxplot shows overexpression of this gene in OVC cohort in compare with OSCCs.

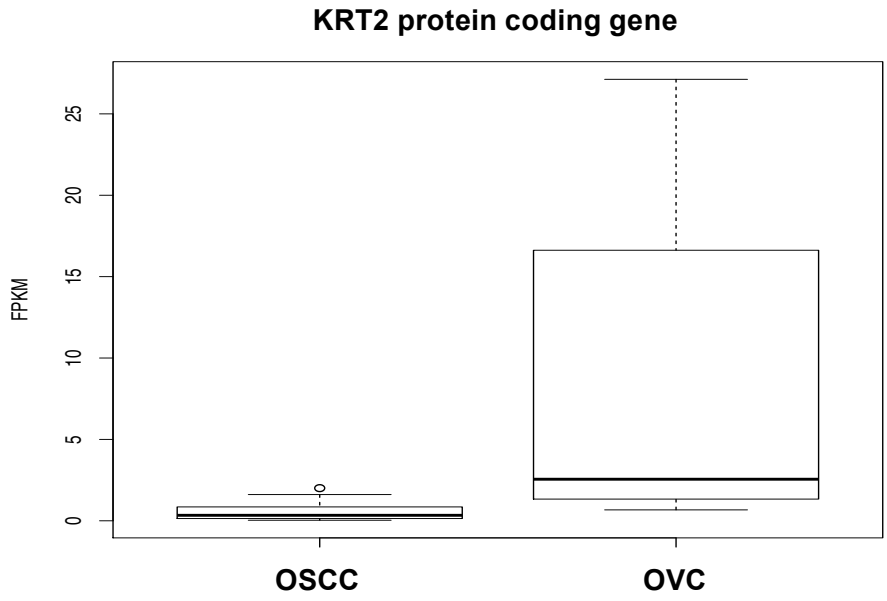


Figure 6.2 *KRT2* differential expression boxplot shows overexpression of this gene in OVC cohort in compare with OSCCs.

6.3.3.2 Functional characterization analysis of the significant DEGs on an individual basis

This section will discuss significant differentially expressed cancer genes that have been reported by previous cancer studies on an individual basis.

6.3.3.2.1 Significant overexpression of HNC genes in the OSCC cohort compared with OVCs (literature search)

The *SMR3B* (also named *PRL3*) protein coding gene was overexpressed in OSCC cohort when compared with OVCs. A study in 2011 by Hassan *et al* reported that *PRL-3* expression was significantly higher in eleven dysplasia and 50 OSCCs than in 12 normal tissues using RT-PCR method. They suggested that *PRL-3* expression have a role in oral cancer development and may be used as a useful marker of oral malignant and pre-malignant lesions (Hassan *et al.*, 2011). *TNFRSF12A* is another protein coding gene that was overexpressed in OSCC samples when compared to OVCs. A recent study in 2012 performed genome-wide expression profiling to reveal differentially expressed genes in 27 gingivo-buccal tumours with identified chromosomal changes (Ambatipudi *et al.*,

2012). A total of 315 putative driver genes were identified including the significantly up-regulated *TNFRSF12A* gene by integrating gene expression and copy number data (Ambatipudi *et al.*, 2012).

Likewise, the *FSTL3* (follistatin-like-3) protein coding gene was overexpressed in OSCC cohort in this study in comparison to OVCs. The encoded FSTL3 protein is a member of the FST-module protein family and suggested to be involved in differentiation and growth regulation (Hayette *et al.*, 1998). A previous study in 2010 reported high mRNA expression of candidate genes, including FSLT3 in 19p13.3 in selected OSCC samples (Freier *et al.*, 2010). Also, *IGFBP6* protein coding gene was overexpressed in OSCC compared to OVCs. Increase in IGFBP6 expression in oral cancers was suggested as a possible prognostic marker of tumour sensitivity that could be used for diagnosis purposes, since a significant association was observed between elevated levels of *IGFBP6* and increased induction of apoptotic cell death in these tumours (Cacalano *et al.*, 2008).

6.3.3.2.2 Significant overexpression of HNC genes in the OVC cohort compared with OSCCs (literature search)

The *DLG2* gene was overexpressed here in OVC samples in compare to OSCCs. A previous study suggested that loss or disruption of *DLG2* besides other adjacent genes might play a role in the development or progression of OSCC (Reshmi *et al.*, 2007). Similarly, *DSC1* protein coding gene was overexpressed in OVC cohort here in compare to OSCCs. High abundance of desmosomes occurs in epithelial tissue as they ensure strong adhesion between epithelial cells (Garrod and Chidgey, 2008). They include proteins originated from three gene families at least: the armadillos, desmosomal cadherins, and the plakins (Desai *et al.*, 2009). Desmosomal cadherins are divided further into desmocollins (DSC1-3) and desmogleins (DSG1-4). Desmosomal component expression is dependent upon cell type, differentiation status and position in the epidermis (Desai *et al.*, 2009). Oral mucosa histology differs from that of epidermis with noticeably more thickly squamous stratified epithelial layers in the first (Teh *et al.*, 2011). Desmosomal gene expression in oral mucosa was examined previously in comparison with epidermis and a

molecular expression maps were generated which constantly showed lack of DSC1 expression in oral mucosa (Donetti *et al.*, 2005), (Teh *et al.*, 2011).

Haemoglobin beta (*HBB*) gene was also overexpressed in OVC samples in this study in compare to OSCCs. *HBB* gene was suggested as a novel tumour suppressor gene in a study in 2005 that reported significant down-expression of this gene in eleven anaplastic thyroid cancer cell lines (Onda *et al.*, 2005). Additionally, *HPGD* protein coding gene was overexpressed here in OVC cohort in compare to OSCCs. A study in 2003 has reported down-regulation of *HPGD* along with seven more genes (not discussed here) in isolated human metastasising oesophageal SCC cell line (Kawamata *et al.*, 2003). They suggested that these genes (including *HPGD*) might control the metastasis process of oesophageal SCC and may be used as prognostic markers for oesophageal SCC lymph node metastasis.

6.3.3.2.3 Significant overexpression of other cancer genes in OSCC versus OVC (literature search)

PDLIM3 protein coding gene was overexpressed in OSCC samples in this study in compare to OVCs. *PDLIM3* gene was found to be up-regulated, expressed 2-fold higher at least in metastasis ovarian serous carcinomas versus primary ovarian serous carcinomas (Bignotti *et al.*, 2007). TCAP was also overexpressed here in OSCC cohort in compare to OVCs. A previous quantitative RT-PCR study revealed that the expression level of *TCAP* was absent or very low in 36 primary breast tumours (Kauraniemi *et al.*, 2003).

Furthermore, *HtrA3* protein coding gene was overexpressed in OSCC cohort in compare to OVCs. The expression status of this gene has not been reported in HNCs. However, a previous study revealed that the expression level of *HtrA3* was absent or reduced *in* primary lung tumours from heavy smokers and in over 50% of lung cancer cell lines (Beleford *et al.*, 2010). *PDK4* is another protein coding gene that was overexpressed in OSCC samples in compare to OVCs. The expression status of this gene has not been reported as well in HNCs. Though, an earlier study suggested that *PDK4* is the most strongly down-regulated gene in all the tested human colon cancer samples

when compared with the neighbouring adjacent healthy colon tissue samples (Blouin *et al.*, 2011).

6.3.3.2.4 Significant overexpression of other cancer genes in OVC and OSSCs (literature search)

HLF is a tumour suppresser that was overexpressed here in OVC samples in compare to OSCCs. *HLF* gene plays a role in the detoxification processes (Woenckhaus *et al.*, 2006), and was also identified to be in cancer gene census dataset. An earlier study reported consistent down-regulated of *HLF* gene in non-small-cell lung cancer samples using the Affymetrix U133A array, real-time PCR and IHC (Woenckhaus *et al.*, 2006). *INPP5f* protein coding gene was also overexpressed in OVC compared to OSCCs. An earlier study suggested that *INPP5f* have a potential role as a suppresser gene in prostate cancer (Ribarska *et al.*, 2012).

6.3.3.3 Principal Component Analysis (PCA) of expressed genes in OVC and OSCC

PCA was performed for all expressed genes using the prcomp function in R. From the produced biplots, plot of PC1 and PC2 resulted in an evident separation of OVC and OSCC tumour samples (Figure 6.3).

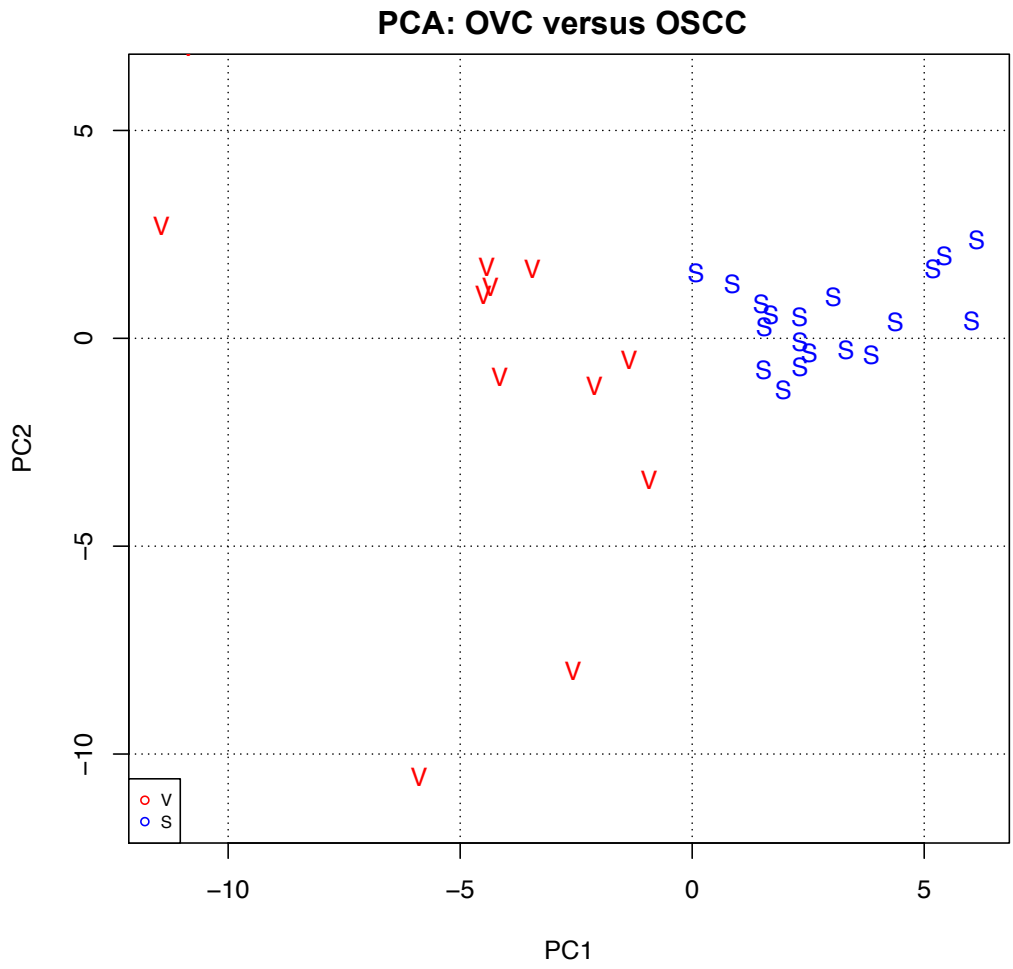


Figure 6.3 Principal Component Analysis (PCA) biplot of PCs 1 and 2 using all expressed genes.

This biplot best separates the two oral tumour groups: V –Verrucous carcinoma, and S – Squamous cell carcinoma.

6.3.4 Integration of copy number analysis and RNAseq gene expression data

To highlight novel genes of possible clinical and biological importance in OVC, the list of genes generated from the copy number analysis data (chapter four, Table 4.3) was integrated with genes from the significant differential expression lists generated from RNAseq analysis (Tables 6.3 and 6.4). The integrated

analysis identified two protein-coding genes: *SERPINE1* and *CDH3*, located in recurrent regions of gain: chromosomes 7q21.3 and 16q22.1, in OVC, with a frequency of ~50% and ~25% of CN changes respectively (Table 6.7). *SERPINE1* was significantly overexpressed at the 1% threshold. The elevated expression levels of this gene could explain the development of OVC carcinogenesis. In addition, *CDH3*, a P-cadherin gene associated with cell-to-cell signalling; was significantly overexpressed at the 1% threshold in OVC. Down regulation of this gene has been previously found in lymph node metastases from oral cancer (Mendez *et al.*, 2009). Similarly, *PLA2G4D* protein coding gene, located in a recurrent region of gain: chromosome 15q15.1, with a frequency ~30% of CN changes (Table 6.7) *was* overexpressed at the 1% threshold in OVC group versus OSCCs.

Table 6.7 genes from the significant differential expression lists, with elevated RNA expression levels, and gain of their corresponding genomic location with the identified copy number aberrant region in OVC.

Gene name	Cytoband	CN change	% of samples with CN change	RNA expression	Log FC	P.adj
SERPINE1	7q21.3	Gain	~50%	Overexpression in OVC	1.49	0.0072
CDH3	16q22.1	Gain	~25%	Overexpression in OVC	1.22	0.0109
PLA2G4D	15q15.1	Gain	~30%	Overexpression in OVC when compared with OSCCs	2.12	0.0031

6.4 Discussion

Clinically, OVC appears as exophytic masses with a verrucous surface and, less often, they may become relatively smooth. However, histological features can be misclassified (Devaney *et al.*, 2011b). For this reason, new experimental approaches are needed to evaluate quantitative biological biomarkers that are required for better understanding of the development and progression of OVCs. This study represents the largest and 'first' study, to date, to inspect the transcriptional changes that occur in OVC and compare them with the transcriptional events in OSCC by performing high coverage RNAseq using strand-specific approaches that capture transcriptional information about coding and non-coding RNA > 200bp.

I present here in Table 6.8 a systematic review that was conducted on OVC aiming to combine the literature on different comparative expression studies. Previous molecular investigations have yielded mixed results. The rarity of verrucous lesions (including OVC and OVH) also makes them difficult to explore. Using different samples, sample numbers, variations in clinical diagnosis, difficulties in defining 'gold-standard' histological criteria for diagnosing verrucous lesions, and different staining procedures and analysis methods may explain the lack of concordance between these studies. Importantly, and after considering all the selected biomarkers tested previously in OVC and OVH (Table 6.7), the genes highlighted as significant DEGs in the current study did not overlap with those listed in the systematic review.

Table 6.8 A systematic review for different comparative expression studies on oral verrucous lesions.

Biomarkers	Verrucous sample numbers	Results	Technique	Reference
p53, Rb, and cyclin D1	29 OVC	p53 accumulation was repeatedly observed in OVC suggesting possible gene mutations. cyclin D1 overexpression and no Rb staining alterations were also observed.	Immunohistochemistry	(GimenezConti et al., 1996)
p16, pRb, p53, p27 and Ki-67	15 OVC	A high expression of pRb and p16 in OVC is relatively different compared with OSCC, which may suggest a potential association between OVC and HPV infection.	Immunohistochemistry	(Saito *et al.*, 1999a)
c-erbB-3	31 OVC and 18 OVH	c-erbB-3 expression was a malignancy index through progression from OVH to OVC and OSCC arising from VC.	Immunohistochemistry	(Sakurai et al., 2000)
iNOS	20 OVC and 20 OVH	Cytoplasmic staining was mostly found in OVH (80%). Nuclear and/or cytoplasmic stainings were observed in OVC (85%).	Immunohistochemistry	(Chen *et al.*, 2002b)
p53, EGFR, TGF-α, and Cyclin D1	26 H&N VC and four H&N VH	p53 was seen in 88% of VC and 100% of VH. EGFR expression was found in 54% of VC and 25% of VH. TGF-α was noted in 88% of VC and 25% of VH. Cyclin-D1 expression was detected in 35% of VC and 75% of VH.	Immunohistochemistry	(Wu *et al.*, 2002b)
MMP-7, MMP-9, MMP-10 MMP-12, MMP-13, MMP-19 and MMP-26	15 OVC and16 OVH	Compared with OSCC, OVC lack epithelial MMP-7, -9, and -12 expression.	Immunohistochemistry, ISH	(Impola et al., 2004)
Cytokeratins 10, 13, 14 and 16	Eight OVC	OVC cytokeratins profile was similar to OSCC cytokeratins profile reported in the literature.	Immunohistochemistry	(Oliveira et al., 2005)
Cyclin D1	30 OVC	Overexpression of cyclin D1 in OVC was similar to OSCC.	Immunohistochemistry	(Angadi and Krishnapillai, 2007)

laminin, laminin-5, collagen IV and fibronectin	20 OVC	Laminin staining pattern was more intensive in OVC compared with OSCC.	Immunohistochemistry	(Arduino et al., 2010)
p53, MDM2, p21, HSP 70, and HPV 16/18 E6	48 OVC and 30 OVH	These biomarkers could not differentiate OVC from OVH. Low p21 expression might explain epithelial abnormal overgrowth in OVC and OVH. MDM2 overexpression and HSP 70 moderate expression may explain the pathogenesis of both verrucous lesions.	Immunohistochemistry	(Lin et al., 2011)
type IV collagen and laminin-332 Y2	15 OVC	Type IV collagen immunohistochemical study did not define if a lesion is non-invasive or invasive. Assessment of Ln-332 Y2 chain expression could be useful as a marker for diagnosis and biological differences of OVC from OSCC.	Immunohistochemistry	(Zargaran et al., 2011a)
microRNAs: miR-21, miR-31, miR-203, miR-125a-5p and miR- 125b and PTEN and p63	30 VC of the H&N	miR-21 and miR-31 were overexpressed in VC and SCC. Levels of miR-203 were unaltered in SCC but elevated in VC. Levels of miR- 125a-5p and miR-125b were unaltered in SCC but reduced in VC. p63 was up-regulated in SCC BUT down-regulated in VC. PTEN was down-regulated in SCC and VC.	Immunohistochemistry and RT-PCR	(Odar et al., 2012a)
αB-crystallin	17 OVC	αB-crystallin expression was noticed in normal mucosa, OVC, and OSCC. (Gradual expression increase from normal to OSCC).	Immunohistochemistry	(Quan et al., 2012)
Ki67	15 OVC	Ki67 was not a reliable marker to evaluate invasion level and differentiate OVC from OSCC.	Immunohistochemistry	(Zargaran et al., 2012)
p53, Ki-67, MMP-2 and MMP-9	20 OVC	Ki-67 was not a good grading marker. OVC less invasive nature in compare to OSCC might be associated with p53, MMP-2, and MMP-9 expression patterns.	Immunohistochemistry	(Mohtasham et al., 2013)
p53, Ki67 and HuR	17 OVC and six OVH	Increased Ki67 and p53 positivity in OVC, while OVH presented rare positive signals. HuR diffuse staining pattern and epithelium expression were observed.	Immunohistochemistry	(Habiba et al., 2014)

The high throughput RNA-Seq data from OVC and adjacent normal tissues transcriptome has successfully identified a list of significant differentially expressed transcripts. However, one of the main limitations in this part of the project was extracting RNA from histologically normal epithelium adjacent to the verrucous malignant area to be used as "normal-matching samples" since no other source of 'normal' material was available for all patient (i.e. blood samples). Occult genetic abnormalities can be harboured within normal-appearing epithelium of cancer patients (Tripathi *et al.*, 2008), and hence, if an expression has already occurred in a certain gene in 'normal' adjacent epithelium, it will not be detected in malignant OVC tissues. Other limitations were the computational difficulties associated with the bioinformatics data analysis and the high costs of the library preparation reagents and kit as well as the high sequencing prices. Also, replication (repetition of the same experimental process using the same sample) is an important step for assessing and decreasing experimental errors. The low extracted RNA amounts for the samples here, which could not be divided for repetition, limited the option of studying biological replicates. Furthermore, in some RNAseq studies, RT-PCR or qPCR has been used for the validation of some of the DEGs identified (Camarena *et al.*, 2010), (Feng *et al.*, 2010). Again, this has not been the case here because of the limited time available for this project. However, the integrated analysis with the CN analysis data and the fact that it showed a correlation between the expression profiles of some genes located within the same CN altered region could be used as a sort of validation.

6.4.1 Implications of DEGs in the current study

Clustering via principle component analysis, using data from expressed genes, demonstrated an evident separation of OVC and OSCC cohorts (Figure 6.3). The use of DAVID (Dennis *et al.*, 2003) functional analysis revealed that significant differentially upregulated genes in OVC versus OSCC are involved in keratinocyte differentiation and epithelium development, while the significant differentially upregulated genes in OSCC versus OVC were more involved in cell growth and migration and angiogenesis (Table 6.6). Significant differences in the expression profiles between OVC versus its adjacent normal epithelium

and versus OSCC suggest the association of some gene candidates with the pathogenesis of OVC. These genes are described in the subsections below. Nonetheless, it must be remembered that these suggested genes, at best, can only report on the likelihood of the development of OVC and requires further investigation.

6.4.1.1 Genes that may have a role in the development of OVC

Specific consistent transcriptional changes and events that occur in OVC drive its phenotype. The differential expression analysis on the 12-matched normal versus OVC samples produced a list consists of 22 significant DEGs. Further functional characterisation of the significant genes on an individual basis has highlighted five HNC genes in OVC (*PAX9*, *TGM3*, *CRNN*, *ADH7*, and *SERPINE1*), in which the expression status of those genes in OVC cohort here matched with their expression status in HNCs.

PAX9 is a transcription factor that regulates the expression of cell proliferation genes, resistance to migration and apoptosis and was significantly down-regulated at the 1% threshold in OVC. In a total of 35 dysplastic lesions and 36 invasive carcinomas of the oesophagus, *PAX9* was a significant marker for down-regulated differentiation of oesophageal keratinocytes (Gerber *et al.*, 2002). *PAX9* expression was significantly reduced or lost in the majority of epithelial dysplasias and invasive carcinomas (Gerber *et al.*, 2002).

TGM3 (transglutaminase family) is normally seen in late differentiating squamous epithelium (Kim *et al.*, 1993), and was significantly down-regulated here at the 1% threshold in OVC. The down-regulation of *TGM3* is consistent with the loss of differentiation and could interrupt a potential important step to apoptosis in HNSCC, and enhance tumour cell survival. *TGM3* has also been shown to be down-regulated in oesophageal cancer (Chen *et al.*, 2000), and in HNSCC cell lines (Gonzalez *et al.*, 2003). Similarly, *CRNN* functions as a tumour suppressor and as a barrier in squamous epithelium in response to injury and was significantly down-regulated at the 1% threshold in OVC. It was demonstrated before that *CRNN* acts as a potential tumour suppressor in oesophageal SCC (Chen *et al.*, 2013).

ADH7 gene was significantly down-regulated at the 1% threshold in OVC, which may explain the development of this malignant lesion (alcohol intake data were not available for patients in this study). Alcohol is a critical risk factor for cancers of the upper aerodigestive tract and is mainly metabolised by *ADH* enzymes (alcohol dehydrogenase enzymes). Six *ADH* genetic variants were investigated in over 5,200 controls and 3,800 aerodigestive cancer cases from three separate studies. In each individual study*, ADH7* gene variant was significantly protective and acted as a suppresser against aerodigestive cancer. Additionally, these protection features became more apparent with more alcohol consumption levels (Hashibe *et al.*, 2008).

On the other hand, the plasma membrane *SERPINE1* gene was significantly overexpressed at the 1% threshold in OVC. The elevated expression levels of this gene might explain the development of OVC carcinogenesis. The integrated analysis also identified *SERPINE1*, located in recurrent regions of gain on chromosomes 7q22.1, with a frequency of ~50% of CN changes in OVC. In 2005, a study revealed that expression of *SERPINE1* gene in primary HNCs was up-regulated in compare to normal mucosa (Chin *et al.*, 2005). *SERPINE1* overexpression was shown to be essential for the progression of HNSCC and was suggested to play a key role in chromosome 7q21.3–22 karyotypic changes and in oral oncogenesis (Chen *et al.*, 2004b). *SERPINE1* plays an important role as a primary inhibitor of plasminogen activators in tumorigenesis and invasion (Ju *et al.*, 2010). However, the responsible mechanism for *SERPINE1* up-regulation in OVC is still not clear.

6.4.1.1.1 Overexpression of keratin genes in OVC

KRT76 and *KRT2* protein coding genes were significantly overexpressed at the 1% threshold in OVC versus OSCCs. Keratins are proteins that form epithelial cells filaments and are needed to maintain normal tissue function and structure (Schweizer *et al.*, 2006). Large genes family clustered at two different chromosomal sites encodes keratins: 17q21.2 for keratins type I (except keratin18), and 12q13.13 for keratins type II, including keratin 18 (Ambatipudi *et al.*, 2013). These also play a key role in protecting epithelial cells from injury

and non-mechanical and mechanical stress (Moll *et al.*, 2008), (Karantza, 2011). Keratins are expressed continuously in epithelial tumour cells, which are also a characteristic of their origin site, and hence; keratins are widely used as an immunohistochemical pathology diagnostic tumour marker (Moll *et al.*, 2008), (Karantza, 2011). Earlier studies have reported variations in keratin expression throughout oral carcinogenesis (Xu *et al.*, 1995), (Gires *et al.*, 2006), (Matthias *et al.*, 2008), (Wei *et al.*, 2009). Moreover, several keratins are reported as independent OSCC prognosis markers (Fillies *et al.*, 2006), (Yanagawa *et al.*, 2007). A complex keratin expression pattern in the oral cavity reflects both, the epithelium type and differentiation stage specific expression (Ambatipudi *et al.*, 2013).

The oral epithelial basal proliferative layer expresses keratin 5, keratin 14 and keratin19. The differentiating suprabasal keratinised epithelial layers express keratin 1 and keratin 10, whereas the non-keratinised epithelial differentiating layers such as the oesophagus and the buccal mucosa produce mainly keratin 4 and keratin 13. Epithelial suprabasal cells of the gingiva and the hard palate express keratin 6, keratin 16, and keratin 76 (Presland and Dale, 2000), (Chu and Weiss, 2002), (Bragulla and Homberger, 2009). Previous studies on oral cancers have described changes in keratin expression patterns and terminal differentiation such as down-regulation of keratin 4, keratin 5, keratin 13 and keratin 19 (Vaidya *et al.*, 1996), (Crowel *et al.*, 1999), (Ohkura *et al.*, 2005) (Yanagawa *et al.*, 2007), (Ambatipudi *et al.*, 2012). Similarly, over-expression of keratin 8/ keratin 18, keratin 17 and keratin 14 have been also reported in oral cancers when compared to normal tissues (Ohkura *et al.*, 2005), (Fillies *et al.*, 2006), (Gires *et al.*, 2006), (Toyoshima *et al.*, 2008), (Wei *et al.*, 2009).

Remarkably, changes of keratin expression points toward the common signature in human oral tumours and experimental oral cancers established in animal models (Ambatipudi *et al.*, 2013). The observed reduced expression of *KRT76* in Ambatipudi *et al* study of 159 gingivo-buccal cancer samples showed a similar trend in the buccal epithelium of DMBA treated hamsters used in the same study. They also reported a strong association of increased risk of oral precancerous lesions and OSCC development with reduced KRT76 expression

(precancerous lesions cases were 61 leukoplakia hyperplastic lesions with focal mild to moderate dysplasia) (Ambatipudi *et al.*, 2013). Accordingly, and since *KRT76* and *KRT2* were reported to be down-regulated in OSCC (Zain *et al.*, 2010), (Ambatipudi *et al.*, 2013), this suggests the potential role of both genes in the development of OVC lesions.

6.4.1.2 Genes that may explain the indolent behaviour of OVC

The distinction between OVC and OSCC is one of the most common problems in oral cancer pathology, and the literature on these borderline tumours is confusing, mainly with regard to their histopathological diagnostic features and treatment options. The clinical behavior of OVC is usually indolent and generally benign. However, till now, the characteristics of OVC were only described and reviewed by the clinicopathological parameters, and there were no available well-defined gene expression profile for this tumour. In the comparative analyses study here of the significant DEGs on an individual basis, eight genes (*DLG2, HBB, TNFRS12A, IGFBP6, FSTL3, DSC1, LAMC2* and *SMR3B*) were highlighted in OVC suggesting its indolent behaviour based on their function in other HNC tumours. *DLG2, HBB,* and *DSC1* genes were overexpressed in OVC compared to OSCCs. Down-regulation or loss of those genes has been reported previously to be associated with the development or progression of HNCs (Onda *et al.*, 2005), (Reshmi *et al.*, 2007), (Teh *et al.*, 2011). Hence, the overexpression *DLG2, HBB,* and *DSC1* genes at the 1% threshold in OVC may explain its more benign behaviour & better prognosis.

On the other hand, *TNFRSF12A*, *IGFBP6, FSTL3, SMR3B* and *LAMC2* protein coding genes were overexpressed at the 1% threshold in OSCC group versus OVCs. Overexpression of those genes has been previously identified and reported to be associated with HNCs (Patel *et al.*, 2002), (Lindberg *et al.*, 2006), (Pyeon *et al.*, 2007), (Cacalano *et al.*, 2008), (Freier *et al.*, 2010), (Hassan *et al.*, 2011), (Ambatipudi *et al.*, 2012). Since overexpression of those genes was associated with HNCs that are considered more aggressive, metastatic tumours in compare to OVC; this suggests a potential involvement of those genes behind the more benign behaviour of OVC.

6.4.1.2.1 Overexpression of tumour suppressor genes in OVC

HLF, HBB and *INPP5f* protein coding genes were significantly overexpressed at the 1% threshold in OVC group in the current study versus OSCCs. *HLF* and *INPP5f* genes have been previously reported as potential suppressor genes in non-small-cell lung cancers (Woenckhaus *et al.*, 2006), and prostate cancers (Ribarska *et al.*, 2012) respectively. Similarly, *HBB* protein coding gene was suggested as a novel tumour suppressor gene in anaplastic thyroid cancer cell lines (Onda *et al.*, 2005). Accordingly, the overexpression of those tumour suppressor genes in OVC may explain the more benign behaviour in these tumours and its good prognosis.

6.4.1.3 Genes that may explain the non-metastatic characteristic of OVC

An important feature associated with OVC is that it is a non-metastatic oral cancer. If the presence of metastasis biomarkers were to contribute to the prognostic factors and distinguish between non-metastasizing OVC and metastasizing OSCC, this would provide important biological information on the metastatic process and may be useful in identifying patients who might benefit from certain treatment options or additional therapies. In this study, the functional characterization analysis of the significant DEGs on an individual basis has highlighted two genes (*CDH3* and *HPGD*), which play a role in cancer metastases. *CDH3*, a P-cadherin gene associated with cell-to-cell signalling was significantly overexpressed in OVC. Furthermore, the integrated analysis identified this gene, located in recurrent regions of gain on chromosomes 16q22.1, with a frequency of ~25% of CN changes in OVC. Importantly, down regulation of *CDH3* has previously been found in lymph node metastasises from oral cancer (Mendez *et al.*, 2009), suggesting that the expression of this gene might be a reason behind metastasis inhibition in OVC. Similarly, *HPGD* protein coding gene was significantly overexpressed at the 1% threshold in OVC versus OSCC. However, down-regulation of *HPGD* was proposed along with seven more genes as prognostic markers for oesophageal SCC lymph node metastasis (Kawamata *et al.*, 2003). Accordingly, it is proposed that the

overexpression of *HPGD* gene here is associated with the non-metastatic behaviour of OVC.

On the other hand, *CXCL5*, *THBS1, MT2A* and *MT1X* protein coding genes were overexpressed at the 1% threshold in OSCC group versus OVCs. Overexpression of the *CXCL5* gene has been previously identified and reported to be associated with HNC nodal metastasis (Miyazaki *et al.*, 2006), overexpression of *THBS1 gene* has been previously reported to be involved with the process of invasion and metastasis in OSCC (Mendez *et al.*, 2009), and overexpression of *MT2A* and *MT1X* genes has been linked with soft tissue sarcoma metastasis (Skubitz *et al.*, 2012). The overexpression of those four genes (that were previously reported to be involved in cancers metastasis) in OSCC cohort when compared with OVCs suggests that they might have a role in the absence of the metastatic behaviour in OVC tumours.

6.4.2 Interpretation of the generated RNAseq results in the current study and a recently published gene profiling analysis paper for patients with OVC and OSCC

During the writing of this chapter, a gene profiling analysis paper was published for patients with OVC and OSCC. In their study, six patients with primary OSCC and five patients with primary OVC including their adjacent matched normal oral mucosa tissue samples were profiled using the HGU133 Plus 2.0 Affymetrix microarray GeneChip platform (Wang *et al.*, 2014b). They found 167 DEGs between OSCC and OVC, 39 genes from the 167 were common DEGs between OSCC and OVC when compared with their adjacent matched normal samples, and eight from the 39 were DEGs between OSCC and OVC (Wang *et al.*, 2014b). The eight genes were *HLF*, *TGFBI*, *SERPINE1*, *MMP1*, *INHBA*, *COL4A2*, *COL4A1*, and *ADAMTS12*. They proposed that the eight genes might differentiate and determine the identity of the two tumours and further studies on much larger samples must be conducted. They also suggested that their results indicate that OVC is a rare variant of OSCC but with different molecular point of view than OSCC (Wang *et al.*, 2014b).

However, five OVC and six OSCC samples were used in Wang *et al* study, while here, I used 12 OVC and included 16 OSCC samples for the performed RNAseq analysis (primary OSCC data belong to the Pre-cancer Genomics Group). Also, and unlike microarray gene expression technique in Wang *et al* study, RNAseq allows measuring gene expression levels of both novel and known RNA transcripts and transcript isoforms with higher dynamic range than microarray, and with a capability to detect low abundance transcripts, if sufficient sequencing depth was provided (Zwemer *et al.*, 2014). Numerous earlier studies comparing both approaches have found RNAseq to be similar to or more sensitive than microarrays when used for differential expression gene analysis of cellular RNA, particularly for high expressed genes (Marioni *et al.*, 2008), (Su *et al.*, 2011), (Sirbu *et al.*, 2012).

Nonetheless, the eight genes (*HLF*, *TGFBI*, *SERPINE1*, *MMP1*, *INHBA*, *COL4A2*, *COL4A1*, and *ADAMTS12*) that were proposed to differentiate between OVC and OSCC in Wang *et al* study were all differentially expressed between OVC and OSCC cohorts in this study as well. Furthermore, all the eight genes were overexpressed in OSCC samples versus OVCs in Wang *et al* study except for *HLF* gene*,* and the same expression status was obtained here from the generated RNAseq results. *HLF* gene was also the only "significant" DEG (with p.adj ≤ 0.01) that was overexpressed in OVC cohort in compare with OSCCs. Additionally, *SERPINE1* was "significantly" overexpressed in OVC when compared to the matched normal samples, and was also located in a recurrent copy number region of gain (chromosomes 7q21.3) as explained earlier. The agreement between the RNAseq generated results here and the data from Wang *et al* study could also emphasise those eight genes as a possible biomarkers to differentiate between OVC and OSCC tumours.

In conclusion, the analysis of the significant DEGs here has highlighted several candidates that might provide important information about the malignant progression and the biological features and could aid prognosis and treatment choices for OVC. This includes *SERPINE1, CDH3* and *PLA2G4D* genes that have been also identified within significant chromosomal gain regions in the copy number analysis data (chapter four).

Chapter 7 Revealing the somatic genomic alterations associated with OVC compared to OSCC

7.1 Introduction

High-throughput Sanger sequencing studies have revealed that candidate cancer gene mutation frequency might be higher than expected, and that certain combination of mutations affect a tumour's characteristics (Sjoblom *et al.*, 2006), (Greenman *et al.*, 2007), (Cancer Genome Atlas Research, 2008). With the advances of NGS technologies, sequencing throughput has radically increased while its cost has decreased. Furthermore, NGS can now be applied to routinely prepared pathological FFPE materials that usually have degraded DNA (Schweiger *et al.*, 2009). Numerous studies have used NGS approaches to identify the underlying mutation in genetic diseases (Ng *et al.*, 2009), (Krawitz *et al.*, 2010). In addition, massively parallel sequencing methods such as whole-exome sequencing have described the genomic variation landscape associated with many different tumours and have also shown biological insights related to clinical aspects (Garraway and Lander, 2013).

OSCC develop in a multistep fashion through a series of genetic and histological changes. Whole-exome sequencing was performed previously to identify the mutational landscape of HNSCC (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). These studies indicated that up to 20% of tumours have *NOTCH1* loss-of-function mutations and >80% contains *TP53* mutations. Similarly and like all solid tumours, OVC is thought to be initiated and develop through a sequence of genetic events. However, the possible link between the biological behaviour and the structural and functional features with the clinical outcome in OVCs is still not clear and requires further investigation. Moreover, understanding the genetic basis of this specific subtype of oral cancer could allow a stratified medicine approach through therapeutic targeted treatment of disturbed pathways.

7.2 Aim

This is the first study up to date that aims to use next generation whole-exome sequencing to investigate the contribution of somatic genomic alteration in the pathogenesis of OVC and gain a comprehensive view of the genetic mutations underlying these lesions and compare them with the somatic genomic mutations underlying OSCCs, which could then aid in the histopathological diagnosis and treatment choices of OVC.

7.3 Results

7.3.1 Characteristics of the study cohort

From the verrucous cohort study described in Chapter 3 and table 3.1, whole exome sequencing of VC lesions and the normal adjacent epithelium tissue was carried out for 12 patients. For the comparison between OVC cohort and OSCC cohort, whole exome sequencing of OSCC lesions and the normal adjacent epithelium tissue was carried out as well for 20 patients (primary OSCC data belong to Pre-cancer Genomics Group). Pathological materials from each OVC case were available in the form of FFPE tumour blocks. Archive ages of FFPE blocks ranged between 10 and 1yr(median age of blocks: 5 years).

7.3.2 Mutated genes in OVC

The high throughput whole exome sequencing data from oral verrucous lesions and adjacent normal tissues DNAs using Illumina HiSeq 2500 NGS technology has successfully identified a list of mutated genes for each sample. Whole exome sequencing results from this parallel sequencing platform were generated for all twelve patients (12 matched normal versus verrucous samples), with an average of 6800453610 sequencing bases ranged between (1673853712 – 10309418584), and an average of 11466717.92 mismatched sequencing bases (Table 7.1). The target sequencing regions were insufficiently covered in two verrucous pair samples (normal and tumour tissue, highlighted in yellow). However, an average of 90.4x of the target sequencing regions were sufficiently covered in the remaining samples tissue for confident variant calling (Table 7.2).

Performing whole exome analyses of DNA isolated from FFPE materials has one major application-specific challenge. Formalin-fixed tissues show a higher rate of non-reproducible DNA sequence changes than frozen tissues. This is expected owing to formalin cross-linking of cytosine (C) nucleotides on either strand, which causes an inability of Taq polymerase in identifying the cytosine and integrating an adenine (A) in place of a guanosine (G) producing C>T or G>A artificial mutation during PCR (Williams *et al.*, 1999) (Srinivasan *et al.*,

2002). In the work presented here, the amount of FFPE damage (C/T artificial mutation) per base ranged between (0.00038 – 0.00085), except for four samples (highlighted in orange in Table 7.1), as they showed the highest amount of FFPE damage.

To get an idea about the C/T amount per base in a 'perfect DNA', the amount of C/T per base for a blood DNA sample (data belong to Pre-cancer Genomics Group) was included also in Table 7.1 (sample PG038-BC, highlighted in green), which was 0.0003 out of the total produced sequencing bases, and this was very low in compare to 0.001 from the orange highlighted four samples. The fact that the four DNA samples (tumour and normal pairs) with the highest amount of C/T artificial mutation were from the same blocks also confirms that the DNA of the whole tissue was exposed to FFPE damage. By looking at the mismatches per base column in Table 7.1, which reflect the amount of mutations in the sequenced samples below, it is also obvious that the same four samples (highlighted in blue) harbour the highest mutation rates when compared with others and with the blood DNA sample (PG038-BC), and this is expected due to the elevated amounts of C/T artificial mutation in each of the four samples.

According to the above-discussed points, and since the gene mutations in samples V-112-T and V-119-T may not be reliable; I decided to exclude V-112-T and V-119-T data from my analysis, since they showed the highest amount of C/T artificial mutation and the highest number of gene mutations.

Table 7.1 Exome capture read results

Name	Bases	Mismatches	Mismatches per base	C/T	C/T per base	C/T per mismatch
V-125-N	7486265906	10498282	0.00140	3198834	0.00042	0.30470
V-125-T	10309418584	13994079	0.00135	4344514	0.00042	0.31045
V-29-N	8776940981	16871436	0.00192	5782143	0.00065	0.34271
V-29-T	6484665376	12574314	0.00193	4316307	0.00066	0.34326
V-108-N	6835858705	7413489	0.00108	2623643	0.00038	0.35390
V-109-T	7468007393	11484447	0.00153	4098928	0.00054	0.35691
V-108-T	6745299542	8744703	0.00129	3160719	0.00046	0.36144
V-123-N	9285726514	13390589	0.00144	4850336	0.00052	0.36221
V-123-T	7965521321	11732772	0.00147	4252427	0.00053	0.36244
V-04-N	8674996779	18280297	0.00210	6627159	0.00076	0.36253
V-04-T	8881438543	20175705	0.00227	7413942	0.00083	0.36746
V-109-N	7510164494	10531154	0.00140	3887331	0.00051	0.36912
V-116-T	7602302896	10718850	0.00140	4072116	0.00053	0.37990
V-14-T	8177158698	14654916	0.00179	5625228	0.00068	0.38384
V-14-N	7663846392	13082551	0.00170	5034663	0.00065	0.38483
V-116-N	6551201624	8478281	0.00129	3283587	0.00050	0.38729
V-124-N	6670629644	9311512	0.00139	3828910	0.00057	0.41120
V-124-T	8105611127	13524638	0.00166	5686421	0.00070	0.42044
V-98-N	8098752008	14231188	0.00175	6860096	0.00084	0.48204
V-98-T	5348618959	9469854	0.00177	4578829	0.00085	0.48351
V-112-N	3048261492	9397622	0.00308	4693239	0.00153	0.49940
V-112-T	1793563186	5597640	0.00312	2803470	0.00156	0.50083
V-119-N	2052782752	6015216	0.00293	3146368	0.00153	0.52306
V-119-T	1673853712	5027695	0.00300	2637332	0.00157	0.52456
Average	6800453610	11466717.92				
PG038-BC	7640943885	9065275	0.00118	2581305	0.00033	0.28474

Table 7.2 Exome capture coverage.

Sample	Mean	Median
V-119-T	19.63	18
V-112-T	22.34	20
V-119-N	25.74	24
V-112-N	37.04	35
V-29-T	85.29	63
V-98-T	83.01	73
V-04-N	115.24	74
V-04-T	114.75	74
V-29-N	110.11	77
V-108-T	92.67	86
V-124-N	92.76	87
V-109-N	102.66	90
V-14-T	114.09	90
V-109-T	99.29	91
V-123-T	112.67	95
V-108-N	111	96
V-116-T	117.32	100
V-98-N	118.43	101
V-124-T	128.67	106
V-123-N	126.68	114
V-125-T	139.03	120

7.3.2.1 Potentially driver somatic mutations in OVC

The Variant Effect Predictor (VEP) computational tool was used (http://www.ensembl.org/info/docs/tools/vep/index.html) to determine variant location and consequence (e.g. frameshift, missense, stop gained, stop lost), and the effect of the variants (e.g. structural variants, deletions and insertions) on genes with providing SIFT and PolyPhen scores (predicts the probability of whether an amino acid substitution is damaging a protein function) (Ng and Henikoff, 2002). Sequence calls were also filtered using a Phred-like consensus quality score, which predicts the probability of incorrect SNPs (Ewing *et al.*, 1998), (Ewing and Green, 1998). From the bioinformatics analysis of the generated whole exome sequencing data, gene mutation lists were produced for each OVC sample. To determine the significant 'driver' mutated ones, they were filtered according to the following:

1. Gene mutations that were found in more than 50% of tumour cells in each OVC sample.
2. Gene mutations that have a consequence on protein function (deleterious, splice variant, probably deleterious, possibly deleterious, stop gained).
3. Gene mutations with Phred quality score of ≥15. (≥15 score means 95% SNP call accuracy).
4. The results were then further analysed by running the gene lists against 13 enriched KEGG pathways (which are more related to HNCs) including: KEGG P13K, KEGG WNT signalling pathway, KEGG cell cycle, KEGG calcium signalling pathway, KEGG VEGF signalling pathway, KEGG MAPK, KEGG DNA replication, KEGG PHOSPHATIDYLINOSITOL signalling system, KEGG P53 signalling pathway, KEGG NOTCH signalling pathway, KEGG JAK STAT signalling pathway, KEGG ERBB signalling and KEGG hedgehog, as well as cancer gene census and Stransky mutation list (76 previously identified genes in HNSCCs harbouring high statistically significant mutations) (Stransky *et al.*, 2011).

The generated driver somatic gene mutations lists for OVC samples are shown below in Table 7.3. From the genes list below, three samples had gene mutations in KEGG P13K pathway, two had gene mutations in KEGG P53 signalling pathway, two had gene mutations in KEGG JAK STAT signalling pathway, two had gene mutations in KEGG calcium signalling pathway, two had gene mutations in KEGG cell cycle pathway, and one sample had mutated genes in KEGG MAPK signalling pathway. However, and importantly, all the potentially driver-mutated genes listed below in Table 7.3 were not shared between the ten OVC samples.

The presence of harmful (deleterious) mutations within the human population has an important impact on human health and can increase the risk of developing a disease. Here, deleterious mutations were detected in six genes, including *IRF6*, *PRIM2*, *HOXC11*, *MAP2K3*, *CDC27* and *STAT3*. Furthermore, two identified genes in this study harboured probably deleterious mutations, including *CREBBP* and *TP63* genes, and two genes harboured possibly deleterious mutations, including *GYS1* and *CASP8*.

In 2011, Stransky *et al*. published whole genome mutational profiling data of HNSCC. Several genes were involved, which include *IRF6* (interferon regulatory factor 6), *NOTCH1*, and *TP63*. They hypothesised that mutations in these genes disturb stratified squamous epithelial differentiation and development program in the precursor cells of this cancer (Stransky *et al.*, 2011). *IRF6* and *TP63* both included deleterious mutations in samples V-004-T and V-116-T respectively. Furthermore, *PRIM2* gene here in sample V-029-T had a deleterious mutation and KEGG pathway analysis revealed enrichment of *PRIM2* gene in KEGG DNA replication pathway. *PRIM2* was reported recently to be associated with breast cancer (Nilsson *et al.*, 2012). In addition, several HOX genes have been described to play a role in the development of many cancers (Bhatlekar *et al.*, 2014). *HOXC11* gene was previously shown as well to be a strong predictor associated with poor survival in breast cancer (McIlroy *et al.*, 2010). Another report shows that progressive and strong overexpression of *HOXC11* gene was detected in metastatic melanoma when compared with normal melanocytes (Cantile *et al.*, 2012). In the current exome sequencing

study, *HOXC11* mutated gene in sample V-116-T harboured a deleterious mutation and was identified too in cancer gene census database (http://www.sanger.ac.uk/genetics/CGP/Census/).

Similarly, another deleterious mutation in *MAP2K3 (*Mitogen-activated protein kinase kinase 3) gene was also detected in sample V-116-T in this study and KEGG pathway analysis revealed enrichment of *MAP2K3* in KEGG MAPK pathway. A previous study reported down-regulation of *MAP2K3* in immortal (proliferate indefinitely) human breast epithelial cells, while overexpression of *MAP2K3* endorsed cell senescence (ageing) (Jia *et al.*, 2010). Sample V-124-T had a deleterious mutation as well in *CDC27* (cell division cycle 27) gene and KEGG pathway analysis revealed enrichment of *CDC27* in KEGG cell cycle pathway. *CDC27* was suggested as a tumour suppresser gene and down-regulation of *CDC27* was associated with several breast cancer cell lines (Pawar *et al.*, 2010).

An earlier study has implicated STAT (signal transduction and activation of transcription) activation, mainly *STAT3*, in tumour progression and transformation (Bromberg and Darnell, 2000). Increased *STAT3* activation has been also previously shown in HNSCC, where *STAT3* contributes to tumours loss of growth through an anti-apoptotic mechanism (Grandis *et al.*, 2000), (Leong *et al.*, 2003). Here, sample V-125-T harboured a deleterious mutation in *STAT3* gene identified in cancer gene census database. KEGG pathway analysis also revealed enrichment of *STAT3* gene in KEGG JAK STAT signalling pathway.

Furthermore, sample V-098-T showed a probable deleterious mutation in *CREBBP* gene (that was also identified in cancer gene census database) and a possible deleterious mutation in *GYS1* (glycogen synthase) gene. KEGG pathway analysis revealed enrichment of *CREBBP* gene in KEGG cell cycle, KEGG JAK STAT, KEGG NOTCH and KEGG WNT pathways, besides enrichment of *GYS1* gene in KEGG P13K pathway. A previous gene expression analysis study revealed that *CREBBP* gene was one of the genes involved in the significant de-regulated KEGG cell cycle and KEGG JAK STAT signalling pathways associated with OSCC (Saleh *et al.*, 2010). Likewise, sample V-116-T

showed a probably deleterious mutation in *TP63* gene and a possibly deleterious mutation in *CASP8* gene. Both genes were listed in Stransky *et al*. mutational profiling data of HNSCC (Stransky *et al.*, 2011), and KEGG pathway analysis revealed as well enrichment of *CASP8* gene in KEGG P53 signalling pathway. 8% of the mutations in Stransky *et al* HNSCC mutational profiling data were in the apoptosis-related gene, *CASP8* (Stransky *et al.*, 2011), and *TP63* gene, located at 3q28, is another HNSCC candidate cancer gene that has been identified previously (de Oliveira *et al.*, 2007), (Tsantoulis *et al.*, 2007), (Leemans *et al.*, 2011b).

Some mutations stop a protein from being made at all by producing a premature stop codon (Pleasance *et al.*, 2010a). In the current study, stop mutations were identified in two genes, including *CACNA1I* in sample V-116-T and *PHLPP1* in sample V-123-T. KEGG pathway analysis revealed enrichment of *CACNA1I* gene in KEGG MAPK and KEGG calcium signalling pathways, while *PHLPP1* gene was enriched in KEGG P13K signalling pathway. A very recent report identified twelve differentially expressed genes, including *CACNA1I*, in a microarray analysis study that was conducted to identify DEGs in para-carcinoma, carcinoma and relapsed human pancreatic cancer (Chang *et al.*, 2014). *PHLPP1* was reported recently also as a tumour suppressor gene in human colorectal cancer (Liao *et al.*, 2013).

Splice variant is a sequence variant in which an alteration has occurred within the splice site region. Expression of tumour-specific splice variants can markedly affect tumour biology including motility &, proliferation (Skotheim and Nees, 2007). Though, it is unclear if the existence of specific splice variants in cancer is a consequence of the malignant phenotype, or if they are contributing to the tumour phenotype (Skotheim and Nees, 2007). Here, splice variant mutations were detected in six genes, including *RFWD2*, *RELN*, *MLLT6*, *ERBB4*, *GPHN* and *FAM135B.* Sample V-098-T showed two splice variant mutations in *ERBB4* gene and *MLLT6* gene (that was also identified in cancer gene census database). The KEGG pathway analysis revealed enrichment of *ERBB4* gene in KEGG ERBB and KEGG calcium signalling pathways. Overexpression ERBB4 receptor family members is common in OSCC; and a

recent IHC analysis study revealed overexpression of ERBB4 with lymph node capsular rupture in OSCC cases (Silva *et al.*, 2014). *MLLT6* was also reported previously as a tumour-related gene in 18 breast cancer cell lines and 47 primary breast tumours (Naylor *et al.*, 2005).

Similarly, Sample V-014-T harboured a splice variant mutation in *RFWD2* gene and KEGG pathway analysis showed enrichment of *RFWD2* gene in KEGG p53 signalling pathway. A previous study revealed overexpression of *RFWD2* gene, also named as *COP1*, in 44% (76 out of 171 samples) of ovarian adenocarcinoma and 81% (25 out of 32 samples) of breast cancer cases. They suggested that overexpression of COP1 plays a role in p53 protein accelerated degradation in cancers and reduces *p53* tumour-suppressor function (Dornan *et al.*, 2004). Sample V-109-T also had a splice variant mutation in *GPHN* gene that was identified as well in cancer gene census database. *GPHN* gene on chromosome 14q24 was reported as a partner gene fused with MLL gene in a leukaemia case (Eguchi *et al.*, 2001). Finally, sample V-116-T showed a splice variant mutation in *FAM135B* gene that has been listed as one of the mutated genes in Stransky *et al*. whole genome mutational profiling data of HNSCC (Stransky *et al.*, 2011). *FAM135B* was reported recently as a novel cancer-associated gene that promotes esophageal SCC malignancy (Islamian *et al.*, 2014).

Again, all the genes mentioned above were not shared between the ten OVC samples in this study. However, these genes were: 1. found in more than 50% of OVC tumour cells, 2. with Phred quality score of ≥15, 3. were either, tumour suppresser or oncogenes previously identified in other cancers, and 4. involved in different pathways. Therefore, mutations that have a consequence on protein function in those genes were expected to be driver mutations in their corresponding samples that they were identified in.

On the other hand, and as seen from Table 7.4 below for the list of potentially driver mutated genes in 20 OSCC cases; 14 OSCC samples had mutations with *TP53* (70% of samples), seven with *CDKN2A* (35% of samples), five with *NOTCH2* (25% of samples), four with *FAT1* (20% of samples) and two with *NOTCH1* (10% of samples) (primary OSCC data belong to Pre-cancer

Genomics Group). Besides that, all mutations in those five genes were shared between two or more OSCC samples, they all also had a consequence on protein functions, were found in more than 50% of OSCC tumour cells and with Phred quality score of ≥15. Importantly, the five genes are significant HNSCC cancer genes that have been identified in two previous whole-exome sequencing studies conducted by two groups independently on a total of approximately 100 HNSCC samples (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). Mutation of the tumour-suppressor, *TP53* gene, is among the earliest identified genetic changes and the most common in HNSCC, arising in over half of all cases (Leemans *et al.*, 2011a). One of the most important findings from the previous HNSCC whole-exome sequencing studies is the discovery of mutations in the *NOTCH1* gene in 12%–15% of the examined samples (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). Also, Stransky *et al*. identified non-synonymous point mutations in *NOTCH3* or *NOTCH2* genes in 11% of the cases. Mutations in *CDKN2A* gene were found as well in about 7% of HNSCC tumours by exome sequencing (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). Additionally, the identification of mutations in the *FAT1* gene in nine HNSCC samples (12%) in Stransky *et al*. study provided a new insight into the potential mechanisms of metastasis and invasion in HNSCC. Interestingly, all the ten OVC samples showed lack of any mutations within *TP53*, *CDKN2A, NOTCH1, NOTCH2* and *FAT1* genes; which has been previously shown to be associated with HNSCC and were found in OSCC cases here as discussed above.

7.3.2.2 Shared potentially driver mutations between OVC and OSCC samples

However, and although none of the above mutated genes was shared between the ten OVC samples, seven mutated genes were shared between OSCC and OVC as revealed in Figure 7.1 and Table 7.4 including: *RFWD2, RELN, CREBBP, CASP8, FAM135B, CACNA1I* and *CDC27*. These genes were shared between one OVC and one OSCC samples except for *CDC27* gene that was shared between one OVC and three OSCC samples (Table 7.4). *CASP8* and *FAM135B* genes were mutated in the same OVC sample (V-116-T) and

they are the only two genes from the seven-shared mutated genes that have been identified in Stransky *et al*. mutational profiling data of HNSCC.

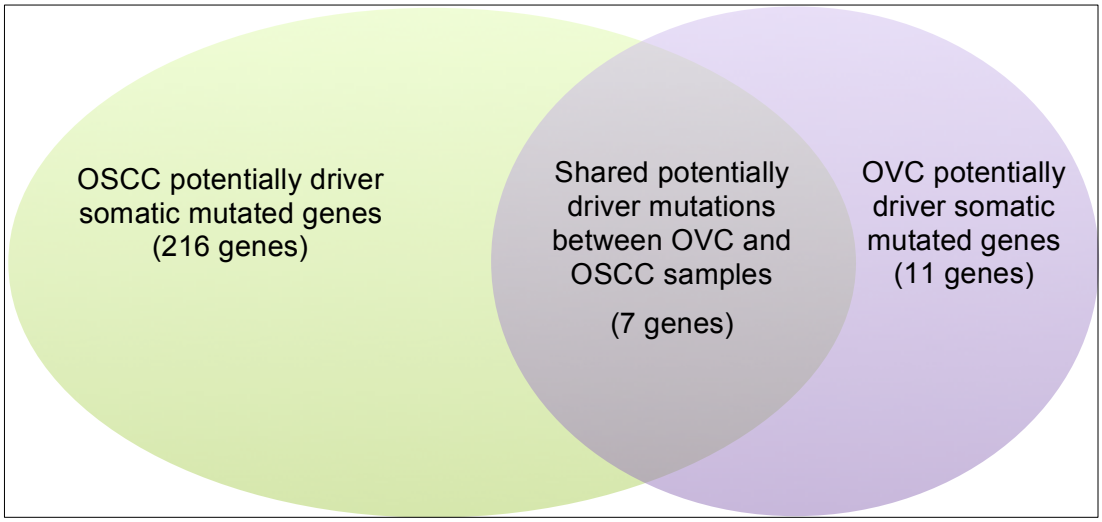


Figure 7.1 Shared potentially driver mutations between OVC and OSCC samples.

Note: the total 18 mutated genes in OVC were not shared between each OVC samples and the number represents the total mutated genes from all the ten OVC samples.

Table 7.3 driver somatic mutations in OVC

Sample	Genes	Position	Depth at position	Mutant reads	Normal depth	Score	Cellularity	Type of mutation	Pathways and Gene lists
V-004-T	IRF6	chr1:209969822	29	9	28	29	0.77	Deleterious	Stransky mutation list
V-014-T	RFWD2	chr1:176050430	29	9	13	16	0.89	Splice variant	KEGG P53 signalling pathway
V-029-T	PRIM2	chr6:57467084	23	6	33	25	0.74	Deleterious	KEGG DNA replication
	RELN	chr7:103363639	27	8	34	31	0.84	Splice acceptor	KEGG P13K
V-098-T	CREBBP	chr16:3808953	22	4	29	15	0.51	Probably deleterious	KEGG cell cycle, KEGG JAK STAT, cancer gene census, KEGG NOTCH, KEGG WNT
	MLLT6	chr17:36878128	21	4	44	20	0.51	Splice variant	Cancer gene census
	GYS1	chr19:49472825	20	4	61	25	0.56	Possibly deleterious	KEGG P13K
	ERBB4	chr2:212578379	21	7	18	21	0.95	Splice variant	KEGG calcium signalling pathway, KEGG ERBB
V-109-T	GPHN	chr14:67391008	58	12	57	29	0.59	Splice variant	Cancer gene census
V-116-T	CASP8	chr2:202131411	108	22	109	71	0.57	Possibly deleterious	KEGG P53 signalling pathway, Stransky mutation list
	TP63	chr3:189586422	118	21	98	58	0.50	Probably deleterious	Stransky mutation list
	FAM135B	chr8:139323081	29	6	22	15	0.54	Splice variant	Stransky mutation list
	HOXC11	chr12:54367454	216	42	111	82	0.55	Deleterious	Cancer gene census
	MAP2K3	chr17:21204192	88	17	31	16	0.54	Deleterious	KEGG MAPK
	CACNA1I	chr22:40058138	23	6	20	17	0.73	Stop	KEGG MAPK, KEGG calcium signalling pathway
V-123-T	PHLPP1	chr18:60383878	90	17	58	30	0.53	Stop	KEGG P13K

Table 7.4 driver somatic mutations in OSCC

Sample	Genes	Genes shared with OVC samples
PG004	BMP8B (deleterious), **NOTCH2** (deleterious), **FAT1** (stop gained), MAP3K4 (deleterious), PKN2 (frame shift), COL4A4 (coding feature truncation), EPHA2 (stop), FBXO11 (possibly deleterious), PIK3CA (possibly deleterious), HDAC2 (probably deleterious), DICER1 (possibly deleterious), FGFR1OP (intron truncation), PAX5 (intron truncation), PRKAR1A (intron truncation)	
PG038	**TP53** (stop), **NOTCH2** (possibly damaging), **NOTCH2** (coding truncation), ANKRD36C (intron elongation), MERTK (intron truncation), **FAT1** (coding truncation), TERF2IP (intron truncation), **CDKN2A** (stop), ZNF521 (stop), BCL2L11 (intron truncation), PI4KA (benign), BMP8A (intron elongation), NCOR2 (in-frame insertion)	
PG049	DRD5 (benign), NOS3 (possibly deleterious), ZFHX4 (probably deleterious), FAM135B (stop), **CDKN2A** (deleterious), **TP53** (deleterious), SRSF2 (benign), ZNF521 (deleterious), LAMA5 (possibly deleterious), ITGA10 (intron truncation), ITPR1 (intron elongation), NSD1 (frame shift), SMC3 (intron elongation), ITGB4 (intron elongation)	FAM135B: V-116-T / splice variant
PG063	**TP53** (splice donor), PTEN (deleterious), PAX7 (intron truncation), EBF1 (5' truncation), KIF5B (intron truncation), SIAH1 (intron truncation)	
PG079	ASPSCR1 (deleterious), RPA3 (intron elongation), **TP53** (deleterious missense), BRCA2 (stop), CSH1 (benign), BRCA2 (stop), CSH1 (benign)	
PG105	CREBBP (benign), ITGA6 (3' truncation), NIN (intron truncation), FGF20 (deleterious), ITGA10 (intron elongation), PBRM1 (splice variant), **TP53** (deleterious missense)	CREBBP: V-098-T / probably deleterious
PG108	NCOA2 (deleterious), **TP53** (stop), NSD1 (stop), LRFN5 (benign), **NOTCH2** (deleterious), RBPJ (splice or possibly deleterious), ADCY8 (deleterious), CACNA2D4 (intron truncation)	
PG122	MAP3K12 (frame shift), COL6A1 (intron truncation), **TP53** (stop), PMS2 (splice variant), PRKDC (splice variant), CDC27 (benign), MAML2 (in-frame deletion)	CDC27: V-124-T / deleterious
PG123	**CDKN2A** (stop), CARD11 (benign), PLA2G2D (deleterious), IL1A (probably deleterious), SENP2 (possibly deleterious), CAMK2A (deleterious), ELN (deleterious), ZFHX4 (deleterious), ZFHX4 (deleterious), CSMD3 (deleterious), **NOTCH1** (deleterious), ARRB1 (stop), ATM (benign), MAP3K12 (deleterious), PPP2R5E (benign), **TP53** (deleterious), NF1 (probably deleterious), RYR1 (probably deleterious), PLCB4 (benign), FGF12 (splice variant), CACNA1G (splice variant), EP300 (intron truncation)	

PG129	**NOTCH2** (coding truncation), MYLK (3' elongation), RELN (start codon variant), PRKACG (5' elongation), ITGB4 (intron elongation), RFWD2 (benign), PIKFYVE (deleterious), FLT1 (possibly deleterious), SYNE1 (intron elongation), NOS1 (intron elongation), TCL1A (splice variant), PER1 (intron truncation)	RFWD2: V-014-T / splice variant RELN: V-029-T / splice variant
PG136	ITGA2 (splice donor), BRIP1 (possibly deleterious), CYLD (intron truncation)	
PG137	CACNA1B (deleterious), LAMC2 (intron truncation), FBXO11 (intron truncation), CLCF1 (intron truncation), **TP53** (coding truncation), RUNX1 (intron truncation), BCL11B (5' elongation), ITPR1 (splice variant), ORC5 (possibly deleterious), CDC27 (stop), NCOA2 (splice variant), CACNA1B (intron truncation), MAML2 (in-frame deletion), PIK3C2G (intron truncation)	CDC27: V-124-T / deleterious
PG144	CSF3R (deleterious), STAT1 (possibly deleterious), **NOTCH1** (deleterious), NUMA1 (benign), CDC27 (benign), **FAT1** (coding truncation), ITPKB (benign), NFE2L2 (deleterious), CASP8 (deleterious), HIST1H3B (deleterious), **FAT1** (stop), SHH (deleterious), FZD3 (stop), RIMS2 (stop), **CDKN2A** (probably deleterious), WNT5B (deleterious), DGKH (deleterious), KAT2A (deleterious), ADORA2A (deleterious), MYH9 (deleterious), EML4 (intron elongation), PPP2R2B (intron elongation), TEK (intron elongation), RBL1 (intron truncation), CACNA1I (intron elongation)	CDC27: V-124-T / deleterious CASP8: V-116-T / possibly deleterious CACNA1I: V-116-T / stop
PG174	CSMD3 (probably deleterious), GNA14 (deleterious), COL5A1 (splice variant), PLCE1 (stop), **TP53** (deleterious), RNF43 (deleterious), **CDKN2A** (stop), ROCK1 (splice variant), LAMA5 (possibly deleterious), IL1R1 (intron truncation), CACNA1A (in-frame deletion)	
PG187	PBRM1 (deleterious), **TP53** (stop), ZFHX4 (possibly deleterious), LRP5 (possibly deleterious), CANT1 (benign), MYH11 (intron truncation), CSMD3 (splice acceptor)	
PG192	HGF (benign), RANBP17 (possibly deleterious), **FAT1** (coding truncation), KCTD8 (possibly deleterious), EPPK1 (benign), MYLK3 (benign), CHN1 (splice variant)	
PG008	**TP53**, ITGB4	
PG019	FANCD2, RASA1, SYNJ2, MAD1L1, **CDKN2A**, PCSK7, MLL2, CHD8, PCK2, PLCB2, CACNA1H, ADCY9, **TP53**, LAMC2, SGK2, EPS15, PIKFYVE	
PG030	TPR, STAM2, **CDKN2A**, CACNA1E, LAMC2, FIP1L1, IVL, FANCD2, HGF, FANCA, ANAPC1, ANAPC10, FGFR4, DGKH, **TP53**, PER1	
PG055	NDRG1, NFAT5, PIM1, ITGA8, MAML2, ESPL1, ASXL1, **NOTCH2**, NSD1, FIP1L1, MAPK13	

7.3.2.3 Other mutated genes in OVC

From the bioinformatics analysis of the generated whole exome sequencing data as explained in section 7.3.2.1, gene mutation lists were produced for each OVC sample. To determine the significant mutated ones, they were filtered according to the following: gene mutations that were found in more than 50% of the tumour cells in each OVC sample, that have a consequence on protein functions, and with Phred quality score of ≥15. A total of 206 genes contained putatively functional variants in OVC cases. These genes were not identified in Stransky *et al*. whole genome mutational profiling data of HNSCC or either the Cancer Gene Census database, or involved in any KEGG signalling pathway. Therefore, I only focused on the shared mutated genes between the ten OVC samples and discussed them further in section 7.3.2.2.1 below.

7.3.2.3.1 Shared mutations between OVC samples

Eleven of the 206 identified mutated genes (Table 7.5) were present in at least two OVC samples: *DSPP* gene was shared between four OVC samples, *MUC4* between three, *NEFH* between three, *ANP32E* between three, and the remaining seven genes were shared between two OVC samples (*FAM194B, TBP, AGAP7, CCDC158, NBPF12, PCDH1* and *HS6ST1*). Additionally, eight out of the eleven-shared genes between OVC samples were mutated as well in at least one OSCC sample (data not shown, primary OSCC data belong to Pre-cancer Genomics Group), except for *AGAP7* and *CCDC158* genes that were mutated only in two OVC samples.

Samples V-004-T, V-109-T, V-116-T and V-029-T harboured inframe deletion mutations in *DSPP* gene, while DAVID functional enrichment analysis showed enrichment of this extracellular gene in calcium ion binding. Interestingly, samples V-004-T and V-109-T had inframe deletion mutations (non-synonymous inframe variant in which bases were deleted from the coding sequence) in exactly the same position in both samples (blue highlighted cells in Table 7.5). A previous study suggested the involvement of *DSPP* gene in prostate carcinogenesis, as they reported significant increase in *DSPP*

expression in 69 prostate cancer lesions as well as three prostatic tumour cell lines (Chaplet *et al.*, 2006). Another earlier study also reported up-regulation of *DSPP* gene along with other genes in histologically aggressive and poorly differentiated OSCC and in oral epithelial dysplasia (Ogbureke *et al.*, 2007), (Ogbureke *et al.*, 2010).

Similarly, samples V-109-T, V-116-T and V-123-T had missense variant mutations (a sequence variant, that alters one or more bases, where the resulting length is preserved but with a different amino acid sequence) in *MUC4* gene, while DAVID functional enrichment analysis revealed enrichment of this plasma membrane gene in biological and cell adhesions. An earlier study revealed overexpression of *MUC4* gene in pancreatic adenocarcinomas and cell lines whereas it remained undetectable in normal pancreatic tissues (Andrianifahanana *et al.*, 2001). Other studies also demonstrated a positive correlation between *MUC4* expression and pancreatic tumour growth and malignant progression (Swartz *et al.*, 2002), (Chaturvedi *et al.*, 2007).

In addition, samples V-109-T, V-125-T and V-124-T harboured inframe deletion mutations in *NEFH* gene, while DAVID functional enrichment analysis showed enrichment of this gene in cytoskeleton and cell projection. Interestingly, samples V-109-T and V-125-T had inframe deletion mutations in exactly the same position in both samples (blue highlighted cells in Table 7.5). A recent microarray study that was conducted in 2012 reported down-regulation of *NEFH* gene along with other genes in nine metastatic squamous cell lung carcinoma samples when compared with eight non-metastatic samples (Wang *et al.*, 2012b).

Likewise, samples V-014-T, V-098-T and V-116-T had missense variant mutations in *ANP32E* gene. Remarkably, all the three samples had missense variant mutations in exactly the same position (blue highlighted cells in Table 7.5). In 2008, Tsukamoto *et al*. indicated that ANP32E expression was up-regulated and more intensely expressed in the cytoplasm of gastric cancer cells when compared with non-neoplastic epithelial cells (Tsukamoto *et al.*, 2008).

FAM194B, TBP, AGAP7, CCDC158, NBPF12, PCDH1 and *HS6ST1* were shared between two OVC samples (Table 7.5). Samples V-014-T and V-123-T had splice acceptor variant mutations (two base region are changed at the 3' end) in the integral membrane protein *PCDH1* gene within exactly the same position (blue highlighted cells in Table 7.5). Furthermore, DAVID functional enrichment analysis revealed enrichment of *PCDH1* plasma membrane gene in cell adhesions. A previous study indicated that epigenetic silencing of *PCDH1* is associated with breast cancer (Novak *et al.*, 2008). Additionally, samples V-116-T and V-123-T had splice acceptor variant mutations in *CCDC158* gene within exactly the same position, and samples V-116-T and V-108-T had inframe deletion mutations in *FAM194B* gene also within exactly the same position (blue highlighted cells in Table 7.5). A recent study reported significant expression of *CCDC158* gene along with other four genes in association with hepatocellular carcinoma (Huang *et al.*, 2012).

Table 7.5 Shared somatic mutations between OVC samples

Genes	Category	Samples	Position	Depth at position	Mutant reads	Normal depth	Score	Cellularity	Consequence
DSPP	Extracellular region, calcium ion binding	V-004-T	chr4:88537297	116	45	112	153	97%	Inframe deletion
		V-109-T	chr4:88537297	48	21	54	81	100%	Inframe deletion
		V-116-T	chr4:88537072	32	19	15	43	100%	Inframe deletion
		V-029-T	chr4:88537486	138	43	165	165	89%	Inframe deletion
MUC4	Plasma membrane part, cell adhesion, biological adhesion	V-109-T	chr3:195509108	34	8	36	19	66%	Missense variant
		V-116-T	chr3:195511897	37	12	23	20	93%	Missense variant
		V-123-T	chr3:195509563	27	5	44	22	53%	Missense variant
NEFH	Cytoskeleton, cell projection	V-109-T	chr22:29885580	92	39	87	134	100%	Inframe deletion
		V-125-T	chr22:29885580	98	38	43	72	100%	Inframe deletion
		V-124-T	chr22:29885598	65	17	51	37	74%	Inframe deletion
ANP32E	Phosphoprotein	V-116-T	chr1:150199042	107	27	62	58	71%	Missense variant
		V-098-T	chr1:150199042	75	19	70	58	72%	Missense variant
		V-014-T	chr1:150199042	59	13	54	30	63%	Missense variant
TBP	Positive regulation of gene expression	V-124-T	chr6:170871013	109	73	47	150	100%	Inframe insertion
		V-109-T	chr6:170871037	21	8	17	16	100%	Inframe deletion
PCDH1	Cell adhesion, plasma membrane part	V-123-T	chr5:141247181	21	7	26	19	95%	Splice acceptor variant
		V-014-T	chr5:141247181	45	13	30	23	81%	Splice acceptor variant
HS6ST1	Plasma membrane part	V-124-T	chr2:129026227	61	12	25	19	56%	Missense variant
		V-125-T	chr2:129025860	52	13	43	37	71%	Missense variant
FAM194B	Polymorphism	V-108-T	chr13:46170719	49	10	54	34	56%	Inframe deletion
		V-116-T	chr13:46170719	42	10	31	26	66%	Inframe deletion
AGAP7	Ion binding, cation binding	V-109-T	chr10:51464976	68	13	50	24	53%	Missense variant
		V-124-T	chr10:51465046	37	8	30	22	62%	Missense variant
NBPF12	-	V-116-T	chr1:146448501	32	15	28	39	100%	Missense variant
		V-124-T	chr1:147579272	36	21	13	28	100%	Missense variant
CCDC158	-	V-116-T	chr4:77305821	38	10	27	18	75%	Splice acceptor variant
		V-123-T	chr4:77305821	29	9	30	22	89%	Splice acceptor variant

7.3.3 Functional analysis of mutated genes in O gene set enrichment analysis)

To further assess the potential biological impact of the mutatio OVC cases, functional enrichment and pathway analysis was pe the Database for Annotation, Visualisation and Integrated Disc (Dennis *et al.*, 2003). A total of 224 genes (18 potentially driver sc genes + 206 other mutated genes in OVC) contained putati variants in the ten OVC cases. I tested for enrichment of Gene categories within each of these genes using DAVID with a thresh (Table 7.6). From the functional annotation analysis and the resu below, 19 significant GO terms were allocated for OVC 224 mutat majority of the enriched mutated genes in the ten OVC cases plasma membrane part (31 genes), participated in cell and biolog (14 genes), and were implemented in calcium ion (Ca^{2+}) bindin Enrichment of these categories is expected in the developmer lesions. DAVID biological processes: keratinocyte prolifera proliferation were also significantly enriched in OVC cohort here the DNA can promote development of a normal keratinocyte int malignant or pre-malignant keratinocyte that is described by th proliferate in a less-controlled manner than normal (Scully and Ba

Additionally, functional enrichment and pathway analysis was OSCC cohort using DAVID computational tool (primary OSCC Pre-cancer Genomics Group). 140 significant gene ontology allocated for OSCC mutated genes (data not shown). The r enriched mutated genes in the 20 OSCC cases located in the cy genes) and non-membrane-bounded organelle parts (119 genes in cell adhesion (38 genes), cell death (37 genes) and were in nucleotide binding (108 genes). Alterations in cellular pro apoptosis pathways cause tumour development and enrichr categories is expected in OSCC lesions.

Remarkably, the gene ontology analysis of the target genes in OVC did not show any involvement of the mutated genes in cell death or apoptosis. Whereas multiple enriched GO terms related to cell death were enriched in OSCC cases, which include GO:0008219 cell death, GO:0043523 regulation of apoptosis and GO:0016265 death. For example, 1.7% of the genes are related to GO:0043523 apoptosis, 5.8% are related to GO:0008219 cell death, and 5.8% are related to GO:0016265 death (data not shown). This can be attributed to the lack of *TP53* mutation (involved in the apoptotic pathway) in OVC cases.

Table 7.6 Gene Ontology (GO) terms for enrichment amongst the 224 genes harbouring putatively functional variants in OVC cases, using DAVID.

Gene Ontology (GO) category:	GO term associated with the gene list	No. of genes overlapping GO term	% of total genes tested overlapping GO term	DAVID P-Value
Molecular function	acetyltransferase activity	4	2.29	0.038
	Monovalent inorganic cation transmembrane transporter activity	5	2.87	0.014
	Inorganic cation transmembrane transporter activity	5	2.87	0.048
	Calcium ion binding	15	8.62	0.039
Cellular component	Apical plasma membrane	5	2.87	0.039
	Membrane raft	6	3.44	0.012
	Plasma membrane part	31	17.81	0.026
Biological Process	Response to tumour cell	2	1.14	0.046
	Keratinocyte proliferation	3	1.72	0.004
	Homophilic cell adhesion	7	4.02	0.001
	Cell morphogenesis involved in differentiation	7	4.02	0.029
	Cell-cell adhesion	8	4.59	0.017
	Cell projection organization	9	5.17	0.025
	Cell proliferation	10	5.74	0.024
	Positive regulation of transcription	11	6.32	0.043
	Positive regulation of gene expression	12	6.89	0.023
	Transmembrane transport	13	7.47	0.008
	Cell adhesion	14	8.04	0.016
	Biological adhesion	14	8.04	0.016

The smaller the P-value, the more enriched are the genes in the GO category.

7.3.4 Integration of exome sequencing data and RI expression data

To highlight genes of possible clinical and biological importance i of significant mutated genes generated from exome sequencing integrated with genes from the significant differential expression coding genes in OVC versus its matched normal, and OVC versu was generated from RNAseq data analysis (chapter 6).

No correlation was found between the mutated genes from analysis of the exome sequencing data and the differentially ex from transcriptome sequencing data, except in one gene: *CXC* 125-T had a missense variant mutation in *CXCL5* gene at c which was also one of the DEGs, overexpressed in OSCC compared to OVC. *CXCL5* has been previously found to be s regulated in head and neck cancer cell line (HN12) that was ol nodal metastasis (Miyazaki *et al.*, 2006). However, it is importar that RNAseq gene expression data were taken from each OVC final (groups) differential expression analysis. While in the exome data here, the missense variant mutation in *CXCL5* gene was only (V-125-T). Also, and as a matter of fact, gene-expressior represent gene-level data, and not all expressed genes ar mutations. Generally, the effect of mutation on the gene dependent on the type of the mutation and the type of the amin For example, with a splice site mutation that affects splicir transcription of the gene, no RNA expression would be expe missence changes might have no effect on the gene expression of the mutation will need functional assays to test the Alternatively, the identified mutated genes could have an indi cause changes in gene expression throughout different set probably act in the same signalling pathway. These points just gene mutations in the exome sequencing study were refle expression levels.

7.4 Discussion

It is known that molecular changes drive the cellular phenotype of any tumour. Until now, all the previously reported molecular studies of OVC lesions have inspected candidate genes rather than taking a complete genome wide approach. This study represents the largest and 'first' study, to date, to inspect somatic gene mutations that occur in OVC and compare them with mutational events in OSCC by performing high coverage whole exome sequencing (WES).

WES have greatly clarified the genetic alteration landscape in several tumour types and provided biological understandings relevant to clinical contexts (Garraway and Lander, 2013). In addition, the affordable cost for achieving higher coverage for large sample numbers makes exome sequencing highly suitable for mutation detection in mixed purity cancer samples (Meyerson *et al.*, 2010). In general, knowledge of changes in the coding exons regions of all genes could suggest treatment selections and further therapeutic options (Garraway and Janne, 2012). The study of HNSCC biology using NGS techniques has guided to a clearer understanding of the aetiological and molecular aspects of HNSCC (Rizzo *et al.*, 2014). The first next generation whole exome sequencing studies were performed in 2011 to identify the mutational landscape and events in key cell cycle components of HNSCC patient tumours (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011).

Exome sequencing data from OVC and adjacent normal tissues in this study has successfully identified a list of significant mutated genes in each OVC case. A total of 224 genes (18 potentially driver somatic mutated genes + 206 other mutated genes) contained putatively functional variants that were identified in OVC cases. However, this study has limitations; most notably, its small sample size due to the limited available funding for this project. Also, one of the main limitations was extracting DNA from histologically normal epithelium adjacent to the OVCs to be used as 'normal-matching samples' since no other source of 'normal' material was available for all patients (i.e. blood samples). Occult genetic abnormalities can arise within normal-appearing epithelium of cancer patients (Tripathi *et al.*, 2008), and hence, if an early mutation has already

occurred in a certain gene in the normal adjacent epithelium, it will not be detected in the malignant OVC tissue. Another very important limitation of WES is the inability to comprehensively signify genomic structural variations, particularly, when the structural variants are important for gene transcriptional regulation (Biesecker *et al.*, 2011). Additionally, exons constitute about 1% of the total human genome (~180,000 exons), while 99% of the genome (which is not yet functionally recognised) is not included in WES. So, if a certain variant is located in a distal regulatory region and has a main influence on a trait, it will be totally missed from the sequencing (Biesecker *et al.*, 2011). However, mutations in the exome regions are more likely to have significant consequences than the 99% remaining genome regions (Kumar *et al.*, 2010).

Moreover, the extracted DNA from each of the 12 OVC cases were divided for copy number analysis and WES experiments, which limited the option of studying biological replicates, for assessing any experimental errors. Furthermore, some exome sequencing studies used Sanger sequencing for the validation of the identified mutated genes (Chatterjee *et al.*, 2012), (Wang *et al.*, 2012a), (Thompson *et al.*, 2012). Unfortunately, this has not been applied here because of the limited available time and funding I have for this project. Again, and as explained in section 7.3.2 above, formalin-fixed tissues show a higher rate of non-reproducible DNA sequence changes than frozen tissues. In the work presented here, four samples (two tumours and their associated normals) were excluded from further analysis due to insufficient coverage and evidence of significant DNA damage caused by FFPE fixation (high mismatch rate, especially C/T changes).

7.4.1 Implications of mutated genes in the current study

Recent advances in cancer tissue genome sequencing have found that individual tumours harbour thousands of somatic variations (Pleasance *et al.*, 2010a), (Beroukhim *et al.*, 2010). These include several genetic changes, such as LOH, rearrangements, single-nucleotide substitutions, deletions, insertions, whole-chromosome deletions or duplications and CN alterations (Pleasance *et al.*, 2010a). It is generally believed that few of these variations trigger tumours

phenotypes; called driver mutations, while the vast majority of the variation events in cancer are called passenger mutations and are thought to have non-significant phenotypes (Beroukhim *et al.*, 2010), (McFarland *et al.*, 2013).

In the current study, 18 potentially driver mutated genes in Table 7.3 were found in more than 50% of OVC tumour cells, with Phred quality score of ≥15 These were either, tumour suppresser or oncogenes, involved in different pathways, and were previously identified in other cancer types. Therefore, mutations that have a consequence on protein function on those genes were expected to be driver mutations in the corresponding samples that they were identified in. However, and importantly, all these genes were not shared between the ten OVC samples, which does not confirm their involvement in the overall development process of OVCs.

In addition, a total of 206 genes contained putatively functional variants in OVC cases. These genes were not identified in Stransky *et al*. whole genome mutational profiling data of HNSCC or the Cancer Gene Census database, or involved in any KEGG signalling pathway. Accordingly, only the shared mutated genes between more than two samples from the ten OVC cases are suggested to have a role in the development of OVC lesions. These genes include: *DSPP* gene (mutated in 40% of OVC cohort), *MUC4*, *NEFH* and *ANP32E* (mutated in 30% of OVC cohort). Also, the four genes were mutated as well in at least one OSCC sample. To emphasise, all the four genes were described in previous cancer studies. Two earlier studies reported up-regulation of *DSPP* gene along with other genes in histologically aggressive and poorly differentiated OSCC and in some oral epithelial dysplasia (Ogbureke *et al.*, 2007), (Ogbureke *et al.*, 2010). Other studies also demonstrated a positive correlation between *MUC4* expression and pancreatic tumour growth and malignant progression (Swartz *et al.*, 2002), (Chaturvedi *et al.*, 2007). The *NEFH* gene was also down-regulated along with other genes in metastatic squamous cell lung carcinoma samples when compared with non-metastatic samples (Wang *et al.*, 2012b), and *ANP32E* expression was up-regulated in gastric cancer cells when compared with non-neoplastic epithelial cells (Tsukamoto *et al.*, 2008). To point out, all the three samples that shared *ANP32E* gene mutation had missense variant

mutations in exactly the same position (blue highlighted cells in Table 7.5), which makes this gene a strong candidate that might have a role in the development of oral verrucous tumours. Two samples shared *DSPP* inframe deletion mutations in exactly the same position, and two samples shared a*NEFH* inframe deletion mutations in exactly the same position as well (blue highlighted cells in Table 7.5). Nonetheless, it must be remembered that these genes, at best, can only report on the likelihood of the development of OVC and requires further investigation.

In addition, and from DAVID functional annotation analysis of the 224 mutated genes in OVC (Table 7.6), the majority of the enriched mutated genes in the ten OVC cases located in the plasma membrane part (31 genes), participated in cell and biological adhesions (14 genes), and were implemented in calcium ion binding (15 genes). The phenotypic alterations associated with cancer development, including cellular proliferation and adhesion, are often initiated or mediated by plasma membrane proteins, making these essential in the biological process (Harvey *et al.*, 2001). Dysregulation in cell adhesion genes is also expected as an early, cancer-related genetic change (Hanahan and Weinberg, 2011). Furthermore, Ca^{2+} signalling is involved in cancer cells tumorigenic dedifferentiation process either through intracellular Ca^{2+} alterations and/or through changes in extracellular Ca^{2+}-sensing receptors (Bikle *et al.*, 2004). Alterations in the activity or expression of Ca^{2+} pumps and channels could have a promoting role in cancer. For example, overexpression of Ca^{2+} plasma membrane channels increases Ca^{2+} influx and endorses Ca^{2+}-dependent proliferative pathways (Berridge *et al.*, 2003). Therefore, these cellular alterations explain the enrichment of the above GO terms in OVC. The biological processes: keratinocyte proliferation and cell proliferation were also significantly enriched in OVC cohort here. This is also expected since oral keratinocytes are assumed to be the cell of origin of squamous cell carcinomas (Scully and Bagan, 2009a).

DAVID functional enrichment analysis was performed as well on OSCC cohort (primary OSCC data belong to Pre-cancer Genomics Group). The majority of the enriched mutated genes in the 20 OSCC cases were located in the

cytoskeleton, participated in cell adhesion and cell death, and were implemented in nucleotide binding. Dysregulation in cellular proliferation and apoptosis leads to cancer development (Hanahan and Weinberg, 2011). In addition, cancers metastatic process requires tumour cells to leave the primary cancer and to gain invasive and migratory capabilities. In epithelial-mesenchymal transition (EMT) process, besides changing their cell adhesion properties, cancer cells involve developmental processes to attain invasive and migratory features that include an intense reorganisation of the actin cytoskeleton and the associated formation of membrane protrusions that are needed for invasive growth (Yilmaz and Christofori, 2009). Consequently, these cellular alterations explain the enrichment of the above GO terms in OSCC. Interestingly, OVC did not show any involvement of the mutated genes in cell death or apoptosis. Whereas multiple enriched gene ontology terms related to cell death were enriched in OSCC cases. This can be attributed to the lack of *TP53* mutation in OVC; which has been significantly involved in the apoptotic pathway, and were found in OSCC cases here.

7.4.2 Exome Sequencing Indicated that OVC tumour development was unrelated to the presence of the main HNSCC Mutations

Interestingly, all the ten OVC samples showed lack of any mutations within *TP53*, *CDKN2A, NOTCH1, NOTCH2* and *FAT1* genes; which has been previously shown to be associated with HNSCC and were frequently detected in OSCC cases here as indicated above (section 7.3.2.1).

7.3.4.1 Lack of *TP53* mutations

In human cancers, alteration in *TP53* gene is one of the most commonly found genetic variations. P53 protein inactivation could be triggered by the binding of p53 protein to other proteins or by mutation in *p53* gene (Lin *et al.*, 2011). Few previous studies have investigated p53 expression of p53 in VC of different anatomic sites including the penis, vulva, and skin (Noel *et al.*, 1996), (Adegboyega *et al.*, 2005a), (Stankiewicz *et al.*, 2009). P53 positive staining rate was reported to be, 87% for 15 penile VCs (Stankiewicz *et al.*, 2009), 75%

for eight skin VCs (Noel *et al.*, 1996). In other studies which have investigated p53 expression in head and neck VCs, P53 positive staining was reported in 40% for ten laryngeal VCs (Lopez-Amado *et al.*, 1996), 52% for 29 OVCs (Gimenez-Conti *et al.*, 1996), 100% for four cases of head and neck VHs (Wu *et al.*, 2002a), 88% for 26 H&N VCs (Wu *et al.*, 2002a), and 50% for ten OVC (Ogawa *et al.*, 2004). Also, from Chang *et al* IHC study in 2002; consistent absence of p53 staining was found in their OVH and OVC cases (Chang *et al.*, 2002). On the other hand, the expression of Ki67 and p53 was assessed by Klieb and Raphael in 28 OVH and 32 OVC cases (Klieb and Raphael, 2007a). More diffuse expression of Ki67 and p53 proteins were found from their study in OVC than that in OVH samples, and they suggested that, in difficult cases, this IHC panel might help in the diagnosis of both lesions (Klieb and Raphael, 2007a). Though, no significant differences were revealed in p53 expression between OVC and OVH samples in this study. In addition, it is unclear if this expression represents stabilized wild type p53 or the mutant protein.

The inconsistencies between the results of the previous studies could be attributed to differences in p53 protein staining techniques and evaluation methods as well as using different verrucous samples that might have incorrect histopathological diagnosis. In the conducted study here, all OVC obtained samples were carefully histopathologically diagnosed based on criteria recommended by WHO (Barnes L, 2005). Though, none of the ten OVC samples showed any mutations in *TP53* gene.

7.3.4.2 Lack of *CDKN2A* mutations

Mutations in *CDKN2A* gene were also detected in about 7% of HNSCC tumours by exome sequencing (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). *CDKN2A* gene, on chromosome 9p21, encodes for p16, the tumour suppressor protein. *CDKN2A* is known to be the second most commonly tumour suppresser mutated gene after *TP53* in HNSCC (Stransky *et al.*, 2011). Loss of *CDKN2A* and *TP53* genes decrease apoptosis and permit abnormal proliferation (Rothenberg and Ellisen, 2012). *CDKN2A* alterations have been identified through a series of different mechanisms including promoter methylation,

homozygous deletion, and absence of protein overexpression (Gonzalez *et al.*, 1997), (Shintani *et al.*, 2001), (Reed *et al.*, 1996a). A very recent conducted research reported frequent alteration in *CDKN2A* gene in oral tongue SCC tumours (Lim *et al.*, 2014). In 1999, the expression of the cell cycle-associated proteins including p16, was investigated in 15 OVCs, 44 OSCCs, 57 dysplasia, and ten normals. This study reported high expression of p16 in OVC (45%) when compared with OSCC (11%), and they attributed that to a possible relationship between HPV infections and OVC. Though, no HPV screening was carried on in this study (Saito *et al.*, 1999b). Importantly, mutation in *CDKN2A* gene has not been identified before in any verrucous carcinoma lesions, and here, all the ten OVC samples did not show any mutations in this gene.

7.3.4.3 Lack of *NOTCH1* and *NOTCH2* mutations

Two previous whole-exome sequencing studies conducted by two research teams in 2011 revealed high rates of NOTCH1 somatic mutations (11%-15%) on a total of approximately 100 HNSCC samples (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). Since the reported mutations in these studies were either, located at domains impacting ligand binding, or leading to potential protein inactivation, *NOTCH1* was suggested as a tumour suppressor gene in HNSCC (Agrawal *et al.*, 2011), (Stransky *et al.*, 2011). *Notch1* mutation was also recently reported to be common in a Chinese OSCC cohort (Song *et al.*, 2014). *NOTCH2* gene has the similar mechanism of action and structure with *NOTCH1* gene. Also, Stransky *et al*. identified non-synonymous point mutations in *NOTCH3* or *NOTCH2* genes in 11% of the HNSCC cases. *NOTCH1* and *NOTCH2* genes have no been inspected before in any verrucous carcinoma lesions, and here, all the ten OVC samples did not show any mutations in either gene.

7.3.4.4 Lack of *FAT1* mutations

Additionally, the identification of mutations in the *FAT1* gene (located on chromosome 4q35) in nine HNSCC samples (12%) in Stransky *et al*. study provided a new insight into the potential mechanisms of metastasis and

invasion in HNSCC. A homozygous loss of *FAT1* was previously identified in 2007 using CGH-array and it was suggested to act as a putative tumour suppressor gene in oral cancer that might also play a role in other SCCs types (Nakaya *et al.*, 2007). Again, mutation in *FAT1* gene has not been identified before in any verrucous carcinoma lesions, and here, all the ten OVC samples did not show any mutations in this gene.

In conclusion, the exome analysis of the significant mutated genes has highlighted a few 'candidates' that provide *prima facie* information about malignant progression for OVC, but additional functional studies as well as re-sequencing these genes in larger patient populations will be needed for more accurate conclusions. The lack of any mutations within the HNSCC suggested tumour suppressor genes (*TP53*, *CDKN2A, NOTCH1, NOTCH2* and *FAT1*) in the ten OVC cases is a significant finding that could underlie the more 'benign' behaviour of this tumour.

Chapter 8 Discussion

Generally speaking, OVC clinico-histopathological diagnosis is usually exclusionary and difficult, though, it has a better prognosis compared to other carcinomas (Ray *et al.*, 2011b). OVC is considered a histological subtype of oral carcinoma with a relatively indolent clinical course (Kademani, 2007, Cabay *et al.*, 2007, Zhu *et al.*, 2012). In 1980, the term verrucous hyperplasia of the oral mucosa was coined by Shear and Pindborg (Shear and Pindborg, 1980b), who demonstrated that 29% of OVH lesions also showed histological features of OVC. Very few studies have been published on OVH and the malignant transformation potential of VH lesions has not been inspected in detail (Shear and Pindborg, 1980b), (Hsue *et al.*, 2007), (Ho *et al.*, 2009). OVH is considered a histological precursor of OVC (Zhu *et al.*, 2012) that may transform into either an OVC or an OSCC (Wang *et al.*, 2009a). Distinguishing OVC from OVH lesions is difficult. In addition, distinguishing OVC from classical OSCC is a common problem for pathologists due to poorly defined diagnostic criteria. The aetiology of OVC is unknown, and the suggested role of HPV as a causative factor remains questionable. The rarity of these lesions makes them difficult to investigate, and most previous studies have been made on small numbers of cases.

Hence, and in light of the above considerations, the specific aims of the project were to use NGS copy number analysis at low coverage to: identify OVC and OVH genomic characteristic features and determine if CN analysis could distinguish between the genomic damage pattern in OVH, OVC, and OSCC. Furthermore, this project aimed to investigate the transcriptional changes and somatic genomic alterations that occur in OVC and compare them with OSCC using next generation RNA sequencing and whole-exome sequencing approaches respectively. Additionally, since a verrucous appearance is suggestive of viral aetiology, this study aimed as well to analyse a subset of oral verrucous lesions (including VC, and VH cases) for the presence of HPV subtypes and all known human viral genomes.

8.1 NGS copy number analysis to identify OVC and OVH genomic characteristic features

It has been previously demonstrated that NGS can provide genomic CN gain and loss details in a cost-effective manner from DNA isolated from FFPE tissue blocks stored after histopathological diagnosis (Wood *et al.*, 2010), (Hayes *et al.*, 2013). It has been also shown that the resolution of NGS CN method has a high degree of correlation and comparable with aCGH, but gives more information for less money with DNA extracted from FFPE materials, when applied at low multiplexing levels, and it is extremely adjustable (Wood *et al.*, 2010). Though it has some limitations (discussed in chapter four, section 4.4).

In this study, the NGS copy number analysis method was used on nanogram quantities of DNA isolated from FFPE tissue, in the largest oral verrucous sample cohort described to date. Visual examination of the 16 individual genomic CN traces as well as OVH frequency karyogram, GISTIC heat map and GISTIC amplification and deletion plots, all revealed considerably less genomic damage in OVH samples, compared to OVC (N: 57), indicating that the genomic profile of these cases has minimal chromosomal abnormalities and is more similar to normal. These findings are surprising since it has been well-known that OVH shares similar clinical and histological morphology to OVC, and the clinical differentiation of OVH from OVC is often difficult (Shear and Pindborg, 1980b), (Poh *et al.*, 2001), (Zhu *et al.*, 2012). From what has been found here, and despite the similar clinical and histological features that OVH and OVC share, the analysis of OVH individual CN karyograms showed that these lesions demonstrate different genomic profiles from OVC in that they present very low, narrow levels of DNA aneuploidy. Gains mapped at chromosome 7q11.2 and 7q22 were noticed in OVHs at a frequency of ~50%, suggesting that this CN alteration may be related to development of OVH.

Visual examination of OVC (N: 57) genomic CN frequency karyogram and GISTIC amplification and deletion plots revealed gains at chromosome arms 7q, 16q and 17q at a frequency of 50%, and these changes have not been previously identified as common CN altered chromosomes in oral cancer.

Nevertheless, deletion trends were minimally found in OVC's frequency karyogram, GISTIC heat map and GISTIC deletion plot, suggesting that overexpression of oncogenes is most likely to be involved in the development of OVC. The CN gain in OVH group at chromosome 7q arm, with a frequency of ~50%, was present as well in OVCs, suggesting that this region might harbour the first CN alteration involved in oral verrucous lesions (OVC and OVH).

8.2 The use of NGS copy number profiles to distinguish between the genomic damage pattern in OVH, OVC and OSCC

Visual examination of the 57 OVC CN traces revealed a lower level of genomic damage when compared with OSCC (N: 45). This suggests that OVC is characterized by a lower degree of chromosomal instability than OSCC. Similar results were obtained in 2011 by another study that investigated differences in chromosomal instability between OVC and OSCC using high-resolution DNA flow-cytometry and reported a lower degree of tumour heterogeneity and chromosomal instability in OVCs compared to OSCCs (Pentenero *et al.*, 2011). In addition, the minimal or absent histological cytological atypia of OVC (Zargaran *et al.*, 2012) could be another reason behind the lower level of chromosomal instability in OVCs. Interestingly, losses were detected more frequently in OSCCs than OVCs, suggesting that these CN alterations may be related to the more aggressive behaviour of OSCC.

Losses on chromosomes 3p, 4q, 9p and 18q, and gains on chromosomes 3q, 5q, 8q, and 20p are chromosomal signatures commonly linked with OSCCs (Baldwin *et al.*, 2005), (Snijders *et al.*, 2005), (Liu *et al.*, 2006), (Jarvinen *et al.*, 2008), (Freier *et al.*, 2010) and were frequently identified in OSCC cohort here but absent in oral verrucous lesions (OVC and OVH); suggesting that these CN alterations may be related to the more aggressive behaviour of OSCC. Additionally, the logistic regression analysis successfully distinguished between the two patient cohorts, correctly identifying 56 out of 57 OVC samples and 41 out of 45 OSCCs. This is further evidence that the two groups had distinct patterns of genomic damage.

Another key point, OVH is believed to be a precancerous lesion and may

'transform' into either OVC or OSCC (Chen *et al.*, 2004a), (Wang *et al.*, 2009a). However, visual examination of OSCC karyograms showed more gain and loss chromosomal abnormalities and much higher degree of aneuploidy across the whole genome when compared to OVCs. With this in mind, and because of significant differences between OVH and OSCC genomic profiles, I therefore suggest that OVH may transform to OVC rather than OSCC. GISTIC amplification and deletion plots of CNAs as well as GISTIC heat maps revealed the same results as the CN accumulative frequency karyograms for OVH, OVC and OSCC lesions in which they all clearly separated the three groups.

CN karyograms helped change initial diagnoses of five cases with an initial diagnosis of OVC. Visual examination of the CN karyograms for those samples revealed OSCC chromosomal signatures (mainly, loss in chromosome 3p arm and gain in 3q). A second, careful, blind, histological revision of those five cases by two different pathologists indicated that the final diagnosis is OSCC with a 'verrucous appearance' and these cases were excluded from this project. Copy number karyograms aided the revision of the pathological diagnosis in these five cases and hence I propose here that it could be useful for differential diagnosis of OVCs from OSCCs.

Considered together the distribution of next generation CN aberrations in OVC and OSCC suggest that there are two different distinct routes to two different oral cancers, one associated with greater genomic rearrangements and acquisition of previously known OSCC chromosomal alterations and signatures, while the other lacks these CN aberrations and with lower level of chromosomal instability. A similar aCGH CN study on oral SCC and dysplasia groups was conducted in 2011 and has distinguished two subtypes that are suggestive for the development of at least two oral cancer pathways, which differ, in chromosomal instability and metastasis risk (Bhattacharya *et al.*, 2011). They revealed two subtypes that were distinguished by acquisition of one or more specific CN alterations at four genomic regions: gain in 3q, loss in 8p, gain in 8q, and gain in chromosome 20 (Bhattacharya *et al.*, 2011). Furthermore, these subtypes were significantly differing in their metastasis risk, and accordingly they suggested that these variations in CN aberrations constitute a biomarker

with clinical value in identifying patients' treatment on the basis of cervical lymph node metastasis risk. Their results showed that neck metastasis was present in 22 of 48 (46%) of +3q-8p+8q+20 tumours and in only one of 15 (7%) of non +3q-8p+8q+20 tumours (Bhattacharya *et al.*, 2011).

Therefore & despite the WHO describing OVC as a variant of classical OSCC, CN results of this research project suggest that OVC is a distinct entity. I also demonstrated here that CN analysis could contribute to differential diagnosis of oral verrucous lesions and classical OSCCs using routine biopsy specimens.

8.3 NGS for detecting human papillomavirus in oral verrucous lesions

Since a verrucous appearance is suggestive of viral aetiology, a number of investigations to study the putative association between HPV and OVC have been undertaken (Stokes *et al.*, 2012, Kari J. Syrjänen, 2000). These have reported a wide range in the incidence of HPV in OVC (30 – 100%) leading to its actual role in OVC pathogenesis being controversial and inconclusive. This variation can be attributed to the deficiency of standardized detection procedures, such as PCR and ISH, and the difficulty in defining complete histological criteria for OVH and OVC cases. Furthermore, the rarity of these types of lesions makes it difficult to study and investigate, and most previous studies or case reports have been made on small number of cases. Next generation sequencing was described here as a novel but validated, powerful, high-throughput method to investigate the presence of HPV and all characterised human viral genomes loads and subtypes in the largest oral verrucous sample cohort described to date, following careful histological definition for OVC and OVH lesions. HPV-16 sequence was identified in one OVH and one OVC, and HPV-2 sequence was detected in one OVC out of 73 oral verrucous samples with low viral load levels [2.24, 8.16, and 0.33 viral genomes per cell] respectively; which confirmed that oral verrucous lesions are not associated with HPV or any other human virus. The method used here provides a digital readout of viral subtypes and loads with high sensitivity and specificity (Conway *et al.*, 2012). Despite that 30% HPV-positive samples have

been reported in a previous review of 5,338 OSCC patients (Kansy *et al.*, 2012); the study here suggested that oral verrucous lesions are not associated with HPV (Samman *et al.*, 2014).

8.4 Transcriptional and mutational events that occur in OVC in compare with the changes that occur in OSCC.

It is known that molecular changes drive cellular phenotype of any tumour. Recent advances in cancer tissue genome sequencing have also found that individual tumours harbour thousands of somatic variations (Pleasance *et al.*, 2010a), (Beroukhim *et al.*, 2010). This study represents the largest and 'first' study, to date, to inspect somatic gene mutations and transcriptional changes that occur in OVC and compare them with the mutational and transcriptional events in OSCC by performing high coverage whole exome sequencing and transcriptome sequencing, respectively. Again, and despite the proven utility of next generation exome sequencing and transcriptome sequencing techniques; it has some limitations, and these were discussed in chapter six, section 6.4, and chapter seven, section 7.4.

All earlier studies on OVC were either case reports or Immunohistochemistry studies. Previous immunohistochemistry studies on oral verrucous lesions have yielded mixed results (GimenezConti *et al.*, 1996), (Saito *et al.*, 1999a), (Sakurai *et al.*, 2000), (Chen *et al.*, 2002b), (Wu *et al.*, 2002b), (Impola *et al.*, 2004), (Oliveira *et al.*, 2005), (Angadi and Krishnapillai, 2007), (Klieb and Raphael, 2007a), (Arduino *et al.*, 2010), (Lin *et al.*, 2011), (Zargaran *et al.*, 2011a), (Odar *et al.*, 2012a), (Quan *et al.*, 2012), (Zargaran *et al.*, 2012), (Mohtasham *et al.*, 2013), (Habiba *et al.*, 2014). Again, the rarity of these lesions (including OVC and OVH) also makes them difficult to explore. Using different samples, sample numbers, variations in clinical diagnosis, difficulties in defining 'gold-standard' histological criteria for diagnosing verrucous lesions, and different staining procedures and analysis methods may explain the lack of concordance between these studies. Nevertheless, only one gene expression study was published in 2014 for patients with OVC and OSCC (discussed in details in chapter 6, section 6.4.2) (Wang *et al.*, 2014b). They reported eight differentially

expressed genes between OSCC and OVC (*HLF*, *TGFBI*, *SERPINE1*, *MMP1*, *INHBA*, *COL4A2*, *COL4A1*, and *ADAMTS12)*. They proposed that the eight genes might differentiate and determine the identity of the two tumours and further studies on much larger samples must be conducted. They also suggested that their results indicate that OVC is a rare variant of OSCC but with different molecular footprint than OSCC (Wang *et al.*, 2014b). Nonetheless, The eight genes that were proposed to differentiate between OVC and OSCC in Wang *et al* study were all also differentially expressed between OVC and OSCC cohorts in this study (refer to appendix 8.1).

The clinical behaviour of OVC is usually indolent and generally benign. However, till now, the characteristics of OVC were only described by the clinicopathological parameters, and no available well-defined gene expression profiles for this tumour. Functional characterisation of the significant expressed genes on an individual basis in this study has highlighted several candidate genes in which the expression status of those genes in the OVC cohort could suggest their potential roles. These genes were discussed in details in chapter six and are summarised in appendix 8.1. In addition, the exome analysis of the significant mutated genes here has highlighted few candidates that might provide important information about the malignant progression for OVC. These genes were discussed in details in chapter seven and are summarized as well in appendix 8.1. Nonetheless, these suggested genes require further investigation and additional functional studies as well as re-sequencing these genes in a larger patient population.

DAVID functional enrichment and pathway analysis on RNA and exome sequencing OVC genes data revealed that a significant number of the enriched differentially expressed genes and mutated genes are located in the plasma membrane, participate in cell adhesions and keratinocyte differentiation, and were involved in calcium ion (Ca^{2+}) binding. On the other hand, DAVID functional enrichment and pathway analysis on RNA and exome sequencing OSCC genes data revealed that a significant number of the enriched differentially expressed genes and mutated genes located in the cytoskeleton,

participated in cell adhesion and migration, angiogenesis and cell death, and were implemented in nucleotide and growth factor binding.

Clustering via principle component analysis, using data from expressed genes in OVC and OSCC similarly revealed an evident separation of the two cohorts. Additionally, and in light of the dissimilarities within the functional enrichment categories between OVC and OSCC, as well as gene expression differences between the two groups, and the lack of any mutations within HNSCC suggested TSGs (*TP53*, *CDKN2A, NOTCH1, NOTCH2* and *FAT1*) in the ten OVC cases indicates that there are two different routes to two different oral cancer types, one associated with greater genomic instability and acquisition of previously known OSCC gene expressions and mutations signatures, while the other lacks these aberrations.

Irrespective of the techniques used in this project (i.e. CN analysis, RNAseq and whole exome sequencing); OVC and OSCC tumours appeared differently, and they clearly separated. As before, and although the WHO describes OVC as a variant of classical OSCC, RNA and exome sequencing results of this project confirms the claim that OVC is a distinct entity.

An aim of the detailed analysis of cancer genomes is to discover genomic features that are prognostic for disease outcome and predictive of treatment response. Historically genomic and genetic changes at specific loci have provided this information (Brodeur *et al.*, 1984), (Slamon *et al.*, 1987) but more recently examination of the whole genome has been shown to have predictive power. Genomic analysis has been used to reclassify breast cancer tumours providing better prediction of disease outcomes (Dawson *et al.*, 2013) and recently, molecular classification using multiple platforms has effectively clustered subtypes across different cancer tissue types, identifying unexpected associations that could influence choice of therapy (Hoadley *et al.*, 2014). The wide application of NGS with powerful bioinformatics tools, mainly through transcriptome and exome sequencing, provides a high-resolution view of the cancer genome and promises improvements in cancer research, diagnosis and therapy (Guan *et al.*, 2012).Though, translation of NGS data into routine clinical practice still remains challenging.

Patients with oral cancer would benefit from a biomarker that would indicate outcome, specifically metastatic spread, as it is common practice to remove associated neck lymph nodes at the same time as the oral tumour resection even without radiological evidence of spread. This procedure can result in considerable morbidity and a clear indication of its requirement would be of advantage to management of this disease. The traditional treatment of OVC is complete surgical excision of the lesion (Ferlito and Recher, 1980), (McDonald *et al.*, 1982), (Kang *et al.*, 2003), (Arduino *et al.*, 2008) with radiation treatment applied in some conditions of either poor surgical candidates or for extensive disease cases (Kolokythas *et al.*, 2010). Unlike OSCC, surgical treatment of VC should not involve neck dissection (Thomas and Barrett, 2009), although lymph nodes enlargement might be palpated, whereas patients with a verrucous lesion that harbour conventional SCC must be treated as if they had invasive SCC (Sheen *et al.*, 2004), (Thomas and Barrett, 2009). A mis-diagnosis of a verrucous SCC [actually an OVC] will get over-treated with more aggressive surgery, radiotherapy possibly and maybe neck dissection, which will result in considerable morbidity. It has been shown here that copy number analysis could be a useful tool for the differential diagnosis of OVCs from OSCCs.

Verrucous carcinoma although rarely metastatic can be locally destructively invasive and may benefit from therapeutic approaches in addition to surgery. The results of this project, including: NGS copy number, RNA and exome-sequencing data should correct our understanding of the definition of OVC and subsequently aid in a more accurate diagnosis, prognosis and treatment of patients. The genomic and transcriptomic changes described here may suggest routes to the identification of drug target specific for these verrucous tumours.

References

Ackerman, L. V. 1948. Verrucous Carcinoma of the Oral Cavity. *Surgery,* 23**,** 670-678.

Ackerman, L. V. & Mc, G. M. 1958. Proliferating benign and malignant epithelial lesions of the oral cavity. *J Oral Surg (Chic),* 16**,** 400-13.

Adegboyega, P. A., Boromound, N. & Freeman, D. H. 2005a. Diagnostic utility of cell cycle and apoptosis regulatory proteins in verrucous squamous carcinoma. *Appl Immunohistochem Mol Morphol,* 13**,** 171-7.

Adegboyega, P. A., Boromound, N. & Freeman, D. H. 2005b. Diagnostic utility of cell cycle and apoptosis regulatory proteins in verrucous squamous carcinoma. *Applied Immunohistochemistry & Molecular Morphology,* 13**,** 171-177.

Adlerstorthz, K., Newland, J. R., Tessin, B. A., Yeudall, W. A. & Shillitoe, E. J. 1986. Human Papillomavirus Type-2 DNA in Oral Verrucous Carcinoma. *Journal of Oral Pathology & Medicine,* 15**,** 472-475.

Agarwal, R., Carey, M., Hennessy, B. & Mills, G. B. 2010. PI3K pathway-directed therapeutic strategies in cancer. *Current Opinion in Investigational Drugs,* 11**,** 615-628.

Agochiya, M., Brunton, V. G., Owens, D. W., Parkinson, E. K., Paraskeva, C., Keith, W. N. & Frame, M. C. 1999. Increased dosage and amplification of the focal adhesion kinase gene in human cancer cells. *Oncogene,* 18**,** 5646-53.

Agrawal, N., Frederick, M. J., Pickering, C. R., Bettegowda, C., Chang, K., Li, R. J., Fakhry, C., Xie, T. X., Zhang, J., Wang, J., Zhang, N., El-Naggar, A. K., Jasser, S. A., Weinstein, J. N., Trevino, L., Drummond, J. A., Muzny, D. M., Wu, Y., Wood, L. D., Hruban, R. H., Westra, W. H., Koch, W. M., Califano, J. A., Gibbs, R. A., Sidransky, D., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., Wheeler, D. A., Kinzler, K. W. & Myers, J. N. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science,* 333**,** 1154-7.

Alkan, A., Bulut, E., Gunhan, O. & Ozden, B. 2010a. Oral verrucous carcinoma: a study of 12 cases. *European journal of dentistry,* 4**,** 202-7.

Alkan, A., Bulut, E., Gunhan, O. & Ozden, B. 2010b. Oral verrucous carcinoma: a study of 12 cases. *Eur J Dent,* 4**,** 202-7.

Alkan, C., Coe, B. P. & Eichler, E. E. 2011. Applications of Next-Generation Sequencing Genome Structural Variation Discovery and Genotyping. *Nature Reviews Genetics,* 12**,** 363-375.

Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., Lehrach, H., Mardis, E. R., Mcvean, G. A., Nickerson, D., Peltonen, L., Schafer, A. J., Sherry, S. T., Wang, J., Wilson, R. K., Gibbs, R. A., Deiros, D., Metzker, M., Muzny, D., Reid, J., Wheeler, D., Wang, J., Li, J. X., Jian, M., Li, G., Li, R. Q., Liang, H. Q., Tian, G., Wang, B., Wang, J., Wang, W., Yang, H. M., Zhang, X. Q., Zheng, H. S., Lander, E. S., Altshuler, D. L., Ambrogio, L., Bloom, T., Cibulskis, K., Fennell, T. J., Gabriel, S. B., Jaffe, D. B., Shefler, E., Sougnez, C. L., Bentley, D. R., Gormley, N., Humphray, S., Kingsbury, Z., Koko-Gonzales, P., Stone, J., Mckernan, K. J., Costa, G. L., Ichikawa, J. K., Lee, C. C., Sudbrak, R., Lehrach, H., Borodina, T. A., Dahl, A., Davydov, A. N., Marquardt, P., Mertes, F., Nietfeld, W., Rosenstiel, P., Schreiber, S., Soldatov, A. V., Timmermann, B., Tolzmann, M., Egholm, M., Affourtit, J., Ashworth, D., Attiya, S., Bachorski, M., Buglione, E., Burke, A., Caprio, A., Celone, C., Clark, S., Conners, D., Desany, B., Gu, L., Guccione, L., Kao, K., Kebbel, A., Knowlton, J., Labrecque, M., Mcdade, L., Mealmaker, C., Minderman, M., Nawrocki, A., Niazi, F., Pareja, K., et al. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-1073.

Amaral, T. M. P., Da Silva Freire, A. R., Carvalho, A. L., Pinto, C. a. L. & Kowalski, L. P. 2004. Predictive factors of occult metastasis and prognosis of clinical stages I and II squamous cell carcinoma of the tongue and floor of the mouth. *Oral Oncology,* 40**,** 780-786.

Ambatipudi, S., Bhosale, P. G., Heath, E., Pandey, M., Kumar, G., Kane, S., Patil, A., Maru, G. B., Desai, R. S., Watt, F. M. & Mahimkar, M. B. 2013. Downregulation of keratin 76 expression during oral carcinogenesis of human, hamster and mouse. *PLoS One,* 8**,** e70688.

Ambatipudi, S., Gerstung, M., Pandey, M., Samant, T., Patil, A., Kane, S., Desai, R. S., Schaffer, A. A., Beerenwinkel, N. & Mahimkar, M. B. 2012. Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Genes Chromosomes & Cancer,* 51**,** 161-173.

Amer, M., Bull, C. A., Daouk, M. N., Mcarthur, P. D., Lundmark, G. J. & Elsenoussi, M. 1985. Shamma Usage and Oral-Cancer in Saudi-Arabia. *Annals of Saudi Medicine,* 5**,** 135-141.

Andrianifahanana, M., Moniaux, N., Schmied, B. M., Ringel, J., Friess, H., Hollingsworth, M. A., Buchler, M. W., Aubert, J. P. & Batra, S. K. 2001. Mucin (MUC) gene expression in human pancreatic adenocarcinoma and chronic pancreatitis: A potential role of MUC4 as a tumor marker of diagnostic significance. *Clinical Cancer Research,* 7**,** 4033-4040.

Angadi, P. V. & Krishnapillai, R. 2007. Cyclin D1 expression in oral squamous cell carcinoma and verrucous carcinoma: correlation with histological differentiation. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology,* 103**,** E30-E35.

Ankathil, R., Mathew, A., Joseph, F. & Nair, M. K. 1996. Is oral cancer susceptibility inherited? Report of five oral cancer families. *Eur J Cancer B Oral Oncol,* 32B**,** 63-7.

Arduino, P. G., Carrozzo, M., Pagano, M., Broccoletti, R., Scully, C. & Gandolfo, S. 2010. Immunohistochemical expression of basement membrane proteins of verrucous carcinoma of the oral mucosa. *Clin Oral Investig,* 14**,** 297-302.

Arduino, P. G., Carrozzo, M., Pagano, M., Gandolfo, S. & Broccoletti, R. 2008. Verrucous oral carcinoma: clinical findings and treatment outcomes in 74 patients in Northwest Italy. *Minerva Stomatol,* 57**,** 335-9, 339-41.

Arwert, E. N., Lal, R., Quist, S., Rosewell, I., Van Rooijen, N. & Watt, F. M. 2010. Tumor formation initiated by nondividing epidermal cells via an inflammatory infiltrate. *Proc Natl Acad Sci U S A,* 107**,** 19903-8.

Bagan, J., Sarrion, G. & Jimenez, Y. 2010a. Oral cancer: Clinical features. *Oral Oncol,* 46**,** 414-417.

Bagan, J., Scully, C., Jimenez, Y. & Martorell, M. 2010b. Proliferative verrucous leukoplakia: a concise update. *Oral Diseases,* 16**,** 328-332.

Bagan, J. V., Jimenez, Y., Murillo, J., Gavalda, C., Poveda, R., Scully, C., Alberola, T. M., Torres-Puente, M. & Perez-Alonso, M. 2007. Lack of association between proliferative verrucous leukoplakia and human papillomavirus infection. *Journal of Oral and Maxillofacial Surgery,* 65**,** 46-49.

Bagan, J. V., Jimenez-Soriano, Y., Diaz-Fernandez, J. M., Murillo-Cortes, J., Sanchis-Bielsa, J. M., Poveda-Roda, R. & Bagan, L. 2011. Malignant transformation of proliferative verrucous leukoplakia to oral squamous cell carcinoma: a series of 55 cases. *Oral Oncol,* 47**,** 732-5.

Balaram, P., Nalinakumari, K. R., Abraham, E., Balan, A., Hareendran, N. K., Bernard, H. U. & Chan, S. Y. 1995. Human papillomaviruses in 91 oral cancers from Indian betel quid chewers--high prevalence and multiplicity of infections. *International Journal of Cancer,* 61**,** 450-4.

Baldwin, C., Garnis, C., Zhang, L. W., Rosin, M. P. & Lam, W. L. 2005. Multiple microalterations detected at high frequency in oral cancer. *Cancer Research,* 65**,** 7561-7567.

Barker, N. 2008. The canonical Wnt/beta-catenin signalling pathway. *Methods Mol Biol,* 468**,** 5-15.

Barker, N. & Clevers, H. 2007. Mining the Wnt pathway for cancer therapeutics (vol 5, pg 997, 2006). *Nature Reviews Drug Discovery,* 6**,** 249-249.

Barnes L, E. J., Reichart P, Sidransky D 2005. *WHO Classification of Tumours, Pathology and genetics of head and neck tumours*, Lyon: IARC Press.

Barretina, J., Taylor, B. S., Banerji, S., Ramos, A. H., Lagos-Quintana, M., Decarolis, P. L., Shah, K., Socci, N. D., Weir, B. A., Ho, A., Chiang, D. Y., Reva, B., Mermel, C. H., Getz, G., Antipin, Y., Beroukhim, R., Major, J. E., Hatton, C., Nicoletti, R., Hanna, M., Sharpe, T., Fennell, T. J., Cibulskis, K., Onofrio, R. C., Saito, T., Shukla, N., Lau, C., Nelander, S., Silver, S. J., Sougnez, C., Viale, A., Winckler, W., Maki, R. G., Garraway, L. A., Lash, A., Greulich, H., Root, D. E., Sellers, W. R., Schwartz, G. K., Antonescu, C. R., Lander, E. S., Varmus, H. E., Ladanyi, M., Sander, C., Meyerson, M. & Singer, S. 2010. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet,* 42**,** 715-21.

Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S. Y., Perner, S., Verhaak, R. G., Kim, S. Y., Wardwell, L., Tamayo, P., Gat-Viks, I., Ramos, A. H., Woo, M. S., Weir, B. A., Getz, G., Beroukhim, R., O'kelly, M., Dutt, A., Rozenblatt-Rosen, O., Dziunycz, P., Komisarof, J., Chirieac, L. R., Lafargue, C. J., Scheble, V., Wilbertz, T., Ma, C. Q., Rao, S., Nakagawa, H., Stairs, D. B., Lin, L., Giordano, T. J., Wagner, P., Minna, J. D., Gazdar, A. F., Zhu, C. Q., Brose, M. S., Cecconello, I., Ribeiro, U., Marie, S. K., Dahl, O., Shivdasani, R. A., Tsao, M. S., Rubin, M. A., Wong, K. K., Regev, A., Hahn, W. C., Beer, D. G., Rustgi, A. K. & Meyerson, M. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature Genetics,* 41**,** 1238-U105.

Beder, L. B., Gunduz, M., Ouchida, M., Fukushima, K., Gunduz, E., Ito, S., Sakai, A., Nagai, N., Nishizaki, K. & Shimizu, K. 2003. Genome-wide analyses on loss of heterozygosity in head and neck squamous cell carcinomas. *Lab Invest,* 83**,** 99-105.

Bedi, G. C., Westra, W. H., Gabrielson, E., Koch, W. & Sidransky, D. 1996. Multiple head and neck tumors: Evidence for a common clonal origin. *Cancer Research,* 56**,** 2484-2487.

Beleford, D., Liu, Z. X., Rattan, R., Quagliuolo, L., Boccellino, M., Baldi, A., Maguire, J., Staub, J., Molina, J. & Shridhar, V. 2010. Methylation Induced Gene Silencing of HtrA3 in Smoking-Related Lung Cancer. *Clinical Cancer Research,* 16**,** 398-409.

Berger, M. F., Levin, J. Z., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., Johnson, L. A., Robinson, J., Verhaak, R. G., Sougnez, C., Onofrio, R. C., Ziaugra, L., Cibulskis, K., Laine, E., Barretina, J., Winckler, W., Fisher, D. E., Getz, G., Meyerson, M., Jaffe, D. B., Gabriel, S. B., Lander, E. S., Dummer, R., Gnirke, A., Nusbaum, C. & Garraway, L. A. 2010. Integrative analysis of the melanoma transcriptome. *Genome Res,* 20**,** 413-27.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J. C., Huang, J. H., Alexander, S., Du, J., Kau, T., Thomas, R. K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R. M., Demichelis, F., Hatton, C., Rubin, M. A., Garraway, L. A., Nelson, S. F., Liau, L., Mischel, P. S., Cloughesy, T. F., Meyerson, M., Golub, T. A., Lander, E. S., Mellinghoff, I. K. & Sellers, W. R. 2007. Assessing the significance of chromosomal aberrations in

cancer: Methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 20007-20012.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y. J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., Weir, B. A., Tanaka, K. E., Chiang, D. Y., Bass, A. J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F. J., Sasaki, H., Tepper, J. E., Fletcher, J. A., Tabernero, J., Baselga, J., Tsao, M. S., Demichelis, F., Rubin, M. A., Janne, P. A., Daly, M. J., Nucera, C., Levine, R. L., Ebert, B. L., Gabriel, S., Rustgi, A. K., Antonescu, C. R., Ladanyi, M., Letai, A., Garraway, L. A., Loda, M., Beer, D. G., True, L. D., Okamoto, A., Pomeroy, S. L., Singer, S., Golub, T. R., Lander, E. S., Getz, G., Sellers, W. R. & Meyerson, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature,* 463**,** 899-905.

Berridge, M. J., Bootman, M. D. & Roderick, H. L. 2003. Calcium signalling: dynamics, homeostasis and remodelling. *Nat Rev Mol Cell Biol,* 4**,** 517-29.

Berx, G., Becker, K. F., Hofler, H. & Van Roy, F. 1998. Mutations of the human E-cadherin (CDH1) gene. *Human Mutation,* 12**,** 226-237.

Bhatlekar, S., Fields, J. Z. & Boman, B. M. 2014. HOX genes and their role in the development of human cancers. *Journal of Molecular Medicine-Jmm,* 92**,** 811-823.

Bhattacharya, A., Roy, R., Snijders, A. M., Hamilton, G., Paquette, J., Tokuyasu, T., Bengtsson, H., Jordan, R. C., Olshen, A. B., Pinkel, D., Schmidt, B. L. & Albertson, D. G. 2011. Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. *Clin Cancer Res,* 17**,** 7024-34.

Bhattacharya, N., Roy, A., Roy, B., Roychoudhury, S. & Panda, C. K. 2009. MYC gene amplification reveals clinical association with head and neck squamous cell carcinoma in Indian patients. *J Oral Pathol Med,* 38**,** 759-63.

Biesecker, L. G., Shianna, K. V. & Mullikin, J. C. 2011. Exome sequencing: the expert view. *Genome Biol,* 12**,** 128.

Bignell, G. R., Greenman, C. D., Davies, H., Butler, A. P., Edkins, S., Andrews, J. M., Buck, G., Chen, L. N., Beare, D., Latimer, C., Widaa, S., Hinton, J., Fahey, C., Fu, B. Y., Swamy, S., Dalgliesh, G. L., Teh, B. T., Deloukas, P., Yang, F. T., Campbell, P. J., Futreal, P. A. & Stratton, M. R. 2010. Signatures of mutation and selection in the cancer genome. *Nature,* 463**,** 893-U61.

Bignell, G. R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K. W., Wei, W., Stratton, M. R., Futreal, P. A., Weber, B., Shapero, M. H. & Wooster, R. 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Research,* 14**,** 287-295.

Bignotti, E., Tassi, R. A., Calza, S., Ravaggi, A., Bandiera, E., Rossi, E., Donzelli, C., Pasinetti, B., Pecorelli, S. & Santin, A. D. 2007. Gene expression profile of ovarian serous papillary carcinomas: identification of metastasis-associated genes. *Am J Obstet Gynecol,* 196**,** 245 e1-11.

Bikle, D. D., Oda, Y. & Xie, Z. 2004. Calcium and 1,25(OH)2D: interacting drivers of epidermal differentiation. *J Steroid Biochem Mol Biol,* 89-90**,** 355-60.

Blouin, J. M., Penot, G., Collinet, M., Nacfer, M., Forest, C., Laurent-Puig, P., Coumoul, X., Barouki, R., Benelli, C. & Bortoli, S. 2011. Butyrate elicits a metabolic switch in human colon cancer cells by targeting the pyruvate dehydrogenase complex. *International Journal of Cancer,* 128**,** 2591-2601.

Bockmuhl, U., Petersen, S., Schmidt, S., Wolf, G., Jahnke, V., Dietel, M. & Petersen, I. 1997. Patterns of chromosomal alterations in metastasizing and nonmetastasizing primary head and neck carcinomas. *Cancer Res,* 57**,** 5213-6.

Bornstein, P. 1995. Diversity of Function Is Inherent in Matricellular Proteins - an Appraisal of Thrombospondin-1. *Journal of Cell Biology,* 130**,** 503-506.

Bornstein, S., White, R., Malkoski, S., Oka, M., Han, G. W., Cleaver, T., Reh, D., Andersen, P., Gross, N., Olson, S., Deng, C. X., Lu, S. L. & Wang, X. J. 2009. Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation. *Journal of Clinical Investigation,* 119**,** 3408-3419.

Bouquot, J. E., Weiland, L. H. & Kurland, L. T. 1988. Leukoplakia and carcinoma in situ synchronously associated with invasive oral/oropharyngeal carcinoma in Rochester, Minn., 1935-1984. *Oral Surg Oral Med Oral Pathol,* 65**,** 199-207.

Braakhuis, B. J. M., Snijders, P. J. F., Keune, W. J. H., Meijer, C. J. L. M., Ruijter-Schippers, H. J., Leemans, C. R. & Brakenhoff, R. H. 2004. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *Journal of the National Cancer Institute,* 96**,** 998-1006.

Brady, R. C. & Bernstein, D. I. 2004. Treatment of herpes simplex virus infections. *Antiviral Research,* 61**,** 73-81.

Bragulla, H. H. & Homberger, D. G. 2009. Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *J Anat,* 214**,** 516-59.

Brennan, J. A., Boyle, J. O., Koch, W. M., Goodman, S. N., Hruban, R. H., Eby, Y. J., Couch, M. J., Forastiere, A. A. & Sidransky, D. 1995a. Association between Cigarette-Smoking and Mutation of the P53 Gene in Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine,* 332**,** 712-717.

Brennan, J. A., Mao, L., Hruban, R. H., Boyle, J. O., Eby, Y. J., Koch, W. M., Goodman, S. N. & Sidransky, D. 1995b. Molecular Assessment of Histopathological Staging in Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine,* 332**,** 429-435.

Brigstock, D. R., Goldschmeding, R., Katsube, K., Lam, S. C. T., Lau, L. F., Lyons, K., Naus, C., Perbal, B., Riser, B., Takigawa, M. & Yeger, H. 2003. Proposal for a unified CCN nomenclature. *Journal of Clinical Pathology-Molecular Pathology,* 56**,** 127-128.

Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. 1984. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science,* 224**,** 1121-4.

Bromberg, J. & Darnell, J. E. 2000. The role of STATs in transcriptional control and their impact on cellular function. *Oncogene,* 19**,** 2468-2473.

Cabay, R. J., Morton, T. H., Jr. & Epstein, J. B. 2007. Proliferative verrucous leukoplakia and its progression to oral carcinoma: a review of the literature. *J Oral Pathol Med,* 36**,** 255-61.

Cabelguenne, A., Blons, H., De Waziers, I., Carnot, F., Houllier, A. M., Soussi, T., Brasnu, D., Beaune, P., Luccourreye, O. & Laurent-Puig, P. 2000. p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma: A prospective series. *Journal of Clinical Oncology,* 18**,** 1465-1473.

Cacalano, N., Le, D., Paranjpe, A., Wang, M. Y., Fernandez, A., Evazyan, T., Park, N. H. & Jewett, A. 2008. Regulation of IGFBP6 gene and protein is mediated by the inverse expression and function of c-jun N-terminal kinase (JNK) and NF kappa B in a model of oral tumor cells. *Apoptosis,* 13**,** 1439-1449.

Califano, J., Vanderriet, P., Westra, W., Nawroz, H., Clayman, G., Piantadosi, S., Corio, R., Lee, D., Greenberg, B., Koch, W. & Sidransky, D. 1996. Genetic progression model for head and neck cancer: Implications for field cancerization. *Cancer Research,* 56**,** 2488-2492.

Califano, J., Westra, W. H., Meininger, G., Corio, R., Koch, W. M. & Sidransky, D. 2000. Genetic progression and clonal relationship of recurrent premalignant head and neck lesions. *Clin Cancer Res,* 6**,** 347-52.

Camarena, L., Bruno, V., Euskirchen, G., Poggio, S. & Snyder, M. 2010. Molecular Mechanisms of Ethanol-Induced Pathogenesis Revealed by RNA-Sequencing. *Plos Pathogens,* 6.

Campisi, G., Giovannelli, L., Ammatuna, P., Capra, G., Colella, G., Di Liberto, C., Gandolfo, S., Pentenero, M., Carrozzo, M., Serpico, R. & D'angelo, M. 2004. Proliferative verrucous vs conventional leukoplakia: no significantly increased risk of HPV infection. *Oral Oncol,* 40**,** 835-40.

Cancer Genome Atlas Research, N. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature,* 455**,** 1061-8.

Cantile, M., Scognamiglio, G., Anniciello, A., Farina, M., Gentilcore, G., Santonastaso, C., Fulciniti, F., Cillo, C., Franco, R., Ascierto, P. A. & Botti, G. 2012. Increased HOX C13 expression in metastatic melanoma progression. *J Transl Med,* 10**,** 91.

Casalini, P., Iorio, M. V., Galmozzi, E. & Menard, S. 2004. Role of HER receptors family in development and differentiation. *J Cell Physiol,* 200**,** 343-50.

Chang, C. C., Shih, J. Y., Jeng, Y. M., Su, J. L., Lin, B. Z., Chen, S. T., Chau, Y. P., Yang, P. C. & Kuo, M. L. 2004. Connective tissue growth factor and its role in lung adenocarcinoma invasion and metastasis. *Journal of the National Cancer Institute,* 96**,** 364-375.

Chang, K. C., Su, I. J., Tsai, S. T., Shieh, D. B. & Jin, Y. T. 2002. Pathological features of betel quid-related oral epithelial lesions in taiwan with special emphasis on the tumor progression and human papillomavirus association. *Oncology,* 63**,** 362-9.

Chang, K. Y., Tsai, S. Y., Chen, S. H., Tsou, H. H., Yen, C. J., Liu, K. J., Fang, H. L., Wu, H. C., Chuang, B. F., Chou, S. W., Tang, C. K., Liu, S. Y., Lu, P. J., Yen, C. Y. & Chang, J. Y. 2013. Dissecting the EGFR-PI3K-AKT pathway in oral cancer highlights the role of the EGFR variant III and its clinical relevance. *Journal of Biomedical Science,* 20.

Chang, Y. C. & Yu, C. H. 2014. Successful treatment of oral verrucous hyperplasia with photodynamic therapy combined with cryotherapy--report of 3 cases. *Photodiagnosis Photodyn Ther,* 11**,** 127-9.

Chang, Z. Y., Sun, R., Ma, Y. S., Fu, D., Lai, X. L., Li, Y. S., Wang, X. H., Zhang, X. P., Lv, Z. W., Cong, X. L. & Li, W. P. 2014. Differential gene expression of the key signalling pathway in para-carcinoma, carcinoma and relapse human pancreatic cancer. *Cell Biochem Funct,* 32**,** 258-67.

Chaplet, M., Waltregny, D., Detry, C., Fisher, L. W., Castronovo, V. & Bellahcene, A. 2006. Expression of dentin sialophosphoprotein in human prostate cancer and its correlation with tumor aggressiveness. *International Journal of Cancer,* 118**,** 850-856.

Chatterjee, R., Ramos, E., Hoffman, M., Vanwinkle, J., Martin, D. R., Davis, T. K., Hoshi, M., Hmiel, S. P., Beck, A., Hruska, K., Coplen, D., Liapis, H., Mitra, R., Druley, T., Austin, P. & Jain, S. 2012. Traditional and targeted exome sequencing reveals common, rare and novel functional deleterious variants in RET-signaling complex in a cohort of living US patients with urinary tract malformations. *Hum Genet,* 131**,** 1725-38.

Chaturvedi, P., Singh, A. P., Moniaux, N., Senapati, S., Chakraborty, S., Meza, J. L. & Batra, S. K. 2007. MUC4 mucin potentiates pancreatic tumor cell proliferation, survival, and invasive properties and interferes with its interaction to extracellular matrix proteins. *Mol Cancer Res,* 5**,** 309-20.

Chen, B. S., Wang, M. R., Xu, X., Cai, Y., Xu, Z. X., Han, Y. L. & Wu, M. 2000. Transglutaminase-3, an esophageal cancer-related gene. *Int J Cancer,* 88**,** 862-5.

Chen, C., Mendez, E., Houck, J., Fan, W. H., Lohavanichbutr, P., Doody, D., Yueh, B., Futran, N. D., Upton, M., Farwell, D. G., Schwartz, S. M. & Zhao, L. P. 2008a.

Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiology Biomarkers & Prevention,* 17**,** 2152-2162.

Chen, H. M., Chen, C. T., Yang, H., Kuo, M. Y. P., Kuo, Y. S., Lan, W. H., Wang, Y. P., Tsai, T. & Chiang, C. P. 2004a. Successful treatment of oral verrucous hyperplasia with topical 5-aminolevulinic acid-mediated photodynamic therapy. *Oral oncology,* 40**,** 630-637.

Chen, K., Li, Y., Dai, Y., Li, J., Qin, Y., Zhu, Y., Zeng, T., Ban, X., Fu, L. & Guan, X. Y. 2013. Characterization of tumor suppressive function of cornulin in esophageal squamous cell carcinoma. *PLoS One,* 8**,** e68838.

Chen, Y. J., Chang, J. T. C., Liao, C. T., Wang, H. M., Yen, T. C., Chiu, C. C., Lu, Y. C., Li, H. F. & Cheng, A. J. 2008b. Head and neck cancer in the betel quid chewing area: recent advances in molecular carcinogenesis. *Cancer Science,* 99**,** 1507-1514.

Chen, Y. J., Lin, S. C., Kao, T., Chang, C. S., Hong, P. S., Shieh, T. M. & Chang, K. W. 2004b. Genome-wide profiling of oral squamous cell carcinoma. *Journal of Pathology,* 204**,** 326-332.

Chen, Y. K., Hsuen, S. S. & Lin, L. M. 2002a. Expression of inducible nitric oxide synthase in human oral premalignant epithelial lesions. *Archives of Oral Biology,* 47**,** 387-392.

Chen, Y. K., Hsuen, S. S. & Lin, L. M. 2002b. Increased expression of inducible nitric oxide synthase for human oral submucous fibrosis, verrucous hyperplasia, and verrucous carcinoma. *Int J Oral Maxillofac Surg,* 31**,** 419-22.

Chen, Y. K., Hsuen, S. S. & Lin, L. M. 2002c. Increased expression of inducible nitric oxide synthase for human oral submucous fibrosis, verrucous hyperplasia, and verrucous carcinoma. *International Journal of Oral and Maxillofacial Surgery,* 31**,** 419-422.

Chen, Y. K. & Lin, L. M. 1995. Immunohistochemical demonstration of epithelial glutathione S-transferase isoenzymes in normal, benign, premalignant and malignant human oral mucosa. *J Oral Pathol Med,* 24**,** 316-21.

Chiang, D. Y., Getz, G., Jaffe, D. B., O'kelly, M. J. T., Zhao, X. J., Carter, S. L., Russ, C., Nusbaum, C., Meyerson, M. & Lander, E. S. 2009. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods,* 6**,** 99-103.

Chiang, D. Y., Villanueva, A., Hoshida, Y., Peix, J., Newell, P., Minguez, B., Leblanc, A. C., Donovan, D. J., Thung, S. N., Sole, M., Tovar, V., Alsinet, C., Ramos, A. H., Barretina, J., Roayaie, S., Schwartz, M., Waxman, S., Bruix, J., Mazzaferro, V., Ligon, A. H., Najfeld, V., Friedman, S. L., Sellers, W. R., Meyerson, M. & Llovet, J. M. 2008. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Research,* 68**,** 6779-6788.

Chin, D., Boyle, G. M., Williams, R. M., Ferguson, K., Pandeya, N., Pedley, J., Campbell, C. M., Theile, D. R., Parsons, P. G. & Coman, W. B. 2005. Novel markers for poor prognosis in head and neck cancer. *International Journal of Cancer,* 113**,** 789-797.

Chu, P. G. & Weiss, L. M. 2002. Keratin expression in human tissues and neoplasms. *Histopathology,* 40**,** 403-39.

Chudova, D., Wilde, J. I., Wang, E. T., Wang, H., Rabbee, N., Egidio, C. M., Reynolds, J., Tom, E., Pagan, M., Rigl, C. T., Friedman, L., Wang, C. C., Lanman, R. B., Zeiger, M., Kebebew, E., Rosai, J., Fellegara, G., Livolsi, V. A. & Kennedy, G. C. 2010. Molecular Classification of Thyroid Nodules Using High-Dimensionality Genomic Data. *Journal of Clinical Endocrinology & Metabolism,* 95**,** 5296-5304.

Chung, C. H., Yang, Y. H., Wang, T. Y., Shieh, T. Y. & Warnakulasuriya, S. 2005. Oral precancerous disorders associated with areca quid chewing, smoking, and alcohol drinking in southern Taiwan. *Journal of Oral Pathology & Medicine,* 34**,** 460-466.

Cirulli, E. T., Singh, A., Shianna, K. V., Ge, D., Smith, J. P., Maia, J. M., Heinzen, E. L., Goedert, J. J. & Goldstein, D. B. 2010. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biology,* 11**,** R57.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., Mckernan, K. J. & Grimmond, S. M. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods,* 5**,** 613-9.

Cloos, J., Spitz, M. R., Schantz, S. P., Hsu, T. C., Zhang, Z. F., Tobi, H., Braakhuis, B. J. M. & Snow, G. B. 1996. Genetic susceptibility to head and neck squamous cell carcinoma. *Journal of the National Cancer Institute,* 88**,** 530-535.

Cogliano, V., Straif, K., Baan, R., Grosse, Y., Secretan, B. & El Ghissassi, F. 2004. Smokeless tobacco and tobacco-related nitrosamines. *Lancet Oncology,* 5**,** 708-708.

Coleman, P. 2002. Improving oral health care for the frail elderly: a review of widespread problems and best practices. *Geriatr Nurs,* 23**,** 189-99.

Conway, C., Chalkley, R., High, A., Maclennan, K., Berri, S., Chengot, P., Alsop, M., Egan, P., Morgan, J., Taylor, G. R., Chester, J., Sen, M., Rabbitts, P. & Wood, H. M. 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.

Coughlin, C. R., Scharer, G. H. & Shaikh, T. H. 2012. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Medicine,* 4.

Courtney, K. D., Corcoran, R. B. & Engelman, J. A. 2010. The PI3K Pathway As Drug Target in Human Cancer. *Journal of Clinical Oncology,* 28**,** 1075-1083.

Cronin, M., Pho, M., Dutta, D., Stephans, J. C., Shak, S., Kiefer, M. C., Esteban, J. M. & Baker, J. B. 2004. Measurement of gene expression in archival paraffin-embedded tissues - Development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *American Journal of Pathology,* 164**,** 35-42.

Crowel, D. L., Milo, G. E. & Shuler, C. F. 1999. Keratin 19 downregulation by oral squamous cell carcinoma lines increases invasive potential. *Journal of Dental Research,* 78**,** 1256-1263.

Da Silva, S. D., Ferlito, A., Takes, R. P., Brakenhoff, R. H., Valentin, M. D., Woolgar, J. A., Bradford, C. R., Rodrigo, J. P., Rinaldo, A., Hier, M. P. & Kowalski, L. P. 2011. Advances and applications of oral cancer basic research. *Oral Oncology,* 47**,** 783-791.

Dallosso, A. R., Hancock, A. L., Szemes, M., Moorwood, K., Chilukamarri, L., Tsai, H. H., Sarkar, A., Barasch, J., Vuononvirta, R., Jones, C., Pritchard-Jones, K., Royer-Pokora, B., Lee, S. B., Owen, C., Malik, S., Feng, Y., Frank, M., Ward, A., Brown, K. W. & Malik, K. 2009. Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. *PLoS Genet,* 5**,** e1000745.

Dame, H., Schuh, F. D. & Hull, D. C. 1974. Verrucous carcinoma of the buccal mucosa. *South Med J,* 67**,** 1070-2.

Daniel, B. T., Damato, K. L. & Johnson, J. 2004. Educational issues in oral care. *Semin Oncol Nurs,* 20**,** 48-52.

Das, B. R. & Nagpal, J. K. 2002. Understanding the biology of oral cancer. *Med Sci Monit,* 8**,** RA258-67.

Dawson, S. J., Rueda, O. M., Aparicio, S. & Caldas, C. 2013. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J,* 32**,** 617-28.

De Bruin, E. C., Van De Pas, S., Lips, E. H., Van Eijk, R., Van Der Zee, M. M. C., Lombaerts, M., Van Wezel, T., Marijnen, C. a. M., Van Krieken, J. H. J. M., Medema, J. P., Van De Velde, C. J. H., Eilers, P. H. C. & Peltenburg, L. T. C. 2005. Macrodissection versus microdissection of rectal carcinoma: minor influence of stroma cells to tumor cell gene expression profiles. *Bmc Genomics,* 6.

De Fraipont, F., Nicholson, A. C., Feige, J. J. & Van Meir, E. G. 2001. Thrombospondins and tumor angiogenesis. *Trends Mol Med,* 7**,** 401-7.

De Oliveira, L. R., Ribeiro-Silva, A. & Zucoloto, S. 2007. Prognostic impact of p53 and p63 immunoexpression in oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine,* 36**,** 191-197.

De Spindula, J. V., Da Cruz, A. D., Oton-Leite, A. F., Batista, A. C., Leles, C. R., Alencar, R. D. G., Saddi, V. A. & Mendonca, E. F. 2011. Oral squamous cell carcinoma versus oral verrucous carcinoma: an approach to cellular proliferation and negative relation to human papillomavirus (HPV). *Tumor Biology,* 32**,** 409-416.

Del Pino, M., Bleeker, M. C. G., Quint, W. G., Snijders, P. J. F., Meijer, C. J. L. M. & Steenbergen, R. D. M. 2012. Comprehensive analysis of human papillomavirus prevalence and the potential role of low-risk types in verrucous carcinoma. *Modern Pathology,* 25**,** 1354-1363.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. & Lempicki, R. A. 2003. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology,* 4.

Desai, B. V., Harmon, R. M. & Green, K. J. 2009. Desmosomes at a glance. *Journal of Cell Science,* 122**,** 4401-4407.

Devaney, K. O., Ferlito, A., Rinaldo, A., El-Naggar, A. K. & Barnes, L. 2011a. Verrucous carcinoma (carcinoma cuniculatum) of the head and neck: what do we know now that we did not know a decade ago? *European Archives of Oto-Rhino-Laryngology,* 268**,** 477-480.

Devaney, K. O., Ferlito, A., Rinaldo, A., El-Naggar, A. K. & Barnes, L. 2011b. Verrucous carcinoma (carcinoma cuniculatum) of the head and neck: what do we know now that we did not know a decade ago? *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies,* 268**,** 477-80.

Ding, L., Ellis, M. J., Li, S. Q., Larson, D. E., Chen, K., Wallis, J., Harris, C. C., Mclellan, M. D., Fulton, R. S., Fulton, L. L., Abbott, R. M., Hoog, J., Dooling, D. J., Koboldt, D. C., Schmidt, H., Kalicki, J., Zhang, Q. Y., Chen, L., Lin, L., Wendl, M. C., Mcmichael, J. F., Magrini, V. J., Cook, L., Mcgrath, S. D., Vickery, T. L., Appelbaum, E., Deschryver, K., Davies, S., Guintoli, T., Lin, L., Crowder, R., Tao, Y., Snider, J. E., Smith, S. M., Dukes, A. F., Sanderson, G. E., Pohl, C. S., Delehaunty, K. D., Fronick, C. C., Pape, K. A., Reed, J. S., Robinson, J. S., Hodges, J. S., Schierding, W., Dees, N. D., Shen, D., Locke, D. P., Wiechert, M. E., Eldred, J. M., Peck, J. B., Oberkfell, B. J., Lolofie, J. T., Du, F. Y., Hawkins, A. E., O'laughlin, M. D., Bernard, K. E., Cunningham, M., Elliott, G., Mason, M. D., Thompson, D. M., Ivanovich, J. L., Goodfellow, P. J., Perou, C. M., Weinstock, G. M., Aft, R., Watson, M., Ley, T. J., Wilson, R. K. & Mardis, E. R. 2010a. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature,* 464**,** 999-1005.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., Mclellan, M. D., Mcmichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., Mcgrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J.,

Graubert, T. A., Mardis, E. R., Wilson, R. K. & Dipersio, J. F. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature,* 481**,** 506-510.

Ding, L., Wendl, M. C., Koboldt, D. C. & Mardis, E. R. 2010b. Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Human Molecular Genetics,* 19**,** R188-R196.

Dolmans, D. E. J. G. J., Fukumura, D. & Jain, R. K. 2003. Photodynamic therapy for cancer. *Nature Reviews Cancer,* 3**,** 380-387.

Donetti, E., Bedoni, M., Boschini, E., Dellavia, C., Barajon, I. & Gagliano, N. 2005. Desmocollin 1 and desmoglein 1 expression in human epidermis and keratinizing oral mucosa: a comparative immunohistochemical and molecular study. *Arch Dermatol Res,* 297**,** 31-8.

Dornan, D., Bheddah, S., Newton, K., Ince, W., Frantz, G. D., Dowd, P., Koeppen, H., Dixit, V. M. & French, D. M. 2004. COP1, the negative regulator of p53, is overexpressed in breast and ovarian adenocarcinomas. *Cancer Research,* 64**,** 7226-7230.

Drummond, S. N., De Marco, L., Pordeus, I. D., Barbosa, A. A. & Gomez, R. S. 2002. TP53 codon 72 polymorphism in oral squamous cell carcinoma. *Anticancer Research,* 22**,** 3379-3381.

Dulak, A. M., Stojanov, P., Peng, S. Y., Lawrence, M. S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S. E., Shefler, E., Mckenna, A., Carter, S. L., Cibulskis, K., Sivachenko, A., Saksena, G., Voet, D., Ramos, A. H., Auclair, D., Thompson, K., Sougnez, C., Onofrio, R. C., Guiducci, C., Beroukhim, R., Zhou, Z. R., Lin, L., Lin, J., Reddy, R., Chang, A., Landrenau, R., Pennathur, A., Ogino, S., Luketich, J. D., Golub, T. R., Gabriel, S. B., Lander, E. S., Beer, D. G., Godfrey, T. E., Getz, G. & Bass, A. J. 2013. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genetics,* 45**,** 478-U37.

Eguchi, M., Eguchi-Ishimae, M., Seto, M., Morishita, K., Suzuki, K., Ueda, R., Ueda, K., Kamada, N. & Greaves, M. 2001. GPHN, a novel partner gene fused to MLL in a leukemia with t(11;14)(q23;q24). *Genes Chromosomes & Cancer,* 32**,** 212-221.

Emilion, G., Langdon, J. D., Speight, P. & Partridge, M. 1996. Frequent gene deletions in potentially malignant oral lesions. *British Journal of Cancer,* 73**,** 809-813.

Ernster, J. A., Sciotto, C. G., O'brien, M. M., Finch, J. L., Robinson, L. J., Willson, T. & Mathews, M. 2007. Rising incidence of oropharyngeal cancer and the role of oncogenic human Papilloma virus. *Laryngoscope,* 117**,** 2115-2128.

Estilo, C. L., O-Charoenrat, P., Talbot, S., Socci, N. D., Carlson, D. L., Ghossein, R., Williams, T., Yonekawa, Y., Ramanathan, Y., Boyle, J. O., Kraus, D. H., Patel, S., Shaha, A. R., Wong, R. J., Huryn, J. M., Shah, J. P. & Singh, B. 2009. Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *Bmc Cancer,* 9.

Etemadmoghadam, D., Defazio, A., Beroukhim, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M. B., Harnett, P., Lade, S., Akslen, L. A., Tinker, A. V., Locandro, B., Alsop, K., Chiew, Y. E., Traficante, N., Fereday, S., Johnson, D., Fox, S., Sellers, W., Urashima, M., Salvesen, H. B., Meyerson, M., Bowtell, D. & Grp, A. S. 2009. Integrated Genome-Wide DNA Copy Number and Expression Analysis Identifies Distinct Mechanisms of Primary Chemoresistance in Ovarian Carcinomas. *Clinical Cancer Research,* 15**,** 1417-1427.

Everly, D., Sharma-Walia, N., Sadagopan, S. & Chandran, B. 2012. Herpesviruses and Cancer. *Cancer Associated Viruses***,** 133-167.

Eversole, L. R. & Papanicolaou, S. J. 1983. Papillary and verrucous lesions of oral mucous membranes. *Journal of oral medicine,* 38**,** 3-13.

Ewing, B. & Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research,* 8**,** 186-194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research,* 8**,** 175-185.

Fakhry, C. & Gillison, M. L. 2006. Clinical implications of human papillomavirus in head and neck cancers. *Journal of Clinical Oncology,* 24**,** 2606-2611.

Fasanmade, A., Kwok, E. & Newman, L. 2007. Oral squamous cell carcinoma associated with khat chewing. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology,* 104**,** E53-E55.

Feldman, M. Y. 1973. Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog Nucleic Acid Res Mol Biol,* 13**,** 1-49.

Feng, L., Liu, H., Liu, Y., Lu, Z. K., Guo, G. W., Guo, S. P., Zheng, H. W., Gao, Y. N., Cheng, S. J., Wang, J., Zhang, K. T. & Zhang, Y. 2010. Power of Deep Sequencing and Agilent Microarray for Gene Expression Profiling Study. *Molecular Biotechnology,* 45**,** 101-110.

Ferlito, A. & Recher, G. 1980. Ackermans Tumor (Verrucous Carcinoma) of the Larynx - a Clinicopathologic Study of 77 Cases. *Cancer,* 46**,** 1617-1630.

Ferreira, D. C., Paiva, S. S. M., Carmo, F. L., Rocas, I. N., Rosado, A. S., Santos, K. R. N. & Siqueira, J. F. 2011. Identification of Herpesviruses Types 1 to 8 and Human Papillomavirus in Acute Apical Abscesses. *Journal of Endodontics,* 37**,** 10-16.

Ferrer, I., Armstrong, J., Capellari, S., Parchi, P., Arzberger, T., Bell, J., Budka, H., Strobel, T., Giaccone, G., Rossi, G., Bogdanovic, N., Fakai, P., Schmitt, A., Riederers, P., Al-Sarraj, S., Ravid, R. & Kretzschmar, H. 2007. Effects of formalin fixation, paraffin embedding, and time of storage on DNA preservation in brain tissue: A BrainNet Europe study. *Brain Pathology,* 17**,** 297-303.

Fettig, A., Pogrel, M. A., Silverman, S., Bramanti, T. E., Da Costa, M. & Regezi, J. A. 2000. Proliferative verrucous leukoplakia of the gingiva. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics,* 90**,** 723-730.

Fillies, T., Werkmeister, R., Packeisen, J., Brandt, B., Morin, P., Weingart, D., Joos, U. & Buerger, H. 2006. Cytokeratin 8/18 expression indicates a poor prognosis in squamous cell carcinomas of the oral cavity. *BMC Cancer,* 6**,** 10.

Firestein, R., Bass, A. J., Kim, S. Y., Dunn, I. F., Silver, S. J., Guney, I., Freed, E., Ligon, A. H., Vena, N., Ogino, S., Chheda, M. G., Tamayo, P., Finn, S., Shrestha, Y., Boehm, J. S., Jain, S., Bojarski, E., Mermel, C., Barretina, J., Chan, J. A., Baselga, J., Tabernero, J., Root, D. E., Fuchs, C. S., Loda, M., Shivdasani, R. A., Meyerson, M. & Hahn, W. C. 2008. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature,* 455**,** 547-U60.

Freier, K., Knoepfle, K., Flechtenmacher, C., Pungs, S., Devens, F., Toedt, G., Hofele, C., Joos, S., Lichter, P. & Radwirnmer, B. 2010. Recurrent Copy Number Gain of Transcription Factor SOX2 and Corresponding High Protein Expression in Oral Squamous Cell Carcinoma. *Genes Chromosomes & Cancer,* 49**,** 9-16.

Fridlyand, J., Snijders, A. M., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B. M., Jain, A. N., Mclennan, J., Ziegler, J., Chin, K., Devries, S., Feiler, H., Gray, J. W., Waldman, F., Pinkel, D. & Albertson, D. G. 2006. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer,* 6**,** 96.

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., Aoki, M., Hosono, N., Kubo, M., Miya, F., Arai, Y., Takahashi, H., Shirakihara, T., Nagasaki, M., Shibuya, T., Nakano, K., Watanabe-Makino, K., Tanaka, H., Nakamura, H., Kusuda, J., Ojima, H., Shimada, K., Okusaka, T., Ueno, M., Shigekawa, Y., Kawakami, Y., Arihiro, K., Ohdan, H., Gotoh, K., Ishikawa, O., Ariizumi, S., Yamamoto, M., Yamada, T., Chayama, K., Kosuge, T., Yamaue,

H., Kamatani, N., Miyano, S., Nakagama, H., Nakamura, Y., Tsunoda, T., Shibata, T. & Nakagawa, H. 2012. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics,* 44**,** 760-U182.

Fujita, S., Senba, M., Kumatori, A., Hayashi, T., Ikeda, T. & Toriyama, K. 2008. Human papillomavirus infection in oral verrucous carcinoma: genotyping analysis and inverse correlation with p53 expression. *Pathobiology,* 75**,** 257-64.

Garnis, C., Buys, T. P. H. & Lam, W. L. 2004. Genetic alteration and gene expression modulation during cancer progression. *Molecular Cancer,* 3.

Garraway, L. A. 2013. Genomics-Driven Oncology: Framework for an Emerging Paradigm. *Journal of Clinical Oncology,* 31**,** 1806-1814.

Garraway, L. A. & Janne, P. A. 2012. Circumventing Cancer Drug Resistance in the Era of Personalized Medicine. *Cancer Discovery,* 2**,** 214-226.

Garraway, L. A. & Lander, E. S. 2013. Lessons from the cancer genome. *Cell,* 153**,** 17-37.

Garrod, D. & Chidgey, M. 2008. Desmosome structure, composition and function. *Biochimica Et Biophysica Acta-Biomembranes,* 1778**,** 572-587.

Gerber, J. K., Richter, T., Kremmer, E., Adamski, J., Hofler, H., Balling, R. & Peters, H. 2002. Progressive loss of PAX9 expression correlates with increasing malignancy of dysplastic and cancerous epithelium of the human oesophagus. *J Pathol,* 197**,** 293-7.

Gibb, E. A., Enfield, K. S., Tsui, I. F., Chari, R., Lam, S., Alvarez, C. E. & Lam, W. L. 2011a. Deciphering squamous cell carcinoma using multidimensional genomic approaches. *Journal of skin cancer,* 2011**,** 541405.

Gibb, E. A., Enfield, K. S., Tsui, I. F., Chari, R., Lam, S., Alvarez, C. E. & Lam, W. L. 2011b. Deciphering squamous cell carcinoma using multidimensional genomic approaches. *J Skin Cancer,* 2011**,** 541405.

Gibcus, J. H., Menkema, L., Mastik, M. F., Hermsen, M., De Bock, G. H., Van Velthuysen, M. L. F., Takes, R. P., Kok, K., Marcos, C. a. A., Van Der Laan, B. F. a. M., Van Den Brekel, M. W. M., Langendijk, J. A., Kluin, P. M., Van Der Wal, J. E. & Schuuring, E. 2007. Amplicon mapping and expression profiling identify the fas-associated death domain gene as a new driver in the 11q13.3 amplicon in Laryngeal/Pharyngeal cancer. *Clinical Cancer Research,* 13**,** 6257-6266.

Gilbert, M. T. P., Haselkorn, T., Bunce, M., Sanchez, J. J., Lucas, S. B., Jewell, L. D., Van Marck, E. & Worobey, M. 2007. The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When? *Plos One,* 2.

Gillison, M. L., Koch, W. M., Capone, R. B., Spafford, M., Westra, W. H., Wu, L., Zahurak, M. L., Daniel, R. W., Viglione, M., Symer, D. E., Shah, K. V. & Sidransky, D. 2000. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst,* 92**,** 709-720.

Gimenez-Conti, I. B., Collet, A. M., Lanfranchi, H., Itoiz, M. E., Luna, M., Xu, H. J., Hu, S. X., Benedict, W. F. & Conti, C. J. 1996. p53, Rb, and cyclin D1 expression in human oral verrucous carcinomas. *Cancer,* 78**,** 17-23.

Gimenezconti, I. B., Collet, A. M., Lanfranchi, H., Itoiz, M. E., Luna, M., Xu, H. J., Hu, S. X., Benedict, W. F. & Conti, C. J. 1996. p53, Rb, and cyclin D1 expression in human oral verrucous carcinomas. *Cancer,* 78**,** 17-23.

Gires, O., Mack, B., Rauch, J. & Matthias, C. 2006. CK8 correlates with malignancy in leukoplakia and carcinomas of the head and neck. *Biochem Biophys Res Commun,* 343**,** 252-9.

Goessel, G., Quante, M., Hahn, W. C., Harada, H., Heeg, S., Suliman, Y., Doebele, M., Von Werder, A., Fulda, C., Nakagawa, H., Rustgi, A. K., Blum, H. E. & Opitz, O. G. 2005. Creating oral squamous cancer cells: A cellular model of oral-

esophageal carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 15599-15604.

Goetz, L., Bethel, K. & Topol, E. J. 2013. Rebooting Cancer Tissue Handling in the Sequencing Era Toward Routine Use of Frozen Tumor Tissue. *Jama-Journal of the American Medical Association,* 309**,** 37-38.

Gonzalez, H. E., Gujrati, M., Frederick, M., Henderson, Y., Arumugam, J., Spring, P. W., Mitsudo, K., Kim, H. W. & Clayman, G. L. 2003. Identification of 9 genes differentially expressed in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg,* 129**,** 754-9.

Gonzalez, M. V., Pello, M. F., Lopez-Larrea, C., Suarez, C., Menendez, M. J. & Coto, E. 1997. Deletion and methylation of the tumour suppressor gene p16/CDKN2 in primary head and neck squamous cell carcinoma. *J Clin Pathol,* 50**,** 509-12.

Gonzalez-Moles, M. A., Ruiz-Avila, I., Rodriguez-Archilla, A. & Martinez-Lara, I. 2000. Suprabasal expression of Ki-67 antigen as a marker for the presence and severity of oral epithelial dysplasia. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 22**,** 658-661.

Gopalakrishnan, R., Weghorst, C. M., Lehman, T. A., Calvert, R. J., Bijur, G., Sabourin, C. L. K., Mallery, S. R., Schuller, D. E. & Stoner, G. D. 1997. Mutated and wild-type p53 expression and HPV integration in proliferative verrucous leukoplakia and oral squamous cell carcinoma. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics,* 83**,** 471-477.

Grandis, J. R., Drenning, S. D., Zeng, Q., Watkins, S. C., Melhem, M. F., Endo, S., Johnson, D. E., Huang, L., He, Y. K. & Kim, J. D. 2000. Constitutive activation of Stat3 signaling abrogates apoptosis in squamous cell carcinogenesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 4227-4232.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D. P., Louis, D. N., Goldstraw, P., Nicholson, A. G., Brasseur, F., Looijenga, L., Weber, B. L., Chiew, Y. E., Defazio, A., Greaves, M. F., Green, A. R., Campbell, P., Birney, E., Easton, D. F., Chenevix-Trench, G., Tan, M. H., Khoo, S. K., Teh, B. T., Yuen, S. T., Leung, S. Y., Wooster, R., Futreal, P. A. & Stratton, M. R. 2007. Patterns of somatic mutation in human cancer genomes. *Nature,* 446**,** 153-8.

Gregg, C., Zhang, J., Butler, J. E., Haig, D. & Dulac, C. 2010a. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science,* 329**,** 682-5.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D. & Dulac, C. 2010b. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science,* 329**,** 643-8.

Guan, Y. F., Li, G. R., Wang, R. J., Yi, Y. T., Yang, L., Jiang, D., Zhang, X. P. & Peng, Y. 2012. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chinese Journal of Cancer,* 31**,** 463-470.

Gupta, P. C., Mehta, F. S., Daftary, D. K., Pindborg, J. J., Bhonsle, R. B., Jalnawalla, P. N., Sinor, P. N., Pitkar, V. K., Murti, P. R., Irani, R. R., Shah, H. T., Kadam, P. M., Iyer, K. S., Iyer, H. M., Hegde, A. K., Chandrashekar, G. K., Shiroff, B. C., Sahiar, B. E. & Mehta, M. N. 1980. Incidence rates of oral cancer and natural history of oral precancerous lesions in a 10-year follow-up study of Indian villagers. *Community Dent Oral Epidemiol,* 8**,** 283-333.

Gupta, P. C. & Ray, C. S. 2003. Smokeless tobacco and health in India and south Asia. *Respirology,* 8**,** 419-431.

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.

Ha, P. K., Pai, S. I., Westra, W. H., Gillison, M. L., Tong, B. C., Sidransky, D. & Califano, J. A. 2002. Real-time quantitative PCR demonstrates low prevalence of human papillomavirus type 16 in premalignant and malignant lesions of the oral cavity. *Clin Cancer Res,* 8**,** 1203-9.

Habiba, U., Kitamura, T., Yanagawa-Matsuda, A., Hida, K., Higashino, F., Ohiro, Y., Totsuka, Y. & Shindoh, M. 2014. Cytoplasmic expression of HuR may be a valuable diagnostic tool for determining the potential for malignant transformation of oral verrucous borderline lesions. *Oncology Reports,* 31**,** 1547-1554.

Haddad, R. I. & Shin, D. M. 2008. Recent advances in head and neck cancer. *New England Journal of Medicine,* 359**,** 1143-1154.

Haimovich, A. D. 2011. Methods, challenges, and promise of next-generation sequencing in cancer biology. *Yale J Biol Med,* 84**,** 439-46.

Hammarstedt, L., Lindquist, D., Dahlstrand, H., Romanitan, M., Onelov, L., Joneberg, J., Creson, N., Lindholm, J., Ye, W. M., Dalianis, T. & Munck-Wikland, E. 2006. Human papillomavirus as a risk factor for the increase in incidence of tonsillar cancer. *International Journal of Cancer,* 119**,** 2620-2623.

Han, J., Kioi, M., Chu, W. S., Kasperbauer, J. L., Strome, S. E. & Puri, R. K. 2009. Identification of potential therapeutic targets in human head & neck squamous cell carcinoma. *Head & Neck Oncology,* 1.

Hanahan, D. & Weinberg, R. A. 2011. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646-674.

Hansen, L. S., Olson, J. A. & Silverman, S. 1985. Proliferative Verrucous Leukoplakia - a Long-Term Study of 30 Patients. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics,* 60**,** 285-298.

Hansson, B. G., Rosenquist, K., Antonsson, A., Wennerberg, J., Schildt, E. B., Bladstrom, A. & Andersson, G. 2005. Strong association between infection with human papillomavirus and oral and oropharyngeal squamous cell carcinoma: A population-based case-control study in southern Sweden. *Acta Oto-Laryngologica,* 125**,** 1337-1344.

Hartwell, L. H. & Kastan, M. B. 1994a. Cell cycle control and cancer. *Science,* 266**,** 1821-8.

Hartwell, L. H. & Kastan, M. B. 1994b. Cell-Cycle Control and Cancer. *Science,* 266**,** 1821-1828.

Harvey, S., Zhang, Y., Landry, F., Miller, C. & Smith, J. W. 2001. Insights into a plasma membrane signature. *Physiological Genomics,* 5**,** 129-136.

Hashibe, M., Brennan, P., Chuang, S. C., Boccia, S., Castellsague, X., Chen, C., Curado, M. P., Dal Maso, L., Daudt, A. W., Fabianova, E., Fernandez, L., Wunsch-Filho, V., Franceschi, S., Hayes, R. B., Herrero, R., Kelsey, K., Koifman, S., La Vecchia, C., Lazarus, P., Levi, F., Lence, J. J., Mates, D., Matos, E., Menezes, A., Mcclean, M. D., Muscat, J., Eluf-Neto, J., Olshan, A. F., Purdue, M., Rudnai, P., Schwartz, S. M., Smith, E., Sturgis, E. M., Szeszenia-Dabrowska, N., Talamini, R., Wei, Q. Y., Winn, D. M., Shangina, O., Pilarska, A., Zhang, Z. F., Ferro, G., Berthiller, J. & Boffetta, P. 2009. Interaction between Tobacco and Alcohol Use and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology Biomarkers & Prevention,* 18**,** 541-550.

Hashibe, M., Mathew, B., Kuruvilla, B., Thomas, G., Sankaranarayanan, R., Parkin, D. M. & Zhang, Z. F. 2000. Chewing tobacco, alcohol, and the risk of erythroplakia. *Cancer Epidemiology Biomarkers & Prevention,* 9**,** 639-645.

Hashibe, M., Mckay, J. D., Curado, M. P., Oliveira, J. C., Koifman, S., Koifman, R., Zaridze, D., Shangina, O., Wunsch, V., Eluf, J., Levi, J. E., Matos, E., Lagiou, P., Lagiou, A., Benhamou, S., Bouchardy, C., Szeszenia-Dabrowska, N., Menezes, A., Dall'agnol, M. M., Merletti, F., Richiardi, L., Fernandez, L., Lence, J., Talamini, R., Barzan, L., Mates, D., Mates, I. N., Kjaerheim, K., Macfarlane, G. J., Macfarlane, T. V., Simonato, L., Canova, C., Holcatova, I., Agudo, A., Castellsague, X., Lowry, R., Janout, V., Kollarova, H., Conway, D. I., Mckinney, P. A., Znaor, A., Fabianova, E., Bencko, V., Lissowska, J., Chabrier, A., Hung, R. J., Gaborieau, V., Boffetta, P. & Brennan, P. 2008. Multiple ADH genes are associated with upper aerodigestive cancers. *Nature Genetics,* 40**,** 707-709.

Hashibe, M., Straif, K., Tashkin, D. P., Morgenstern, H., Greenland, S. & Zhang, Z. F. 2005. Epidemiologic review of marijuana use and cancer risk. *Alcohol,* 35**,** 265-275.

Hassan, N. M. M., Hamada, J., Kameyama, T., Tada, M., Nakagawa, K., Yoshida, S., Kashiwazaki, H., Yamazaki, Y., Suzuki, Y., Sasaki, A., Nagatsuka, H., Inoue, N. & Moriuchi, T. 2011. Increased Expression of the PRL-3 Gene in Human Oral Squamous Cell Carcinoma and Dysplasia Tissues. *Asian Pacific Journal of Cancer Prevention,* 12**,** 947-951.

Haya-Fernandez, M. C., Bagan, J. V., Murillo-Cortes, J., Poveda-Roda, R. & Calabuig, C. 2004. The prevalence of oral leukoplakia in 138 patients with oral squamous cell carcinoma. *Oral Dis,* 10**,** 346-8.

Hayes, J. L., Tzika, A., Thygesen, H., Berri, S., Wood, H. M., Hewitt, S., Pendlebury, M., Coates, A., Willoughby, L., Watson, C. M., Rabbitts, P., Roberts, P. & Taylor, G. R. 2013. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics,* 102**,** 174-181.

Hayette, S., Gadoux, M., Martel, S., Bertrand, S., Tigaud, I., Magaud, J. P. & Rimokh, R. 1998. FLRG (follistatin-related gene), a new target of chromosomal rearrangement in malignant blood disorders. *Oncogene,* 16**,** 2949-54.

Hennessey, P. T., Westra, W. H. & Califano, J. A. 2009. Human Papillomavirus and Head and Neck Squamous Cell Carcinoma: Recent Evidence and Clinical Implications. *Journal of Dental Research,* 88**,** 300-306.

Hittelman, W. N., Voravud, N., Shin, D. M., Lee, J. S., Ro, J. Y. & Hong, W. K. 1993. Early Genetic Changes during Upper Aerodigestive Tract Tumorigenesis. *Journal of Cellular Biochemistry***,** 233-236.

Ho, P. S., Chen, P. L., Warnakulasuriya, S., Shieh, T. Y., Chen, Y. K. & Huang, I. Y. 2009. Malignant transformation of oral potentially malignant disorders in males: a retrospective cohort study. *BMC Cancer,* 9**,** 260.

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., Leiserson, M. D. M., Niu, B. F., Mclellan, M. D., Uzunangelov, V., Zhang, J. S., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A. A., Van't Veer, L. J., Lopez-Bigas, N., Laird, P. W., Raphael, B. J., Ding, L., Robertson, A. G., Byers, L. A., Mills, G. B., Weinstein, J. N., Van Waes, C., Chen, Z., Collisson, E. A., Benz, C. C., Perou, C. M., Stuart, J. M. & Network, C. G. a. R. 2014. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell,* 158**,** 929-944.

Hogg, R. P., Honorio, S., Martinez, A., Agathanggelou, A., Dallol, A., Fullwood, P., Weichselbaum, R., Kuo, M. J., Maher, E. R. & Latif, F. 2002. Frequent 3p allele loss and epigenetic inactivation of the RASSF1A tumour suppressor gene from region 3p21.3 in head and neck squamous cell carcinoma. *European Journal of Cancer,* 38**,** 1585-1592.

Hostetter, G., Kim, S. Y., Savage, S., Gooden, G. C., Barrett, M., Zhang, J., Alla, L., Watanabe, A., Einspahr, J., Prasad, A., Nickoloff, B. J., Carpten, J., Trent, J., Alberts, D. & Bittner, M. 2010. Random DNA fragmentation allows detection of single-copy, single-exon alterations of copy number by oligonucleotide array CGH in clinical FFPE samples. *Nucleic Acids Research,* 38.

Hrstka, R., Coates, P. J. & Vojtesek, B. 2009. Polymorphisms in p53 and the p53 pathway: roles in cancer susceptibility and response to treatment. *Journal of Cellular and Molecular Medicine,* 13**,** 440-453.

Hsue, S. S., Wang, W. C., Chen, C. H., Lin, C. C., Chen, Y. K. & Lin, L. M. 2007. Malignant transformation in 1458 patients with potentially malignant oral mucosal disorders: a follow-up study based in a Taiwanese hospital. *J Oral Pathol Med,* 36**,** 25-9.

Huang, D. W., Sherman, B. T. & Lempicki, R. A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols,* 4**,** 44-57.

Huang, G. L., Li, B. K., Zhang, M. Y., Wei, R. R., Yuan, Y. F., Shi, M., Chen, X. Q., Huang, L., Zhang, H. Z., Liu, W., Huang, B. J., Li, H., Zheng, X. F., Luo, X. R. & Wang, H. Y. 2012. Allele loss and down-regulation of heparanase gene are associated with the progression and poor prognosis of hepatocellular carcinoma. *PLoS One,* 7**,** e44061.

Huang, S. & He, X. 2011. The role of microRNAs in liver cancer progression. *British Journal of Cancer,* 104**,** 235-240.

Ibrahim, E. M., Satti, M. B., Al Idrissi, H. Y., Higazi, M. M., Magbool, G. M. & Al Quorain, A. 1986. Oral cancer in Saudi Arabia: the role of alqat and alshammah. *Cancer Detect Prev,* 9**,** 215-8.

Impola, U., Uitto, V. J., Hietanen, J., Hakkinen, L., Zhang, L., Larjava, H., Isaka, K. & Saarialho-Kere, U. 2004. Differential expression of matrilysin-1 (MMP-7), 92 kD gelatinase (MMP-9), and metalloelastase (MMP-12) in oral verrucous and squamous cell cancer. *J Pathol,* 202**,** 14-22.

Inaba, T., Matsushime, H., Valentine, M., Roussel, M. F., Sherr, C. J. & Look, A. T. 1992. Genomic organization, chromosomal localization, and independent expression of human cyclin D genes. *Genomics,* 13**,** 565-74.

Islamian, J. P., Mohammadi, M. & Baradaran, B. 2014. Inhibition of human esophageal squamous cell carcinomas by targeted silencing of tumor enhancer genes: an overview. *Cancer Biol Med,* 11**,** 78-85.

Izzo, J. G., Papadimitrakopoulou, V. A., Liu, D. D., Den Hollander, P. L. C., Babenko, I. M., Keck, J., El-Naggar, A. K., Shin, D. M., Lee, J. J., Hong, W. K. & Hittelman, W. N. 2003. Cyclin D1 genotype, response to biochemoprevention, and progression rate to upper aerodigestive tract cancer. *Journal of the National Cancer Institute,* 95**,** 198-205.

Janssen, A. M., Van Duijn, W., Kubben, F. J., Griffioen, G., Lamers, C. B., Van Krieken, J. H., Van De Velde, C. J. & Verspaget, H. W. 2002. Prognostic significance of metallothionein in human gastrointestinal cancer. *Clin Cancer Res,* 8**,** 1889-96.

Jarvinen, A. K., Autio, R., Kilpinen, S., Saarela, M., Leivo, I., Grenman, R., Maekitie, A. A. & Monni, O. 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**,** 500-509.

Jemal, A., Siegel, R., Ward, E., Hao, Y. P., Xu, J. Q. & Thun, M. J. 2009. Cancer Statistics, 2009. *Ca-a Cancer Journal for Clinicians,* 59**,** 225-249.

Jia, M., Souchelnytskyi, N., Hellman, U., O'hare, M., Jat, P. S. & Souchelnytskyi, S. 2010. Proteome profiling of immortalization-to-senescence transition of human breast epithelial cells identified MAP2K3 as a senescence-promoting protein

which is downregulated in human breast cancer. *Proteomics Clinical Applications,* 4**,** 816-828.

Jiang, W. W., Fujii, H., Shirai, T., Mega, H. & Takagi, M. 2001. Accumulative increase of loss of heterozygosity from leukoplakia to foci of early cancerization in leukoplakia of the oral cavity. *Cancer,* 92**,** 2349-2356.

Jin, R. X., Chow, V. T. K., Tan, P. H., Dheen, S. T., Duan, W. & Bay, B. H. 2002. Metallothionein 2A expression is associated with cell proliferation in breast cancer. *Carcinogenesis,* 23**,** 81-86.

Johnson, N. W., Jayasekara, P. & Amarasinghe, A. a. H. K. 2011. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. *Periodontology 2000,* 57**,** 19-37.

Jordan, R. C. 1995. Verrucous carcinoma of the mouth. *Journal,* 61**,** 797-801.

Jordan, R. C. K., Bradley, G. & Slingerland, J. 1998. Reduced levels of the cell-cycle inhibitor p27(Kip1) in epithelial dysplasia and carcinoma of the oral cavity. *American Journal of Pathology,* 152**,** 585-590.

Ju, H. S., Lim, B., Kim, M., Noh, S. M., Kim, W. H., Ihm, C., Choi, B. Y., Kim, Y. S. & Kang, C. W. 2010. SERPINE1 Intron Polymorphisms Affecting Gene Expression Are Associated With Diffuse-Type Gastric Cancer Susceptibility. *Cancer,* 116**,** 4248-4255.

Kademani, D. 2007. Oral cancer. *Mayo Clin Proc,* 82**,** 878-87.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. 1992. Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors. *Science,* 258**,** 818-821.

Kang, C. J., Chang, J. T., Chen, T. M., Chen, I. H. & Liao, C. T. 2003. Surgical treatment of oral verrucous carcinoma. *Chang Gung Med J,* 26**,** 807-12.

Kansy, K., Thiele, O. & Freier, K. 2012. The role of human papillomavirus in oral squamous cell carcinoma: myth and reality. *Oral Maxillofac Surg*.

Karagozoglu, K. H., Buter, J., Leemans, C. R., Rietveld, D. H. F., Van Den Vijfeijken, S. & Van Der Waal, I. 2012. Subset of patients with verrucous carcinoma of the oral cavity who benefit from treatment with methotrexate. *British Journal of Oral & Maxillofacial Surgery,* 50**,** 513-518.

Karantza, V. 2011. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene,* 30**,** 127-138.

Kari J. Syrjänen, S. S., Stina M. Syrjänen 2000. *Papillomavirus infections in human pathology,* New York, J Wiley & Sons.

Katoh, M. & Katoh, M. 2003. Identification and characterization of human TIPARP gene within the CCNL amplicon at human chromosome 3q25.31. *International Journal of Oncology,* 23**,** 541-547.

Kauraniemi, P., Kuukasjarvi, T., Sauter, G. & Kallioniemi, A. 2003. Amplification of a 280-kilobase core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. *American Journal of Pathology,* 163**,** 1979-1984.

Kawamata, H., Furihata, T., Omotehara, F., Sakai, T., Horiuchi, H., Shinagawa, Y., Imura, J., Ohkura, Y., Tachibana, M., Kubota, K., Terano, A. & Fujimori, T. 2003. Identification of genes differentially expressed in a newly isolated human metastasizing esophageal cancer cell line, T.Tn-AT1, by cDNA microarray. *Cancer Science,* 94**,** 699-706.

Kim, I. G., Gorman, J. J., Park, S. C., Chung, S. I. & Steinert, P. M. 1993. The deduced sequence of the novel protransglutaminase E (TGase3) of human and mouse. *J Biol Chem,* 268**,** 12682-90.

Kim, S. K., Fan, Y. H., Papadimitrakopoulou, V., Clayman, G., Hittelman, W. N., Hong, W. K., Lotan, R. & Mao, L. 1996. DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. *Cancer Research,* 56**,** 2519-2521.

Klieb, H. B. & Raphael, S. J. 2007a. Comparative study of the expression of p53, Ki67, E-cadherin and MMP-1 in verrucous hyperplasia and verrucous carcinoma of the oral cavity. *Head Neck Pathol,* 1**,** 118-22.

Klieb, H. B. E. & Raphael, S. 2007b. Comparative study of the expression of ki67, p53, MMP-1 and E-cadherin in verrucous hyperplasia and verrucous carcinoma of the oral cavity. *Laboratory Investigation,* 87**,** 224A-225A.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., Mclellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research,* 22**,** 568-76.

Koliopanos, A., Friess, H., Di Mola, F. F., Tang, W. H., Kubulus, D., Brigstock, D., Zimmermann, A. & Buchler, M. W. 2002. Connective tissue growth factor gene expression alters tumor progression in esophageal cancer. *World Journal of Surgery,* 26**,** 420-427.

Kolokythas, A., Rogers, T. M. & Miloro, M. 2010. Hybrid verrucous squamous carcinoma of the oral cavity: treatment considerations based on a critical review of the literature. *Journal of oral and maxillofacial surgery : official journal of the American Association of Oral and Maxillofacial Surgeons,* 68**,** 2320-4.

Kondi-Paphitis, A., Deligeorgi-Politi, H., Liapis, A. & Plemenou-Frangou, M. 1998. Human papilloma virus in verrucus carcinoma of the vulva: an immunopathological study of three cases. *European Journal of Gynaecological Oncology,* 19**,** 319-320.

Kowalski, L. P., Carvalho, A. L., Priante, A. V. M. & Magrin, J. 2005. Predictive factors for distant metastasis from oral and oropharyngeal squamous cell carcinoma. *Oral Oncology,* 41**,** 534-541.

Kozarewa, I., Rosa-Rosa, J. M., Wardell, C. P., Walker, B. A., Fenwick, K., Assiotis, I., Mitsopoulos, C., Zvelebil, M., Morgan, G. J., Ashworth, A. & Lord, C. J. 2012. A Modified Method for Whole Exome Resequencing from Minimal Amounts of Starting DNA. *Plos One,* 7.

Krawitz, P. M., Schweiger, M. R., Rodelsperger, C., Marcelis, C., Kolsch, U., Meisel, C., Stephani, F., Kinoshita, T., Murakami, Y., Bauer, S., Isau, M., Fischer, A., Dahl, A., Kerick, M., Hecht, J., Kohler, S., Jager, M., Grunhagen, J., De Condor, B. J., Doelken, S., Brunner, H. G., Meinecke, P., Passarge, E., Thompson, M. D., Cole, D. E., Horn, D., Roscioli, T., Mundlos, S. & Robinson, P. N. 2010. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet,* 42**,** 827-9.

Kreimer, A. R., Clifford, G. M., Boyle, P. & Franceschi, S. 2005. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev,* 14**,** 467-75.

Kridel, R., Meissner, B., Rogic, S., Boyle, M., Telenius, A., Woolcock, B., Gunawardana, J., Jenkins, C., Cochrane, C., Ben-Neriah, S., Tan, K., Morin, R. D., Opat, S., Sehn, L. H., Connors, J. M., Marra, M. A., Weng, A. P., Steidl, C. & Gascoyne, R. D. 2012. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood,* 119**,** 1963-71.

Ku, T. K., Nguyen, D. C., Karaman, M., Gill, P., Hacia, J. G. & Crowe, D. L. 2007. Loss of p53 expression correlates with metastatic phenotype and transcriptional profile in a new mouse model of head and neck cancer. *Mol Cancer Res,* 5**,** 351-62.

Kulkarni, V. & Saranath, D. 2004. Concurrent hypermethylation of multiple regulatory genes in chewing tobacco associated oral squamous cell carcinomas and adjacent normal tissues. *Oral Oncology,* 40**,** 145-153.

Kumar, A., Turner, E. & Shendure, J. 2010. Targeted Capture and Massively Parallel Sequencing of the Human Exome. *Journal of Investigative Medicine,* 58**,** 123-123.

Kushner, J., Bradley, G. & Jordan, R. C. K. 1997. Patterns of p53 and Ki-67 protein expression in epithelial dysplasia from the floor of the mouth. *Journal of Pathology,* 183**,** 418-423.

Lee, J. J., Hong, W. K., Hittelman, W. N., Mao, L., Lotan, R., Shin, D. M., Benner, S. E., Xu, X. C., Lee, J. S., Papadimitrakopoulou, V. M., Geyer, C., Perez, C., Martin, J. W., El-Naggar, A. K. & Lippman, S. M. 2000. Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin Cancer Res,* 6**,** 1702-10.

Lee, J. S., Kim, S. Y., Hong, W. K., Lippman, S. M., Ro, J. Y., Gay, M. L. & Hittelman, W. N. 1993. Detection of Chromosomal Polysomy in Oral Leukoplakia, a Premalignant Lesion. *Journal of the National Cancer Institute,* 85**,** 1951-1954.

Leemans, C. R., Braakhuis, B. J. & Brakenhoff, R. H. 2011a. The molecular biology of head and neck cancer. *Nat Rev Cancer,* 11**,** 9-22.

Leemans, C. R., Braakhuis, B. J. M. & Brakenhoff, R. H. 2011b. The molecular biology of head and neck cancer. *Nature Reviews Cancer,* 11**,** 9-22.

Leong, P. L., Andrews, G. A., Johnson, D. E., Dyer, K. F., Xi, S. C., Mai, J. C., Robbins, P. D., Gadiparthi, S., Burke, N. A., Watkins, S. F. & Grandis, J. R. 2003. Targeted inhibition of Stat3 with a decoy oligonucleotide abrogates head and neck cancer cell growth. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 4138-4143.

Lewis, F., Maughan, N. J., Smith, V., Hillan, K. & Quirke, P. 2001. Unlocking the archive - gene expression in paraffin-embedded tissue. *Journal of Pathology,* 195**,** 66-71.

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., Mclellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., Mcgrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D. E., Koboldt, D. C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L. W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J. R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W. D., Nagarajan, R., Walter, M. J., Link, D. C., Graubert, T. A., Dipersio, J. F. & Wilson, R. K. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature,* 456**,** 66-72.

Li, H., Cai, Q., Wu, H., Vathipadiekal, V., Dobbin, Z. C., Li, T. Y., Hua, X., Landen, C. N., Birrer, M. J., Sanchez-Beato, M. & Zhang, R. G. 2012. SUZ12 Promotes Human Epithelial Ovarian Cancer by Suppressing Apoptosis via Silencing HRK. *Molecular Cancer Research,* 10**,** 1462-1472.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 60.

Liao, W. T., Li, T. T., Wang, Z. G., Wang, S. Y., He, M. R., Ye, Y. P., Qi, L., Cui, Y. M., Wu, P., Jiao, H. L., Zhang, C., Xie, Y. J., Wang, J. X. & Ding, Y. Q. 2013. microRNA-224 promotes cell proliferation and tumor growth in human colorectal cancer by repressing PHLPP1 and PHLPP2. *Clin Cancer Res,* 19**,** 4662-72.

Lim, A. M., Do, H., Young, R. J., Wong, S. Q., Angel, C., Collins, M., Takano, E. A., Corry, J., Wiesenfeld, D., Kleid, S., Sigston, E., Lyons, B., Fox, S. B., Rischin, D., Dobrovic, A. & Solomon, B. 2014. Differential mechanisms of CDKN2A (p16) alteration in oral tongue squamous cell carcinomas and correlation with patient outcome. *International Journal of Cancer,* 135**,** 887-895.

Lin, B. R., Chang, C. C., Che, T. F., Chen, S. T., Chen, R. J. C., Yang, C. Y., Jeng, Y. M., Liang, J. T., Lee, P. H., Chang, K. J., Chau, Y. P. & Kuo, M. L. 2005. Connective tissue growth factor inhibits metastasis and acts as an independent prognostic marker in colorectal cancer. *Gastroenterology,* 128**,** 9-23.

Lin, H. P., Chen, H. M., Yu, C. H., Yang, H. A., Wang, Y. P. & Chiang, C. P. 2010. Topical photodynamic therapy is very effective for oral verrucous hyperplasia and oral erythroleukoplakia. *Journal of Oral Pathology & Medicine,* 39**,** 624-630.

Lin, H. P., Wang, Y. P. & Chiang, C. P. 2011. Expression of p53, MDM2, p21, heat shock protein 70, and HPV 16/18 E6 proteins in oral verrucous carcinoma and oral verrucous hyperplasia. *Head Neck,* 33**,** 334-40.

Lin, W. M., Baker, A. C., Beroukhim, R., Winckler, W., Feng, W., Marmion, J. M., Laine, E., Greulich, H., Tseng, H., Gates, C., Hodi, F. S., Dranoff, G., Sellers, W. R., Thomas, R. K., Meyerson, M., Golub, T. R., Dummer, R., Herlyn, M., Getz, G. & Garraway, L. A. 2008. Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Research,* 68**,** 664-673.

Lindberg, P., Larsson, A. & Nielsen, B. S. 2006. Expression of plasminogen activator inhibitor-1, urokinase receptor and laminin gamma-2 chain is an early coordinated event in incipient oral squamous cell carcinoma. *International Journal of Cancer,* 118**,** 2948-2956.

Link, D. C., Schuettpelz, L. G., Shen, D., Wang, J. L., Walter, M. J., Kulkarni, S., Payton, J. E., Ivanovich, J., Goodfellow, P. J., Le Beau, M., Koboldt, D. C., Dooling, D. J., Fulton, R. S., Bender, R. H. F., Fulton, L. L., Delehaunty, K. D., Fronick, C. C., Appelbaum, E. L., Schmidt, H., Abbott, R., O'laughlin, M., Chen, K., Mclellan, M. D., Varghese, N., Nagarajan, R., Heath, S., Graubert, T. A., Ding, L., Ley, T. J., Zambetti, G. P., Wilson, R. K. & Mardis, E. R. 2011. Identification of a Novel TP53 Cancer Susceptibility Mutation Through Whole-Genome Sequencing of a Patient With Therapy-Related AML. *Jama-Journal of the American Medical Association,* 305**,** 1568-1576.

Lippman, S. M., Shin, D. M., Lee, J. J., Batsakis, J. G., Lotan, R., Tainsky, M. A., Hittelman, W. N. & Hong, W. K. 1995. p53 and retinoid chemoprevention of oral carcinogenesis. *Cancer Res,* 55**,** 16-9.

Liu, C. J., Lin, S. C., Chen, Y. J., Chang, K. M. & Chang, K. W. 2006. Array-comparative genomic hybridization to detect genomewide changes in microdissected primary and metastatic oral squamous cell carcinomas. *Molecular Carcinogenesis,* 45**,** 721-731.

Liu, W., Wang, Y. F., Zhou, H. W., Shi, P., Zhou, Z. T. & Tang, G. Y. 2010. Malignant transformation of oral leukoplakia: a retrospective cohort study of 218 Chinese patients. *Bmc Cancer,* 10.

Lopez-Amado, M., Garcia-Caballero, T., Lozano-Ramirez, A. & Labella-Caballero, T. 1996. Human papillomavirus and p53 oncoprotein in verrucous carcinoma of the larynx. *J Laryngol Otol,* 110**,** 742-7.

Lui, V. W. Y., Hedberg, M. L., Li, H., Vangara, B. S., Pendleton, K., Zeng, Y., Lu, Y. L., Zhang, Q. H., Du, Y., Gilbert, B. R., Freilino, M., Sauerwein, S., Peyser, N. D., Xiao, D., Diergaarde, B., Wang, L., Chiosea, S., Seethala, R., Johnson, J. T., Kim, S., Duvvuri, U., Ferris, R. L., Romkes, M., Nukui, T., Ng, P. K. S., Garraway, L. A., Hammerman, P. S., Mills, G. B. & Grandis, J. R. 2013. Frequent Mutation of the PI3K Pathway in Head and Neck Cancer Defines Predictive Biomarkers. *Cancer Discovery,* 3**,** 761-769.

Lumerman, H., Freedman, P. & Kerpel, S. 1995. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod,* 79**,** 321-9.

Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. & Chinnaiyan, A. M. 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature,* 458**,** 97-101.

Mao, L. 1997. Leukoplakia: molecular understanding of pre-malignant lesions and implications for clinical management. *Molecular Medicine Today,* 3**,** 442-448.

Mao, L., Hong, W. K. & Papadimitrakopoulou, V. A. 2004. Focus on head and neck cancer. *Cancer Cell,* 5**,** 311-316.

Mao, L., Lee, J. S., Fan, Y. H., Ro, J. Y., Batsakis, J. G., Lippman, S., Hittelman, W. & Hong, W. K. 1996. Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. *Nature Medicine,* 2**,** 682-685.

Marchetti, A., Felicioni, L. & Buttitta, F. 2006. Assessing EGFR mutations. *New England Journal of Medicine,* 354**,** 526-527.

Marcu, L. G. & Yeoh, E. 2009. A review of risk factors and genetic alterations in head and neck carcinogenesis and implications for current and future approaches to treatment. *J Cancer Res Clin Oncol,* 135**,** 1303-14.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., Mclellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., Mcgrath, S. D., Fulton, L. A., Locke, D. P., Magrini, V. J., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F. Y., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., Mcmichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X. Q., Lin, L., Schmidt, H., Tang, Y. Z., Haipek, C., Wiechert, M. E., Ivy, J. V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath, S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., Dipersio, J. F., Wilson, R. K. & Ley, T. J. 2009. Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *New England Journal of Medicine,* 361**,** 1058-1066.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res,* 18**,** 1509-17.

Marklund, L. & Hammarstedt, L. 2011. Impact of HPV in Oropharyngeal Cancer. *J Oncol,* 2011**,** 509036.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17.

Mashberg, A., Merletti, F., Boffetta, P., Gandolfo, S., Ozzello, F., Fracchia, F. & Terracini, B. 1989. Appearance, Site of Occurrence, and Physical and Clinical Characteristics of Oral-Carcinoma in Torino, Italy. *Cancer,* 63**,** 2522-2527.

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.

Matthias, C., Mack, B., Berghaus, A. & Gires, O. 2008. Keratin 8 expression in head and neck epithelia. *BMC Cancer,* 8**,** 267.

Mccoy, J. M. & Waldron, C. A. 1981. Verrucous carcinoma of the oral cavity. A review of forty-nine cases. *Oral Surg Oral Med Oral Pathol,* 52**,** 623-9.

Mccullough, M. J. & Farah, C. S. 2008. The role of alcohol in oral carcinogenesis with particular reference to alcohol-containing mouthwashes. *Australian Dental Journal,* 53**,** 302-305.

Mcdonald, J. S., Crissman, J. D. & Gluckman, J. L. 1982. Verrucous carcinoma of the oral cavity. *Head Neck Surg,* 5**,** 22-8.

Mcfarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. & Mirny, L. A. 2013. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A,* 110**,** 2910-5.

Mcilroy, M., Mccartan, D., Early, S., P, O. G., Pennington, S., Hill, A. D. & Young, L. S. 2010. Interaction of developmental transcription factor HOXC11 with steroid receptor coactivator SRC-1 mediates resistance to endocrine therapy in breast cancer [corrected]. *Cancer Res,* 70**,** 1585-94.

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & Depristo, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research,* 20**,** 1297-303.

Mclaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. & Cunningham, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26**,** 2069-70.

Medina, J. E., Dichtel, W. & Luna, M. A. 1984. Verrucous-Squamous Carcinomas of the Oral Cavity - a Clinicopathologic Study of 104 Cases. *Archives of Otolaryngology-Head & Neck Surgery,* 110**,** 437-440.

Mehra, M. R., Kobashigawa, J. A., Deng, M. C., Fang, K. C., Klingler, T. M., Lal, P. G., Rosenberg, S., Uber, P. A., Starling, R. C., Murali, S., Pauly, D. F., Dedrick, R., Walker, M. G., Zeevi, A., Eisen, H. J. & Investigators, C. 2007. Transcriptional signals of T-cell and corticosteroid-sensitive genes are associated with future acute cellular rejection in cardiac allografts. *Journal of Heart and Lung Transplantation,* 26**,** 1255-1263.

Mehra, M. R., Kobashigawa, J. A., Deng, M. C., Fang, K. C., Klingler, T. M., Lal, P. G., Rosenberg, S., Uber, P. A., Starling, R. C., Murali, S., Pauly, D. F., Dedrick, R., Walker, M. G., Zeevi, A., Eisen, H. J. & Investigators, C. 2008. Clinical implications and longitudinal alteration of peripheral blood transcriptional signals indicative of future cardiac allograft rejection. *Journal of Heart and Lung Transplantation,* 27**,** 297-301.

Mendez, E., Fan, W., Choi, P., Agoff, S. N., Whipple, M., Farwell, D. G., Futran, N. D., Weymuller, E. A., Zhao, L. P. & Chen, C. 2007. Tumor-specific genetic expression profile of metastatic oral squamous cell carcinoma. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 29**,** 803-814.

Mendez, E., Houck, J. R., Doody, D. R., Fan, W. H., Lohavanichbutr, P., Rue, T. C., Yueh, B., Futran, N. D., Upton, M. P., Farwell, D. G., Heagerty, P. J., Zhao, L. P., Schwartz, S. M. & Chen, C. 2009. A Genetic Expression Profile Associated with Oral Cancer Identifies a Group of Patients at High Risk of Poor Survival. *Clinical Cancer Research,* 15**,** 1353-1361.

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 Biology,* 12.

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

Miller, C. S. & Johnstone, B. M. 2001. Human papillomavirus as a risk factor for oral squamous cell carcinoma: A meta-analysis, 1982-1997. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics,* 91**,** 622-635.

Mithani, S. K., Mydlarz, W. K., Grumbine, F. L., Smith, I. M. & Califano, J. A. 2007. Molecular genetics of premalignant oral lesions. *Oral Diseases,* 13**,** 126-133.

Mitsuishi, T., Ohara, K., Kawashima, M., Kobayashi, S. & Kawana, S. 2005. Prevalence of human papillomavirus DNA sequences in verrucous carcinoma of the lip: genomic and therapeutic approaches. *Cancer Lett,* 222**,** 139-43.

Miyamoto, R., Uzawa, N., Nagaoka, S., Hirata, Y. & Amagasa, T. 2003. Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas. *Oral Oncology,* 39**,** 610-618.

Miyazaki, H., Patel, V., Wang, H. X., Ensley, J. F., Gutkind, J. S. & Yeudall, W. A. 2006. Growth factor-sensitive molecular targets identified in primary and metastatic head and neck squamous cell carcinoma using microarray analysis. *Oral Oncology,* 42**,** 240-256.

Mohapatra, G., Betensky, R. A., Miller, E. R., Carey, B., Gaumont, L. D., Engler, D. A. & Louis, D. N. 2006. Glioma test array for use with formalin-fixed, paraffin-embedded tissue - Array comparative genomic hybridization correlates with loss of heterozygosity and fluorescence in situ hybridization. *Journal of Molecular Diagnostics,* 8**,** 268-276.

Mohtasham, N., Babakoohi, S., Shiva, A., Shadman, A., Kamyab-Hesari, K., Shakeri, M. T. & Sharifi-Sistani, N. 2013. Immunohistochemical study of p53, Ki-67, MMP-2 and MMP-9 expression at invasive front of squamous cell and verrucous carcinoma in oral cavity. *Pathology Research and Practice,* 209**,** 110-114.

Moll, R., Divo, M. & Langbein, L. 2008. The human keratins: biology and pathology. *Histochemistry and Cell Biology,* 129**,** 705-733.

Mork, J., Lie, A. K., Glattre, E., Clark, S., Hallmans, G., Jellum, E., Koskela, P., Moller, B., Pukkala, E., Schiller, J. T., Youngman, L., Lehtinen, M. & Dillner, J. 2001. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *New England Journal of Medicine,* 344**,** 1125-1131.

Morlan, J. D., Qu, K. & Sinicropi, D. V. 2012. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One,* 7**,** e42882.

Morozova, O., Hirst, M. & Marra, M. A. 2009. Applications of New Sequencing Technologies for Transcriptome Analysis. *Annual Review of Genomics and Human Genetics,* 10**,** 135-151.

Morrison, C. D., Liu, P. Y., Woloszynska-Read, A., Zhang, J. M., Luo, W., Qin, M. C., Bshara, W., Conroy, J. M., Sabatini, L., Vedell, P., Xiong, D. H., Liu, S., Wang, J. M., Shen, H., Li, Y. W., Omilian, A. R., Hill, A., Head, K., Guru, K., Kunnev, D., Leach, R., Eng, K. H., Darlak, C., Hoeflich, C., Veeranki, S., Glenn, S., You, M., Pruitt, S. C., Johnson, C. S. & Trump, D. L. 2014. Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** E672-E681.

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621-8.

Mullis, T. C., Tang, X. & Chong, K. T. 2008. Expression of connective tissue growth factor (CTGF/CCN2) in head and neck squamous cell carcinoma. *Journal of Clinical Pathology,* 61**,** 606-610.

Murugan, A. K., Hong, N. T., Fukui, Y., Munirajan, A. K. & Tsuchida, N. 2008. Oncogenic mutations of the PIK3CA gene in head and neck squamous cell carcinomas. *International Journal of Oncology,* 32**,** 101-111.

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**,** 2709-2716.

Nakamura, E., Kozaki, K. I., Tsuda, H., Suzuki, E., Pimkhaokham, A., Yamamoto, G., Irie, T., Tachikawa, T., Amagasa, T., Inazawa, J. & Imoto, I. 2008. Frequent silencing of a putative tumor suppressor gene melatonin receptor 1 A (MTNR1A) in oral squamous-cell carcinoma. *Cancer Science,* 99**,** 1390-1400.

Nakaya, K., Yamagata, H. D., Arita, N., Nakashiro, K., Nose, M., Miki, T. & Hamakawa, H. 2007. Identification of homozygous deletions of tumor suppressor gene FAT in oral cancer using CGH-array. *Oncogene,* 26**,** 5300-5308.

Nasman, A., Attner, P., Hammarstedt, L., Do, J., Eriksson, M., Giraud, G., Ahrlund-Richter, S., Marklund, L., Romanitan, M., Lindquist, D., Ramqvist, T., Lindholm, J., Sparen, P., Ye, W. M., Dahistrand, H., Munck-Wikland, E. & Dalianis, T. 2009. Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: An epidemic of viral-induced carcinoma? *International Journal of Cancer,* 125**,** 362-366.

Natsugoe, S., Xiangming, C., Matsumoto, M., Okumura, H., Nakashima, S., Sakita, H., Ishigami, S., Baba, M., Takao, S. & Aikou, T. 2002. Smad4 and transforming growth factor beta1 expression in patients with squamous cell carcinoma of the esophagus. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 8**,** 1838-42.

Navin, N., Krasnitz, A., Rodgers, L., Cook, K., Meth, J., Kendall, J., Riggs, M., Eberling, Y., Troge, J., Grubor, V., Levy, D., Lundin, P., Maner, S., Zetterberg, A., Hicks, J. & Wigler, M. 2010. Inferring tumor progression from genomic heterogeneity. *Genome Res,* 20**,** 68-80.

Naylor, T. L., Greshock, J., Wang, Y., Colligon, T., Yu, Q. C., Clemmer, V., Zaks, T. Z. & Weber, B. L. 2005. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Research,* 7**,** R1186-R1198.

Nelson, W. J. & Nusse, R. 2004. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science,* 303**,** 1483-1487.

Neville, B. W. & Day, T. A. 2002. Oral cancer and precancerous lesions. *Ca-a Cancer Journal for Clinicians,* 52**,** 195-215.

Ng, P. C. & Henikoff, S. 2002. Accounting for human polymorphisms predicted to affect protein function. *Genome Research,* 12**,** 436-446.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A. & Shendure, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature,* 461**,** 272-6.

Nguyen, D. Q., Webber, C. & Ponting, C. P. 2006. Bias of selection on human copy-number variants. *Plos Genetics,* 2**,** 198-207.

Nilsson, E. M., Brokken, L. J., Narvi, E., Kallio, M. J. & Harkonen, P. L. 2012. Identification of fibroblast growth factor-8b target genes associated with early and late cell cycle events in breast cancer cells. *Mol Cell Endocrinol,* 358**,** 104-15.

Nobletopham, S. E., Fliss, D. M., Hartwick, R. W. J., Mclachlin, C. M., Freeman, J. L., Noyek, A. M. & Andrulis, I. L. 1993. Detection and Typing of Human Papillomavirus in Verrucous Carcinoma of the Oral Cavity Using the Polymerase Chain-Reaction. *Archives of Otolaryngology-Head & Neck Surgery,* 119**,** 1299-1304.

Noel, J. C., Peny, M. O., De Dobbeleer, G., Thiriar, S., Fayt, I., Haot, J. & Heenen, M. 1996. p53 Protein overexpression in verrucous carcinoma of the skin. *Dermatology,* 192**,** 12-5.

Northcott, P. A., Nakahara, Y., Wu, X. C., Feuk, L., Ellison, D. W., Croul, S., Mack, S., Kongkham, P. N., Peacock, J., Dubuc, A., Ra, Y. S., Zilberberg, K., Mcleod, J., Scherer, S. W., Rao, J. S., Eberhart, C. G., Grajkowska, W., Gillespie, Y., Lach, B., Grundy, R., Pollack, I. F., Hamilton, R. L., Van Meter, T., Carlotti, C. G., Boop, F., Bigner, D., Gilbertson, R. J., Rutka, J. T. & Taylor, M. D. 2009. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genetics,* 41**,** 465-472.

Novak, P., Jensen, T., Oshiro, M. M., Watts, G. S., Kim, C. J. & Futscher, B. W. 2008. Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res,* 68**,** 8616-25.

Nozad-Mojaver, Y., Mirzaee, M. & Jafarzadeh, A. 2009. Synergistic effects of cigarette smoke and saliva. *Medicina Oral Patologia Oral Y Cirugia Bucal,* 14**,** E217-E221.

O-Charoenrat, P., Rhys-Evans, P. H., Modjtahedi, H. & Eccles, S. A. 2002. The role of c-erbB receptors and ligands in head and neck squamous cell carcinoma. *Oral Oncology,* 38**,** 627-640.

Odar, K., Bostjancic, E., Gale, N., Glavac, D. & Zidar, N. 2012a. Differential expression of microRNAs miR-21, miR-31, miR-203, miR-125a-5p and miR-125b and proteins PTEN and p63 in verrucous carcinoma of the head and neck. *Histopathology,* 61**,** 257-265.

Odar, K., Zidar, N., Bonin, S., Gale, N., Cardesa, A. & Stanta, G. 2012b. Desmosomes in verrucous carcinoma of the head and neck. *Histology and Histopathology,* 27**,** 467-474.

Ogawa, A., Fukuta, Y., Nakajima, T., Kanno, S. M., Obara, A., Nakamura, K., Mizuki, H., Takeda, Y. & Satoh, M. 2004. Treatment results of oral verrucous carcinoma and its biological behavior. *Oral Oncol,* 40**,** 793-7.

Ogbureke, K. U. E., Abdelsayed, R. A., Kushner, H., Li, L. & Fisher, L. W. 2010. Two Members of the SIBLING Family of Proteins, DSPP and BSP, May Predict the Transition of Oral Epithelial Dysplasia to Oral Squamous Cell Carcinoma. *Cancer,* 116**,** 1709-1717.

Ogbureke, K. U. E., Nikitakis, N. G., Warburton, G., Ord, R. A., Sauk, J. J., Waller, J. L. & Fisher, L. W. 2007. Up-regulation of SIBLING proteins and correlation with cognate MMP expression in oral cancer. *Oral Oncology,* 43**,** 920-932.

Ohkura, S., Kondoh, N., Hada, A., Arai, M., Yamazaki, Y., Sindoh, M., Takahashi, M., Matsumoto, I. & Yamamoto, M. 2005. Differential expression of the keratin-4, -13, -14, -17 and transglutaminase 3 genes during the development of oral squamous cell carcinoma from leukoplakia. *Oral Oncology,* 41**,** 607-613.

Okami, K., Wu, L., Riggins, G., Cairns, P., Goggins, M., Evron, E., Halachmi, N., Ahrendt, S. A., Reed, A. L., Hilgers, W., Kern, S. E., Koch, W. M., Sidransky, D. & Jen, J. 1998. Analysis of PTEN/MMAC1 alterations in aerodigestive tract tumors. *Cancer Research,* 58**,** 509-511.

Oliveira, D. T., De Moraes, R. V., Fiamengui Filho, J. F., Fanton Neto, J., Landman, G. & Kowalski, L. P. 2006. Oral verrucous carcinoma: a retrospective study in Sao Paulo Region, Brazil. *Clinical oral investigations,* 10**,** 205-9.

Oliveira, M. C., Silveira, E. J. D., Godoy, G. P., Amorim, R. F. B., Costa, A. L. L. & Queiroz, L. M. G. 2005. Immunohistochemical evaluation of intermediate filament proteins in squamous papilloma and oral verrucous carcinoma. *Oral Diseases,* 11**,** 288-292.

Onda, M., Akaishi, J., Asaka, S., Okamoto, J., Miyamoto, S., Mizutani, K., Yoshida, A., Ito, K. & Emi, M. 2005. Decreased expression of haemoglobin beta (HBB) gene in anaplastic thyroid cancer and recovory of its expression inhibits cell growth. *British Journal of Cancer,* 92**,** 2216-2224.

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S. & Weinberg, R. A. 2008. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Research,* 68**,** 3645-3654.

Orvidas, L. J., Olsen, K. D., Lewis, J. E. & Suman, V. J. 1998. Verrucous carcinoma of the larynx: A review of 53 patients. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 20**,** 197-203.

Pai, S. I. & Westra, W. H. 2009. Molecular Pathology of Head and Neck Cancer: Implications for Diagnosis, Prognosis, and Treatment. *Annual Review of Pathology-Mechanisms of Disease,* 4**,** 49-70.

Patel, K. R., Chernock, R. D., Zhang, T. R., Wang, X. W., El-Mofty, S. K. & Lewis, J. S. 2013. Verrucous carcinomas of the head and neck, including those with associated squamous cell carcinoma, lack transcriptionally active high-risk human papillomavirus. *Human Pathology,* 44**,** 2385-2392.

Patel, V., Aldridge, K., Ensley, J. F., Odell, E., Boyd, A., Jones, J., Gutkind, J. S. & Yeudall, W. A. 2002. Laminin-gamma 2 overexpression in head-and-neck squamous cell carcinoma. *International Journal of Cancer,* 99**,** 583-588.

Pawar, S. A., Sarkar, T. R., Balamurugan, K., Sharan, S., Wang, J., Zhang, Y. H., Dowdy, S. F., Huang, A. M. & Sterneck, E. 2010. C/EBP delta targets cyclin D1 for proteasome-mediated degradation via induction of CDC27/APC3 expression. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 9210-9215.

Peano, C., Pietrelli, A., Consolandi, C., Rossi, E., Petiti, L., Tagliabue, L., De Bellis, G. & Landini, P. 2013. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microbial informatics and experimentation,* 3**,** 1.

Pedersen, M. O., Larsen, A., Stoltenberg, M. & Penkowa, M. 2009. The role of metallothionein in oncogenesis and cancer prognosis. *Prog Histochem Cytochem,* 44**,** 29-64.

Pentenero, M., Donadini, A., Di Nallo, E., Maffei, M., Marino, R., Familiari, U., Broccoletti, R., Castagnola, P., Gandolfo, S. & Giaretti, W. 2011. Distinctive Chromosomal Instability Patterns in Oral Verrucous and Squamous Cell Carcinomas Detected by High-Resolution DNA Flow Cytometry. *Cancer,* 117**,** 5052-5057.

Pentenero, M., Meleti, M., Vescovi, P. & Gandolfo, S. 2014. Oral proliferative verrucous leucoplakia: are there particular features for such an ambiguous entity? A systematic review. *British Journal of Dermatology,* 170**,** 1039-1047.

Pereira, M. C., Oliveira, D. T., Landman, G. & Kowalski, L. P. 2007. Histologic subtypes of oral squamous cell carcinoma: Prognostic relevance. *Journal of the Canadian Dental Association,* 73**,** 339-344.

Perez-Ordonez, B., Beauchemin, M. & Jordan, R. C. K. 2006. Molecular biology of squamous cell carcinoma of the head and neck. *Journal of Clinical Pathology,* 59**,** 445-453.

Petti, S. 2003. Pooled estimate of world leukoplakia prevalence: a systematic review. *Oral Oncology,* 39**,** 770-780.

Pickering, C. R., Zhang, J., Yoo, S. Y., Bengtsson, L., Moorthy, S., Neskey, D. M., Zhao, M., Ortega Alves, M. V., Chang, K., Drummond, J., Cortez, E., Xie, T. X., Zhang, D., Chung, W., Issa, J. P., Zweidler-Mckay, P. A., Wu, X., El-Naggar, A. K., Weinstein, J. N., Wang, J., Muzny, D. M., Gibbs, R. A., Wheeler, D. A., Myers, J. N. & Frederick, M. J. 2013. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer Discov,* 3**,** 770-81.

Pindborg Jj, W. P. 1997. *Histological typing of cancer and precancer of the oral mucosa,* Berlin; New York, springer.

Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W. & Albertson, D. G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics,* 20**,** 207-211.

Pleasance, E. D., Cheetham, R. K., Stephens, P. J., Mcbride, D. J., Humphray, S. J., Greenman, C. D., Varela, I., Lin, M. L., Ordonez, G. R., Bignell, G. R., Ye, K., Alipaz, J., Bauer, M. J., Beare, D., Butler, A., Carter, R. J., Chen, L. N., Cox, A. J., Edkins, S., Kokko-Gonzales, P. I., Gormley, N. A., Grocock, R. J., Haudenschild, C. D., Hims, M. M., James, T., Jia, M. M., Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L. J., Ning, Z. M., Royce, T., Schulz-Trieglaff, O. B., Spiridou, A., Stebbings, L. A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M. T., Campbell, P. J., Bentley, D. R., Futreal, P. A. & Stratton, M. R. 2010a. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature,* 463**,** 191-U73.

Pleasance, E. D., Stephens, P. J., O'meara, S., Mcbride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C., Varela, I., Nik-Zainal, S., Davies, H. R., Ordonez, G. R., Mudie, L. J., Latimer, C., Edkins, S., Stebbings, L., Chen, L. N., Jia, M. M., Leroy, C., Marshall, J., Menzies, A., Butler, A., Teague, J. W., Mangion, J., Sun, Y. A., Mclaughlin, S. F., Peckham, H. E., Tsung, E. F., Costa, G. L., Lee, C. C., Minna, J. D., Gazdar, A., Birney, E., Rhodes, M. D., Mckernan, K. J., Stratton, M. R., Futreal, P. A. & Campbell, P. J.

2010b. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature,* 463**,** 184-U66.

Poage, G. M., Christensen, B. C., Houseman, E. A., Mcclean, M. D., Wiencke, J. K., Posner, M. R., Clark, J. R., Nelson, H. H., Marsit, C. J. & Kelsey, K. T. 2010. Genetic and epigenetic somatic alterations in head and neck squamous cell carcinomas are globally coordinated but not locally targeted. *PLoS One,* 5**,** e9651.

Poh, C. F., Zhang, L. W., Lam, W. L., Zhang, X. L., An, D., Chau, C., Priddy, R., Epstein, J. & Rosin, M. P. 2001. A high frequency of allelic loss in oral verrucous lesions may explain malignant risk. *Laboratory Investigation,* 81**,** 629-634.

Poveda-Roda, R., Bagan, J. V., Jimenez-Soriano, Y., Margaix-Munoz, M. & Sarrion-Perez, G. 2010. Changes in smoking habit among patients with a history of oral squamous cell carcinoma (OSCC). *Medicina Oral Patologia Oral Y Cirugia Bucal,* 15**,** E721-E726.

Presland, R. B. & Dale, B. A. 2000. Epithelial structural proteins of the skin and oral cavity: Function in health and disease. *Critical Reviews in Oral Biology & Medicine,* 11**,** 383-408.

Puente, X. S., Pinyol, M., Quesada, V., Conde, L., Ordonez, G. R., Villamor, N., Escaramis, G., Jares, P., Bea, S., Gonzalez-Diaz, M., Bassaganyas, L., Baumann, T., Juan, M., Lopez-Guerra, M., Colomer, D., Tubio, J. M., Lopez, C., Navarro, A., Tornador, C., Aymerich, M., Rozman, M., Hernandez, J. M., Puente, D. A., Freije, J. M., Velasco, G., Gutierrez-Fernandez, A., Costa, D., Carrio, A., Guijarro, S., Enjuanes, A., Hernandez, L., Yague, J., Nicolas, P., Romeo-Casabona, C. M., Himmelbauer, H., Castillo, E., Dohm, J. C., De Sanjose, S., Piris, M. A., De Alava, E., San Miguel, J., Royo, R., Gelpi, J. L., Torrents, D., Orozco, M., Pisano, D. G., Valencia, A., Guigo, R., Bayes, M., Heath, S., Gut, M., Klatt, P., Marshall, J., Raine, K., Stebbings, L. A., Futreal, P. A., Stratton, M. R., Campbell, P. J., Gut, I., Lopez-Guillermo, A., Estivill, X., Montserrat, E., Lopez-Otin, C. & Campo, E. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature,* 475**,** 101-5.

Pyeon, D., Newton, N. A., Lambert, P. F., Den Boon, J. A., Sengupta, S., Marsit, C. J., Woodworth, C. D., Connor, J. P., Haugen, T. H., Smith, E. M., Kelsey, K. T., Turek, L. P. & Ahlquist, P. 2007. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Research,* 67**,** 4605-4619.

Qin, G. Z., Park, J. Y., Chen, S. Y. & Lazarus, P. 1999. A high prevalence of p53 mutations in pre-malignant oral erythroplakia. *Int J Cancer,* 80**,** 345-8.

Quan, H., Tang, Z., Zhao, L., Wang, Y., Liu, O., Yao, Z. & Zuo, J. 2012. Expression of alphaB-crystallin and its potential anti-apoptotic role in oral verrucous carcinoma. *Oncol Lett,* 3**,** 330-334.

Radhakrishnan, R. 2011. Inherited proliferative oral disorder: a reductionist approach to proliferative verrucous leukoplakia. *Indian J Dent Res,* 22**,** 365-6.

Rautava, J., Luukkaa, M., Heikinheimo, K., Alin, J., Grenman, R. & Happonen, R. P. 2007. Squamous cell carcinomas arising from different types of oral epithelia differ in their tumor and patient characteristics and survival. *Oral oncology,* 43**,** 911-919.

Ray, J. G., Mukherjee, S., Pattanayak Mohanty, S. & Chaudhuri, K. 2011a. Oral verrucous carcinoma--a misnomer? Immunohistochemistry based comparative study of two cases. *BMJ case reports,* 2011.

Ray, J. G., Mukherjee, S., Pattanayak Mohanty, S. & Chaudhuri, K. 2011b. Oral verrucous carcinoma--a misnomer? Immunohistochemistry based comparative study of two cases. *BMJ Case Rep,* 2011.

Redon, R., Muller, D., Caulee, K., Wanherdrick, K., Abecassis, J. & Du Manoir, S. 2001. A simple specific pattern of chromosomal aberrations at early stages of head and neck squamous cell carcinomas: PIK3CA but not p63 gene as a likely target of 3q26-qter gains. *Cancer Research,* 61**,** 4122-4129.

Reed, A. L., Califano, J., Cairns, P., Westra, W. H., Jones, R. M., Koch, W., Ahrendt, S., Eby, Y., Sewell, D., Nawroz, H., Bartek, J. & Sidransky, D. 1996a. High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma. *Cancer Res,* 56**,** 3630-3.

Reed, A. L., Califano, J., Cairns, P., Westra, W. H., Jones, R. M., Koch, W., Ahrendt, S., Eby, Y., Sewell, D., Nawroz, H., Bartek, J. & Sidransky, D. 1996b. High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma. *Cancer Research,* 56**,** 3630-3633.

Regezi Ja, S. J. 1993. *Oral pathology: clinical-pathologic correlations,* Philadelphia, Saunders.

Reibel, J. 2003. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. *Crit Rev Oral Biol Med,* 14**,** 47-62.

Reichart, P. A. & Philipsen, H. P. 2005. Oral erythroplakia - a review. *Oral Oncology,* 41**,** 551-561.

Rekha, K. P. & Angadi, P. V. 2010. Verrucous carcinoma of the oral cavity: a clinico-pathologic appraisal of 133 cases in Indians. *Oral Maxillofac Surg,* 14**,** 211-8.

Reshmi, S. C., Huang, X., Schoppy, D. W., Black, R. C., Saunders, W. S., Smith, D. I. & Gollin, S. M. 2007. Relationship between FRA11F and 11q13 gene amplification in oral cancer. *Genes Chromosomes & Cancer,* 46**,** 143-154.

Rheinwald, J. G., Hahn, W. C., Ramsey, M. R., Wu, J. Y., Guo, Z., Tsao, H., De Luca, M., Catricala, C. & O'toole, K. M. 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.

Ribarska, T., Bastian, K. M., Koch, A. & Schulz, W. A. 2012. Specific changes in the expression of imprinted genes in prostate cancer--implications for cancer progression and epigenetic regulation. *Asian J Androl,* 14**,** 436-50.

Rizzo, G., Black, M., Mymryk, J., Barrett, J. & Nichols, A. 2014. Defining the genomic landscape of head and neck cancers through next-generation sequencing. *Oral Dis*.

Rocco, J. W., Leong, C. O., Kuperwasser, N., Deyoung, M. P. & Ellisen, L. W. 2006. p63 mediates survival in squamous cell carcinoma by suppression of p73-dependent apoptosis. *Cancer Cell,* 9**,** 45-56.

Rodrigo, J. P., Lazo, P. S., Ramos, S., Alvarez, I. & Suarez, C. 1996. MYC amplification in squamous cell carcinomas of the head and neck. *Archives of Otolaryngology-Head & Neck Surgery,* 122**,** 504-507.

Roepman, P., Wessels, L. F., Kettelarij, N., Kemmeren, P., Miles, A. J., Lijnzaad, P., Tilanus, M. G., Koole, R., Hordijk, G. J., Van Der Vliet, P. C., Reinders, M. J., Slootweg, P. J. & Holstege, F. C. 2005. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet,* 37**,** 182-6.

Rosin, M. P., Cheng, X., Poh, C., Lam, W. L., Huang, Y. Q., Lovas, J., Berean, K., Epstein, J. B., Priddy, R., Le, N. D. & Zhang, L. W. 2000. Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. *Clinical Cancer Research,* 6**,** 357-362.

Rosin, M. P., Lam, W. L., Poh, C., Le, N. D., Li, R. J., Zeng, T., Priddy, R. & Zhang, L. W. 2002. 3p14 and 9p21 loss is a simple tool for predicting second oral malignancy at previously treated oral cancer sites. *Cancer Research,* 62**,** 6447-6450.

Ross, J. S., Wang, K., Elkadi, O. R., Tarasen, A., Foulke, L., Sheehan, C. E., Otto, G. A., Palmer, G., Yelensky, R., Lipson, D., Chmielecki, J., Ali, S. M., Elvin, J., Morosini, D., Miller, V. A. & Stephens, P. J. 2014. Next-generation sequencing reveals frequent consistent genomic alterations in small cell undifferentiated lung cancer. *J Clin Pathol*.

Rothenberg, S. M. & Ellisen, L. W. 2012. The molecular pathogenesis of head and neck squamous cell carcinoma. *Journal of Clinical Investigation,* 122**,** 1951-1957.

Ruiz, M. I. G., Floor, K., Rijmen, F., Grunberg, K., Rodriguez, J. A. & Giaccone, G. 2007. EGFR and K-ras mutation analysis in non-small cell lung cancer: Comparison of paraffin embedded versus frozen specimens. *Cellular Oncology,* 29**,** 257-264.

Saito, T., Nakajima, T. & Mogi, K. 1999a. Immunohistochemical analysis of cell cycle-associated proteins p16, pRb, p53, p27 and Ki-67 in oral cancer and precancer with special reference to verrucous carcinomas. *J Oral Pathol Med,* 28**,** 226-32.

Saito, T., Nakajima, T. & Mogi, K. 1999b. Immunohistochemical analysis of cell cycle-associated proteins p16, pRb, p53, p27 and Ki-67 in oral cancer and precancer with special reference to verrucous carcinomas. *Journal of Oral Pathology & Medicine,* 28**,** 226-232.

Sajnani, M. R., Patel, A. K., Bhatt, V. D., Tripathi, A. K., Ahir, V. B., Shankar, V., Shah, S., Shah, T. M., Koringa, P. G., Jakhesara, S. J. & Joshi, C. G. 2012. Identification of novel transcripts deregulated in buccal cancer by RNA-seq. *Gene,* 507**,** 152-158.

Sakurai, K., Urade, M., Takahashi, Y., Kishimoto, H., Noguchi, K., Yasoshima, H. & Kubota, A. 2000. Increased expression of c-erbB-3 protein and proliferating cell nuclear antigen during development of verrucous carcinoma of the oral mucosa. *Cancer,* 89**,** 2597-2605.

Salahshourifar, I., Vincent-Chong, V. K., Kallarakkal, T. G. & Zain, R. B. 2014. Genomic DNA copy number alterations from precursor oral lesions to oral squamous cell carcinoma. *Oral Oncology,* 50**,** 404-412.

Saleh, A., Zain, R. B., Hussaini, H., Ng, F., Tanavde, V., Hamid, S., Chow, A. T., Lim, G. S., Abraham, M. T., Teo, S. H. & Cheong, S. C. 2010. Transcriptional profiling of oral squamous cell carcinoma using formalin-fixed paraffin-embedded samples. *Oral Oncology,* 46**,** 379-386.

Samman, M., Wood, H., Conway, C., Berri, S., Pentenero, M., Gandolfo, S., Cassenti, A., Cassoni, P., Al Ajlan, A., Barrett, A. W., Chengot, P., Maclennan, K., High, A. S. & Rabbitts, P. 2014. Next-generation sequencing analysis for detecting human papillomavirus in oral verrucous carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol,* 118**,** 117-125 e1.

Samman, M. A., Bowen, I. D., Taiba, K., Antonius, J. & Hannan, M. A. 1998. Mint prevents shamma-induced carcinogenesis in hamster cheek pouch. *Carcinogenesis,* 19**,** 1795-801.

Santoro, A., Pannone, G., Contaldo, M., Sanguedolce, F., Esposito, V., Serpico, R., Lo Muzio, L., Papagerakis, S. & Bufo, P. 2011. A Troubling Diagnosis of Verrucous Squamous Cell Carcinoma ("the Bad Kind" of Keratosis) and the Need of Clinical and Pathological Correlations: A Review of the Literature with a Case Report. *Journal of skin cancer,* 2011**,** 370605.

Sarkaria, I., O-Charoenrat, P., Talbot, S. G., Reddy, P. G., Ngai, I., Maghami, E., Patel, K. N., Lee, B., Yonekawa, Y., Dudas, M., Kaufman, A., Ryan, R., Ghossein, R., Rao, P. H., Stoffel, A., Ramanathan, Y. & Singh, B. 2006. Squamous cell carcinoma related oncogene/DCUN1D1 is highly conserved and activated by amplification in squamous cell carcinomas. *Cancer Research,* 66**,** 9437-9444.

Sawair, F. A., Al-Mutwakel, A., Al-Eryani, K., Al-Surhy, A., Maruyama, S., Cheng, J., Al-Sharabi, A. & Saku, T. 2007. High relative frequency of oral squamous cell

carcinoma in Yemen: Qat and tobacco chewing as its aetiological background. *International Journal of Environmental Health Research,* 17**,** 185-195.

Scheifele, C., Nassar, A. & Reichart, P. A. 2007. Prevalence of oral cancer and potentially malignant lesions among shammah users in Yemen. *Oral Oncology,* 43**,** 42-50.

Schepman, K. P., Van Der Meij, E. H., Smeele, L. E. & Van Der Waal, I. 1998. Malignant transformation of oral leukoplakia: a follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. *Oral Oncology,* 34**,** 270-275.

Schmidt, B. L., Dierks, E. J., Homer, L. & Potter, B. 2004. Tobacco smoking history and presentation of oral squamous cell carcinoma. *Journal of Oral and Maxillofacial Surgery,* 62**,** 1055-1058.

Schuuring, E., Verhoeven, E., Mooi, W. J. & Michalides, R. J. 1992. Identification and cloning of two overexpressed genes, U21B31/PRAD1 and EMS1, within the amplified chromosome 11q13 region in human carcinomas. *Oncogene,* 7**,** 355-61.

Schwartz, S. M., Daling, J. R., Doody, D. R., Wipf, G. C., Carter, J. J., Madeleine, M. M., Mao, E. J., Fitzgibbons, E. D., Huang, S. X., Beckmann, A. M., Mcdougall, J. K. & Galloway, D. A. 1998. Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection. *J Natl Cancer Inst,* 90**,** 1626-1636.

Schweiger, M. R., Kerick, M., Timmermann, B., Albrecht, M. W., Borodina, T., Parkhomchuk, D., Zatloukal, K. & Lehrach, H. 2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. *PLoS One,* 4**,** e5548.

Schweizer, J., Bowden, P. E., Coulombe, P. A., Langbein, L., Lane, E. B., Magin, T. M., Maltais, L., Omary, M. B., Parry, D. a. D., Rogers, M. A. & Wright, M. W. 2006. New consensus nomenclature for mammalian keratins. *Journal of Cell Biology,* 174**,** 169-174.

Scully, C. & Bagan, J. 2009a. Oral squamous cell carcinoma overview. *Oral Oncol,* 45**,** 301-8.

Scully, C. & Bagan, J. V. 2009b. Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. *Oral Diseases,* 15**,** 388-399.

Seiwert, T. Y., Jagadeeswaran, R., Faoro, L., Janamanchi, V., Nallasura, V., El Dinali, M., Yala, S., Kanteti, R., Cohen, E. E. W., Lingen, M. W., Martin, L., Krishnaswamy, S., Klein-Szanto, A., Christensen, J. G., Vokes, E. E. & Salgia, R. 2009. The MET Receptor Tyrosine Kinase Is a Potential Novel Therapeutic Target for Head and Neck Squamous Cell Carcinoma. *Cancer Research,* 69**,** 3021-3031.

Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R. A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G. A., Teschendorff, A. E., Tse, K., Turashvili, G., Varhol, R., Warren, R. L., Watson, P., Zhao, Y. J., Caldas, C., Huntsman, D., Hirst, M., Marra, M. A. & Aparicio, S. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature,* 461**,** 809-U67.

Sharp, A. J., Locke, D. P., Mcgrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. A., Schwartz, S., Segraves, R., Oseroff, V. V., Albertson, D. G., Pinkel, D. & Eichler, E. E. 2005. Segmental duplications and copy-number variation in the human genome. *American Journal of Human Genetics,* 77**,** 78-88.

Shaw, R. J., Hall, G. L., Lowe, D., Liloglou, T., Field, J. K., Sloan, P. & Risk, J. M. 2008. The role of pyrosequencing in head and neck cancer epigenetics: correlation of

quantitative methylation data with gene expression. *Arch Otolaryngol Head Neck Surg,* 134**,** 251-6.

Shear, M. & Pindborg, J. J. 1980a. Verrucous hyperplasia of the oral mucosa. *Cancer,* 46**,** 1855-62.

Shear, M. & Pindborg, J. J. 1980b. Verrucous Hyperplasia of the Oral-Mucosa. *Cancer,* 46**,** 1855-1862.

Sheen, M. C., Sheu, H. M., Lai, F. J., Lin, S. D., Wu, C. F., Wang, Y. W. & Lan, C. C. E. 2004. A huge verrucous carcinoma of the lower lip treated with intra-arterial infusion of methotrexate. *British Journal of Dermatology,* 151**,** 727-729.

Sheu, J. J. C., Hua, C. H., Wan, L., Lin, Y. J., Lai, A. T., Tseng, H. C., Jinawath, N., Tsai, M. H., Chang, N. W., Lin, C. F., Lin, C. C., Hsieh, L. J., Wang, T. L., Shih, I. M. & Tsai, F. J. 2009. Functional Genomic Analysis Identified Epidermal Growth Factor Receptor Activation as the Most Common Genetic Event in Oral Squamous Cell Carcinoma. *Cancer Research,* 69**,** 2568-2576.

Shintani, S., Nakahara, Y., Mihara, M., Ueyama, Y. & Matsumura, T. 2001. Inactivation of the p14(ARF), p15(INK4B) and p16(INK4A) genes is a frequent event in human oral squamous cell carcinomas. *Oral Oncology,* 37**,** 498-504.

Shroyer, K. R., Greer, R. O., Fankhouser, C. A., Mcguirt, W. F. & Marshall, R. 1993. Detection of Human Papillomavirus DNA in Oral Verrucous Carcinoma by Polymerase Chain-Reaction. *Modern Pathology,* 6**,** 669-672.

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.

Silva, S. D., Cunha, I. W., Younes, R. N., Soares, F. A., Kowalski, L. P. & Graner, E. 2010. ErbB receptors and fatty acid synthase expression in aggressive head and neck squamous cell carcinomas. *Oral Dis,* 16**,** 774-80.

Silverman, S. 1988. Early Diagnosis of Oral-Cancer. *Cancer,* 62**,** 1796-1799.

Silverman, S., Jr. 1999. Oral cancer: complications of therapy. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod,* 88**,** 122-6.

Sinicropi, D., Qu, K., Collin, F., Crager, M., Liu, M. L., Pelham, R. J., Pho, M., Dei Rossi, A., Jeong, J., Scott, A., Ambannavar, R., Zheng, C., Mena, R., Esteban, J., Stephans, J., Morlan, J. & Baker, J. 2012. Whole Transcriptome RNA-Seq Analysis of Breast Cancer Recurrence Risk Using Formalin-Fixed Paraffin-Embedded Tumor Tissue. *Plos One,* 7.

Sirbu, A., Kerr, G., Crane, M. & Ruskin, H. J. 2012. RNA-Seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. *PLoS One,* 7**,** e50986.

Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science,* 314**,** 268-74.

Skotheim, R. I. & Nees, M. 2007. Alternative splicing in cancer: Noise, functional, or systematic? *International Journal of Biochemistry & Cell Biology,* 39**,** 1432-1449.

Skubitz, K. M., Francis, P., Skubitz, A. P. N., Luo, X. H. & Nilbert, M. 2012. Gene expression identifies heterogeneity of metastatic propensity in high-grade soft tissue sarcomas. *Cancer,* 118**,** 4235-4243.

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A. & Mcguire, W. L. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science,* 235**,** 177-82.

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.

Slootweg, P. J. & Muller, H. 1983. Verrucous hyperplasia or verrucous carcinoma. An analysis of 27 patients. *J Maxillofac Surg,* 11**,** 13-9.

Smeets, S. J., Braakhuis, B. J. M., Abbas, S., Snijders, P. J. F., Ylstra, B., Van De Wiel, M. A., Meijer, G. A., Leemans, C. R. & Brakenhoff, R. H. 2006. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. *Oncogene,* 25**,** 2558-2564.

Smeets, S. J., Brakenhoff, R. H., Ylstra, B., Van Wieringen, W. N., Van De Wiel, M. A., Leemans, C. R. & Braakhuis, B. J. 2009. Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis. *Cell Oncol,* 31**,** 291-300.

Smeets, S. J., Hesselink, A. T., Speel, E. J., Haesevoets, A., Snijders, P. J., Pawlita, M., Meijer, C. J., Braakhuis, B. J., Leemans, C. R. & Brakenhoff, R. H. 2007. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. *International journal of cancer. Journal international du cancer,* 121**,** 2465-72.

Smeets, S. J., Van Der Plas, M., Schaaij-Visser, T. B. M., Van Veen, E. a. M., Van Meerloo, J., Braakhuis, B. J. M., Steenbergen, R. D. M. & Brakenhoff, R. H. 2011. Immortalization of oral keratinocytes by functional inactivation of the p53 and pRb pathways. *International Journal of Cancer,* 128**,** 1596-1605.

Snijders, A. M., Schmidt, B. L., Fridlyand, J., Dekker, N., Pinkel, D., Jordan, R. C. K. & Albertson, D. G. 2005. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. *Oncogene,* 24**,** 4232-4242.

Somers, K. D., Merrick, M. A., Lopez, M. E., Incognito, L. S., Schechter, G. L. & Casey, G. 1992. Frequent P53 Mutations in Head and Neck-Cancer. *Cancer Research,* 52**,** 5997-6000.

Song, X., Xia, R., Li, J., Long, Z., Ren, H., Chen, W. & Mao, L. 2014. Common and complex Notch1 mutations in Chinese oral squamous cell carcinoma. *Clin Cancer Res,* 20**,** 701-10.

Spiro, R. H. 1998. Verrucous carcinoma, then and now. *American Journal of Surgery,* 176**,** 393-397.

Srinivasan, M., Sedmak, D. & Jewell, S. 2002. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol,* 161**,** 1961-71.

Stankiewicz, E., Kudahetti, S. C., Prowse, D. M., Ktori, E., Cuzick, J., Ambroisine, L., Zhang, X., Watkin, N., Corbishley, C. & Berney, D. M. 2009. HPV infection and immunochemical detection of cell-cycle markers in verrucous carcinoma of the penis. *Mod Pathol,* 22**,** 1160-8.

Stokes, A., Guerra, E., Bible, J., Halligan, E., Orchard, G., Odell, E. & Thavaraj, S. 2012. Human papillomavirus detection in dysplastic and malignant oral verrucous lesions. *Journal of Clinical Pathology,* 65**,** 283-286.

Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., Kryukov, G. V., Lawrence, M. S., Sougnez, C., Mckenna, A., Shefler, E., Ramos, A. H., Stojanov, P., Carter, S. L., Voet, D., Cortes, M. L., Auclair, D., Berger, M. F., Saksena, G., Guiducci, C., Onofrio, R. C., Parkin, M., Romkes, M., Weissfeld, J. L., Seethala, R. R., Wang, L., Rangel-Escareno, C., Fernandez-Lopez, J. C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Winckler, W., Ardlie, K., Gabriel, S. B., Meyerson, M., Lander, E. S., Getz, G., Golub, T. R., Garraway, L. A. & Grandis, J. R. 2011. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. *Science,* 333**,** 1157-1160.

Strojan, P., Smid, L., Cizmarevic, B., Zagar, T. & Auersperg, M. 2006. Verrucous carcinoma of the larynx: Determining the best treatment option. *Ejso,* 32**,** 984-988.

Su, Z., Li, Z., Chen, T., Li, Q. Z., Fang, H., Ding, D., Ge, W., Ning, B., Hong, H., Perkins, R. G., Tong, W. & Shi, L. 2011. Comparing next-generation sequencing and microarray technologies in a toxicological study of the effects of aristolochic acid on rat kidneys. *Chem Res Toxicol,* 24**,** 1486-93.

Sun, P. C., Uppaluri, R., Schmidt, A. P., Pashia, M. E., Quant, E. C., Sunwoo, J. B., Gollin, S. M. & Scholnick, S. B. 2001. Transcript map of the 8p23 putative tumor suppressor region. *Genomics,* 75**,** 17-25.

Swartz, M. J., Batra, S. K., Varshney, G. C., Hollingsworth, M. A., Yeo, C. J., Cameron, J. L., Wilentz, R. E., Hruban, R. H. & Argani, P. 2002. MUC4 expression increases progressively in pancreatic intraepithelial neoplasia. *Am J Clin Pathol,* 117**,** 791-6.

Swerdlow, A. J., Marmot, M. G., Grulich, A. E. & Head, J. 1995. Cancer Mortality in Indian and British Ethnic Immigrants from the Indian Subcontinent to England and Wales. *British Journal of Cancer,* 72**,** 1312-1319.

Szelachowska, J., Dziegiel, P., Jelen-Krzeszewska, J., Jelen, M., Tarkowski, R., Spytkowska, B., Matkowski, R. & Kornafel, J. 2009. Correlation of metallothionein expression with clinical progression of cancer in the oral cavity. *Anticancer Res,* 29**,** 589-95.

Tabor, M. P., Brakenhoff, R. H., Ruijter-Schippers, H. J., Kummer, J. A., Leemans, C. R. & Braakhuis, B. J. 2004. Genetically altered fields as origin of locally recurrent head and neck cancer: a retrospective study. *Clin Cancer Res,* 10**,** 3607-13.

Tabor, M. P., Brakenhoff, R. H., Ruijter-Schippers, H. J., Van Der Wal, J. E., Snow, G. B., Leemans, C. R. & Braakhuis, B. J. M. 2002. Multiple head and neck tumors frequently originate from a single preneoplastic lesion. *American Journal of Pathology,* 161**,** 1051-1060.

Tabor, M. P., Brakenhoff, R. H., Van Houten, V. M., Kummer, J. A., Snel, M. H., Snijders, P. J., Snow, G. B., Leemans, C. R. & Braakhuis, B. J. 2001. Persistence of genetically altered fields in head and neck cancer patients: biological and clinical implications. *Clin Cancer Res,* 7**,** 1523-32.

Takebayashi, S., Ogawa, T., Jung, K. Y., Muallem, A., Mineta, H., Fisher, S. G., Grenman, R. & Carey, T. E. 2000. Identification of new minimally lost regions on 18q in head and neck squamous cell carcinoma. *Cancer Research,* 60**,** 3397-3403.

Takes, R. P., Rinaldo, A., Rodrigo, J. P., Devaney, K. O., Fagan, J. J. & Ferlito, A. 2008. Can biomarkers play a role in the decision about treatment of the clinically negative neck in patients with head and neck cancer? *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 30**,** 525-538.

Takeshima, M., Saitoh, M., Kusano, K., Nagayasu, H., Kurashige, Y., Malsantha, M., Arakawa, T., Takuma, T., Chiba, I., Kaku, T., Shibata, T. & Abiko, Y. 2008. High frequency of hypermethylation of p14, p15 and p16 in oral pre-cancerous lesions associated with betel-quid chewing in Sri Lanka. *Journal of Oral Pathology & Medicine,* 37**,** 475-479.

Tamura, T., Shomori, K., Haruki, T., Nosaka, K., Hamamoto, Y., Shiomi, T., Ryoke, K. & Ito, H. 2010. Minichromosome maintenance-7 and geminin are reliable prognostic markers in patients with oral squamous cell carcinoma: immunohistochemical study. *Journal of Oral Pathology & Medicine,* 39**,** 328-334.

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., Reva, B., Antipin, Y., Mitsiades, N., Landers, T., Dolgalev, I., Major, J. E., Wilson, M., Socci, N. D., Lash, A. E., Heguy, A.,

Eastham, J. A., Scher, H. I., Reuter, V. E., Scardino, P. T., Sander, C., Sawyers, C. L. & Gerald, W. L. 2010. Integrative genomic profiling of human prostate cancer. *Cancer Cell,* 18**,** 11-22.

Teh, M. T., Parkinson, E. K., Thurlow, J. K., Liu, F., Fortune, F. & Wan, H. 2011. A molecular study of desmosomes identifies a desmoglein isoform switch in head and neck squamous cell carcinoma. *J Oral Pathol Med,* 40**,** 67-76.

Thomas, G., Hashibe, M., Jacob, B. J., Ramadas, K., Mathew, B., Sankaranarayanan, R. & Zhang, Z. F. 2003. Risk factors for multiple oral premalignant lesions. *Int J Cancer,* 107**,** 285-91.

Thomas, G. J. & Barrett, A. W. 2009. Papillary and verrucous lesions of the oral mucosa. *Diagnostic Histopathology,* 15**,** 279-285.

Thomas, R. K., Nickerson, E., Simons, J. F., Janne, P. A., Tengs, T., Yuza, Y., Garraway, L. A., Laframboise, T., Lee, J. C., Shah, K., O'neill, K., Sasaki, H., Lindeman, N., Wong, K. K., Borras, A. M., Gutmann, E. J., Dragnev, K. H., Debiasi, R., Chen, T. H., Glatt, K. A., Greulich, H., Desany, B., Lubeski, C. K., Brockman, W., Alvarez, P., Hutchison, S. K., Leamon, J. H., Ronan, M. T., Turenchalk, G. S., Egholm, M., Sellers, W. R., Rothberg, J. M. & Meyerson, M. 2006. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med,* 12**,** 852-5.

Thompson, E. R., Doyle, M. A., Ryland, G. L., Rowley, S. M., Choong, D. Y. H., Tothill, R. W., Thorne, H., Barnes, D. R., Li, J., Ellul, J., Philip, G. K., Antill, Y. C., James, P. A., Trainer, A. H., Mitchell, G., Campbell, I. G. & Kconfab 2012. Exome Sequencing Identifies Rare Deleterious Mutations in DNA Repair Genes FANCC and BLM as Potential Breast Cancer Susceptibility Alleles. *Plos Genetics,* 8.

Thomson, P. J., Soames, J. V., Booth, C. & O'shea, J. A. 2002. Epithelial cell proliferative activity and oral cancer progression. *Cell Proliferation,* 35**,** 110-120.

Toyoshima, T., Vairaktaris, E., Nkenke, E., Schlegel, K. A., Neukam, F. W. & Ries, J. 2008. Cytokeratin 17 mRNA expression has potential for diagnostic marker of oral squamous cell carcinoma. *Journal of Cancer Research and Clinical Oncology,* 134**,** 515-521.

Tripathi, A., King, C., De La Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., Hirsch, E. F., Kavanah, M., Mendez, J., Stone, M., Gerry, N. P., Lenburg, M. E. & Rosenberg, C. L. 2008. Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *International Journal of Cancer,* 122**,** 1557-1566.

Tsai, C. C., Chen, C. C., Lin, C. C., Chen, C. H., Lin, T. S. & Shieh, T. Y. 1999. Interleukin-1 beta in oral submucous fibrosis, verrucous hyperplasia and squamous cell carcinoma tissues. *Kaohsiung J Med Sci,* 15**,** 513-9.

Tsai, Y. S., Lee, K. W., Huang, J. L., Liu, Y. S., Juo, S. H. H., Kuo, W. R., Chang, J. G., Lin, C. S. & Jong, Y. J. 2008. Arecoline, a major alkaloid of areca nut, inhibits p53, represses DNA repair, and triggers DNA damage response in human epithelial cells. *Toxicology,* 249**,** 230-237.

Tsantoulis, P. K., Kastrinakis, N. G., Tourvas, A. D., Laskaris, G. & Gorgoulis, V. G. 2007. Advances in the biology of oral cancer. *Oral oncology,* 43**,** 523-534.

Tsui, I. F. L., Rosin, M. P., Zhang, L. W., Ng, R. T. & Lam, W. L. 2008. Multiple Aberrations of Chromosome 3p Detected in Oral Premalignant Lesions. *Cancer Prevention Research,* 1**,** 424-429.

Tsukamoto, Y., Uchida, T., Karnan, S., Noguchi, T., Nguyen, L. T., Tanigawa, M., Takeuchi, I., Matsuura, K., Hijiya, N., Nakada, C., Kishida, T., Kawahara, K., Ito, H., Murakami, K., Fujioka, T., Seto, M. & Moriyama, M. 2008. Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer. *J Pathol,* 216**,** 471-82.

Tuch, B. B., Laborde, R. R., Xu, X., Gu, J., Chung, C. B., Monighetti, C. K., Stanley, S. J., Olsen, K. D., Kasperbauer, J. L., Moore, E. J., Broomer, A. J., Tan, R. Y., Brzoska, P. M., Muller, M. W., Siddiqui, A. S., Asmann, Y. W., Sun, Y. M., Kuersten, S., Barker, M. A., De La Vega, F. M. & Smith, D. I. 2010. Tumor Transcriptome Sequencing Reveals Allelic Expression Imbalances Associated with Copy Number Alterations. *Plos One,* 5.

Vaidya, M. M., Borges, A. M., Pradhan, S. A. & Bhisey, A. N. 1996. Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa. *Oral Oncology-European Journal of Cancer Part B,* 32B**,** 333-336.

Van Allen, E. M., Wagle, N., Stojanov, P., Perrin, D. L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D. C., Kryukov, G., Carter, S. L., Mckenna, A., Sivachenko, A., Rosenberg, M., Kiezun, A., Voet, D., Lawrence, M., Lichtenstein, L. T., Gentry, J. G., Huang, F. W., Fostel, J., Farlow, D., Barbie, D., Gandhi, L., Lander, E. S., Gray, S. W., Joffe, S., Janne, P., Garber, J., Macconaill, L., Lindeman, N., Rollins, B., Kantoff, P., Fisher, S. A., Gabriel, S., Getz, G. & Garraway, L. A. 2014. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nature Medicine,* 20**,** 682-688.

Van Heerden, W. F. & Van Zyl, A. W. 2009. Surgical pathology of oral cancer. *Diagnostic histopathology,* 15**,** 296-302.

Vanderriet, P., Nawroz, H., Hruban, R. H., Corio, R., Tokino, K., Koch, W. & Sidransky, D. 1994. Frequent Loss of Chromosome 9p21-22 Early in Head and Neck-Cancer Progression. *Cancer Research,* 54**,** 1156-1158.

Vartanian, J. G., Carvalho, A. L., De Arajo, M. J., Hattori, M., Magrin, J. & Kowalski, L. P. 2004. Predictive factors and distribution of lymph node metastasis in lip cancer patients and their implications on the treatment of the neck. *Oral Oncology,* 40**,** 223-227.

Vermeulen, S., Vanmarck, V., Vanhoorde, L., Vanroy, F., Bracke, M. & Mareel, M. 1996. Regulation of the invasion suppressor function of the cadherin/catenin complex. *Pathology Research and Practice,* 192**,** 694-707.

Virgilio, L., Shuster, M., Gollin, S. M., Veronese, M. L., Ohta, M., Huebner, K. & Croce, C. M. 1996. FHIT gene alterations in head and neck squamous cell carcinomas. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 9770-9775.

Walvekar, R. R., Chaukar, D. A., Deshpande, M. S., Pai, P. S., Chaturvedi, P., Kakade, A., Kane, S. V. & D'cruz, A. K. 2009. Verrucous carcinoma of the oral cavity: A clinical and pathological study of 101 cases. *Oral oncology,* 45**,** 47-51.

Wang, K., Yuen, S. T., Xu, J. C., Lee, S. P., Yan, H. H. N., Shi, S. T., Siu, H. C., Deng, S. B., Chu, K. M., Law, S., Chan, K. H., Chan, A. S. Y., Tsui, W. Y., Ho, S. L., Chan, A. K. W., Man, J. L. K., Foglizzo, V., Ng, M. K., Chan, A. S., Ching, Y. P., Cheng, G. H. W., Xie, T., Fernandez, J., Li, V. S. W., Clevers, H., Rejto, P. A., Mao, M. & Leung, S. Y. 2014a. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nature Genetics,* 46**,** 573-582.

Wang, L., Tsutsumi, S., Kawaguchi, T., Nagasaki, K., Tatsuno, K., Yamamoto, S., Sang, F., Sonoda, K., Sugawara, M., Saiura, A., Hirono, S., Yamaue, H., Miki, Y., Isomura, M., Totoki, Y., Nagae, G., Isagawa, T., Ueda, H., Murayama-Hosokawa, S., Shibata, T., Sakamoto, H., Kanai, Y., Kaneda, A., Noda, T. & Aburatani, H. 2012a. Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res,* 22**,** 208-19.

Wang, N., Zhou, F., Xiong, H., Du, S., Ma, J., Okai, I., Wang, J., Suo, J., Hao, L., Song, Y., Hu, J. & Shao, S. 2012b. Screening and identification of distant metastasis-

related differentially expressed genes in human squamous cell lung carcinoma. *Anat Rec (Hoboken),* 295**,** 748-57.

Wang, X. S., Wang, F. W., Taniguchi, K., Seelan, R. S., Wang, L., Zarfas, K. E., Mcdonnell, S. K., Qian, C. P., Pan, K. F., Lu, Y. Y., Shridhar, V., Couch, F. J., Tindall, D. J., Beebe-Dimmer, J. L., Cooney, K. A., Isaacs, W. B., Jacobsen, S. J., Schaid, D. J., Thibodeau, S. N. & Liu, W. G. 2006. Truncating variants in p53AIP1 disrupting DNA damage-induced apoptosis are associated with prostate cancer risk. *Cancer Research,* 66**,** 10302-10307.

Wang, Y. H., Tian, X., Liu, O. S., Fang, X. D., Quan, H. Z., Xie, S., Gao, S. & Tang, Z. G. 2014b. Gene profiling analysis for patients with oral verrucous carcinoma and oral squamous cell carcinoma. *Int J Clin Exp Med,* 7**,** 1845-52.

Wang, Y. P., Chen, H. M., Kuo, R. C., Yu, C. H., Sun, A., Liu, B. Y., Kuo, Y. S. & Chiang, C. P. 2009a. Oral verrucous hyperplasia: histologic classification, prognosis, and clinical implications. *Journal of Oral Pathology & Medicine,* 38**,** 651-6.

Wang, Y. P., Chen, H. M., Kuo, R. C., Yu, C. H., Sun, A. D., Liu, B. Y., Kuo, Y. S. & Chiang, C. P. 2009b. Oral verrucous hyperplasia: histologic classification, prognosis, and clinical implications. *Journal of Oral Pathology & Medicine,* 38**,** 651-656.

Wang, Z., Gerstein, M. & Snyder, M. 2009c. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics,* 10**,** 57-63.

Warnakulasuriya, S. 2009. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol,* 45**,** 309-16.

Wei, K. J., Zhang, L., Yang, X., Zhong, L. P., Zhou, X. J., Pan, H. Y., Li, J., Chen, W. T. & Zhang, Z. Y. 2009. Overexpression of cytokeratin 17 protein in oral squamous cell carcinoma in vitro and in vivo. *Oral Dis,* 15**,** 111-7.

Wei, Q., Sheng, L., Shui, Y., Hu, Q., Nordgren, H. & Carlsson, J. 2008. EGFR, HER2, and HER3 expression in laryngeal primary tumors and corresponding metastases. *Ann Surg Oncol,* 15**,** 1193-201.

Weinlich, G., Eisendle, K., Hassler, E., Baltaci, M., Fritsch, P. O. & Zelger, B. 2006. Metallothionein - overexpression as a highly significant prognostic factor in melanoma: a prospective study on 1270 patients. *British Journal of Cancer,* 94**,** 835-841.

Weir, B. A., Woo, M. S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W. M., Province, M. A., Kraja, A., Johnson, L. A., Shah, K., Sato, M., Thomas, R. K., Barletta, J. A., Borecki, I. B., Broderick, S., Chang, A. C., Chiang, D. Y., Chirieac, L. R., Cho, J., Fujii, Y., Gazdar, A. F., Giordano, T., Greulich, H., Hanna, M., Johnson, B. E., Kris, M. G., Lash, A., Lin, L., Lindeman, N., Mardis, E. R., Mcpherson, J. D., Minna, J. D., Morgan, M. B., Nadel, M., Orringer, M. B., Osborne, J. R., Ozenberger, B., Ramos, A. H., Robinson, J., Roth, J. A., Rusch, V., Sasaki, H., Shepherd, F., Sougnez, C., Spitz, M. R., Tsao, M. S., Twomey, D., Verhaak, R. G. W., Weinstock, G. M., Wheeler, D. A., Winckler, W., Yoshizawa, A., Yu, S. Y., Zakowski, M. F., Zhang, Q. Y., Beer, D. G., Wistuba, I. I., Watson, M. A., Garraway, L. A., Ladanyi, M., Travis, W. D., Pao, W., Rubin, M. A., Gabriel, S. B., Gibbs, R. A., Varmus, H. E., Wilson, R. K., Lander, E. S. & Meyerson, M. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature,* 450**,** 893-U22.

Williams, C., Ponten, F., Moberg, C., Soderkvist, P., Uhlen, M., Ponten, J., Sitbon, G. & Lundeberg, J. 1999. A high frequency of sequence alterations is due to formalin fixation of archival specimens. *American Journal of Pathology,* 155**,** 1467-1471.

Woenckhaus, M., Klein-Hitpass, L., Grepmeier, U., Merk, J., Pfeifer, M., Wild, P., Bettstetter, M., Wuensch, P., Blaszyk, H., Hartmann, A., Hofstaedter, F. & Dietmaier, W. 2006. Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J Pathol,* 210**,** 192-204.

Wong, K. K., Engelman, J. A. & Cantley, L. C. 2010. Targeting the PI3K signaling pathway in cancer. *Current Opinion in Genetics & Development,* 20**,** 87-90.

Wong, W. C., Kim, D., Carter, H., Diekhans, M., Ryan, M. C. & Karchin, R. 2011. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics,* 27**,** 2147-8.

Wood, H. M., Belvedere, O., Conway, C., Daly, C., Chalkley, R., Bickerdike, M., Mckinley, C., Egan, P., Ross, L., Hayward, B., Morgan, J., Davidson, L., Maclennan, K., Ong, T. K., Papagiannopoulos, K., Cook, I., Adams, D. J., Taylor, G. R. & Rabbitts, P. 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Research,* 38.

Woolgar, J. A. & Triantafyllou, A. 2009. Pitfalls and procedures in the histopathological diagnosis of oral and oropharyngeal squamous cell carcinoma and a review of the role of pathology in prognosis. *Oral oncology,* 45**,** 361-85.

Wu, C. F., Chen, C. M., Shen, Y. S., Huang, I. Y., Chen, C. H., Chen, C. Y., Shieh, T. Y. & Sheen, M. C. 2008. Effective eradication of oral verrucous carcinoma with continuous intraarterial infusion chemotherapy. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck,* 30**,** 611-617.

Wu, M., Putti, T. C. & Bhuiya, T. A. 2002a. Comparative study in the expression of p53, EGFR, TGF-alpha, and cyclin D1 in verrucous carcinoma, verrucous hyperplasia, and squamous cell carcinoma of head and neck region. *Appl Immunohistochem Mol Morphol,* 10**,** 351-6.

Wu, M. X., Putti, T. C. & Bhuiya, T. A. 2002b. Comparative study in the expression of p53, EGFR, TGF-alpha, and cyclin D1 in verrucous carcinoma, verrucous hyperplasia, and squamous cell carcinoma of head and neck region. *Applied Immunohistochemistry & Molecular Morphology,* 10**,** 351-356.

Xie, C. & Tammi, M. T. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *Bmc Bioinformatics,* 10.

Xie, D., Nakachi, K., Wang, H. M., Elashoff, R. & Koeffler, H. P. 2001. Elevated levels of connective tissue growth factor, WISP-1, and CYR61 in primary breast cancers associated with more advanced features. *Cancer Research,* 61**,** 8917-8923.

Xu, X. C., Lee, J. S., Lippman, S. M., Ro, J. Y., Hong, W. K. & Lotan, R. 1995. Increased Expression of Cytokeratins Ck8 and Ck19 Is Associated with Head and Neck Carcinogenesis. *Cancer Epidemiology Biomarkers & Prevention,* 4**,** 871-876.

Yanagawa, T., Yoshida, H., Yamagata, K., Onizawa, K., Tabuchi, K., Koyama, Y., Iwasa, S., Shimoyamada, H., Harada, H. & Omura, K. 2007. Loss of cytokeratin 13 expression in squamous cell carcinoma of the tongue is a possible sign for local recurrence. *Journal of Experimental & Clinical Cancer Research,* 26**,** 215-220.

Yang, M. H., Lin, B. R., Chang, C. H., Chen, S. T., Lin, S. K., Kuo, M. Y. P., Jeng, Y. M., Kuo, M. L. & Chang, C. C. 2012. Connective tissue growth factor modulates oral squamous cell carcinoma invasion by activating a miR-504/FOXP1 signalling. *Oncogene,* 31**,** 2401-2411.

Ye, H., Yu, T. W., Temam, S., Ziober, B. L., Wang, J. G., Schwartz, J. L., Mao, L., Wong, D. T. & Zhou, X. F. 2008a. Transcriptomic dissection of tongue squamous cell carcinoma. *Bmc Genomics,* 9.

Ye, Y. Q., Lippman, S. M., Lee, J. J., Chen, M., Frazier, M. L., Spitz, M. R. & Wu, X. F. 2008b. Genetic Variations in Cell-Cycle Pathway and the Risk of Oral Premalignant Lesions. *Cancer,* 113**,** 2488-2495.

Yeh, C. J. 2000. Simple cryosurgical treatment for oral lesions. *International Journal of Oral and Maxillofacial Surgery,* 29**,** 212-216.

Yeh, C. J. 2003. Treatment of verrucous hyperplasia and verrucous carcinoma by shave excision and simple cryosurgery. *International Journal of Oral and Maxillofacial Surgery,* 32**,** 280-283.

Yilmaz, M. & Christofori, G. 2009. EMT, the cytoskeleton, and cancer cell invasion. *Cancer and Metastasis Reviews,* 28**,** 15-33.

Yoshimura, Y., Mishima, K., Obara, S., Nariai, Y., Yoshimura, H. & Mikami, T. 2001. Treatment modalities for oral verrucous carcinomas and their outcomes: contribution of radiotherapy and chemotherapy. *Int J Clin Oncol,* 6**,** 192-200.

Yousef, A. & Hashash, M. 1983. Common features and surgical interference in a prevalent oral cancer in Saudi Arabia (a preliminary report). *J Laryngol Otol,* 97**,** 837-43.

Yu, K. K., Zanation, A. M., Moss, J. R. & Yarbrough, W. G. 2002. Familial head and neck cancer: molecular analysis of a new clinical entity. *Laryngoscope,* 112**,** 1587-93.

Zaahl, M. G., Warnich, L., Victor, T. C. & Kotze, M. J. 2005. Association of functional polymorphisms of SLC11A1 with risk of esophageal cancer in the South African Colored population. *Cancer Genetics and Cytogenetics,* 159**,** 48-52.

Zabarovsky, E. R., Lerman, M. I. & Minna, J. D. 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene,* 21**,** 6915-6935.

Zain, R. B., Anwar, A., Karen-Ng, L. P., Cheong, S. C., Zaini, Z., Prepagaran, N., Hercus, R., Idris, N. M., Saidin, A., Chia, J. Y., Merican, A. F., Abraham, M. T., Tay, K. K., Mustaffa, W. M. W., Norain, A. T., Rahman, Z. a. A. & Jallaludin, A. 2010. Selected transcriptome profiles of oral cancer suggestive of field cancerisation using second generation sequencing. *Oral Diseases,* 16**,** 521-521.

Zargaran, M., Eshghyar, N., Baghaei, F. & Moghimbeigi, A. 2012. Assessment of Cellular Proliferation in Oral Verrucous Carcinoma and Well-Differentiated Oral Squamous Cell Carcinoma Using Ki67: A Non-Reliable Factor for Differential Diagnosis? *Asian Pacific Journal of Cancer Prevention,* 13**,** 5811-5815.

Zargaran, M., Eshghyar, N., Vaziri, P. B. & Mortazavi, H. 2011a. Immunohistochemical evaluation of type IV collagen and laminin-332 gamma2 chain expression in well-differentiated oral squamous cell carcinoma and oral verrucous carcinoma: a new recommended cut-off. *J Oral Pathol Med,* 40**,** 167-73.

Zargaran, M., Eshghyar, N., Vaziri, P. B. & Mortazavi, H. 2011b. Immunohistochemical evaluation of type IV collagen and laminin-332 gamma 2 chain expression in well-differentiated oral squamous cell carcinoma and oral verrucous carcinoma: a new recommended cut-off. *Journal of Oral Pathology & Medicine,* 40**,** 167-173.

Zhang, L. & Rosin, M. P. 2001. Loss of heterozygosity: a potential tool in management of oral premalignant lesions? *Journal of Oral Pathology & Medicine,* 30**,** 513-520.

Zhang, Q., Zhang, J., Jin, H. & Sheng, S. T. 2013. Whole transcriptome sequencing identifies tumor-specific mutations in human oral squamous cell carcinoma. *Bmc Medical Genomics,* 6.

Zhao, X. J., Li, C., Paez, J. G., Chin, K., Janne, P. A., Chen, T. H., Girard, L., Minna, J., Christiani, D., Leo, C., Gray, J. W., Sellers, W. R. & Meyerson, M. 2004. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Research,* 64**,** 3060-3071.

Zhu, L. K., Ding, Y. W., Liu, W., Zhou, Y. M., Shi, L. J. & Zhou, Z. T. 2012. A clinicopathological study on verrucous hyperplasia and verrucous carcinoma of the oral mucosa. *Journal of Oral Pathology & Medicine,* 41**,** 131-135.

Zwemer, L. M., Hui, L., Wick, H. C. & Bianchi, D. W. 2014. RNA-Seq and expression microarray highlight different aspects of the fetal amniotic fluid transcriptome. *Prenat Diagn*.

Appendices

Chapter 2 Appendices

Appendix 2.1 List of suppliers.

Supplier	Addresses	E.mail addresses
Agencourt Bioscience	MA, USA	webinfo@agencourt.com
Agilent technologies	CA, USA	
BDH	Poole, UK	uksales@uk.vwr.com
BioLabs	Ipswich, MA, USA	orderinfo@neb.com
Covaris Inc.	Massachusetts, USA	info@covarisinc.com
Epicentre	USA	
Gibco	Paisley, UK	euroinfo@invitrogen.com
Illumina Inc.	UK	info@illumina.com
Integrated DNA technologies	UK	custcare@idtdna.com
Invitrogen	Paisley, UK	euroinfo@invitrogen.com
Nikon	Florida, USA	
Olympus	Essex, UK	Microscope@olympus.uk.com
Promega	Madison, WI, USA	custserv@promega.com
Qiagen	West Sussex, UK	customercare@qiagen.com
Sigma-Aldrich	St Louis Mo, USA	ukorders@sial.com
Solmedia	UK	labsupplies@solmedialtd.com
Thermo Fisher Scientific	USA	TSsupport@thermofisher.com

Appendix 2.2 Ethical approval letter

NHS

National Research Ethics Service
Leeds (East) Research Ethics Committee

Yorkshire and Humber REC Office
First Floor, Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Tel: 0113 3050108

Professor Pamela Rabbitts
Professor of Experimental Respiratory Research
University of Leeds
St James's University Hospital NHS Trust
Beckett ST
LEEDS
LS9 7TF

01 November 2010

Dear Professor Rabbitts

Study title:	**Molecular abnormalities associated with tumour development in early cancer: characteristics and consequences**
REC reference:	**07/Q1206/30**
Amendment number:	**12**
Amendment date:	**23 September 2010**

The above amendment was reviewed at the meeting of the Sub-Committee held on 21 October 2010.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
CV of Manar Samam		
Notice of Substantial Amendment (non-CTIMPs)		23 September 2010

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

07/Q1206/30:	**Please quote this number on all correspondence**

Yours sincerely

pp.

Miss Jade Thorpe
Assistant Committee Co-ordinator

E-mail: jade.thorpe@leedspft.nhs.uk

Copy to: *Rachel de Souza, University of Leeds*

Appendix 2.3 GISTIC2.0 parameters

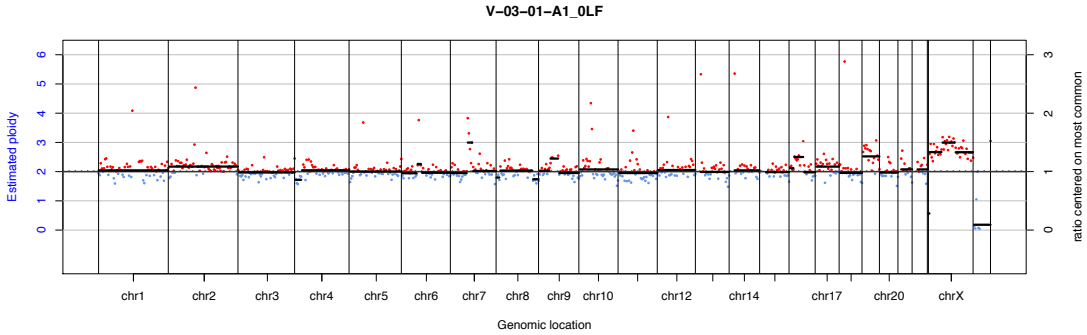
Name	Option String	Description
refgene.file (required)	-refgene	The reference file including cytoband and gene location information. (Default: hg19)
seg.file (required)	-seg	The segmentation file contains the segmented data for all the samples identified by GLAD, CBS, or some other segmentation algorithm. (See GLAD file format for more information.) It is a six- column, tab-delimited file with an optional first line identifying the columns. Positions are in base pair units.
markers.file (required)	-mk	The markers file identifies the marker names and positions of the markers in the original dataset before segmentation. It is a three-column, tab- delimited file with an optional header. Markers are sorted by genomic position if the file is not already in that order.
gene.gistic (required)	-genegistic	Flag indicating that the gene GISTIC algorithm should be used to calculate the significance of deletions at a gene level instead of a marker level. (Default: yes)
amplifications. threshold (required)	-ta	Threshold for copy number amplifications. Regions with a log2 ratio above this value are considered amplified. (Default: 0.1)
deletions. threshold (required)	-td	Threshold for copy number deletions. Regions with a log2 ratio below the negative of this value are considered deletions. (Default: 0.1)
join.segment. size (required)	-js	Threshold for copy number deletions. Regions with a log2 ratio below the negative of this value are considered deletions. (Default: 0.1)
qv.thresh (required)	-qvt	Significance threshold for q-values. Regions with q-values below this number are considered significant. (Default: 0.25)
remove.x (required)	-rx	Flag indicating whether to remove data from the X chromosome before analysis. (Default: yes)
confidence.level (required)	-conf	Confidence level used to calculate the region containing a driver. (Default: 0.75)
run.broad. analysis (required)	-broad	Flag indicating whether an additional broad-level analysis should be performed. (Default: no).
broad.length. cutoff (required)	-brlen	Threshold used to distinguish broad from focal events, given in units of fraction of chromosome arm. (Default: 0.98)
max.sample. segs	-maxseg	Maximum number of segments allowed for a sample in the input data. Samples with more segments than this threshold are excluded from the analysis. (Default: 2500)
arm.peel	-armpeel	Whether to perform arm level peel off, which helps separate peaks and clean up noise. (Default: no)
output.prefix (required)	-fname	The prefix for the output file name.

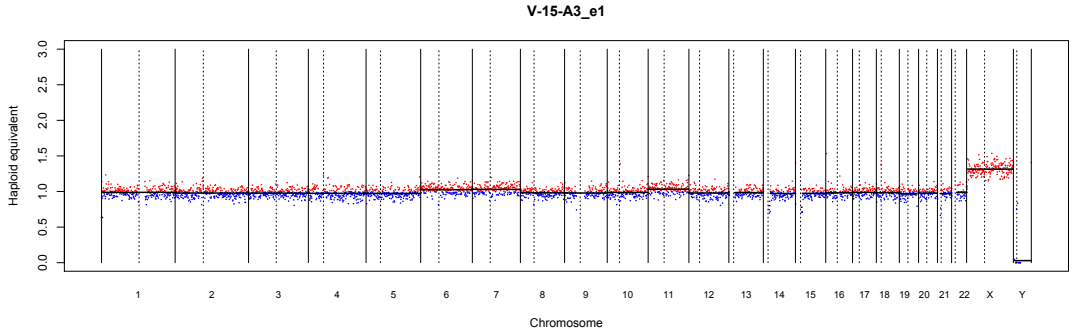
Chapter 4 Appendices

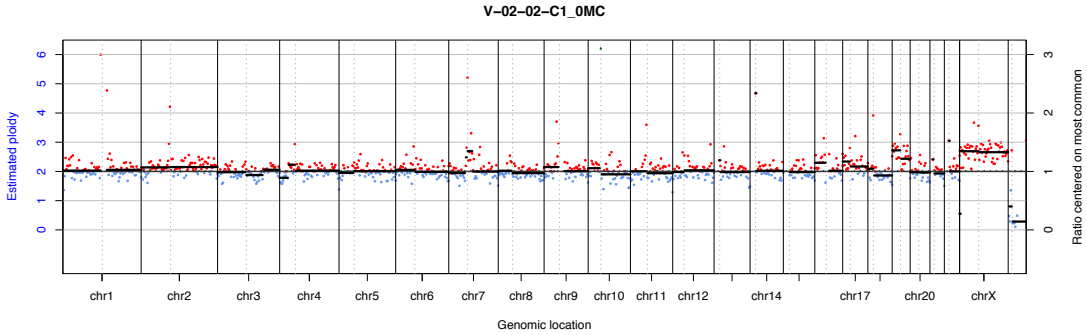
Appendix 4.1 Failed verrucous samples.

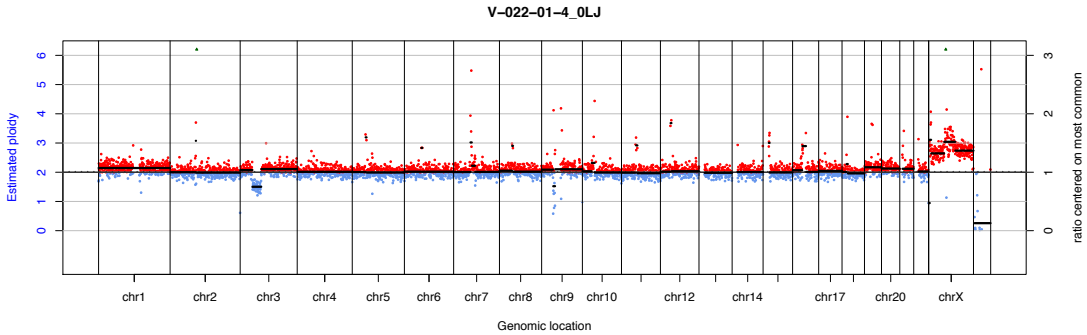
Sample ID	Fail reason	Additional comments
V-06-01	Library preparation failed	Illumina GA Library preparation protocol, Library DNA too low
V-09-01	Library preparation failed	Illumina GA Library preparation protocol
V-13-01	Library preparation failed	Illumina GA Library preparation protocol, Library DNA too low
V-16-01	Library preparation failed	Illumina GA Library preparation protocol, Library DNA too low
V-17-01	Library preparation failed	Illumina GA Library preparation protocol
V-24-01	Library preparation failed	Illumina GA Library preparation protocol, Library DNA too low
V-32-01	Library preparation failed	Illumina GA Library preparation protocol, Library DNA too low
V-34-01	Low DNA yield after extraction	
V-35-01	Library preparation failed	Illumina GA Library preparation protocol
V-39-01	Low DNA yield after extraction	
V-43-01	Library preparation failed	Illumina GA Library preparation protocol
V-45-01	Library preparation failed	Illumina GA Library preparation protocol
V-56-01	Low DNA yield after extraction	
V-59-01	Low DNA yield after extraction	
V-64-01	Low DNA yield after extraction	
V-76-01	Low DNA yield after extraction	
V-81-01	Library preparation failed	Very old sample, 1993
V-82-01	Low DNA yield after extraction	
V-103-01	Low DNA yield after extraction	

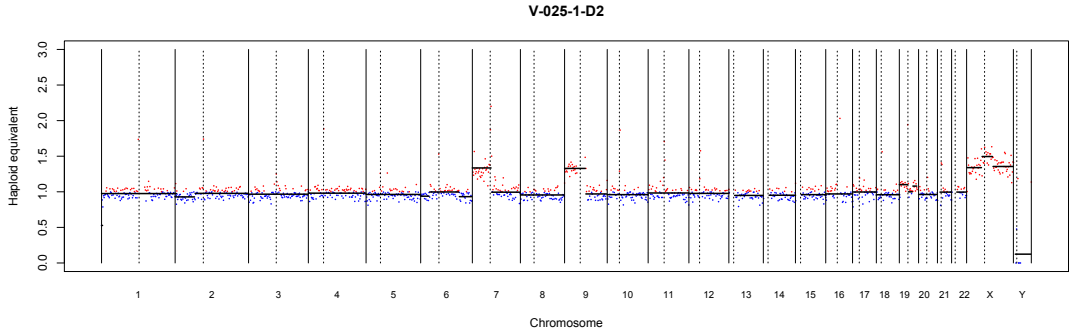
Appendix 4.2 OVH CN keryograms

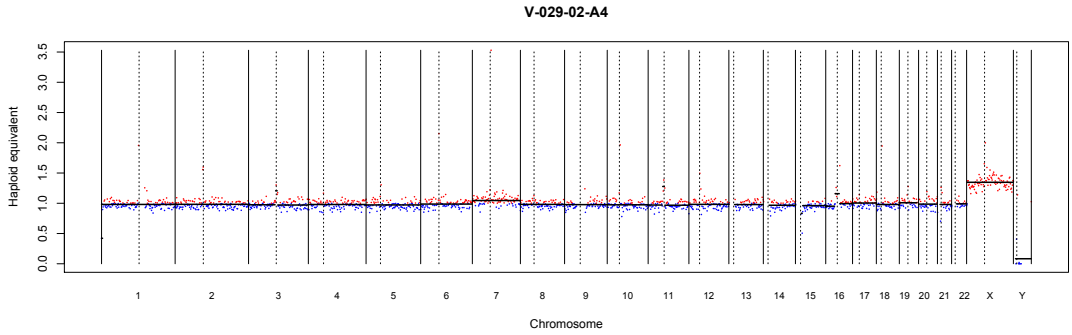


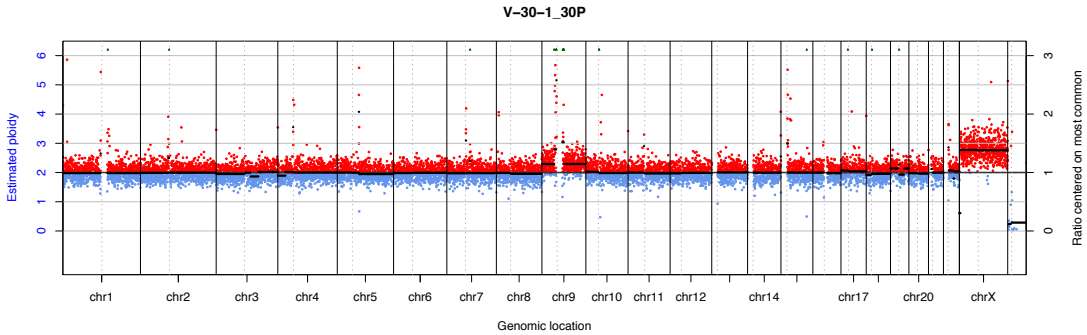


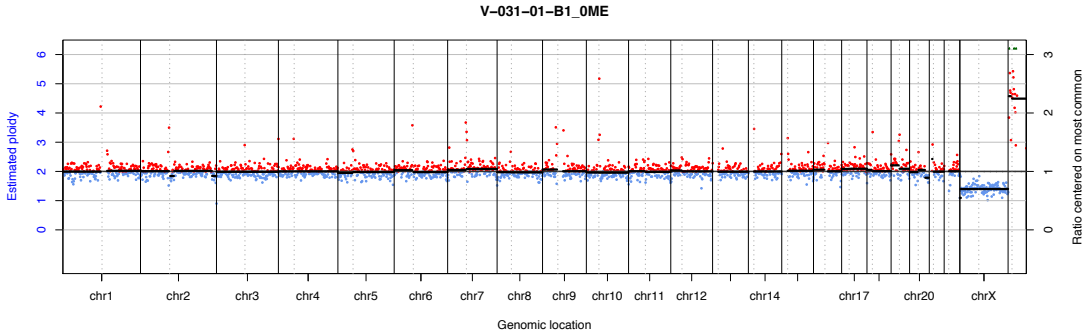


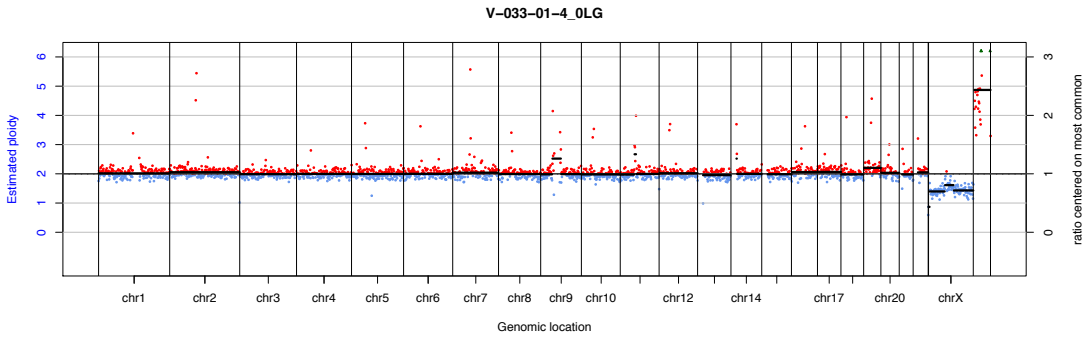


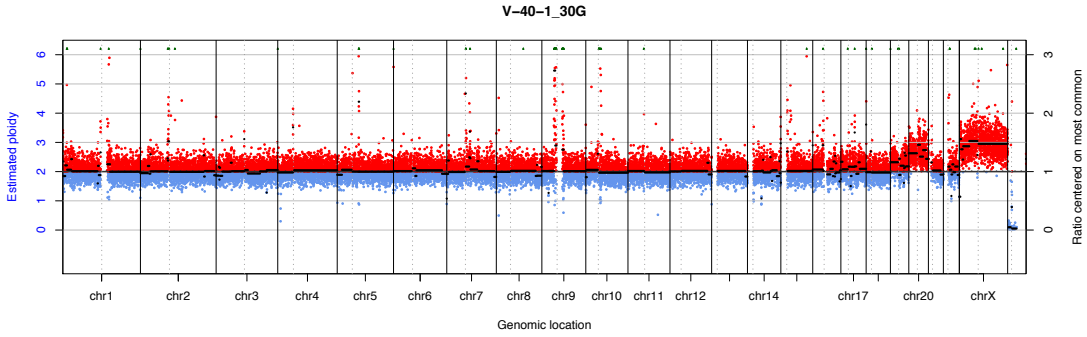

V-025-1-D2
Haploid equivalent
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Chromosome

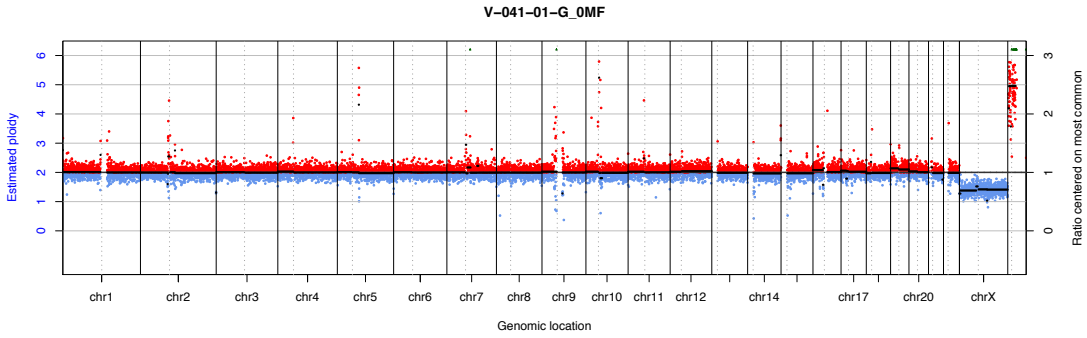

V-029-02-A4
Haploid equivalent
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Chromosome

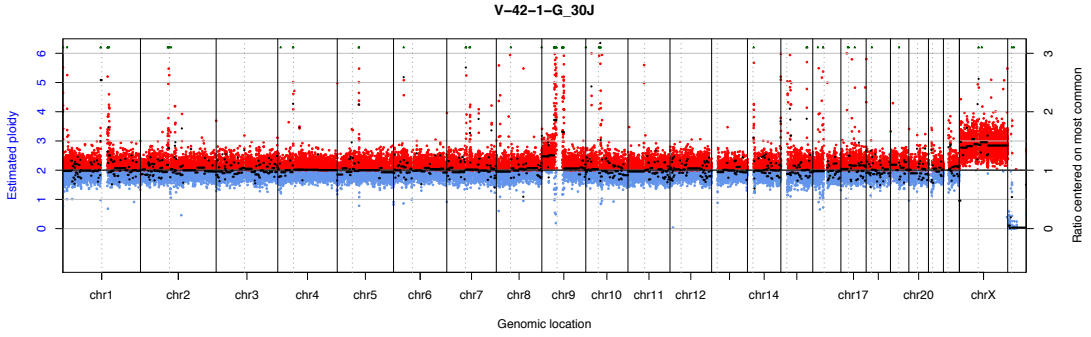

V-30-1_30P
Estimated ploidy
0
1
2
3
4
5
6
Ratio centered on most common
0
1
2
3
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

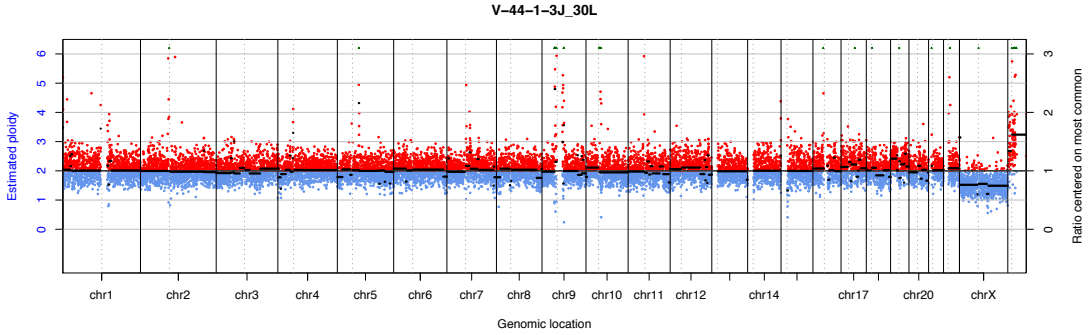

V-031-01-B1_0ME
Estimated ploidy
0
1
2
3
4
5
6
Ratio centered on most common
0
1
2
3
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

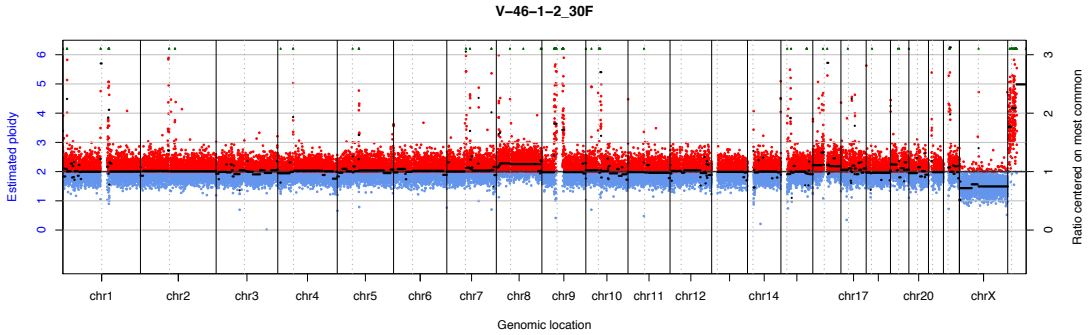

V−033−01−4_0LG
Estimated ploidy
ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr12 chr14 chr17 chr20 chrX
Genomic location

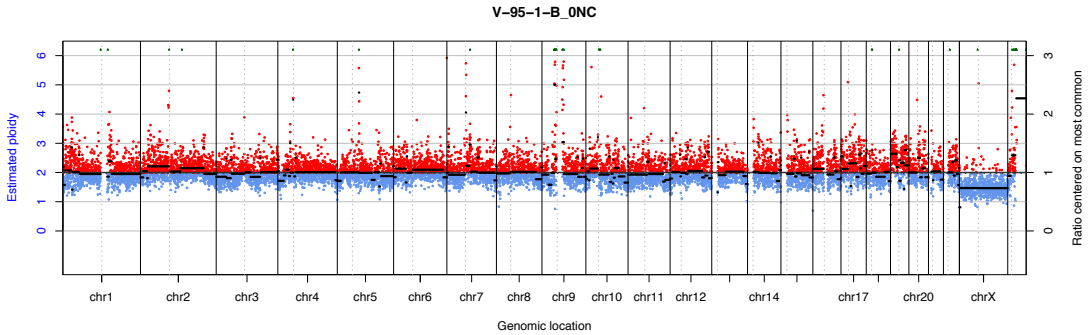

V−40−1_30G
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

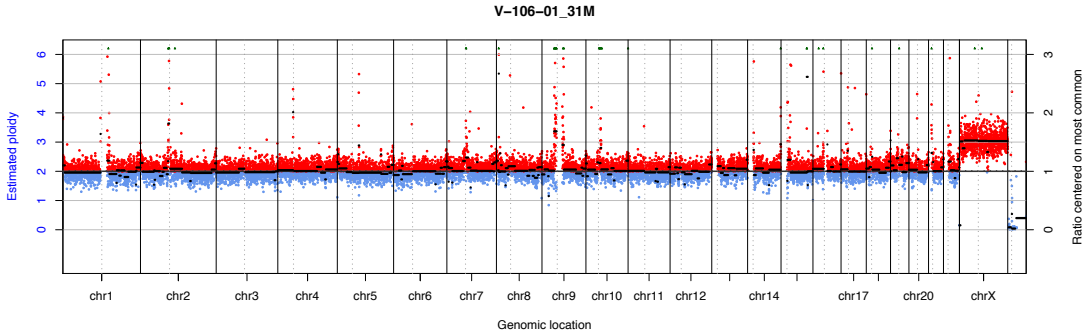

V−041−01−G_0MF
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location


V−42−1−G_30J
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

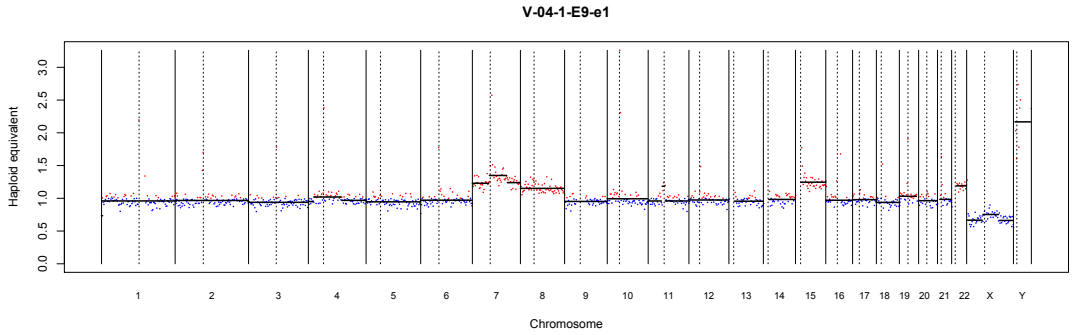

V−44−1−3J_30L
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

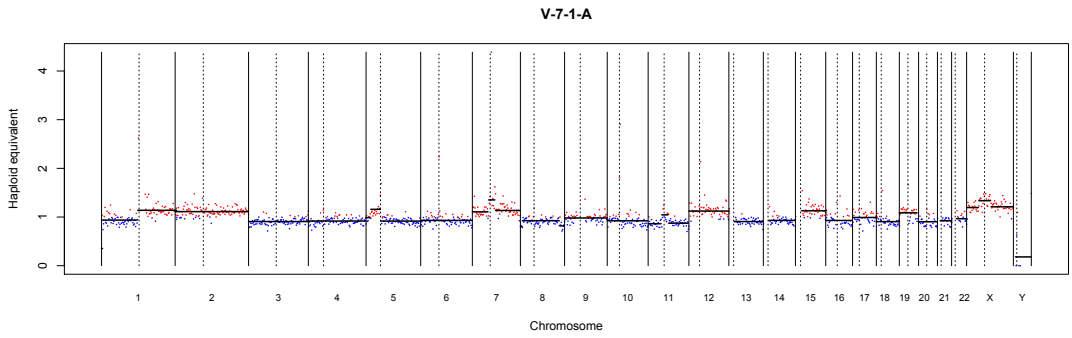

V−46−1−2_30F
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

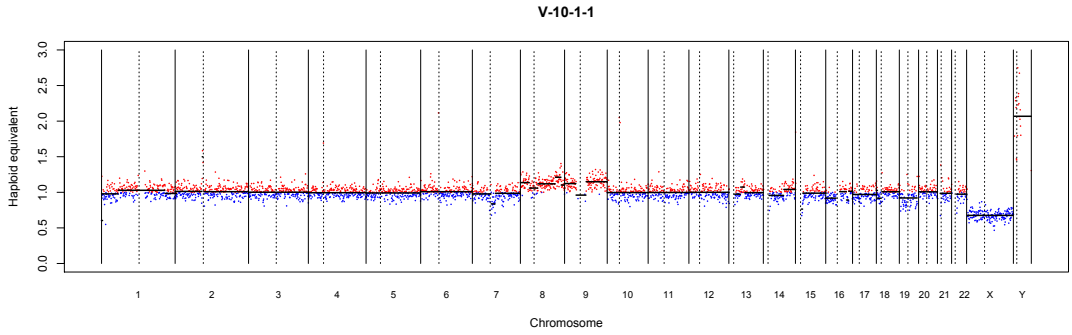

V−95−1−B_0NC
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

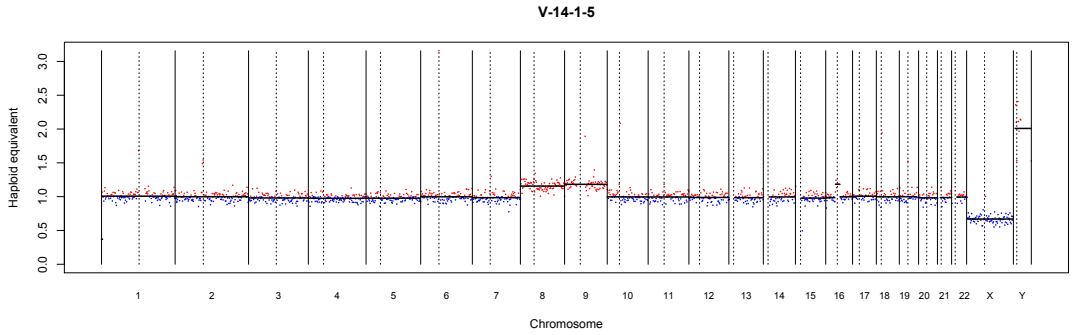

V−106−01_31M
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

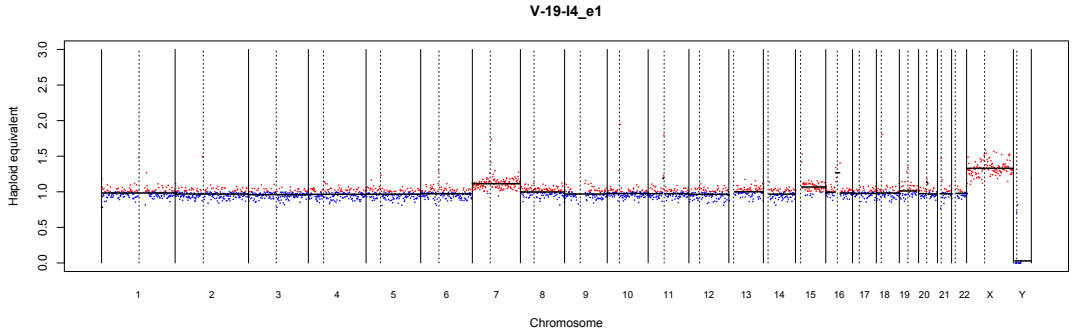
Appendix 4.3 OVC CN keryograms

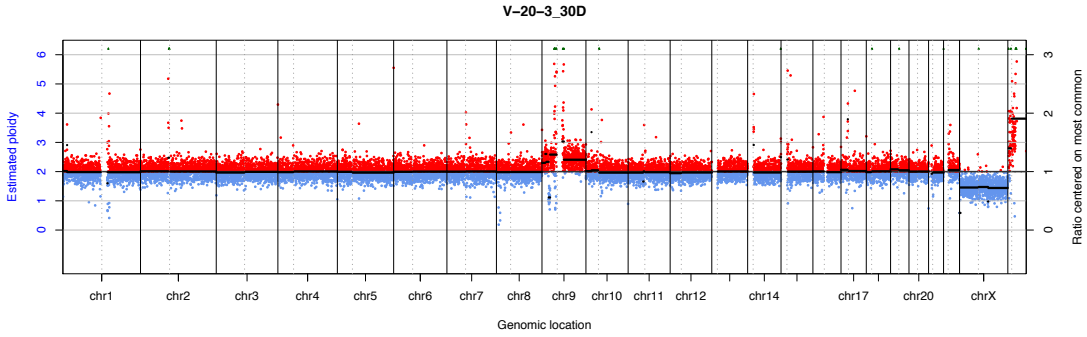


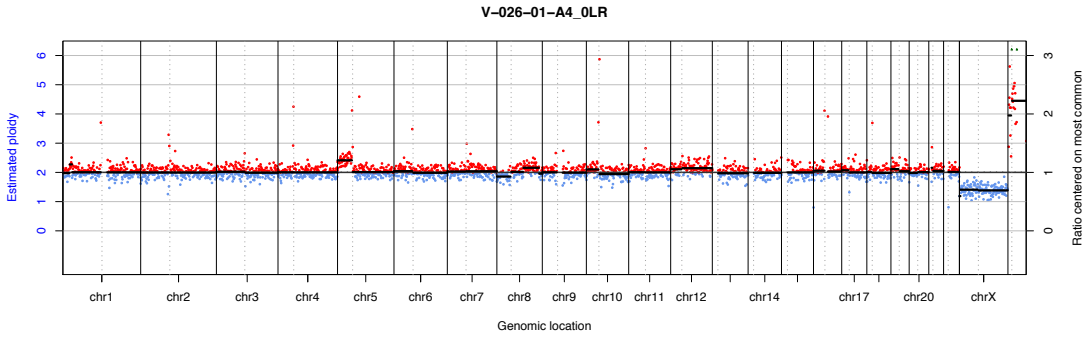


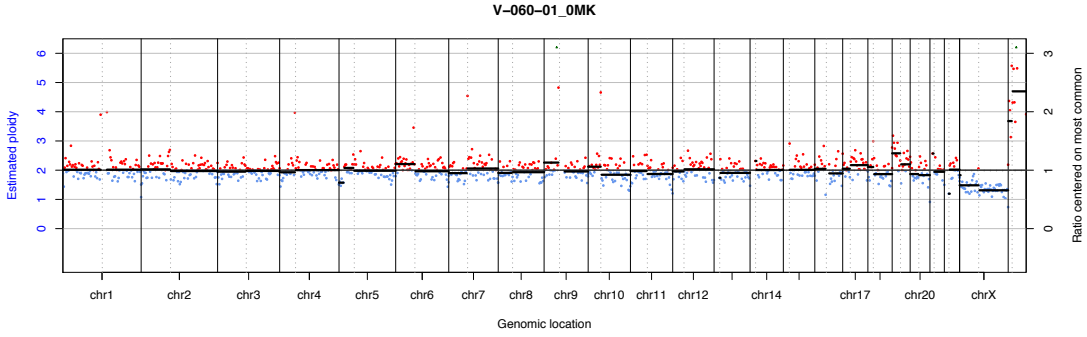


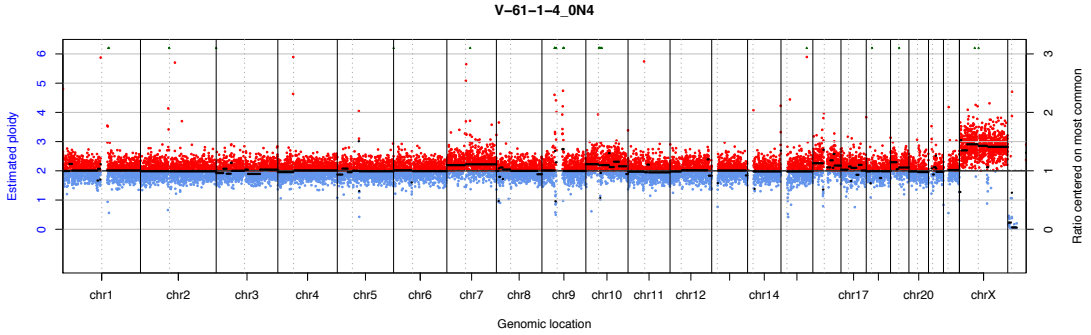


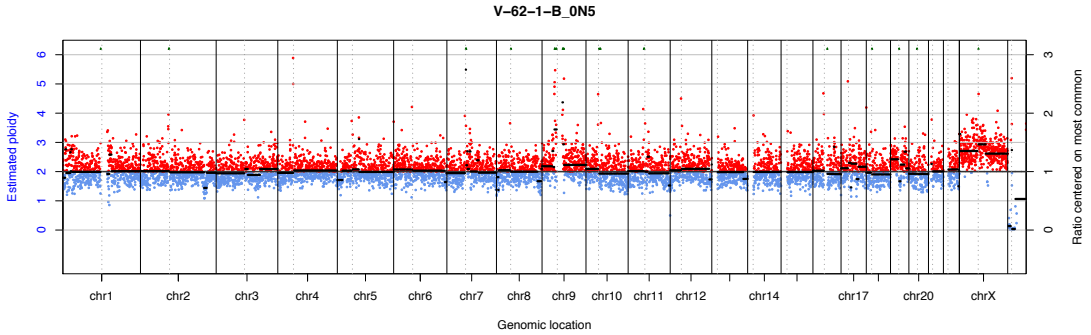

V-19-I4_e1
Haploid equivalent
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Chromosome

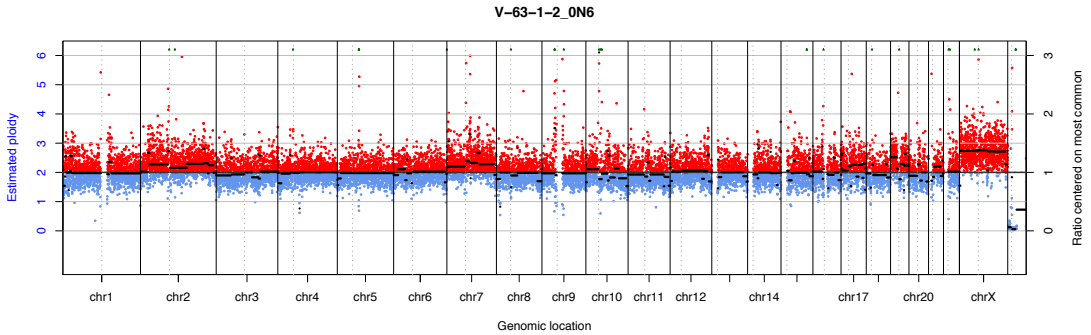

V-20-3_30D
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

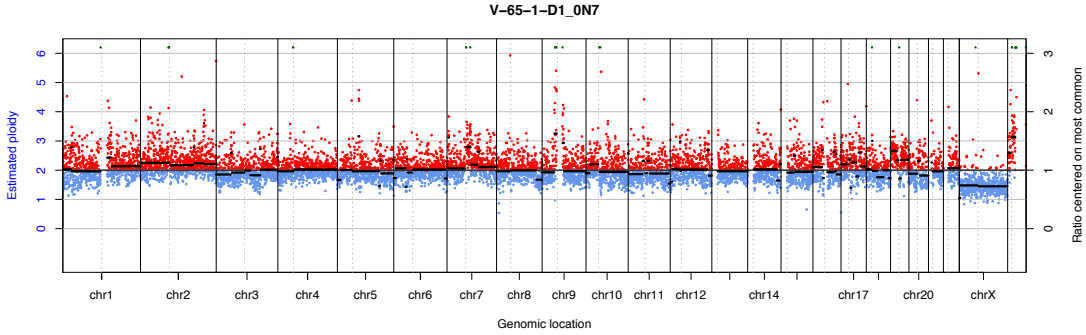

V-026-01-A4_0LR
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

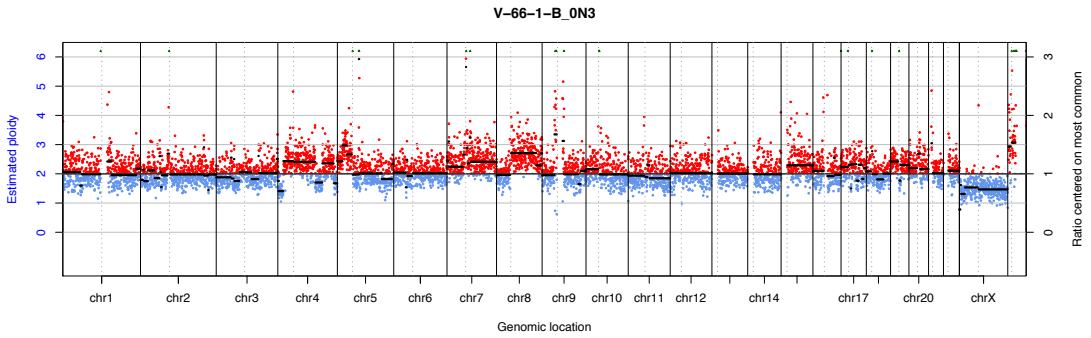

V-060-01_0MK
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

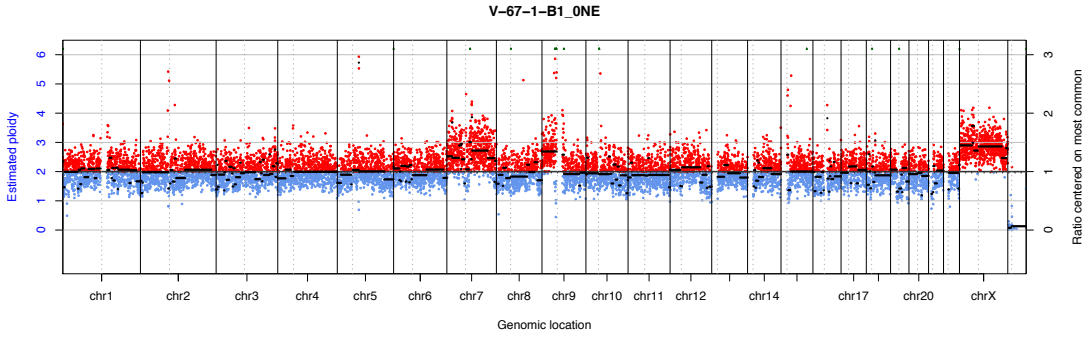

V−61−1−4_0N4
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

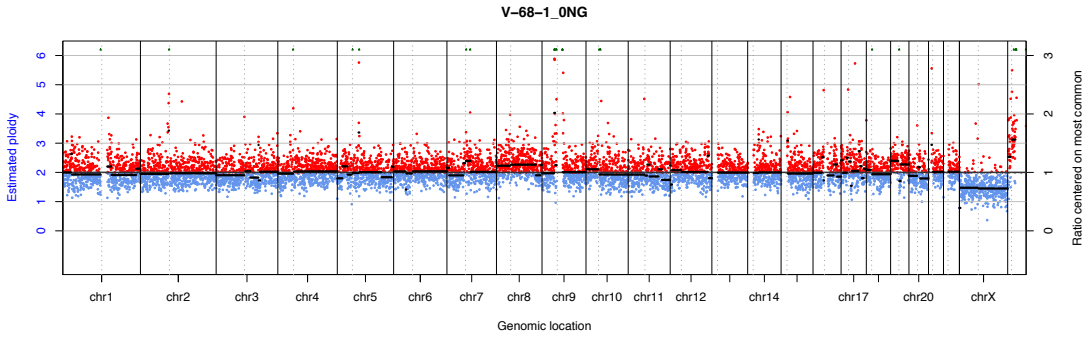

V−62−1−B_0N5
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

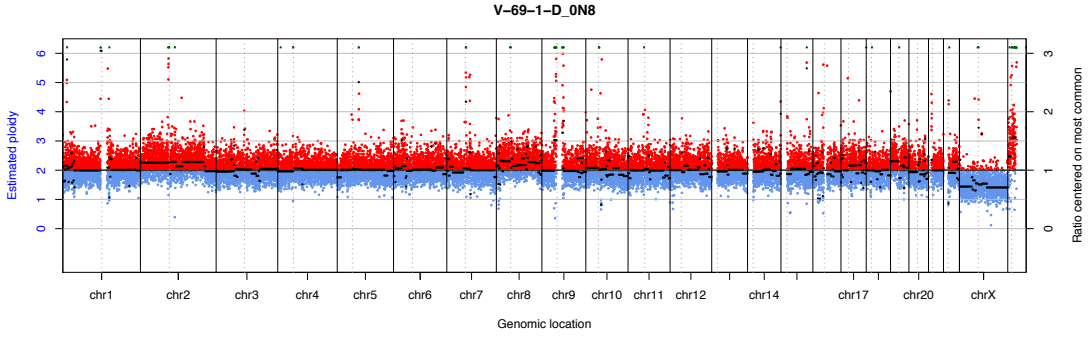

V−63−1−2_0N6
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

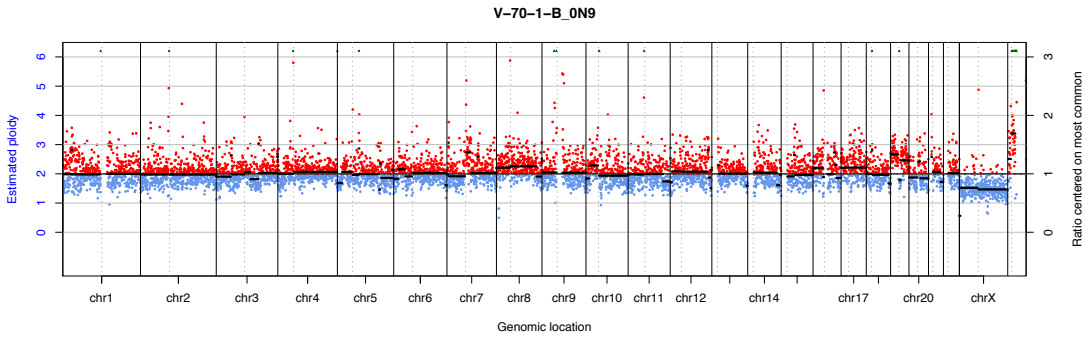

V−65−1−D1_0N7
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

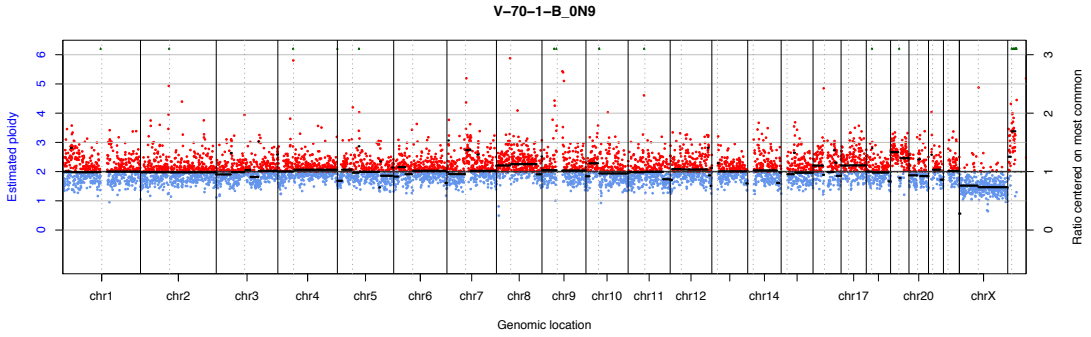

V−66−1−B_0N3
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

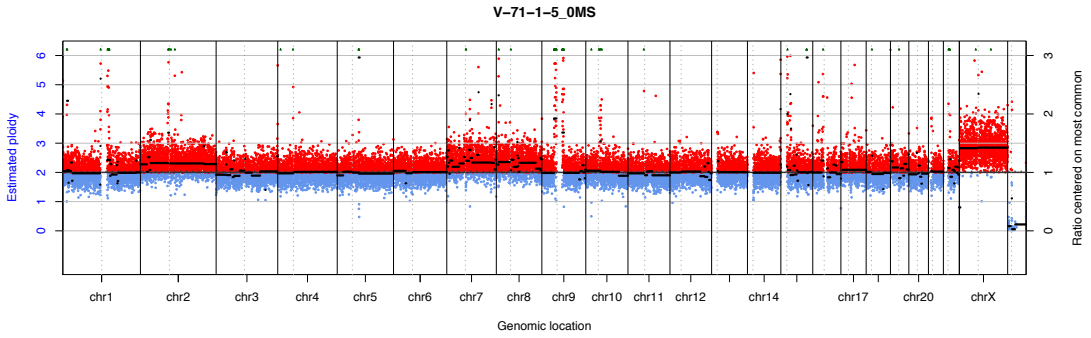

V−67−1−B1_0NE
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

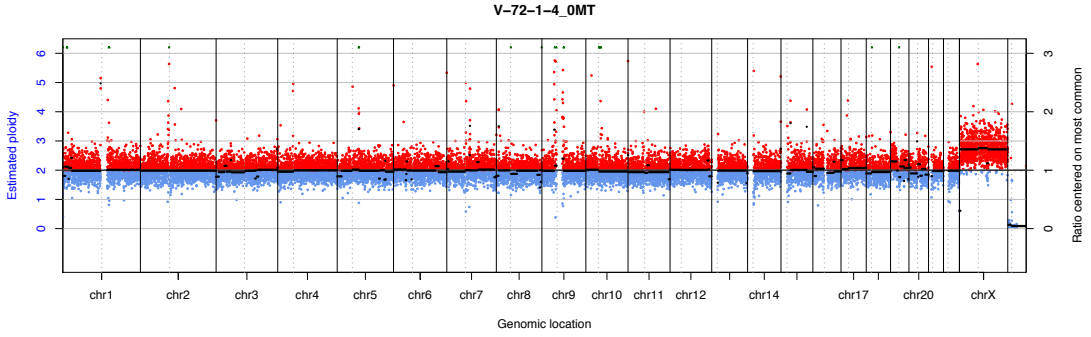

V−68−1_0NG
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

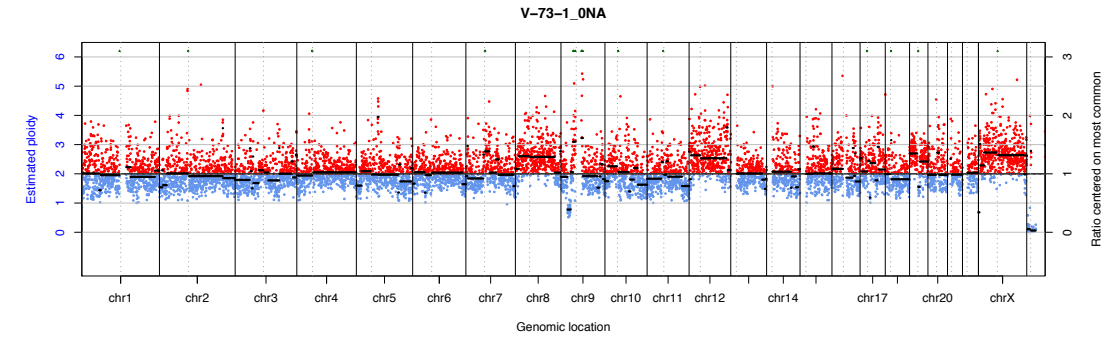

V−69−1−D_0N8
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

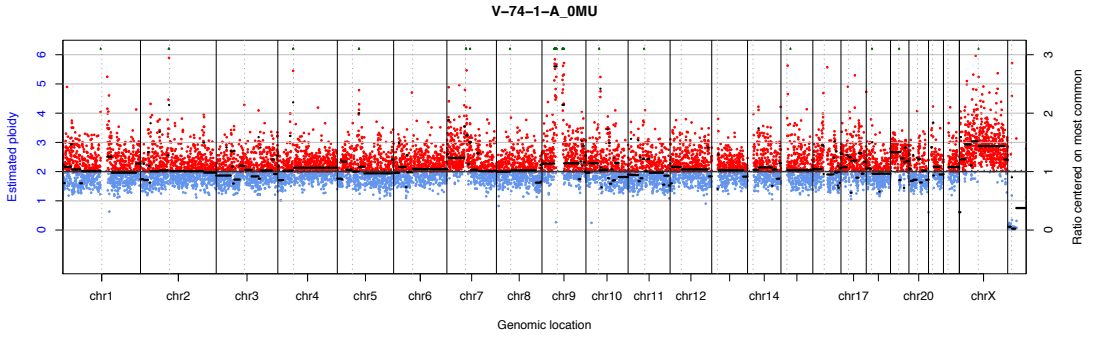

V−70−1−B_0N9
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

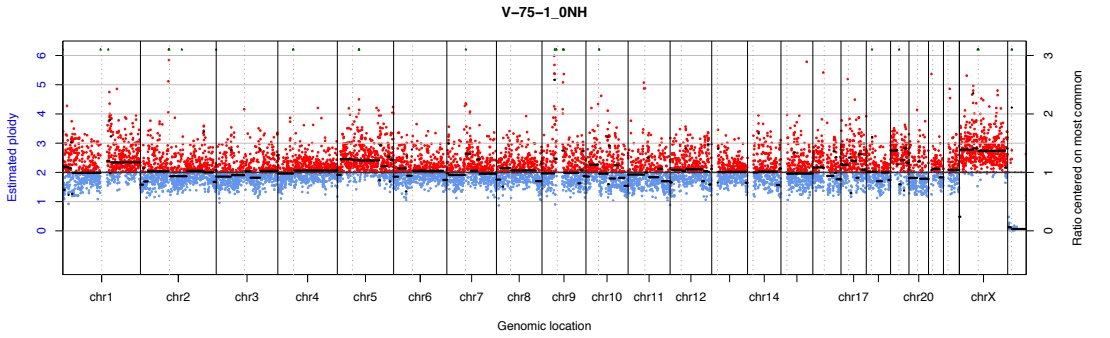

V−70−1−B_0N9
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

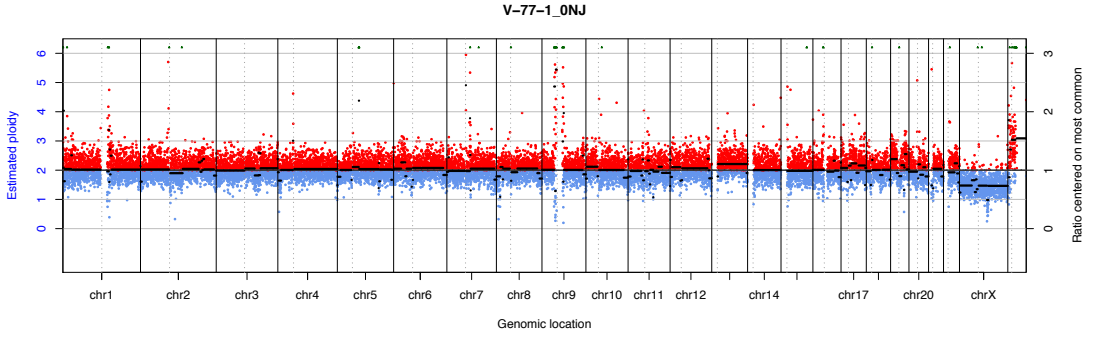

V−71−1−5_0MS
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

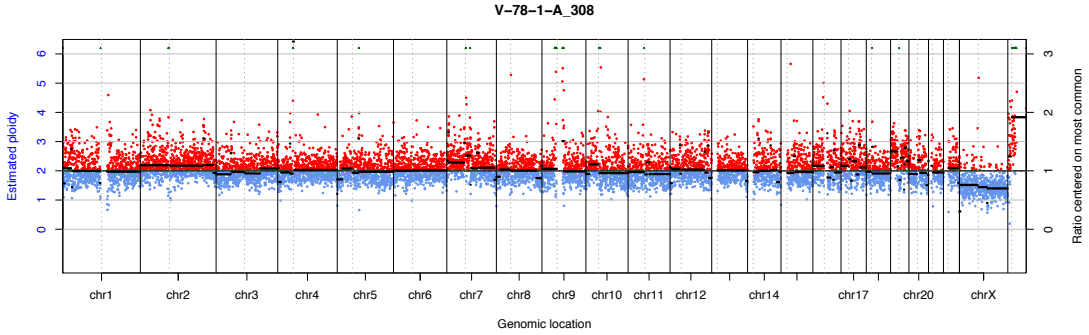

V−72−1−4_0MT
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

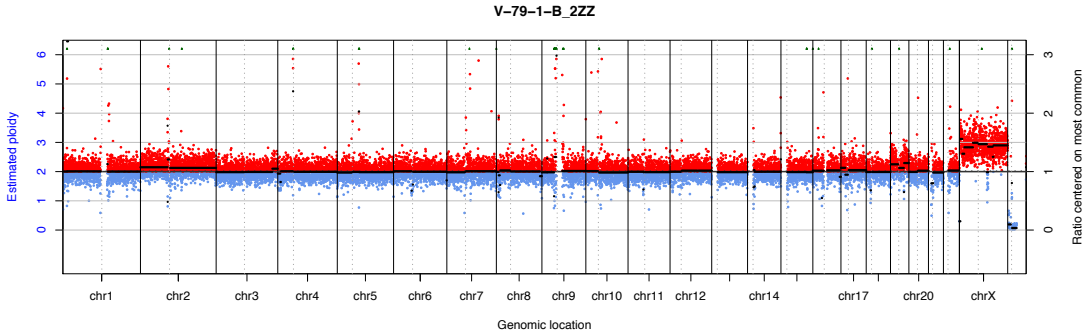

V−73−1_0NA
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

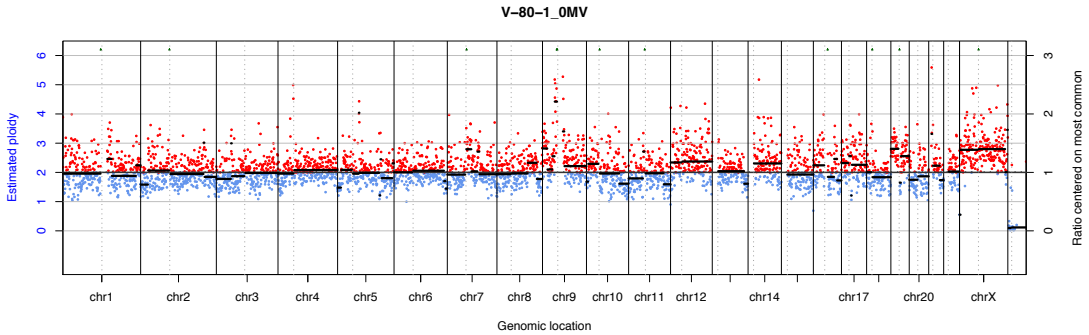

V−74−1−A_0MU
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

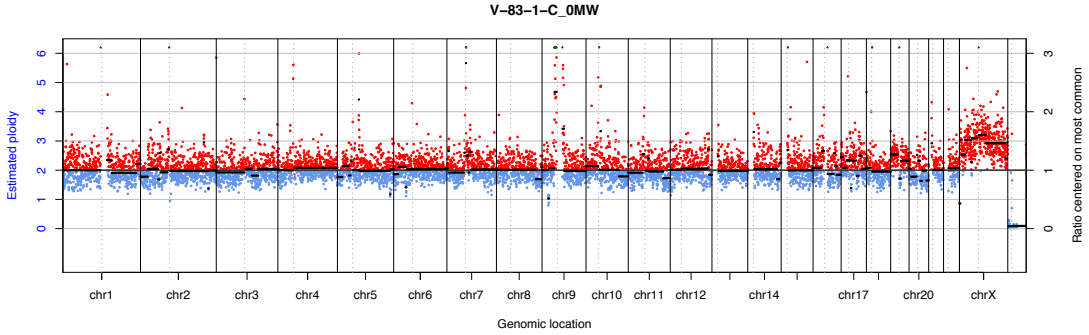

V−75−1_0NH
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

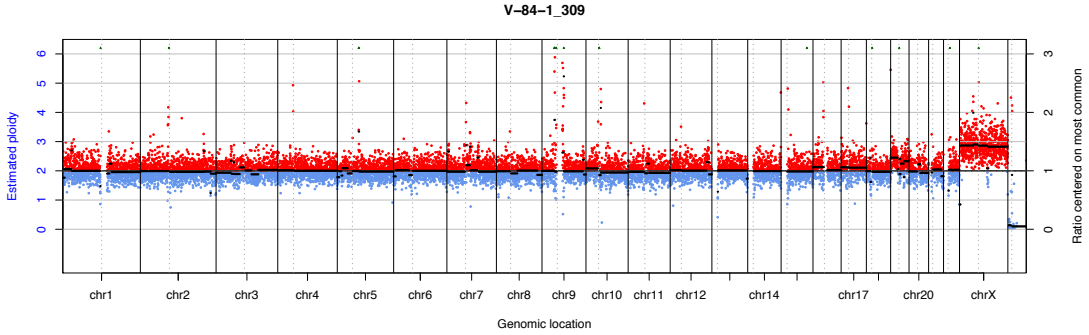

V−77−1_0NJ
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

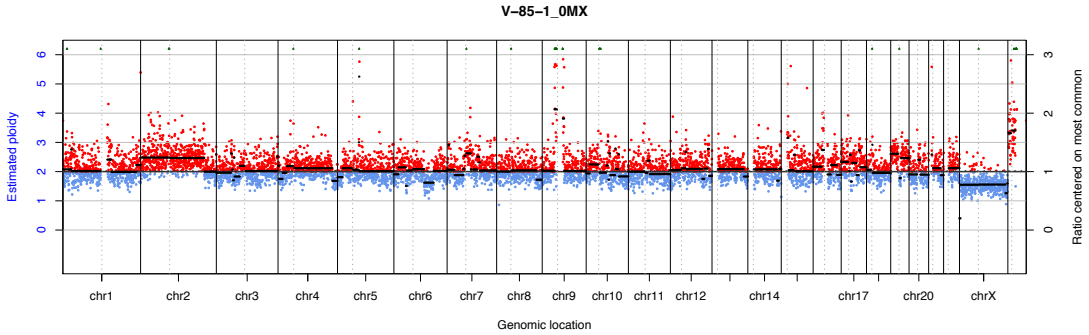

V−78−1−A_308
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

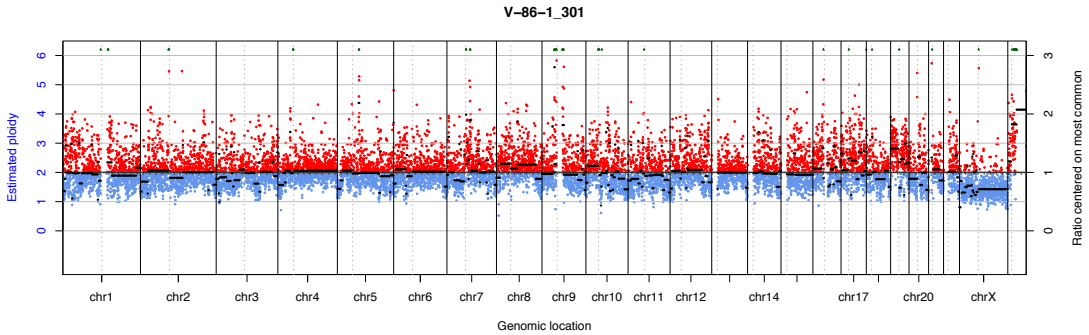

V−79−1−B_2ZZ
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

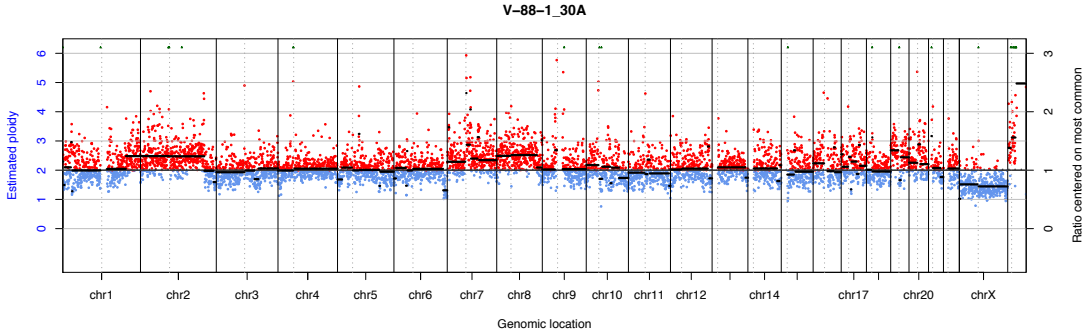

V−80−1_0MV
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

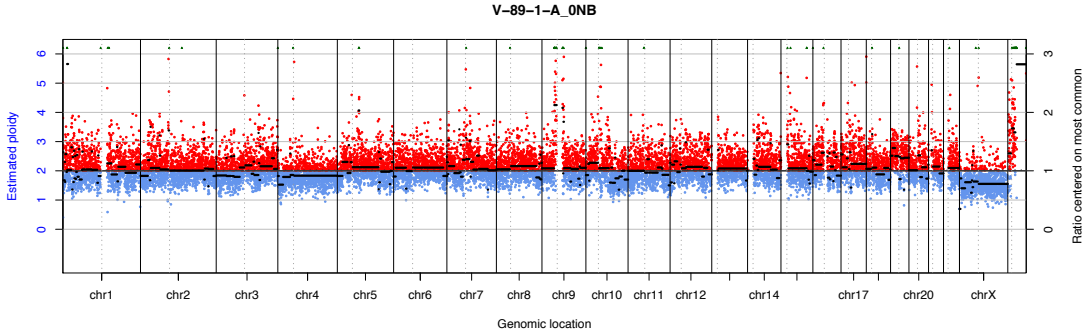

V−83−1−C_0MW
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

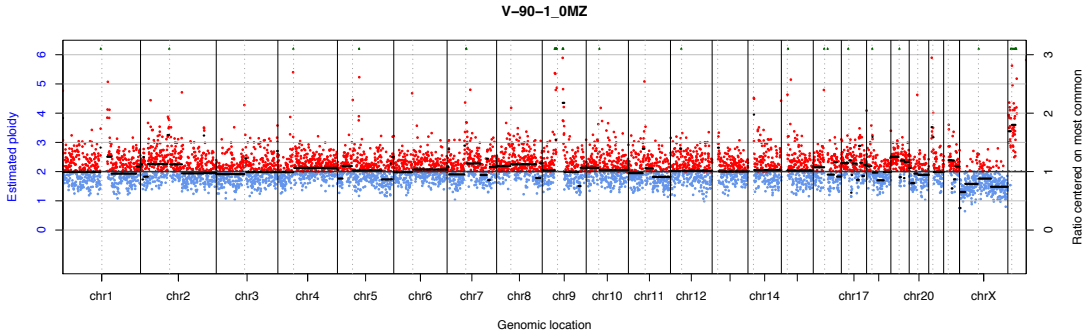

V−84−1_309
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

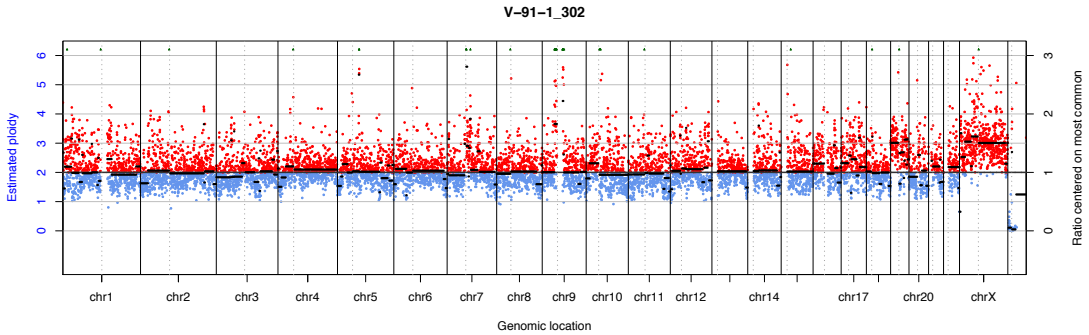

V−85−1_0MX
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

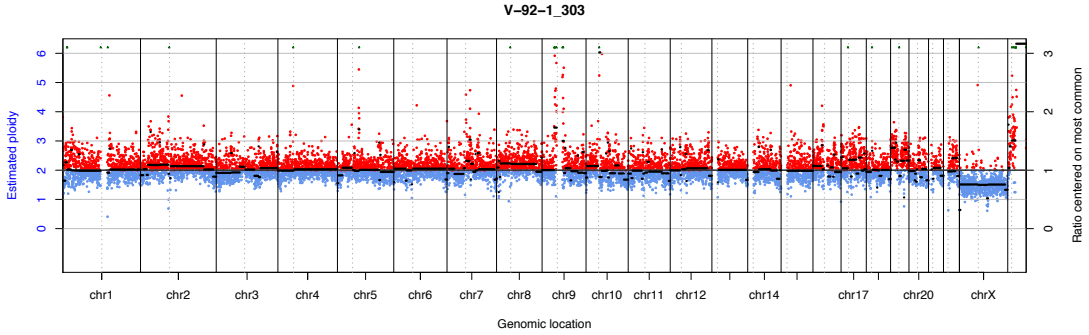

V−86−1_301
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

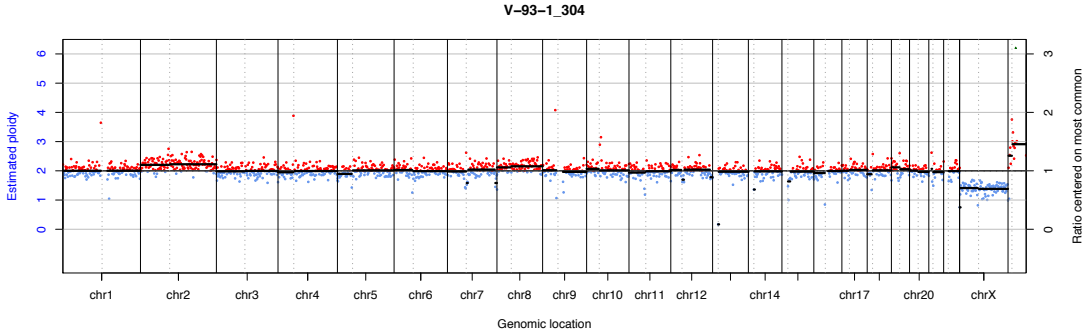

V−88−1_30A
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

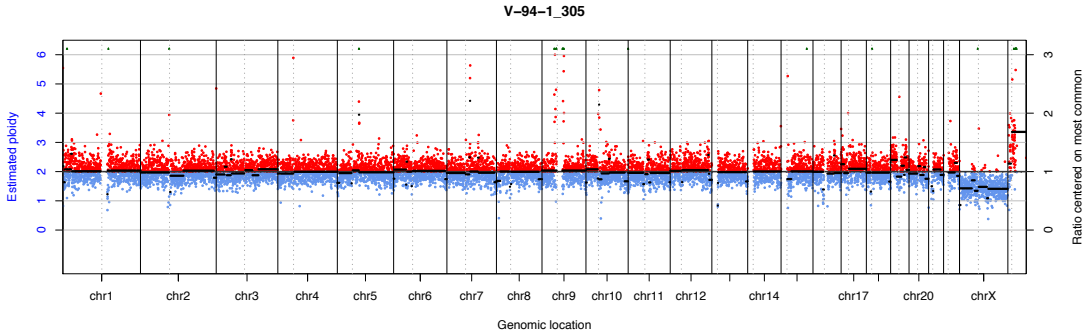

V−89−1−A_0NB
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

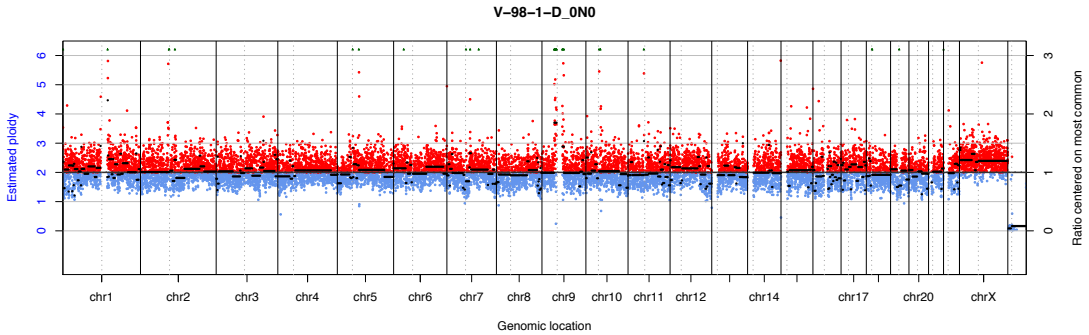

V−90−1_0MZ
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

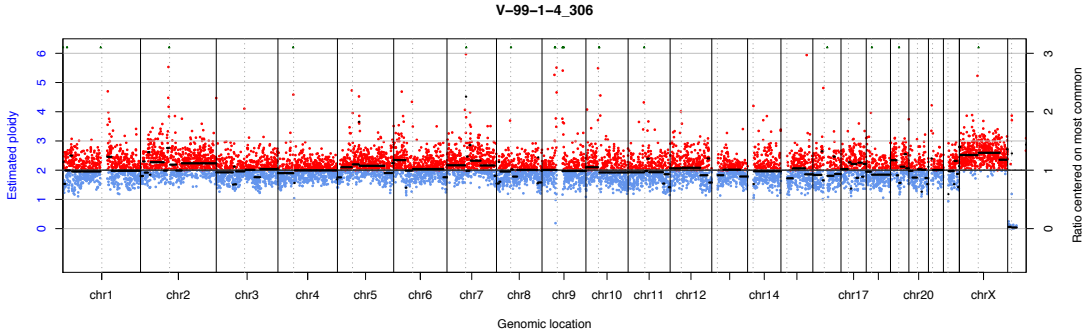

V−91−1_302
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

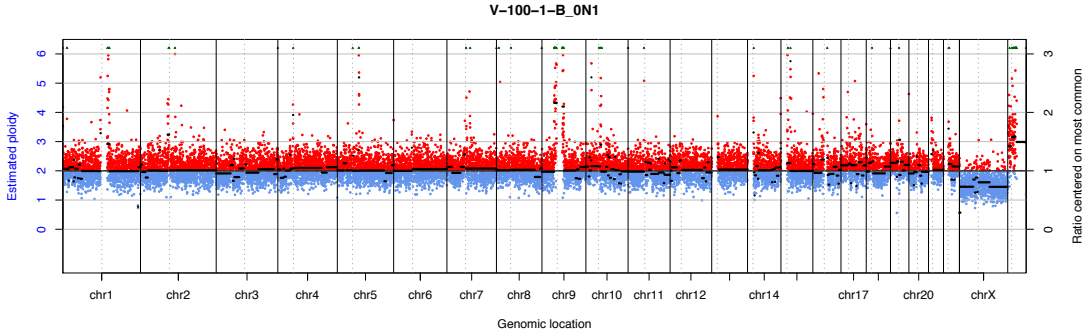

V−92−1_303
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

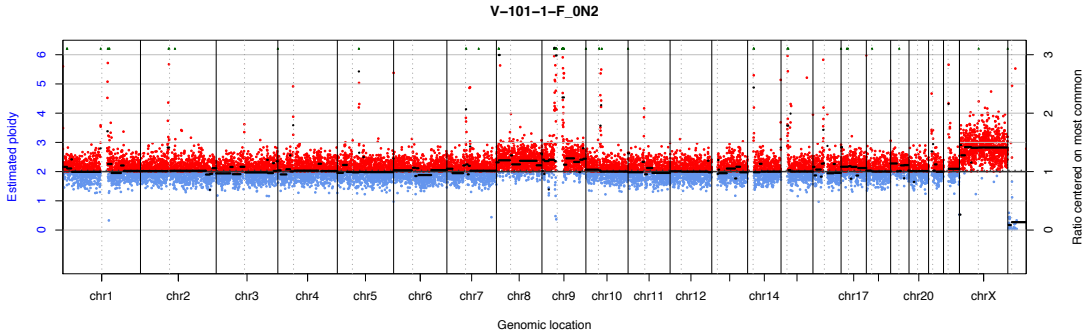

V−93−1_304
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

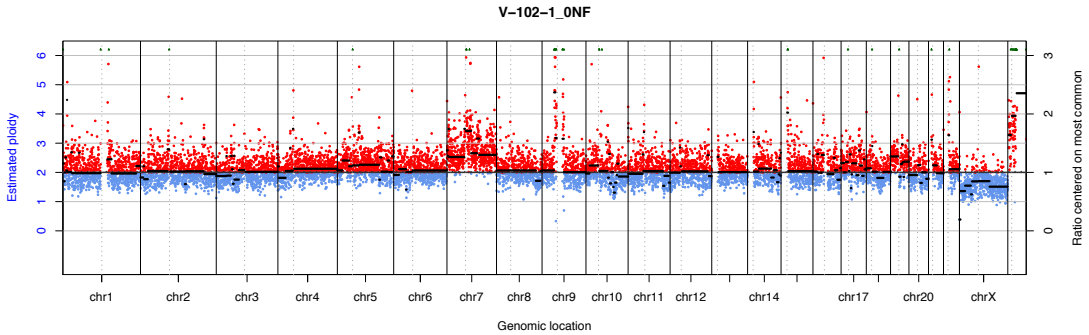

V−94−1_305
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

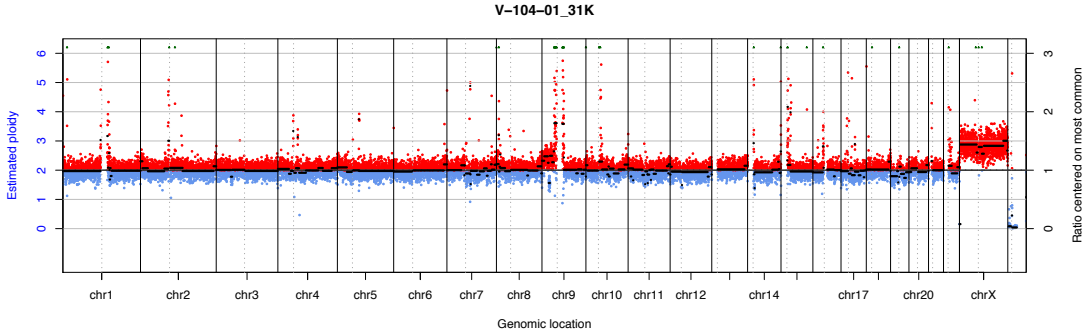

V−98−1−D_0N0
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

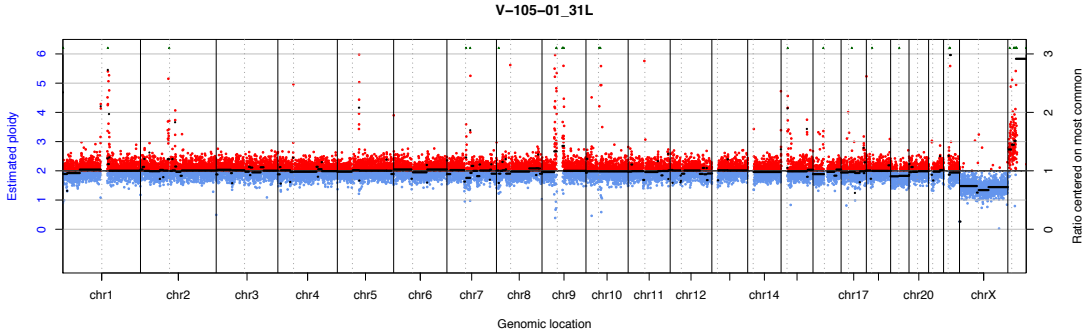

V−99−1−4_306
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

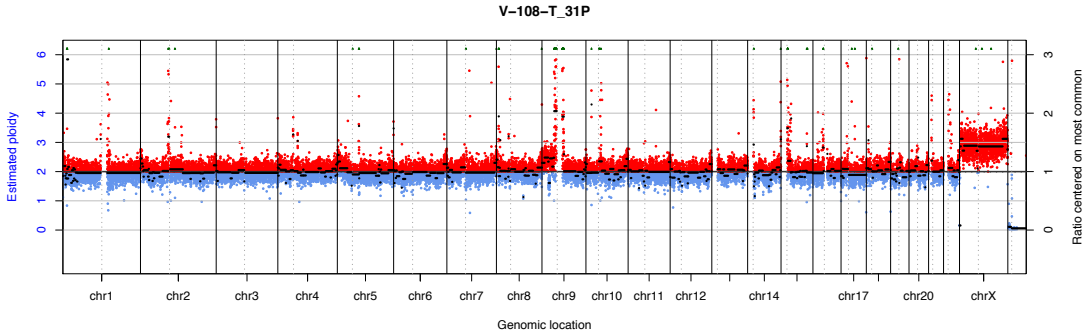

V−100−1−B_0N1
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

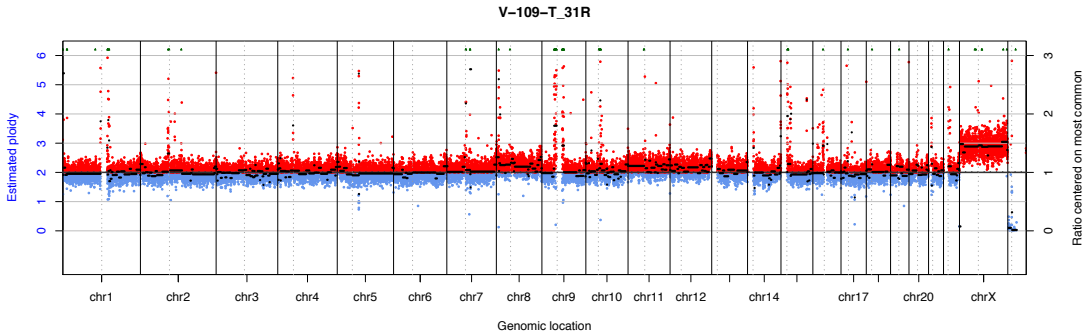

V−101−1−F_0N2
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

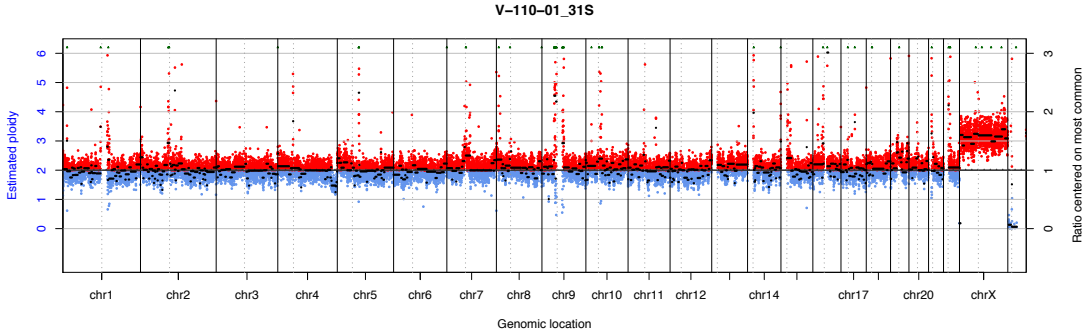

V−102−1_0NF
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

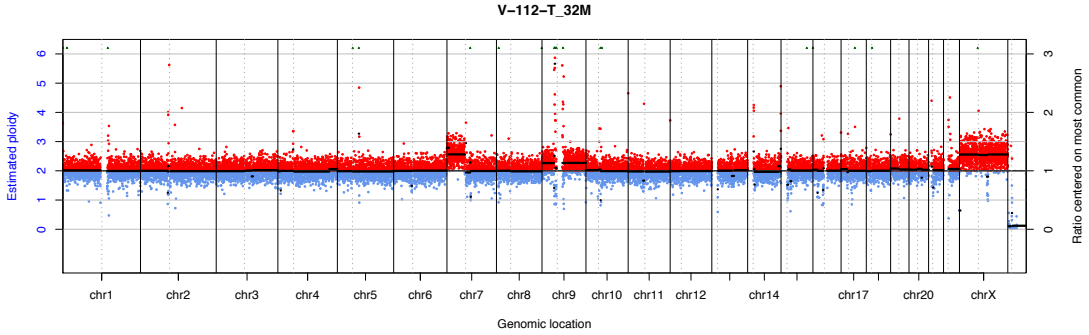

V−104−01_31K
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

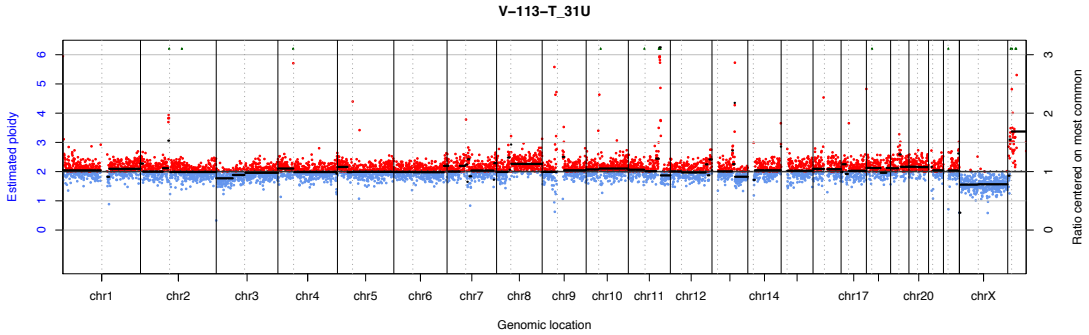

V−105−01_31L
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

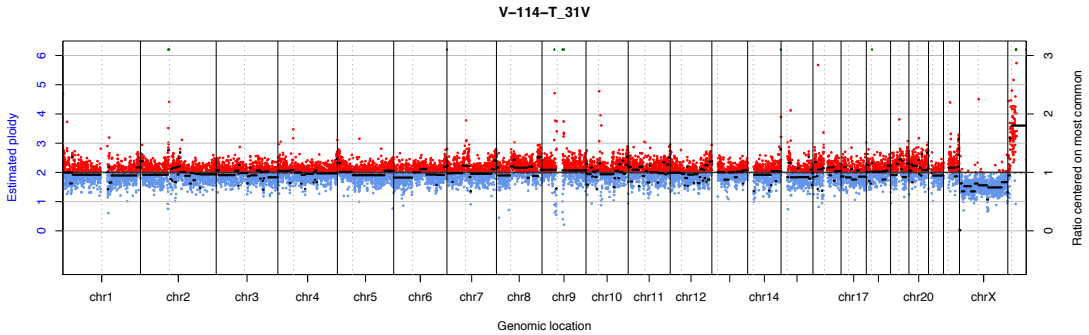

V−108−T_31P
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

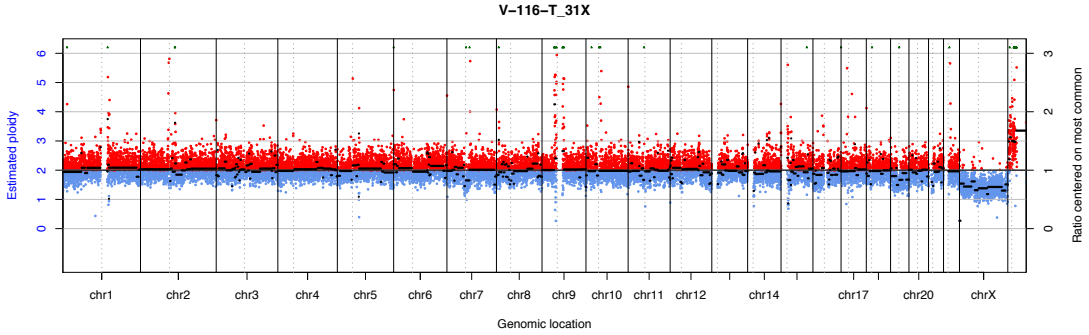

V−109−T_31R
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

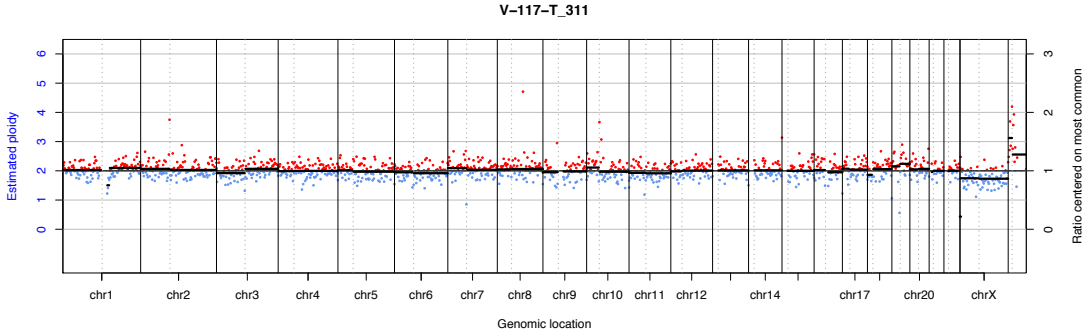

V−110−01_31S
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

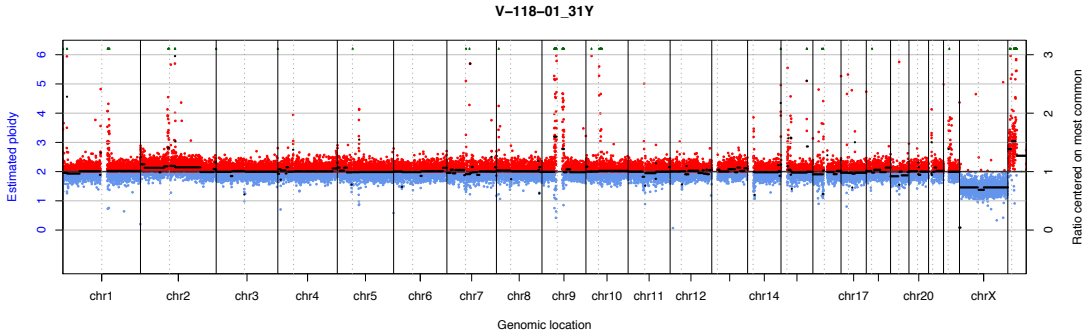

V−112−T_32M
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

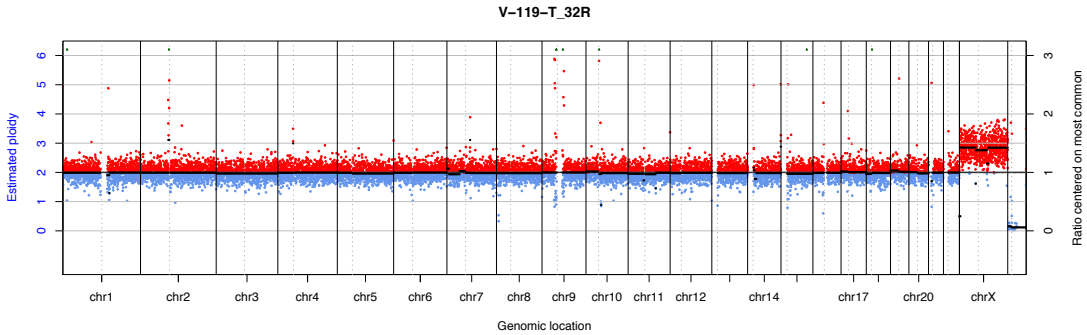

V−113−T_31U
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

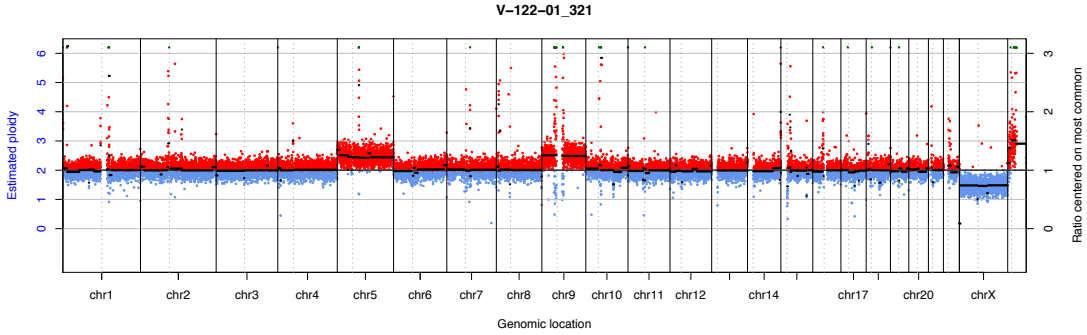

V−114−T_31V
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

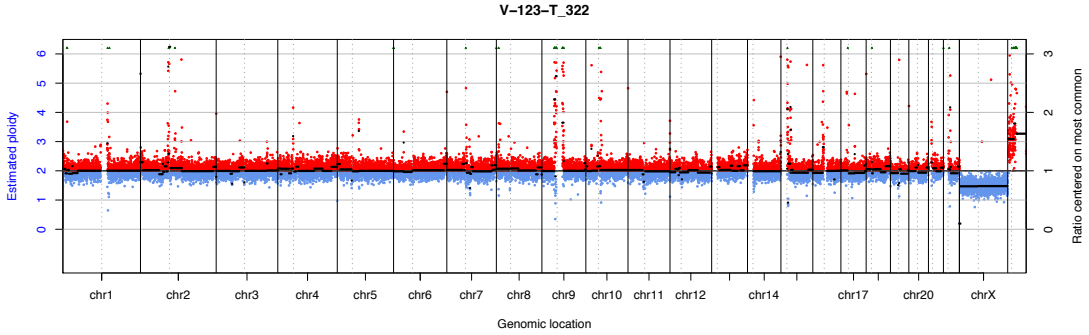

V−116−T_31X
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

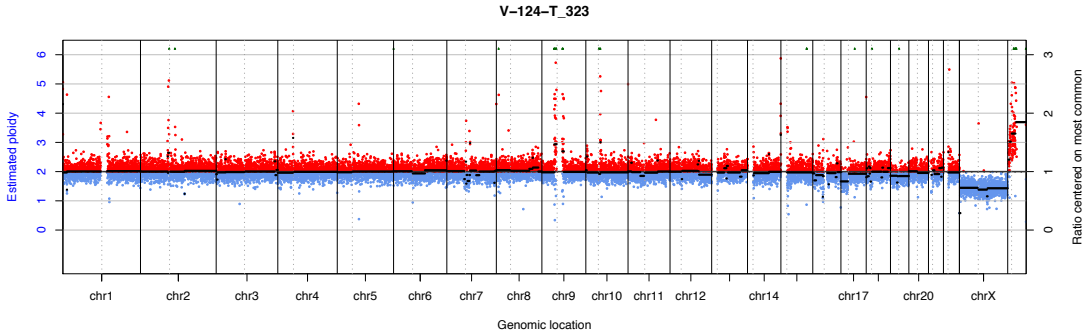

V−117−T_311
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

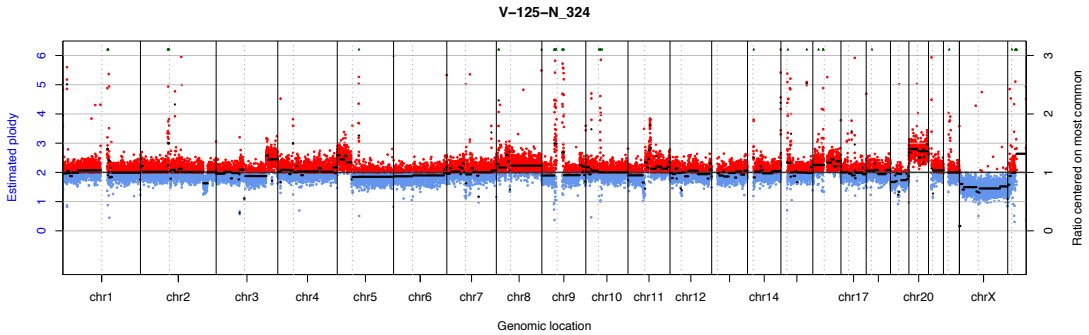

V−118−01_31Y
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location


V−119−T_32R
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location


V−122−01_321
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location


V−123−T_322
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location


V−124−T_323
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

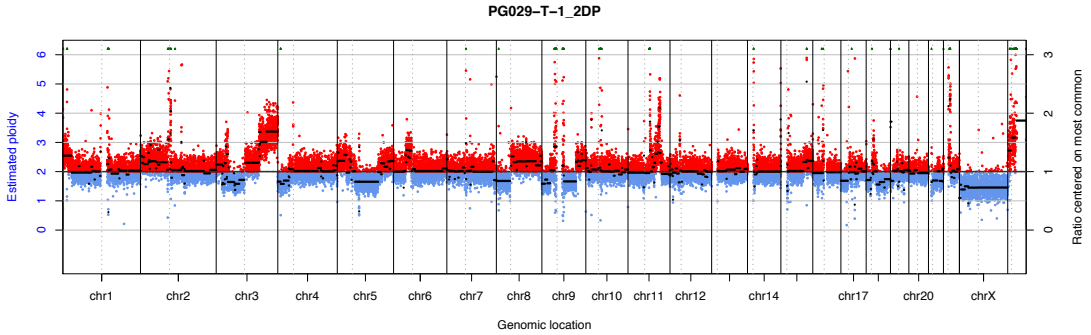

V−125−N_324
Estimated ploidy
Ratio centered on most common
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr14
chr17
chr20
chrX
Genomic location

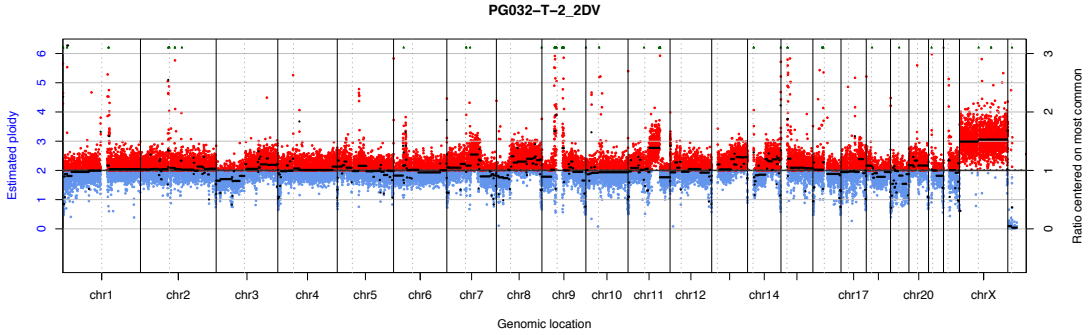
Appendix 4.4 The three study groups for copy number analysis, and OSCC clinicopathological data.

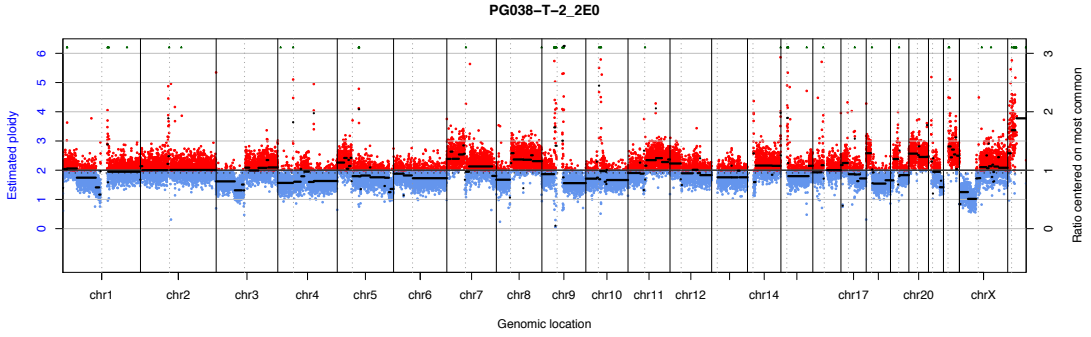
OVH (N: 16)	OVC (N: 57)	OSCC (N: 45)		
Sample ID	**Sample ID**	**Sample ID**	**Sex**	**Age**
V-95-1-B	V-04-01-E9	PG001-T-1-EL1-DNA	Male	72
V-02-02-C1	V-07-01-A	PG004-T-3-EL1-DNA	Male	80
V-03-01-A1	V-10-01-1 e1	PG013-T-1-DNA	Male	57
V-015-01-A3	V-014-01-5	PG015-T-1-DNA	Female	53
V-022-01-4	V-019-01-I4	PG016-T-1-DNA	Male	63
V-025-01-D2	V-20-3	PG017-T-1-DNA	Male	75
V-029-02-A3	V-026-01-A4	PG018-T-1-DNA	Male	72
V-30-1	V-060-01	PG021-T-1-DNA	Female	53
V-031-01-B1	V-61-1-4	PG023-T-1-DNA	Female	52
V-033-01-4	V-62-1-B	PG027-T-1-DNA	Female	53
V-40-1	V-63-1-2	PG028-T-1-DNA	Male	84
V-041-01-G	V-65-1-D1	PG029-T-1-DNA	Male	54
V-42-1-G	V-66-1-B	PG031-T-1-DNA	Male	60
V-44-1-3J	V-67-1-B1	PG032-T-2-DNA	Female	60
V-46-1-2	V-68-1	PG036-T-1-DNA	Male	53
V-106-01	V-69-1-D	PG038-T-2-DNA	Male	52
	V-70-1-B	PG049-T-1-DNA	Female	56
	V-71-1-5	PG059-T-1-DNA	Female	73
	V-72-1-4	PG063-T-1-DNA	Female	57
	V-73-1	PG071-T-1-DNA	Male	63
	V-74-1-A	PG079-T-1-DNA	Male	54
	V-75-1	PG083-T-1-DNA	Male	52
	V-77-1	PG104-T-2-DNA	Male	61
	V-78-1-A	PG105-T-1-DNA	Female	73
	V-79-1-B	PG108-T-1-DNA	Male	72
	V-80-1	PG118-T-1-DNA	Male	57
	V-83-1-C	PG121-T-5-DNA	Male	54
	V-84-1	PG122-T-1-DNA	Male	43
	V-85-1	PG123-T-3-DNA	Female	53
	V-86-1	PG129-T-1-DNA	Male	71
	V-88-1	PG130-T-1-DNA	Female	47
	V-89-1-A	PG133-T-1-DNA	Male	65
	V-90-1	PG135-T-1-DNA	Male	66
	V-91-1	PG137-T-1-DNA	Male	52
	V-92-1	PG139-T-4-DNA	Male	68
	V-93-1	PG141-T-1-DNA	Male	46
	V-94-1	PG142-T-1-DNA	Male	71

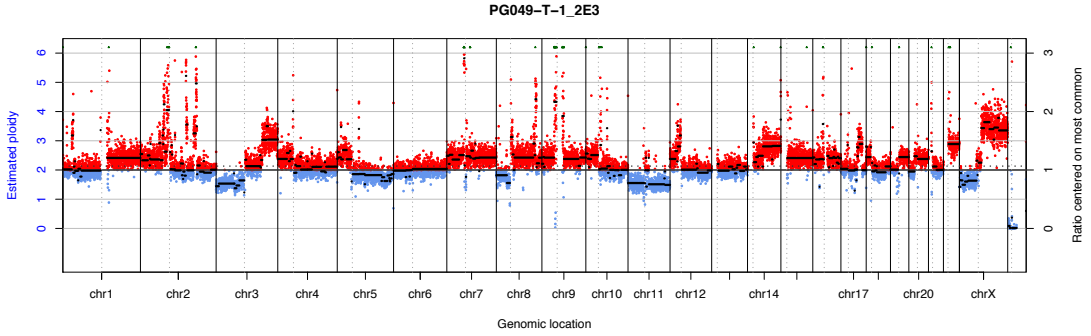
	V-98-1-D	PG144-T-1-DNA	Male	91
	V-99-1-4	PG156-T-1-DNA	Male	62
	V-100-1-B	PG187-T-1-DNA	Male	40
	V-101-1-F	PG192-T-1-DNA	Male	74
	V-102-1	PG195-T-2-DNA	Male	69
	V-105-01	PG197-T-1-DNA	Male	56
	V-108-01	PG198-T-1-DNA	Male	54
	V-109-T	PG200-T-1-DNA	Male	68
	V-110-01			
	V-112-T			
	V-113-T			
	V-114-T			
	V-116-T			
	V-117-T			
	V-118-01			
	V-119-T			
	V-122-01			
	V-123-T			
	V-124-T			
	V-125-T			

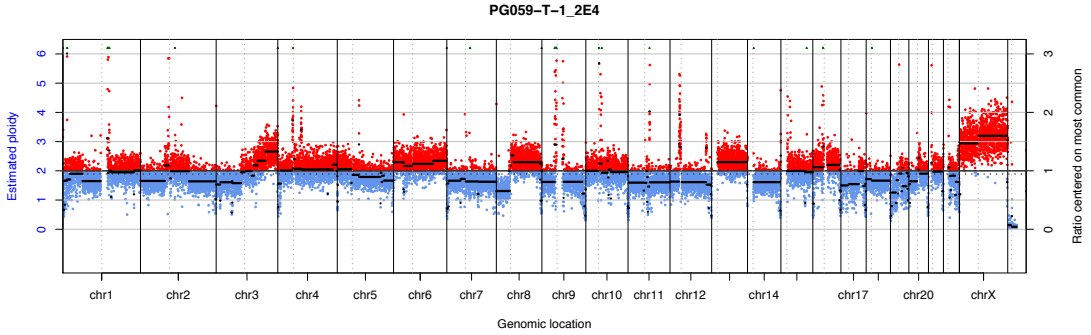
Appendix 4.5 OSCC CN keryograms

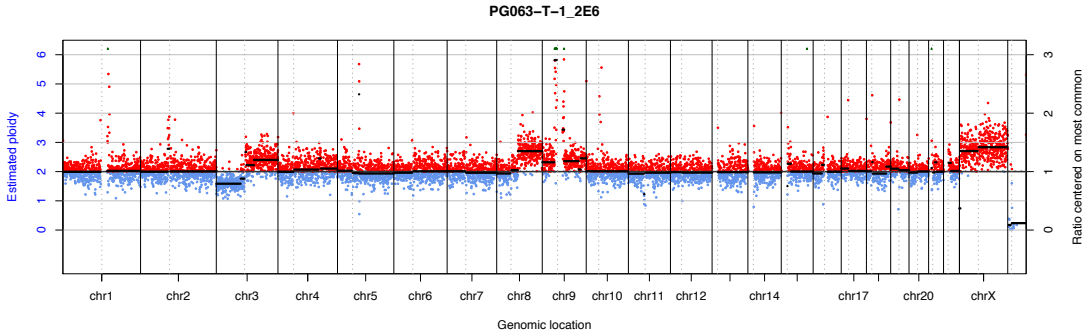


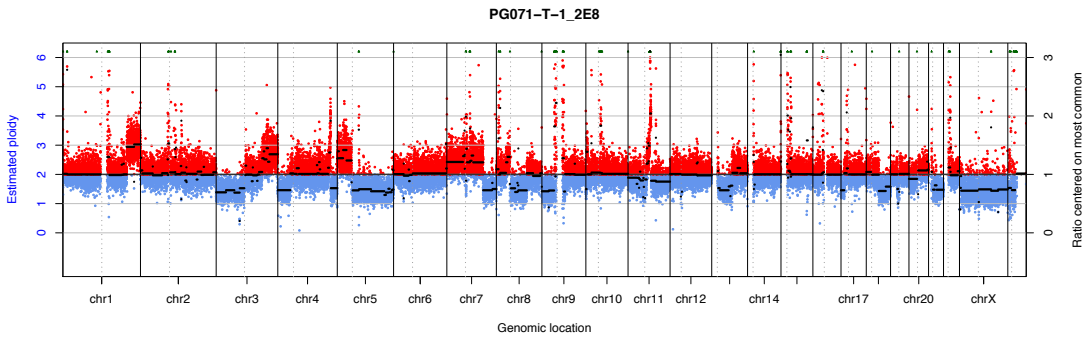


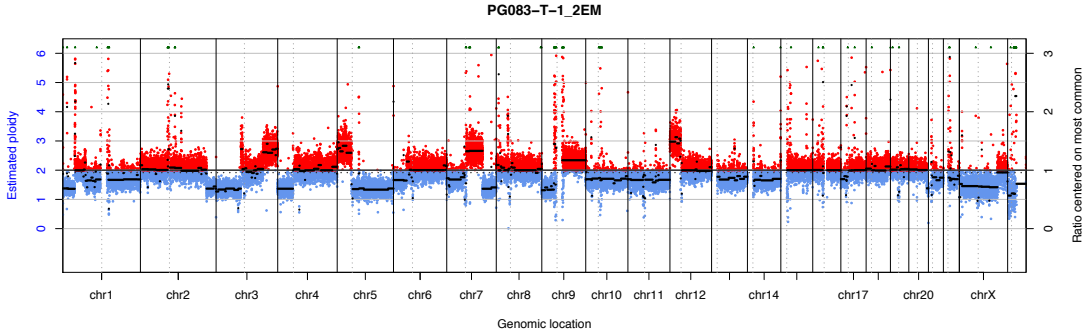




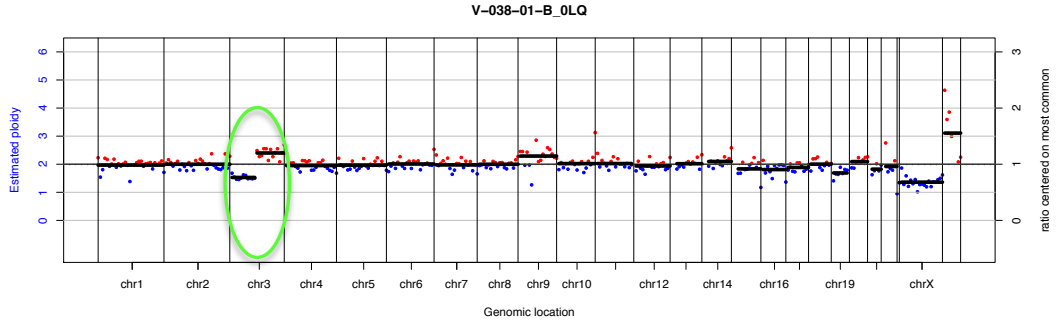

PG059−T−1_2E4
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

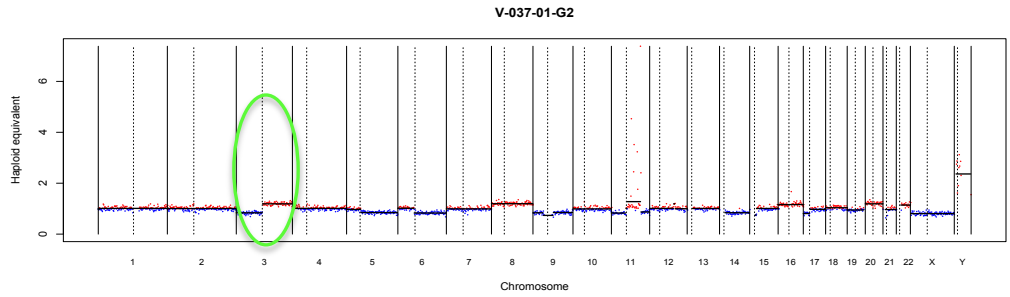

PG063−T−1_2E6
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

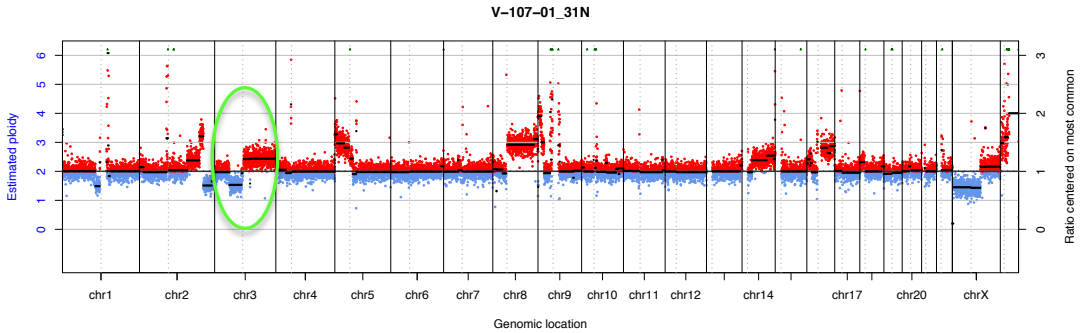

PG071−T−1_2E8
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

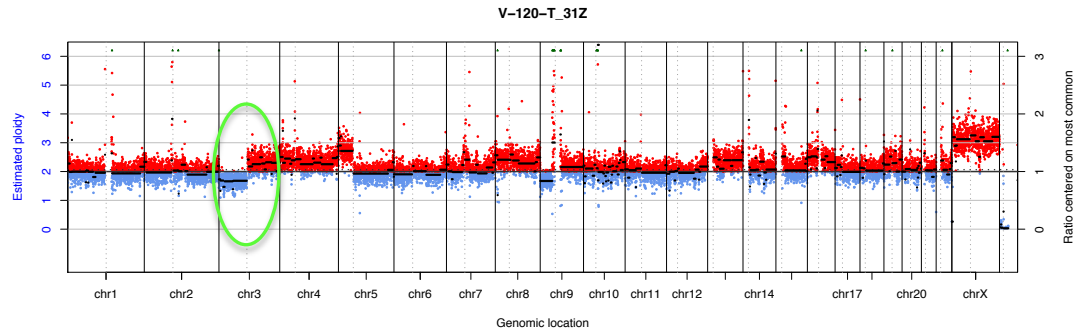

PG083−T−1_2EM
Estimated ploidy
Ratio centered on most common
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr14 chr17 chr20 chrX
Genomic location

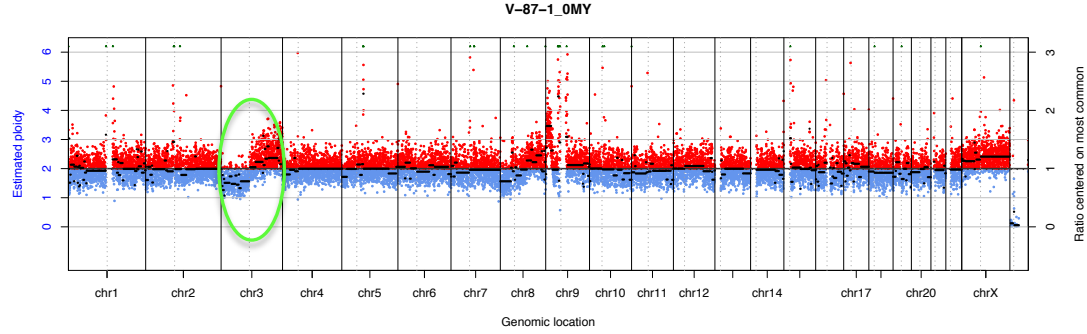
Appendix 4.6 Copy number karyograms for misrepresented OVC cases.





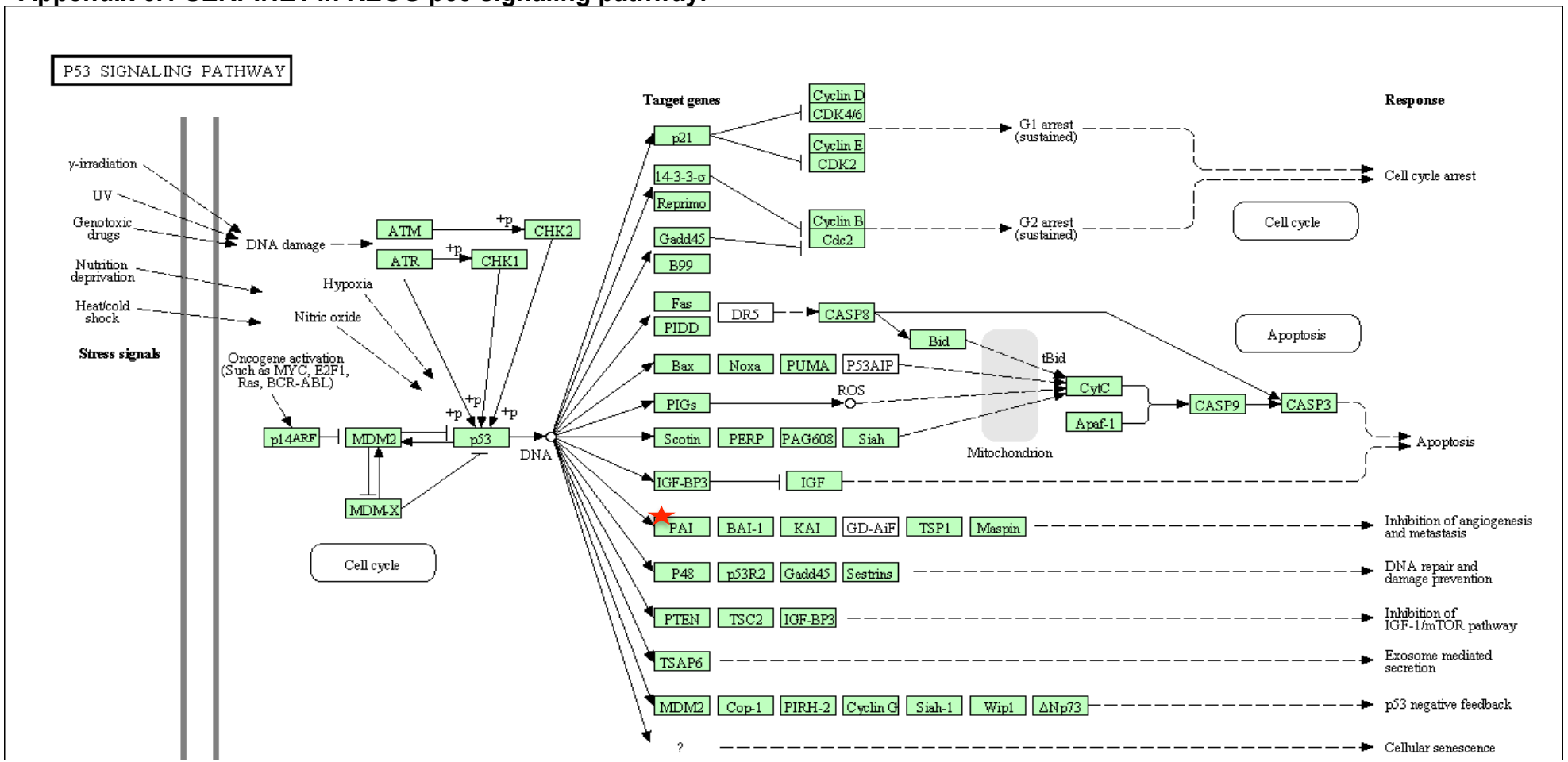






Chapter 6 Appendices

Appendix 6.1 SERPINE1 in KEGG p53 signaling pathway.



Chapter 8 Appendices

Appendix 8.1 Implications of mutated and differentially expressed genes in this project.

Genes	Genes status	NGS technique	Suggested impact in the biology of OVC	Reason
PAX9 *TGM3* *CRNN* *ADH7* *SERPINE1*	Significantly down-regulated in OVC Significantly down-regulated in OVC Significantly down-regulated in OVC Significantly down-regulated in OVC Significantly overexpressed in OVC	RNAseq	Genes that could have a role in the development of OVC	Expression status of those genes in OVC cohort matched with their expression status in other HNCs studies. (Gerber et al., 2002), (Chen et al., 2000), (Chen et al., 2013), (Hashibe et al., 2008), (Chin et al., 2005).
KRT76 *KRT2*	Significantly overexpressed in OVC Significantly overexpressed in OVC	RNAseq	Genes that could have a role in the development of OVC	Keratins are expressed continuously in epithelial tumour cells, which are also a characteristic of their origin site. Both genes were previously down-regulated in OSCC (Zain et al., 2010), (Ambatipudi et al., 2013).
DLG2 *HBB* *DSC1*	Significantly overexpressed in OVC Significantly overexpressed in OVC Significantly overexpressed in OVC	RNAseq	Genes that could be a reason behind the indolent behavior of OVC	Down-regulation or loss of those genes has been reported previously to be associated with the development or progression of HNCs (Onda et al., 2005), (Reshmi et al., 2007), (Teh et al., 2011).
LAMC2 *SMR3B* *TNFRS12* *IGFBP6* *FSTL3*	Significantly overexpressed in OSCC Significantly overexpressed in OSCC Significantly overexpressed in OSCC Significantly overexpressed in OSCC Significantly overexpressed in OSCC	RNAseq	Genes that could be a reason behind the indolent behavior of OVC	Overexpression of those genes has been previously reported in HNCs (Patel et al., 2002), (Lindberg et al., 2006), (Pyeon et al., 2007), (Cacalano et al., 2008), (Freier et al., 2010), (Hassan et al., 2011), (Ambatipudi et al., 2012).
HLF *HBB* *INPP5f*	Significantly overexpressed in OVC Significantly overexpressed in OVC Significantly overexpressed in OVC	RNAseq	Overexpression of those TSGs in OVC could explain its benign behavior	These genes were suggested as TSGs in different cancers (Woenckhaus et al., 2006), (Onda et al., 2005), (Ribarska et al., 2012).
CDH3 *HPGD*	Significantly overexpressed in OVC Significantly overexpressed in OVC	RNAseq	Overexpression of both genes in OVC could be a reason behind its non-metastatic feature	Down-regulation of both genes has previously been found in lymph node metastasises (Mendez et al., 2009), (Kawamata et al., 2003).
CXCL5 *THBS1* *MT2A*	Significantly overexpressed in OSCC Significantly overexpressed in OSCC Significantly overexpressed in OSCC	RNAseq	Absence of overexpression of those genes in OVC could explain its non-	Genes were previously reported to be overexpressed and involved in cancers metastasis (Miyazaki et al., 2006), (Mendez et

MT1X	Significantly overexpressed in OSCC		metastatic feature	al., 2009), (Skubitz et al., 2012).
HLF *TGFBI* *SERPINE1* *MMP1* *INHBA* *COL4A2* *COL4A1* *ADAMTS12*	Significantly overexpressed in OVC Overexpressed in OSCC Overexpressed in OSCC Overexpressed in OSCC Overexpressed in OSCC Overexpressed in OSCC Overexpressed in OSCC Overexpressed in OSCC	RNAseq	Possible biomarkers to differentiate between OVC and OSCC tumours.	Were all proposed to differentiate between OVC and OSCC in Wang et al study. All genes were overexpressed in OSCC samples versus OVCs in Wang et al study except for *HLF* gene, and the same expressions status were also obtained here from the generated RNAseq results.
SERPINE1 *CDH3* *PLA2G4D*	Significantly overexpressed in OVC Significantly overexpressed in OVC Significantly overexpressed in OVC	RNAseq	Possible biomarkers	Have been also identified within significant chromosomal gain regions in the CN analysis data (chapter four).
DSPP *MUC4* *NEFH* *ANP32E*	Mutated in 40% of OVC cohort Mutated in 30% of OVC cohort Mutated in 30% of OVC cohort Mutated in 30% of OVC cohort	Exome sequencing	Mutated genes could have a role in the development of OVC	All the four genes were described in previous cancer studies.
TP53 *CDKN2A* *NOTCH1* *NOTCH2* *FAT1*	Mutated in 70% of OSCC cohort Mutated in 35% of OSCC cohort Mutated in 10% of OSCC cohort Mutated in 25% of OSCC cohort Mutated in 20% of OSCC cohort	Exome sequencing	The lack of any mutations within those TSGs in the ten OVC cases could be behind the benign behavior of this tumour.	The five genes are significant HNSCC cancer genes (Agrawal et al., 2011), (Stransky et al., 2011).