THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Evaluating cellular microRNA expression in human papillomavirus associated head and neck squamous cell carcinoma

Dr. Sarah Louise Emmett

Bachelor of Science (UQ)

Bachelor of Medicine, Bachelor of Surgery (UQ, 2012)

A thesis submitted for the degree of Master of Philosophy at

The University of Queensland in 2017

Faculty of Medicine

Abstract

Squamous cell carcinoma of mucosal head and neck sites (HNSCC) is the sixth most common cause of cancer world-wide, and is associated with a significant morbidity and mortality despite advances in therapeutic options over the past decades. HNSCC includes tumours of the oral cavity, oropharynx, larynx and hypopharynx. While there has been a decline in tumours occurring as a result of exposure to the traditional risk factors of tobacco smoking and alcohol consumption, a significant increase has been observed in the number of tumours attributed to the Human Papillomavirus (HPV). This has primarily been associated with causation of tumours in the oropharynx, however, a smaller number of tumours at other subsites are also likely to be related to the virus. Although HPV-related tumours tend to have an improved prognosis and a better response to chemo- and radiotherapy, the pathobiology underlying these differences is not yet fully understood.

MiRNAs are short, non-coding single strands of RNA which regulate gene expression at the post-transcriptional level. They have been identified to be dysregulated and function as tumour suppressor or oncogenes in a number of human cancers, including head and neck cancers. MiRNAs have also been associated with prognostication including associations with recurrence of disease and disease-free survival. As they are readily available across a range of body tissues including solid tissue, saliva, and plasma, these biomarkers have potential utility in relation to screening for disease diagnosis, recurrence or progression. Although there is an established difference between HPV-positive and negative tumours in the oropharynx clinically, the majority of studies investigating miRNA expression in HNSCC have not taken this factor into account. In the small number of studies which have examined the effect of HPV on miRNA expression in HNSCC, there is a lack of consistency in the results described.

This thesis aimed to assess the difference in miRNA expression between HPV-positive and negative tumours, through use of miRNA microarray followed by validation with real-time PCR using TaqMan® assays combined with a sensitive detection method (Fluidigm HD) in a larger cohort of patients. It also investigated differences in miRNA expression in association with lifestyle factors, demographics and clinical outcomes. A set of seven miRNAs was identified to be differentially expressed between HPV-positive and negative tumours in the oropharynx. No difference in miRNA expression was observed between the two tumour cohorts at other subsites in the head and neck. Differences in miRNA expression were also observed with lifestyle factors such as smoking status and alcohol consumption. Recurrence of the primary tumour and overall disease-free survival were further associated with altered miRNA expression. The findings outlined in this thesis add to the present literature regarding differences in miRNA expression between HPV-positive and negative

tumours in the oropharynx, and provide a basis for further research to focus on the use of miRNAs for diagnosis or prognosis in HNSCC.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Conference abstracts:
Neville Davis Prize 2016 – Oral presentation
ASOHNS 2017 – Accepted for poster presentation

Publications included in this thesis

No publications included.

Contributions by others to the thesis

Dr. Mitchell Stark (Translational Research Institute) kindly lent his expertise in the use of the Fluidigm plates for RT-PCR, as well as assisting with the data analysis obtained from both the microarray and the Fluidigm data.

Dr. Leesa Wockner and Dr. Rebecca Leeks (QIMR Berghofer Medical Research Institute) kindly supervised the statistical analysis in this thesis.

Associate Professor Lutz Krause and Dr. Harald Oley kindly lent their expertise for supervision of the bioinformatics section of this thesis.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I would like to thank my supervisors Professor Ben Panizza, Dr. Annika Antonsson and Professor David Whiteman for their generosity, guidance, patience and support during the completion of this thesis.

I would also like to very gratefully acknowledge the contribution of Dr. Mitchell Stark (Translational Research Institute) for his support with use of the Fluidigm plates and data analysis; Dr. Leesa Wockner and Dr. Rebecca Leeks (Statistics, QIMR Berghofer Medical Research Institute) for their assistance with the statistical analysis of the data; and Dr. Harald Oley and Associate Professor Lutz Krause (Department/Institute) for their guidance in relation to bioinformatics.

The Cancer Control laboratory at the QIMR Berghofer Medical Research Institute also were invaluable in their assistance in providing me with the necessary laboratory bench space, instruments, general reagents and supplies. In particular, I would like to thank Mrs. Lani Knight and Mr. Blake Ferguson for their help with general laboratory issues. I would also like to thank Dr. Glen Boyle, Dr. Antonia Pritchard and Dr. Nicole Cloonan (QIMR Berghofer Medical Research Institute) for their kind reviews and directives.

This project would not have occurred without significant financial support, and therefore I would like to sincerely thank the Queensland Centre of Excellence for research in Head and Neck cancer (NHMRC) and the Princess Alexandra Hospital Research Foundation, without whom completion of this thesis would not have been possible.

Finally, I would like to acknowledge and thank my friends, family and colleagues who have supported me during the completion of this thesis.

Keywords

Human papillomavirus, HPV, Squamous cell carcinoma, SCC, oropharynx, miRNA

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANSRC Code: 060109 Proteomics and Intermolecular Interactions, 30%

ANZSRC Code: 060102 Bioinformatics, 30%

ANSRC Code: 111299 Oncology and Carcinogenesis not elsewhere classified, 40%

Field of Research (FoR) Classification

FoR Code 1199, Medicine and Health Sciences, not otherwise specified, 100%

TABLE OF CONTENTS

List of Tables

Chapter One

Chapter Three

Chapter Four

Appendix One

miRNAs which are differentially expressed; one of which was down-regulated in patients without a recurrence, and ten of which were up-regulated.

List of Figures

Chapter One

Chapter Three

Appendix One

List of Abbreviations

HNSCC	Head and neck mucosal squamous cell carcinoma
SCC	Squamous cell carcinoma
HPV	Human papillomavirus
miRNA	MicroRNA
RNA	Ribonucleic acid
PCR	Polymerase chain reaction
CT	Computed tomography
MRI	Magnetic resonance imaging
PET	Positive emission tomography
FDG	Fluoro-2-deoxyglucose
AJCC	American Joint Committee on Cancer
TNM	Tumour Node Metastasis staging system
DNA	Deoxyribonucleic acid
mRNA	mature RNA
CLL	Chronic lymphocytic leukaemia
FFPE	Formalin fixed paraffin embedded
HREC	Human research ethics committee
RFS	Recurrence free survival
DFS	Disease free survival
TGCA	The cancer genome atlas

Please note that all microRNAs referred to in this thesis are human microRNAs, unless otherwise specified.

CHAPTER ONE: Introduction

1.1 Mucosal Head and Neck Cancer

1.1.1 Epidemiology

Squamous cell carcinoma (SCC) of the head and neck cancer (HNSCC) is the sixth most common cause of cancer worldwide [1, 2]. In Australia, it affects approximately 4000 persons per year (3.4% of all cancer diagnoses per year in Australia), which has increased over the past two decades [3, 4]. This includes SCC of mucosal anatomic sub-sites: oral cavity (tongue, buccal mucosa), oropharynx (base of tongue, palatine tonsil), hypopharynx (posterior pharyngeal wall, post-cricoid region) and larynx (glottis, supraglottis, subglottis). Mortality rates remain high despite advances in conventional therapy, with an approximately 68% 5-year survival rate in Australia at present [3]. Further, significant morbidity is associated not only with diagnosis but also with treatments. This may range from physical disfigurement and low self-esteem to radiation induced swallowing difficulties and xerostomia, to difficulty with speech and overall reduced quality of life. The majority of HNSCCs continue to be diagnosed at a later stage, resulting in more extensive disease, and contributing to the overall low survival rate [4]. Traditionally, tobacco smoking and alcohol consumption are risk factors for this disease, having synergistic carcinogenic effects, and tumours related to these risk factors have demonstrated a stabilization or decline in incidence [1, 2, 5, 6]. Human papillomavirus (HPV) infection is also now an established risk factor for a subset of HNSCC, particularly HNSCC occurring in the oropharynx (base of tongue and palatine tonsils), following its discovery in these tumours in 1983 by Syrjanen *et al* [1, 7-9]. Since this time, there has been a clear and significant increase in the incidence of oropharyngeal tumours related to HPV of 225% between 1984 and 2004 in America, as well as a decrease of approximately 50% of HPV-negative tumours [10].

1.1.2 Staging

A tissue diagnosis remains the gold standard for a diagnosis of HNSCC, and can include either a biopsy of the primary tumour site or a fine needle aspirate of a neck lymph node. Clinical examination of the head and neck, in combination with imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI) and positive emission tomography (PET) with fluoro-2-deoxyglucose (FDG) are used to accurately stage patients. Mucosal HNSCC is staged according to the 7th American Joint Committee on Cancer (AJCC) TNM classification [11]. The T stage is used to refer to the extent of the primary tumour, while the N stage indicates whether there is any involvement of loco-regional lymph nodes. The M stage refers to the presence or absence of distant metastatic disease. Accurate staging of patients at diagnosis allows for suitable treatment regimens to be instituted by a multidisciplinary team and permits appropriate prognostication.

1.2 Human papillomavirus associated head and neck cancer

HPV has been well studied in the context of cervical carcinoma in women, but the role of HPV as an etiologic agent in HNSCC was first postulated in 1983 by Syrjanen *et al* [9, 12]. Since this time, the oncogenic role of HPV-16 in oropharyngeal SCC has been established by the International Agency for Research on Cancer [7]. In oropharyngeal SCCs, HPV has an estimated prevalence of 50%-60%, but HPV DNA has been detected in up to 90% of oropharyngeal tumors [2, 5]. When the presence of HPV-DNA is combined with $p16^{INK4a}$ positivity or E6/E7 mRNA, the attributable fraction of HPV in the oropharynx is approximately 40% according to a meta-analysis in 2014; however, at present clinically, up to 70% of patients presenting with oropharyngeal SCCs have HPV-associated tumours [2].

1.2.1 Clinical and Pathologic features

Patients with HPV-positive disease in the oropharynx clinically have distinctly different disease to HPV-negative tumors. These tumors tend to occur in younger (40-55 years, compared with >60 years), male, Caucasian patients, as well as be associated with higher levels of education achieved, higher income and higher socio-economic status [5, 8, 13, 14]. The patients tend to have lower levels of tobacco and alcohol consumption [1, 8]. The tumors are more likely to be of non-keratinizing or basaloid morphology histologically, and while they present at an earlier T-stage, there is an association with advanced nodal disease at diagnosis [5, 14]. An improved responsiveness to chemotherapy and radiotherapy has been reported, as well as an overall improved survival rate [5, 13, 14].

1.2.2 Role of HPV in non-oropharyngeal tumours

While high-risk HPV has been isolated from HNSCC at sites outside the oropharynx, its role in carcinogenesis is much less certain. High-risk HPV is potentially associated with the development of a small subset of oral and laryngeal SCCs [5, 15], although the probability of a HPV-attributable cancer at these sites is at least five times lower than of SCCs occurring in the oropharynx [13]. A number of studies have investigated the prevalence of HPV at other HNSCC sites (excluding oropharynx), with widely varying results, potentially as a result of different geography, genetic interactions and differing methodology employed [14-16].

A meta-analysis in 2014 has identified a HPV-DNA prevalence of 24% in the oral cavity, and when this is combined with $p16^{INK4a}$ status and E6/E7 mRNA the attributable fractions of disease are 7% and 16% respectively [2]. Similarly for laryngeal and hypopharyngeal SCC (combined), a HPV-DNA

prevalence of 22% was identified, with attributable fractions of 19% when combined with $p16^{INK4a}$ positivity, and 9% when combined with E6/E7 mRNA [2]. These findings were correlated with a systematic review [15]. There is discrepancy between HPV-DNA prevalence and attributable fraction of disease, and the use of E6/7 mRNA analysis to confirm biologically active HPV infection has been utilized [2, 16]. It is clear that the presence of HPV DNA in a tumor does not signify that the tumor has been caused by HPV [2, 5, 15]. This has contributed to the theory that oncogenic HPV infection can sometimes be a 'bystander' infection, without contributing to carcinogenesis in a subset of the population.

1.2.3 Prognosis

Prognostic factors in head and neck cancer include the tumour stage (T4) and nodal stage (N2b-3) at diagnosis, pack year history of smoking, poor performance status, and HPV status of the tumour [6, 17]. Overall, HPV-positive tumours in the oropharynx have an improved survival rate when compared with HPV-negative tumours, regardless of the treatment modality employed: 75% 3-year survival rate compared with 50% in HPV-negative tumours [1, 6, 8]. These patients also demonstrate a lower rate of recurrence (approximately 50% lower than HPV-negative tumours), and a lower rate of metastatic disease [1, 6]. This has been postulated to be due to the patient's overall younger age at diagnosis, improved functional status and lower rates of morbidities usually associated with higher rates of substance abuse [1]. While an underlying molecular mechanism has not been identified yet, this remains possible. To date, no association between the presence of HPV and prognosis has been demonstrated outside of the oropharynx [5, 15]. Unfortunately, approximately 20% of HPV-positive tumours in the oropharynx (those staged as T4 or N3 at diagnosis) do not have a favourable prognosis, and have a higher rate of metastatic and recurrent disease [1, 8, 18, 19].

1.2.4 Treatment

Presently, therapeutic interventions for HNSCC patients include a combination of surgery, radiotherapy, chemotherapy and targeted molecular therapies [6]. Despite the differences observed clinically between tumours associated with traditional risk factors and those related to HPV, clinical guidelines distinguishing treatment strategies between these two patient cohorts do not exist currently. The therapeutic options for each patient are determined by a multi-disciplinary team, which takes into consideration the stage at diagnosis, HPV status of the tumour and tumour site; as well as the patient's functional status and wishes [4]. Consideration of the expected functional outcomes, including the level of organ preservation, and the patient's quality of life is also important. Early-stage (T1-2) tumours are usually treated with single modality therapy, such as surgery or radiotherapy

alone. More locally advanced, recurrent or metastatic tumours are treated with a combination of therapies such as surgical resection followed by post-operative radiotherapy (possibly combined with chemotherapy) or concomitant chemo-radiotherapy primarily; with surgical resection reserved for residual disease [4].

As HPV-related oropharyngeal tumours have an overall improved prognosis, focus is increasingly shifting towards differential therapeutic regimens for HPV-positive and HPV-negative tumours. It has been well established that HPV-positive tumours have an improved responsiveness to chemotherapy and radiotherapy treatment regimens, when compared with other treatment modalities [1, 4, 6, 8]. Surgical excision of small (T1-2) primary oropharyngeal tumours has been shown to have oncologic outcomes equivalent to primary radiotherapy in properly selected patients [20, 21]. Contention exists in the literature regarding the appropriateness of de-intensification of therapy for HPV-positive tumours given their better survival rates, with the intent of minimizing the morbidity associated with treatment [1, 6]. However, it will be increasingly important for clinicians to be able to distinguish the subset of patients with HPV-positive tumours who are candidates for therapeutic de-escalation but will have a poorer prognosis, so that the offered treatments can be appropriately potent.

1.3 Human Papillomavirus

1.3.1 Structure

Human papillomaviruses are non-enveloped, double-stranded DNA viruses of the P*apillomaviridae* family, with a genome of approximately 8000bp [1, 22]. There are over 200 distinct HPV genotypes, which are divided into high- and low-risk groups based on their demonstrated oncogenic potential [8, 16, 23]. Low-risk HPVs include Type 6 and 11, which may cause anogenital warts or laryngeal papillomas respectively [16, 23]. The most common oncogenic HPV is Type 16, followed closely by 18 [8, 16, 23]. Others which have demonstrable ability to cause cancer of the cervix, head and neck, and anogenital regions include Types 31, 33, 35, 45, 51, 52 and 56 [8, 16, 23]. The mucosal HPV genome (Refer to Figure 1) encodes six early proteins (E1-2 and E4-7), two late proteins (L1, L2), and a non-coding control region [16]. Across papillomaviruses, L1-2 and E1-2 are well conserved, containing genes involved in replication and viral packaging; whereas there can be significant variability in the remaining genes, which function in cell cycle entry, viral release and host immune evasion [23].

HPVs are transmitted through mucosal and non-mucosal skin epithelium via small tears which allow the virus access to the basal layer of the epithelium, where cell infection is followed by a period of

genome amplification. Through the E6 and E7 proteins, oncogenic HPVs are able to control entry into the cell cycle, which is necessary for completion of its lifecycle [1, 22, 23]. The virus takes advantage of the cell differentiation and proliferation between the basal and upper layers of the epithelium in order to achieve this.

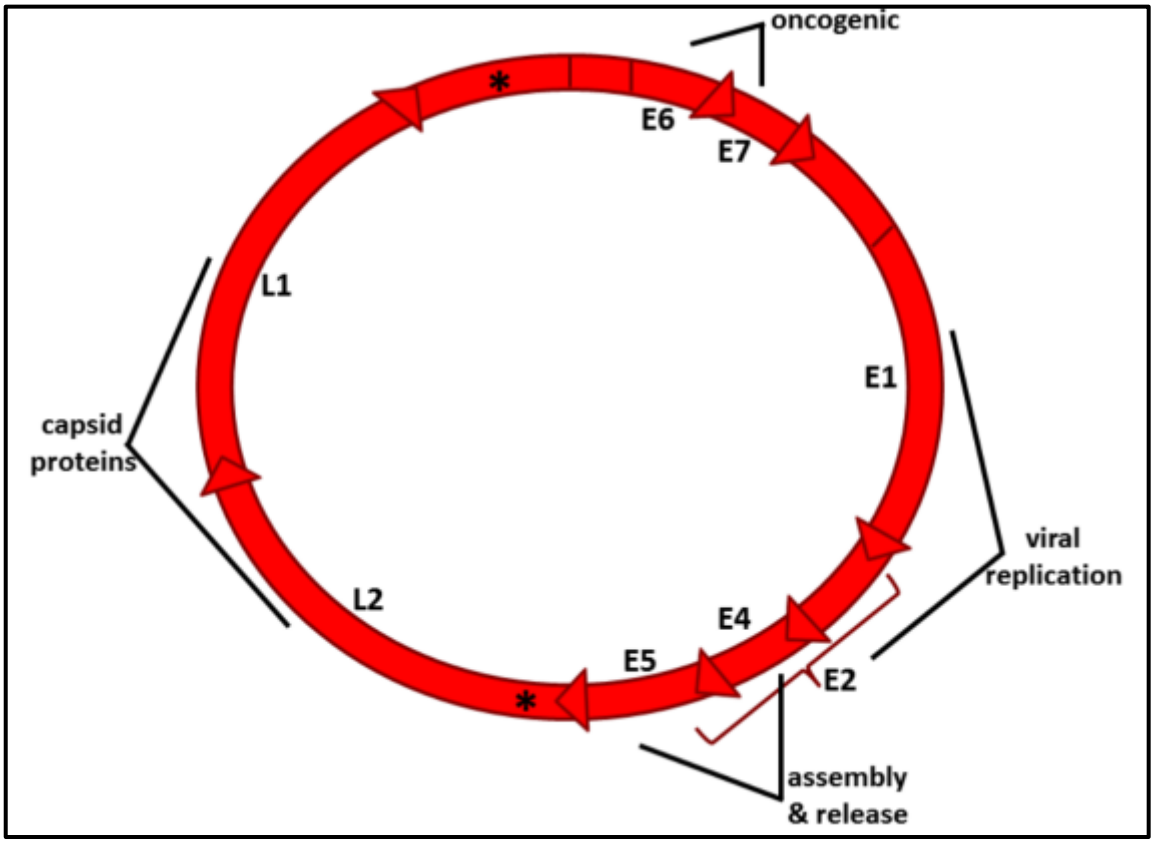


Figure 1.1: Schematic version of the Human Papillomavirus genome (Adapted from [24])

1.3.2 Natural history

HPV has been extensively studied in the context of cervical carcinoma in women, and its transmission in the genitals by sexual contact has been firmly established [1, 12]. The natural history of HPV infection in healthy oral mucosa is the subject of ongoing research. In a healthy Australian population, the estimated point prevalence of high-risk HPV was 2.3% (7/307 participants), and the majority of infections were HPV-18 [25]. The presence of HPV was significantly associated with male gender, and sexual behaviors such as increased number of oral sexual partners [25]. The HIM study [26] has demonstrated an incidence of 1.7% of oncogenic HPV infections in healthy oral mucosa; and Gillison *et al* also corroborated these results, finding an overall prevalence of HPV infection in the oral cavity of 6.9%, with a higher prevalence in men than women [27].

There is increasing evidence that changing sexual behaviors is associated with the rising incidence of HPV-associated HNSCC, and subsequently for oral-oral and genital-oral transmission routes [1, 5]. For example, oral HPV infection is strongly correlated with sexual practices including ever having had oral sex, increased lifetime number of sexual partners, French kissing, history of casual sex and younger age at sexual debut [1, 5, 27]. Additionally, both patients with cervical carcinomas and their partners have been identified to have a higher risk of developing a future head and neck cancer [1, 28].

Given that HPV infection persistence is necessary in the development of cervical cancer, it is likely that this is important for HPV-mediated carcinogenesis in the oropharynx as well. The majority of oral HPV infections clear spontaneously, with 66% clearing within 12 months and 90% clearing within 24 months [1, 26, 27]. The natural history of HPV infection in the oral cavity therefore appears to be of shorter duration than in the cervix, and it is clear that oral HPV infection is frequently cleared, and does not always result in malignancy [5, 26]. However, the presence of an oral HPV infection has been associated with a five- to fifty-fold increased risk of developing a future HNSCC [1, 27]. Unanswered question still remain regarding which patients with oncogenic oral HPV infections will go on to develop a cancer, and why. While the fraction of patients that will develop HPV-associated HNSCC is small, it remains feasible that prophylactic HPV vaccination may prevent the majority or all of these cancers from developing if given prior to viral exposure.

1.3.3 Oncogenic pathways of HPV oncoproteins

HPV encoded E6 and E7 proteins are the dominant oncoproteins, and act via their effects on the p53 and pRB cellular pathways respectively (Refer to Figure 2). p53 is a well-studied transcription factor which induces apoptosis or cell-cycle arrest in response to cellular damage, and is mutated in many human cancers, with resultant carcinogenic effects. HPV-E6 binds to E6AP (E6-associated protein), and this complex ubiquitinates p53, resulting in its degradation [16]. This then results in prevention of cell-cycle arrest and/or apoptosis [16]. E6 also interacts with molecules involved in the extrinsic apoptosis pathway, such as Fas-associated Death Domain (FADD), resulting in their degradation, and as a result prevents apoptosis [1]. pRB binds the transcription factor E2F, to control progression of the cell-cycle through the G_1-S checkpoint [16]. HPV-E7 interacts with and destabilizes pRB, causing release of E2F and subsequent uncontrolled progression through the cell cycle. Inactivation of pRB also indirectly results in over-expression of $p16^{INK4a}$, a commonly used marker in clinical practice for the presence of oncogenic HPV [5, 16].

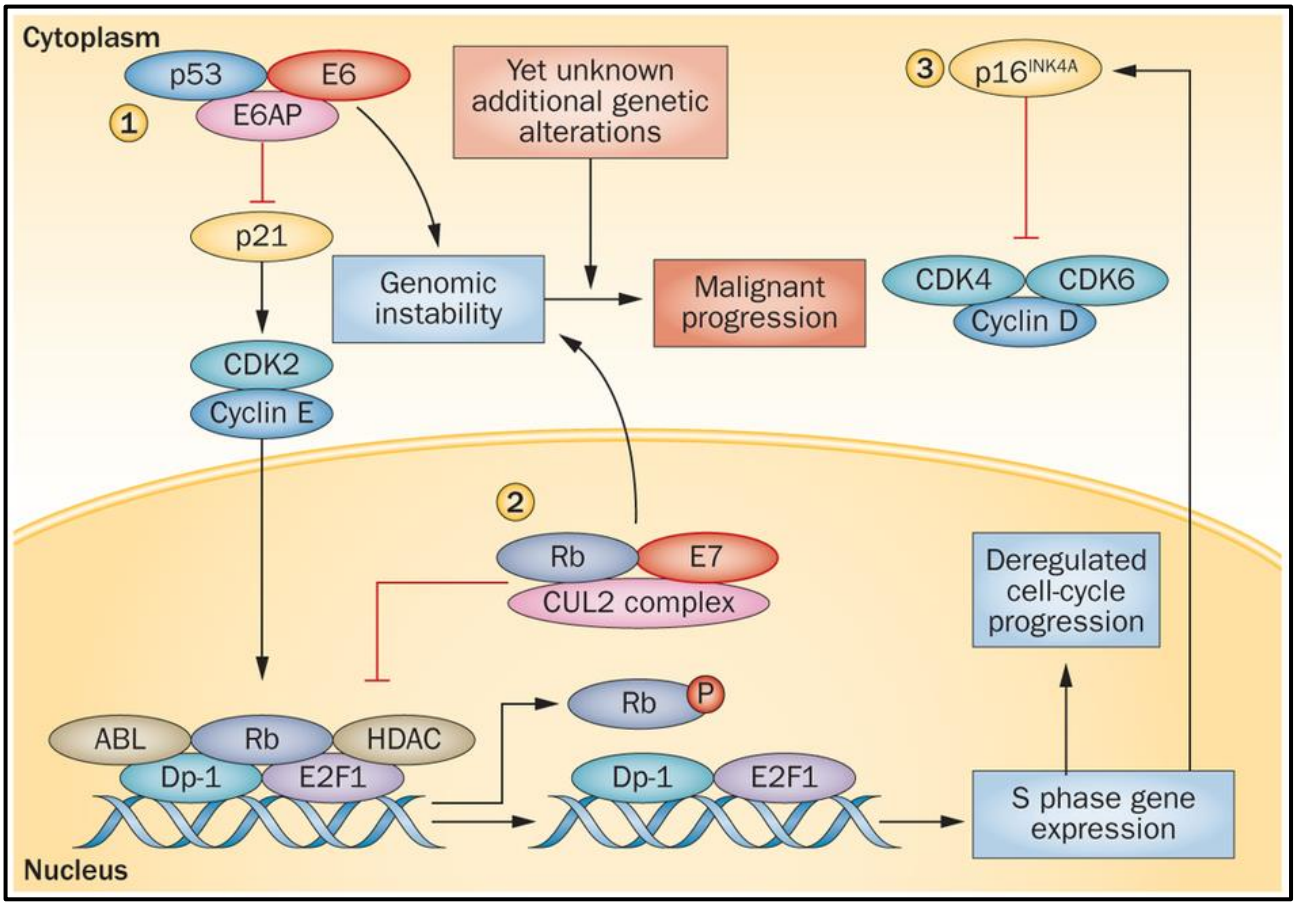


Figure 1.2: Schematic version of E6 and E7 interactions with TP53 and pRB (Adapted from [29])

1.4 MicroRNAs

1.4.1 Introduction

MiRNAs were first discovered in *Caenorhabditis Elegans* where lin-4 and let-7 were found to control timing of larval development [30], and since have been discovered in both plant and animal cells as highly-specific tissue biomarkers [31]. They are short (18-25 nucleotide), regulatory, non-coding single strands of RNA, which have diverse roles in cellular functions, including: development, differentiation, proliferation, inflammation, immune responses, stress responses, and apoptosis [32, 33]. MiRNAs were initially linked to human cancers because of their proximity to chromosomal breakpoints or areas of instability, and dysregulated expression levels in several human cancers [31]. Altered expression of specific miRNAs have been associated with initiation or prognosis in a number of human cancers, and functional studies performed in cancer cell lines, or mouse models, involving over-expression or knock-down of miRNAS, have supported a role for miRNAs in carcinogenesis [31, 34].

1.4.2 Nomenclature

The Faculty of Life Sciences at the University of Manchester, United Kingdom, now hosts and maintains miRBase, a searchable database of all published miRNA sequences and annotations [35-39]. The latest version of this database was released in 2015 (version 21); which contains 1881 precursors and 2588 mature miRNAs. MiRNAs are identified using the prefix 'miR', followed by a specific number. As miRNAs exist in three forms during their biogenesis, the prefix can be slightly altered to reflect the stage of development: primary miRNA transcript ('pri-miR'), stem-loop precursor miRNA transcript ('pre-miR') and mature miRNA ('miR). The species to which the miRNA belongs is designated by three letters which precedes the 'miR' prefix: in *Homo Sapiens* this is 'hsa'.

The numbering of miRNAs is sequential with regards to when they were discovered. Numbered suffixes are utilised to denote distinct precursor sequences which express identical mature sequences: for example, hsa-miR-16-1 and hsa-miR-16-2. Lettered suffixes identify sequences which are closely related and differ by several nucleotides only: for example, hsa-miR-20a and hsa-miR-20b [40]. As miRNA terminology and naming has evolved over the years, each miRNA sequence entry in miRBase also has a unique accession number, which remains constant and allows changes to be tracked in the database, providing the user with full access to the history.

Occasionally there are two miRNA sequences which originate from the same precursor; and sequencing technology has indicated that there is usually a dominant sequence, and a secondary sequence which originate from opposite arms of the precursor miRNA sequence. Previously these have been referred to as miR-142 (dominant) and miR-142* (from opposite arm); however, this has shifted to naming each strand for the arm from which it originated: for example: miR-142-3p (from the 3' arm) and miR-142-5p (from the 5' arm) [35, 39-42].

1.4.3 MicroRNA Biogenesis

MiRNA biogenesis (Refer to Figure 3) begins with nuclear transcription of a large pri-miRNA, transcribed by RNA polymerase II or III [43]. These fold to produce one or more ~80 nucleotide (nt) hairpin structures, consisting of a ~32bp imperfect stem loop and large terminal loop [43]. This is recognized and processed by two enzymes, RNase III, Drosha (RNASEN) and a double-stranded RNA binding protein, Pasha (DGCR8), to form one or more ~60-70nt pre-miRNAs containing ~2nt 3' overhangs [31, 44, 45]. These are then actively transported to the cytoplasm by the RAN GTP dependent transporter Exportin 5 (XPO5), where they are processed by another RNase III enzyme,

Dicer, to yield single-stranded ~22 (16-29) nucleotide structures: the miRNA:miRNA* duplex [31, 43-45]. Dicer removes the terminal base pairs and loop at the opposite end to the cleavage by Drosha, and therefore this defines the second end of the mature miRNA [45].

One strand of the miRNA:miRNA* duplex is then incorporated into a miRNA-containing ribouclеoprotein complex (miRNP), also referred to as a RNA induced silencing complex (RISC), as it is similar in structure and function to the RISC [45]. The miRNA* (antisense strand) is released and degraded [45]. This complex contains one member of the Argonaute protein family (Ago 1-4), which recognizes and binds the miRNA, regardless of its sequence, and facilitates mRNA target recognition [31]. Ago2 is the only argonaute protein in humans which is able to directly cleave target mRNAs with near-perfect complementarity [31].

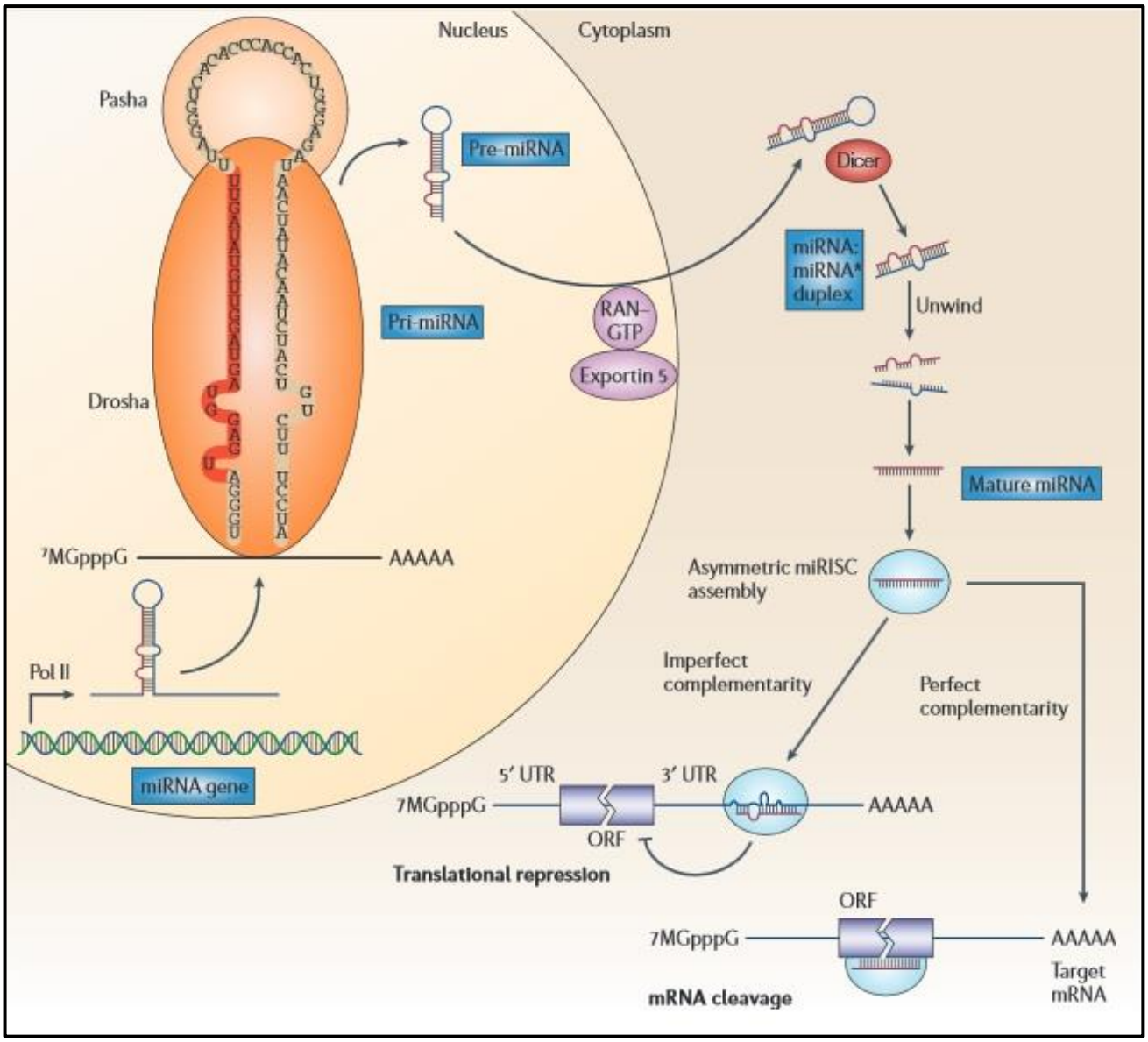


Figure 1.3: Schematic version of miRNA biogenesis (Adapted from [44])

1.4.4 MiRNA::mRNA target interaction

The mRNA target is recognized by base pairing of miRNA nucleotides in positions 2 to 8 (the 'seed' sequence), to complementary sequences located predominantly in the 3' UTR or coding regions of the target mRNA (Refer to Figure 4) [31, 43, 45]. The fate of the targeted mRNA is dependent on the degree of binding complementarity, and can be either mRNA cleavage, or translational repression [43]. In humans and other mammalian cells, this pairing is usually with imperfect complementarity, resulting in translational repression of the mRNA transcript and mRNA destabilization (e.g. de-capping, de-adenylation) [31]. In plants, where there is usually pairing with near-perfect complementarity, the miRNA will specifically cleave the target mRNA [45]. However, the full extent of the mechanistic miRNA::mRNA interactions are not understood at present. There is evidence to support alternative modes of miRNA binding (Refer to Figure 4), and it has been estimated that up to 40% of miRNA target sites are mediated by non-canonical interactions; indicating that there is far more variability in functionally interactive binding sites than previously thought [46].

While the full extent of the mechanistic miRNA interaction with its target mRNA is not understood yet, it is estimated that miRNAs are able to regulate the expression of over 90% of human gene transcripts. Therefore, they have the potential to significantly influence a vast number of biological processes.

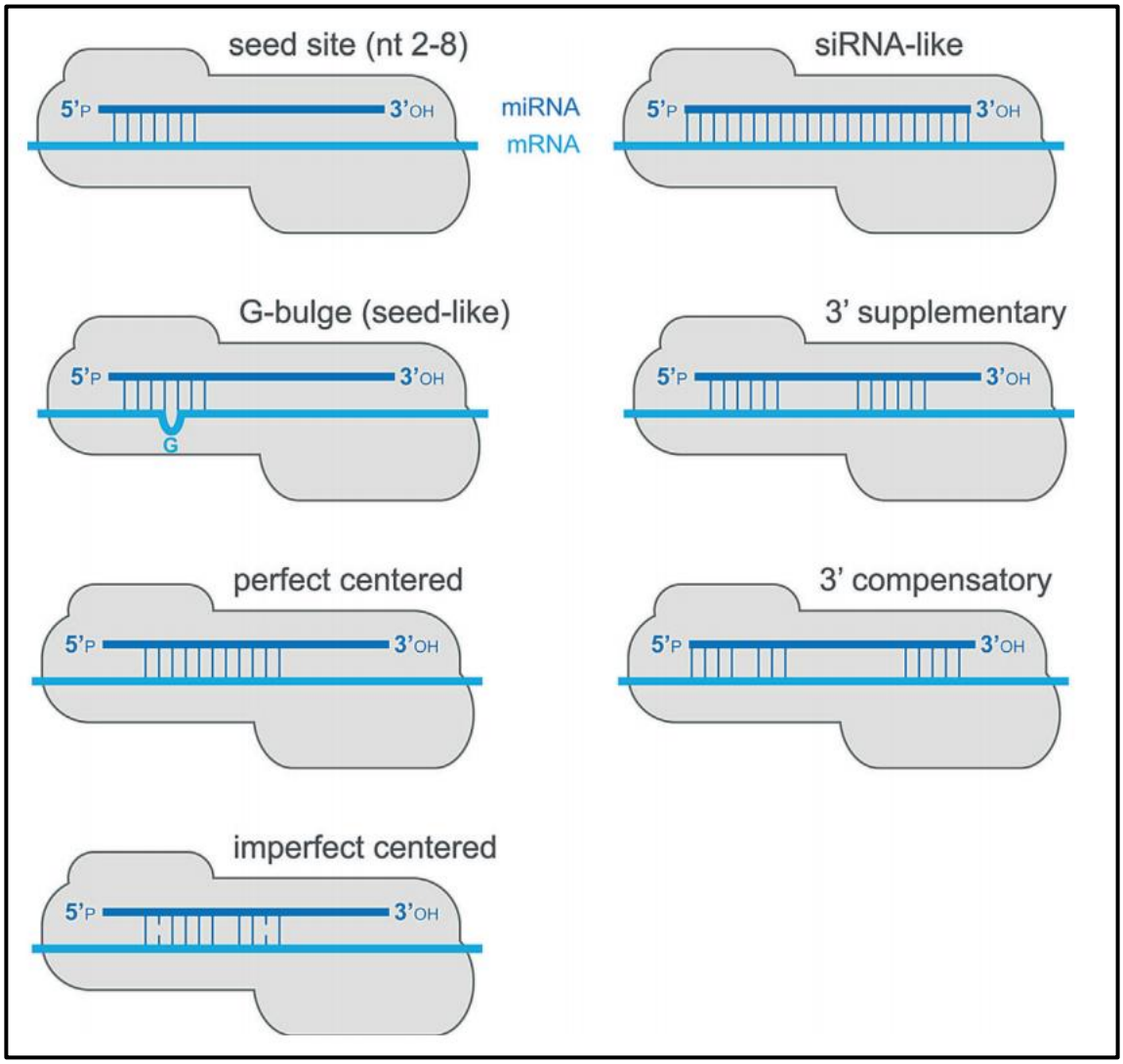


Figure 1.4: Variations of miRNA::mRNA interactions (Adapted from [46])

Given the imperfect complementarity of nucleotide matching between miRNAs and their target mRNA, one miRNA is able to target many mRNAs (some studies estimate this may be up to 200 per miRNA), and in this way miRNAs potentially control an estimated 60% of human mRNAs [44]. It is evident that miRNAs are capable of huge effects on cellular functions, and have the potential to significantly affect a vast number of biological processes. In addition, miRNAs not only target many mRNAs, but also may act cooperatively with other miRNAs to influence a particular cellular pathway. Because miRNAs have such a broad influence on cellular processes, deletion or amplification of expression is likely to contribute to disease, including cancer [44].

1.4.5 Regulation of miRNA expression

MiRNA expression is regulated through a number of different mechanisms. Genomic profiling has suggested that miRNAs are frequently located in chromosomal areas most involved in alterations in human cancer; which therefore potentially influences miRNA expression [31, 47]. Defects in miRNA

processing, such as changes in Drosha, Dicer or cofactors involved in biogenesis; as well as altered expression of transcription factors, also impact miRNA expression. Epigenetic changes such as hyper-methylation or hypo-methylation have also been shown to result in aberrant miRNA expression. Complicating this, miRNAs are also able to regulate their own expression through negative feedback loops, and regulation of components of the epigenetic machinery [48, 49].

1.5 MicroRNAs in human cancer

The first indication that miRNAs may be involved in human carcinogenesis came from studies on chronic lymphocytic leukaemia (CLL) in 2005, where Calin *et al* reported down-regulation of miR-15a and miR-16-1 [50]. More recently, the function of these two miRNAs has been elucidated: negative regulation of the BCL2 gene, which functions as an anti-apoptotic gene [44]. MiR-15a and miR-16-1 therefore function as tumor suppressors, and their down-regulation in CLL allows for unchecked cell cycle progression. MiRNA profiles have demonstrated aberrant miRNA expression across a number of tumor types, which are tissue specific, and functional analysis of various dysregulated miRNAs through overexpression or silencing has supported their role as either oncogenes or tumor suppressors [34, 51, 52]. Further, miRNAs demonstrate hierarchical clustering according to developmental origin, and in work by Lu *et al*, have been used to obtain an accurate diagnosis in 12 of 17 patients with metastatic tumors of unknown primary [52] [53].

In addition, genomic mapping has facilitated the identification of loci encoding miRNAs or miRNA families. miRNAs that are frequently dysregulated in human cancers tend to be located at or in proximity of sites of genomic instability, such as areas of chromosomal breakpoints or deletions [31, 47]. Common genetic aberrations found in oral cavity SCC, such as amplification of chromosome 11q13, and loss of distal 11q, have also been associated with altered miRNA expression (such as downregulation of miR-125b and miR-100, which map to these regions) [54]. This indicates that miRNA expression in cancers may be altered secondary to genetic aberrations occurring in each particular cancer, leading to altered miRNA expression profiles and subsequent effects on target pathways, promoting carcinogenesis.

1.6 MicroRNAs in head and neck cancer

1.6.1 Introduction

Many studies in both head and neck cancer tissue as well as head and neck cancer cell lines have investigated miRNA expression profiles of HNSCC (Refer to Table 1.1), and demonstrated aberrant expression when compared with normal squamous epithelial tissue. While many different miRNAs

have been identified, there are several dysregulations that have been consistent across reported studies, and include miR-21, miR-31, miR-106b-25 cluster, miR-155, miR-363 and let-7 (up-regulated), and miR-125, and miR-375 (down-regulated) [4, 32, 55, 56]. Table 1.1 details the studies which have comprehensively profiled miRNA expression in HNSCC. Interestingly, Avissar *et al* identified that the ratio between miR-221: miR-375 had a sensitivity of 92% and specificity of 93% when differentiating between normal and tumor tissue in the head and neck, although these results have not been further validated to date [57].

1.6.2 Variations in literature

Clearly there are differences in the miRNA expression profiles identified by the various studies, and variation again in which miRNAs are found to be differentially regulated in HPV-positive and HPV-negative tumors, as discussed below. This may be due to differing methodology used (microarray, next-generation sequencing and quantitative real-time PCR), differences in tumors collected such as between anatomical subsites, and the use of HNSCC cell lines versus FFPE tumor tissue or fresh frozen tissue, sample size, and differences in analysis of results, including the use of different versions of miRBase. The latter may affect results as newer versions of miRBase not only contain novel miRNAs, but also include updated sequences of particular miRNAs. There is obvious variation in sample type utilized by different studies (cell lines, fresh frozen, FFPE). Several studies [40, 45] compared FFPE and fresh frozen tissue with SCC cell lines. Avissar *et al* [40] found that HNSCC cell lines show distinct miRNA clustering, different from that observed in primary tumors, and therefore cautioned that cell lines have limited utility as a model for determining clinically relevant miRNAs.

1.6.3 Differences across subsites

There is ongoing conflicting evidence regarding whether there are differing expression profiles in tumors occurring at different subsites in the head and neck. Hui *et al* did not identify any difference in miRNA expression between HNSCC subsites (oropharynx, hypopharynx, larynx) [55]. Lajer *et al* identified a number of differentially expressed miRNAs across oral cavity and oropharyngeal subsites compared with normal tissue, and when comparing oral cavity and oropharyngeal tumors separately, although the fold-change in miRNA expression was only moderate, and none of the miRNAs were specific to any subsite [58]. However, many of the studies do not comment on site specific miRNA differences, making it difficult to assess at this stage whether there are comprehensive differences in expression profiles between anatomic subsites or not.

Table 1.1: Published studies reporting microRNA alterations in HNSCC according to HPV tumour status

Study	Sample size	Subsite	Normal tissue	Tissue type	Methodology	Up-regulated miRs	Down-regulated miRs
Lajer 2011	30 19	Oral cavity Oropharynx Hypopharynx	39	Fresh frozen	Affymetrix miRNA microarray (miRBase v.11) qRT-PCR	miR-195, miR-363	miR-26b, miR-101, miR-125a, miR-126, miR-127, miR-143, miR-145, miR-199a, miR-199b, miR-379, miR-381, miR-409, miR-432, miR-433, miR-517a, miR-1201
Lajer 2012	46	Oral cavity Orophayrnx	13	Fresh frozen FFPE	Affymetrix GeneChip miRNA array qRT-PCR	Let-7f, let-7g, miR-15a, miR-15b, miR-16, miR-20b, miR-21, miR-29a, miR-34a, miR-125b, miR-146a, miR-146b, miR-150, miR-155, miR-342, miR-363, miR-596, miR-598, miR-625, miR-768	Let-7e, miR-31, miR-99b, miR-125a, miR-181b, miR-193b, miR-200c, miR-324, miR-423, miR-744, miR-877, miR-1189, miR-1275
Gao 2013	150	Oropharynx		FFPE	qRT-PCR	miR-9, miR-155	miR-18a, miR-31, miR-223
Hui 2013	88	Oropharynx	7	FFPE	TaqMan low density array Human miRNA panel V1.0 qRT-PCR	miR-9, miR-20b, miR-422a, miR-492, miR-545, miR-591	miR-107, miR-193b
Wald 2011	6	Oral cavity Oropharynx Hypopharynx	NHOK	Cell lines	mirVana miRNA Bioarray V2 qRT-PCR	miR-33, miR-363, miR-497	miR-29a, miR-142, miR-155, miR-181a, miR-181b, miR-218, miR-221, miR-222,
Miller 2015	53 357	Oropharynx		FFPE FFPE TMA Cell lines	Qiagen miRNome qPCR miR ISH	miR-9, miR-15b, miR-16-2, miR-20b, miR-25, miR-29c, miR-93, miR-106a, miR-106b, miR-107, miR-148a, miR-150, miR-222,miR-320a, miR-335, miR-363, miR-378, miR-598, miR-625,	miR-126, miR-143, miR-145, miR-193b, miR-199a, miR-199b, miR-214, miR-337, miR-574
Mirghani 2016	26	Oropharynx		Fresh frozen	3D-Gene human miRNA microarray (Toray Industries) qRT-PCR	miR-7-2	miR-18a, miR-18b, miR-99b, miR-107, miR-135b, miR-138-1, miR-212, miR-324, miR-584, miR-615, miR-668, miR-675, miR-857, miR-1234, miR-1246, miR-3144, miR-3176, miR-3177, miR-4267, miR-4418, miR-4764

1.6.4 Associations with survival and prognosis

Several miRNAs have also been correlated with survival and prognosis in HNSCC patients. In a study by Gao *et al*, 6 miRNAs were found to significantly correlate with prognosis: miR-142-3p, miR-146a and miR-26b over-expression were associated with improved survival, while miR-31, miR-24 and miR-193b over-expression were associated with poorer patient outcome [59]. miR-451 has also been identified as significantly over-expressed in HNSCC patients without recurrent disease [55]. Reduced expression of miR-137 miR-205 and let-7d has also been significantly associated with increased frequency of loco-regional recurrence and poorer overall survival across all subsites [60, 61]. Up-regulated miR-21 and down-regulated miR-375 have been associated with poorer survival in laryngeal SCC [62]. Over-expression of miR-21 in SCC of the tongue has also been associated with a poorer prognosis [63]. Further, up-regulation of miR-21 has also been associated with a poorer 5-year survival rate in HNSCC across all subsites [64]. While it is clear that different studies are discovering different miRNAs which are associated with patient outcomes, this nevertheless highlights the significance of miRNA expression profiles, and demonstrates a potential future role of miRNAs in providing information regarding prognostic outcomes for patients with HNSCC.

Currently, the inability to identify a primary tumor has significant effects on treatment, including a wider radiotherapy field, which in turn has significant effects on treatment-associated morbidity [51]. Barker *et al* attempted to classify tumors of unknown primary in six patients, with tumors from base of tongue, tonsil and post-nasal space, and found that each tumor maintained its miRNA profile between primary site and lymph node metastasis. In addition, the miRNA tumor profile differed between each anatomic site, allowing its potential identification through tissue obtained from the nodal metastasis [51]. Approximately 10-20% of HNSCC patients have occult metastasis to cervical lymph nodes at diagnosis, and this is one of the strongest predictors of prognosis [32, 65]. Using the differential expression of miR-205 between HNSCC and normal tissue, Fletcher *et al* were able to distinguish between benign lymph nodes and metastatic SCC to cervical lymph nodes [65]. While fine-needle aspiration of a lymph node, in combination with imaging, remains the gold standard of metastatic nodal disease detection; miRNAs have potential use as adjunct predictors of nodal disease.

1.7 MicroRNAs in HPV-associated head and neck cancer

While there are a number of studies into miRNA expression profiles in HNSCC, eight have examined HPV as an independent factor (Refer to Table 1). Lajer *et al* identified an up-regulation in miR-363, and down-regulation of miR-125a, miR-126, miR-127-3p, miR-145 and miR-379 which is associated with HPV-positive HNSCC in the oropharynx [58]. The same authors compared miRNA expression

profiles between HPV positive HNSCC and cervical cancers, and found significant similarity between groups, particularly of miR-15, miR-16, miR-195 and miR-497 and families, miR-143, miR-145 and the miR-106-363 cluster [12]. Gao *et al* identified five miRNAs which are differentially expressed in HPV-positive and HPV-negative oropharyngeal SCCs: miR-9 and miR-155 (up-regulated), and miR-18a, miR31, and miR-223 (down-regulated) [59]. Differential expression between HPV-positive and HPV-negative cell lines has also been demonstrated: miR-33, miR-363 and miR-497 are up-regulated in HPV-positive HNSCC cell lines; and miR-29a, miR-142, miR-155, miR181a, miR-181b, miR-218, miR-221 and miR-222 are down-regulated in HPV-positive HNSCC cell lines when compared with HPV-negative lines [66].

1.8 Detection of microRNAs in body fluids

miRNAs have been isolated from a number of bodily fluids, including saliva, serum, urine, sweat; as well as from solid tissues [67]. They are typically present in differing concentrations across the tissues, with the highest concentration of miRNAs occurring in saliva, breast milk, seminal fluid, tears and plasma; with the most highly expressed miRNAs conserved across fluid types [67]. The changes in miRNA expression in body tissue in association with a medical condition such as a neoplasm are reflected to varying degrees in accessible body fluids [68, 69]. The most accessible of these in HNSCC patients are plasma and saliva. Several studies have focused on miRNA profiling saliva of oral cavity tumour patients in the search for a diagnostic tool, with promising results [70, 71]. Liu and colleagues compared miRNA detection in both plasma and saliva in oral cavity tumours, finding expression levels to be higher in saliva [70]. The non-invasive collection of body fluids such as saliva for measurement of miRNA expression will provide a useful avenue for miRNA profiling, and diagnostic or prognostic tool development, in the future.

1.9 Objectives

The aims of this project were firstly to compare the expression profile of miRNAs in mucosal head and neck squamous cell carcinomas, according to HPV tumour status. Secondly, it aimed to determine any variability of miRNA expression in HNSCC in association with lifestyle factors, clinical factors, cell cycle regulation or clinical outcomes. The final aim of this project was to investigate potential genetic targets for specific miRNAs identified to be differentially expressed, and to link these if possible to the biological function of genes known to be dysregulated in HNSCC.

CHAPTER TWO: Methodology

2.1 Ethics approvals

Ethics approvals were sought and obtained from the University of Queensland (2016001065), the QIMR Berghofer Medical Research Institute HREC (1364) and Queensland Health Metro South HREC (HREC/11/QPAH/93).

2.2 Participant recruitment and data collection

Participants included in this study were recruited from a study previously performed [72]. Included participants were aged older than 18 years, with a new diagnosis of mucosal head and neck squamous cell carcinoma, presenting to the Princess Alexandra Hospital Head and Neck Clinic between 2004 – 2010. The diagnosis was confirmed with histopathology. Up to date (to January 2017) patient and clinical information was obtained from patient records. A total of 240 patients had sufficient tissue for inclusion in this study, and one formalin-fixed paraffin-embedded (FFPE) block was available for each patient. For the collection of smoking and alcohol consumption data, participants were categorised as a non-smoker or drinker (having never regularly consumed the product), current (continuing to consume either product) or ex-smoker or drinker (having ceased consumption at least six months prior to diagnosis).

2.3 HPV Status and Immunohistochemistry

HPV (Human papillomavirus) status, type and $p16^{INK4a}$ immunohistochemistry information was obtained from a previous study, where the methods are described [72].

2.4 RNA extraction

Total RNA was extracted from one block of FFPE tissue per patient using the RecoverAll Total Nucleic Acid Isolation Kit (ThermoFisher), as described by the manufacturer. MiRNAs have been demonstrated to be well preserved in FFPE tissue, and the RecoverAll kit has been shown to be as effective in recovering total RNA from FFPE tissue when compared with total RNA extracted from fresh frozen tissue [73]. Quality and quantity of RNA was established with the Nanodrop 1000 Spectrometer (A260/280 >1.9 was taken to indicate RNA purity) and RNA Integrity Number (RIN). Total RNA was stored at -80°C until further use.

2.5 Microarray

Fifty-two tonsillar SCCs were chosen for the microarray. Oropharyngeal tumours were chosen as HPV has the highest prevalence in this subset of tumours. 1μg of total RNA for each patient was reconstituted in NaOAc (3M, pH 5.2) and 100% ethanol according to the shipping instructions for LC Sciences (Texas, USA). The specimens were shipped on dry ice to LC Sciences, where a comprehensive miRNA microarray was performed, using miRBase v.21 [37-39], and Cy5 dye colour.

2.6 Selection of microRNAs for further validation

The top eighteen miRNAs which were differentially expressed between HPV positive and negative tumours on the microarray, with a fold change of ≥2.0, were chosen for further validation in a larger cohort of tumours. RNU-48, a small, non-coding RNA molecule, was chosen as the endogenous control. RNU-48 is expressed widely in human tissues and is stable; and is widely used as an endogenous control for microRNA expression profiling [74]. Further, this molecule has been utilised commonly as an endogenous control in the study of microRNA expression in head and neck cancer [60, 75, 76], as well as squamous cell carcinoma at other sites such as the lung [77].

2.7 Real-Time PCR Validation

The following steps were performed for all specimens. TaqMan® assays (Life Technologies) combined with a sensitive method of detection (Fluidigm, HD Biomark) was used. cDNA synthesis was performed with a mastermix composed of 6.75μL of a custom TaqMan RT primer pool (composed of 10μL each RT primer probe (5X) and 820μL 1xTE), 0.34μL dNTP, 1.69μL 10X RT buffer, 1.125μL Multiscribe, and 3.38μL Nuclease-free water). The RT primers included were miRNA specific RT primers contained within each TaqMan® Assay. These were: hsa-miR-15a-5p (000389), hsa-miR-15b-5p (000390), hsa-miR-16-2-3p (002171), hsa-miR-20b-5p (001014), hsa-miR-29a-3p (002112), hsa-miR-29c-3p (000587), hsa-miR-30e-5p (000421), hsa-miR-142-3p (000464), hsa-miR-146a-5p (000468), hsa-miR-148b-3p (000471), hsa-miR-150-5p (000473), hsa-miR-155-5p (002623), hsa-miR-205-3p (241040_mat), hsa-miR-342-3p (002260), hsa-miR-361-3p (002116), hsa-miR-363-3p (002171), hsa-miR-3607-3p (463814_mat), hsa-miR-4484 (464264_mat) and RNU-48 (001006; FFPE endogenous control). On ice, 3μL of total RNA (concentration of 20ng/μL) and 12μL mastermix were added to a 96-well plate. In a S100 Thermal Cycler (BioRad), the plate was incubated at 16°C for 30mins, 42°C for 30 mins and 85°C for 5 mins. The plate was placed on ice and then stored at -20°C.

Preamplification was performed on a 384-well plate on ice. 2.5μL cDNA and 7.5μL of mastermix were added to each well. The mastermix was composed of 1.69μL of assay probe pool (composed of 10μL each assay probe (20X) and 820μL 1xTE), 5.63μL preamp mastermix and 1.125μL nuclease-free water. On a BioRad PCR machine, the plate was incubated at 95°C for 10 mins, 55°C for 2 mins, 72°C for 2 mins, 14 cycles of 95°C for 15s and then 60°C for 4 mins; followed by 99.99°C for 10 mins. The plate was immediately removed and incubated on ice for 5 mins. Each preamp reaction was diluted 1:7 with 0.1xTE prior to the Fluidigm reaction.

The assay wells of the Fluidigm plate were prepared with 3μL 20X assay probe, and 3μL 2X loading reagent in each well. Each assay was performed in quadruplicate. Empty wells were prepared with 3μL nuclease-free water and 3μL 2X loading reagent. The sample wells of the Fluidigm plate were prepared with 2.7μL of diluted preamp product, and 3.3μL of gene expression mastermix (composed of 360μL gene expression buffer, and 36μL of sample loading reagent). In empty wells, nuclease-free water replaced the sample.

The Fluidigm plates were primed in the BioMark HX machine, and then each assay and sample mix was added to the plate. The plate was replaced in the BioMark HX machine for mixing of assay and samples; and then transferred to the BioMark HD machine. ROX was used as a passive reference, and FAM-MGB was used as a single probe. The plate was incubated at 50°C for 2mins, 70°C for 30mins, 25°C for 10mins, 50°C for 2mins, and then 95°C for 10mins; followed by PCR cycling (40 cycles) of 95°C for 15s and 60°C for 1min. PCR data was collected using BioMark HD Data Collection software, and analysed using the Fluidigm Real-Time PCR Analysis v.4.0.1 package. A serial dilution (30ng, 15ng, 3ng, 0.5ng) of a sample that highly expressed all miRNAs chosen for validation was included on each plate to test the limits of detection.

2.8 Median normalisation of Fluidigm data

Four Ct values were obtained for each microRNA for each sample (between 1 and 40, the number of PCR cycles). These were averaged to obtain one averaged Ct value for each microRNA for each sample. The median Ct value for RNU-48 was calculated from all Ct values obtained across all samples for RNU-48. The median value for RNU-48 Ct values across all samples was 11.60448563. Then, the difference between the median Ct value for RNU-48 and the Ct value for RNU-48 for each sample were calculated. Finally, this difference was added to each average Ct value for each microRNA for each sample. This yielded a Ct value for each microRNA, for each sample, which was normalised to the median RNU-48 Ct value. This data was then used for further analysis. This method

has been described and utilised previously in studies pertaining to the use of the Fluidigm system for analysis of microRNA expression [78, 79].

2.9 Statistical analysis

The microarray results were analysed using Genespring GX. A Mann-Whitney U-test was performed between HPV positive and negative groups to yield differentially expressed miRNAs. Analyses were also performed for clinical and pathological factors such as smoking status, recurrence of disease and $p16^{INK4a}$ status. Results from Fluidigm IFC real-time PCR were interpreted using the Fluidigm Real-Time PCR Analysis software, and normalised data was calculated as outlined above in Section 2.8. SPSS v.24 was used to analyse the results. The differences between individual groups (univariate analysis of each miRNA in each cohort) was assessed using the Mann-Whitney U-test, and adjusted for multiple comparisons using the Benjamini and Hochberg method, and ANOVA where appropriate. Univariate and multivariate logistic regression were used to determine miRNAs associated with disease stage at diagnosis and recurrence. Survival analysis was performed for recurrence-free (RFS) and disease-free survival (DFS) with univariate and multivariate cox proportional hazard model with forwards covariate search. Analyses were performed for oropharyngeal and oral cavity tumours separately, for all subsites combined, and for all non-oropharyngeal tumours (oral cavity, larynx, hypopharynx).

2.10 MicroRNA target prediction

2.10.1 Introduction

As discussed earlier, miRNAs function by targeting and inhibiting specific mRNAs. Given the imperfect complementarity in miRNA:mRNA binding in mammalian cells, an enormous number of potential target binding sites exist for each miRNA. Computational approaches for miRNA target site prediction are utilised in this context to limit the potential target genes for laboratory validation, as this can be expensive and laborious. Computational databases for miRNA target site prediction are notoriously lacking in sensitivity and specificity, and although each algorithm has predictive power, each has limitations based on inclusion or exclusion of features which may predict target sites [80]. Combining multiple prediction tools can overcome the limitations of use of a solitary algorithm, and is currently the best method for target identification and prediction [80, 81].

2.10.2 Methodology

In this study, a combination of the three most frequently updated and used online target prediction algorithms were utilised, TargetScan (version 7.1), DIANA-microT-CDS (version 5.0) and miRANDA-mirSVR (2010 Release). The specifics of each individual database will be further discussed below. Only predicted targets identified across all three online databases were considered possible targets. Each online algorithm was mined and sequentially filtered for potential target genes for each miRNA. A combination of scores from each database was utilized to give a final score for each predicted target gene. A further list of potential target genes was obtained from miRSystem (version 20160513), an online database which searches seven different online target prediction algorithms and condenses the results [82]. This list was limited to target genes identified across at least three algorithms.

Both the list of top twenty predicted gene targets for each miRNA and the full list of putative targets identified across at least three algorithms in miRSystem were utilised for further network analysis, using both the Database for Annotation, Visualisation and Integrated Discovery (DAVID; version 6.8) and Ingenuity Pathway Analysis (IPA®, Qiagen). MiRTarBase [83] was also interrogated for genes identified in the present study as the top twenty predicted targets, to identify those which have already been experimentally validated.

2.10.3 TargetScan

The TargetScan database was one of the first online computational methods for use in the identification of potential target mRNA:miRNA interactions, and is well maintained with the most recent update in 2015 [81]. This algorithm utilises a total context score (total context ++) based on 14 features. These include miRNA factors such as the P_{CT} score, or probability of conserved targeting, estimates the probability of a site being conserved due to targeting by a miRNA; seed-pairing stability (SPS), which predicts thermodynamic stability of pairing to the seed sequence; target site type; local AU content; minimum distance from the stop codon; predicted structural accessibility (SA); 3' supplementary pairing and target site abundance in the 3' UTR [84]. Predicted target mRNA factors are also considered by the algorithm, including length and number of sites in the 3' UTR and ORF [84]. The aggregate P_{CT} score is provided as one 3' UTR may contain multiple potential target sites [80, 84]. The main limitations of TargetScan currently include that the algorithm is focused to detect targets in the 3'UTR region of mRNA transcripts, and therefore does not include potential binding sites elsewhere in the target mRNA, and at the present time does not include alternative isoforms of target 3' UTRs [81, 84].

2.10.4 miRANDA-mirSVR

The miRanda-mirSVR algorithm additionally was one of the earlier online target prediction tools, and was last updated in 2010 [80, 85, 86]. In this algorithm, miRanda identifies 3' UTR binding sites and complementarity in binding, and predicts the free energy of a mRNA:miRNA interaction. Conservation of both the binding site and position in 3' UTR sequence is also taken into account. The mirSVR algorithm is then used to score potential duplexes, and includes a number of other factors such as target site accessibility, AU flanking content, position of target site, base pairing, conservation and 3' UTR length [80, 85]. Additionally, the advantage to this predictive algorithm is that it takes into account the strength of a miRNA's regulatory or repressive effect on a potential target mRNA [80, 81, 85, 86].

2.10.5 DIANA-microT-CDS

The DIANA-microT-CDS database is the latest version of another of the first online miRNA target prediction tools, and was most recently updated in 2012 [80, 87]. This algorithm takes into account both potential binding sites located in the 3'UTR as well as in coding regions (CDS), and currently remains the only online database to do so [80, 81]. The inclusion of CDS binding sites significantly increases the sensitivity of the obtained dataset [88]. Other features contained in the analysis include accessibility of target sites; binding category weight, an estimate of the efficacy of binding in an extended seed sequence; conservation; paired stability; distance to adjacent binding sites; AU content; and the predicted free energy of the mRNA:miRNA binding [80, 81]. The algorithm uses individual models for binding sites in each 3' UTR and CDS, and then combines them into a single score.

2.11 Gene ontology and gene network analysis

For each miRNA, the Database for Annotation, Visualisation and Integrated Discovery (DAVID; version 6.8) was interrogated using the list of putative target genes obtained from miRSystem for both gene ontology analysis and functional enrichment of gene clusters. Gene ontology aims to describe gene functions and unify the relationships between these concepts, with three aspects. These include molecular functions, or the molecular activities of a gene; and biological activities, such as the larger pathways which are composed of multiple functions and genes. For this analysis, *Homo Sapiens* was used as the species and background. A p value of <0.0001 was considered significant for over-represented functions. Functional gene classification was also performed using DAVID. The list obtained from miRSystem for each miRNA, as well as the list of top gene targets identified in the previous study, were classified into functionally enriched gene clusters. This was performed using

Homo Sapiens as the species and background, and the highest stringency filter. An enrichment score of ≥2.0 was considered to be significantly enriched.

Ingenuity Pathway Analysis (IPA®, Qiagen) was then used to interrogate the list of putative target genes, including both the top twenty predicted targets identified, and a full list of predicted targets obtained from miRSystem, for the top gene networks targeted. This software transforms a list of genes into a large network, which optimizes interconnectivity between a set of user-specific genes, which may represent significant biological functions. The significance of a particular pathway is identified with a Fischer's Exact Test, and a p value of <0.05 was considered statistically significant. Up-stream regulators of these pathways were noted, as was interactions with canonical pathways. The genes involved in the top twenty-five target gene networks identified were interrogated for the genes identified in the top twenty miRNA targets in the present study.

CHAPTER THREE: MiRNA Profiling

3.1 Patient characteristics and HPV status

A summary of patient and tumour characteristics, separated by HPV status, are presented in Table 3.1. A total of 240 patients had sufficient tissue for RNA extraction and were included in this study. Forty-six of the tumours were HPV positive, and the highest prevalence of HPV was in the oropharynx (n=36, p<0.0001). All HPV-positive tumours contained oncogenic HPV types. The majority of HPV positive tumours expressed HPV-16 (n=38), followed in frequency by HPV-18 (n=5), HPV-33 (n=2) and HPV-69 (n=1). The majority of patients with HPV positive tumours were younger than those with HPV-unrelated tumours. In both groups, the majority of patients were male, and current or ex-smokers and consumers of alcohol. As expected, a strong correlation existed between $p16^{INK4a}$ (p<0.0001) and Cyclin D1 (p<0.0001) status in HPV-associated tumours. HPV-positive tumours tended to present at clinically more advanced nodal stages than their HPV-negative counterparts (p<0.0001). HPV-negative tumours had higher rates of recurrence of the primary tumour and overall poorer survival than the HPV-associated tumours included in this study [72].

Table 3.1: Patient and tumour characteristics of HPV-positive and negative tumours

Characteristics	**HPV Negative n (%) 194**	**HPV Positive n (%) 46**	**P value**
Age			
<40 years	5 (2.6%)	3 (6.5%)	**0.0010**
40-49 years	21 (10.8%)	7 (15.2%)	
50-59 years	52 (26.8%)	24 (52.2%)	
60-69 years	49 (25.3%)	9 (19.6%)	
70-79 years	47 (24.2%)	2 (4.3%)	
>80 years	20 (10.3%)	1 (2.2%)	
Gender			
Male	138 (71.1%)	40 (87.0%)	**0.0275**
Female	56 (28.9%)	6 (13.0%)	
Smoking			
Non-smoker	19 (10.0%)	6 (13.0%)	**0.0427**
Current smoker	105 (55.3%)	16 (34.8%)	
Ex-smoker	66 (34.7%)	24 (52.2%)	
Alcohol			
Non-drinker	18 (9.9%)	5 (11.1%)	0.9356
Current drinker	146 (80.2%)	35 (77.8%)	
Ex-drinker	18 (9.9%)	5 (11.1%)	
Tumour site			**<0.0001**
Oral cavity	83 (42.8%)	5 (10.9%)	
Oropharynx	39 (20.1%)	36 (78.3%)	
Larynx	52 (26.8%)	3 (6.5%)	
Hypopharynx	20 (10.3%)	2 (4.3%)	
Tumour differentiation			
Well	26 (14.1%)	1 (2.3%)	**0.0300**
Mod	106 (57.6%)	23 (53.5%)	
Poor	52 (28.3%)	19 (44.2%)	
Recurrence			
Yes	63 (36.0%)	5 (11.6%)	**0.0167**
No	106 (60.6%)	38 (88.4%)	
Not stated	1 (0.6%)	0 (0.0%)	

Other	1 (0.6%)	0 (0.0%)	
N/A	4 (2.2%)	0 (0.0%)	
T Stage			0.1880
T1	44 (22.9%)	14 (31.1%)	
T2	51 (26.6%)	16 (35.6%)	
T3	36 (18.7%)	4 (8.9%)	
T4	61 (31.8%)	11 (24.4%)	
N Stage			**<0.0001**
N0	114 (59.4%)	7 (15.5%)	
N1	22 (11.5%)	7 (15.5%)	
N2	51 (26.5%)	27 (60.0%)	
N3	5 (2.6%)	4 (9.0%)	
M Stage			0.7949
M0	187 (99.0%)	43 (100.0%)	
M1	1 (0.5%)	0 (0.0%)	
Mx	1 (0.5%)	0 (0.0%)	
Disease status at last follow-up			**0.0033**
Alive without disease	80 (41.3%)	32 (69.6%)	
Alive with disease	7 (3.6%)	2 (4.3%)	
Dead without disease	5 (2.6%)	0 (0.0%)	
Dead with disease	67 (34.5%)	5 (10.9%)	
Lost to F/U	35 (18.0%)	7 (15.2%)	
p16 status			
Positive	19 (10.7%)	41 (89.1%)	**<0.0001**
Negative	159 (89.3%)	5 (10.9%)	
Cyclin D1 status			
Positive	175 (97.2%)	35 (79.5%)	**<0.0001**
Negative	5 (2.8%)	9 (20.5%)	

3.2 Microarray results

The microarray (performed for fifty-two oropharyngeal tumours) found a significant difference in miRNA expression between HPV positive and negative tumours, as defined by HPV DNA positivity in the tumour (Refer to Figure 3.1 and Table 3.2). Seventeen miRNAs were differentially expressed in HPV DNA positive tumours, with a fold change (FC) of >2.0 ($p < 0.001$). Six of these retained their significance with a $p < 0.0001$: hsa-miR-16-2-3p, hsa-miR-146a-5p, hsa-miR-150-5p, hsa-miR-155-5p, hsa-miR-342-3p and hsa-miR-3607-3p. Eighteen miRNAs were also identified to be differentially expressed between tumours that were HPV DNA/$p16^{INK4a}$ positive and those that were HPV DNA/$p16^{INK4a}$ negative (Refer to Table 3.3). Differential miRNA expression was also identified between smokers and non-smokers (Refer to Supplementary Table 3.4), patients who had recurrence of the primary tumour (Refer to Supplementary Table 3.5) and with $p16^{INK4a}$ status (Refer to Supplementary Tables 3.2 and 3.3).

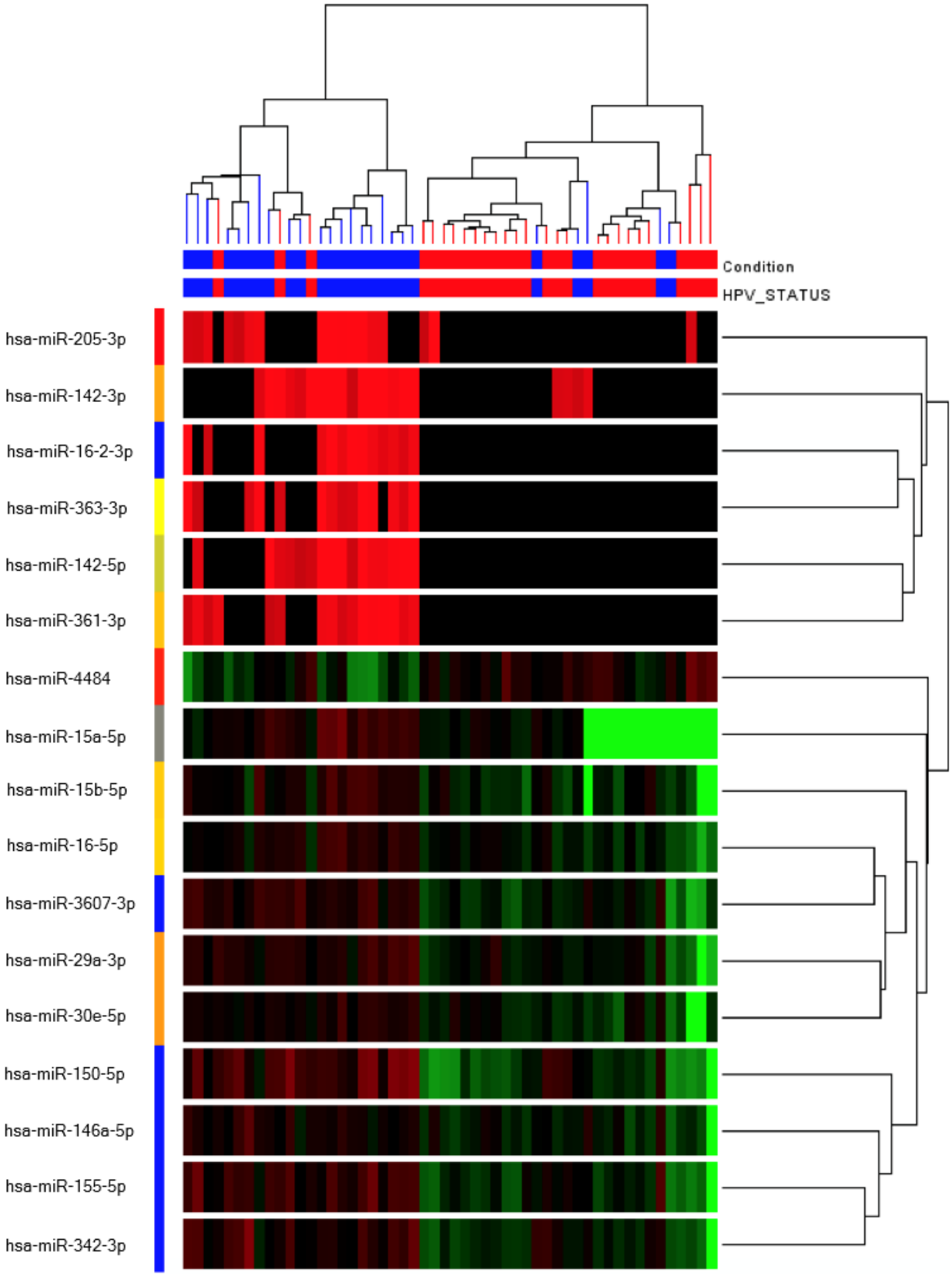
Table 3.2: Differential microRNA expression in HPV DNA positive oropharyngeal tumours observed on microarray (Fold change >2.0, p <0.001). Seventeen miRNAs were differentially expressed in HPV DNA positive tumours, one down-regulated, and sixteen up-regulated.

miRNA	**FC**	**P value**	**Direction**
hsa-miR-142-3p	9.87	3.54E-4	Up
hsa-miR-142-5p	8.77	1.49E-4	Up
hsa-miR-16-2-3p	8.73	3.70E-5	Up
hsa-miR-363-3p	7.42	1.65E-4	Up
hsa-miR-361-3p	7.12	2.95E-4	Up
hsa-miR-150-5p	6.55	2.10E-5	Up
hsa-miR-205-3p	6.50	7.93E-4	Up
hsa-miR-15a-5p	5.64	1.26E-4	Up
hsa-miR-155-5p	4.13	5.82E-6	Up
hsa-miR-342-3p	3.74	9.47E-6	Up
hsa-miR-146a-5p	3.06	3.12E-6	Up
hsa-miR-4484	3.06	6.44E-4	Down
hsa-miR-29a-3p	2.75	3.90E-4	Up
hsa-miR-15b-5p	2.67	2.80E-4	Up
hsa-miR-30e-5p	2.60	3.90E-4	Up
hsa-miR-3607-3p	2.39	5.72E-5	Up
hsa-miR-16-5p	2.21	2.61E-4	Up

Table 3.3: Differential microRNA expression in HPV DNA and $p16^{INK4a}$ positive oropharyngeal tumours observed on microarray (Fold change >2.0, p <0.01). Eighteen miRNAs were differentially expressed in positive tumours, one down-regulated, and seventeen up-regulated.

miRNA	**FC**	**P value**	**Direction**
hsa-miR-15a-5p	5.66	4.42E-4	Up
hsa-miR-16-2-3p	7.44	1.02E-4	Up
hsa-miR-16-5p	2.10	0.001	Up
hsa-miR-29a-3p	2.74	7.10E-4	Up
hsa-miR-29c-3p	4.30	4.73E-4	Up
hsa-miR-30e-5p	2.53	0.001	Up
hsa-miR-142-3p	10.48	2.55E-4	Up
hsa-miR-142-5p	7.14	4.97E-4	Up
hsa-miR-146a-5p	2.87	1.88E-5	Up
hsa-miR-150-5p	7.23	1.21E-5	Up
hsa-miR-155-5p	4.14	5.88E-6	Up
hsa-miR-342-3p	3.33	9.24E-5	Up
hsa-miR-342-5p	4.04	0.001	Up
hsa-miR-361-3p	6.94	1.69E-4	Up
hsa-miR-363-3p	6.13	3.71E-4	Up
hsa-miR-3607-3p	2.40	6.16E-5	Up
hsa-miR-3653-5p	5.18	3.56E-4	Up
hsa-miR-4484	2.72	0.001	Down

Figure 3.1: Heat map corresponding to Table 3.2 visually demonstrating the difference in miRNA expression between HPV positive (blue) and negative (red) tumours in the oropharynx, with a fold change of >2.0 ($p<0.0001$). Up-regulated microRNAs are designated red, and down-regulated microRNAs are designated green.



3.3 Validation of microRNAs

A larger patient cohort was used to validate the microarray findings, including tumours from all four head and neck subsites: oropharynx, oral cavity, larynx and hypopharynx (n = 240). Eighteen miRNAs were chosen for further validation: hsa-miR-15a-5p, hsa-miR-15b-5p, hsa-miR-16-2-3p, hsa-miR-20b-5p, hsa-miR-29a-3p, hsa-miR-29c-5p, hsa-miR-30e-5p, hsa-miR-142-3p, hsa-miR-146a-5p, hsa-miR-148b-3p, hsa-miR-150-5p, hsa-miR-155-5p, hsa-miR-205-3p, hsa-miR-342-3p, hsa-miR-361-3p, hsa-miR-363-3p, hsa-miR-3607 and hsa-miR-4484. hsa-miR-4484 failed the assay quality control, and was therefore not included in any further analyses.

3.3.1 HPV Status

Between HPV positive and negative tumours, five miRNAs demonstrated differential expression in the oropharynx (Refer to Table 3.4). These were: hsa-miR-16-2-3p, hsa-miR-20b-5p, hsa-miR-142-3p, hsa-miR-150-5p and hsa-miR-363-3p (up-regulated). There was no difference between HPV positive and negative tumours at other individual subsites (oral cavity, larynx, hypopharynx). Across all HNSCC subsites combined, seven miRNAs were significantly differentially expressed between the two groups. These were: hsa-miR-29a-3p, hsa-miR-29c-3p, hsa-miR-148b-3p and hsa-miR-205-3p (down-regulated), and hsa-miR-142-3p, hsa-miR-150-5p and hsa-miR-363-3p (up-regulated).

Table 3.4: Differential microRNA expression between HPV DNA positive tumours across all subsites (All), and in the oropharynx (OP) identified on further validation of miRNA expression.

miRNA	**HPV Positive MnCt (Range)**	**HPV Negative MnCt (Range)**	**P value**	**Log2 Fold change**
hsa-miR-16-2-3p				
OP	25.95 (22.40-40.00)	39.31 (23.67-40.00)	**0.0069**	13.36
All	26.04 (22.40-40.00)	32.03 (20.90-40.00)	ns	
hsa-miR-20b-5p				
OP	19.91 (16.20-29.64)	21.11 (18.36-37.19)	**0.0016**	1.20
All	19.86 (16.20-29.64)	20.48 (12.96-39.57)	ns	
hsa-miR-29a-3p				
OP	12.67 (10.25-14.83)	12.64 (10.43-15.84)	ns	
All	12.70 (10.10-14.83)	11.72 (2.80-23.70)	**<0.0001**	-0.98
hsa-miR-29c-3p				
OP	22.03 (18.13-29.64)	20.82 (18.39-40.00)	ns	
All	21.96 (18.04-29.64)	20.36 (8.30-37.19)	**<0.0001**	-1.60
hsa-miR-142-3p				
OP	14.62 (12.05-19.75)	16.54 (12.83-22.56)	**0.0006**	1.92
All	15.38 (11.69-19.75)	15.79 (7.52-27.87)	**0.0397**	0.41
hsa-miR-148b-3p				
OP	24.54 (21.19-37.65)	24.75 (20.50-40.00)	ns	
All	24.54 (21.19-34.05)	23.48 (17.04-40.00)	**0.0363**	-1.06
hsa-miR-150-5p				
OP	10.49 (9.03-13.55)	11.52 (8.16-15.41)	**0.0117**	1.03
All	10.67 (8.30-13.58)	11.64 (1.78-26.13)	**0.0039**	0.97
hsa-miR-205-3p				
OP	24.10 (19.53-37.06)	24.60 (18.57-40.00)	ns	

All	24.04 (21.29-37.06)	22.67 (12.93-40.00)	**0.0043**	-1.37
hsa-miR-363-3p				
OP	25.61 (22.41-40.00)	37.20 (22.45-40.00)	**0.0004**	11.59
All	25.68 (22.41-40.00)	28.53 (16.20-40.00)	**0.0304**	2.85

3.3.2 Analysis of HPV DNA status and p16^{INK4a} status

In the oropharynx, seven miRNAs were differentially expressed in HPV DNA positive and p16^{INK4a} positive tumours as compared with HPV DNA negative and p16^{INK4a} negative tumours. These were: hsa-miR-15b-5p, hsa-miR-16-2-3p, hsa-miR-20b-5p, hsa-miR-30e-5p, hsa-miR-142-3p, hsa-miR-361-3p and hsa-miR-363-3p, and were all up-regulated in HPV DNA and p16^{INK4a} positive tumours (Refer to Table 3.5). Across all subsites, only four miRNAs were significantly differentially expressed: hsa-miR-16-2-3p, hsa-miR-20b-5p (up-regulated), hsa-miR-29a-3p and hsa-miR-29c-3p (down-regulated). There was no difference observed at other subsites of the head and neck (oral cavity, larynx and hypopharynx).

Table 3.5: Differential microRNA expression between HPV DNA positive and p16^{INK4a} positive tumours across all subsites (All) and in the oropharynx (OP) identified on further validation of miRNA expression.

miRNA	**HPV Positive MnCt (Range)**	**HPV Negative MnCt (Range)**	**P value**	**Fold change (Log2)**
hsa-miR-15b-5p				
OP	17.04 (14.69-29.64)	17.85 (15.67-21.28)	**0.0249**	0.80
All	17.23 (14.69-29.64)	17.33 (8.79-27.21)	ns	
hsa-miR-16-2-3p				
OP	24.61 (22.40-40.00)	39.73 (23.67-40.00)	**0.0008**	15.11
All	25.91 (22.40-40.00)	35.64 (16.36-40.00)	**0.0126**	9.73
hsa-miR-20b-5p				
OP	18.80 (16.20-29.64)	21.26 (18.36-37.19)	**<0.0001**	2.45
All	19.12 (16.20-29.64)	20.36 (11.42-39.57)	**0.0187**	1.24
hsa-miR-29a-3p				
OP	12.46 (10.25-14.83)	12.69 (10.43-15.84)	ns	
All	12.67 (10.25-14.83)	11.60 (2.80-23.70)	**0.0406**	**-1.09**
hsa-miR-29c-3p				
OP	21.57 (18.13-29.64)	20.79 (18.62-40.00)	ns	
All	21.61 (18.13-29.64)	20.41 (8.30-40.00)	**0.0104**	-1.18
hsa-miR-30e-5p				
OP	18.89 (17.48-29.64)	20.69 (17.27-40.00)	**0.0121**	1.80
All	18.97 (17.08-29.64)	19.38 (11.75-40.00)	ns	
hsa-miR-142-3p				
OP	13.76 (12.05-17.62)	16.48 (12.83-22.56)	**0.0005**	2.72
All	14.08 (11.69-19.29)	15.74 (7.52-27.87)	ns	
hsa-miR-361-3p				
OP	24.79 (21.27-40.00)	27.50 (22.06-40.00)	**0.0456**	**2.70**
All	25.12 (21.27-40.00)	25.64 (12.61-40.00)	ns	
hsa-miR-363-3p				
OP	25.13 (22.41-40.00)	37.20 (22.45-40.00)	**0.0027**	**12.07**
All	25.68 (22.41-40.00)	28.46 (16.20-40.00)	ns	

3.3.3 Cell cycle regulation factors

There was no difference in miRNA expression between $p16^{INK4a}$ positive and negative specimens in any individual subsite of the head and neck. hsa-miR-16-2-3p was significantly differentially expressed when all subsites were combined (p=0.0351). In the oropharynx, four miRNAs were up-regulated between tumours positive and negative for Cyclin D1 immunohistochemistry. These were: hsa-miR-146a-5p (p=0.0012), hsa-miR-150-5p (p=0.0366), hsa-miR-342-3p (p=0.0454) and hsa-miR-3607-3p (p=0.038). There was no difference in miRNA expression with Cyclin D1 status across all subsites combined.

3.3.4 Lifestyle factors

Across all subsites, hsa-miR-363-3p expression was significantly up-regulated in non-smokers as compared with smokers (p = 0.0032). Several other miRNAs trended towards significantly different expression, but this did not achieve statistical significance. These included: hsa-miR-16-2-3p, hsa-miR-20b-5p, hsa-miR-205-3p and hsa-miR-3607-3p. In the oropharynx alone, hsa-miR-20b-5p was significantly differentially expressed between smokers and non-smokers (p<0.0001). In a two-way ANOVA comparison of the effect of HPV status and smoking status on miRNA expression, only hsa-miR-150-5p (p=0.029) was significantly different in the oropharynx. There was no difference at other individual subsites or across all sites.

There was no difference in miRNA expression between patients who consumed alcohol and those who abstained at individual and across all subsites. Across all subsites, and in the oral cavity, larynx and hypopharynx individually, there was no difference in miRNA expression between HPV positive and negative tumours in association with alcohol consumption status. In the oropharynx, a number of miRNAs were up-regulated between HPV positive and negative tumours in association with alcohol consumption status. These were: hsa-miR-15a-5p (p<0.0001), hsa-miR-20b-5p (p<0.0001), hsa-miR-29c-3p (p<0.0001), hsa-miR-30e-5p (p<0.0001), hsa-miR-142-3p (p=0.0005), hsa-miR-146a-5p (p=0.032), hsa-miR-148b-3p (p=0.011), hsa-miR-155-5p (p=0.041) and hsa-miR-205-3p (p=0.002).

3.3.5 Recurrence of primary tumour

Irrespective of HPV status, in the oropharynx, down-regulation of hsa-miR-16-2-3p was associated with recurrence of the primary tumour (p = 0.0377). Up-regulation of four miRNAs was associated with primary tumour recurrence across all four head and neck subsites: hsa-miR-15a-5p (p=0.0247), hsa-miR-29a-3p (p=0.0007), hsa-miR-29c-3p (p=0.001) and hsa-miR-3607-3p (p=0.0487). In

oropharyngeal HPV positive tumours compared with HPV negative tumours, hsa-miR-205-3p (p=0.006), hsa-miR-142-3p (p=0.050) and hsa-miR-363-3p (p=0.044) were up-regulated in HPV positive tumour without recurrence.

3.3.6 Stage at diagnosis

There was no difference observed in miRNA expression with increasing T-stage at diagnosis across all subsites, with and without consideration of HPV status.

In the oropharynx, an analysis of HPV status and N-stage found down-regulation of hsa-miR-15b-5p (p=0.012) and hsa-miR-342-3p (p=0.009), and up-regulation of hsa-miR-15a-5p (p=0.012), hsa-miR-30e-5p (p=0.034) and hsa-miR-155-5p (p=0.046) to be significantly associated with increasing N-stage at diagnosis. Irrespective of HPV-status, increasing N-stage at diagnosis was associated with significant up-regulation of hsa-miR-20b-5p (p=0.024), hsa-miR-142-3p (p=0.038), hsa-miR-146a-5p (p=0.043) and hsa-miR-363-3p (p=0.023). Across all subsites, down-regulation of hsa-miR-29c-3p (p=0.045) and hsa-miR-150-5p (p=0.019), and up-regulation of hsa-miR-142-3p (p=0.016) were associated with increasing nodal stage at diagnosis. There was no difference with the inclusion of HPV status across all subsites.

With increasing M-stage at diagnosis, in the oropharynx, hsa-miR-155-5p was up-regulated (p=0.018). Across all subsites, increasing M-stage at diagnosis correlated with up-regulation of hsa-miR-15b-5p (p=0.026) and hsa-miR-150-5p (p=0.025). There was no difference when analysed for HPV tumour status in the oropharynx or across all subsites.

3.3.7 Survival

A statistically significant difference in survival was only observed between HPV-positive and negative tumours in the oropharynx. The Kaplan-Meier curve for oropharyngeal tumours and corresponding statistics are given in Figure 3.2 and Table 3.6.

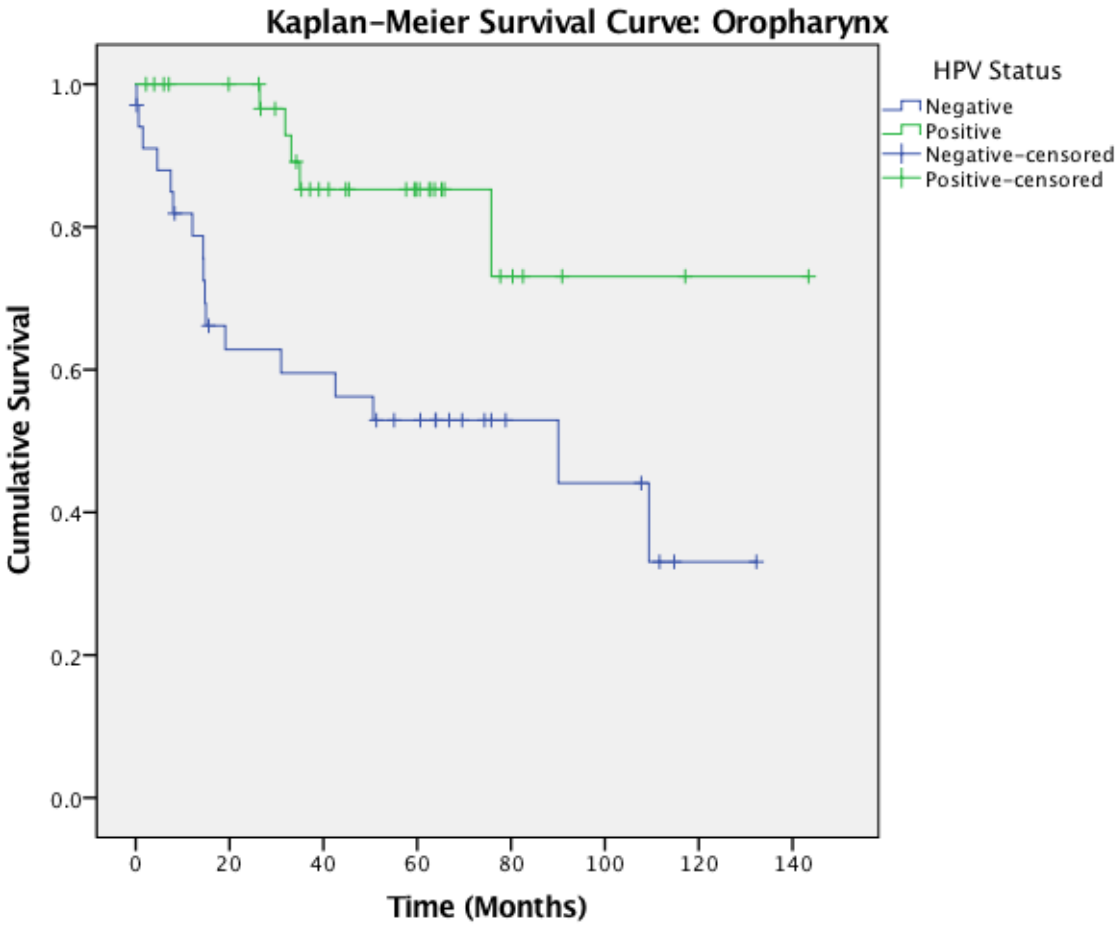


Figure 3.2: The Kaplan-Meier survival curve demonstrates a significant difference in disease-specific survival in HPV-positive oropharyngeal tumours (green) compared with HPV-negative tumours (blue).

Table 3.6: Overall comparison statistics corresponding to the Kaplan-Meier survival curve for oropharyngeal tumours

	Chi-Square	**df**	**Sig**
Log Rang (Mantel-Cox)	8.078	1	0.004
Breslow (Generalized Wilcoxon)	10.106	1	0.001
Tarone-Ware	9.332	1	0.002

Disease-free survival (DFS)

Across all subsites combined, miR-20b was significantly associated with DFS (p=0.002) when HPV status, stage at diagnosis and smoking status were taken into account. No miRNAs were significantly associated with DFS when HPV status was not considered. In the oropharynx, no miRNAs were significantly associated with DFS when HPV status was taken into account, along with stage at diagnosis and smoking status. However, when HPV status was not considered, hsa-miR-20b-5p was significantly associated with DFS (p=0.008). When evaluating tumours of the oral cavity, along with HPV status and smoking status, hsa-miR-20b-5p (p=0.021) and hsa-miR-363-3p (p=0.018) were significantly associated with DFS. In HPV-negative tumours of the oral cavity, hsa-miR-363-3p was once again significantly associated with DFS (p=0.015).

Recurrence-free survival (RFS)

Across all subsites combined, three miRNAs were significantly associated with RFS when HPV status and smoking status were taken into account. These were: hsa-miR-29a-3p (p<0.001), hsa-miR-30e-5p (p=0.037) and hsa-miR-342-3p (p=0.019). No miRNAs were significant when HPV status was not considered. In oropharyngeal tumours, hsa-miR-16-2-3p was significantly associated with RFS (p=0.01) when HPV status was considered, however, retained its significance when HPV status was removed from the analysis (p=0.011). Two other miRNAs were also associated with RFS when HPV status was not considered: hsa-miR-29a-3p (p=0.004), and hsa-miR-155-5p (p=0.001). No miRNAs were significantly associated with RFS in the oral cavity with and without consideration of HPV tumour status.

CHAPTER FOUR: Predicted target functionality

4.1 hsa-miR-15a-5p and hsa-miR-15b-5p

The top twenty target genes identified across all four target prediction programs are given, along with their corresponding scores from each program, in Table 4.1. TargetScan interrogation revealed 968 putative targets, with 1071 conserved and 324 poorly conserved sites. DIANA microT-CDS investigation revealed 1611 potential target genes, and the miRANDA-mirSVR database found 8785 potential target genes.

DAVID Analysis of these genes identified two distinct functional clusters: Group 1 contained CBX4, PURA, MYBL1, MYB, MED26 and PLAG1 (enrichment score 1.36), and Group 1 contained PLSCR4, PCMT1, SEMA6D, GPR63 and UNC80 (enrichment score 0.38). Investigation of the DAVID functional classification database using miRSystem identified genes (number 1198), identified five enriched pathways (Refer to Table 4.2) including those associated with carcinogenesis such as regulation of transcription and cell proliferation. Of these, TLK1 was identified in cluster 1, and PLAG1 was identified in cluster 3. GO analysis revealed over-representation of molecular functions such as ubiquitination, protein kinase activity and transcription regulation ($p<0.0001$), and of biological functions including apoptosis, ubiquitination, transcription regulation, cell signaling and proliferation ($p<0.0001$).

Table 4.1: Top twenty predicted targets for hsa-miR-15a-5p and hsa-miR-15b-5p

Target gene	**Full Gene Name**	**TargetScan Aggregate P_{CT}**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally Validated [83]**
USP15	Uniquitin specific peptidase 15	0.950	0.977	-2.570	Yes
UNC80	Unc-80 homolog, NALCN activator	0.540	0.982	-2.830	No
MYB	MYB proto-oncogene	0.980	0.998	-2.230	Yes
TLK1	Tousled like kinase 1	0.980	1.000	-2.210	Yes
ARIH1	Ariadne RBR E3 ubiquitin protein ligase 1	0.860	0.765	-2.430	Yes
MYBL1	MYB proto-oncogene like 1	0.980	0.987	-2.070	Yes
CCNE1	Cyclin E1	0.970	1.000	-2.020	Yes
DNAJB4	DnaJ head shock protein (Hsp40) member B4	0.950	0.981	-2.040	No
ANO3	Anoctamin 3	0.790	0.959	-2.130	No
WEE1	WEE1 G2 checkpoint kinase	0.980	1.000	-1.850	Yes
C1QL3	Complement C1q like 3	0.760	0.814	-2.250	No

COL12A1	Collagen type XII alpha 1 chain	0.900	0.837	-2.080	No
PLSCR4	Phospholipid scramblase 4	0.460	0.932	-2.420	Yes
SHOC2	Leucine rich repeat scaffold protein	0.940	0.935	-1.900	Yes
C2orf42	Chromosome 2 open reading frame 42	0.770	0.893	-2.070	Yes
PLAG1	PLAG1 zinc finger	0.980	1.000	-1.740	Yes
PPP1R11	Protein phosphatase 1 regulatory inhibitor subunit 11	0.820	0.986	-1.910	Yes
SLC9A6	Solute carrier family 9 member A6	0.750	0.884	-2.060	Yes
CBX4	Chromobox 4	0.840	0.991	-1.860	Yes
GPR63	G-protein coupled receptor 63	0.990	0.928	-1.770	No

Table 4.2: Functionally enriched target gene groups of hsa-miR-15a-5p and hsa-miR-15b-5p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Protein phosphorylation, ATP binding, Nucleotide binding, Intracellular signal transduction	9.53	18
Cell membrane transport and signaling, Cell proliferation	5.64	13
Metal ion binding, DNA binding, Transcription regulation	4.56	32
Cadherin, cell adhesion, cell signaling, calcium binding	3.80	15
Kelch, Actin binding, Cytoskeleton	2.42	5

IPA network analysis identified 25 gene networks involved in critical cellular pathways (Refer to Table 4.3). The top up-stream regulators identified were miR-16 (p=5.71E-16), miR-15 (p=4.24E-7), miR-34 (p=9.22E-7), TP53 (5.25E-6) and ESR1 (1.91E-5). Other regulators included MYC, Histone H2, E2F1, WNT3A, FGF2 and IGF1. The top canonical pathways targeted include insulin receptor signaling, the PTEN and STAT3 pathways, axonal guidance and molecular mechanisms of cancer pathway. Comparison of the top 20 targets identified in this study with those involved in the top 10 IPA networks, revealed significant cross-over. ANO3 was found to be a member of Group 1; MYB, MYBL1, DNAJB4 and WEE1 were members of Group 2; CCNE1 is in Group 3; and CBX4 was contained in Group 4. Group 8 contained PPP1R11 and SHOC2; while Group 9 contained COL12A1.

Table 4.3: Top ten networks for hsa-miR-15 target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Post-translational modification, cell-to-cell signaling and interaction, hair and skin development and function	33	Cellular development, cellular growth and proliferation, haematological system development and function	30
Gene expression, cancer, gastrointestinal disease	32	Carbohydrate metabolism, cellular function and maintenance, molecular transport	29
Haematological system development and function, haematopoiesis, lymphoid tissue structure and development	32	Lipid metabolism, molecular transport, small molecule biochemistry	27
Cancer, haematological disease, immunological disease	31	Cancer, organismal injury and abnormalities, respiratory disease	27
Embryonic development, organismal development, cellular development	30	Immunological disease, gastrointestinal disease, hepatic system disease	26

4.2 hsa-miR-16

The top twenty target genes identified across all four target prediction programs are given, along with their corresponding scores from each program, in Table 4.4. Interrogation of TargetScan revealed 4497 possible target mRNA transcripts, containing a total of 6699 possible binding sites. The DIANA microT-CDS database identified 840 possible target genes, while miRANDA-mirSVR found 7754 potential target genes.

Analysis (DAVID) of these genes identified two clusters of functionally related genes: Group 1 contained MYEF2, CPEB2, PURA and CPEB3 (enrichment score 1.61), and Group 2 contained MIER3, ZNF624 and MYEF2 (enrichment score 1.49). DAVID analysis of all putative target genes (1147) identified across at least three databases by MiRSystem interrogation revealed over-representation of molecular functions such as protein kinase activity, ATP binding, and protein and ion transport ($p<0.0001$), as well as biological functions such as ubiquitination, intra-cellular signaling, cell differentiation and cytoskeleton functions ($p<0.0001$). Functional gene classification identified four enriched pathways (Refer to Table 4.5). Of these clusters, TLK1 was identified in cluster 1, and PLAG1 was identified in cluster 3.

Table 4.4: Top twenty predicted targets for hsa-miR-16

Target gene	Full Gene Name	TargetScan Total context ++	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
TLK1	Tousled like kinase 1	-0.060	0.977	-2.220	Yes
NGLY1	N-glycanase 1	-0.490	0.729	-1.560	No
PURA	Prune rich element binding protein A	-0.280	0.744	-1.740	Yes
PLAG1	PLAG1 zinc finger	-0.170	0.744	-1.710	Yes
NAA30	N(alpha)-acetyltransferase 30	-0.350	0.978	-1.260	No
ACADM	Acyl-CoA dehydrogenase, C-4 to C-12 straight chain	-0.520	0.825	-1.240	No
MIER3	MIER family member 3	-0.230	0.802	-1.450	No
TMEM7B	Transmembrane protein 7B	-0.080	0.915	-1.460	No
CDK17	Cyclin dependent kinase 17	-0.130	0.956	-1.350	Yes
MYEF2	Myelin expression factor 2	-0.100	0.771	-1.540	No
CPEB3	Cytoplasmic polyadenylation element binding protein 3	-0.040	0.956	-1.410	Yes
CUL2	Cullin 2	-0.310	0.912	-1.160	Yes
ANKIB1	Ankyrin repeat and IBR domain containing 1	-0.050	0.719	-1.610	No
ECAFB1		-0.500	0.768	-1.110	No
PAFAH1B1	Platelet activating factor acetylhydrolase 1b regulatory subunit 1	-0.140	0.980	-1.250	Yes
CPEB2	Cytoplasmic polyadenylation element binding protein 2	-0.150	0.988	-1.210	Yes
CYP26B1	Cytochrome P450 family 26 subfamily B member 1	-0.210	0.933	-1.200	Yes
B2M	Beta-2-microglobulin	-0.500	0.706	-1.110	Yes
NOVA1	NOVA alternative splicing regulator 1	-0.060	0.978	-1.200	No
GALNT1	Polypeptide N-acetyl galactosaminyltransferase 1	-0.240	0.895	-1.090	Yes

Table 4.5: Functionally enriched target gene groups of hsa-miR-16 identified by gene functional classification in DAVID

Functional classification group	Enrichment score	Number of target genes
Protein kinase activity, ATP binding, phosphorylation, cell cycle	9.87	18
Cell membrane structure and function	5.97	16
Metal ion binding, nucleotide binding, transcription regulation	4.52	30
Cadherin, cell adhesion, cell signaling, calcium binding	3.87	15

IPA pathway analysis revealed 25 target gene networks. The top 10 are listed in Table 4.6. The top up-stream regulators of these pathways were identified as miR-16 (p=3.39E-16), miR-15 (p=3.23E-7), ESR1 (3.19E-6), miR-34 (p=6.47E-6) and TP53 (p=6.87E-6). Other statistically significant (p<0.0001) upstream regulators include MYC, E2F1, VEGFA, IGF1 and WNT3A. The top canonical pathways targeted by miR-16 are PTEN, STAT2, Insulin and B-cell receptor signaling and molecular mechanisms of cancer pathway. Within these top 10 gene networks, several of the top 20 target genes identified in this study were found, including MYEF2 in Group 1; CUL2 in Group 2; and PLAG1 and CPEB2 in Group 7.

Table 4.6: Top ten networks for hsa-miR-16 target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Gene expression, lipid metabolism, molecular transport	33	Cellular development, cellular growth and proliferation, embryonic development	29
Post-translational modification, cell death and survival, gene expression	32	Cancer, organismal injury and abnormalities, respiratory disease	28
Gene expression, DNA replication, recombination and repair, cancer	32	Endocrine system disorders, metabolic disease, carbohydrate metabolism	27
Cellular compromise, cell morphology, molecular transport	31	Embryonic development, organ development, organismal development	25
Cancer, organismal injury and abnormalities, reproductive system disease	30	Nervous system development and function, organ morphology, organismal development	25

4.3 hsa-miR-20b-5p

hsa-miR-20b-5p belongs to the miR-17 family, which also includes miR-20a, miR-93 and miR-106a/b. This study has identified up-regulation of hsa-miR-20b-5p in HPV-related tumours. The top twenty predicted targets for hsa-miR-20b-5p are listed in Table 4.7. TargetScan identified 1384 possible target gene transcripts, with 1649 conserved and 898 poorly conserved possible binding sites. DIANA microT-CDS found 480 possible targets, and miRANDA-mirSVR identified 9122 potential target genes.

Table 4.7: Top twenty predicted targets for hsa-miR-20b-5p

Target Gene	Full Gene Name	TargetScan Aggregate P_{CT}	DIANA miTG	miRANDA mirSVR	Experimentally validated [83]
DYNC1LI2	Dynein cytoplasmic 1 light intermediate chain 2	0.980	0.986	-2.930	Yes
ZNFX1	Zinc finger NFX1-type containing 1	>0.990	1.000	-2.770	Yes
KPNA3	Karyopherin subunit alpha 3	0.810	0.918	-2.510	No
C14orf28	Chromosome 14 open reading frame 28	0.740	0.952	-2.390	Yes
SCN1A	Sodium voltage-gated channel alpha subunit 1	>0.990	0.976	-2.300	No
PKD2	Polycystin 2, transient receptor potential cation channel (polycystic kidney disease),	>0.990	0.999	-2.250	No
LHX8	LIM homeobox 8 (neural tube development)	0.740	0.969	-2.290	No
STAT3	Signal transducer and activator of transcription 3	0.930	0.920	-2.100	Yes
PFN2	Profilin 2	0.920	0.995	-1.990	No
TXNIP	Thioredoxin interacting protein	0.970	0.993	-1.910	Yes
ZFYVE26	Zinc finger FYVE-type containing 26	>0.990	0.925	-1.890	Yes
PCDHA4	Protocadherin alpha 4	0.760	0.911	-2.110	No
SERP1	Stress associated endoplasmic reticulum protein 1	0.960	0.961	-1.840	No
PCDHA6	Protocadherin alpha 6	0.760	0.910	-1.970	No
PCDHA2	Protocadherin alpha 2	0.760	0.910	-1.970	No
PCDHA3	Protocadherin alpha 3	0.760	0.910	-1.970	No
NR4A3	Nuclear receptor subfamily 4 group A member 3	0.930	0.986	-1.580	No
FCHO2	FCH domain only 2	0.920	0.996	-1.500	Yes
RAB11FIP5	RAB11 family interacting protein 5	0.960	0.998	-1.330	No
PCNP	PEST proteolytic signal containing nuclear protein	0.740	0.759	-1.740	No

DAVID analysis of these genes identified one cluster of genes containing PCDHA6, PRRG1 and PCDHA4 (enrichment score 0.85). To further assess the biological relevance of predicted targets for miR-20b, gene ontologies (GO) were explored for all putative target genes identified across at least three distinct databases, using miRSystem. miRSystem interrogation revealed 1121 target genes identified across at least three databases. These functional gene group classifications are listed in Table 4.8. Three target genes identified in the top twenty target genes were contained in Group 4: PCDHA2, 3, 4, and 6. GO enrichment analysis identified over-representation of biological processes such as regulation of cell cycle progression, cell proliferation and apoptosis; as well as cell migration, adhesion and signaling ($p < 0.0001$). Molecular functions were also over-represented, related to protein and ion binding; cytoskeleton activity; and transcription factor activation ($p < 0.0001$).

Table 4.8: Enriched hsa-miR-20b-5p target gene groups identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Protein kinase activity, phosphorylation, intra-cellular signalling	10	22
Protein kinase activity, intra-cellular signaling, nucleotide binding, protein phosphorylation	6.82	5
Transcription regulation, DNA binding	6.44	50
Cadherin, cell adhesion, calcium binding	4.77	16
Ubiquitin protein activity, cytoskeleton organization, Kelch	3.29	7
GTP-binding, GTPase activity, intra-cellular signalling	2.66	6
Inter-cellular signaling, cell secretion	2.3	6

Network analysis of the 1121 genes identified by miRSystem identified 25 gene networks involved in critical cellular pathways, and the top ten are detailed in Table 4.9. The top regulators included SP3 (p=1.68E-7), Histone H4 (p=1.28E-6), NR3C1 (p=1.86E-6), TFAP2A (p=3.03E-6) and TP63 (6.33E-6). However, up-stream regulators also included E2F1 (p=1.82E-4), TP53 (p=8.33E-4), CDKN2A (p=3.25E-3) and EGF1 (p=1.98E-3). The top canonical pathways targeted include molecular mechanisms of cancer pathways, NGF signaling, HGF signaling and cell cycle regulation by the BTG family of proteins. Of the genes involved in the top gene networks identified in IPA, several were also found in the top 20 putative targets identified in this study. These included PCDHA4 in Group 5; TXNIP in Group 6; PFN2 and NR4A3 in Group 7; and STAT3 in Group 8.

Table 4.9: Top ten networks for hsa-miR-20b-5p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Cardiovascular disease, organismal injury & abnormalities, reproductive system disease	34	Cellular growth & proliferation, tissue development, cancer	31
Cellular growth and proliferation, gene expression, survival	33	Cellular development, cellular growth & proliferation, cellular movement	31
Gene expression, cellular movement, survival	31	Cell death & survival, cellular movement, gene expression	27
Cardiovascular system development & function, connective tissue disorders	31	Cell cycle, connective tissue development & function, cellular development	25
Cancer, hematological disease, organismal injury & abnormalities	31	Gene expression, protein synthesis, infectious diseases	20

4.4 hsa-miR-29a-3p

The top twenty predicted gene targets for hsa-miR-29a-3p are listed in Table 4.10. TargetScan identified 1256 possible gene targets, with a total of 1455 conserved and 319 poorly conserved sites. DIANA microT-CDS identified 991 potential target genes, and miRANDA-mirSVR found 6777 potential targets. DAVID analysis revealed one functional cluster of genes: COL4A1, PXDN, COL3A1, ELN, COL5A2, COL5A1, COL1A2 and PDGFC (enrichment score 4.76).

miRSystem identified 875 potential target genes across at least three databases. GO interrogation of this list revealed over-representation of biological functions such as cell adhesion, extracellular matrix formation and function, and cytoskeleton organization and function ($p<0.0001$), and of molecular functions including protein binding, ubiquitination, transcription factor activity and calcium binding ($p<0.0001$). Seven clusters of genes were functionally enriched, and these are listed in Table 4.11. Five genes from the top twenty targets identified in this study were found in cluster 2: COL3A1, COL11A1, COL1A2, COL5A1, and COL5A2.

Table 4.10: Top twenty predicted targets for hsa-miR-29a-3p

Target Gene	Full Gene Name	TargetScan score Aggregate P_{CT}	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
COL3A1	Collagen type III alpha 1 chain	>0.990	1.000	-2.650	Yes
TDG	Thymine DNA glycosylase	0.900	0.967	-2.410	Yes
PDGFC	Platelet derived growth factor C	0.920	0.971	-2.360	No
HBP1	HMG-box transcription factor 1	0.950	1.000	-2.330	Yes
COL11A1	Collagen type XI alpha 1 chain	0.970	0.986	-2.270	No
ATAD2B	ATPase family, AAA domain containing 2B	>0.990	1.000	-2.220	No
TET2	Tet methylcytosine dioxygenase 2	>0.990	0.994	-2.030	Yes
ELN	Elastin	>0.990	1.000	-1.920	No
COL1A2	Collagen type I alpha 2 chain	0.990	0.990	-1.880	No
REV3L	REV2 like, DNA directed polymerase zeta catalytic subunit	0.990	0.901	-1.950	No
JARID2	Jumonji and AT-rich interaction domain containing 2	0.850	0.724	-2.250	No
HMCN1	Hemicentin 1	0.750	0.994	-2.060	No
COL5A1	Collagen type V alpha 1 chain	>0.990	0.986	-1.760	No
PTEN	Phosphatase and tensin homolog	>0.990	0.983	-1.740	Yes
MAPRE1	Microtubule associated protein RP/EB family member 1	0.840	0.948	-1.920	No
TET1	Tet methylcytosine dioxygenase 1	>0.990	0.991	-1.710	Yes
OTUD4	OUT deubiquitinase 4	0.880	0.980	-1.630	Yes
CCNJ	Cyclin J	0.740	0.998	-1.750	No
COL5A2	Collagen type V alpha 2 chain	>0.990	0.992	-1.450	Yes

ARRDC3	Arrestin domain containing 3	0.950	0.941	-1.450	No

Table 4.11: Functionally enriched target gene groups of hsa-miR-29a-3p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Extracellular matrix, basement membrane and collagen structure and function, angiogenesis, cell proliferation, cell differentiation	6.25	6
Extracellular matrix formation and function, calcium metabolism, collagen	6.20	8
Cadherin, calcium regulation, cell adhesion	5.27	16
Protein kinase activity, phosphorylation, ATP & nucleotide binding	3.20	12
Transcription, metal binding, nucleic acid binding	3.17	27
Peptidase, metal ion binding, cell membrane function	3.06	9
Translation, chromosome arrangement, protein binding	2.97	6

IPA analysis identified 25 gene networks, and the top 10 are listed in Table 4.12. The top up-stream regulators are miR-29 (p=2.52E-9), TGFB1 (p=1.19E-9), ERF (p=1.15E-7) and NR3C1 (p=2.27E-7). Other significant regulators (p<0.0001) are PGF, ERBB2, ESR1, IFN-B, GH, TNF, TP63, and PI3K. Top canonical pathways involved included axonal guidance signaling, PAK signaling and glioma signaling. Several of the top twenty predicted target genes for miR-29a were identified in the leading gene networks identified in IPA. These include: COL5A1 and COL5A2 in Group 1; PDGFC in Group 5; COL3A1 and ELN in Group 6; HBP1 in Group 8; and PTEN in Group 10.

Table 4.12: Top ten networks for hsa-miR-29a-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Connective tissue disorders, organismal injury and abnormalities, cancer	32	Cellular assembly and orgnisation, tissue development, embryonic development	27
Gene expression, cardiovascular disease, haematological disease	31	Post-translational modification, cellular movement, cellular development	27
Cancer, organismal injury and abnormalities, reproductive system disease	31	Cell-to-cell signaling and interaction, cellular assembly and organization, tissue development	27
Haematological disease, immunological disease, antimicrobial response	28	Cell morphology, cell-to-cell signaling and interaction, nervous system development and function	26
Cardiovascular disease, tissue morphology, DNA replication, recombination and repair	28	Cancer, organismal injury and abnormalities, tissue morphology	25

4.5 hsa-miR-29c-3p

The top twenty predicted gene targets for hsa-miR-29c-3p are given in Table 4.13. miRANDA-mirSVR identified a possible 6856 gene targets, while DIANA microT-CDS identified 23 possible targets. TargetScan identified 409 possible transcripts, with a total of 413 possible binding sites. DAVID analysis identified one functional genetic cluster in these genes: ZNF346, YY1 and ZNF468 (enrichment score 0.59).

MiRSystem interrogation revealed 875 putative target genes identified across three databases. GO analysis revealed over-representation of biological functions including extracellular matrix organization, cell adhesion, and regulation of cell proliferation ($p<0.0001$), as well as over-representation of molecular functions such as protein binding, ubiquitination, growth factor activity and calcium binding ($p<0.0001$). Functional gene clustering of these identified seven enriched functional clusters (Refer to Table 4.14). None of the top twenty target genes identified in this study were identified in these clusters.

Table 4.13: Top twenty predicted targets for hsa-miR-29c-3p

Target gene	**Full Gene Name**	**TargetScan Total context ++**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally validated [83]**
PTBP2	Polypyrimidine tract binding protein 2	-0.350	0.492	-2.010	No
ELN	Elastin	-0.370	0.454	-1.920	No
MYCN	v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog	-0.370	0.751	-1.080	Yes
DNMT3A	DNA methyltransferase 3 alpha	-0.290	0.481	-1.220	Yes
BZW1	Basic leucine zipper and W2 domains 1	-0.180	0.505	-1.050	No
CAPRIN1	Cell cycle associated protein 1	-0.230	0.435	-1.070	No
LIMS1	LIM zinc finger domain containing 1	-0.230	0.522	-0.940	Yes
YY1	YY1 transcription factor	-0.210	0.431	-1.010	No
ATP6V1A	ATPase H+ transporting V1 subunit A	-0.480	0.596	-0.560	No
DOLPP1	Dolichyldiphosphatase 1	-0.250	0.435	-0.94	No
CLEC7A	C-type lectin domain family 7 member A	-0.240	0.410	-0.960	No
COL22A1	Collagen type XXII alpha 1 chain	-0.220	0.529	-0.840	No
MYO1D	Myosin 1D	-0.200	0.454	-0.920	No
CRISPLD2	Cysteine rich secretory protein LCCL domain containing 2	-0.230	0.408	-0.880	No
CCNYL1	Cyclin Y like 1	-0.240	0.407	-0.860	No
KCTD5	Potassium chanel tetramerization domain containing 5	-0.290	0.412	-0.750	No
ZNF468	Zinc finger protein 468	-0.180	0.375	-0.750	No
ZNF346	Zinc finger protein 346	-0.290	0.446	-0.560	No
FMN1	Formin 1	-0.200	0.443	-0.580	No
XKR4	XK related 4	-0.120	0.401	-0.700	No

Table 4.14: Functionally enriched target gene groups of hsa-miR-29c-3p identified by gene functional classification in DAVID

Functional classification group	Enrichment score	Number of target genes
Collagen, cell adhesion	6.54	6
Collagen, extracellular matrix organization, cell signalling	6.37	8
Cadherin, calcium ion binding, cell adhesion	5.31	16
Peptidase activity, metal ion binding,	3.19	9
Protein kinase activity, phosphorylation, ATP binding	3.13	12
Zinc finger, metal ion binding, transcription regulation, nucleic acid binding	3.08	27
WD40, phosphorylation, translation initiation	2.92	6

IPA analysis found 25 target gene networks, the top 10 of which are listed in Table 4.15. The top canonical pathways targeted are axonal guidance signaling, PAK and glioma signaling pathways. Top regulators of these networks include miR-29 (p=2.54E-9), TGFB1 (p=3.47E-10), Estrogen receptor (p=8.85E-8) and ERK (p=1.17E-7). Other regulators (p<0.0001) include IFN-B, TFAP2A, NR3C1, SPDEF, TNF and TP53. Several of the top 20 predicted targets identified in this study were found to be involved in the top networks identified in IPA. These include: MYCN and YY1 in Group 1; ZNF346 in Group 3; DNMT3A in Group 4; and ELN in Group 10.

Table 4.15: Top ten networks for hsa-miR-29c-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Gene expression, organismal survival, embryonic development	32	Cancer, organismal injury and abnormalities, cellular assembly and organization	27
Cell death and survival, haematological disease, immunological disease	30	Cellular assembly and organization, tissue development, cancer	25
Cellular growth and proliferation, connective tissue development and function, tissue development	30	Port-translational modification, protein degradation, protein synthesis	25
Cardiovascular disease, tissue morphology, DNA replication, recombination and repair	29	Organismal development, cell death and survival, gastrointestinal disease	24
Cell-to-cell signaling and interaction, nervous system development and function, cardiovascular system development and function	28	Cell death and survival, cellular development, cellular growth and proliferation	24

4.6 hsa-miR-30e-5p

The top twenty predicted targets for hsa-miR-30e-5p are given in Table 4.16. TargetScan found a total of 1566 transcripts, with 1848 conserved and 565 poorly conserved sites. DIANA microT-CDS identified 1787 possible gene targets, and miRANDA-miR-SVR found 6323 potential target genes. DAVID analysis of the genes did not identify any functional clusters.

GO analysis revealed over-representation in molecular functions such as protein binding, ubiquitination, chromatin binding and transcription factor activity ($p<0.0001$), as well as biological functions such as transcription regulation, signal transduction, intracellular transport, and apoptosis ($p<0.0001$). Investigation of miRSystem targets for miR-30e found 1150 target genes across at least three databases. Functional classification of these genes identified 8 enriched clusters (Refer to Table 4.17). Of the top target genes identified in this study, RARB was identified in cluster 1, WDR7 in cluster 3 and SCN2A in cluster 7.

Table 4.16: Top twenty predicted targets for hsa-miR-30e-5p

Target Gene	**Full Gene Name**	**TargetScan Aggregate P_{CT}**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally validated [83]**
CELSR3	Cadherin EGF LAF seven-pass G-type receptor 3	>0.990	1.000	-4.050	Yes
RAP1B	RAP1B, member of RAS oncogene family	0.810	0.993	-3.710	Yes
RFX6	Regulatory factor X6	0.990	1.000	-3.480	No
XPO1	Exportin 1	0.940	0.991	-3.200	Yes
TNRC6A	Trinucleotide repeat containing 6A	>0.990	1.000	-3.110	Yes
ANKRA2	Ankyrin repeat family A member 2	0.650	0.991	-3.200	Yes
MEX3B	Mex-3 RNA binding family member B	0.960	1.000	-2.840	No
RAD23B	RAD23 homolog B	0.960	0.962	-2.770	Yes
SCN2A	Sodium voltage-gated channel subunit 2	0.950	1.000	-2.700	No
B3GNT5	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	0.990	0.997	-2.630	No

RARB	Retinoic acid receptor beta	0.900	0.950	-2.750	No
LHX8	LIM homeobox 8	0.980	1.000	-2.590	No
WDR7	WD repeat domain 7	0.980	0.942	-2.640	No
UBE2V2	Ubiquitin conjugating enzyme E2 V2	0.960	0.982	-2.610	No
PDE7A	Phosphodiesterase 7A	>0.990	1.000	-2.520	No
PTPN13	Protein tyrosine phosphatase, non-receptor type 13	0.830	0.983	-2.690	No
FOXG1	Forkhead box G1	0.930	0.986	-2.560	Yes
CBLB	Cbl proto-oncogene B	0.870	0.972	-2.540	No
PAPD4	Poly(A) RNA polymerase D4	-0.970	0.997	-2.360	No
RFX7	Regulatory factor X7	0.880	0.993	-2.450	No

Table 4.17: Functionally enriched target gene groups of hsa-miR-30e-5p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
DNA binding, steroid hormone receptor function, transcription regulation	7.56	9
Zinc finger, metal ion binding, nucleotide binding, transcription regulation	6.10	38
Protein kinase activity, nucleotide binding, ATP binding, intracellular signal transduction	5.26	15
Ubiquitination, ATP binding	4.15	5
WD40, cytoskeleton structure and function, calcium binding	3.97	6
GTP binding, methylation, intracellular vesicle function, endocytosis	3.66	8
Sodium channel function, Action potential, cell membrane function	3.11	5
Immunoglobulin, cell adhesion, cell membrane structure and function	2.14	6

25 target gene networks were identified on IPA analysis. The top 10 are listed in Table 4.18. The top canonical pathways involved included axonal guidance signaling, Reelin signaling in neurons, cardiac hypertrophy signaling, integrin signaling and B-cell receptor signaling. The predominant up-stream regulators included miR-34 (p=6.39E-7), estrogen receptors (p=5.03E-6), FSH (p=6.93E-6), miR-339 (p=7.96E-6) and miR-8 (p=9.08E-6). Others included TNF, CCND1, MAPK1, and E2F1 (p<0.0001). Within these top 10 networks, several genes identified in Table 4.16 were identified, including FOXG1 in Group 2; PTPN13 in Group 4; RARB in Group 5; RAD23B in Group 6; and RAP1B in Group 9.

Table 4.18: Top ten networks for hsa-miR-30e-5p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Infectious diseases, cell death and survival, embryonic development	32	Gene expression, respiratory disease, cancer	30
Gene expression, developmental disorder, gastrointestinal disease	31	Cell morphology, organ morphology, skeletal and muscular disorders	30
Amino acid metabolism, small molecule biochemistry, connective tissue development and function	30	Gene expression, cell death and survival, embryonic development	29
Cellular assembly and organization, cellular function and maintenance, cardiovascular disease	30	Cellular growth and proliferation, nervous system development and function, tissue morphology	28
Cellular growth and proliferation, tissue development, developmental disorder	30	Cardiac arrhythmia, cardiovascular disease, hereditary disorder	28

4.7 hsa-miR-142-3p

The top twenty potential target genes for hsa-miR-142-3p are listed in Table 4.19. TargetScan identified 599 possible target gene transcripts, with 662 conserved and 216 poorly conserved binding sites. miRANDA-mirSVR identified 4334 possible targets and DIANA microT-CDS identified 517 possible target genes. DAVID analysis identified one functional genetic cluster, which included SLC35F5, TMEM200B, DIRC2, SLC38A4 and STX12 (enrichment factor 0.34).

GO analysis revealed over-representation of molecular functions such as protein, chromatin, metal ion and calmodulin binding, GTPase activity, transcription factor activity and cadherin binding ($p<0.0001$). Biological functions including regulation of transcription and gene expression, endocytosis, intracellular protein transportation, and regulation of apoptosis were also over-represented ($p<0.0001$). MiRSystem identified 336 target genes across at least three databases. Functional gene clustering of these target genes identified one functional gene cluster (Refer to Table 4.20). No genes identified in the top 20 putative targets were identified in this cluster.

Table 4.19: Top twenty predicted targets for hsa-miR-142-3p

Target gene	Full Gene Name	TargetScan Aggregate P_{CT}	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
TMEM200B	Transmembrane protein 200B	0.780	0.804	-2.770	No
ARNTL	Aryl hydrocarbon receptor nuclear translocator like	0.780	0.999	-2.520	Yes
USP6NL	USP6 N-terminal like	0.960	0.989	-2.190	Yes
WASL	Wiskott-Aldrich syndrome like	0.840	0.999	-2.240	Yes
C20orf194	Chromosome 20 open reading frame 194	0.940	0.986	-2.130	No
FNDC3A	Fibronectin type III domain containing 3A	0.720	0.924	-2.140	Yes
RHOBTB3	Rho related BTB domain containing 3	0.770	0.704	-2.200	No
ASH1L	ASH1 like histone lysine methyltransferase	0.800	0.987	-1.870	No
BOD1	Biorientation of chromosomes in cell division 1	0.980	0.993	-1.680	Yes
C9orf72	Chromosome 9 open reading frame 72	0.850	0.997	-1.730	Yes
HMGA2	High mobility group AT-hook 2	0.700	0.717	-2.050	Yes
ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif 3	0.730	0.832	-1.900	No
STAU1	Staufen double-stranded RNA binding protein 1	<0.100	0.949	-2.350	No
KDM6A	Lysine demethylase 6A	0.760	0.824	-1.760	No
SLC35F5	Solute carrier family 35 member F5	0.910	0.971	-1.440	Yes
SLC38A4	Solute carrier family 38 member A4	0.360	0.845	-2.070	No
RAB2A	RAB2A, member RAS oncogene family	0.870	0.999	-1.340	Yes
CCNJ	Cyclin J	0.730	0.887	-1.590	Yes
MAP4K3	Mitogen-activated protein kinase kinase kinase kinase 3	0.930	0.865	-1.340	No
STX12	Syntaxin 12	0.730	0.921	-1.460	No

Table 4.20: Enriched hsa-miR-142-3p target gene groups identified by gene functional classification in DAVID

Functional classification group	Enrichment score	Number of target genes
Zinc finger, metal ion binding, nucleic acid binding, transcription regulation	3.36	13

25 target gene networks were identified with IPA analysis, and the top 10 are listed in Table 4.21. The top canonical pathways include clathrin-mediated endocytosis signaling, protein-kinase A signaling, epithelial adherens junction signaling, and molecular mechanisms of cancer. The top regulators include miR-142 (p=1.51E-5), MLK1 (p=3.79E-5), UPF1 (p=5.26E-5), miR-8 (p=6.74E-5) and FOXO3 (p=1.27E-4). Other regulators (p<0.0001) include RAB1B, ELK1, IL-2, FOXO1, BTG1, STAT5, AKT3, STAT3, TP53 and MTDH. Within these top 10 networks (Refer to Table 4.21), several target genes identified in this study were found. These include ASH1L and RAB2A in Group 1; HMGA2 in Group 2; ARNTL in Group 3; USP6NL, STAU1 and KDM6A in Group 4; WASL in Group 5; FNDC3A, RHOBTB3 and STX12 in Group 6; WASL in Group 7; and BOD1 in Group 8.

Table 4.21: Top ten networks for hsa-miR-142-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Cancer, organismal injury and abnormalities, post-translational modification	27	Cellular development, inflammatory disease, organismal injury and abnormalities	16
Cellular growth and proliferation, cellular development, cell death and survival	25	Cellular function and maintenance, neurological disease, organismal injury and abnormalities	14
Cellular growth and proliferation, cellular assembly and organization, tissue development	25	Cell death and survival, cellular movement, cellular growth and proliferation	14
Neurological disease, connective tissue disorders, haematological disease	22	Cell morphology, cellular assembly and organization, cancer	13
Cell morphology, cellular assembly and organization, cellular function and maintenance	16	Cell death and survival, cell cycle, embryonic development	13

4.8 hsa-miR-146a-5p

The top twenty predicted gene targets for hsa-miR-146a-5p are listed in Table 4.22. DIANA microT-CDS identified 854 possible targets, and miRANDA-mirSVR found 6798 possible target genes. TargetScan found 275 possible target gene transcripts, with a total of 290 conserved and 124 poorly conserved possible binding sites. Functional clustering analysis (DAVID) identified two clusters of target genes: Group 1 contained RARB, KLF7, MYBL1, HNRNPD and ZBTB2 (enrichment score 1.2); and Group 2 contained SLC19A3, LFNG, IGSF1 and SLC10A3 (enrichment score 0.14).

Table 4.22: Top twenty predicted targets for hsa-miR-146a-5p

Target Gene	**Full Gene Name**	**TargetScan Total context ++**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally validated [83]**
NOVA1	NOVA alternative splicing regulator 1	-0.490	0.892	-2.980	No
TRAF6	TNF receptor associated factor 6	-1.000	1.000	-2.320	Yes
IGSF1	Immunoglobulin superfamily member 1	-0.670	1.000	-2.320	No
IRAK1	Interleukin 1 receptor associated kinase 1	-0.560	1.000	-1.670	Yes
ZBTB2	Zinc finger and BTB domain containing 2	-0.570	0.998	-1.490	No
WWC2	WW and C2 domain containing 2	-0.650	0.953	-1.380	No
ZHDDC13		-0510	0.983	-1.310	No
HNRNPD	Heterogeneous nuclear ribonucleoprotein D	-0.520	0.983	-1.300	No
CKDN2AIP		-0.580	0.931	-1.240	No
ZNRF2	Zinc and ring finger 2	-0.520	0.719	-1.440	No
NUMB	NUMB, endocytic adaptor protein	-0.380	0.938	-1.340	Yes
RARB	Retinoic acid receptor beta	-0.390	0.973	-1.290	No
EIF4G2	Eukaryotic translation initiation factor 4 gamma 2	-0.350	0.975	-1.320	No
SLC10A3	Solute carrier family 10 member 3	-0.530	0.985	-1.120	No
SLC19A3	Solute carrier family 19 member 3	-0.310	0.812	-1.500	No
STRBP	Spermatid perinuclear RNA binding protein	-0.330	0.900	-1.330	No
KLF7	Kruppel like factor 7	-0.310	0.998	-1.250	No
MYBL1	MYB proto-oncogene like 1	-0.180	0.968	-1.390	No
RFTN2	Raftlin family member 2	-0.160	0.749	-1.550	No

LFNG	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	-0.390	0.872	-1.100	No

340 putative target genes were identified across at least three databases using miRSystem. GO analysis found that biological functions such as regulation of cell proliferation, DNA transcription, cell migration, ubiquitination and control of the cell cycle were over-represented ($p<0.0001$), as were molecular functions such as transcription factor activity, co-receptor binding, protein kinase activity and ubiquitin-protein activity ($p<0.0001$). Gene functional classification identified one enriched cluster of genes (Refer to Table 4.23). Of these, KLF7 and ZBTB2 were also identified in our top twenty putative target genes for miR-146a.

Table 4.23: Functionally enriched target gene groups of hsa-miR-146a-5p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Transcription regulation, zinc finger, metal ion binding, DNA binding, cell division	3.21	17

Of the 25 target gene networks identified, the top 10 are listed in Table 4.24. Top canonical pathways included a-adrenergic signaling, FMLP signaling in neutrophils, 3-phosphoinositide biosynthesis and Fcy receptor-mediated phagocytosis. The top regulators include HNRNPA2B1 (p=1.73E-5), TSC22D3 (p=8.02E-5), ERBB4 (p=2.19E-4), NGF (p=6.23E-4) and LNX2 (p=6.23E-4). Others included CDK4, let-7, TP53, BTG2, and CCND1 ($p<0.0001$). Within these top 10 networks, several of our top target genes were identified to be involved. These were inclusive of: TRAF6, IGSF1 and IRAK1 in Group 1; HNRNPD in Group 2; RARB in Group 3; RFTN2 in Group 4; and NOVA1 in Group 5.

Table 4.24: Top ten networks for hsa-miR-146a-5p target genes from Ingenity Pathway Analysis

Network functions	**Number of Genes**	**Network functions**	**Number of Genes**
Haematological system development and function, lymphoid tissue structure and development, tissue morphology	26	Gene expression, cell morphology, organismal injury and abnormalities	15
Cancer, endocrine system disorders, organismal injury and abnormalities	24	Cell cycle, cell death and survival, cellular growth and proliferation	14
Cancer, endocrine system disorders, organismal injury and abnormalities	23	Developmental disorder, cancer, organismal injury and abnormalities	14

Cell morphology, cellular function and maintenance, cell cycle	16	Humoral immune response, immunological disease, protein synthesis	13
Organismal injury and abnormalities, cardiovascular disease, connective tissue disorders	15	Cell death and survival, nervous system development and function, cellular growth and proliferation	13

4.9 hsa-miR-148b-3p

The top twenty predicted genetic targets for hsa-miR-148b-3p are listed in Table 4.25. TargetScan identified 795 possible target genes, with a total of 882 conserved, and 263 poorly conserved possible binding sites. miRANDA-mirSVR identified 7413 possible targets, and DIANA-microT CDS found 954 putative targets. Functional clustering analysis (DAVID) identified one genetic cluster, which contained S1PR1, ROBO1, ATP6AP2, B4GALT5 and PTPRA (enrichment score 0.45). GO analysis revealed over-representation of molecular functions such as protein binding, protein kinase activity, transcription factor activity and growth factor binding ($p<0.0001$). Biological functions including regulation of transcription, phosphorylation, cell cycle arrest, regulation of cell adhesion and DNA damage response regulation ($p<0.0001$). MiRSystem identified 679 putative target genes across at least three databases for miR-148b. Functional clustering of these genes found three enriched functional clusters (Refer to Table 4.26). None of the top twenty genes identified in this study were identified in these clusters.

Table 4.25: Top twenty predicted targets for hsa-miR-148b-3p

Target Gene	**Full Gene Name**	**TargetScan Aggregate P_{CT}**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally validated [83]**
SOS2	SOS Ras/Rho guanine nucleotide exchange factor 2	0.890	1.000	-2.670	Yes
ATP6AP2	ATPase H+ transporting accessory protein 2	0.960	0.999	-2.430	Yes
BAI3		0.790	0.999	-2.510	No
MEOX2	Mesenchyme homeobox 2	0.960	1.000	-2.160	No
RBM24	RNA binding motif protein 24	0.820	0.900	-2.350	No
C5orf30	Chromosome 5 open reading frame 30	0.690	0.997	-2.370	No
ROBO1	Roundabout guidance receptor 1	0.940	0.985	-2.060	No

OSBPL11	Oxysterol binding protein like 11	0.820	1.000	-2.100	Yes
NPTN	Neuroplastin	0.890	0.998	-2.030	No
CDK19	Cyclin dependent kinase 19	0.980	0.992	-1.890	Yes
USP33	Ubiquitin specific peptidase 33	0.760	0.977	-2.110	Yes
PTEN	Phosphatase and tensin homolog	0.990	0.863	-1.970	No
SNAP91	Synaptosome associated protein 91	0.760	0.999	-1.990	No
UBE2D1	Ubiquitin conjugating enzyme E2 D1	0.580	0.825	-2.320	No
HOMER1	Homer scaffolding protein 1	0.790	0.879	-1.990	No
DMXL1	Dmx like 1	0.890	0.992	-1.740	No
WDR47	WD repeat domain 47	0.660	0.931	-2.030	No
S1PR1	Sphingosine 1-phosphate receptor 1	0.940	0.997	-1.440	No
ROCK1	Rho associated coiled-coil containing protein kinase 1	>0.990	0.998	-1.290	Yes
MITF	Melanogenesis associated transcription factor	0.980	0.991	-1.270	No

Table 4.26: Functionally enriched target gene groups of hsa-miR-148b-3p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Protein kinase activity, ATP binding, intracellular signaling, phosphorylation	5.83	11
Zinc finger, transcription regulation, metal ion binding, DNA binding	3.87	7
GTP binding, protein transport, autophagy	3.80	6

IPA analysis identified 25 target gene networks, the top ten of which are detailed in Table 4.27. The predominant canonical pathways are PPARα/RXRα activation, estrogen receptor signaling, Neuregulin signaling, B-cell receptor signaling and G-protein coupled receptor signaling. The main up-stream regulators included TP63 (p=2.20E-8), miR-148 (p=1.71E-7), STATa/b (p=1.52E-6), TP53 (p=2.33E-6) and NOX4 (p=3.06E-5). Others included MAP2K1, IGF1, MITF, AKT3, MYC, and PTEN (p<0.0001). Within the top 10 gene networks identified in IPA, MITF and ROCK1 were involved in Group 1; PTEN in Group 2; S1PR1 in Group 5; and CDK19 in Group 8.

Table 4.27: Top ten networks for hsa-miR-148b-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Celllular growth and proliferation, developmental disorder, gastrointestinal disease	30	Gene expression, organismal survival, cellular growth and proliferation	25
Cardiovascular system development and function, organismal development, tissue morphology	28	Cellular development, connective tissue development and function, tissue development	25
Protein synthesis, carbohydrate metabolism, molecular transport	28	Cellular growth and proliferation, connective tissue development and function, tissue development	22
Cell death and survival, gene expression, cellular movement	27	RNA damage and repair, molecular transport, RNA trafficking	19
Cardiovascular system development and function, organismal development, cellular movement	25	Embryonic development, organismal development, tissue development	16

4.10 hsa-miR-150-5p

The top twenty predicted target genes for hsa-miR-150-5p are given in Table 4.28. TargetScan identified 351 possible targets, with a total of 370 conserved and 404 poorly conserved binding sites. miRANDA-mirSVR found 9583 putative target genes, and DIANA microT-CDS found 1121 possible targets. Functional clustering analysis (DAVID) found two clusters of genes: Group 1 contained ZBTB4, NR2F2, FOXD3, ZEB1, MYB, MDM4 and RORB (enrichment score 1.90); while Group 2 included CMTM6, MTCH2 and SGMS1 (enrichment score 0.52).

Table 4.28: Top twenty predicted targets for hsa-miR-150-5p

Target gene	Full Gene Name	TargetScan Total context ++	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
MYB	MYB proto-oncogene	-1.140	1.000	-3.210	Yes
ADIPOR2	Adiponectin receptor 2	-0.750	0.996	-2.050	Yes
MTCH2	Mitochondrial carrier 2	-0.660	0.990	-1.690	No
MDM4	MDM4, p53 regulator	-1.150	0.929	-1.210	No
PDCD4	Programmed cell death 4	-0.400	0.926	-1.930	Yes
ACO1	Aconitase 1	-0.790	0.870	-1.320	No
ELOVL3	ELOVL fatty acid elongase 3	-0.480	0.980	-1.190	No

SKP1	S-phase kinase associated protein 1	-0.670	0.877	-1.080	No
TADA1	Transcriptional adaptor 1	-0.580	0.998	-1.020	No
ZBTB4	Zinc finger and BTB domain containing 4	-0.380	0.994	-1.200	No
FOXD3	Forkhead box D3	-0.310	0.996	-1.250	No
PDE7A	Phosphodiesterase 7A	-0.380	0.980	-1.150	No
BASP1	Brain abundant membrane attached signal protein 1	-0.380	0.892	-1.200	No
SGMS1	Sphingomyelin synthase 1	-0.260	0.883	-1.280	No
CMTM6	CKLF like MARVEL transmembrane domain containing 6	-0.600	0.949	-0.860	No
ZSWIM6	Zinc finger SWIM-type containing 6	-0.180	0.932	-1.290	No
NKX2-4	NK2 homeobox 4	-0.300	0.917	-1.180	No
ZEB1	Zinc finger E-box binding homeobox 1	-0.220	0.996	-1.180	Yes
ADAM19	ADAM metallopeptidase domain 19	-0.090	0.984	-1.260	No
TEK	TEK receptor tyrosine kinase	-0.450	0.866	-0.980	No

Interrogation of miRsystem revealed 290 possible target genes for hsa-miR-150-5p across at least three databases. Analysis of gene ontology of these genes found over-representation of molecular functions such as transcription factor activity, ubiquitin protein activity, protein kinase activity and DNA binding ($p<0.0001$), as well as of biological functions including regulation of transcription, cell migration, gene expression, Wnt signaling pathway and cell secretion ($p<0.0001$). Two functionally enriched clusters were identified (Refer to Table 4.29), however, none of these genes were identified in the top twenty predicted targets from this study.

Table 4.29: Functionally enriched target gene groups of hsa-miR-150-5p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Nucleic acid binding, zinc finger, meta ion binding, transcription regulation	2.20	7

IPA analysis identified 25 target gene networks, and the top 10 of these are detailed in Table 4.30. The highest ranking canonical pathways included molecular mechanisms of cancer, GNRH signaling, HGF signaling, CDK5 signaling and Wnt/β-catenin signaling. The predominant up-stream regulators included ZFAS1 (p=2.40E-6), miR-185 (p=1.28E-4), RPLP2 (p=1.80E-4), ITGA6 (p=2.68E-4) and FOXO3 (p=5.45E-4). Others included TP63, EGFR, E2F1, IL7R, and MYC (p<0.0001). Several of the top 20 gene targets identified in this study are involved in the top 10 networks identified in IPA. These include: MYB in Group 1; ZEB1 in Group 3; ADAM19 in Group 4; ADIPOR2 and SGMS1 in Group 6; and PDCD4 and ELOL3 in Group 9.

Table 4.30: Top ten networks for hsa-miR-150-5p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Digestive system development and function, hepatic system development and function, organ morphology	24	Behaviour, nervous system development and function, small molecule biochemistry	14
Organismal functions, gene expression, behavior	23	Cancer, organismal injury and abnormalities, cellular development	14
Cellular development, cellular growth and proliferation, haematological system development and function	21	Endocrine system disorders, organ morphology, organismal injury and abnormalities	12
Cellular growth and proliferation, embryonic development, cellular development	15	Protein synthesis, cancer, cell death and survival	12
Cell death and survival, cellular development, cellular function and maintenance	15	Cell morphology, organ morphology, skeletal muscular system development and function	12

4.11 hsa-miR-155-5p

The top twenty potential target genes for hsa-miR-155-5p are listed in Table 4.31. DIANA microT-CDS identified 1090 possible targets, and miRANDA-mirSVR found 5455 possible gene targets. TargetScan identified 552 possible targets, with 592 conserved and 239 poorly conserved potential binding sites. Functional clustering analysis (DAVID) revealed one genetic functional cluster, containing BACH1, TSHZ3, HIVEP2, CEBPB, IRF2BP2, JARID2, ARID2 and ZIC3 (enrichment score 1.61).

Table 4.31: Top twenty predicted targets for hsa-miR-155-5p

Target Gene	Full Gene Name	TargetScan Aggregate P_{CT}	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
ARID2	AT-rich interaction domain 2	0.710	0.999	-2.970	Yes
HIVEP2	Human immunodeficiency virus type I enhancer binding protein 2	0.610	0.977	-2.810	Yes
BACH1	BTB domain and CNC homolog 1	0.790	0.994	-2.400	Yes
PICALM	Phosphatidylinositol binding clathrin assembly protein	0.610	0.933	-2.440	Yes
ZIC3	Zic family member 3	0.620	0.995	-2.240	Yes
GABRA1	Gamma-aminubutyric acid type A receptor alpha 1 subunit	0.370	0.959	-2.460	No
TSHZ3	Teashirt zinc finger homeobox 3	0.810	1.000	-1.950	Yes
JARID2	Jumonji and AT-rich interaction domain containing 2	0.820	1.000	-1.900	Yes
LRP1B	LDL receptor related protein 1B	0.380	0.999	-2.220	No
RBMS3	RNA binding motif single stranded interacting protein 3	0.590	0.894	-2.110	No
FOS	Fos proto-oncogene, AP-1 transcription factor subunit	0.550	0.998	-1.970	Yes
WEE1	WEE1 G2 checkpoint kinase	0.940	0.997	-1.410	Yes
FBXO30	F-box protein 30	0.500	0.997	-1.340	No
MYB1		0.410	0.766	-2.070	No
DHX40	DEAH-box helicase 40	0.710	0.998	-1.280	Yes
CEBPB	CCAAT/enhancer binding protein beta	0.770	0.990	-1.200	Yes
TMEM202	Transmembrane protein 202	<0.100	0.881	-1.960	No
TRIM23	Tripartitie motif containing 23	<0.100	0.842	-1.940	No
FBXO33	F-box protein 33	0.500	0.997	-1.340	No
IRF2BP2	Interferon regulatory factor 2 binding protein 2	0.390	1.000	-1.320	Yes

429 potential target genes were identified for hsa-miR-155-5p using miRSystem, across at least three databases. GO analysis of these genes identified over-representation of molecular functions such as transcription factor activity, protein and DNA binding, and protein kinase activity ($p<0.0001$), as well as biological functions including regulation of transcription and the cell cycle, and intracellular signaling ($p<0.0001$). Functional clustering analysis identified one enriched functional cluster (Refer to Table 4.32). Within this cluster, IRF2BP2 and HIVEP2 were also identified in the top twenty

targets identified in this study. A further gene, TMEM202, was found in another functionally enriched cluster (enrichment score 0.14).

Table 4.32: Functionally enriched target gene groups of hsa-miR-155-5p identified by gene functional classification in DAVID

Functional classification group	Enrichment score	Number of target genes
Transcription regulation, zinc finger, metal binding, DNA binding	5.25	18

IPA analysis identified 25 gene networks, and the top 10 are listed in Table 4.33. The predominant canonical pathways involved include T-cell receptor signaling, molecular mechanisms of cancer, TGF-β-signaling, IGF-1 signaling and regulation of IL-2 expression. The main up-stream regulators include miR-155 (p=7.00E-10), COL18A1 (p=6.16E-6), TGFBR2 (p=4.86E-5), miR-17 (p=1.78E-4) and PSMD10 (p=1.78E-4). Others include TGFB1, Akt, PI3K, CCND1, CDK4, MAP2K1, and EGFR (p<0.0001). Several of our top 20 predicted targets were identified to be involved in gene networks identified in IPA. These include TSHZ3 and IRF2BP2 in Group 1; FOS in Group 2; ARID2 and WEE1 in Group 3; and PICALM in Group 4. Several of the top 20 predicted targets identified in this study were identified in networks 11 (BACH1) and 12 (GABRA1 and FBXO33). Network 11 is involved in cellular functions such as cell cycle, DNA replication, recombination and repair, cellular assembly and organization, and contained 13 genes. Network 12 is involved in amino acid metabolism, small molecule biochemistry, and DNA replication, recombination and repair; and also contained 13 genes.

Table 4.33: Top ten networks for hsa-miR-155-5p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Gene expression, cellular development, haematoogical system development and function	29	Nervous system development and function, tissue morphology, cellular movement	20
Cell cycle, hair and skin development and function, organ development	25	Cancer, haemtological disease, immunological disease	16
Gene expression, cellular compromise, haematological system development and function	25	Cell morphology, renal and urological system development and function, cellular assembly and organization	15
Developmental disorder, cardiovascular disease, haematological disease	24	Cellular development, cell death and survival, embryonic development	15

Cellular development, cellular growth and proliferation, cardiovascular system development and function	24	Cancer, organismal injury and abnormalities, cell morphology	14

4.12 hsa-miR-205-3p

The top twenty predicted target genes for hsa-miR-205-3p are listed in Table 4.34. TargetScan investigation revealed 6551 possible targets, with a total of 10996 possible binding sites. miRANDA-mirSVR found 7894 potential gene targets, while DIANA microT-CDS found 2283 possible targets. Functional clustering analysis (DAVID) found two clusters of genes: Group 1 included SUZ12, NR3C2, MDM4 and MSL2 (enrichment score 0.75), while Group 2 contained PTCHD1, CDH11, TMEM144 and SLC35A3 (enrichment score 0.24).

Table 4.34: Top twenty predicted targets for hsa-miR-205-3p

Target gene	**Full Gene Name**	**TargetScan Total context ++**	**DIANA score miTG**	**miRANDA score mirSVR**	**Experimentally validated [83]**
GCC2	GRIP and coiled-coil domain containing 2	-0.060	0.964	-1.930	No
ZFYVE16	Zinc finger FYVE-type containing 16	-0.030	0.839	-1.930	No
TMEM144	Transmembrane protein 144	-0.060	0.813	-1.820	No
PLCB1	Phospholipase C beta 1	-0.060	0.854	-1.770	No
DHFR	Dihydrofolate reductase	-0.040	0.714	-1.820	No
CUL5	Cullin 5	-0.050	0.963	-1.550	Yes
NR3C2	Nuclear receptor subfamily 3 group C member 2	-0.030	0.745	-1.730	No
PTCHD1	Patched domain containing 1	-0.060	0.795	-1.620	No
SUZ12	SUZ12 polycomb repressive complex 2 subunit	-0.120	0.996	-1.350	No
MSL2	Male-specific lethal 2 homolog	-0.050	0.709	-1.700	No
FGD4	FYVE, RhoGEF and PH domain containing 4	-0.040	0.797	-1.610	No
SH3GL3	SH3 domain containing GRB2 like 3, endophilin A3	-0.170	0.944	-1.320	No
CDK19	Cyclin dependent kinase 19	-0.090	0.807	-1.440	No
MDM4	MDM4, p53 regulator	-0.050	0.940	-1.330	No

FAM35A	Family with sequence similarity 35 member A	-0.050	0.975	-1.250	No
FAM19A2	Family with sequence similarity 19 member A2	-0.020	0.772	-1.840	No
CDH11	Cadherin 11	-0.060	0.861	-1.300	No
SLC35A3	Solute carrier family 35 member 3	-0.080	0.858	-1.300	No
TRNT1	tRNA nucleotidyl transferase 1	-0.030	0.754	-1.390	No
SBF2	SET binding factor 2	-0.010	0.716	-1.320	No

miRSystem interrogation revealed 497 potential target genes for hsa-miR-205-3p across three databases. GO analysis of these genes found that biological functions such as regulation of transcription and gene expression, protein kinase activity and intracellular signal transduction were over-represented ($p<0.0001$), as well as molecular functions including enzyme and protein biding, transcription coactivator and transcription factor binding, and protein transport ($p<0.0001$). Two functionally enriched gene clusters were identified (Refer to Table 4.35), however, there were no similarities between genes in these clusters and the top twenty potential targets identified here.

Table 4.35: Functionally enriched target gene groups of hsa-miR-205-3p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Protein kinase activity, phosphorylation, intracellular signal transduction, ATP binding	4.04	7
Zinc finger, DNA binding, transcription regulation, cell division	3.18	15

IPA analysis identified 25 gene networks, and the top 10 are listed in Table 4.36. The top canonical pathways involved include RAR activation, Neuregulin signaling, HIPPO signaling and Fcy receptor-mediated phagocytosis. The top up-stream regulators include TP53 (p=1.86E-5), Estrogen receptors (p=5.43E-5), miR-34a (p=8.30E-5), CTBP2 (p=1.21E-4) and miR-205 (p=2.37E-4). Others included KRAS, MNT5A, TGFB1, and TNF ($p<0.0001$). Several genes identified in this study were similarly identified to be involved in the top gene networks identified in IPA. These include SUZ12 in Group 1; CUL5 and NR3C2 in Group 5; CDH11 in Group 6; MSL2 in Group 9; and ZFYVE16 in Group 10.

Table 4.36: Top ten networks for hsa-miR-205-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Cellular assembly and organization, RNA post-transcriptional modification, developmental disorder	31	Cellular development, cellular growth and proliferation, embryonic development	19
Cellular movement, tissue development, cardiovascular system development and function	29	Cell death and survival, connective tissue disorders, haematological disease	19
Cellular assembly and organization, cell cycle, DNA replication, recombination and repair	28	Infectious diseases, cancer, organismal injury and abnormalities	16
Lymphoid tissue structure and development, cell death and survival, cellular growth and proliferation	26	Cellular response to therapeutics, DNA replication, recombination and repair, cell morphology	15
Cancer, neurological disease, cell morphology	23	Embryonic development, organismal development, tissue development	15

4.13 hsa-miR-342-3p

The top twenty predicted gene targets of hsa-miR-342-3p are given in Table 4.37. miRANDA-mirSVR investigation yielded 7724 possible gene targets, while DIANA microT-CDS found 992 possible targets. TargetScan found 315 potential transcript targets, with a total of 317 conserved and 211 poorly conserved possible binding sites. Functional analysis (DAVID) identified one functional genetic cluster, containing BTN2A1, GOLM1, PRIMA1, MTDH, FXYD3 and CA12 (enrichment score 0.38).

Table 4.37: Top twenty predicted targets for hsa-miR-342-3p

Target gene	Full Gene Name	TargetScan Total context ++	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
ID4	Inhibitor of DNA binding 4, HLH protein	-0.510	0.977	-1.530	Yes
BTN2A1	Butyrophilin subfamily 2 member A1	-0.690	0.940	-1.370	No
TIA1	TIA1 cytotoxic granule associated RNA binding protein	-0.540	0.998	-1.420	No
UBE2D2	Uniquitin conjugating enzyme E2 D2	-0.650	0.946	-1.220	No
KDM6B	Lysine demethylase 6B	-0.410	1.000	-1.300	No

FUT8	Fucosyltransferase 8	-0.380	0.977	-1.320	No
MATN1	Matrilin 1, cartilage matrix protein	-0.520	0.933	-1.240	No
MRFAP1	Morf4 family associated protein 1	-0.530	0.801	-1.230	No
EP300	E1A binding protein p300	-0.280	0.999	-1.260	No
FXYD3	FXYD domain containing ion transport regulator 3	-0.500	0.889	-1.100	No
ZFR	Zinc finger RNA binding protein	-0.220	0.960	-1.280	No
RGS4	Regulator of G-protein signalling 4	-0.730	0.886	-0.840	Yes
FOSB	FosB proto-oncogene, AP-1 trnascription factor subunit	-0.710	0.978	-0.730	No
NBEA	Neurobeachin	-0.210	0.957	-1.200	No
GOLM1	Golgi membrane protein 1	-0.320	0.975	-0.990	No
PDCFRA		-0.230	1.000	-0.980	No
MTDH	Metadherin	-0.340	0.877	-0.910	No
CA12	Carbonic anhydrase 12	-0.370	0.997	-0.740	No
NKX2-2	NKX2-2 antisense RNA 1	-0.250	0.767	-1.080	No
TRMT2A	tRNA methyltransferase 2 homolog A	-0.020	0.854	-1.190	No

276 genes across at least three databases were identified using miRSystem to be potential targets of hsa-miR-342-3p. GO analysis of these genes found that molecular functions such as protein kinase binding, DNA binding, transcription factor activity and receptor signaling were over-represented ($p<0.0001$), as were biological functions such as cell migration, regulation of transcription, and cell proliferation and differentiation ($p<0.0001$). One functionally enriched gene cluster was identified (Refer to Table 4.38), but there were no similarities with the top twenty predicted targets identified in this study.

Table 4.38: Functionally enriched target gene groups of hsa-miR-342-3p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Transcription regulation, zinc finger, metal binding, DNA binding	2.24	10

IPA analysis identified 25 gene networks, and the top 10 are detailed in Table 4.39. The top canonical pathways included ERK5 signaling, Tissue factor in cancer, molecular mechanisms of cancer, and Ephrin receptor signaling. The predominant up-stream regulators included GH (p=2.99E-5), EXH2 (p=3.01E-5), PDGF BB (p=4.88E-5), SMAD4 (p=1.25E-4) and CTNNB1 (p=1.37E-4). Others

included IL6, BCL6, EGFR, NR3C1, BCL2 and IL6R ($p<0.0001$). Within the top 10 gene networks identified in IPA, several of our top target genes were found to be involved. These include: FOSB in Group 1; EP300 and CA12 in Group 2; MTDH and NKX2-2 in Group 3; ID4 and UBE2D2 in Group 5; MRFAP1 in Group 6; UBE2D2 and KDM6B in Group 8; and MATN1 in Group 10.

Table 4.39: Top ten networks for hsa-miR-342-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Cellular movement, cellular growth and proliferation, cell death and survival	29	Cell cycle, connective tissue development and function, hepatic system development and function	13
Cell cycle, cardiovascular disease, organismal injury and abnormalities	25	Cancer, cellular movement, organismal injury and abnormalities	13
Post-translational modification, cardiovascular disease, cardiovascular system development and function	21	Cell death and survival, connective tissue disorders, developmental disorder	13
Organ morphology, organismal development, reproductive system development and function	15	Cellular function and maintenance, molecular transport, small molecule biochemistry	13
Cancer, neurological disease, organismal injury and abnormalities	13	Embryonic development, organismal development, gastrointestinal disease	12

4.14 hsa-miR-361-3p

The top twenty predicted gene targets for hsa-miR-361-3p are listed in Table 4.40. miRANDA-mirSVR identified 7814 possible target genes, while DIANA microT-CDS found 238 potential targets. TargetScan identified 5163 possible gene targets, with a total of 7850 potential binding sites. Functional gene clustering analysis (DAVID) identified one cluster, which included ENTPD3, UGT8, MFSD6, KCNMA1, CADM4, PTGER2 and GRID1 (enrichment score 0.8).

Table 4.40: Top twenty predicted targets for hsa-miR-361-3p

Target Gene	Full Gene Name	TargetScan Total context ++	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
CADM4	Cell adhesion molecule 4	-0.800	0.982	-1.280	No
PTPLA	C-hydroxyacyl-CoA dehydratase 1	-0.730	0.999	-1.320	No
KCNMA1	Potassium calcium-activated channel subfamily M alpha 1	-0.010	0.997	-1.730	No

ABI1	Abl interactor 1	-0.510	0.920	-1.190	No
HOXA10	Homeobox A10	-0.510	0.882	-1.220	No
PRAMEF8	PRAME family member 8	-0.650	0.959	-0.850	No
GGCT	Gamma-glutamylcyclotransferase	-0.580	0.798	-1.060	No
MFSD6	Major facilitator superfamily domain containing 6	-0.340	0.942	-1.140	Yes
SETD6	SET domain containing 6	-0.550	0.949	-0.890	No
MAP3K10	Mitogen-activated protein kinase kinase kinase 10	-0.550	0.795	-1.030	No
BATF1	Basuc leucine zipper ATF-like transcription factor	-0.340	0.865	-1.080	No
UGT8	UDP glycosyltransferase 8	-0.380	0.884	-1.010	No
NUBPL	Nucleotide binding protein like	-0.390	-0.981	-0.820	No
PTGER2	Prostaglandin E receptor 2	-0.360	0.720	-1.080	No
OSR2	Odd-skipped related transcription factor 2	-0.520	0.737	-0.890	No
TMEM189-YBE2V1	TMEM189-UBE2V1 readthrough	-0.330	0.790	-1.000	No
SENP1	SUMO1/sentrin specific peptidase 1	-0.420	0.869	-0.830	No
TOMM20	Translocase of outer mitochondrial membrane 20	-0.570	0.927	-0.550	No
TMEM35	Transmembrane protein 35	-0.470	0.905	-0.670	No
HOXC6	Homeobox C6	-0.320	0.932	-0.790	No

132 putative target genes were identified across at least three databases using miRSystem. Biological functions including transcription regulation, cell adhesion, synaptic transmission, and response to growth factors were over-represented on GO analysis ($p<0.0001$). Molecular functions were also over-represented including transcription factor and coactivator activity, DNA binding and protein kinase activity ($p<0.0001$). No functionally enriched gene clusters were identified.

IPA analysis found 16 gene networks, the top 10 of which are listed in Table 4.41. The top canonical pathways identified included glucocorticoid receptor signaling, calcium signaling, Ephrin receptor signaling, B-cell receptor signaling and RAR activation. Top regulators of these networks include miR-34 (p=6.27E-4), RUNX1 (p=1.35E-3), ESR2 (p=1.86E-3), miR-218 (p=2.90E-3) and DDX5 (p=2.90E-3). Others include CCDC8, JAK2, and PTEN ($p<0.0001$). Several of the top 20 target genes identified in this study were found to be involved in the top gene networks (Refer to Table 4.41). These included KCNMA1 in Group 1; MFSD6 in Group 2; OSR2 in Group 4; HOXA10 in Group 5; and ABI1 and HOXC6 in Group 7.

Table 4.41: Top ten networks for hsa-miR-361-3p target genes from Ingenuity Pathway Analysis

Network functions	Number of Genes	Network functions	Number of Genes
Gene expression, connective tissue development and function, tissue development	22	Cellular movement, cardiovascular system development and function, organismal development	11
Organismal injury and abnormalities, renal degeneration, renal and urological disease	14	Cellular function and maintenance, molecular transport, small molecule biochemistry	10
Nervous system development and function, embryonic development, organismal development	12	Developmental disorder, hereditary disorder, nervous system development and function	1
Cell cycle, cell-to-cell signaling and interaction, cellular growth and proliferation	12	Cell signaling, cancer, organismal injury and abnormalities	1
Cell morphology, cell cycle, embryonic development	11	Cell cycle, cell-to-cell signaling and interaction, cellular assembly and organization	1

4.15 hsa-miR-363-3p

The top twenty predicted gene targets for hsa-miR-363-3p are listed in Table 4.42. TargetScan found 1037 possible target genes, with a total of 1163 conserved and 323 poorly conserved binding sites. DIANA microT-CDS found 861 possible target genes, while miRANDA-mirSVR found 6386 putative gene targets. Functional clustering analysis (DAVID) identified one functional cluster, including MAN2A1, LHFPL2, CHST7, B3GALT2, SLC17A6, SLC12A5, CD69 and BTLA (enrichment score 1.02).

Table 4.42: Top twenty predicted targets for hsa-miR-363-3p

Target Gene	Full Gene Name	Targetscan Aggregate P_{CT}	DIANA score miTG	miRANDA score mirSVR	Experimentally validated [83]
CD69	CD69 molecule	>0.990	0.998	-3.660	No
MAN2A1	Mannosidase alpha class 2A member 1	>0.990	0.980	-2.740	No
FBXW7	F-box and WD repeat domain containing 7	0.980	0.992	-2.440	No
SLC12A5	Solute carrier family 12 member 5	0.980	0.992	-2.340	Yes
KLF4	Kruppel like factor 4	0.950	0.985	-2.360	No
FNIP1	Folliculin interacting protein 1	0.990	0.999	-2.180	Yes

MYO1B	Myosin 1B	0.970	0.992	-2.150	Yes
MORC3	MORC family CW-type zinc finger 3	0.910	0.971	-2.200	No
KIAA1432	RIC1 homolog, RAB6A GEF complex partner 1	0.990	0.974	-2.110	No
MAP2K4	Mitogen-activated protein kinase kinase 4	>0.990	0.989	-2.080	Yes
SLC17A6	Solute carrier family 17 member 6	0.740	0.974	-2.220	No
RBM47	RNA binding motif protein 47	0.730	0.976	-2.190	No
RNF38	Ring finger protein 38	0.860	0.974	-1.850	No
KLHL14	Kelch like family member 14	0.950	0.894	-1.820	No
RAB23	RAB23, member RAS oncogene family	0.990	0.891	-1.760	No
ARMC1	Armadillo repeat containing 1	0.940	0.956	-1.650	No
BTG2	BTG anti-proliferation factor 2	>0.990	0.943	-1.440	Yes
BTLA	B and T lymphocyte associated	0.460	0.833	-2.070	No
HIPK3	Homeodomain interacting protein kinase 3	>0.990	1.000	-1.370	No
LHFPL2	Lipoma HMGIC fusion partner-like 2	>0.990	0.955	-1.400	Yes

miRSystem identified 573 potential target genes for hsa-miR-363-3p across at least three databases. GO analysis of this gene list found that biological functions such as cell proliferation, transcription, cell adhesion and synaptic transmission ($p<0.0001$), as well as molecular functions including protein binding, protein kinase activity, transcription factor activity and DNA binding ($p<0.0001$) were over-represented. Three functionally enriched gene clusters were identified (Refer to Table 4.43). Of the predicted targets identified in this study, HIPK3 was identified in functional cluster 2.

Table 4.43: Functionally enriched target gene groups of hsa-miR-363-3p identified by gene functional classification in DAVID

Functional classification group	**Enrichment score**	**Number of target genes**
Membrane composition and function, intracellular signal transduction	4.76	10
Protein kinase activity, nucleotide binding, cytoskeleton, cell membrane	4.66	5
Zinc finger, transcription regulation, metal ion binding, cell division	3.65	10

IPA analysis identified 25 networks, and the top 10 are listed in Table 4.44. Top canonical pathways involved include synaptic long term potentiation, calcium signaling, molecular mechanisms of cancer

and 3-phosphoinositide biosynthesis. The top regulators were miR-486 (p=4.58E-10), miR-25 (p=1.51E-6), MYOC (p=9.10E-6), CST5 (p=5.47E-5) and miR-32 (p=7.35E-5). Others included TGFB1, CTBP2, JARID2 TBPL1, NR3C1, TNF and IL12B (p<0.0001). Within the top 10 gene networks identified in IPA, several of the top 20 genes identified in this study were found to be involved. These include: MAP2K4 and BTG2 in Group 1; CD69 in Group 2; FBXW7 in Group 3; RBM47 in Group 8; and KLF4 and LHFPL2 in Group 10.

Table 4.44: Top ten networks for hsa-miR-363-3p target genes from Ingenuity Pathway Analysis

Network functions	**Number of Genes**	**Network functions**	**Number of Genes**
Cellular growth and proliferation, cancer, organismal injury and abnormalities	29	Cancer, organismal injury and abnormalities, neurological disease	18
Connective tissue disorders, organismal injury and abnormalities, skeletal muscular disorders	27	Cellular assembly and organization, cellular function and maintenance, protein trafficking	18
Gene expression, cell cycle, developmental disorder	26	Dermatological diseases and conditions, inflammatory disease, inflammatory response	14
Cancer, haematological disease, organismal injury and abnormalities	25	Cardiovascular disease, congenital heart anomaly, developmental disorder	16
Cellular movement, cell-to-cell signaling and interaction, haematological system development and function	21	Gene expression, cellular growth and proliferation, organismal development	16

CHAPTER FIVE: DISCUSSION

5.1 Introduction

Squamous cell carcinoma of mucosal sites in the head and neck is a complex disease. Clinically, patients with HPV-associated tumours in the oropharynx tend to have a distinct profile: younger, and higher functioning. Further, they have also demonstrated an enhanced response to chemotherapy and radiotherapy than patients with HPV-negative tumours; and improved overall survival rates [5, 13, 14, 22]. While we have a firm understanding of the clinical consequences of a HPV-associated tumour in the oropharynx, the pathobiology of these tumours remains poorly understood. As clinical trials aimed at ensuring appropriateness of therapy, including de-escalation of therapy when suitable, are presently ongoing [1, 6], it is imperative that we fully understand the molecular mechanisms underlying carcinogenesis in both HPV-positive and HPV-negative tumours, to allow identification of patients who may require more intensive therapy. While a significant proportion of HNSCC patients tend to present with either more locally advanced disease (T stage III/IV) or with regional nodal disease at diagnosis [4, 5], as yet there is no clinically applicable tool such as a population wide screening tool, to enable an earlier diagnosis. There is enormous potential for miRNAs to fill these gaps as clinical biomarkers in head and neck cancer.

5.2 Discussion of profiling results

5.2.1 HPV-positive tumours demonstrate altered microRNA expression

In this study, a HPV-positive tumour was defined as positive for HPV DNA with positive staining of $p16^{INK4a}$ on immunohistochemistry. A number of miRNAs were differentially expressed between HPV-positive and HPV-negative tumours in the oropharynx. These were: hsa-miR-15b-5p, hsa-miR-16-2-3p, hsa-miR-20b-5p, hsa-miR-30e-5p, hsa-miR-142-3p, hsa-miR-361-3p and hsa-miR-363-3p. Several of these, hsa-miR-15b-5p, hsa-miR-16-2-3p, hsa-miR-20b-5p and hsa-miR-363-3p have consistently been reported as upregulated across previous studies investigating HPV-associated HNSCC (Refer to Table 5.1) [12, 55, 58, 66, 89]. Two miRNAs which have been identified as differentially expressed in this study have not been previously reported in HPV-positive head and neck cancers, hsa-miR-30e-5p and hsa-miR-361-3p. Finally, hsa-miR-142-3p was identified as up-regulated in HPV-positive tumours in the present study, but has only previously been reported as down-regulated in these tumours in one study on head and neck cell lines [66]. This needs to be interpreted with caution as studies comparing cell lines with human tissue have identified differences in cellular miRNA expression; as the cells may not accurately depict in-vivo intra-cellular changes

[57]. Further studies will determine whether miR-142 is up- or down-regulated in HPV-associated HNSCC.

Table 5.1: Comparison of miRNA profiling results with previous literature

miRNA	Direction of Change	Up-regulated	Down-regulated
hsa-miR-15b	Up	Lajer et al [12], Miller et al [89]	
hsa-miR-16	Up	Lajer et al [12], Miller et al [89]	
hsa-miR-20b	Up	Lajer et al [12], Hui [90], Miller et al [89]	
hsa-miR-30e	Up		
hsa-miR-142	Up		Wald et al [66]
hsa-miR-361	Up		
hsa-miR-363	Up	Lajer et al [58], Lajer et al [12], Wald 2011 [66], Miller et al [89]	

Two miRNAs were identified as up-regulated across all subsites in HPV-positive tumours when compared with HPV-negative tumours. These were: hsa-miR-16-2-3p and hsa-miR-20b-5p. The majority of studies looking at the differing miRNA expression in HPV-positive tumours have only included tumours of the oropharynx (Refer to Table 1.1), but Lajer et al [12] also identified these two miRNAs to be differentially expressed across all sites, as in the present study.

5.2.2 MicroRNA alteration and $p16^{INK4a}$ status

In the literature, the definition of a HPV-associated tumour is variable, and can include the use of $p16^{INK4a}$ status alone, in combination with HPV DNA status, HPV DNA status alone, or E6 and E7 mRNA status. The latter of these is considered the gold standard for defining HPV status in research, although use of a combination of HPV DNA and $p16^{INK4a}$ immunohistochemistry has been shown to be equivalent in detecting clinically significant infections [91]. Clinically however, only the $p16^{INK4a}$ status of a tumour is used as a surrogate marker for HPV status. This study identified only one miRNA which was dysregulated between tumours positive and negative for $p16^{INK4a}$ across all subsites: miR-16. There was no difference in miRNA expression with $p16^{INK4a}$ status at any individual subsite, including the oropharynx. Hui et al [90] defined a HPV-positive tumour as one that is positive for $p16^{INK4a}$ only, but only identified miR-9 and miR-20b to be dysregulated.

5.2.3 HPV-positive smokers demonstrate altered microRNA expression

HPV-positive tumours occurring in patients with a history of heavy tobacco smoking are a distinct entity. A history of cigarette smoking both increases the risk of developing a HPV-related tumour [7, 92], and negatively influences survival [8, 93, 94]. The present study has identified hsa-miR-20b-5p to be over-expressed in smokers with oropharyngeal tumours, and hsa-miR-363-3p to be over-expressed in smokers across all subsites. In a multi-variate analysis of HPV status and smoking status, only hsa-miR-150-5p was significantly differentially expressed, being up-regulated in oropharyngeal HPV-positive tumours. To date, potential differences in miRNA expression with smoking status in combination with HPV tumour status has not specifically been addressed in the literature. However, several studies have investigated the difference in miRNA expression between smokers and non-smokers. Some authors [95] have identified no difference between the two groups, while many, with particular focus on lung cancers, have identified a number of differences in tumours that have exposure to nicotine [96].

5.2.4 Alcohol and association with microRNA expression

A number of miRNAs were identified to be differentially regulated in tumours of patients who were heavy alcohol consumers, compared with patients who abstained from alcohol consumption. Several studies have investigated the relationship of miRNA expression to alcohol use in HNSCC patients. In head and neck cancers, miR-375 has previously been shown to be correlate significantly with alcohol consumption [64]. Another study found long-term exposure of tumour cells to ethanol resulted in up-regulation of miR-30a, miR-934, miR-3164 and miR-3178 [97]. However, while we identified several miRNAs which were differentially expressed in HPV-related tumours in patients who also consumed alcohol in a multivariate analysis, there is no previous literature with which to compare these findings. Further research is therefore warranted to confirm or refute these results.

5.2.5 Association of microiRNA expression with recurrence of disease

Several miRNAs were identified to be associated with recurrence of the primary tumour. When compared to the previous literature, only miR-205 has previously been reported as dysregulated in association with loco-regional recurrence of disease [60, 61]. However, very few studies have evaluated miRNA expression in the context of recurrence of the primary tumour, and therefore although miR-205 demonstrates promise in that it has been consistently reported across several studies, further research in this area is needed.

5.2.6 MicroRNAs and association with survival

Both disease-free and recurrence-free survival were analysed in this study, and a number of miRNAs were identified to be differentially expressed in each analysis. Unfortunately, as this study focused on evaluating differential expression between HPV-positive and negative tumours, the miRNAs chosen for further validation in this study are not ones which have been reported to correlate with survival and prognosis previously in HNSCC studies. Therefore, it is difficult to draw conclusions from comparison of the current results with those of previous studies. However, this study has identified several miRNAs which demonstrate promise as prognostic markers. In particular, hsa-miR-16-2-3p, hsa-miR-20b-5p and hsa-miR-363-3p. hsa-miR-16-2-3p was found to correlate with RFS in oropharyngeal tumours both when HPV status was considered and when it was not, possibly indicating that this miRNA may be suitable for further research as a prognostic biomarker in oropharyngeal SCCs. DFS was shown to correlate with expression of hsa-miR-20b-5p when HPV tumour status is accounted for across all subsites and in the oral cavity; however, interestingly, only when HPV status was not considered in analysis of oropharyngeal tumours. Again, this warrants further study, and consideration as a potential prognostic biomarker. Finally, in HPV-negative oral cavity tumours; apart from smoking status, currently, there is clinically no way of determining which patients will have a poorer prognosis. This study identified that miR-363 correlated significantly with DFS in these patients, again potentially indicating that this miRNA may hold promise for use as a prognostic biomarker in this tumour cohort.

5.3 The genomic landscape of HNSCC

A number of studies have focused on genomic profiling of head and neck cancers in recent years, and a smaller number again have investigated the difference in genomic differences between HPV-positive and HPV-negative tumours. In particular, analysis of the cancer genome atlas of head and neck cancers (TGCA), which is ongoing currently, will provide the most comprehensive and integrative genomic profile of head and neck cancers [98-100]. There are a number of genes which are differentially expressed between HPV-positive and HPV-negative tumours; as well as several commonly dysregulated intracellular signaling pathways which are abnormally regulated in HPV-associated HNSCC when compared with HPV-unrelated tumours. These pathways are important for future research to focus on, as if potentially drugable gene targets are identified, this may have important clinical implications for therapeutics in HPV-related head and neck cancer. It is clear from both this study and the literature that as a result of the nature of miRNA::mRNA interactions, one miRNA is able to target a number of mRNA strands, and similarly, one mRNA is potentially targeted by a number of miRNAs. This biological redundancy enables multiple points within a single cellular

pathway to be disrupted, with subsequently multiple points at which steps towards carcinogenesis may occur.

In this study, IPA pathway analysis was performed in order to determine the most commonly potentially targeted intra-cellular pathways by the miRNAs evaluated. Across all miRNAs, there were commonalities across targeted pathways. These include networks of DNA replication, recombination and repair; gene expression; cellular organization; intracellular and cell-to-cell signaling; cell growth and differentiation; and cell movement. The normal function of these biological processes within cells is essential in maintaining normal cellular function; and it is clear that disruption of these processes eventually results in the hallmarks of cancer, and carcinogenesis. Dysregulation of miRNAs, as demonstrated in HPV-associated tumours in this study, have the ability to disrupt these cellular functions through the pathways described, among others, and therefore potentially represent a mechanism of carcinogenesis in these tumours. This deserves further study in detail with experimental demonstration of gene expression in relation to miRNA expression.

Many genes identified as potential miRNA targets in this study participate in intracellular networks known to be disrupted in HPV-related HNSCC, including cell cycle regulation and p53 signaling pathways, the PI3KCA-AKT pathway, NOTCH signaling pathway, and β-catenin-Wnt pathway [98-104]; and these pathways were identified in the top canonical pathways affected by several miRNAs investigated in this study. Many of these putative targets have also been experimentally validated by other authors, as listed in MiRTarBase (Refer to Table 5.2) [83]. However, whether miRNA dysregulation alone acts as a driver of carcinogenesis in head and neck cancers; or occurs as a function of genomic instability or HPV oncoprotein function with subsequent cell cycle dysfunction, remains to be experimentally determined. Regardless of this however, the evident differences in miRNA expression and gene expression between HPV-related and unrelated tumours allows for further study into the clinical applicability of these differences in the diagnostics, therapeutics and prognostics of head and neck cancer.

Table 5.2: A comparison of major genes participating in cellular pathways commonly dysregulated in HPV-related HNSCC, comparing results obtained from IPA interrogation (Refer to Chapter Four), with experimentally validated target genes for each miRNA (obtained from MiRTarBase [83]).

Cellular pathway	**Target genes (IPA) and corresponding putative target miRNAs**		**MiRTarBase identified experimentally validated miRNAs for corresponding target gene**
	RB	miR-20b	miR-20b

Cell cycle regulation	E2F	miR-15, miR-16, miR-20b, miR-30e, miR-148b, miR-150, miR-205, miR-363	miR-15, miR-16, miR-20b, miR-205
	WEE1	miR-15, miR-16, miR-20b	miR-15, miR-16, miR-20b
	Cyclin E	miR-15, miR-16, miR-30e, miR-363	miR-15, miR-16
	Cyclin D1	miR-15, miR-16, miR-20b, miR-29a, miR-29c, miR-150, miR-155	miR-15, miR-16, miR-20b, miR-29a, miR-155
	CDK4/6	miR-15, miR-16, miR-29a, miR-29c, miR-148b, miR-363	miR-15, miR-16, miR-29a
NOTCH	NOTCH	miR-30e, miR-205, miR-150, miR-16, miR-29c, miR-29a, miR-15,	miR-15, miR-16, miR-30e, miR-150, miR-205
	MAM	miR-30e,	miR-30e
	NUMB	miR-146a,	miR-146a
	FURIN	miR-16, miR-15, miR-20b,	miR-15, miR-16
PI3KCA-Akt	AKT	miR-15, miR-16, miR-20b, miR-29c, miR-150	miR-15, miR-16, miR-20b, miR-29c
	PI3K	miR-15, miR-16, miR-20b, miR-29a, miR-29c, miR-30e, miR-142, miR-148b, miR-150, miR-363	
	PTEN	miR-20b, miR-29a, miR-29c, miR-148b, miR-205, miR-363	miR-20b, miR-29a, miR-29c, miR-205
	JAK	miR-20b, miR-30e	miR-20b, miR-30e
	BCL-2	miR-15, miR-16, miR-20b, miR-30e, miR-205	miR-15, miR-16, miR-20b, miR-205
	Integrin	miR-15, miR-16, miR-20b, miR-29a, miR-29c, miR-30e, miR-148b, miR-150, miR-363	miR-15, miR-16, miR-20b, miR-29a, miR-29c, miR-30e, miR-148b, miR-150, miR-363
B-catenin-Wnt	Frizzled	miR-363, miR-342, miR-30e, miR-150, miR-148b, miR-146a, miR-16, miR-29c, miR-29a, miR-15, miR-20b,	miR-15, miR-16, miR-20b, miR-30e, miR-148b, miR-363
	TGFBR	miR-363, miR-342, miR-30e, miR-155, miR-150, miR-148b, miR-142, miR-16, miR-29c, miR-29a, miR-15, miR-20b,	miR-15, miR-16, miR-20b, miR-142
	SOX	miR-363, miR-342, miR-30e, miR-155, miR-150, miR-148b, miR-142, miR-16, miR-29c, miR-29a, miR-15, miR-20b,	miR-15, miR-16, miR-20b, miR-30e, miR-148b, miR-150, miR-363
	Wnt	miR-342, miR-155, miR-148b, miR-16, miR-15, miR-20b,	miR-15, miR-16, miR-148b, miR-155

	Groucho	miR-20b, miR-15, miR-16, miR-155, miR-150, miR-205,	miR-15, miR-16
	RAR	miR-361, miR-30e, miR-205, miR-146a, miR-142, miR-16, miR-29c, miR-29a, miR-15, miR-20b,	miR-15, miR-16

5.4 Future directions of research

There are a number of directions in which future research in this area can go. One of the first projects to arise from this study would be the validation of the putative target genes put forward in this thesis as experimentally validated targets of each respective miRNA. In squamous cell carcinoma cell lines, after confirming the expression of the miRNA in question using qRT-PCR, a number of assays can be performed to determine whether the gene in question is a valid target. Western blot analysis can be used to determine the presence and level of gene expression. Dual luciferase reporter assays can be utilised following this to confirm the gene as a miRNA target, by measuring and comparing luciferase activity when cells are transfected with anti-miR-x (or double stranded siRNA) and miR-x. Western blot can further be applied to measure gene expression in cells transfected with anti-miR-x (knockdown of the miR in question). Cell proliferation and apoptosis assays could also be performed to determine any effect of miR-x expression or knockdown on gene expression and subsequently on cellular function.

5.4.1 Diagnostics

Another possible avenue for further research in this area involves the development of a screening tool. There is an urgent need for a screening tool to be developed as although HPV associated tumours have an improved survival when compared with HPV unrelated tumours, a small number of these patients do not [8]. The majority of patients continue to present at a more advanced stage at diagnosis. As a consequence, the overall mortality for head and neck cancer remains at approximately 50%, despite advances in treatment over the past decade; as well as the ongoing significant morbidity associated with treatment [1, 6]. A screening tool which can be widely applied to the general population has the potential to dramatically improve both morbidity and mortality rates by enabling earlier detection of tumours.

The ability for a set of miRNAs to distinguish normal human tissue from tumour tissue has been well demonstrated [52, 105]. MiRNAs are unique as tissue biomarkers as they are secreted by cells as exosomes or extracellular vesicles, and remain stable in body fluids [106, 107]. The combination of

these discoveries means that miRNAs are well suited as readily available and less invasive (than tissue biopsies) biomarkers for disease diagnosis. In several human malignancies, studies have been performed to demonstrate the effectiveness of miRNAs as diagnostic biomarkers. For example, in cholangiocarcinoma, a panel of miRNAs in bile predicts early stage disease with a sensitivity of 67% and a specificity of 96%; and when combined with CA19-9 the sensitivity increased to 96% [108]. In lung cancer a miRNA signature has been found to have a similar sensitivity and specificity to CT imaging for the detection of primary tumours in at-risk cohorts; while a further set of miRNAs has been found to distinguish between early tumours and benign nodules [109-111].

In HNSCC, studies have demonstrated differences between tumour and normal tissue in the head and neck [57]; as well as the ability to differentiate HNSCC tumour origin in lymph node metastasis of unknown primary [51, 65]. Focus on the search for a screening tool is shifting to saliva as a readily accessible body fluid, and although research lacks behind that of other human malignancies, there is promise in the search for an accurate screening tool [70, 112]. To determine whether a panel of miRNAs can be utilised as a screening tool, the panel identified in tumours would first need to be compared with a cohort of normal tissue, such as normal tonsillar tissue. However, this is difficult to obtain because of the risks obtained with removing tissue from a normal, control patient. Another way to get around this would be to identify a panel of miRNAs in a body fluid such as saliva or plasma, with a comparison of expression between patients with HPV associated and non-associated tumours, and normal controls. If a difference is observed between the three cohorts, pre-malignant lesions can be introduced.

5.4.2 Therapeutics

The obvious potential application of target gene validation for any given miRNA is the future development of novel therapeutic agents. The most appealing property of miRNAs therapeutically is their ability to target a number of different mRNAs and intracellular pathways, resulting in efficient regulation of particular biological processes. There are two mechanisms currently in practice for therapeutic targeting of miRNA expression in cancer. Indirectly, drugs that target miRNA transcription or processing can be used to modulate miRNA expression; or oligonucleotides can be used to either introduce a lost tumour suppressor miRNA or silence the expression of an oncogenic miRNA [113]. There are a number of siRNA-based therapeutics currently in pre-clinical development or undergoing clinical trials. For example, a systemically delivered siRNA targeting p53 has completed a phase I trial in prevention of delayed graft function in patients undergoing kidney transplantation, and is currently in phase II trials (results unpublished, Quark Pharmaceuticals) [113]. A systemically delivered siRNA first commenced clinical trials for a solid human cancer, melanoma,

in 2008; however, this trial was ceased due to toxicities [114]. Since then, a number of siRNA therapeutics have been formulated. In pancreatic adenocarcinoma, a siRNA targeting PKN3 has been shown to cause stabilization or regression of disease with no toxicities (Silence Therapeutics) [115]. Clinical trials are also underway for siRNAs targeting MYC in hepatocellular carcinoma, solid tumours and non-Hodgkin lymphoma (Dicerna Pharmaceuticals) [113]. Although this area has yet to develop in head and neck cancer, there has been clear progress in miRNA therapeutics in a number of human cancers; and the therapies currently undergoing clinical trials hopefully will also be applicable to this specialty in the future.

5.4.3 Prognostication

This study has identified several miRNAs which are associated with recurrence of primary disease and disease-specific survival in HPV-associated head and neck cancer. Several other studies, as outlined above, have also reported similar findings although the literature lacks consistency given the small number of studies. Regardless, miRNAs have clear promise as prognostic biomarkers in head and neck cancer, and this deserves further study. In an era of therapy de-intensification, a miRNA profile associated with recurrence, progression of disease or survival can be utilised in clinical practice to risk-stratify patients, allowing accurate identification of those patients who still require maximal therapy in order to improve their clinical outcomes. This cohort for example, would include the approximately 20% of HPV-positive tumours which have a higher recurrence rate and a poorer prognosis [1, 18].

Research into the use of miRNAs as prognostic biomarkers is far more advanced in other solid human cancers, such as melanoma, breast cancer and lung cancer. For example, Australian studies in Stage IV melanoma patients have identified a panel of miRNAs which associate with progression of disease and prognosis [78, 79]. This panel is readily identifiable in plasma, and could potentially be applied clinically [78]. Similarly, studies in lung cancer are far more advanced, and sets of prognostic miRNAs which are identifiable in plasma have been identified and correlated across multiple studies [116-118]. Over-expression of miR-21 has been identified as a predictor of poorer prognosis across a range of solid and haematological human cancers, including pancreatic and colon adenocarcinoma [119, 120], chronic lymphocytic leukaemia [50, 121], lung cancer [122], as well as HNSCC [62, 64].

In order to develop a prognostic biomarker panel from the results of the present study, the first step will be to identify both the presence of these miRNAs in the saliva or plasma of patients with a recurrence of their tumour, and their absence in normal controls and patients without recurrence. Then a feasibility study could be performed, measuring a miRNA panel in the saliva or plasma of HPV-

associated HNSCC patients, with collections at diagnosis, and at spaced intervals (with clinical follow-up) for a period of over two years. The majority of recurrences occur within the first two years, and therefore to maximize the number of patients with a recurrence of their disease, the study should be continued for at least this period of time.

5.5 Discussion of study

5.5.1 Strengths of study

The main strength of the present study is the sample size, being the largest investigating miRNA expression in head and neck squamous cell carcinoma to date, as well as the largest study investigating miRNA expression in HPV positive and negative tumours. It is also the first study to use the Fluidigm PCR system for miRNA validation in the study of head and neck cancers; and the first study of this kind in Australia. The methods utilized in these experiments are robust and well validated in the literature; and the use of RNU-48 as an endogenous control and performance of all assays in quadruplicate for each sample enhanced this. We used the most recent and up to date version of miRBase (v.21), enabling identification on microarray of all identified human miRNAs to date. Finally, the data collection was robust and the medical questionnaires completed by patients were extensive in the number and detail of variables obtained. Medical data was also thoroughly collected to ensure that recurrence and mortality data was up to date (2017) at the time of submission. Longer follow-up of patients included in the study, of over ten years from initial data collection (2004 – 2017), ensured that a higher number of events were detected, lending quality to our results.

5.5.2 Limitations of study

Although large cohort numbers is a strength of this study, we had smaller numbers of HPV positive tumours at subsites outside of the oropharynx, which although unfortunate, is a reflection of the clinical demographics of HPV associated tumours. However, this does significantly limit the interpretation of results at these other subsites. Secondly, no normal human tissue was included for comparison with the tumour miRNA expression results. Previous studies have used a variation of tissues as controls, including cell lines, adjacent 'normal' tissue, and additionally sampled 'normal' mucosa. The difficulty with using 'normal' tissue obtained from directly adjacent to the tumour is that there have not been any studies investigating whether the miRNA expression in this 'normal' appearing tissue is significantly similar to healthy squamous oral mucosa. For this reason, and given that the aim of this project was to directly compare the HPV-positive and negative tumours for miRNA expression, normal tissue controls were not included in the study design.

Finally, the identification of putative target genes for each miRNA was heavily reliant on three online target gene prediction algorithms, which are detailed above. Although we used a combination of these databases with the aim of improving target prediction accuracy, these are notoriously inaccurate, and are not all equally up to date. As a result, any identified potential target genes should be experimentally validated before being considered a target for any given miRNA; and the results of target gene prediction for each investigated miRNA therefore open up many future avenues for experimental target gene validation.

5.6 Conclusions

In conclusion, this study has identified a number of miRNAs which are differentially expressed in HPV-positive tumours, when compared with HPV-negative tumours. Differential miRNA expression was also demonstrated with lifestyle factors such as tobacco smoking and alcohol consumption, as well as with dysregulation of cell cycle factors including $p16^{INK4a}$ and cyclin D1. MiRNAs were also identified to be associated with clinical features including increasing stage at diagnosis, recurrence of disease and disease-specific survival. There are a number of directions in which future research can progress on the basis of these findings, as described above, although two important avenues would be the development of a screening tool to aid earlier diagnosis; and the development of a miRNA panel for prognostication, in order to identify those patients who will require intensive therapy as opposed to those who would be suitable for treatment de-intensification. HPV-related tumours of mucosal head and neck sites are clearly distinct in both clinical behavior and in the molecular mechanisms underlying tumour development, and this study has added to our understanding of the latter.

Supplementary Table 3.1: Microarray results for differentially expressed miRNAs between HPV-positive and negative tumours at a fold change of >2.0 and p value <0.01. This revealed forty-nine biologically significant miRNAs which are differentially expressed, nine of which are down-regulated in HPV-related tumours, and forty of which are up-regulated.

miRNA	Fold Change	P value	Direction of change	miRNA	Fold Change	P value	Direction of change
hsa-miR-et-7g-5p	2.16	0.001	Up	hsa-miR-205-3p	6.50	7.94E-4	Up
hsa-miR-15a-5p	5.64	1.26E-4	Up	hsa-miR-221-5p	2.81	0.005	Up
hsa-miR-15b-3p	2.80	0.005	Up	hsa-miR-340-5p	3.06	0.006	Up
hsa-miR-15b-5p	2.67	2.81E-4	Up	hsa-miR-342-3p	3.74	9.47E-6	Up
hsa-miR-16-2-3p	8.73	3.70E-5	Up	hsa-miR-342-5p	4.03	0.001	Up
hsa-miR-16-5p	2.21	2.62E-4	Up	hsa-miR-361-3p	7.12	2.96E-4	Up
hsa-miR-20b-5p	5.17	0.001	Up	hsa-miR-363-3p	7.42	1.66E-4	Up
hsa-miR-26b-5p	2.28	0.005	Up	hsa-miR-425-5p	2.86	0.002	Up
hsa-miR-29a-3p	2.75	3.90E-4	Up	hsa-miR-497-5p	2.45	0.008	Up
hsa-miR-29b-3p	4.43	0.001	Up	hsa-miR-664b-3p	3.11	0.006	Up
hsa-miR-29c-3p	3.85	0.001	Up	hsa-miR-1224-5p	3.47	0.006	Down
hsa-miR-29c-5p	5.18	0.005	Up	hsa-miR-1268a	2.33	0.001	Down
hsa-miR-30e-5p	2.60	3.90E-4	Up	hsa-miR-1271-5p	3.27	0.009	Up
hsa-miR-34a-5p	2.16	0.002	Up	hsa-miR-1275	2.09	0.002	Down
hsa-miR-101-3p	4.35	0.009	Up	hsa-miR-3607-3p	2.39	5.72E-5	Up
hsa-miR-140-5p	2.70	0.004	Up	hsa-miR-3653-5p	4.64	0.003	Up
hsa-miR-142-3p	9.87	3.55E-4	Up	hsa-miR-3937	3.85	0.005	Up
hsa-miR-142-5p	8.77	1.50E-4	Up	hsa-miR-4484	3.06	6.44E-4	Down
hsa-miR-146a-5p	3.06	3.12E-6	Up	hsa-miR-6722-3p	2.74	0.003	Down
hsa-miR-146b-5p	2.03	0.006	Up	hsa-miR-6754-5p	6.59	0.002	Down
hsa-miR-148b-3p	4.56	0.006	Up	hsa-miR-6756-5p	3.22	0.006	Down
hsa-miR-150-5p	6.55	2.10E-5	Up	hsa-miR-6763-5p	6.42	0.004	Down
hsa-miR-155-5p	4.13	5.83E-6	Up	hsa-miR-6771-5p	3.81	0.006	Down
hsa-miR-192-5p	4.27	0.003	Up	hsa-miR-7848-3p	3.51	0.005	Up

Supplementary Table 3.2: Microarray results for differentially expressed miRNAs in $p16^{INK4a}$ positive compared with $p16^{INK4a}$ negative tumours in the oropharynx, with a fold change of >2.0 and p value of <0.05. This revealed forty-five biologically significant miRNAs which are differentially expressed; three of which are down-regulated in $p16^{INK4a}$ positive tumours, and forty-two of which are up-regulated.

miRNA	Fold change	p value	Direction of change	miRNA	Fold change	p value	Direction of change
hsa-let-7g-5p	2.11	6.93E-03	Up	hsa-miR-150-5p	7.23	5.06E-05	Up
hsa-miR-10a-5p	3.43	3.21E-02	Up	hsa-miR-155-5p	4.14	2.38E-05	Up
hsa-miR-15a-5p	5.67	1.93E-03	Up	hsa-miR-192-5p	3.52	3.84E-02	Up
hsa-miR-15b-3p	2.81	2.14E-02	Up	hsa-miR-195-5p	2.93	3.63E-02	Up
hsa-miR-15b-5p	2.40	9.58E-03	Up	hsa-miR-205-3p	5.23	6.04E-03	Up
hsa-miR-16-2-3p	7.44	4.12E-04	Up	hsa-miR-221-5p	2.82	1.37E-02	Up
hsa-miR-16-5p	2.10	4.33E-03	Up	hsa-miR-331-3p	2.92	4.29E-02	Up
hsa-miR-19b-3p	3.84	4.81E-02	Up	hsa-miR-340-5p	2.63	2.57E-02	Up
hsa-miR-20b-5p	3.97	2.67E-02	Up	hsa-miR-342-3p	3.33	2.41E-04	Up
hsa-miR-26b-5p	2.22	4.81E-02	Up	hsa-miR-342-5p	4.04	2.91E-03	Up
hsa-miR-29a-3p	2.74	3.42E-03	Up	hsa-miR-361-3p	6.94	4.40E-04	Up
hsa-miR-29b-3p	4.46	9.37E-03	Up	hsa-miR-363-3p	6.13	9.65E-04	Up
hsa-miR-29c-3p	4.30	2.27E-03	Up	hsa-miR-374c-5p	3.81	3.87E-02	Up
hsa-miR-29c-5p	4.73	1.85E-02	Up	hsa-miR-425-5p	2.13	1.65E-02	Up
hsa-miR-30e-3p	3.67	2.50E-02	Up	hsa-miR-664b-3p	3.27	3.51E-02	Up
hsa-miR-30e-5p	2.54	5.26E-03	Up	hsa-miR-664b-5p	4.97	2.65E-02	Up
hsa-miR-34a-5p	2.22	1.08E-02	Up	hsa-miR-3607-3p	2.40	3.23E-04	Up
hsa-miR-125b-2-3p	3.29	4.83E-02	Up	hsa-miR-3653-5p	5.18	1.86E-03	Up
hsa-miR-125b-5p	2.29	4.98E-02	Up	hsa-miR-5001-5p	2.24	3.98E-02	Down
hsa-miR-140-5p	2.34	4.72E-02	Up	hsa-miR-5006-5p	3.37	2.19E-02	Down
hsa-miR-142-3p	10.48	1.24E-03	Up	hsa-miR-6516-3p	3.09	4.49E-02	Up
hsa-miR-142-5p	7.14	2.40E-03	Up	hsa-miR-7848-3p	2.27	4.05E-02	Down
hsa-miR-146a-5p	2.88	8.31E-05	Up				

Supplementary Table 3.3: Microarray results for differentially expressed miRNAs in $p16^{INK4a}$ positive compared with $p16^{INK4a}$ negative tumours in the oropharynx, with a fold change of >2.0 and p value of <0.001. This revealed eighteen biologically significant miRNAs which are differentially expressed; one of which was down-regulated in $p16^{INK4a}$ positive tumours, and seventeen of which are up-regulated.

miRNA	Fold Change	p value	Direction of change
hsa-miR-15a-5p	5.66	0.005	Up
hsa-miR-16-2-3p	7.43	0.002	Up
hsa-miR-16-5p	2.10	0.009	Up
hsa-miR-29a-3p	2.73	0.007	Up
hsa-miR-29c-3p	4.30	0.005	Up
hsa-miR-30e-5p	2.53	0.009	Up
hsa-miR-142-3p	10.48	0.004	Up
hsa-miR-142-5p	7.14	0.005	Up
hsa-miR-146a-5p	2.87	9.42E-4	Up
hsa-miR-150-5p	7.22	9.09E-4	Up
hsa-miR-155-5p	4.14	8.82E-4	Up
hsa-miR-342-3p	3.32	0.002	Up
hsa-miR-342-5p	4.03	0.009	Up
hsa-miR-361-3p	6.93	0.003	Up
hsa-miR-363-3p	6.13	0.005	Up
hsa-miR-3607-3p	2.40	0.002	Up
hsa-miR-3653-5p	5.18	0.005	Up
hsa-miR-4484	2.72	0.009	Down

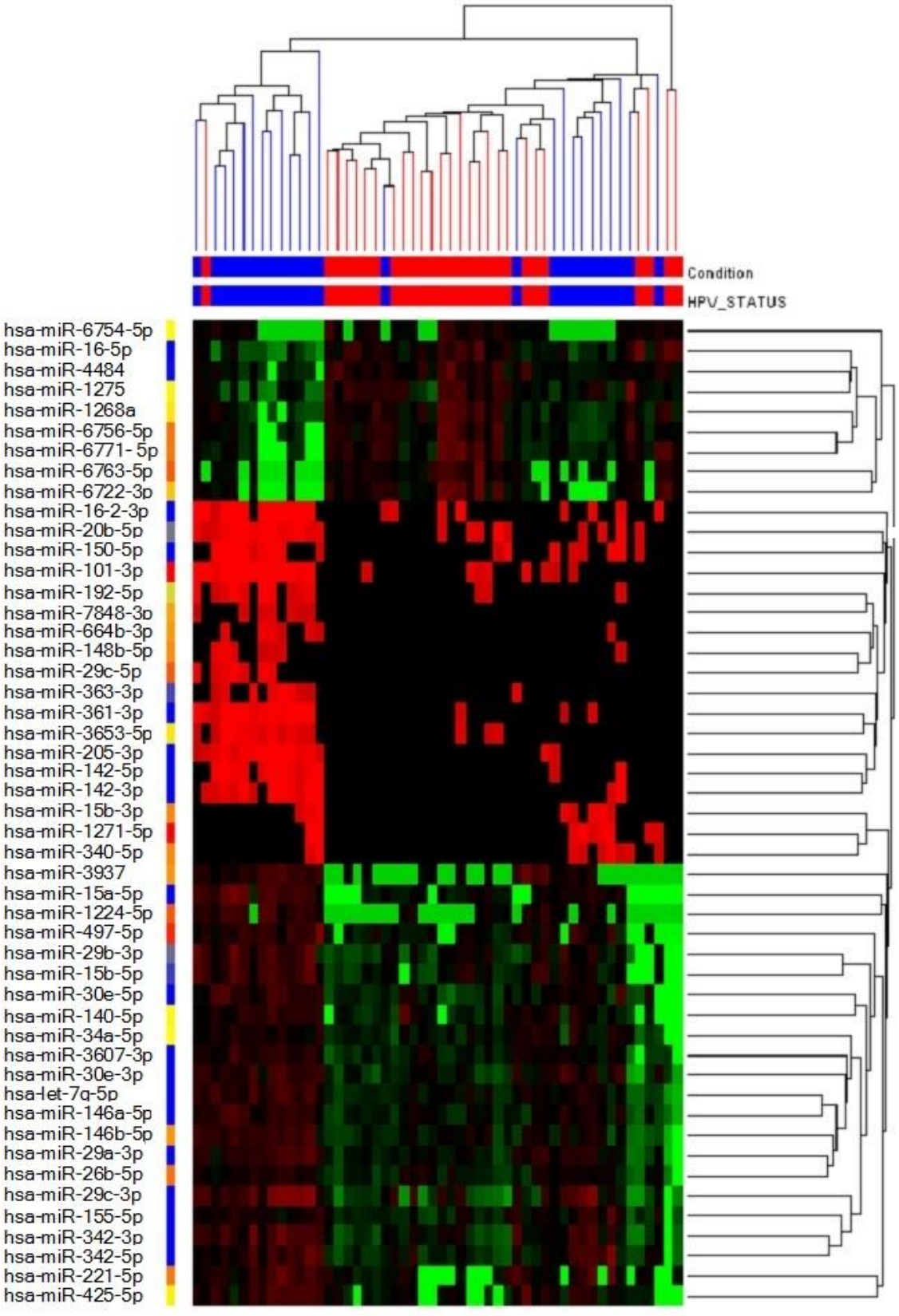
Supplementary Table 3.4: Microarray results for differentially expressed miRNAs in non-smokers, compared with smokers, in the oropharynx, with a fold change of >2.0 and p value of <0.01. This revealed fifteen biologically significant miRNAs which are differentially expressed; all of which are up-regulated in non-smokers.

miRNA	Fold Change	p value	Direction of change
hsa-miR-19a-3p	2.62	9.44E-4	Up
hsa-miR-34c-5p	2.34	8.51E-4	Up
hsa-miR-196b-3p	2.36	0.001	Up
hsa-miR-301a-3p	2.66	0.001	Up
hsa-miR-584-5p	2.41	6.05E-4	Up
hsa-miR-1245a	2.43	7.00E-4	Up
hsa-miR-1914-3p	4.81	0.002	Up
hsa-miR-3679-5p	2.40	0.003	Up
hsa-miR-4289	2.78	0.003	Up
hsa-miR-4327	2.32	0.004	Up
hsa-miR-5011-3p	2.36	0.001	Up
hsa-miR-6068	9.59	0.005	Up
hsa-miR-6735-3p	2.35	0.001	Up
hsa-miR-7155-5p	2.48	0.001	Up
hsa-miR-7162-3p	2.38	0.001	Up

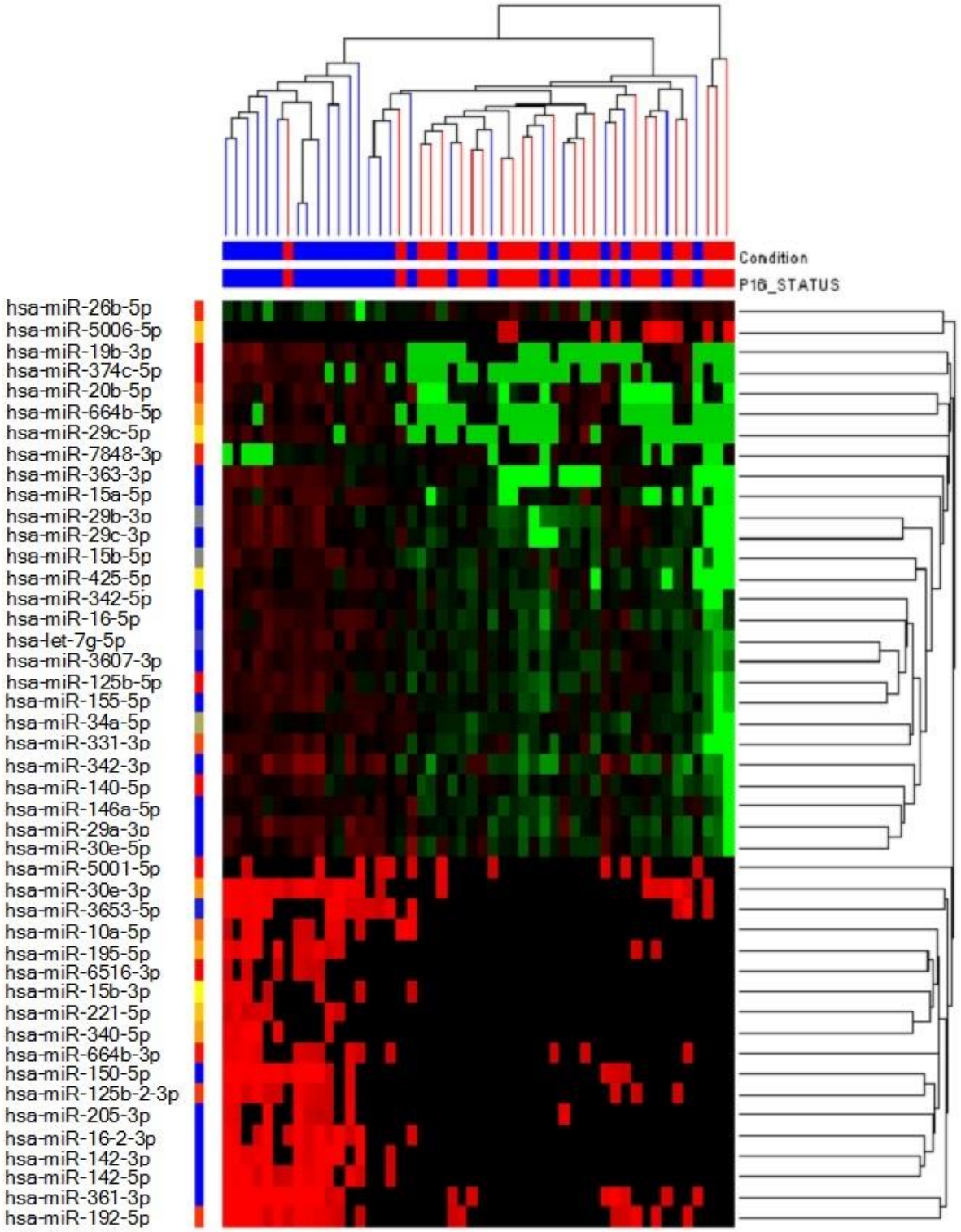
Supplementary Table 3.5: Microarray results for differentially expressed miRNAs in oropharyngeal tumours who did not have a recurrence of their primary tumour, compared with those who did; with a fold change of >2.0 and p value of <0.05. This revealed eleven biologically significant miRNAs which are differentially expressed; one of which was down-regulated in patients without a recurrence, and ten of which were up-regulated.

miRNA	Fold Change	p value	Direction of change
hsa-miR-16-2-3p	3.30	0.044	Up
hsa-miR-142-3p	4.37	0.041	Up
hsa-miR-146a-5p	2.14	0.003	Up
hsa-miR-150-5p	4.14	0.016	Up
hsa-miR-155-5p	3.06	0.007	Up
hsa-miR-205-3p	3.57	0.043	Up
hsa-miR-342-3p	2.58	0.011	Up
hsa-miR-483-5p	4.36	0.011	Down
hsa-miR-1301-5p	2.91	0.030	Up
hsa-miR-3607-3p	2.18	0.013	Up
hsa-miR-3607-5p	2.31	0.042	Up

Supplementary Figure 3.1: Heat map corresponding to Supplementary Table 3.1, visually demonstrating the difference in miRNA expression between HPV positive (blue) and HPV negative (red) oropharyngeal tumours, with a fold change of >2.0 and p <0.01. miRNAs which are up-regulated in HPV positive tumours are designated red and those which are down-regulated are designated green.



Supplementary Figure 3.2: Heat map corresponding to Supplementary Table 3.2, visually demonstrating the difference in miRNA expression between $p16^{INK4a}$ positive (red) and $p16^{INK4a}$ negative (blue) oropharyngeal tumours, with a fold change of >2.0 and p <0.05. miRNAs which are up-regulated in $p16^{INK4a}$ tumours are designated green and those which are down-regulated are designated red.



REFERENCE LIST

1. Whang, S.N., M. Filippova, and P. Duerksen-Hughes, *Recent Progress in Therapeutic Treatments and Screening Strategies for the Prevention and Treatment of HPV-Associated Head and Neck Cancer.* Viruses, 2015. **7**(9): p. 5040-65.
2. Ndiaye, C., et al., *HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis.* Lancet Oncol, 2014. **15**(12): p. 1319-31.
3. AIHW, *Head and neck cancers in Australia*, in *Cancer Series 83*. 2014: Canberra.
4. Suh, Y., et al., *Clinical update on cancer: molecular oncology of head and neck cancer.* Cell Death Dis, 2014. **5**: p. e1018.
5. Benson, E., et al., *The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas.* Oral Oncol, 2014. **50**(6): p. 565-74.
6. Dok, R. and S. Nuyts, *HPV Positive Head and Neck Cancers: Molecular Pathogenesis and Evolving Treatment Strategies.* Cancers (Basel), 2016. **8**(4).
7. Herrero, R., et al., *Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study.* J Natl Cancer Inst, 2003. **95**(23): p. 1772-83.
8. Spence, T., et al., *HPV Associated Head and Neck Cancer.* Cancers (Basel), 2016. **8**(8).
9. Syrjanen, K., et al., *Morphological and immunohistochemical evidence suggesting human papillomavirus (HPV) involvement in oral squamous cell carcinogenesis.* Int J Oral Surg, 1983. **12**(6): p. 418-24.
10. Chaturvedi, A.K., et al., *Human papillomavirus and rising oropharyngeal cancer incidence in the United States.* J Clin Oncol, 2011. **29**(32): p. 4294-301.
11. Edge, S.B. and C.C. Compton, *The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM.* Ann Surg Oncol, 2010. **17**(6): p. 1471-4.
12. Lajer, C.B., et al., *The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer.* Br J Cancer, 2012. **106**(9): p. 1526-34.
13. Tran, N., B.R. Rose, and C.J. O'Brien, *Role of human papillomavirus in the etiology of head and neck cancer.* Head Neck, 2007. **29**(1): p. 64-70.
14. Pannone, G., et al., *The role of human papillomavirus in the pathogenesis of head & neck squamous cell carcinoma: an overview.* Infect Agent Cancer, 2011. **6**: p. 4.
15. Isayeva, T., et al., *Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review.* Head Neck Pathol, 2012. **6 Suppl 1**: p. S104-20.

16. Zaravinos, A., *An updated overview of HPV-associated head and neck carcinomas.* Oncotarget, 2014. **5**(12): p. 3956-69.
17. Ang, K.K., et al., *Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III to IV head and neck carcinoma: RTOG 0522.* J Clin Oncol, 2014. **32**(27): p. 2940-50.
18. O'Sullivan, B., et al., *Deintensification candidate subgroups in human papillomavirus-related oropharyngeal cancer according to minimal risk of distant metastasis.* J Clin Oncol, 2013. **31**(5): p. 543-50.
19. Porceddu, S.V.M., R.; Brown, E.; Bernard, A.; Rahbari, R.; Foote, M.; Cartmill, B.; McGrath, M.; Coward, J.; Panizza, B., *Validation of the ICON-S staging for HPV-associated oropharyngeal carcinomas using a pre-defined treatment policy.* Oral Oncol, 2017. **Mar**(66): p. 81-86.
20. de Almeida, J.R., et al., *Oncologic Outcomes After Transoral Robotic Surgery: A Multi-institutional Study.* JAMA Otolaryngol Head Neck Surg, 2015. **141**(12): p. 1043-51.
21. Sinha, P., et al., *Transoral laser microsurgery for oral squamous cell carcinoma: oncologic outcomes and prognostic factors.* Head Neck, 2014. **36**(3): p. 340-51.
22. Vidal, L. and M.L. Gillison, *Human papillomavirus in HNSCC: recognition of a distinct disease type.* Hematol Oncol Clin North Am, 2008. **22**(6): p. 1125-42, vii.
23. Doorbar, J., et al., *Human papillomavirus molecular biology and disease association.* Rev Med Virol, 2015. **25 Suppl 1**: p. 2-23.
24. Miller, D.L., M.D. Puricelli, and M.S. Stack, *Virology and molecular pathogenesis of HPV (human papillomavirus)-associated oropharyngeal squamous cell carcinoma.* Biochem J, 2012. **443**(2): p. 339-53.
25. Antonsson, A., et al., *Prevalence and risk factors for oral HPV infection in young Australians.* PLoS One, 2014. **9**(3): p. e91761.
26. Kreimer, A.R., et al., *Incidence and clearance of oral human papillomavirus infection in men: the HIM cohort study.* Lancet, 2013. **382**(9895): p. 877-87.
27. Gillison, M.L., et al., *Prevalence of oral HPV infection in the United States, 2009-2010.* JAMA, 2012. **307**(7): p. 693-703.
28. Steinau, M., et al., *Prevalence of cervical and oral human papillomavirus infections among US women.* J Infect Dis, 2014. **209**(11): p. 1739-43.
29. Kang, H., A. Kiess, and C.H. Chung, *Emerging biomarkers in head and neck cancer in the era of genomics.* Nat Rev Clin Oncol, 2015. **12**(1): p. 11-26.
30. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.

31. Farazi, T.A., et al., *miRNAs in human cancer.* J Pathol, 2011. **223**(2): p. 102-15.
32. Janiszewska, J., M. Szaumkessel, and K. Szyfter, *microRNAs are important players in head and neck carcinoma: a review.* Crit Rev Oncol Hematol, 2013. **88**(3): p. 716-28.
33. Nagadia, R., et al., *miRNAs in head and neck cancer revisited.* Cell Oncol (Dordr), 2013. **36**(1): p. 1-7.
34. Yu, S.L., et al., *Unique MicroRNA signature and clinical outcome of cancers.* DNA Cell Biol, 2007. **26**(5): p. 283-92.
35. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Res, 2006. **34**(Database issue): p. D140-4.
36. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic Acids Res, 2008. **36**(Database issue): p. D154-8.
37. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Res, 2011. **39**(Database issue): p. D152-7.
38. Kozomara, A. and S. Griffiths-Jones, *miRBase: annotating high confidence microRNAs using deep sequencing data.* Nucleic Acids Res, 2014. **42**(Database issue): p. D68-73.
39. Griffiths-Jones, S., *The microRNA Registry.* Nucleic Acids Res, 2004. **32**(Database issue): p. D109-11.
40. Ambros, V., et al., *A uniform system for microRNA annotation.* RNA, 2003. **9**(3): p. 277-9.
41. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units.* Genome Res, 2004. **14**(10A): p. 1902-10.
42. Desvignes, T., et al., *miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants.* Trends Genet, 2015. **31**(11): p. 613-26.
43. Skalsky, R.L. and B.R. Cullen, *Viruses, microRNAs, and host interactions.* Annu Rev Microbiol, 2010. **64**: p. 123-41.
44. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer.* Nat Rev Cancer, 2006. **6**(4): p. 259-69.
45. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
46. Cloonan, N., *Re-thinking miRNA-mRNA interactions: intertwining issues confound target discovery.* Bioessays, 2015. **37**(4): p. 379-88.
47. Calin, G.A., et al., *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2999-3004.
48. Corcoran, D.L., et al., *Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data.* PLoS One, 2009. **4**(4): p. e5279.

49. Ozsolak, F., et al., *Chromatin structure analyses identify miRNA promoters.* Genes Dev, 2008. **22**(22): p. 3172-83.
50. Calin, G.A., et al., *A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia.* N Engl J Med, 2005. **353**(17): p. 1793-801.
51. Barker, E.V., et al., *microRNA evaluation of unknown primary lesions in the head and neck.* Mol Cancer, 2009. **8**: p. 127.
52. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-8.
53. Jay, C., et al., *miRNA profiling for diagnosis and prognosis of human cancer.* DNA Cell Biol, 2007. **26**(5): p. 293-300.
54. Henson, B.J., et al., *Decreased expression of miR-125b and miR-100 in oral cancer cells contributes to malignancy.* Genes Chromosomes Cancer, 2009. **48**(7): p. 569-82.
55. Hui, A.B., et al., *Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas.* Clin Cancer Res, 2010. **16**(4): p. 1129-39.
56. Liu, X., et al., *MicroRNA profiling and head and neck cancer.* Comp Funct Genomics, 2009: p. 837514.
57. Avissar, M., et al., *MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma.* Clin Cancer Res, 2009. **15**(8): p. 2850-5.
58. Lajer, C.B., et al., *Different miRNA signatures of oral and pharyngeal squamous cell carcinomas: a prospective translational study.* Br J Cancer, 2011. **104**(5): p. 830-40.
59. Gao, G., et al., *A microRNA expression signature for the prognosis of oropharyngeal squamous cell carcinoma.* Cancer, 2013. **119**(1): p. 72-80.
60. Childs, G., et al., *Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma.* Am J Pathol, 2009. **174**(3): p. 736-45.
61. Langevin, S.M., et al., *MicroRNA-137 promoter methylation is associated with poorer overall survival in patients with squamous cell carcinoma of the head and neck.* Cancer, 2011. **117**(7): p. 1454-62.
62. Hu, A., et al., *miR-21 and miR-375 microRNAs as candidate diagnostic biomarkers in squamous cell carcinoma of the larynx: association with patient survival.* Am J Transl Res, 2014. **6**(5): p. 604-13.
63. Li, J., et al., *MiR-21 indicates poor prognosis in tongue squamous cell carcinomas as an apoptosis inhibitor.* Clin Cancer Res, 2009. **15**(12): p. 3998-4008.
64. Avissar, M., et al., *MicroRNA expression in head and neck cancer associates with alcohol consumption and survival.* Carcinogenesis, 2009. **30**(12): p. 2059-63.

65. Fletcher, A.M., A.C. Heaford, and D.K. Trask, *Detection of metastatic head and neck squamous cell carcinoma using the relative expression of tissue-specific mir-205.* Transl Oncol, 2008. **1**(4): p. 202-8.
66. Wald, A.I., et al., *Alteration of microRNA profiles in squamous cell carcinoma of the head and neck cell lines by human papillomavirus.* Head Neck, 2011. **33**(4): p. 504-12.
67. Weber, J.A., et al., *The microRNA spectrum in 12 body fluids.* Clin Chem, 2010. **56**(11): p. 1733-41.
68. Iorio, M.V. and C.M. Croce, *MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review.* EMBO Mol Med, 2012. **4**(3): p. 143-59.
69. Moldovan, L., et al., *Methodological challenges in utilizing miRNAs as circulating biomarkers.* J Cell Mol Med, 2014. **18**(3): p. 371-90.
70. Liu, C.J., et al., *Exploiting salivary miR-31 as a clinical biomarker of oral squamous cell carcinoma.* Head Neck, 2012. **34**(2): p. 219-24.
71. Momen-Heravi, F., et al., *Genomewide Study of Salivary MicroRNAs for Detection of Oral Cancer.* J Dent Res, 2014. **93**(7 Suppl): p. 86S-93S.
72. Antonsson, A., et al., *Human papillomavirus status and p16(INK4A) expression in patients with mucosal squamous cell carcinoma of the head and neck in Queensland, Australia.* Cancer Epidemiol, 2015. **39**(2): p. 174-81.
73. Li, J., et al., *Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells.* BMC Biotechnol, 2007. **7**: p. 36.
74. Lussier, Y.A., et al., *MicroRNA expression characterizes oligometastasis(es).* PLoS One, 2011. **6**(12): p. e28650.
75. Harris, T., et al., *Low-level expression of miR-375 correlates with poor outcome and metastasis while altering the invasive properties of head and neck squamous cell carcinomas.* Am J Pathol, 2012. **180**(3): p. 917-28.
76. Severino, P., et al., *MicroRNA expression profile in head and neck cancer: HOX-cluster embedded microRNA-196a and microRNA-10b dysregulation implicated in cell proliferation.* BMC Cancer, 2013. **13**: p. 533.
77. Skrzypski, M., et al., *Prognostic value of microRNA expression in operable non-small cell lung cancer patients.* Br J Cancer, 2014. **110**(4): p. 991-1000.
78. Stark, M.S., et al., *The Prognostic and Predictive Value of Melanoma-related MicroRNAs Using Tissue and Serum: A MicroRNA Expression Analysis.* EBioMedicine, 2015. **2**(7): p. 671-80.

79. Tembe, V., et al., *MicroRNA and mRNA expression profiling in metastatic melanoma reveal associations with BRAF mutation and patient prognosis.* Pigment Cell Melanoma Res, 2015. **28**(3): p. 254-66.
80. Peterson, S.M., et al., *Common features of microRNA target prediction tools.* Front Genet, 2014. **5**: p. 23.
81. Vlachos, I.S. and A.G. Hatzigeorgiou, *Online resources for miRNA analysis.* Clin Biochem, 2013. **46**(10-11): p. 879-900.
82. Lu, T.P., et al., *miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets.* PLoS One, 2012. **7**(8): p. e42390.
83. Chou, C.H., et al., *miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database.* Nucleic Acids Res, 2016. **44**(D1): p. D239-47.
84. Agarwal, V., et al., *Predicting effective microRNA target sites in mammalian mRNAs.* Elife, 2015. **4**.
85. Betel, D., et al., *Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites.* Genome Biol, 2010. **11**(8): p. R90.
86. Betel, D., et al., *The microRNA.org resource: targets and expression.* Nucleic Acids Res, 2008. **36**(Database issue): p. D149-53.
87. Paraskevopoulou, M.D., et al., *DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W169-73.
88. Reczko, M., et al., *Functional microRNA targets in protein coding sequences.* Bioinformatics, 2012. **28**(6): p. 771-6.
89. Miller, D.L., et al., *Identification of a human papillomavirus-associated oncogenic miRNA panel in human oropharyngeal squamous cell carcinoma validated by bioinformatics analysis of the Cancer Genome Atlas.* Am J Pathol, 2015. **185**(3): p. 679-92.
90. Hui, A.B., et al., *Potentially prognostic miRNAs in HPV-associated oropharyngeal carcinoma.* Clin Cancer Res, 2013. **19**(8): p. 2154-62.
91. Smeets, S.J., et al., *A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen.* Int J Cancer, 2007. **121**(11): p. 2465-72.
92. Smith, E.M., et al., *Tobacco and alcohol use increases the risk of both HPV-associated and HPV-independent head and neck cancers.* Cancer Causes Control, 2010. **21**(9): p. 1369-78.
93. Smith, E.M., et al., *Complex etiology underlies risk and survival in head and neck cancer human papillomavirus, tobacco, and alcohol: a case for multifactor disease.* J Oncol, 2012. **2012**: p. 571862.

94. Wittekindt, C., et al., *Basics of tumor development and importance of human papilloma virus (HPV) for head and neck cancer.* GMS Curr Top Otorhinolaryngol Head Neck Surg, 2012. **11**: p. Doc09.
95. Kolokythas, A., et al., *Similar Squamous Cell Carcinoma Epithelium microRNA Expression in Never Smokers and Ever Smokers.* PLoS One, 2015. **10**(11): p. e0141695.
96. Momi, N., et al., *Smoking and microRNA dysregulation: a cancerous combination.* Trends Mol Med, 2014. **20**(1): p. 36-47.
97. Saad, M.A., et al., *Alcohol-dysregulated miR-30a and miR-934 in head and neck squamous cell carcinoma.* Mol Cancer, 2015. **14**: p. 181.
98. Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-82.
99. Sass, S., et al., *MicroRNA-Target Network Inference and Local Network Enrichment Analysis Identify Two microRNA Clusters with Distinct Functions in Head and Neck Squamous Cell Carcinoma.* Int J Mol Sci, 2015. **16**(12): p. 30204-22.
100. Stransky, N., et al., *The mutational landscape of head and neck squamous cell carcinoma.* Science, 2011. **333**(6046): p. 1157-60.
101. Chung, C.H., et al., *Genomic alterations in head and neck squamous cell carcinoma determined by cancer gene-targeted sequencing.* Ann Oncol, 2015. **26**(6): p. 1216-23.
102. Lohavanichbutr, P., et al., *Genomewide gene expression profiles of HPV-positive and HPV-negative oropharyngeal cancer: potential implications for treatment choices.* Arch Otolaryngol Head Neck Surg, 2009. **135**(2): p. 180-8.
103. Martinez, I., et al., *Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas.* Eur J Cancer, 2007. **43**(2): p. 415-32.
104. Pyeon, D., et al., *Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers.* Cancer Res, 2007. **67**(10): p. 4605-19.
105. Rosenfeld, N., et al., *MicroRNAs accurately identify cancer tissue origin.* Nat Biotechnol, 2008. **26**(4): p. 462-9.
106. Chen, X., et al., *Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases.* Cell Res, 2008. **18**(10): p. 997-1006.
107. Thery, C., *Exosomes: secreted vesicles and intercellular communications.* F1000 Biol Rep, 2011. **3**: p. 15.
108. Li, L., et al., *Human bile contains microRNA-laden extracellular vesicles that can be used for cholangiocarcinoma diagnosis.* Hepatology, 2014. **60**(3): p. 896-907.

109. Pastorino, U., et al., *Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial.* Eur J Cancer Prev, 2012. **21**(3): p. 308-15.
110. Shen, J., et al., *Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers.* BMC Cancer, 2011. **11**: p. 374.
111. Sozzi, G., et al., *Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study.* J Clin Oncol, 2014. **32**(8): p. 768-73.
112. Salazar, C., et al., *A novel saliva-based microRNA biomarker panel to detect head and neck cancers.* Cell Oncol (Dordr), 2014. **37**(5): p. 331-8.
113. Ozcan, G., et al., *Preclinical and clinical development of siRNA-based therapeutics.* Adv Drug Deliv Rev, 2015. **87**: p. 108-19.
114. Zuckerman, J.E., et al., *Correlating animal and human phase Ia/Ib clinical data with CALAA-01, a targeted, polymer-based nanoparticle containing siRNA.* Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11449-54.
115. Strumberg, D., et al., *Phase I clinical development of Atu027, a siRNA formulation targeting PKN3 in patients with advanced solid tumors.* Int J Clin Pharmacol Ther, 2012. **50**(1): p. 76-8.
116. Boeri, M., et al., *MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3713-8.
117. Landi, M.T., et al., *MicroRNA expression differentiates histology and predicts survival of lung cancer.* Clin Cancer Res, 2010. **16**(2): p. 430-41.
118. Qi, J. and D. Mu, *MicroRNAs and lung cancers: from pathogenesis to clinical implications.* Front Med, 2012. **6**(2): p. 134-55.
119. Dillhoff, M., et al., *MicroRNA-21 is overexpressed in pancreatic cancer and a potential predictor of survival.* J Gastrointest Surg, 2008. **12**(12): p. 2171-6.
120. Schetter, A.J., et al., *MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma.* JAMA, 2008. **299**(4): p. 425-36.
121. Rossi, S., et al., *microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival.* Blood, 2010. **116**(6): p. 945-52.
122. Gao, W., et al., *MiR-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis.* J Cancer Res Clin Oncol, 2011. **137**(4): p. 557-66.