# Minimally Invasive Detection of Head and Neck Squamous Cell Carcinoma

by

Nuwan Shyanaka Dharmawardana
BSc, PGDipSci (Merit), MSc (Hons), BMBS, MClinSci

Thesis submitted to Flinders University
for the degree of

**Doctor of Philosophy**

Department of Otorhinolaryngology, Head and Neck Surgery
College of Medicine and Public Health
Flinders University

May 2021

"When you have exhausted all possibilities, remember this: you haven’t."

~ Thomas Edison

## Acknowledgements

I would like to acknowledge the Australian Government Research Training Program Scholarship, the Garnett Passe and Rodney Williams Memorial Foundation's Academic Surgeon-Scientist Research Scholarship, and the Australia and New Zealand Head and Neck Cancer Society Foundation for their financial support during this candidature.

I am honoured and privileged to have been supervised by Associated Professor Eng Ooi, Dr Roger Yazbeck and Dr Damian Hussey. As my primary supervisor, I would like to thank Associate Professor Ooi for his guidance, support, and motivation throughout my candidature towards becoming a surgeon-scientist in the field of Otorhinolaryngology-Head and Neck Surgery. I extend my thanks to Dr Yazbeck for his close supervision, positive attitude, and the constant encouragement to think big and push further. I would also like to thank Dr Hussey for his incredible knowledge and advice in meticulous experimental design. Thank-you for allowing me to be part of this diverse research group and to foster skills for the future. Special thank you to Professor Geraint Rogers and Dr Jocelyn Choo for supervision and support during the microbiome component of the thesis.

Completion of this work would not have been possible without the support of Dr Charmaine Woods. I thank you for believing in me and my work; your encouragement and praise throughout this candidature has made a significant positive impact on my mental health as well as my academic progress.

I thank Dr George Mayne for taking the time to discuss intricacies of mathematical modelling and broader research concepts. Thank you to Ms Michelle Parsons for

technical support during initial phases of breath analysis. My sincere thanks to all work colleagues and mentors at Flinders Medical Centre and Royal Adelaide Hospital for the opportunity to meet and recruit patients with head and neck cancers. Thank you for being flexible with sample collection procedures and the processes involved with clinical research.

I am grateful to have met these amazing and capable higher degree students: Dr Thomas Goddard, Dr Kyra Sierakowski, Ms Tamara Crittenden, Ms Mistyka Schar, Dr Anushia Ashokan, Dr Simone Jaenisch, Dr Ravi Vissapragada and Dr Lucy Huang. Thank-you for your friendship, candour, intellect, work ethic and joy. Special thank you to Dr Goddard for thought provoking discussions throughout our candidature that significantly contributed to my academic progress.

Thank you to my family for their unwavering support and encouragement. To Holly, my better half in life, thank you for believing in me and making it possible to pursue a career as a surgeon-scientist. To my children Kira and Luka, thank you for reminding me - on a daily basis - what is really important in life; I look forward to spending more time with you. To Ammi (Mum) and Thaththi (Dad), thank you for nurturing this curious brain of mine, words cannot express the love and gratitude I have for you.

Last but not least, I would like to thank all the patients who participated in this research project. At your most vulnerable time with a pending diagnosis of cancer, you selflessly said ‘yes’ to me, knowing that the outcome of this project may never help you directly. I cannot thank you enough for participating in this research, without your contribution, completion of this work would have been impossible.

## Abstract

This PhD thesis investigated extracellular vesicle microRNA and exhaled breath volatile organic compounds as minimally invasive biomarkers for the detection of head and neck squamous cell carcinoma (HNSCC).

HNSCC is a debilitating disease with poor patient outcomes associated with advanced stages of disease. There is a paucity of clinically validated biomarkers capable of early detection of HNSCC. Therefore, an initial systematic review of recent literature was conducted to identify reasons for poor translation of biomarkers into clinical practice. These reviews indicated significant variability in methodology in both circulating microRNA and breath analysis research that contributed to the lack of clinically validated biomarkers for detection of HNSCC.

A key focus of this thesis was the development of a predictive model using a minimally invasive biomarker to detect patients with HNSCC. Standardised protocols were developed to collect blood, breath, and microbiome samples from patients in a clinical setting. Subsequently, extracellular vesicle microRNAs and exhaled breath volatile organic compounds were investigated for their potential as a biomarker for detecting HNSCC. Both modalities were accurate at detecting patients with HNSCC with high sensitivity and specificity.

The breath test described in Chapter five was the first and the largest study to describe the use of raw mass spectra from a selected ion flow tube mass spectrometer to accurately detect HNSCC. This thesis was also the first to describe a HNSCC dependent intestinal dysbiosis that was detected using exhaled breath hydrogen and methane analysis.

A range of complex algorithms available for predictive analysis were explored in this thesis including the assessment of their applicability to certain datasets. Binary logistic regression was the most intuitive algorithm for predictive analysis with minimal assumptions regarding the dataset. However, machine learning algorithms should be utilised for large datasets with complex interactions.

Core relationships between circulating microRNAs, exhaled breath compounds, patient factors, environmental factors and disease risk factors were also explored in this thesis. The importance of measuring ambient volatile gas levels during breath sample collection to recognise potential sample contamination was further highlighted. It also identified a measurable marker (acetonitrile) of smoking status which is important as a key risk factor for the development HNSCC as well as treatment outcomes for HNSCC.

Collectively, this thesis provides evidence for the use of circulating microRNAs and exhaled breath volatile compounds as biomarkers for detecting HNSCC. They are both accurate and potentially able to implement into a clinical setting. However, they require in-depth cost-effectiveness analysis and large-scale clinical trials in a primary care setting prior to effective clinical translation.

## Thesis Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and
2. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Nuwan Dharmawardana

7th December 2020

# Publication, Presentations and Awards

## Publications related to this thesis.

1. **Dharmawardana, N.**, Ooi, E.H., Woods, C., Hussey, D. Circulating microRNAs in head and neck cancer: A scoping review of methods. Clin Exp Metastasis 36, 291-302 doi.org/10.1007/s10585-019-09961-6 (2019)

2. **Dharmawardana, N.**, Woods, C., Watson, D.I., Yazbeck, R., Ooi, E.H. A review of breath analysis techniques in head and neck cancer. Oral Oncol 104, 104654 doi.org/10.1016/j.oraloncology.2020.104654 (2020)

3. Mayne, G., Woods, C., **Dharmawardana, N.**, Wang, T., Krishnan, S., Hodge, J. *et al*. Cross validated serum small extracellular vesicle micrornas for the detection of oropharyngeal squamous cell carcinoma. Journal of translational medicine 18, 1-12 (2020)

4. **Dharmawardana, N.**, Goddard, T., Woods, C., Watson, D.I., Butler, R., Ooi, E.H. *et al*. Breath methane to hydrogen ratio as a surrogate marker of intestinal dysbiosis in head and neck cancer. Scientific Reports 10, 1-8 (2020)

5. **Dharmawardana, N.**, Goddard, T., Woods, C., Watson, D.I., Ooi, E.H., Yazbeck, R. Development of a non-invasive exhaled breath test for the diagnosis of head and neck cancer. British Journal of Cancer 1-7 (2020)

## Other publications during candidature

1. Kao, S.S., Peters, M.D.J., **Dharmawardana, N.**, Stew, B., Ooi, E.H. Scoping review of pediatric tonsillectomy quality of life assessment instruments. Laryngoscope 127, 2399-2406 doi.org/10.1002/lary.26522 (2017)

2. **Dharmawardana, N.**, Campbell, J.M., Carney, A.S., Boase, S. Effectiveness of primary surgery versus primary radiotherapy on unknown primary head and neck squamous cell carcinoma: A systematic review protocol. JBI Evidence Synthesis 16, 308-315 (2018)

3. **Dharmawardana, N.**, Chandran, D., Elias, A., Kao, S.S.-T. Management of post tonsillectomy secondary haemorrhage: Flinders experience. Australian Journal of Otolaryngology 1, (2018)

4. Kao, S.S.-T., Marshall-Webb, M., **Dharmawardana, N.**, Foreman, A., Ooi, E.H. Gastrostomy tube insertion outcomes in South Australian head and neck cancer patients. Australian Journal of Otolaryngology 1, (2018)

5. Noor, A., Stepan, L., Kao, S.S., **Dharmawardana, N.**, Ooi, E., Hodge, J. *et al*. Reviewing indications for panendoscopy in the investigation of head and neck squamous cell carcinoma. The Journal of Laryngology & Otology 132, 901-905 (2018)

6. Stew, B., **Dharmawardana, N.**, Kao, S.S., Elias, A.J., Ooi, E.H. Endoscopic smartphone adaptors in otolaryngology: How do they compare? Clin Otolaryngol 43, 1599-1602 doi.org/10.1111/coa.13186 (2018)

7. Stew, B., Kao, S.S., **Dharmawardana, N.**, Ooi, E.H. A systematic review of validated sinus surgery simulators. Clin Otolaryngol 43, 812-822 doi.org/10.1111/coa.13052 (2018)

8. Huang, L., Athanasiadis, T., Woods, C., **Dharmawardana, N.**, Ooi, E.H. The use of transnasal humidified rapid insufflation ventilatory exchange in laryngeal and pharyngeal surgery: Flinders case series. Australian Journal of Otolaryngology 2, (2019)

9. Krishnan, G., Stepan, L., Du, C., Padhye, V., Bassiouni, A., **Dharmawardana, N.** *et al*. Tonsillectomy using the Bizact: A pilot study in 186 children and adults. Clin Otolaryngol 44, 392-396 (2019)

10. Huang, L., **Dharmawardana, N.**, Badenoch, A., Ooi, E.H. A review of the use of transnasal humidified rapid insufflation ventilatory exchange for patients undergoing surgery in the shared airway setting. Journal of Anesthesia 34, 134-143 (2020)

**Presentations during candidature**

1. **Scoping review of selected ion flow mass spectrometry methodologies for human breath analysis**. Australian Society for Medical Research, Regional Annual Scientific Meeting, Adelaide, Australia, June 2017

2. **Breath volatile organic compounds in head and neck squamous cell cancer.** Australian Society for Medical Research, Regional Annual Scientific Meeting, Adelaide, Australia, June 2018

3. **Characterisation of breath volatile organic compounds in patients with upper aerodigestive mucosal squamous cell cancer.** International Association for Breath Research, Maastricht, Netherlands, July 2018

4. **Breath testing for cancer detection.** Australia and New Zealand Head and Neck Cancer Society, Annual Scientific Meeting, Melbourne, Australia, July 2018

5. **Breath testing for cancer detection.** The Garnett Passé and Rodney Williams Memorial Foundation, Frontiers Conference, Cairns, Australia, August 2018

6. **Characterisation of breath volatile organic compounds in patients with upper aerodigestive mucosal squamous cell cancer.** 6th World Congress of the International Federation of Head and neck Oncologic Societies, Buenos Aires, Argentina, September 2018

7. **Towards Standardisation: Circulation microRNAs in Head and Neck Cancer.** Australian Society of Otolaryngology Head and Neck Surgery, Annual Scientific Meeting, Brisbane, Australia, March 2019

8. **Breath, Bugs and Cancer.** Australia and New Zealand Head and Neck Cancer Society, Annual Scientific Meeting, Adelaide, Australia, September 2019

9. **Minimally Invasive Detection of Head and Neck Cancer.** Ron Gristwood Medal Presentation, South Australian Otorhinolaryngology, Head and Neck Society Scientific Evening. Adelaide, Australia

10. **Minimally Invasive Detection of Head and Neck Cancer.** R.P. Jepson Medal Presentation, Royal Australasian College of Surgeons, Annual Scientific Presentations Evening. Adelaide, Australia

**Scholarships/Awards/Grants**

1. **Academic Surgeon-Scientist Research Scholarship**, The Garnett Passé and Rodney Williams Memorial Foundation, Jan 2018

2. **People's Choice Award for Best Oral Presentation**, South Australian Annual Scientific Meeting, Australian Society for Medical Research, June 2018

3. **Best Oral Presentation**, Annual Scientific Meeting, Australian and New Zealand Head and Neck Cancer Society and International Society for Maxillofacial Rehabilitation, July 2018

4. **Best Science Poster Presentation**, 6th World Congress of the International Federation of Head and neck Oncologic Societies, September 2018

5. **Best Poster Presentation**, The Garnett Passé and Rodney Williams Memorial Foundation, Frontiers Conference, August 2018

6. **Student Travel Grant**, The Hospital Research Foundation, South Australia, 2018

7. **Student Publication Prize**, College of Medicine and Public Health, Flinders University, 2019

8. **Student Poster Prize**, Flinders Cancer Research Day, College of Medicine and Public Health, Flinders University, September 2019

9. **Student Conference Travel Grant,** College of Medicine and Public Health, Flinders University, 2019

10. **R. P. Jepson Medal**, Royal Australasian College of Surgeons, South Australia, 2020

## Abbreviations

| | |
|---|---|
| AIHW | Australian Institute of Health and Welfare |
| AJCC | American Joint Committee on Cancer |
| ALDH | Alcohol Dehydrogenase |
| CNR | Cannabinoid Receptor |
| COX | Cycloxygenase |
| CS | Current Smoker |
| EGFR | Epidermal Growth Factor Receptor |
| ES | Ex-Smoker |
| FOM | Floor of Mouth |
| HNSCC | Head and Neck Squamous Cell Carcinoma |
| HPV | Human Pappilomavirus |
| ICOR | International Consortium for Outcomes Research |
| IMRT | Intensity Modulated Radiation Therapy |
| MALT | Mucosally Associated Lymphoid Tissues |
| MAPK | MAP - Kinase |
| NS | Non-Smoker/Never Smoked |
| PCA | Principal Component Analysis |
| RF | Random Forest |
| SBRT | Stereotactic Body Radiation Therapy |
| SCC | Squamous Cell Carcinoma |
| SPLSDA | Sparse Partial Least Squares Discriminant Analysis |
| SVM | Support Vector Machine |
| TNM | Tumour, Node and Metastasis |
| UICC | International Union Agaist Cancer |
| VEGF | Vascular Endothelial Growth Factor |
| VMAT | Volumetric Modulated Arc Therapy |
| VOC | Volatile Organic Compound |
| FMC | Flinders Medical Centre |
| RAH | Royal Adelaide Hospital |
| RCF | Relative Centrifugation Force |
| RNA | Ribonucleic Acid |

# List of Figures and Tables

## Figures

## Appendix Figures

## Tables

## Appendix Tables



# Equations

# Table of Contents

# 1 Introduction

## 1.1 Structure of the thesis

**Chapter 1** provides an overview of HNSCC and establishes the need for biomarkers of early detection. Existing biomarkers in the context of HNSCC and their utility are also described here.

**Chapter 2** presents a systematic scoping review of existing literature investigating variability in methods in the context of serum-based microRNA biomarker discovery for HNSCC.

**Chapter 3** investigates the use of serum-based small extracellular vesicle microRNA to detect patients with HNSCC. Multiple predictive modelling algorithms are explored here including logistic regression, unsupervised principal component analysis, sparse partial least squares discriminant analysis, support vector machines and random forest. Individual microRNAs that significantly contributed to these models are further discussed in detail.

**Chapter 4** presents a narrative review of current literature specific to breath analysis in head and neck cancer. This chapter describes a) technical limitations, b) methodological variability, and c) multitude of technologies available to conduct exhaled breath analysis.

**Chapter 5** describes the development of a predictive model using binary logistic regression to accurately identify HNSCC using exhaled breath analysis.

**Chapter 6** investigates the relationship between exhaled breath compounds and the host microbiome in the context of HNSCC.

**Chapter 7** provides a summary of findings, clinical and research implications, future direction, and limitations.

**Appendices I, II and III** provide supplementary material related to chapters described above.

**Appendix IV** investigates the influence of environmental gases on exhaled breath markers.

**Appendix V** provides a summary of patient dependent factors and their relationship between microRNA and VOC biomarkers.

**Appendix VI, VII and VIII** provides supplementary information that includes participant information sheet, participant consent forms and participant questionnaires.

## 1.2 Aims of the thesis.

The main aim of this thesis was to identify minimally invasive biomarkers for the detection of HNSCC. Specifically, this thesis aimed to:

- Determine the sources of variability related to methods in circulating microRNA biomarkers specific to HNSCC.
- Investigate and understand the implementation of advanced classification models appropriate for discovery of microRNA biomarkers.
- Determine the sources of variability related to methods in exhaled breath biomarkers specific to HNSCC.
- Develop a predictive model based on breath biomarkers capable of accurately identifying patients with HNSCC.
- Investigate the relationship between exhaled breath markers and the host microbiome.
- Define the demographic dependent variability in minimally invasive biomarkers.
- Determine the influence of environmental gases on exhaled breath markers.

## 1.3 Head and Neck Squamous Cell Carcinoma

Head and neck cancer is a debilitating disease that primarily consists of squamous cell carcinoma (SCC) (Cohen *et al*. 2018b). Head and neck squamous cell carcinomas (HNSCC) affect mucosal, cutaneous, and adnexal subsites in the head and neck region. Therefore, patients with HNSCC develop a multitude of sensory deficits related to sight, smell, taste, and hearing. They also suffer from upper aerodigestive dysfunction with dysphagia and respiratory distress that at times could be immediately life threatening (Smith 2000).

Advanced HNSCC lesions in particular can severely disfigure patients leading to subsequent psychosocial disadvantages (Cohen *et al*. 2018a) and cause significant issues with communication due to complete removal of the larynx (voice box) or resection of vital soft tissues required for speech (Patterson 2019). Advances in public health campaigns, diagnostic modalities and treatment paradigms have improved the overall prognosis and associated morbidity of this disease (Kim *et al*. 2010). However, oncological outcomes remain poor in advanced stage disease (Cadoni *et al*. 2017).

Furthermore, geographical, socioeconomic, and racial disadvantages related to access to appropriate health care have been reported (Mukherjee *et al*. 2020). Therefore, significant efforts are needed to formulate methods of early detection of HNSCC. Community-based screening (clinical examination) and educational programs have shown an increased awareness of HNSCC and related risk factors (Harris *et al*. 2013). Therefore, these screening programs have the potential to increase HNSCC related health literacy and the potential to influence a reduction in exposure to risk factors while allowing identification of HNSCCs at an early

cancer stage. However, further cost-effective, and accurate screening tools are needed for effective detection of early stage HNSCC.

In current clinical practice, primary health care systems rely heavily on patient reported symptoms to prompt referral for specialist head and neck cancer assessment. The positive predictive value of diagnosing a HNSCC based on upper aerodigestive symptoms alone is 77% (Tikka *et al*. 2016). While this is comparable to other diseases, the presence of clinical signs and symptoms with high specificity (such as cranial neuropathy or a neck mass) is consistent with advanced disease (Tikka *et al*. 2016). In addition, 23% of the patients referred to tertiary care assessment have no malignancy (Tikka *et al*. 2016), adding a significant cost to the health care system with patients receiving unnecessary investigations for suspected HNSCC. Therefore, providing effective screening tools to primary care setting could improve early diagnosis of HNSCC and potentially reduce tertiary health care costs.

### 1.3.1 Mucosal and lymphatic anatomy of the head and neck region

This thesis will focus on mucosal HNSCC with the exclusion of salivary gland, sinonasal, nasopharyngeal and cutaneous carcinomas. Oncological anatomy of the head and neck region is complex, and this section describes the oncology specific subsite anatomy highlighting the complexities associated with examining these regions of the human body.

#### 1.3.1.1 Oral cavity oncological anatomy and physiology

The distinct (oncologically relevant) anatomical subsites of the oral cavity prone to SCC are defined as the vermilion border of the lips, inferior and superior alveolar

ridges, buccal mucosa, retromolar trigones, oral tongue (anterior 2/3) (Figure 2) and the FOM (Montero and Patel 2015).

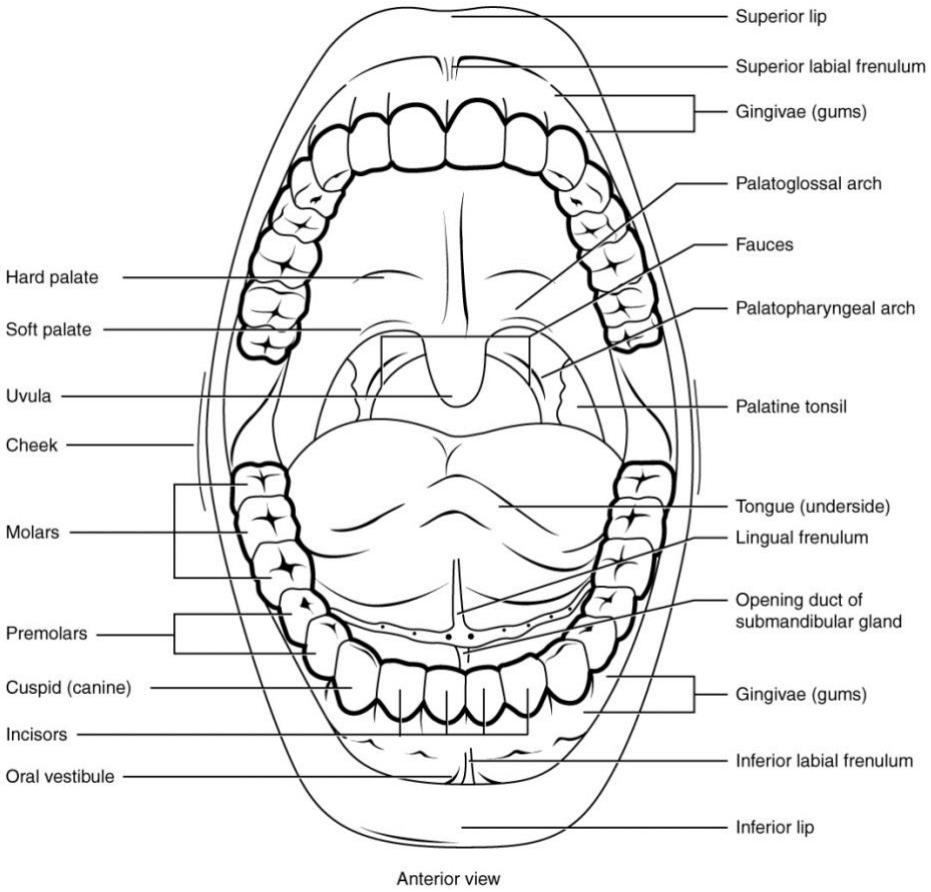



***Figure 1. Oral cavity. Image from [https://opentextbc.ca](https://opentextbc.ca) used according to permission from Creative Commons Attribution 4.0 International License.***

It is important to consider the deep soft tissue structures surrounding the oral cavity as the invasion of these structures determines the overall prognosis of HNSCC and cause significant, potentially longstanding cosmetic and functional deficits.

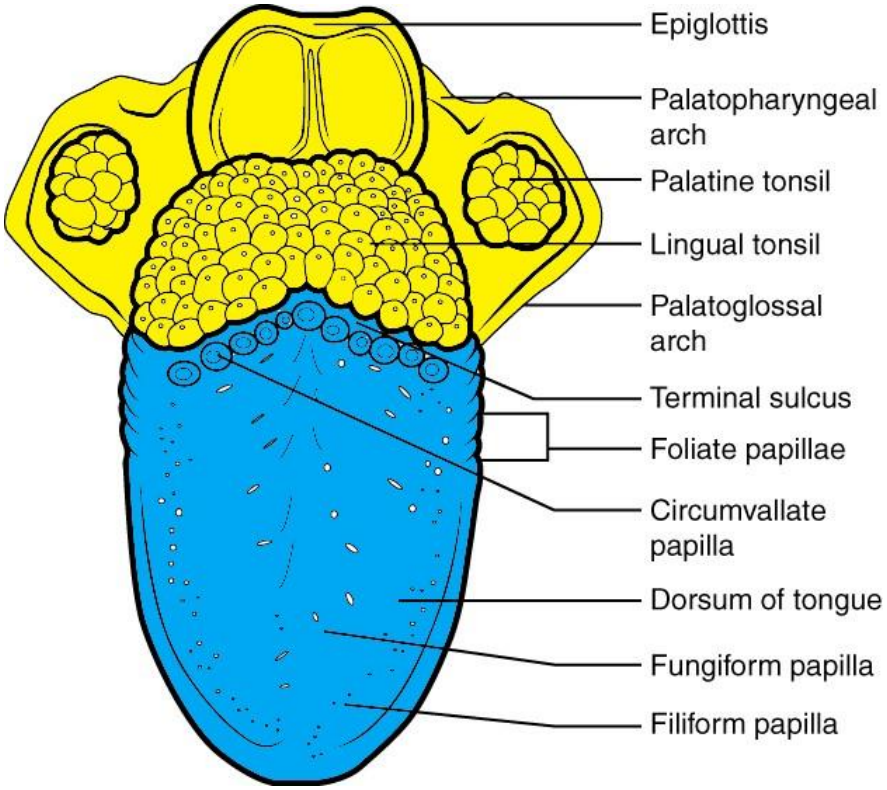



***Figure 2. Oblique view of the oral tongue and the adjacent oropharynx. Highlighted in yellow and blue represents oropharynx and oral cavity, respectively. Image from [https://opentextbc.ca](https://opentextbc.ca) modified according to permission from Creative Commons Attribution 4.0 International License.***

### 1.3.1.2 Oropharyngeal oncological anatomy and physiology

The oropharynx is the conduit between the nasopharynx, oral cavity, and the laryngopharynx. Therefore, any dysfunction of this region because of a malignancy can affect swallowing as well as breathing. Cancer related subsites were defined as the palatine tonsils, base of tongue (posterior 1/3 of the tongue), posterior pharyngeal wall and the soft palate including the uvula (Lydiatt *et al*. 2017).

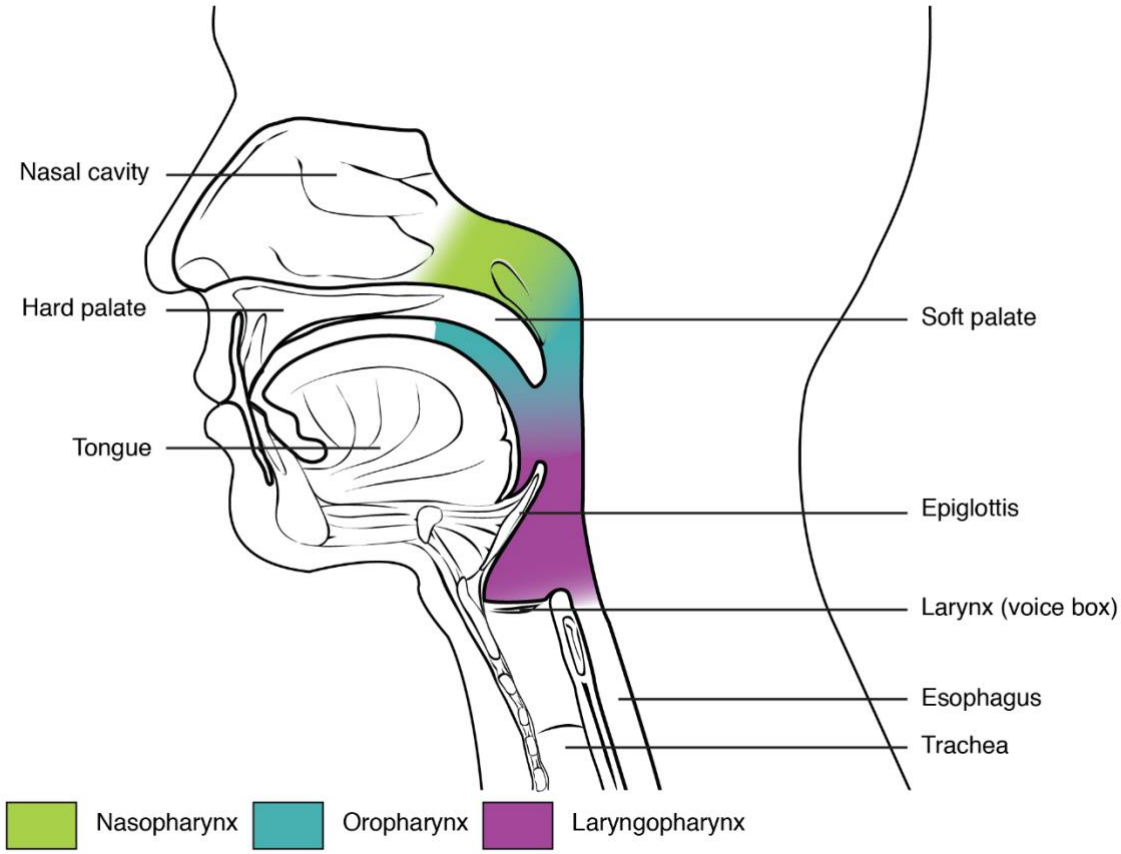



***Figure 3. Parasagittal schematic of the nasal, oral and pharyngeal anatomy. Image from [https://opentextbc.ca](https://opentextbc.ca) modified according to permission from Creative Commons Attribution 4.0 International License.***

Palatine tonsils and the lymphoid tissue overlying the base of tongue (lingual tonsils) belong to the Waldeyer's ring of mucosal associated lymphoid tissues (MALT) (Hellings *et al*. 2000). While the palatine tonsils can be visualised during routine clinical examination via the oral cavity, the base of tongue pose a challenge in the primary care setting without having access to specialist equipment such as a flexible nasendoscope (Figure 3). Epithelial malignancies as well as lymphomas are common to these specialised lymphoid organs. Unique to pharyngeal carcinomas, patients with advanced oropharyngeal HNSCC often requires an alternate method of nutrition such as nasogastric tube feeding or gastrostomy feeding, during and after treatment due to physical and functional obstruction of the pharyngeal swallowing (Kao *et al*. 2018). The soft palate

including the uvula can be a challenging subsite with two mucosal surfaces (one facing the oral cavity, another facing the nasopharynx), that is often difficult to visualise adequately in primary care setting without specialist equipment.

#### 1.3.1.3 Laryngeal oncological anatomy and physiology

The human larynx is divided into three subsites:1) the supraglottis which includes the epiglottis, aryepiglottic folds, false folds, and the vestibule lying above the true vocal folds. 2) the glottis which includes true vocal folds and arytenoid cartilages, and 3) the subglottis, 1 cm below the glottis which is predominantly formed by the cricoid cartilage. The trachea lies inferior to the cricoid cartilage and does not form part of the larynx (Figure 4). Patients with laryngeal cancers are likely to present with voice change, dysphagia, and respiratory distress (when advanced). However, it is not possible for primary care physicians or patients to examine or screen for any suspicious lesions in the larynx (Figure 5). Flexible nasendoscopy by an otorhinolaryngologist is the primary method of visualising this area in clinical practice.

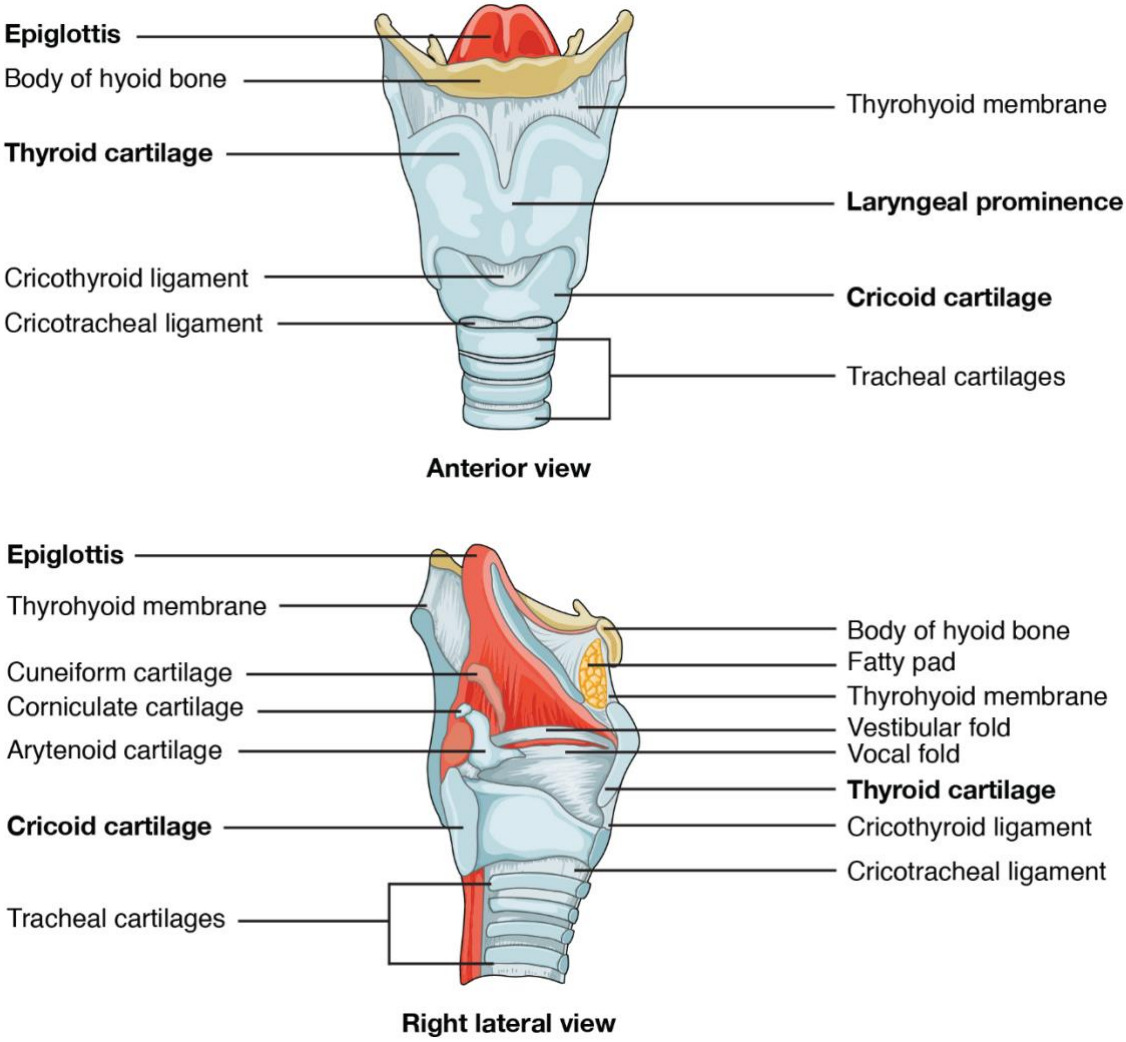



***Figure 4. Anterior and parasagittal views of the larynx. Image from [https://opentextbc.ca](https://opentextbc.ca) modified according to permission from Creative Commons Attribution 4.0 International License.***

### 1.3.1.4 Hypopharynx

The hypopharynx subsite includes the lateral pharyngeal walls, posterior pharyngeal wall surrounding the larynx and the piriform sinuses on either side of the larynx converging into the cervical oesophagus. While the oesophagus itself is not defined as a head and neck subsite, the post cricoid space of the cervical oesophagus is considered part of the hypopharyngeal subsite.

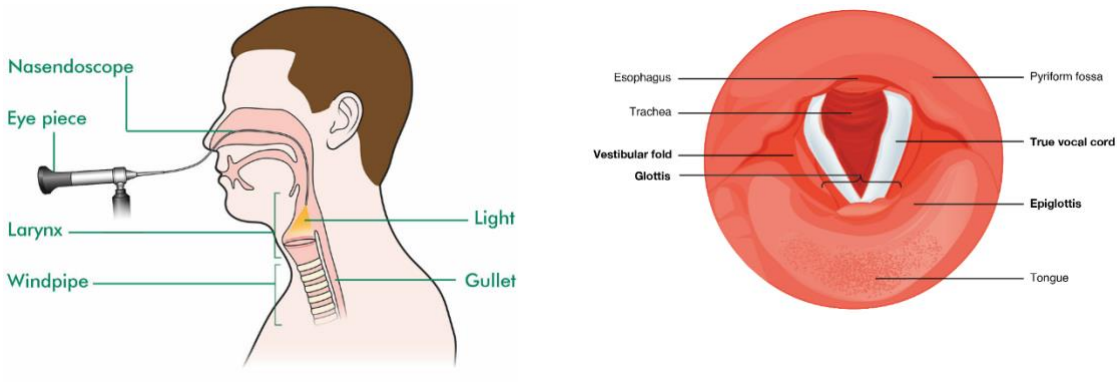



***Figure 5. Schematic of a fibre optic view of the glottis and hypopharynx. Image from [https://opentextbc.ca](https://opentextbc.ca) modified according to permission from Creative Commons Attribution 4.0 International License.***

### 1.3.1.5 The neck and related lymphoid anatomy

The neck extends from the base of skull superiorly to the root of the neck bounded by the first cervical ribs inferiorly. The superior border of the neck ventrally is the mandible and dorsally it is the basiocciput. From an oncological perspective, the neck contains numerous lymphoid tissues that include lymph nodes and lymph channels such as the thoracic and lymphatic ducts.

In head and neck oncology six lymph node levels are defined based on identifiable anatomical landmarks that also relate to biologically significant pathways of locoregional tumour metastasis (Robbins *et al*. 2002). This classification is also important when describing the extent of lymph node dissection during oncological surgery. The most recent iteration by the American Academy of Otolaryngology-Head and Neck Surgery (AAOHNS) will be referred to in this thesis as summarised in Figure 6 and

Table 1 (Robbins *et al*. 2002). The improvements in this classification system have been driven by the clinical applicability and the ability for radiological discrimination of node levels. In the most recent iteration, lymph nodes have been divided into six main levels in the neck (Table 1) (Gregoire *et al*. 2014).

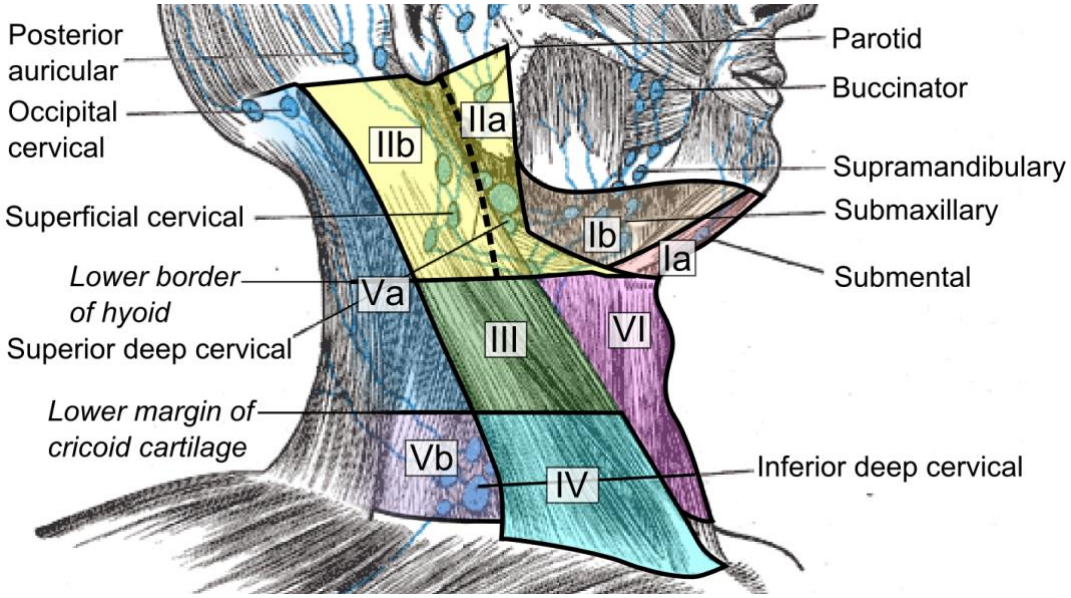



***Figure 6. Lymph node levels of the right neck, oblique view. Image from https://commons.wikimedia.org/wiki/ modified according to permission from Creative Commons Attribution 1.0 International License.***

***Table 1. Description of surgically distinct lymph node as divided into "levels" with specific anatomical boundaries.***

| Lymph node level | Boundaries |
|---|---|
| Level 1a | Mandibular symphysis<br>Anterior belly of contralateral and ipsilateral digastric muscle<br>Body of Hyoid bone |
| Level 1b | Body of mandible<br>Anterior belly of digastric muscle<br>Stylohyoid muscle<br>Body of Hyoid bone |
| Level 2a | Skull base<br>Inferior border of the hyoid bone<br>Stylohyoid muscle<br>Spinal accessory nerve |
| Level 2b | Skull base<br>Horizontal plane defined by the inferior border of the hyoid bone<br>Spinal accessory nerve<br>Lateral border of sternocleidomastoid muscle |
| Level 3 | Horizontal plane defined by the inferior border of the hyoid bone<br>Horizontal plane defined by the inferior border of the cricoid cartilage<br>Lateral border of the sternohyoid muscle<br>Lateral border of the sternocleidomastoid muscle |
| Level 4 | Inferior border of the cricoid cartilage<br>Clavicle<br>Lateral border of the sternohyoid muscle<br>Lateral border of the sternocleidomastoid muscle |
| Level 5a | Apex of the sternocleidomastoid and trapezius muscle<br>Horizontal plane defined by the inferior border of cricoid cartilage<br>Lateral border of the sternocleidomastoid muscle<br>Medial border of the trapezius muscle |
| Level 5b | Horizontal plane defined by the inferior border of the cricoid cartilage<br>Clavicle<br>Lateral border of the sternocleidomastoid muscle<br>Medial border of the trapezius muscle |
| Level 6 | Hyoid bone<br>Suprasternal notch<br>Common carotid artery either side |

### 1.3.2 Epidemiology of Head and Neck Squamous Cell Carcinoma

SCC is the predominant (90%) histopathological diagnosis of head and neck cancers (Cohen *et al*. 2018b). It is the seventh most common cancer worldwide with an incidence of up to 890,000 cases and 450,000 deaths per year (Ferlay *et al*. 2018). Globally, oral cavity SCC is the most common subsite accounting for 354,864 new cancers and 177,384 related deaths per annum (Ferlay *et al*. 2018). Historically, tobacco and alcohol use were the predominant causative agents for HNSCC (Argiris *et al*. 2008). More recently, Human Papilloma Virus (HPV) has emerged as a common causative agent for oropharyngeal SCC (Gillison 2004). HPV positive cancers tend to affect younger patients of middle to high socio-economic status compared to HPV negative patient cohorts (Deschler *et al*. 2014). A geographical variance in HPV positive HNSCC is also noted where the developing world reports relatively low incidence of HPV positive oropharyngeal SCCs compared to the rest of the world despite having an overall increase in HNSCC (Elrefaey *et al*. 2014).

#### 1.3.2.1 Australian head and neck cancer statistics

The Australian Institute of Health and Welfare (AIHW) in 2014 described an increase in incidence from 2475 cases in 1982 to 3896 cases in 2009. Although, the age standardised incidence rate decreased from 19.3 to 16.8 per 100000 persons in the same period (Figure 7). This discrepancy likely represents the aging Australian population (AIHW 2014). Similar to the incidence rate, the age standardised mortality rate reduced from 6.1 to 3.8 per 100000 persons while the absolute number of deaths from this disease increased from 752 to 944 (AIHW 2014) (Figure 8).

Head and neck cancer predominantly affect the Australian male population with 78% of new cases in 2009 reported in males compared to 26% of new cases in females (AIHW 2014). An updated report in 2017 estimated a further increase to the incidence of 4955 (3625 male and 1330 female). The estimated mortality is reported as 1026 (777 male and 249 female) in 2017 (AIHW 2017). Older males continue to have a higher incidence rate and a poorer mortality rate compared to females (AIHW 2013; AIHW 2017).

One in thirty-two adult men under 85 years were estimated to develop a head and neck malignancy in 2017. The five-year relative survival for head and neck cancer in males is 68% (AIHW 2017). While the five-year survival rates for both sexes have improved over the past three decades, females maintain a five-year survival advantage of 4% compared to males for years 2009-2013 (AIHW 2013).

Australian prevalence data indicates that 41550 people were living with a head and neck cancer diagnosis at the end of 2012. These patients were diagnosed between 1982 and 2012 and only 3953 of these patients were diagnosed in 2012. This large group of patients surviving after the diagnosis of head and neck cancer indicates the need for effective surveillance methods for early detection of cancer recurrence.

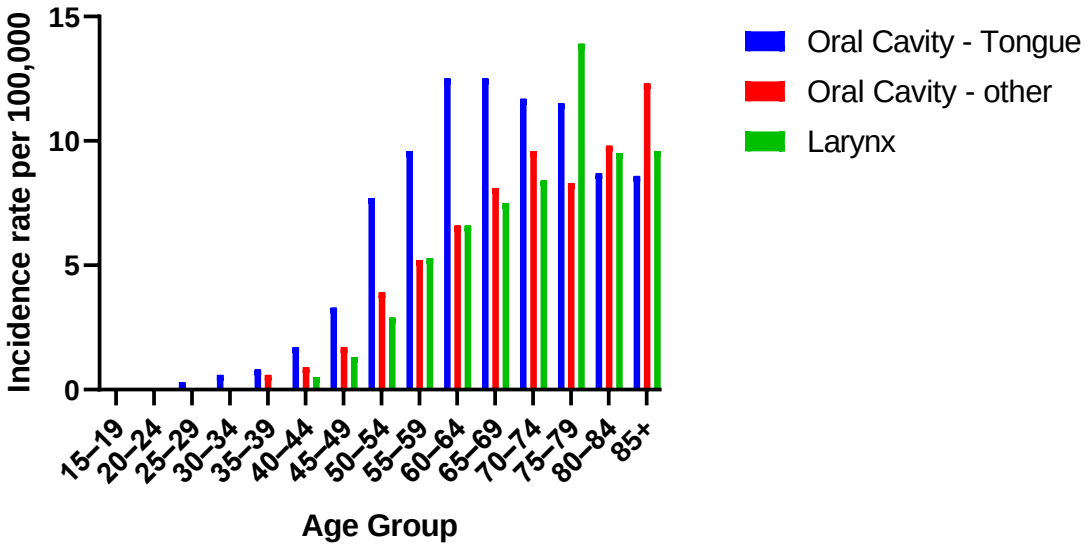



***Figure 7. Incidence rate of specific head and neck cancer subsites per 100,000 persons in Australia. Data obtained from 2016 Australian Institute of Health and Welfare Cancer Database.***

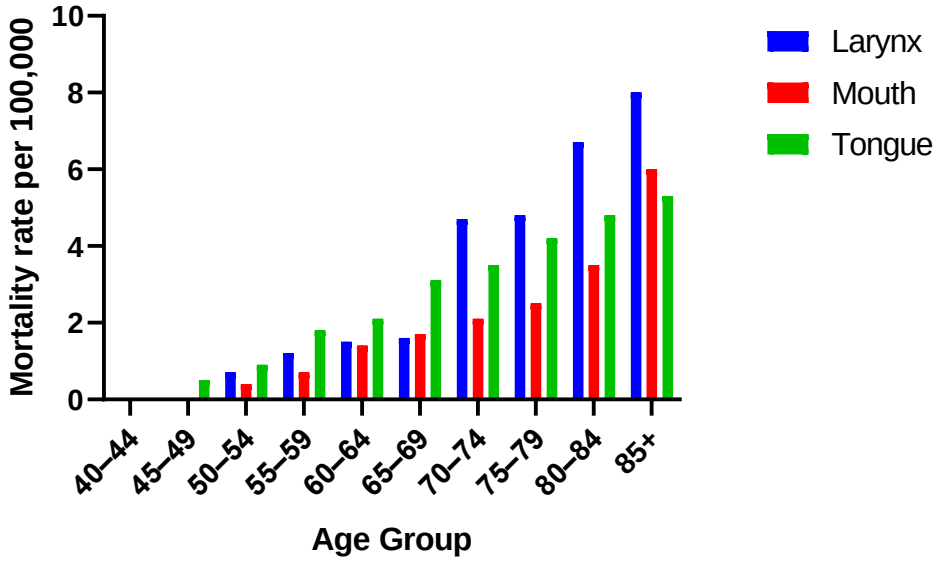



***Figure 8. Mortality rate of specific head and neck cancer subsites per 100,000 persons in Australia. Data obtained from 2016 Australian Institute of Health and Welfare Cancer Database.***

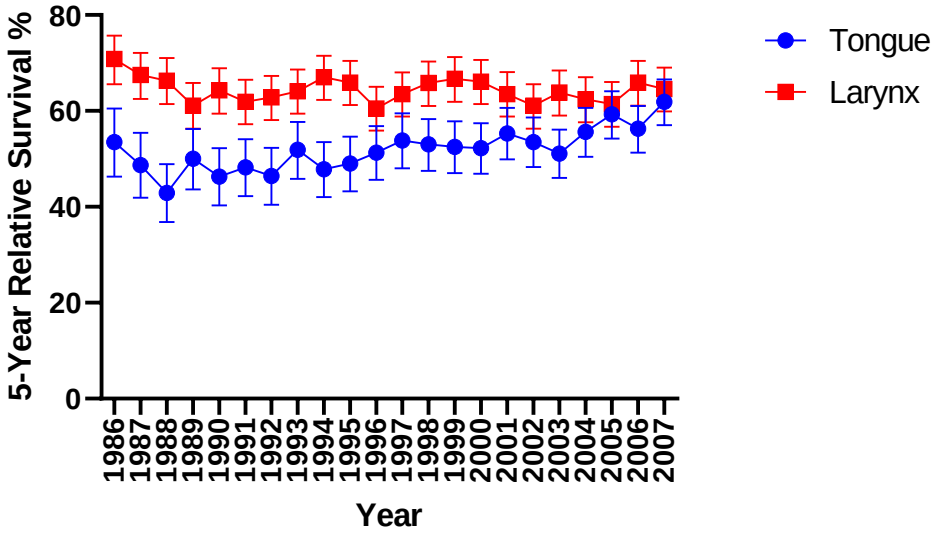



***Figure 9. Five-year relative survival for tongue and laryngeal carcinoma in Australia. Data obtained from 2014 Australian Institute of Health and Welfare Cancer Database.***

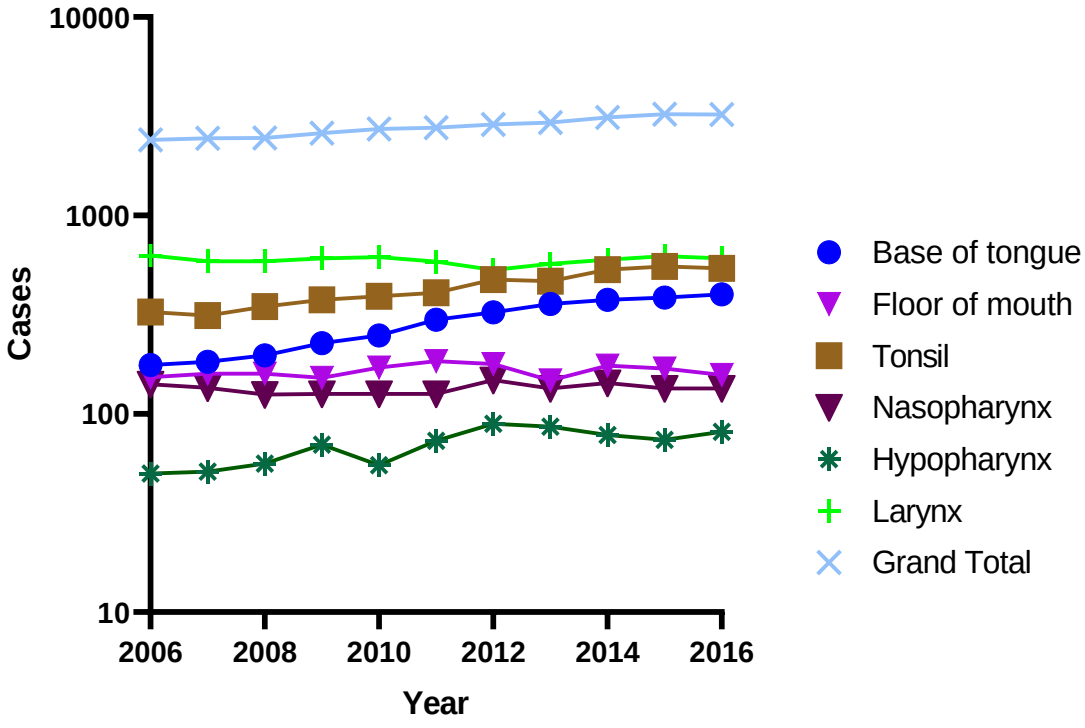



***Figure 10. New cases of head and neck cancer based on subsite for ten years in Australia. Data obtained from 2016 Australian Institute of Health and Welfare Cancer Database.***

### 1.3.3 Risk factors for head and neck squamous cell carcinoma

Primary and secondary prevention of mucosal HNSCC has focused principally on reducing tobacco smoking and alcohol consumption (Shaw and Beasley 2016). HPV (detailed in section 1.5.2) is an increasingly recognised risk factor for HNSCC with the incidence of HPV associated HNSCC increasing from 16.3% to 72.7% over 20 years in in North America (Chow 2020). Other important risk factors described in current literature are summarised in Table 2.

***Table 2. Modifiable risk factors for HNSCC and described markers.***

| Risk Factor | Molecular Mechanisms | Geographic High Incidence | Quantifiable Biomarkers | References |
|---|---|---|---|---|
| Alcohol | P53, ALDH1B, ALDH2 | Worldwide | Acetaldehyde | (Huang *et al*. 2017; Seitz and Stickel 2007) |
| Cannabis use | P38, MAPK | Worldwide | CNR1, CNR2 | (Liu *et al*. 2020a) |
| Chewing betel quid and areca nut | COX2, NF-kB, MAOA, | India, China | Arecoline | (Chang *et al*. 2014) |
| Immunocompromise | DC, APC, MP, TC, EP | Worldwide | VEGF | (Duray *et al*. 2010) |
| Ionizing radiation | DNA repair pathways | Uncommon | | (Gilbert 2009) |
| Malnutrition | DNA repair pathways | NL | Selenium | (Gromadzińska *et al*. 2008) |
| Poor oral hygiene | NL | NL | Missing teeth | (Hashim *et al*. 2016) |
| Reflux disease | NL | NL | NL | (Langevin *et al*. 2013) |
| Tobacco smoking | P53, EGFR | Worldwide | Nitrosamine Ketones, Arecoline | (Yang *et al*. 2018) |

References here indicated in this table are not exhaustive and represents an up to date study or review relevant to the factor of interest. CNR, cannabinoid receptor. ALDH, alcohol dehydrogenase. COX, cyclooxygenase. MAOA, monoamine oxidase. MAPK, . EGFR, epidermal growth factor receptor. DC, dendritic cells. APC, antigen presenting cells. MP, macrophages. TC, T-cells. ES, eosinophils. DNA, deoxyribonucleic acid. NL, not extensively described in literature. VEGF, vascular endothelial growth factor.

Tobacco smoking was identified as an independent risk factor for head and neck cancer in 1957 (Wynder and Bross 1957). Subsequently, studies identified a dose

dependent relationship between tobacco smoking, alcohol consumption and HNSCC (Brennan *et al*. 1995; Choi and Kahyo 1991). They also described the synergistically detrimental effects of these two common factors (Castellsague *et al*. 2004; Talamini *et al*. 2002). In Australia, the proportion of active smokers are low (Figure 11). However, there are many ex-smokers with an associated significant risk for HNSCC (Cao *et al.* 2016). Australian alcohol consumption is highest in the middle age group (35 – 65 years old) with males consuming alcohol at more than twice the rate of females (Figure 12).

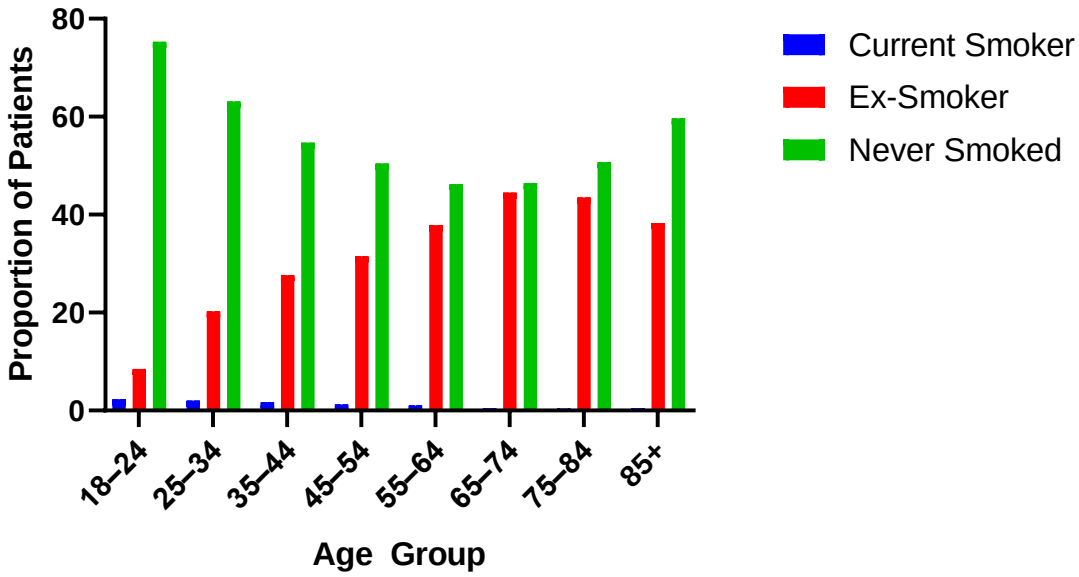



***Figure 11. Smoking status of patients with cancer in Australia. Data obtained from 2016 Australian Institute of Health and Welfare Cancer Database.***

Betel quid (also known as 'pan' in the Indian subcontinent) is the combination of betel leaf, areca-nut, tobacco, and other ingredients. Chewing of betel quid is a common social practice in large parts of India and China (Mehrtash *et al*. 2017; Sharma 2003). This accounts for approximately 600 million people with a large portion of this population at risk of oral and oropharyngeal SCC (Gupta and

Warnakulasuriya 2002). Macroscopic changes noted with betel quid chewing include ulceration, thickened epithelium, brown discolouration, submucosal fibrosis and pseudomembranous wrinkle (Chewer's mucosa) (Reichart and Phillipsen 1998). These correspond to histological characteristics such as ballooning, epithelial hyperplasia, inflammatory infiltrates, basal nuclei hyperkeratosis, pyknosis and varying degrees of dysplasia (Reichart and Phillipsen 1998).

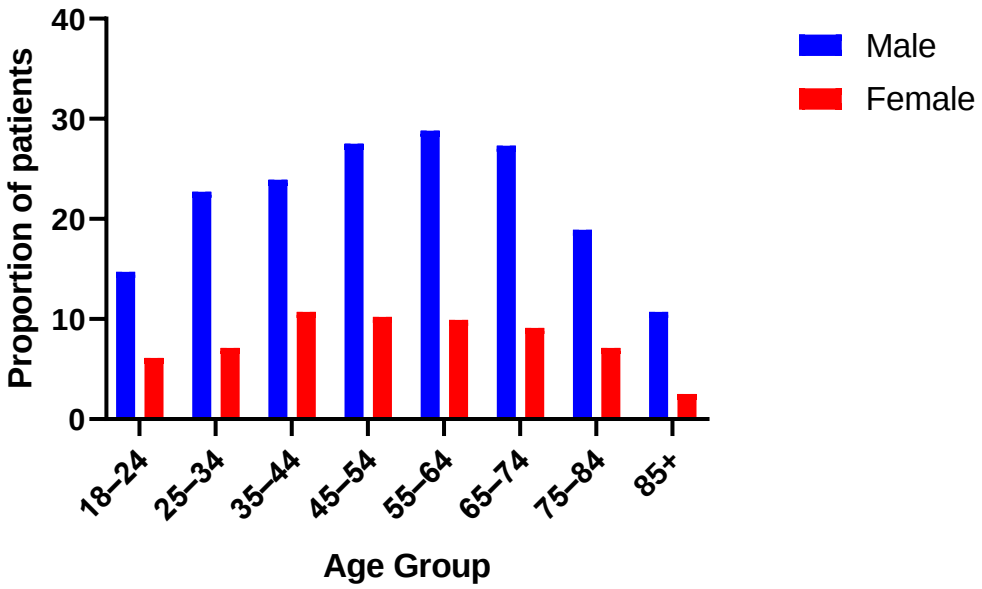



***Figure 12. Proportion of patients with cancer exceeding recommended alcohol intake guidelines. Data obtained from 2016 Australian Institute of Health and Welfare Cancer Database.***

## 1.4 Classification of Head and Neck Squamous Cell Carcinoma

Classification of HNSCC is important for prognostication and clinical decision making. HNSCCs can be classified according to anatomical subsites, macroscopic primary tumour characteristics, extent of metastasis, histological characteristics, molecular markers, and patient response to treatment.

### 1.4.1 Clinical and Pathological Staging

Clinical staging involves physical examination of the patient with the aid of radiological findings, whereas pathological staging is based on the analysis of histological specimens of patients treated with primarily surgical resection.

The American Joint Committee on Cancer (AJCC) regularly reviews staging systems to appropriately revise the staging based on current evidence. Until 2017, HNSCC was staged using the seventh edition of AJCC staging manual (Edge and Compton 2010). Staging of HNSCC is currently based on the International Union Against Cancer (UICC) TNM (Tumour, Nodes and Metastasis) classification of malignant tumours, 8th Edition or the AJCC staging manual, 8th Edition (Lydiatt *et al*. 2017). The eighth edition was released in mid-2016 and first used in clinical practice from January 2017.

Both systems (UICC and AJCC) classify T-stage based on the primary tumour size, N-stage based on the extent of locoregional metastasis to lymph nodes in the neck and M-stage based on the presence of distant metastases, beyond the head and neck region. Combination of the TNM staging allows for an overall four-tier prognostic staging system with stages 1-2 considered early and stages 3-4 considered advanced stage malignancy (Lydiatt *et al*. 2017). This staging allows for differentiation of patients between the various stages, with stage 1 having a better survival than stage 4 (Table 3 and Table 4). Staging is also used to guide treatment with single-modality versus multiple-modality treatment regimens for early stage and advanced stage HNSCC. However, treatment de-escalation based on new trial data continues to evolve and influence overall treatment decision making (Rosenberg and Vokes 2020).

There are several important changes between the 7th and the 8th editions of the TNM staging system for HNSCC. The inclusion of prognostic advantage of HPV for oropharyngeal SCC is the most significant change. Patients with multiple positive malignant nodes between 3 and 6 centimetres in maximal size are categorised as N1 in the 8th edition of TNM staging compared to the 7th edition where N1 was only for a single ipsilateral lymph node less than 3 cm in the greatest dimension (Lydiatt *et al*. 2017). In addition, there is a new pathological lymph node staging system introduced for surgically treated patients. A significant difference in treatment outcomes were noted depending on the absolute number of SCC positive lymph nodes in the neck (Table 5). Where, malignant lymph nodes between 1 and 4 had a better prognosis compared to patients with 5 or more SCC positive lymph nodes (Table 6). However, this staging criterion is clearly different in HPV negative tumours (Table 7).

***Table 3. Clinical staging of T-category for HPV positive oropharyngeal SCC (Lydiatt et al. 2017)***

| T Category | Clinical Criteria |
|---|---|
| T0 | No primary tumour identified |
| T1 | Primary tumour is 2cm* or smaller |
| T2 | Primary tumour is larger than 2cm* but smaller than 4cm* |
| T3 | Primary tumour is larger than 4cm* OR extends to the lingual surface of the epiglottis |
| T4 | Primary tumour invades following surrounding anatomy: larynx OR extrinsic tongue muscles OR medial pterygoid OR hard palate OR mandible OR beyond these anatomical landmarks |

HPV, human papilloma virus. SCC, squamous cell carcinoma. *, sizes mentioned above correlates to maximum tumour diameter. Note there is no carcinoma *in situ* category for HPV positive tumours. See section 1.3.1.2 for the description of anatomical landmarks.

***Table 4. Clinical staging of T-category for HPV negative oropharyngeal SCC (Lydiatt et al. 2017)***

| T Category | Clinical Criteria |
|---|---|
| Tx | Unable to assess the primary tumour |
| Tis | Carcinoma *in situ* only |
| T1 | Primary tumour is 2cm* or smaller |
| T2 | Primary tumour is larger than 2cm* but smaller than 4cm* |
| T3 | Primary tumour is larger than 4cm* OR extends to the lingual surface of the epiglottis |
| T4a | Primary tumour invades following surrounding anatomy: larynx OR extrinsic tongue muscles OR medial pterygoid OR hard palate OR mandible |
| T4b | Primary tumour invades following surrounding anatomy: lateral pterygoid muscle OR pterygoid plates OR lateral nasopharynx OR skull base OR encases internal carotid artery |

HPV, human papilloma virus. SCC, squamous cell carcinoma. *, sizes mentioned above correlates to maximum tumour diameter. See section 1.3.1.2 for the description of anatomical landmarks.

***Table 5. Clinical staging of N-category for HPV positive oropharyngeal SCC (Lydiatt et al. 2017)***

| N Category | Clinical Criteria |
|---|---|
| Nx | Cannot assess regional lymph nodes |
| N0 | No metastasis to regional lymph nodes |
| N1 | One or more, ipsilateral lymph nodes, less than or equal to 6cm*, positive for SCC |
| N2 | One or more, contralateral or bilateral lymph nodes, less than or equal to 6cm*, positive for SCC |
| N3 | Any lymph node larger than 6cm* |

HPV, human papilloma virus. SCC, squamous cell carcinoma. *, sizes mentioned above correlates to maximum tumour diameter.

***Table 6. Pathological staging of N-category for HPV positive oropharyngeal SCC (Lydiatt et al. 2017)***

| N Category | Pathological Criteria |
|---|---|
| Nx | Cannot assess regional lymph nodes |
| pN0 | No lymph nodes positive for SCC |
| pN1 | 4 or less lymph nodes positive for SCC |
| pN2 | More than 4 lymph nodes positive for SCC |

HPV, human papilloma virus. SCC, squamous cell carcinoma. Note pathological N-category staging is only applicable for surgically treated patients following histological analysis of neck dissection.

***Table 7. Clinical staging of N-category for HPV negative oropharyngeal SCC (Lydiatt et al. 2017)***

| N Category | Clinical Criteria |
|---|---|
| Nx | Cannot assess regional lymph nodes |
| N0 | No metastasis to regional lymph nodes |
| N1 | ECS negative, single, ipsilateral lymph node, less than or equal to 3cm*, |
| N2a | ECS negative, single, ipsilateral lymph node, between 3cm* and 6cm* |
| N2b | ECS negative, multiple ipsilateral lymph nodes, less than or equal to 6cm* |
| N2c | ECS negative, any number of bilateral or contralateral lymph nodes, less than or equal to 6cm* |
| N3a | ECS negative, any lymph node larger than 6cm* |
| N3b | ECS positive, any lymph node |

HPV, human papilloma virus. SCC, squamous cell carcinoma. ECS, extra-capsular spread. *, sizes mentioned above correlates to maximum tumour diameter.

Primary tumour depth of invasion (DOI) was also introduced as a fundamental criterion for upstaging the T-category for oral SCC (Table 8). This staging modification was as a result of improved statistical modelling of disease specific survival when depth of tumour invasion is incorporated into the survival model compared to using maximum primary tumour size alone (International Consortium for Outcome Research in *et al*. 2014). Previously, absolute tumour thickness was considered to have an impact on survival. However, evidence from International Consortium for Outcome Research (ICOR) indicated depth of invasion is more

congruent with survival than tumour thickness (International Consortium for Outcome Research in *et al*. 2014).

There has been considerable ambiguity and inconsistency in the measurement of tumour thickness of oral cavity SCC. Some earlier studies referred to DOI as the tumour thickness whereas others measured tumour thickness as the maximal vertical diameter perpendicular to the organ surface (Faisal *et al*. 2018). Therefore, largely exophytic tumours with a small DOI were considered to have the same risk as a largely endophytic tumour with large DOI. The 8$^{th}$ edition of TNM staging has defined DOI as the perpendicular measurement from a line drawn at the level of the basement membrane of normal adjacent tissue to the deepest level of invading tumour (Faisal *et al*. 2018). This multi-national research consortium recommended a 10mm DOI threshold to differentiate early and advanced oral cavity SCCs. Prior research by Yuen *et al* and Howaldt *et al* had also recommended 3mm and 5mm DOI respectively for the differentiation of T1 and T2 oral cavity tumours (Howaldt *et al*. 1999; Yuen *et al*. 2000). The 10mm DOI for primary oral tongue tumours has a distinct prognostic modelling advantage compared to the previously recommended DOI of 3 to 5 mm. Therefore, current staging system reflecting a) T1 lesions to be less than or equal to 5 mm DOI, b) T2 lesions to be between 5 and 10 mm DOI and c) T3 lesions to be larger than 10 mm DOI. Clinical and pathological staging of nodal status for oral cavity SCC is also separated to distinct clinical and pathological staging systems (Table 9).

***Table 8. Clinical staging of T-category for oral cavity SCC (Lydiatt et al. 2017)***

| T Category | Clinical Criteria |
|---|---|
| Tx | Unable to assess the primary tumour |
| Tis | Carcinoma *in situ* only |
| T1 | DOI less than or equal to 5mm[#], primary tumour is less than or equal to 2cm* |
| T2 | DOI between 5mm and 10mm primary tumour is less than 2cm* OR primary tumour is between 2cm* and 4cm* with DOI less than or equal to 10mm[#] |
| T3 | Primary tumour is larger than 4cm* OR DOI is larger than 10mm[#] |
| T4a | Primary tumour invades following surrounding anatomy: cortical bone (mandible/maxilla) OR inferior alveolar nerve OR floor of mouth OR skin of face |
| T4b | Primary tumour invades following surrounding anatomy: masticator space OR pterygoid plates OR skull base OR encase the internal carotid artery |

HPV, human papilloma virus. SCC, squamous cell carcinoma. DOI, depth of invasion. *, sizes mentioned above correlates to maximum tumour diameter. [#], depth of invasion is not equivalent to tumour thickness (See section 1.4.1). See section 1.3.1.2 for the description of anatomical landmarks.

***Table 9. Clinical staging of N-category for oral cavity SCC (Lydiatt et al. 2017)***

| N Category | Clinical Criteria |
|---|---|
| Nx | Cannot assess regional lymph nodes |
| N0 | No metastasis to regional lymph nodes |
| N1 | ECS negative, single, ipsilateral lymph node, less than or equal to 3cm*, |
| N2a | ECS positive, single, ipsilateral lymph node, less than or equal to 3cm* **OR** ECS negative, single, ipsilateral lymph node between 3cm* and 6cm* |
| N2b | ECS negative, multiple ipsilateral lymph nodes, less than or equal to 6cm* |
| N2c | ECS negative, any number of bilateral or contralateral lymph nodes, less than or equal to 6cm* |
| N3a | ECS negative, any lymph node larger than 6cm* |
| N3b | ECS positive, single, ipsilateral lymph node, larger than 3cm* **OR** any bilateral or contralateral lymph nodes that were ECS positive. |

HPV, human papilloma virus. SCC, squamous cell carcinoma. ECS, extra-capsular spread. *, sizes mentioned above correlates to maximum tumour diameter.

Unknown primary head and neck cancer (HNCUP) has its own staging criteria separated to clinical (Table 10) and pathological (Table 11) staging. Overall prognostic staging for HNCUP is separated to non-virally mediated (Table 12),

EBV mediated (Table 13) and HPV mediated (Table 14) sections with a clear survival benefit observed in virally mediated cancers.

***Table 10. HNCUP Clinical Staging of Lymph Nodes AJCC/UICC 2017 (Huang and O'Sullivan 2017)***

| Nodal Category | Clinical Criteria |
|---|---|
| Nx | Regional lymph nodes cannot be assessed |
| N0 | No regional lymph node metastasis |
| N1 | Metastasis in a single ipsilateral lymph node<br>3cm or smaller in greatest dimension<br>No extra-nodal extension |
| N2 | |
| N2a | Metastasis in a single ipsilateral lymph node<br>Larger than 3cm but smaller than 6cm<br>No extra-nodal extension |
| N2b | Metastasis in multiple ipsilateral lymph nodes<br>All smaller than 6cm<br>No extra-nodal extension |
| N2c | Metastasis in bilateral or contralateral lymph nodes<br>All smaller than 6cm<br>No extra-nodal extension |
| N3 | |
| N3a | Metastasis in a lymph node larger than 6cm<br>No extra-nodal extension |
| N3b | Metastasis in any lymph node with clinically overt extra-nodal extension or invasion of skin overlying the lymph node |

HNCUP – head and neck carcinoma of unknown primary, AJCC – American Joint Committee on Cancer, UICC – International Union Against Cancer, N – Neck

***Table 11. HNCUP Pathological Staging of Lymph Nodes AJCC/UICC 2017 (Huang and O'Sullivan 2017)***

| Nodal Category | Pathological Criteria |
|---|---|
| Nx | Regional lymph nodes cannot be assessed |
| N0 | No regional lymph node metastasis |
| N1 | Metastasis in a single ipsilateral lymph node<br>3cm or smaller in greatest dimension<br>No extra-nodal extension |
| N2 | |
| N2a | Metastasis in a single ipsilateral node<br>3cm or smaller<br>With extra-nodal extension<br>OR<br>Metastasis in a single ipsilateral lymph node<br>Larger than 3cm but smaller than 6cm<br>No extra-nodal extension |
| N2b | Metastasis in multiple ipsilateral lymph nodes<br>All smaller than 6cm<br>No extra-nodal extension |
| N2c | Metastasis in bilateral or contralateral lymph nodes<br>All smaller than 6cm<br>No extra-nodal extension |
| N3 | |
| N3a | Metastasis in a lymph node larger than 6cm<br>No extra-nodal extension |
| N3b | Metastasis in a single ipsilateral lymph node larger than 3cm<br>With extra-nodal extension<br>OR<br>Multiple ipsilateral, contralateral, or bilateral nodes of any size<br>With extra-nodal extension<br>OR<br>A single contralateral node smaller than 3cm<br>With extra-nodal extension |

HNCUP – head and neck carcinoma of unknown primary, AJCC – American Joint Committee on Cancer, UICC – International Union Against Cancer, N – Neck.

***Table 12. Non-virally mediated HNCUP Prognostic Staging AJCC/UICC 2017 (Huang and O'Sullivan 2017)***

| T – stage | N – stage | M – stage | Overall Stage |
|---|---|---|---|
| T0 | N1 | M0 | III |
| T0 | N2 | M0 | IV-A |
| T0 | N3 | M0 | IV-B |
| T0 | N – Any | M1 | IV-C |

HNCUP – head and neck carcinoma of unknown primary, AJCC – American Joint Committee on Cancer, UICC – International Union Against Cancer, N – Neck.

***Table 13. Epstein Barr Virus (EBV) mediated HNCUP Prognostic Staging AJCC/UICC 2017 (Huang and O'Sullivan 2017)***

| T – stage | N – stage | M – stage | Overall Stage |
|---|---|---|---|
| T0 | N1 | M0 | II |
| T0 | N2 | M0 | III |
| T0 | N3 | M0 | IV-A |
| T0 | N – Any | M1 | IV-B |

HNCUP – head and neck carcinoma of unknown primary, AJCC – American Joint Committee on Cancer, UICC – International Union Against Cancer, T – Primary tumour, N – Neck, M – Distant metastasis, EBV – Epstein Barr Virus.

***Table 14. Human Papilloma Virus (HPV) mediated HNCUP Prognostic Staging AJCC/UICC 2017 (Huang and O'Sullivan 2017)***

| T – stage | N – stage | M – stage | Overall Stage |
|---|---|---|---|
| T0 | N1 | M0 | I |
| T0 | N2 | M0 | II |
| T0 | N3 | M0 | III |
| T0 | N – Any | M1 | IV |

HNCUP – head and neck carcinoma of unknown primary, AJCC – American Joint Committee on Cancer, UICC – International Union Against Cancer, T – Primary tumour, N – Neck, M – Distant metastasis, HPV – Human Papilloma Virus.

### 1.4.2 Current Diagnostic Modalities of Head and Neck Squamous Cell Carcinoma

In current practice, the diagnostic approach to locating the primary cancer site relies on clinical examination with flexible nasendoscopic assessment, radiology,

rigid endoscopic biopsies and biomarker identification of cancer from fine needle aspirates or core biopsies of the cervical nodal disease (Weber *et al*. 2001). Radiologically, whole body positron emission tomography (PET) (Rudmik *et al*. 2011), high resolution contrast-enhanced computed tomography (CT) and magnetic resonance imaging (MRI) are utilised to clinically assess the size and location of the primary tumour as well as metastases (Cianchetti *et al*. 2009). Surrogate indirect and direct biomarkers such as HPV and Epstein Barr virus (EBV) from fine needle aspirates or core biopsies of the cervical nodal disease have also been used to identify possible primary cancer sites (Cheol Park *et al*. 2017). Tonsillectomy, targeted biopsies, or tongue base mucosectomy with trans-oral robotic surgery (TORS) are increasingly used to identify the primary cancer site in HPV positive HNSCC (Berta *et al*. 2014)

## 1.5 Pathophysiology of Head and Neck Squamous Cell Carcinoma

Known risk factors mentioned earlier as well as other sporadic causes lead to molecular insults that accumulate over time, and these contribute to the development of these highly aggressive malignancies. Established molecular pathways and those under investigation are reviewed in this section.

### 1.5.1 The molecular basis of squamous cell carcinomas

A complex cascade of molecular events gives rise to HNSCC. Numerous genes have been implicated in the pathogenesis of HNSCCs; these include NOTCH1, NOTCH2, NOTCH3, IRF6, TP53, CDKN2A, HRAS, PTEN, SYNE1, SYNE2, RIMS2, PCLO, Rb/INK4/ARF and PIK3CA (Stransky *et al*. 2011). The TP53 (p53) gene pathway is the most common cited mutation in more than half of all HNSCC

(Leemans *et al*. 2011). The p53 gene has been identified in malignant as well as pre-malignant lesions such as leucoplakia (Braakhuis *et al*. 2004). This has led to the "patch-field" progression theory of HNSCC, where a field of genetically abnormal mucosal tissue gains growth advantage and further mutations that progress to carcinoma (Rothenberg and Ellisen 2012). In addition, p53 mutations in adjacent tissue could be unique at a base pair level from the primary neoplasm, indicating the potential for metachronous tumours in the same patient after accumulation of further mutations in adjacent tissues (Nees *et al*. 1993). There is also an independent association with poor survival in patients with a truncating or function disrupting p53 mutation compared to patients without (Poeta *et al*. 2007).

Another key tumour suppressor gene pathway of importance is Rb/INK4/ARF (Rothenberg and Ellisen 2012). Under normal physiological conditions, the CDKN2A gene encodes for cell cycle regulators that include p16/INK4A and p14/ARF/INK4B among many others. Inactivation of the CDKN2A gene via the retinoblastoma (Rb) pathway has been reported in up to 30% of HNSCC (Ai *et al*. 2003). The p16/INK4A pathway is particularly important in HPV positive HNSCC that will be further explored in section 1.5.2.

The NOTCH pathway has been considered an evolutionarily highly conserved intracellular signalling pathway that is also implicated in HNSCC differentiation (Sun *et al*. 2014). In benign tissue, TP63 (p63) and NOTCH1 controls the squamous morphogenesis of mucosa (Agrawal *et al*. 2011). The transcription factor p63 is expressed in keratinocytes of the basal layer and maintains their potential for proliferation, while expression of NOTCH1 causes terminal differentiation into spinous and granular layers of the mucosa. Loss of NOTCH1

and mutated expression of p63 remove further barriers to neoplastic proliferation and survival of malignant cells (Agrawal *et al*. 2011).

Up to 40% of HNSCC has a phosphatase and tensin homolog (PTEN) mutation, causing the activation of the PI3K signalling pathway. PI3K pathway is negatively regulated by PTEN and positively regulated by PIK3CA. This is important in HPV positive HNSCCs as the combination of HPV E6 and E7 proteins with PIK3CA can lead to more invasive oropharyngeal carcinomas (Henken *et al*. 2011). The loss of p53, CDKN2A, TGFBR2/SMAD4 and amplification of CCND1 promotes progression and stops apoptosis of HNSCC (Rothenberg and Ellisen 2012). Invasive features of HNSCC are promoted by the loss of cell adhesion molecules such as FAT1, SMAD3 and TFGB1 (Rothenberg and Ellisen 2012).

The molecular genetics of HNSCC implicates various stages of cell differentiation, tumour genesis, tumour progression and tumour suppression, giving rise to a largely heterogeneous group of neoplasms despite a common and contiguous anatomical location. However, these molecular markers and their regulatory molecules are a source of biomarkers for detecting HNSCC. MicroRNAs play a key regulatory role in many molecular pathways including cancer. MicroRNAs and their regulatory role or associations with above molecular pathways and specific genes important for HNSCC are further discussed in sections 1.7.1 and 3.4.2.

### 1.5.2 The role of human papillomavirus

The *human papillomavirus* belongs to a large group of DNA viruses that forms the *Papillomaviridae family* with more than hundred subtypes identified in humans alone (Burd and Dean 2016). These non-enveloped, double stranded DNA

viruses have a predilection towards infecting epithelial cells such as the upper aerodigestive mucosal epithelium. The carcinogenic potential of HPV was first described by Harald zur Hausen in relation to cervical cancer (zur Hausen 1996). High risk HPV subtypes 16 and 18 have demonstrated a strong causal relationship to SCC (Pyeon *et al*. 2007). These high-risk HPV subtypes are associated with up to 30% of head and neck cancers (Gillison *et al*. 2015). The oncogenic potential of HPV is due to viral oncogenes E6 and E7, blocking the function of the tumour suppressor genes p53 and Rb, respectively (Hebner and Laimins 2006).

Initial epidemiological evidence of the oncogenic nature of HPV for HNSCC was first comprehensively described by Hobbs *et al* (Hobbs *et al*. 2006). Their systematic review and meta-analysis concluded a strong association exists between HPV subtype 16 and tonsil SCC with an odds ratio of 15.1 (95% CI 6.8 - 33.7) (Hobbs *et al*. 2006). Importantly, a positive HPV status in oropharyngeal HNSCC gave an independent prognostic advantage compared to HPV negative mucosal carcinomas (Gillison 2004). HPV status may also determine treatment response, as HPV positive patients responded to radiotherapy more effectively compared to HPV negative patients (Pyeon *et al*. 2007). A recent prospective trial identified a therapeutic and survival advantage for HPV positive oropharyngeal HNSCC patients (Schache *et al*. 2013). Although, this survival advantage is limited to non-smokers, the new evidence suggests that the survival outcome of current smokers was poor and independent of the HPV status (Grønhøj *et al*. 2019).

## 1.6 Current treatment modalities for head and neck squamous cell carcinoma

The treatment of HNSCC can be broadly divided into primary, adjuvant and salvage options (van Weert and Leemans 2020). The two primary treatment modalities are surgical excision of the primary tumour (with dissection of the locoregional lymph nodes) or primary radiation therapy (Shukla *et al*. 2020). Adjuvant radiotherapy may be required following initial surgery depending on histopathological high-risk features in the primary tumour, excision margin status and overall stage of the tumour (Ivaldi *et al*. 2019). Adjuvant radio-chemotherapy is reserved for specific patient groups postoperatively or as a primary treatment modality in combination with radiotherapy (Dauzier *et al*. 2019). However, emerging evidence suggests that treatment of early stage tumours with uni-modality or deescalated treatment protocols can provide equivalent oncological and functional outcomes to current treatment protocols (Schutte *et al*. 2020).

### 1.6.1 Surgery

The underlying principle behind the surgical treatment of HNSCC is to excise the tumour to obtain histologically clear margins while preserving or reconstructing the functional tissues of the head and neck (Bungum *et al*. 2020). Other than in exceptional circumstances of palliative surgery, there is no oncological role for primary debulking procedures in HNSCC (Jang *et al*. 2013). However, when patients present with extremis due to impending airway obstruction, procedures to debulk and secure the airway are conducted to preserve life (Fang *et al*. 2015).

The type and extent of surgery is subsite and primary tumour dependent. In the oral cavity, small tumours (early stage) can be primarily excised with adequate

margins whilst preserving function without the need for complex reconstruction. However, larger, and more involved tumours require extensive ablative surgery with soft tissue or bony reconstruction to rehabilitate functional outcomes (You *et al*. 2020). Trans oral surgery (with or without the aid of a robot) has improved surgical outcomes for oropharyngeal surgery as a primary treatment option with lower morbidity compared to open surgery conducted decades ago (Meccariello *et al*. 2019). A recent systematic review comparing transoral robotic surgery to primary radiotherapy has also indicated a survival advantage for robotic surgery patients (De Virgilio *et al*. 2020).

Regional neck metastasis of HNSCC can be treated primarily with surgery or radiotherapy. Surgically, the type of neck dissection performed is dependent on the site of the primary disease and radiological evidence of the burden of the nodal disease. Lymph nodes of the neck has been divided in to anatomically and oncologically distinct levels as detailed in Table 1. The comprehensive dissection of levels one to five is referred to as the modified radical neck dissection. This is distinctly 'modified' from the older descriptions of radical neck dissections where, functionally important structures such as the sternocleidomastoid muscle, the internal jugular vein and at times the spinal accessory nerve were sacrificed (Nahum *et al*. 1961). In the modern era, radical neck dissections are rarely performed and only indicated when there is clinical or radiological evidence of invasion into those important structures.

Elective neck dissection in the radiologically node negative neck continues to be debated, especially for very small primary tumours (D'Cruz *et al*. 2009). In the absence of a sensitive biomarker to predict locoregional involvement, most of the

patients receive at least a selective neck dissection (lymph node levels at risk are dissected) for histopathological staging of nodal disease status.

### 1.6.2 Radiation oncology

In the first half of the 20th century, radiotherapy predominated as the treatment modality for HNSCC (Cognetti *et al*. 2008) and continues to be an option for primary treatment, adjuvant therapy and as a palliative option for incurable disease (De Felice *et al*. 2018).

The aim of radiation therapy is to deliver high radiation energy to a specified field of tissue to kill malignant cells. This is achieved through the energy emitted as electrons or photons from a linear accelerator (Holsti 1995). Photons can penetrate deeper tissues while, electrons are able deliver effective radiation to superficial sites without causing injury to underlying tissues (Malicki 2015). This energy causes the ionisation of DNA at an atomic level, subsequently disrupting the molecular structure of DNA, initiating a lethal cascade. However, the effects are largely non-specific to healthy and malignant tissues (Bentzen *et al*. 2001). While healthy cells can repair their DNA to maintain cell survival to a degree, there are certain thresholds of radiation induced cell damage that are unable to be salvaged. Therefore, the fundamental principle of radiation oncology is the therapeutic ratio between malignant tumour lethality and damage to healthy tissues (Baskar *et al*. 2014). Fractionation of the total radiation dosage has improved this therapeutic ratio by allowing time for healthy tissues to repair in the midst of receiving radiation therapy (Arcangeli *et al*. 1979).

Since the advent of radiation therapy for treatment of solid cancers, there has been numerous technological improvements to increase effectiveness while reducing patient morbidity. The discovery of super-voltage radiation, where linear accelerators or telecobalt units could deliver adequate external beam radiation to deeper anatomical structures while causing limited harm to the superficial structures allowed for radiation oncology as a viable option for the treatment of HNSCC (Bernier and Bentzen 2006). However, in the last two decades, there has been considerable improvement to topographical planning, conformity, and effective delivery of radiation. Intensity modulated radiation therapy (IMRT) (Nutting *et al*. 2011), three-dimensional conformal radiotherapy (3D-cRT), volumetric modulated arc therapy (VMAT) (Brown *et al*. 2016) and image guided radiation therapy (IGRT) are some of the novel conformal approaches to effectively delivering radiation energy to the tumour while reducing injury to healthy tissues. Stereotactic radiosurgery (SRS) and stereotactic body radiation therapy (SBRT) refers to delivery of total dose of radiation in a single fraction or limited multi-fractions respectively (Chen and Girvigian 2006).

### 1.6.3 Medical oncology

Chemotherapy can be administered for specific groups of patients with HNSCC in multiple formats: a) concurrent chemoradiotherapy, b) induction chemotherapy prior to definitive chemoradiotherapy (Haddad and Shin 2008), c) adjuvant concomitant chemoradiotherapy and d) finally as palliative treatment of recurrent or metastatic disease (Chow 2020). The chemotherapy agents commonly used are platinum-based agents (cisplatin, carboplatin) (Chow 2020), 5-Fluorouracil, taxane-based agents (paclitaxel, docetaxel), cetuximab (Merlano *et al*. 2020), anti-metabolites (methotrexate, capecitabine) (Iqbal and Pan 2016), EGFR inhibitors

(afatinib) (Machiels *et al*. 2015) and immunotherapy agents (pembrolizumab, nivolumab) (Ho and Mehra 2019; Yen *et al*. 2020). These agents can be used as a mono-chemotherapy or poly-chemotherapy depending on disease status and intent of treatment guided by institutional protocols or in a clinical trial.

The individual patient data meta-analysis conducted by the MACH-NC study updated in 2011 has provided collated evidence for the use of chemotherapy in HNSCC (Blanchard *et al*. 2011). This meta-analysis of 87 trials with 16,485 patients showed that addition of concomitant chemotherapy to locoregional treatment of advanced HNSCC improved overall survival with a five-year risk of death decrease by 13% (Blanchard *et al*. 2011). However, the overall effect is dependent on the timing of chemotherapy and the anatomical subsite of HNSCC. Concomitant chemotherapy had the greatest effect (7% absolute benefit in two years and 8% absolute benefit in five years) compared to other delivery forms. In addition, platinum-based agents performed the best (Bourhis *et al*. 2007). However, the toxicity associated with the addition of chemotherapy is also greatest with platinum-based agents (Trotti *et al*. 2007). Interestingly, addition of chemotherapy mainly benefited locoregional control and had no effect on distant failure rates (Blanchard *et al*. 2011; Dauzier *et al*. 2019).

### 1.6.4 Integrated treatment approach

Treatment decision making in HNSCC is multidisciplinary and patient centred. It is vital to have this approach due to the heterogeneity of patient factors at presentation and variable tumour behaviour (Lo Nigro *et al*. 2017). Up to 40% of patients present with early-stage localised disease (Stage 1 or 2) and depending on the tumour subsite, these patients are treated with primary surgery or primary

radiotherapy. Patients with early stage HNSCC generally have a five-year overall survival up to 90% with low treatment related morbidity. However, in locoregionally advanced disease (Stage 3 or 4) multimodality treatment is required. A combination of primary surgery followed by adjuvant radiotherapy (with or without chemotherapy) or primary radiotherapy with concurrent chemotherapy is generally used in the treatment of advanced stage cancers (Homer and Fardy 2016).

### 1.6.5 Treatment related morbidity

Varying degrees of treatment related morbidity are associated with each of the treatment modalities described above. These can be generally divided to acute, short-term, and long-term issues. Surgery poses potential risk for acute anaesthetic and surgical complications. These complications range from the potential injury to neurovascular structures such as great vessels of the neck, facial nerve, vagus nerve, recurrent laryngeal nerve, superior laryngeal nerve, spinal accessory nerve and the hypoglossal nerve that can impede long term functional status of the patient, to airway compromise and potential death (Harréus 2013). Radiotherapy can cause short term oral mucositis and potential delayed radio-necrosis of bone or soft tissues (Bentzen *et al*. 2001). Vital structures such as the brain, spinal cord and eyes are also at risk of radiation injury during treatment. Delayed emergence of radiotherapy related secondary malignancies have also been documented. Chemotherapy can induce systemic immunological responses as well as an immunocompromised state increasing opportunistic infections and secondary malignancies. Therefore, it is recognised that cancers detected at an earlier stage can receive curative treatment with less treatment related morbidity due to de-escalation of treatment paradigms. This is emerging

as a treatment paradigm for HPV related HNSCC, where new trials are being conducted to determine the de-escalation of treatment to reduce patient morbidity, while maintaining treatment effectiveness (Mehanna *et al*. 2020).

HNSCC is an incredibly heterogeneous cancer with patient and tumour dependent outcomes. The emergence of prognostic surrogate biomarkers such HPV associated HNSCC with p16 immunohistochemistry has allowed for identification of improved prognosis and de-escalation of treatment to improve overall oncological and functional outcomes. This highlights the value of clinical biomarkers to improve patient outcomes from HNSCC, and the necessity to identify new biomarkers capable of a) detecting early stage cancers (potentially used for screening specific patients), b) minimally invasive diagnosis, c) prognosticating and d) of monitoring response to treatment. These would likely improve the ability to deliver personalised treatment plans to HNSCC patients.

## 1.7 Biomarkers in Head and Neck Squamous Cell Carcinoma

Biological markers or biomarkers are defined as accurate, reliable, and measurable biological markers predictive of a defined health status (Strimbu and Tavel 2010). The National Institutes of Health Biomarkers Definitions Working Group added to this definition and highlighted the importance of objectivity in biomarker measurements (Biomarkers Definitions Working 2001). Further expansions to the definition were conducted by the International Programme on Chemical Safety, Inter-Organisation Programme for the Sound Management of Chemicals and the World Health Organisation (Strimbu and Tavel 2010). Based on their expansion a biomarker is now defined as any objective measurement describing the interaction between a biological system and a potential hazard

where this measured response may be physiological, biochemical, or a result of a molecular level interaction (Strimbu and Tavel 2010).

In the setting of cancer, the ability to detect high risk individuals, the ability to diagnose cancer with high certainty and the ability to detect recurrence or predict prognosis are useful attributes of a biomarker. As described earlier various biological measurements (Section 1.4) are used to classify and prognosticate HNSCC patients but only after the diagnosis of cancer. Whilst extensive molecular mechanisms of HNSCC have been described (Section 1.5.1), only a handful of these attributes are used as a biomarker in routine clinical practice.

There are numerous reasons for the lack of useful biomarkers for early detection of HNSCC. The incredibly heterogeneous clinical, anatomical, histological, and molecular groupings of HNSCC, fundamentally impairs the ability to validate these biomarkers in large groups (Kim *et al*. 2014). However, methodological inconsistencies during the discovery phase of biomarkers have also been identified as a key impediment for reproducibility and clinical validation (Dharmawardana *et al*. 2019; Dharmawardana *et al*. 2020c; Lampri *et al*. 2015). This has led to the development of methodological and reporting standardisation to improve the consistency of biomarker research. Recently, the transparent reporting of multivariable prediction model for individual prognosis or diagnosis (TRIPOD) initiative (Moons *et al*. 2015) and the reporting recommendations for tumour marker prognostic studies (REMARK) (Sauerbrei *et al*. 2018) has paved the way for clinicians, primary researchers, bioinformaticians and journal editors to work collaboratively towards publishing high quality biomarker studies. The sources of variability relevant to circulating microRNA and exhaled breath

biomarker research is further discussed in Chapter 2 (Dharmawardana *et al*. 2019) and Chapter 4 (Dharmawardana *et al*. 2020c) respectively.

The earliest investigations into a HNSCC specific biomarkers included carcinoembryonic antigen (CEA) (Alsarraf *et al*. 1981), tumour associated antigen (TA-4) (Kato *et al*. 1979), squamous cell carcinoma antigen (SCC-antigen) (Fischbach *et al*. 1990; Imai *et al*. 2015) and carbohydrate antigen (CA-19-9) (Gustafsson *et al*. 1988). However, despite early promise, these markers were never clinically validated for detection or prognostication of HNSCC.

The measurement of HPV status is mainly via p16, a surrogate marker for HPV, where the extent of p16 protein expression provides indirect evidence of transcriptionally active HPV (Lassen *et al*. 2014). The p16 marker can be detected in cell blocks of fine needle aspirated samples and in resected samples using cytological techniques and immunohistochemistry, respectively (Wells *et al*. 2015). The sensitivity of detecting transcriptionally active HPV is significantly higher when 70% of tumour cells or the tissue sample is stained positive for p16 (Gronhoj Larsen *et al*. 2014). Another method of measuring HPV status is direct quantification of HPV DNA using polymerase chain reaction (PCR) or in-situ hybridization techniques (Wells *et al*. 2015). A meta-analysis of studies with combined identification of HPV status reported three survival groups. There is a high survival benefit in HPV positive and p16 positive patients, intermediate survival benefit for HPV negative but p16 positive patients, and a limited survival benefit for patients with p16 negative status regardless of HPV positivity (Coordes *et al*. 2016). The AJCC has recommended this combined approach for reporting HPV status for prognostic staging (Lydiatt *et al*. 2017). In Australia, it is rare for pathology institutes to conduct both types of testing as the prognostic advantage

described earlier does not alter the prescribed treatment. Presumably, future direction of precision medicine would necessitate the detailed testing of HPV status to provide less morbid and more effective personalised treatments.

The classical risk factors for HNSCC described earlier (Section 1.3.3) can be used as screening and prognostic markers. However, there are no clinically useful biomarkers to objectively measure a risk factor such as smoking. Tobacco smokers develop HNSCC with an odds ratio of 2.37 compared to non-smokers, and this increases to 5.73 when tobacco smoking is combined with alcohol consumption (Jethwa and Khariwala 2017). Therefore, current smokers with upper aerodigestive symptoms are at higher risk of having a HNSCC and should be closely monitored or referred for specialist assessment. However, this attribute is not included in the current HNSCC staging system. Additionally, the treatment efficacy of HNSCC is also impaired in current smokers. Where the five-year overall survival (OS) of smokers were significantly less compared to matched non-smokers (22% difference in 5 year -OS) (Chen *et al*. 2011). Chen *et al* also reported a higher radiotherapy related toxicity rate and a significantly higher rate of post-surgical complications in current smokers (Chen *et al*. 2011). There has been a significant reduction in tobacco smoking in Australia, United Kingdom, and the United States of America (Jethwa and Khariwala 2017). This reduced prevalence of smoking in the community, raises smoking status as unique marker with the potential of detecting high-risk patients for developing HNSCC and a poorer response to treatment. A biomarker capable of objectively assessing the smoking status beyond the patients' history may be useful to include in future staging systems.

The advent of immunotherapy for the treatment of cancer has uncovered further useful biomarkers in the form of immune check point modulators. In advanced oral and laryngeal SCC, increased Programmed Cell Death-ligand 1 (PD-L1) expression had a direct correlation with poor progression free survival (PFS) (Lin *et al*. 2015; Vassilakopoulou *et al*. 2016). Recently, PD-L1 has become a clinically relevant biomarker because of the FDA approval of pembrolizumab (a blocker of PD-L1 and PDL2) as a single agent, for the treatment of recurrent or metastatic HNSCC, requires patients to have a PD-L1 combined positive score (PDLCPS) of more than one (Rischin *et al*. 2019). PDLCPS is the number of PD-L1 staining tumour cells divided by the total number of viable tumour cells in sample multiplied by hundred. Therefore, a score of one or more indicates that at least 1% of the tumour cells are expressing PD-L1 (Cohen *et al.* 2019). This immunohistochemical method of quantifying a biomarker requires invasive sampling of tumour tissues (O'Malley *et al*. 2019). While useful in the setting of prescribing immunotherapy, a biomarker for early detection requires careful consideration of the sampling method for the type of biomarker.

Liquid biopsy has been described as an attractive option, due to ease of access and established serum sampling methods in medicine. Liquid biopsies are not limited to simple 'blood tests' and can also include serum, saliva, urine, stools, sweat and their respective sub-components (Bianchi 2015). These liquid biopsies can be obtained from various populations with ease. The ability to collect these samples pre-treatment, during treatment and post-treatment in a standardised manner allows for applications such as screening, prognostication, and treatment efficacy (Economopoulou *et al*. 2017; Holdenrieder 2016). However, it is critical to understand that the molecular make up of these samples are highly dependent on normal physiology and homeostasis (Hocking *et al*. 2016). The expectation has

been that cancer related changes to the tumour microenvironments would likely affect components within some of these circulating samples (Huang and Hoon 2016). In current clinical practice, serum-based monitoring of disease status remains in the forefront. Hence, circulating biomarker research forms a key component of cancer biomarker discovery. HNSCC specific serum based circulating biomarkers are further described in sections 1.7.1 and 2.1.

The analysis of human breath for potential biomarkers of cancer is an evolving science with promising research in the field of cancer (Abderrahman 2019a; Hanna *et al*. 2019; Markar *et al*. 2019) and other disease states (Alkhouri *et al*. 2014; Bayrakli *et al*. 2016; Li and Duan 2015; Pereira *et al*. 2015). It can be considered as an extension of a liquid biopsy and the non-invasive nature of breath analysis is the most attractive feature for screening and detection. It provides an overall metabolic profile of the subject based on exhaled breath content with variation dependent on content in a) serum (Grove *et al*. 2012; Turner *et al*. 2009; Wilson *et al*. 2001; Wilson *et al*. 2002), b) upper and lower respiratory tract (Rocco 2018), and c) digestive tract (Sloan *et al*. 2018) including the resident human microbiome (Kerlin and Wong 1988; Zhu *et al*. 2013b). The ability to sample the exhaled volatilome representative of several homeostatic systems is advantageous, while analytically challenging (Dharmawardana *et al*. 2020c).

### 1.7.1 Circulating Biomarkers of Cancer

Circulating cancer biomarkers are defined as cancer specific tumour cell components (DNA, RNA, proteins, and metabolites) detected in circulating body fluids (Rapisuwon *et al*. 2016). The ability to detect circulating tumour cells in serum was described, more than four decades ago, as associated with poor

disease specific survival (Kuper and Bignall 1966). More recently, circulating cell free DNA and RNA in serum is correlated to tumour burden (Stroun and Anker 2005; Swiecicki *et al*. 2019). Circulating antigens specific to squamous cell carcinoma were also detectable in serum showing promise as a biomarker (Wang *et al*. 2020). An overview of circulating cancer biomarkers is described here.

#### 1.7.1.1 Circulating Tumour Cells

Circulating tumour cells (CTCs) are defined as cells derived from a primary or metastatic tumour mass that has entered the circulating blood volume. Multiple mechanisms have been postulated for the presence of CTCs: 1) direct shedding into leaky vasculature surrounding the tumour, 2) lymphatic spread and 3) active migration of differentiated tumour cells into the vascular space (Dong *et al*. 2013; Gorges and Pantel 2013; Kulasinghe *et al*. 2015). CTCs were thought to represent a precursor to tumour metastasis (Kulasinghe *et al*. 2019; Watanabe 1954). CTCs are now classified into 3 types: Type 1 CTCs have identical genomic copy number changes to the primary tumour, Type 2 CTCs have genomic copy number changes present in the CTCs but not identical to the primary tumour, and Type 3 CTCs where no genomic copy number changes are detected (Chemi *et al*. 2019; Lampignano *et al*. 2019). This highlights the novel complexity in analysing CTCs as cancer biomarkers. Based on current understanding, CTCs possibly allows for biomarkers of diagnosis with advanced stage cancers. However, whether CTCs fundamentally allow for the diagnosis of early-stage cancers remains debatable.

#### 1.7.1.2 Circulating Cell-Free DNA

In 1977, circulating cell-free DNA was noted to be elevated in patients with cancer (Leon *et al*. 1977). Subsequent research discovered the ability to detect tumour

specific changes such as DNA methylation (Fujiwara *et al*. 2005) and microsatellite instability (Shaw *et al*. 2000). The size of the DNA molecules appears to be cancer dependent, and the primary mode of its release is thought to be via apoptosis (Stroun and Anker 2005). Many cancers have rapid cell turnover with simultaneous apoptosis, necrosis, and impaired phagocytosis, which leads to accumulation of cellular debris that is likely circulated as cell free DNA (Bellairs *et al*. 2017). Circulating tumour DNA (ctDNA) only represents less than 1% of circulating free DNA (Leary *et al*. 2012). Next generation sequencing platforms have significantly enhanced the ability to detect such low levels of ctDNA (Galot and Machiels 2020).

Positive correlations have been reported for circulating HPV DNA with an overall higher disease stage (Based on AJCC 7$^{th}$ edition) and higher neck staging, with no clear association to primary tumour size (Dahlstrom *et al*. 2015). Oropharyngeal HNSCC, patients with advanced stage disease and no smoking history had the highest HPV-16 DNA detectable in serum (Mazurek *et al*. 2019). Interestingly, undetectable levels of circulating HPV DNA prior to treatment resulted in better progression free survival (Dahlstrom *et al*. 2015). These findings were in keeping with a direct relationship between circulating cancer markers and tumour burden. Contrary to the above studies that suggest ctDNA as a precursor to metastasis, smaller case series have reported patients with clinically proven metastasis without evidence of ctDNA (Egyud *et al*. 2019). Also, at a population level HPV DNA and HPV specific antibodies are detectable in serum without any associated disease (Liu *et al*. 2016). Therefore, it is important to carefully interpret circulating HPV DNA findings in the specific context of disease. Circulating tumour DNA may also provide a pre-treatment measurement of radio-sensitivity (Willers *et al*. 2004), where DNA damage response foci can be analysed to determine the radio-

sensitivity of patients prior to receiving radiation therapy, thus individualising their treatment (Willers *et al*. 2015). Overall, circulating tumour DNA analysis shows potential as a diagnostic or a prognostic biomarker for HNSCC. However, in current clinical practise, circulating DNA testing is only utilised in foetal-maternal medicine, where germline mutations were detected by analysing circulating foetal DNA in maternal plasma (Fan *et al*. 2008).

#### 1.7.1.3 Circulating extracellular vesicles

Soluble mediators present in the tumour microenvironment are thought to be pertinent for cell-to-cell communication and critical for tumour growth while interacting with the host immune system. These extracellular vesicles were initially considered to be a cellular by product. However, recent findings highlighted an important role of these extracellular vesicles for intercellular communication in physiological and pathological states (Bellmunt À *et al*. 2019). They include exosomes and micro-vesicles (Whiteside 2018) with the distinction between micro-vesicle subtypes and exosomes defined as per the updated minimal information for studies of extra cellular vesicles (MISEV-2018) position paper (Thery *et al*. 2018).

Exosomes are defined as small vesicles that contains proteins, RNA species and lipids. These distinct biological entities range in nanoscale sizes between 30nm to 100nm and have been found in various body fluids and intracellular spaces. They are thought to be molecules capable of acting through cell-to-cell adhesion molecules or secretion of soluble factors (Bock *et al*. 2001). They are formed by the invagination of cell endosomal membrane into the cytoplasm forming multi-vesicular bodies. The proteins and RNA initially contained in exosomes were from

host cytoplasm. However, additional material can be incorporated into these vesicles following their formation. These mechanisms of protein transport into exosomes have been extensively reviewed (McKelvey *et al*. 2015) and not discussed in further detail here. However, the dynamic ability to alter the contents of exosomes points to their potential in detecting malignant transformation of healthy cells.

Exosomes forms an integral component of cancer specific circulating biomarkers due to their unique features. They are detectable in various body fluids accessible for analysis. These exosomes circulate for a limited period with a half-life in plasma of approximately two minutes. These vesicles contain molecules that are highly specific to the parent cell of origin on the surface (antigens) as well as within (mRNA/microRNA) the exosome. They are thought to represent markers of active cancers. In contrast, circulating cell free DNA, that is thought to arise from cell apoptosis (Rapisuwon *et al*. 2016) may not be specific to a viable malignancy.

Recent literature presents many biomarkers associated with extracellular vesicles and HNSCC (Yap *et al*. 2020). However, none of them were clinically utilised, in part, due to rapid evolution of methods and analytical techniques as discussed in Chapter 2.

### 1.7.2 Saliva biomarkers of cancer

Human saliva is produced by six major salivary glands and hundreds of minor salivary glands that drain into the oral cavity and the pharynx. Saliva samples can be obtained non-invasively with or without stimulation. While cancers can directly develop from salivary glands it is rare to have primary SCCs arising from salivary

gland parenchyma. However, locoregional effects of mucosal HNSCCs can affect the microenvironment of salivary glands and their physiology. Saliva is comprised of water, electrolytes, proteins and microbes, which can be analysed for biomarker discovery (Proctor 2016). Initial studies have focused on saliva volume, metabolites, and antibodies. However, more recently, the analysis of saliva based nucleic acids (Jiang *et al*. 2005; Rapado-González *et al*. 2018; Roesch *et al*. 2020) and microbiome (Furquim *et al*. 2017b; Guerrero-Preston *et al*. 2017) have been investigated in the context of HNSCC.

#### 1.7.2.1 Saliva microbiome and cancer

The vast array of human oral microflora has been described prior to the term microbiome being coined (Dewhirst *et al*. 2010). Microbiome refers to the total community of microorganisms within a specified niche that may have a symbiotic or pathological role. The term encompasses all microbes, whether or not they can be grown in culture. However, there remains debate as to whether the term microbiome include viral species present in the same ecosystems (Neu and Mainou 2020). This relatively new method of defining the human microbiome is made possible mainly using 16S rRNA genomic platforms, that are able to identify in excess of 600 phylotypes (Dewhirst *et al*. 2010). The 16S rRNA methodology and its utility in HNSCC is further described in section 5.3. The Human Microbiome Project and various other large population level projects have highlighted the importance of the human microbiome and its association to disease states including cancer (Allaband *et al*. 2019). The oral cavity microbiome is unique where individual subsites of the oral cavity (Section 1.3.1.1), despite being limited to a small anatomical space, form distinct ecosystems of microbes (Li *et al*. 2013). Saliva does exhibit extensive individual variability, potentially

influenced by the increased exposure to the external environment, other humans and animals (air, diet, and the exchange of saliva and skin flora) as well as individual physiology (Bik *et al*. 2010; Dassi *et al*. 2014). However, the salivary microbiome has been considered representative of the oral cavity, as it provides a homogeneous representation from a biomass perspective compared to swabs of individual subsites.

The association between cancer and microbes is well established in various organ systems. For example, 1) HPV in cervical cancer (Chang and Parsonnet 2010; zur Hausen 2009), 2) *Helicobacter pylori* in gastric adenocarcinoma (Shiota and Yamaoka 2014), 3) *Neisseria* species in oesophageal carcinoma (Peters *et al*. 2017) and 4) *Fusobacterium nucleatum* in colorectal carcinoma (Furquim *et al*. 2017a). In HNSCC, salivary microbes such as *Fusobacterium nucleatum*, *Lactobascillus gasseri/johnsonii* and *Lactobacillus vaginalis* are associated with oral carcinomas (Gholizadeh *et al*. 2016; Guerrero-Preston *et al*. 2017). Oral microbiome panels have shown to be capable of accurately classifying HNSCC patients from controls (Lim *et al*. 2018).

### 1.7.3 Breath Biomarkers of Cancer

The use of human olfaction to detect human diseases date back to 400bc, when odours emitted from various sources such as breath, skin, urine, or faeces were associated with various diseases (Curtis 2015). Interestingly, canine studies have demonstrated the ability for dogs to detect human conditions such as cancer with olfaction alone (Gordon *et al*. 2008). In the context of HNSCC, various reports have described the pungent halitosis noted in patients with oral cavity cancers (Uppal and Singh 2016). These ‘odours’ comprise a complex mixture of volatile

organic compounds (VOCs) that have been broadly termed the ‘volatilome’. It is likely that a complex interaction between human metabolism, dietary intake and the resident microbiome contribute to the measured volatilome.

A VOC is broadly defined as an organic molecule present in a gas state at ambient temperature (Võ and Morris 2014). In humans, the source of VOCs can be vast with significant contribution from the microbiome and other environmental factors have been identified (Opitz and Herbarth 2018). Exhaled VOCs comprise of endogenous compounds that are a product of cellular respiration and metabolism as well as products from the resident microbiota and their fermentation products. They also include a fraction of exogenous or environmental VOCs that have been inhaled at the time of collection. Exhaled breath is described as mixed alveolar gas, where the composition is largely nitrogen, oxygen, argon, and carbon dioxide. It also contains water vapour and approximately one percent of the mixed alveolar gas are VOCs. Therefore, individual VOCs are only present in trace concentrations (Lourenco and Turner 2014). Recent report indicates up to 872 distinct VOCs have been identified in human exhaled breath and up to 1840 total VOCs identified from all human biological products (de Lacy Costello *et al*. 2014).

A link between exhaled breath and cancer was first implied in the setting of colorectal cancer and exhaled methane and hydrogen analysis (Haines *et al*. 1977; Sivertsen *et al*. 1992). However, both the analytical technology as well as our understanding of the volatilome has advanced significantly with various exhaled compounds reported to be associated with human cancers (Abderrahman 2019a; Hanna *et al*. 2019). Exhaled breath analysis has drawn interest as a possible source for novel biomarkers of HNSCC, and several VOCs have been associated with HNSCC (Table 15). In addition, various pattern recognition

instruments and statistical models have identified accurate HNSCC specific exhaled breath profiles (Table 16).

***Table 15. VOCs associated with head and neck squamous cell carcinoma.***

| Study | Method | Reported VOC name | CAS | Molecular Weight (g/mol) | m/z |
|---|---|---|---|---|---|
| (Chandran *et al*. 2019) | SIFT | Hydrogen Cyanide | 74-90-8 | 27.03 | NR |
| (Hartwig *et al*. 2017) | GC | Dibutyl hydroxytoluene | | 220.35 | NR |
| | | Dimethyl disulphide | 624-92-0 | 94.20 | |
| | | Decamethylcycopentasiloxane | 541-02-6 | 370.77 | |
| | | Methyl ethyl ketone (MEK) | 78-93-3 | 72.11 | |
| | | n-heptane | 142-82-5 | 100.20 | |
| | | p-xylene | 106-42-3 | 106.16 | |
| | | toluene | 108-88-3 | 92.14 | |
| | | 1-heptene | 592-76-7 | 98.19 | |
| (Gruber *et al*. 2014) | GC | Ethanol | 64-17-5 | 46.07 | 31 |
| | | 2-Propanenitrile | 107-13-1 | 53.06 | 53 |
| | | Undecane | 1120-21-4 | 156.31 | 57 |
| (García *et al*. 2013) | GC | Ethanol | | | 46 |
| | | 2-Butanone | 64-17-5 | 46.07 | 72 |
| | | 2,3-Butanediol | 78-93-3 | 72.11 | 90 |
| | | 9-Tetradecen-1-ol | 513-85-9 | 90.12 | 212 |
| | | Octene derivative | 52957-16-1 | 212.37 | 126 |
| | | Cycloheptane derivative | | | 108 |
| | | Cyclononane derivative | | | 136 |
| (Hakim *et al*. 2011) | GC | 4,6-Dimethyldodecane | 611-72-8 | 198.39 | 57 |
| | | 2,2-Dimethylpropanoic acid | 75-98-9 | 102.13 | 57 |
| | | 5-Methyl-3-hexanone | 623-56-3 | 114.19 | 57 |
| | | 2,2-Dimethyldecane | 17302-37-3 | 170.33 | 57 |
| | | Limonene | 138-86-3 | 136.23 | 68 |
| | | 2,2,3-Trimethyl-, exo-, bicyclo[2.2.1]heptane | 20536-41-8 | 138.25 | 95 |

CAS, Chemical abstracts service. VOC, volatile organic compounds. SIFT, selected ion flow tube. GC, gas chromatography. m/z, mass to charge ratio. NR, not reported.

***Table 16. Reported sensitivity and specificity of VOC biomarkers to differentiate between HNSCC and healthy controls.***

| Study | Method | Sample Size (Cancer / Control) | Sensitivity (%) | Specificity (%) | NPV (%) | PPV (%) |
|---|---|---|---|---|---|---|
| Chandran 2019 | Syft® SIFT-MS | 23 / 21 | 91 | 76 | 88.6 | 80.4 |
| Gruber 2014 | Sensor-Array | 22 / 19 | 77 | 90 | NR | NR |
| Hakim 2011 | NA-Nose® | 22 / 40 | 100 | 92 | NR | NR |
| Leunis 2014 | DiagNose® | 36 / 23 | 90 | 80 | NR | NR |

SIFT-MS = selected ion flow tube mass spectrometry, NPV = negative predictive value, PPV = positive predictive value, NR = not reported

A recent review has also identified the importance of exhaled breath condensate (EBC) in HNSCC biomarker research (Mäkitie *et al*. 2020). EBC occurs because of cooling and condensation of exhaled aerosols that largely originating from the lower respiratory tract (Horváth *et al.* 2005). This exhaled condensate is versatile and can be used to study respiratory epithelial function, molecular pathways, respiratory microbiome, genomic and metabolomic function (Konstantinidi *et al.* 2015). In the context of HNSCC, there are no studies exploring the utility of EBC as a biomarker of disease detection.

# 2 Sources of variability in circulating biomarker research methodology

## 2.1 Introduction

The utility of microRNAs as biomarkers in cancer diagnosis and prognosis has been an emerging field in various solid organ tumours (Tiberio *et al*. 2015). A microRNA is defined as single-stranded noncoding RNA molecule (approximately 22 nucleotides in length) first described in the context of developmental timing in bacteria (Lee *et al*. 1993). These highly conserved molecules play a significant role in post-transcriptional regulation of genes (Lee *et al*. 1993; Reinhart *et al*. 2000). microRNAs regulate gene expression by binding to messenger RNA (mRNA) and this microRNA-mRNA complex may undergo degradation or cause translational repression (Bartel 2004). Mature single stranded microRNA binds to multiprotein RNA-induced silencing complex (miRNAISC) while being guided to target mRNA (Trang *et al*. 2009). The ability of microRNAs to regulate gene expression has been linked to normal physiology and pathological conditions, including carcinogenesis (Bartel 2004).

A stable form of microRNAs present in human plasma and serum has been extensively investigated as minimally invasive cancer biomarkers (Mitchell *et al*. 2008). In the setting of HNSCC, circulating microRNA have been described in multiple anatomical sub-sites including oral cavity, nasopharynx, larynx, salivary glands and cutaneous malignancies (Poel *et al*. 2018). The various methods used to collect, store, process and interpret these microRNAs are likely to introduce bias and contribute to the inconsistent validation results (Poel *et al*. 2018).

The purpose of conducting this scoping review of the recent literature is to identify the available evidence on technical details and challenges encountered during the

study of circulating microRNAs from clinical samples in the context of head and neck cancer.

## 2.2 Methods

### 2.2.1 Data sources and search strategy

This scoping review was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines and scoping review guidelines from Joanna Briggs Institute (Munn *et al*. 2018). The search strategy captured studies in English language published between April 2007 to May 2018 (Ten Years) and excluded non-human studies. The last ten years were included as the landmark study describing circulating microRNAs for detecting cancer was published in 2008 (Mitchell *et al*. 2008). PubMed and Embase databases were searched for titles and abstracts using the keywords listed in Appendix II. Trial searches indicated that searching for only head and neck cancer related keywords missed key studies that were found previously using free hand search likely due to non-descriptive titles. Therefore, cancer related keywords were used for the database search, then titles and abstracts were individually screened for studies relevant to head and neck cancer. The primary author conducted the search in consultation with other authors and disputes of inclusion or exclusion of studies were resolved with discussion. The PRISMA diagram is shown in Figure 13 with the breakdown of systematic search results.

### 2.2.2 Inclusion criteria

Studies in English language describing microRNA markers for detection of head and neck cancer (without discriminating for specific subsite or pathological entity)

in a circulating specimen (whole blood, plasma, or serum) were included. A critical appraisal for risk of bias was not conducted, hence did not determine the inclusion or exclusion of studies as guided by Munn *et al* 2018 (Munn *et al*. 2018). However, potential sources of variability and bias were detailed in results and discussion sections. Search strategy is listed in Appendix II.

### 2.2.3 Exclusion criteria

Citations of literature reviews, conference abstracts, letters to the Editor, retracted studies, posters, or theses were excluded. When a study by the same author group describing the same technique were published at multiple time points, the earliest published study was included, and the rest excluded if the full text review didn't identify any variations in experimental technique. Studies that referred to a previously published article for their detailed methodology were excluded and the original article was included for review, provided it was within the same search time frame. Non-human studies, benign diseases, malignancies outside of the head and neck region and studies referring only to tissue microRNAs were excluded from further review.

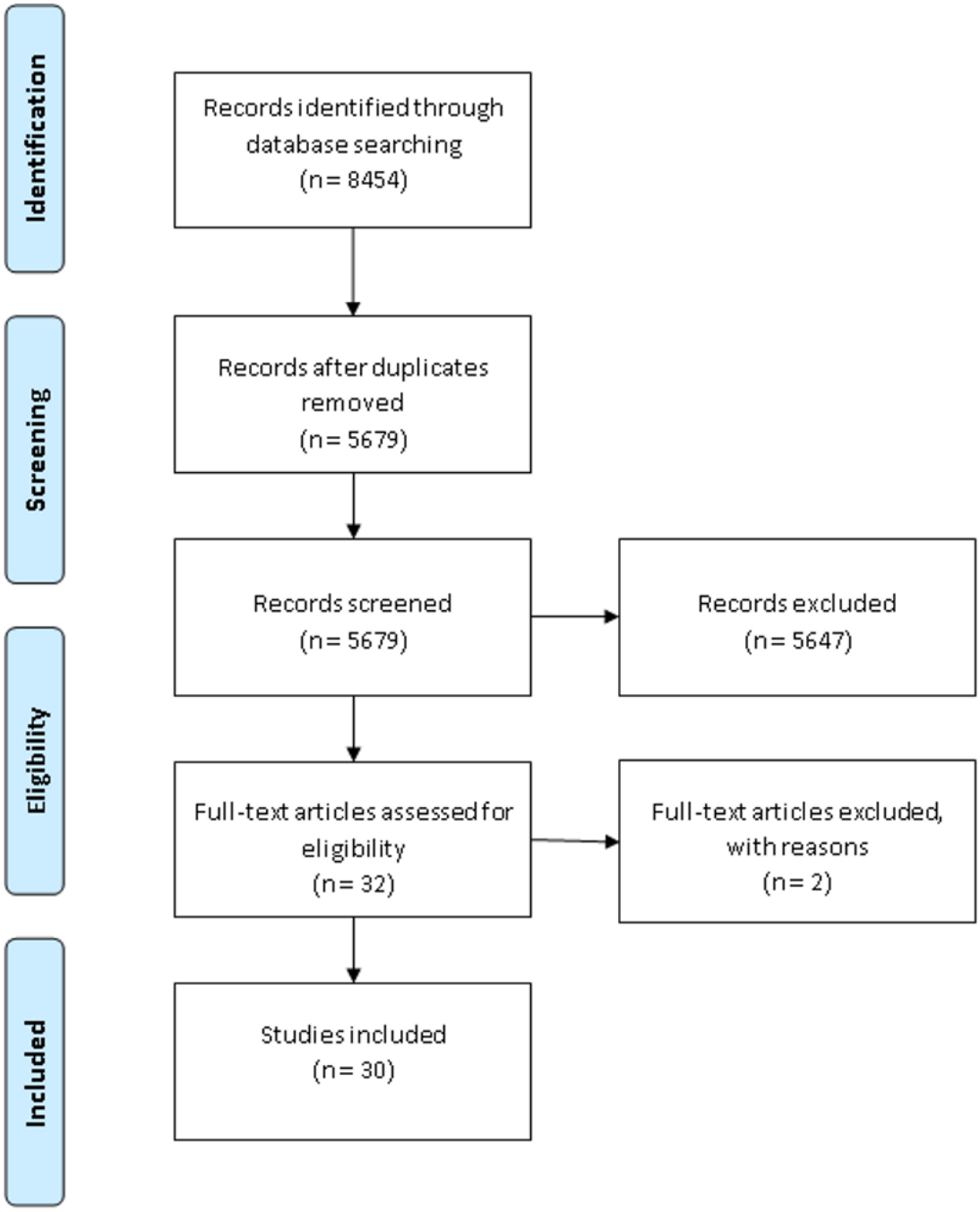



***Figure 13. PRISMA flow diagram of systematic search and inclusion of studies for further review***

### 2.2.4 Data Extraction

The key information extracted included protocols used for patient preparation (i.e. fasting status, pre/post anaesthetic, pre/post treatment), sample collection, sample storage, sample pre-processing, centrifugation, RNA isolation, RNA quantification

and RNA normalisation methods. The details of reagent kits used for each of the RNA processing steps were also extracted.

## 2.3 Results

The search identified 5679 articles following removal of duplicate results. Title and abstract screening against the inclusion criteria excluded 5647 studies. Two further studies were excluded following full text review as they referred to another study for their detailed methodology. Finally, thirty studies were included for data extraction. The study institutes and patient populations were from 13 countries with majority of publications resulting from China. Publication years ranged from 2010 to 2018. Pathology of interest included mucosal malignancies of the following subsites a) oral cavity, b) nasopharynx, c) oropharynx, d) hypopharynx, e) upper oesophagus, and f) the larynx. Only one study focused on primary salivary gland malignancies (Cinpolat *et al*. 2017). Majority of the studies used patient populations with mucosal squamous cell carcinomas. There were no studies that included cutaneous malignancies.

### 2.3.1 Patient Preparation

All except one study (Ricieri Brito *et al*. 2010) collected pre-treatment blood samples as well as sampling at various post treatment timepoints. Majority of studies did not describe the fasting status, the anaesthetic status, or the clinical setting of sample collection. Only Xu *et al* described the collection of fasting samples in an outpatient setting (Xu *et al*. 2016a).

### 2.3.2 Sample Collection, Incubation and Storage

Overall, details of sample collection process were poorly reported (Table 17). The use of vacutainers with EDTA was described in thirteen studies (Cinpolat *et al*. 2017; Gourzones *et al*. 2013; Gourzones *et al*. 2010; Hsu *et al*. 2012; Martinez *et al*. 2015; Poel *et al*. 2018; Rabinowits *et al*. 2017; Ricieri Brito *et al*. 2010; Summerer *et al*. 2013; Summerer *et al*. 2015; Wei *et al*. 2016a; Xu *et al*. 2018b; Yan *et al*. 2017). Four studies (Lerner *et al*. 2016; Ries *et al*. 2017; Ries *et al*. 2014a; Ries *et al*. 2014b) (three studies by the same author) reported the use of PAXgene RNA Tube (Switzerland). However, it was unclear if the bloods samples were directly collected into RNA tubes in a clinical setting or later aliquoted in the lab. Other studies (Hou *et al*. 2015a; Schneider *et al*. 2018; Sun *et al*. 2018a; Sun *et al*. 2016a; Tachibana *et al*. 2016; Xu *et al*. 2016a; Yu *et al*. 2015a) reported collection of a certain volume of blood or plasma but did not detail the process any further (

Figure 14).


***Figure 14. Patient preparation and venous sampling***

***Table 17. Types of blood collection tubes used by individual studies. Majority of the studies in this review were using EDTA containing collection tubes.***

| Collection Tube | Author (Year) |
|---|---|
| **PAXgene® RNA Tube** | Lerner C, *et al*. (2016)<br>Ries J, *et al*. (2014) |
| **Tube Containing EDTA** | Poel D, *et al*. (2018)<br>Xu X, *et al*. (2018)<br>Yan Y, *et al*. (2017)<br>Rabinowits G, *et al*. (2017)<br>Cinpolat O, *et al*. (2017)<br>Wei L, Mao M, Liu H (2016)<br>Martinez BV, *et al*. (2015)<br>Summerer I, *et al*. (2013)<br>Hsu CM, *et al*. (2012)<br>Gourzones C, *et al*. (2010)<br>Ricieri Brito JA, *et al*. (2010) |
| **Undefined** | Sun G, *et al*. (2018)<br>Schneider A, *et al*. (2018)<br>Tachibana H, *et al*. (2016)<br>Xu H, *et al*. (2016)<br>Sun L, *et al*. (2016)<br>Yu Q, *et al*. (2015)<br>Hou B, *et al*. (2015) |

EDTA, ethylenedaminetetraacetic acid. RNA, ribonucleic acid

All studies reviewed here stored their samples in a -80°C freezer prior to RNA isolation. However, the process taken prior to the freezing step varied between studies (Figure 15). Six studies specified exact incubation times prior to centrifugation steps of their sample. These incubation periods ranged from 15 minutes to 24 hours (Table 18). Poel *et al* experimented pre-processing time intervals of 1, 2, 4- and 24-hours post sample collection. All their samples were stored upright at RT for 1 hour and then on ice for various time intervals until centrifugation (Poel *et al*. 2018). Xu *et al* incubated their samples in a 4°C refrigerator for 1 hour prior to centrifugation (Xu *et al*. 2016b). Martinez *et al* incubated their samples at RT for 15 minutes prior to centrifugation steps

(Martinez *et al*. 2015). Liu *et al* incubated their samples for 1-hour post collection (Liu *et al*. 2014). Zeng *et al* described incubation of collected samples for 1 hour at room temperature followed by centrifugation steps to isolate serum (Zeng *et al*. 2012). Maclellan *et al* described incubation for 30 minutes at room temperature followed by centrifugation steps and storage of supernatant (Maclellan *et al*. 2012).

***Table 18. Various incubation periods for blood samples prior to freezing or centrifugation.***

| Author (Year) | Incubation Time and Temperature |
|---|---|
| | Multiple Incubation Time Comparisons |
| | 1 hour – at RT – no ice |
| Poel et al (2018) | 2 hours – 1 hour at RT then on ice |
| | 4 hours – 1 hour at RT then on ice |
| | 24 hours – 1 hour at RT then on ice |
| Xu et al (2018) | Variable up to 4 hours at RT |
| Xu et al (2016) | 1 hour at 4◦C |
| Martinez et al (2015) | 15 minutes at RT |
| Liu et al (2014) | 1 hour at RT |
| Summerer et al (2013) | Variable up to 2 hours at RT |
| Gourzones et al (2013) | Variable up to 2 hours unknown temperature |
| Zeng et al (2012) | 1 hour at RT |
| Maclellan et al (2012) | 30 minutes at RT |
| Hsu et al (2012) | Variable up to 4 hours at 4◦C |

RT, room temperature

Four studies provided a maximum time to centrifuge from the time of collection. However, there was no indication of an average time of incubation. While some studies indicated the incubation temperatures others did not. Xu *et al* stored their blood samples at RT until centrifugation and all samples were processed within 4 hours (Xu *et al*. 2018b). Summerer *et al* centrifuged their sample tubes within 2 hours of collection but did not specify incubation temperature (Summerer *et al*. 2013; Summerer *et al*. 2015). Gourzones *et al* centrifuged the sample within 2 hours of collection (Gourzones *et al*. 2013). Hsu *et al* described storage of

samples at 4 degrees immediately after collection for a maximum of 4 hours prior to centrifugation (Hsu *et al*. 2012).

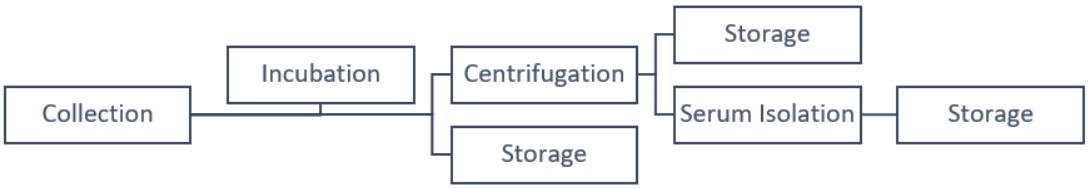



***Figure 15. Sample collection, incubation, centrifugation, and storage***

Ricieri Brito *et al* was the only study to describe the storage of blood samples in EDTA tubes at -80°C prior to any processing or serum isolation (Ricieri Brito *et al*. 2010). Six of the studies reviewed indicated separation of serum or plasma from the whole blood but did not adequately describe the process between collection and freezing of supernatant (Gourzones *et al*. 2010; Liang *et al*. 2016; Sun *et al*. 2016b; Tachibana *et al*. 2016; Wei *et al*. 2016a; Yu *et al*. 2015a). Five further studies had very poor descriptions of the process prior to RNA isolation steps (Hou *et al*. 2015a; Plieskatt *et al*. 2014b; Rabinowits *et al*. 2017; Schneider *et al*. 2018; Xu *et al*. 2016a). Studies that utilised whole blood stored in PAXgene blood RNA tubes froze the samples at -80°C immediately after collection (Lerner *et al*. 2016; Ries *et al*. 2017; Ries *et al*. 2014a; Ries *et al*. 2014b).

### 2.3.3 Centrifugation

Centrifugation steps ranged from a single step up to three steps (Table 19). The relative centrifugation force (RCF) for the first spinning cycle ranged from 120g to 10000g. The second spinning cycle RCF ranged between 360g and 16000g. The studies that had a third spinning cycle the RCF ranged from 2700g to 14000g.

The temperature during centrifugation ranged from 4°C to 20°C, although, four studies did not report the temperature. Nine of the reviewed studies did not adequately report the centrifugation steps used for isolation of specific blood components prior to RNA isolation (Cinpolat *et al*. 2017; Hou *et al*. 2015b; Plieskatt *et al*. 2014a; Rabinowits *et al*. 2017; Ricieri Brito *et al*. 2010; Schneider *et al*. 2018; Sun *et al*. 2018b; Tachibana *et al*. 2016; Xu *et al*. 2016a). Xu *et al* 2016 and Yan *et al* 2017 only reported centrifugation in revolutions per minute without describing the details of the centrifuge to calculate a comparable RCF.

***Table 19. Summary of various centrifugation steps and configurations***

| Author | Year | Centrifugation Steps |
|---|---|---|
| Poel *et al* | 2018 | <u>Serum tests:</u><br>1500g for 10 min at RT<br><u>Whole blood tests:</u><br>120g for 20 min at RT<br>Supernatant further spun 360g for 20min at RT<br>Supernatant further spun 2700g for 10 min at RT |
| Xu *et al* | 2018 | 1500g for 10 min at RT |
| Yan *et al* | 2017 | 10000 rpm for 10 min at 4°C |
| Xu *et al* | 2016 | 500 rpm for 3 min<br>Temperature not reported |
| Wei *et al* | 2016 | 2000g for 10 min at 4°C<br>Supernatant further spun 12000g for 10 min 4°C |
| Sun *et al* | 2016 | 3500g for 5 min at RT |
| Liang *et al* | 2016 | 1000g for 5 min at RT<br>Supernatant further spun 12000g for 5 min at 4°C |
| Yu *et al* | 2015 | 3000 rpm for 5 min at RT<br>Supernatant further spun 12000 g for 5 min at 4°C |
| Martinez *et al* | 2015 | 1300g for 10 min<br>Supernatant further spun 16000g for 15 min<br>Temperature not reported |
| Liu *et al* | 2014 | 500g for 10 min at 4°C<br>Supernatant further spun 10000g for 30 min at 4°C |
| Summerer *et al* | 2013 | 350g for 10 min<br>Supernatant further spun 1200g for 3 min<br>Supernatant further spun 14000g for 10 min<br>Temperature not reported |
| Zeng *et al* | 2012 | 1600g for 10 min at 4°C |
| Maclellan *et al* | 2012 | 1500g for 15 min at RT |
| Hsu *et al* | 2012 | 3500g for 15 min<br>Temperature not reported |
| Gourzones *et al* | 2010 | 1700g for 15 min at 20°C |

### 2.3.4 RNA isolation

A wide range of reagents and RNA isolation kits were described in the studies reviewed (Table 20). Earlier studies used TRIzol™ Reagent (Thermo Fisher

Scientific®) and miRVana™ PARIS™ kit (Thermo Fisher Scientific®). Variations of miRNeasy™ RNA isolation kits (QIAGEN®) were predominantly used after 2015. Studies that used whole blood as the primary sample source predominantly used the PAXgene™ Blood microRNA kit (PreAnalytiX®, QIAGEN®, AMBION®) (Lerner *et al*. 2016; Ries *et al*. 2017; Ries *et al*. 2014a; Ries *et al*. 2014b; Yu *et al*. 2015a). Only one study (Rabinowits *et al*. 2017) isolated exosome fraction for analysis using the ExoQuick™ Exosome Precipitation Solution (System Biosciences, California, USA).

Other kits utilised for RNA isolation included QIAamp™ Circulating Nucleic Acid Kit (QIAGEN®), QIAamp™ RNA Blood Mini kit (QIAGEN®), 3D-Gene™ RNA extraction reagent (TORAY®), High-Pure™ microRNA Isolation Kit (Roche® Diagnostics GmbH) and miRCURY™ RNA Isoloation kit (Exiqon®, QIAGEN®). The quantity of raw sample used varied substantially and ranged between 100 – 1000 microlitre for plasma, 200 – 500 microlitre for serum and 250 – 2500 microlitre for whole blood. Seven of these studies did not report exact quantity of raw sample used for the RNA isolation step (Cinpolat *et al*. 2017; Lerner *et al*. 2016; Liang *et al*. 2016; Rabinowits *et al*. 2017; Sun *et al*. 2016b; Xu *et al*. 2016b; Yu *et al*. 2015a). There was no consistency in the quantity of raw sample input based on the specific kit used for microRNA isolation.

***Table 20. Various RNA isolation methods described in literature.***

| Author | Year | Volume | Raw Sample | Reagents/Kits |
|---|---|---|---|---|
| Poel *et al* | 2018 | 300microL | Serum or plasma | miRCURY RNA isolation kit |
| Scheider *et al* | 2018 | 300microL | Serum | miRNeasy serum/plasma kit |
| Sun *et al* | 2018 | 200microL | Plasma | miRNeasy Mini kit |
| Xu *et al* | 2018 | *500microL | Plasma | RNeasy Mini kit, TRIzol LS reagent |
| Rabinowits *et al* | 2017 | Unknown | Plasma | TRIzol reagent<br>ExoQuick serum exosome precipitation solution |
| Yan *et al* | 2017 | 1000microL | Plasma | miRNeasy serum/plasma kit |
| Cinpolat *et al* | 2016 | Unknown | Serum | High-Pure miRNA isolation kit |
| Lerner *et al* | 2016 | Unknown | Whole Blood | PAXgene blood miRNA kit |
| Liang *et al* | 2016 | Unknown | Serum | QIAamp RNA Blood Mini kit |
| Sun *et al* | 2016 | Unknown | Serum | miRNeasy kit |
| Tachibana *et al* | 2016 | 1000microL | Plasma | 3D-Gene RNA extraction kit – for micro-array |
| | | 100microL | Plasma | miRNeasy kit – for qRT-PCR |
| Wei *et al* | 2016 | 200microL | Plasma | miRNeasy serum/plasma kit |
| Xu, H *et al* | 2016 | Unknown | Serum | miRNeasy RNA isolation kit |
| Xu, Y *et al* | 2016 | 1500microL | Serum | miRNeasy RNA isolation kit |
| Hou *et al* | 2015 | 200microL | Plasma | miRNeasy serum/plasma kit |
| Martinez *et al* | 2015 | 500microL | Serum | miRNeasy kit |
| Yu *et al* | 2015 | Unknown | Whole Blood | QIAamp RNA Blood Mini kit |
| Liu *et al* | 2014 | 400microL | Serum | TRIzol LS reagent |
| Plieskatt *et al* | 2014 | 500microL | Serum | QIAamp Circulating Nucleic Acid kit |
| Ries *et al* | 2014 | 2500microL | Whole Blood | PAXgene blood miRNA kit |
| Gourzones *et al* | 2013 | 300microL | Plasma | miRVana PARIS kit |
| Summerer *et al* | 2013 | 300microL | Plasma | miRVana miRNA isolation kit |
| Hsu *et al* | 2012 | 500microL | Plasma | miRVana PARIS kit |
| MacLellan *et al* | 2012 | 200microL | Serum | miRNeasy Mini kit |
| Zeng *et al* | 2012 | 500microL | Serum | miRVana PARIS kit |
| Gourzones *et al* | 2010 | 100microL | Plasma | miRVana miRNA isolation kit |
| Ricieri-Brito *et al* | 2010 | 250microL | Whole Blood | TRIzol reagent |

*Xu *et al* 2018 reported the use of 500mL of raw sample volume – assumed to be an error, hence reported here as 500microL to fit with normal range of sample volume.

### 2.3.5 Quantification and Profiling of microRNA

Table 21, Table 22 and Table 23 illustrates the range of reagents and instruments used for quantification of microRNA in these studies. There was minimal standardisation in reporting of measured units, reagent, and instrument nomenclature.

A large array of reagent kits was utilised for single tube quantitative reverse transcription polymerase chain reaction (qRT-PCR). Studies prior to 2012 exclusively used TaqMan based assays, and these continue to be used at present. However, miScript and miRCURY RT-PCR kits with SYBR-Green master mixes were more commonly used since 2012. Wei *et al* 2016 described the use of a droplet digital PCR method not described in other studies reviewed here (Wei *et al*. 2016a).

***Table 21. Single tube based qRT-PCR techniques used for relative quantification of microRNAs – Part 1.***

| Author | Year | Total RNA | qRT-PCR kits |
|---|---|---|---|
| Poel *et al* | 2018 | 2microL | miRCURY LNA Universal RT microRNA PCR kit<br>ExiLENT SYBR Green master mix |
| Xu *et al* | 2018 | | All-in One miRNA first strand cDNA synthesis kit<br>All-in One miRNA qPCR kit |
| Sun *et al* | 2018 | 5microL | TaqMan microRNA assays |
| Yan *et al* | 2017 | | Universal cDNA synthesis kit II<br>Pick-&-Mix microRNA PCR Panels |
| Xu, H *et al* | 2016 | | One-Step PrimeScript® miRNA cDNA synthesis kit<br>ABI 7500 system<br>SYBR Green PCR Master Mix |
| Xu, Y *et al* | 2016 | 2microL | TIANGEN RT PCR reaction system<br>SYBR Green I<br>CFX96 Touch RT PCR Detection System |
| Wei *et al* | 2016 | | M-MLV reverse transcriptase, stem-loop RT primer for miRNA<br>TaqMan universal master mix II |
| Tachibana *et al* | 2016 | 5microL | High capacity cDNA RT kit<br>7500 Fast Real-Time PCR system |
| Sun *et al* | 2016 | | miScript SYBR Green PCR kit<br>7500 Real-Time PCR System |
| Liang *et al* | 2016 | 1microg | miRNA reverse transcription kit<br>miRNA Q-PCR detection kit<br>ABI 7500 thermocycler |
| Lerner *et al* | 2016 | | TaqMan miRNA assays<br>TaqMan miRNA RT kit<br>TaqMan gene expression master mix |

***Table 22. Single tube based qRT-PCR techniques used for relative quantification of microRNAs – Part 2.***

| Author | Year | Total RNA | qRT-PCR kits |
|---|---|---|---|
| Yu *et al* | 2015 | 2microg | SYBR PrimeScript miRNA RT-PCR kit<br>7500 Real-Time PCR System |
| Hou *et al* | 2015 | 3microL | Universal RT reaction (LNA) PCR Primers<br>miRCURY LNA RT miRNA PCR<br>miRCURY LNA Universal RT cDNA synthesis kit<br>ABI Step One Plus Real-time PCR System<br>Applied Biosystems SDS |
| Reis *et al* | 2014 | 500ng | miScript II RT Kit<br>miScript SYBR-Green PCR Kit<br>ABI-7300 Sequence Detection System |
| Liu *et al* | 2014 | 200ng | Bulge-loop miRNA-specific RT primers<br>M-MLV reverse transcriptase<br>Platinum SYBR-Green qPCR SuperMix-UDG reagents<br>PRISM 7900HT System |
| Summerer *et al* | 2013 | *3microL of cDNA | TaqMan miRNA RT Kit<br>miRNA-specific stem-loop primers<br>Cyclone PCR system<br>ViiA 7 real-time PCR system |
| Zeng *et al* | 2012 | 5microL | miScript RT Kit<br>miScript SYBR-Green PCR Kit |
| Maclellan *et al* | 2012 | 19.2microL | miRCURY LNA Universal RT miRNA PCR, Polyadenylation and cDNA synthesis Kit<br>miRNA Ready-to-Use PCR, Human panel I and panel II<br>SYBR-Green master mix<br>Viia7 Software |
| Hsu *et al* | 2012 | 10ng | Stem-loop miRNA specific primers<br>TaqMan miRNA assay<br>TaqMan Universal PCR Master Mix |
| Ricieri-Brito *et al* | 2010 | 1microg | Stem-loop miRNA specific primers<br>TaqMan miRNA assay<br>Step-One real-time PCR 48-well plate<br>TaqMan Universal PCR master mix |
| Gourzones *et al* | 2010 | 9.16microL | TaqMan miRNA RT Kit<br>TaqMan miRNA Assay<br>Stem-loop miRNA specific primers<br>FastStart Universal Probe Master mix<br>StepOnePlus Detection System |

The use of microarrays for microRNA profiling and quantification was described since 2012 within the reviewed articles (Table 23). Various technologies were utilised including TaqMan, miScript, miRCURY, 3D-Gene, SurePrint and Agilent. Three studies also described Illumina TruSeq library preparation kit for microRNA sequencing (Martinez *et al*. 2015; Schneider *et al*. 2018; Yan *et al*. 2017) and

discussion of the analytical variables involved in sequencing is outside of the scope of this review.

***Table 23. Array-based PCR techniques used for relative quantification and profiling of microRNAs.***

| Author | Year | Total RNA | qRT-PCR kits |
|---|---|---|---|
| Xu *et al* | 2018 | | miRCURY Hy3/Hy5 power labelling kit<br>miRCURY LNA array |
| Rabinowits *et al* | 2017 | 60ng | TaqMan array human miRNA platform<br>Agieltn 2100 Bioanalyzer<br>Human 384-well TaqMan Array 2-card set v3.0<br>7900HT Real-Time PCR system |
| Cinpolat *et al* | 2016 | | miScript II RT kit<br>miScript Microfluidics preamp kit<br>miScript miRNA Assays<br>Dynamic Array 96.96<br>BioMark System |
| Xu, H *et al* | 2016 | | miRCURY LNA Array<br>Axon GenePix 4000B microarray scanner<br>GenePix Pro 6.0 |
| Tachibana *et al* | 2016 | 5microL | miRCURY LNA Array<br>3D-Gene Human miRNA Oligochips<br>3D-Gene Scanner<br>3D-Gene Extraction software |
| Liang *et al* | 2016 | 1microg | miRNA reverse transcription kit<br>miRNA Q-PCR detection kit |
| Lerner *et al* | 2016 | | SurePrint G3 x 60 K miRNA microarray |
| Yu *et al* | 2015 | 2microg | SYBR PrimeScript miRNA RT-PCR kit |
| Hou *et al* | 2015 | 3microL | Sureprint G3 Human v16 miRNA microarray<br>GeneSpring GX |
| Reis *et al* | 2014 | 500ng | GeniomH real-time analyser<br>Agilent SurePrint G3 Human v16 miRNA Array kit<br>Geniom Wizard Software |
| Plieskatt *et al* | 2014 | 32-125ng | miScript RT II Kit<br>miScript SYBR-Green PCR Kit<br>Custom printed 96 well miScript miRNA arrays<br>BioRad iCycler iQ5 |
| Liu *et al* | 2014 | 100ng | Agilent human miRNA v12.0 microarray<br>Agilent microarray scanner<br>Gene Spring Software 11.0 |
| Summerer *et al* | 2013 | 12microL | TaqMan Array Human MicroRNA A Cards v2.0<br>TaqMan miRNA RT Kit<br>Stem-loop megaplex primer pool set A v2.1<br>Applied Biosystems 7900HT |
| Zeng *et al* | 2012 | 3microL | TaqMan miRNA RT Kit<br>TaqMan low density array<br>TaqMan Human miRNA arrays<br>Applied Biosystems 7900HT thermocycler<br>SDS version 2.0 |

Normalisation for data analysis was conducted through endogenous controls, exogenous controls, or a combination of both. snRNA U6 was the most abundantly used endogenous control and spiked in cel-miR-39 is the most commonly exogenous control. Only two studies reviewed utilised both endogenous and exogenous controls for normalisation (Table 24).

***Table 24. Endogenous and exogenous control microRNAs used for experiment normalisation.***

| Author | Year | Normalisation |
|---|---|---|
| | | Endogenous Control |
| Xu *et al* | 2018 | U6 snRNA, miR-634, miR-1228-3p (average) |
| Rabinowits *et al* | 2017 | Global mean value of all miRNA |
| Lerner *et al* | 2016 | RNU48 |
| Liang *et al* | 2016 | U6 snRNA |
| Sun *et al* | 2016 | RNU6B |
| Tachibana *et al* | 2016 | Let7a, RNU6B |
| Wei *et al* | 2016 | U6 snRNA |
| Xu, H *et al* | 2016 | U6 snRNA |
| Xu, Y *et al* | 2016 | U6 snRNA |
| Hou *et al* | 2015 | Average of miR-93, miR103a, miR-191, miR-423 |
| Yu *et al* | 2015 | RNU6B |
| Liu *et al* | 2014 | Endogenous miR-16 |
| Plieskatt *et al* | 2014 | SNORD and RT controls |
| Ries *et al* | 2014 | SNORD44, U6 snRNA |
| Hsu *et al* | 2012 | U6 snRNA |
| Maclellan *et al* | 2012 | ROX reference dye |
| Gourzones *et al* | 2010 | hsa-miR-21, hsa-miR-146a, RNU44 |
| Ricieri-Brito *et al* | 2010 | RNU48, U47, RNU44 |
| | | Spiked-in Control |
| Yan *et al* | 2017 | Spike-in Control (Qiagen) |
| Summerer *et al* | 2013 | Spiked in cel-miR-39 |
| Zeng *et al* | 2012 | Spiked in cel-miR-39, cel-miR238 |
| Gourzones *et al* | 2010 | Spiked in cel-miR-39 |
| | | Combination of Controls |
| Poel *et al* | 2018 | miR-16-5p, cel-miR-39-3p (and their average) |
| Sun *et al* | 2018 | cel-miR-39, U6 snRNA |

miR, microRNA. snRNA, small nuclear ribonucleic acid

## 2.4 Discussion

Discovery and development of circulating biomarkers for head and neck cancer has several important steps to consider. These include discovery phase, verification, and validation phase, and lastly, the clinical application and revalidation phase. In order for circulating microRNA to progress as a biomarker (discrete microRNAs or panel of microRNAs) into standard clinical practice, the findings from discovery and validation phases need to be reproducible and stable (Poel *et al*. 2018).

In the discovery phase of identifying a circulating microRNA biomarker for head and neck cancer, high-throughput techniques are used. However, the lack of standardisation of sample collection and processing techniques has led to inconsistent reports and failure of any microRNA biomarker to enter validation phases or subsequent translation into routine clinical practice. This scoping review demonstrates the large variability from a technical viewpoint in the setting of circulating microRNA biomarker discovery for head and neck cancer. Previous reviews have identified the type of sample, the methods used for measurement and the strategies used for normalisation as potential sources of inconsistency (He *et al*. 2015). The potential contribution to bias from patient preparation, sample handling and inconsistency in microRNA measurement and profiling techniques are reported here. In the setting of circulating microRNA, there is only one study (Poel *et al*. 2018) that directly compared various methods using the same set of patient samples.

Most studies reviewed here collected samples from fasted patients, prior to commencing any treatment. However, patient preparation and categorisation may

need be more stringent. Length of fasting, type of food ingested, and their hydration status may affect the expression of circulating microRNA (Lopez *et al*. 2018). However, practically, above standardisation process is difficult to achieve. Nonetheless, it needs to be comprehensively reported in order to consider the potential that the microRNA difference is due to patient preparation factors rather than the disease of interest.

Patient comorbidities such as diabetes status may also affect the expression of these microRNA (Zhang *et al*. 2016). While in large patient groups such bias may be negligible, smaller cohorts may require strict inclusion and exclusion criteria in discovery and validation phases to remove such bias. The contribution from anaesthetic agents on circulating microRNA should also be considered, as some study protocols collected blood samples from an anaesthetised patient compared to pre-anaesthetic. Propofol a common anaesthetic agent, has been well documented in cell culture models to affect the expression of certain microRNA (Su *et al*. 2014; Zhang *et al*. 2013). Volatile anaesthetic agents, such as sevoflurane, can also alter expression of microRNA in animal studies and human cell lines (Tanaka *et al*. 2012; Yi *et al*. 2016). It is important to note that the evidence of volatile anaesthetic agents and alterations to microRNA expression has not been shown in human studies, however, there is potential for this factor to introduce bias in small patient samples.

Human blood samples consist of specific components a) red blood cells, b) platelets, c) white blood cells and d) plasma. Expression of microRNA in fractionated human blood is demonstrated to be significantly different based on the blood components assessed (Glinge *et al*. 2017). Therefore, it is important to be consistent with the use of certain blood collection tubes and to maintain a high-

quality standard for sample collection and handling to prevent contamination of various blood components (due to haemolysis, anticoagulation, or clotting). All studies reviewed here collected venous blood samples, predominantly in EDTA containing blood collection tubes using a vacutainer system. A comparative study by Glinge *et al* 2017 reports no measurable difference in microRNA quantification with tubes containing serum separator or anticoagulants, except reporting the lack of microRNA detection in lithium-heparin plasma containing tubes (Glinge *et al*. 2017).

The interval between sample collection and serum/plasma separation and storage could potentially contribute to higher rates of haemolysis and hence further inconsistency in microRNA quantification (Poel *et al*. 2018). Incubation time for blood samples in the studies reviewed here ranged from 30 minutes to 24 hours at various temperatures including RT, 4°C or -80°C. Significant changes in microRNA levels after 72 hours were reported when incubated at room temperature, while remaining stable for at least 24 hours (Glinge *et al*. 2017). A sample dependent significant fluctuation in haemolytic index is also demonstrated to increase over the incubation time for serum samples (Poel *et al*. 2018). The same study reported that plasma samples that were depleted with platelets did not demonstrate the same variation in haemolytic index, while a microRNA dependent variation is observed based on incubation time (Poel *et al*. 2018). Long term storage of samples at -80°C is also reported to introduce variability depending on fractionation, where storage of plasma is stable over-time, up to 9 months compared to whole blood where the microRNA concentration increased over-time likely due to microRNA release from platelets and red blood cells (Glinge *et al*. 2017). The effect of freeze thaw cycles has also been studied suggesting

unpredictable and inconsistent changes in microRNA levels (Zhao *et al*. 2014). A single freeze-thaw cycle was reported in all the studies reviewed here.

Mechanical disturbances to the blood samples during sample handling and transportation could introduce variation to microRNA levels, where Glinge *et al* 2017 reported a microRNA dependent (but not a sample fractionation dependent) change in relative levels following eight hours of mechanical disturbance by using a sample shaker (Glinge *et al*. 2017). While this was an interesting finding, samples are practically unlikely to encounter eight hours of mechanical stress in a clinical research setting. However, the results add to the knowledge that fraction separation prior to sample transportation does not affect the variation in microRNA observed. Centrifugation steps contribute to intra-assay variation as reported by Binderup *et al* 2018, only when the quantification is normalised to endogenous controls (Binderup *et al*. 2018). Studies in this review had centrifugation protocols ranging between one and three steps, with variations in centrifugation time and force potentially contributing to inconsistency in microRNA levels.

Haemolysis during sample preparation clearly contributes to variability in detected circulating microRNA levels (Kirschner *et al*. 2011). However, from the studies included here only Poel *et al* reported the use of haemolytic index to quantify the level of haemolysis in their samples (Poel *et al*. 2018). Maclellan *et al* mentions the exclusion of haemolysed samples but did not describe the method of quantification or identification of haemolysis (Maclellan *et al*. 2012). Tachibana *et al* commented on avoiding haemolysis but did not describe how this was achieved (Tachibana *et al*. 2016). Measurement of the level of sample haemolysis to either exclude samples or to normalise based on a marker such as microRNA-16 that

changes with the level of haemolysis (Kirschner *et al*. 2011) is crucial for the methodological validity in this setting.

RNA isolation method is a potential source of variation in measured microRNA levels. Binderup *et al* 2018 reports up to 73% of the variation is due to differential microRNA isolation methods (Binderup *et al*. 2018). Eight distinct RNA isolation protocols by different manufacturers were described in the studies reviewed here. It is difficult to comment on the superiority of one protocol over another, mainly because no direct comparisons have been reported in current literature. The methods may vary in their ability to liberate nucleic acid contained within extracellular vesicles, which contribute an important fraction of circulating microRNA (Urabe *et al*. 2017). An ideal comparative study would use aliquots from the same blood samples with standardised collection and pre-processing to isolate RNA using various methods and report on the yield. Interestingly, the number of commercially available kits for microRNA isolation has been reduced due to QIAGEN® company purchasing other branded RNA isolation products such as Exiqon® and incorporating these high-throughput technologies into a single protocol or product series. This unification of various commercially available kits contributes positively to the standardisation process while likely reducing competitiveness in the commercial market and possibly increasing costs for laboratories.

Binderup *et al* 2018 reports good correlation between single tube qRT-PCR and digital droplet PCR (ddPCR). However, this is only true when exogenous normalisation is used (Binderup *et al*. 2018). Comparison of single tube qRT-PCR to microarray-based PCR by Binderup *et al* 2018 indicated a poor correlation of relative microRNA levels. They utilised TaqMan based single tube PCR and low-

density arrays for this experiment. The similar technology used here provides for a valid comparison and the variability observed is potentially related to different reaction volumes and other variables such as pre-amplification steps used in microarray-based PCR. There also appears to be a microRNA dependent variability in the correlation between single tube and array-based PCR techniques (Binderup *et al*. 2018; Jorge *et al*. 2017). The thirty studies reviewed here used three different microRNA quantification methods that included single tube qRT-PCR, microarray-based PCR and ddPCR. Many of the studies that used microarray-based PCR also conducted single tube qRT-PCR for validation of results.

Normalisation is a key component of relative quantification of microRNA of interest regardless of the various methodological steps described above. However, there are no guidelines on normalising standards specific to circulating microRNA in the setting of HNC. Studies reviewed here predominantly used an endogenous control (example: U6 snRNA) for normalisation. However, other methods included 1) exogenous spike-in controls (example: synthetic cel-miR-39), 2) geometric mean of previously characterised microRNAs or 3) the mean of all quantified microRNA. Poel *et al* 2018 reported averaging the endogenous and exogenous controls to provide a better normalisation (For plasma and serum samples) based on their analysis of variability in cel-miR-39-3p and miR-16-5p $C_q$ values (Poel *et al*. 2018). There was large variability in statistical methods and relative quantification methods. However, detailed analysis of statistical methodology and relative quantification methods is beyond the scope of this chapter.

### 2.4.1 Summary

In summary, the circulating microRNA discovery studies reviewed here identified many potential biomarkers for detection and prognostication of head and neck cancer. However, there is considerable variability in all methodological aspects explored here. It ranged from patient preparation to relative quantification. There appears to be no standardisation observed in reporting or methodology. It is difficult to recommend one technique over another due to the lack of direct comparisons reported in the literature. It is important to discuss and standardise the incubation time, temperature, centrifugation protocol, RNA isolation protocol and relative quantification methods (at least in a head and neck cancer society level). If it is not possible to reach a consensus, then adherence to detailed reporting of methodology for each step would allow for reproducibility of experiments and appropriate meta-analysis in the future.

# 3 Circulating small extracellular vesicle microRNA as a predictor of oropharyngeal squamous cell carcinoma – exploring methods of predictive modelling using microRNAs

## 3.1 Introduction

Squamous cell carcinoma is the most abundant cancer affecting the oropharynx with an increasing incidence (Khan and Tomar 2020). HPV associated oropharyngeal squamous cell carcinomas (OPSCC) has dramatically increased compared to HPV negative OPSCC (Economopoulou *et al*. 2020). However, HPV associated OPSCC are reported to have a favourable prognosis compared to HPV negative cancers of the head and neck (Grønhøj *et al*. 2019). Although, this survival advantage is not as clear in the group of patients who smoked (House *et al*. 2018). In addition, the treatment success has an inverse relationship to the tumour stage at diagnosis and the treatment related morbidity is directly proportional to the tumour stage (Chow 2020). Therefore, methods of earlier diagnosis of OPSCC warrants further investigation.

In current clinical practice, other than categorising cancers based on HPV status none of the other biomarkers are routinely utilised in the detection of OPSCC (Basheeth and Patil 2019). Furthermore, in the primary care setting the examination of the oropharynx, especially the tongue base region is limited. Therefore, patients may not be diagnosed at an early-stage despite being symptomatic. The common symptoms of sore throat due to chronic or recurrent upper respiratory tract infections can often overlap with OPSCC with the risk of missing early diagnosis (Bannister and Ah-See 2014). Therefore, minimally invasive biomarkers have been investigated to diagnose these patients early as possible (Lange *et al*. 2010).

Serum based microRNAs are reported in various disease states as a minimally invasive molecular marker of disease. The unique ability of microRNAs to regulate

multiple target genes makes them a prime biomarker candidate. Expression levels of microRNAs contained in small extracellular vesicles are thought to be more disease specific compared to free serum microRNAs (Peacock *et al*. 2018). The ability to profile a patient with a unique set of molecular markers paves the way for precision medicine where cancer treatment can be truly individualised. Recently published results from this thesis indicates the ability to clearly identify patients with OPSCC from a pool of patients with cancer, inflammation, and benign disease (Mayne *et al*. 2020). As mentioned in Mayne *et al*, appropriate statistical modelling is a key component of successfully delivering a biomarker from bench to clinic. Therefore, some of the available methods of advanced predictive modelling for biological data are described below.

### 3.1.1 Introduction to methods of statistical modelling

Robust statistical modelling is a key component of identifying appropriate biomarkers for predicting a disease. While basic descriptive statistics forms the foundation of complex classification methods, there has been no clear guidance or standardisation of data-analytical methodology (Shariat *et al*. 2010). The ability to accurately classify experimental datasets depends on a multitude of factors. One of the primary considerations is the sample size (Bujang and Adnan 2016). The prevalence of HNSCC is relatively low compared to other common malignancies (Cohen *et al*. 2018b). Therefore, it is important to consider strategic patient recruitment protocols while maintaining ethical standards for generating appropriate sample sizes. As mentioned in the previous chapter, standardisation of all steps involved, from patient preparation to data analysis can affect the final outcome of a biomarker study (Dharmawardana *et al*. 2019). However, a negative effect of strict inclusion and exclusion criteria is the limited recruitment opportunity,

leading to a smaller sample size. Options to improve the sample size includes, a) the recruitment of patients over a long period of time while maintaining a strict standard and storage of samples/data in a biobank for future analysis. Here, the expected heterogeneity associated with time dependent changes in clinical protocols, patient factors and environmental factors would be mitigated by the potential to generate a significantly large sample size; b) the collection at a high-throughput centre; or c) the use of appropriate statistical methods that would be less influenced by the sample size. Practically a combination of such methods can be used.

Another important element of data analysis in biomarker studies is the predictive model generation (Di Ciaccio *et al*. 2012). In this chapter multiple classification methods were explored to analyse the same dataset. These mathematical models are described in detail below. Overfitting of mathematical models also needs to be considered as a priority. This is where parameters for a given model are extremely tweaked to classify the training sample with 100% accuracy and when this model is tested on an independent cohort there is a significant reduction in model accuracy. Indicating that the model is overfitted to the features of the training sample characteristics alone. Therefore, the total sample size needs to be large enough to randomly split them into a training and a testing cohort with similar characteristics (i.e. age matched). Also, the recruitment of participants needs to be homogeneous as practically possible. In an ideal scenario, the tested model should be re-tested in a large prospectively collected validation cohort. If only a small sample is available due to clinical or logistical reasons, statistical cross-validation procedures can be conducted to improve the accuracy of the model without overfitting parameters (Bull 1993). However, statistical manipulation of data (such as nested cross validation) can inherently introduce bias that may

prove significant only when these models are tested in an independent cohort. Therefore, every attempt should be made to recruit a larger sample before reverting to cross-validation procedures.

Binary logistic regression is one of the most commonly utilised regression methods to predict data points to a binary outcome (e.g. cancer versus control). Mathematically, logistic regression falls within generalised linear models and uses one or more numerical independent variables to describe the relationship to the dependent binary variable (Equation 1) (Bonney 1987). The independent variables here can be nominal, ordinal or ratios. The regression equation calculates the probability of each sample, belonging to one of the two dependent variables. This method of data classification assumes that the independent variables do not correlate well with each other and there are no outliers. Pearson chi-square test or Hosmer-Lemeshow test statistic are generally utilised to test the overall goodness-of-fit on the two-by-two classification tables. This goodness-of-fit can be altered based on estimated model parameters (β). These parameters are estimated based on maximum likelihood of the generalised linear model.

$$\boldsymbol{\pi_i = \Pr(Y_i = 1 | X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}}$$

Y = binary response trait
$Y_i$ = 1 (if the trait is present)
$Y_i$ = 0 (if the trait is not present)
X = independent variable ($X_1$, $X_2$, $X_3$… for multiple variables)
Pr = probability estimate
β0 = model parameter (odds of characteristic present at baseline)
β1 = model parameter (odds of a characteristic is present is multiplied by exp(β1) for each increase in independent variable.

***Equation 1. Binary logistic regression equation***

Principal component analysis (PCA) is an unsupervised, explorative, multivariate, data classification method. This algorithm is often used when a large number of independent variables (dimensions) are present for describing the dependent variable (Gray 2017). PCA amongst other regression methods have become popular in health sciences. This is mainly due to relatively small sample sizes with many independent variables. This analysis reduces the number of dimensions (the number of independent variables) by mathematically transforming a group of independent variables into a single abstract variable while maintaining the important information that describe the dataset. This method is especially important for machine learning algorithms that classify datasets, because a small number of independent variables can more accurately describe a dataset without overfitting the generated model (Husson *et al*. 2011). It also reduces computational time required for such classification algorithms as the number of iterations required to fit the number of variables are reduced.

Overall, the fundamental function of a PCA is to reduce the number of independent variables (dimensions) describing a dataset while preserving the significant information that contribute to the variance. A key feature of PCA is that the classification is based purely on the variance present in individual independent variables and does not require a dependent variable. Once the dimension reduction is conducted, the data points can be labelled (based on any available categorical information) to understand the meaning of the classification (see Figure 20). However, this feature can also be a limitation, because the PCA selects for components that describe the most variance in a given dataset, but these components may not have any biologically meaningful information relevant to the research question. Therefore, resulting classifications based on a PCA needs to be carefully interpreted.

PCA also requires the data to be standardised and a subsequent covariance matrix calculated to identify variables with high correlation to each other. Variable standardisation is conducted by subtracting the column mean from the data point and then divided by its standard deviation (Gray 2017). This allows all independent variables to be represented in the same scale. The reason for identifying highly correlated variables is to remove ones with the same information (redundant) describing a given dataset (Taguchi 2016).

Eigenvalues are matrix equations calculated from the covariance matrix to determine the principal components of the dataset (Xanthopoulos *et al*. 2013). A principal component is a new variable constructed from linear integration of a mixture of the original variables (Xanthopoulos *et al*. 2013). The first few components contain majority of the information that described the original dataset. Maximum 'amount' of the original information is contained in the first component, then the remainder in the second and so on until a selected number of components are generated to describe 100% of the variance seen in a given dataset displayed as a scree plot (Figure 18). Mathematically, the largest variance of a dataset is accounted for by the first principal component and repeatedly added to by new components until the total variance is zero. A limitation of such dimension reduction is that principal components themselves have no 'real' meaningful measure. However, the original variables used to calculate such a principal component can be identified along with matrix equation parameters to reproduce a principal component vector (Sell *et al*. 2020). Therefore, enabling the extraction of biologically relevant information.

Sparse partial least squares discriminant analysis (SPLSDA) is a supervised classification method based on partial least squares regression (Lee *et al*. 2018). Compared to PCA described earlier, the SPLSDA method incorporates the categorical information about the dependent variable (supervised) into the selected components (Lee *et al*. 2018). Hence, a more efficient classification of the data. However, this method requires a categorical dependent variable (e.g., diagnosis: cancer versus control) and therefore sensitive to overfitting of the model to the training dataset (Le Cao *et al*. 2011). This potential to overfit can be mitigated by cross-validation methods with inherent accuracy calculations and variable selection. K-fold or leave-one-out cross validation are validated methods described in current literature (Rodriguez-Perez *et al*. 2018). These cross-validation methods continually test the initial model with numerous iterations omitting some of the training sample data. In the case of leave-one-out cross validation, a single sample is left out while the model is built and tested on the remaining samples. Multiple iterations (folds) of this can be conducted until the accuracy is improved and the balanced error rate (BER) is stable (Chung and Keles 2010). In principle, component (known as the latent variable) creation in SPLSDA is very similar to PCA with similar prediction accuracies. In contrast to PCA, SPLSDA is able to achieve the same accuracy with a lower number of variables, because it preferentially selects for minor components based on the independent variables that best correlated with the dependent variable (Sutton *et al*. 2018).

Support vector machines (SVM) are supervised machine learning algorithms with the ability to classify or regress large datasets (Steeb 2008). This algorithm attempts to separate datapoints based on a hyperplane (Equation 2). This hyperplane can be described as a straight line in two-dimensions and classification

of data is based on which side of the hyperplane a particular data point falls. The advantages of SVM include a) computational efficiency, b) works well with relatively low sample sizes and c) works extremely well when there is a clear separation in data points (Yang 2004). The tuning of the SVM attempts to calculate the distance from the hyperplane to the closest data points. The best model is then chosen based on the hyperplane that has the furthest distance from the closest data point in each classification group. However, SVM functions in multi-dimensions, such that a hyperplane can be difficult to visualise. In addition, biological data have rare linear separations, therefore, SVM uses a 'soft margin' or 'kernel tricks' to separate data in a non-linear fashion (Jiang *et al*. 2013).

In the soft margin method, the same concept as a linear separation is utilised, however, with less stringent criteria for misclassification. This 'tolerance' for misclassification therefore becomes an SVM parameter (known as a 'C' hyper-parameter) (Statnikov *et al*. 2008). Where a larger C value penalises the model more for misclassification and vice-versa. In contrast, the kernel trick method uses non-linear boundaries functions such as polynomial, sigmoid, or radial bias function (Steeb 2008). When non-linear boundary functions are utilised, further model parameters need to be implemented. For an example gamma is a parameter implemented in the radial bias function where higher the gamma value, the more sensitive the model is to the training dataset.

The most important concern with SVM modelling is its poor handling of noisy data, where overlapping dependent variables generates unusable models (Statnikov *et al*. 2008). In addition, SVMs does not natively estimate predictive probabilities for each sample, while these can be calculated with cross-validation methods, it can be computationally intensive and time consuming (Statnikov *et al*. 2008).

$$\boldsymbol{\beta_0} + (\boldsymbol{\beta_1} \times \boldsymbol{x_1}) + (\boldsymbol{\beta_2} \times \boldsymbol{x_2}) + \cdots + (\boldsymbol{\beta_n} \times \boldsymbol{x_n}) = \mathbf{0}$$

***Equation 2. Description of any hyperplane for support vector machines***

Random forest (RF) classifier is a decision tree-based machine learning algorithm. It uses a large number of decision trees (the 'forest' component) in order to decide on whether to classify a data point to a known category (dependent variable) (Statnikov *et al*. 2008). Because a single decision tree is often poor at producing an effective model for a given set of data. These decision trees are made at 'random', where each decision tree uses a random number of independent variables to make an output decision and ultimately combine the output of individual trees to compute the overall decision tree (the model) (Chen and Ishwaran 2012). The performance of the model relies on the ability to create multiple decision trees that do not correlate with each other. Hence providing a higher level of overall information to describe the dataset and provide an accurate classification. The features (individual microRNA levels in this case) used to create these decision trees need to be poorly correlating with each other in order to produce an accurate decision tree (Dias and Torkamani 2019). However, a key limitation of this method is its sensitivity to the training dataset. Therefore, a misrepresented training dataset creates a model that is unusable. To mitigate this a bootstrapping procedure is often recommended and this is termed bagging in data science specifically for RF classification (Fabris *et al*. 2018). RF dependent methods are also used for imputing missing values (Tang and Ishwaran 2017) as well as calculating cost effectiveness models using clinical data (Yang *et al*. 2009).

## 3.2 Methods

### 3.2.1 Patients, blood collection and storage

#### 3.2.1.1 Patients

Blood specimens and related patient demographic information were accessed from three sources a) the SA ENT Tissue Bank (Flinders Medical Centre, Adelaide, Australia), b) PROBE-NET (Flinders Medical Centre, Adelaide, Australia) and c) Victorian Cancer Biobank (Melbourne, Australia). Inclusion criteria for specimen selection was the diagnosis of OPSCC (with or without p16 positivity). The specimens collected from the Flinders Medical Centre and the Royal Adelaide Hospital were stored in the SA ENT Tissue Bank, and the Victorian Cancer Biobank collected samples from all tertiary hospitals in greater Victoria. The diagnosis of these patients was confirmed and documented at multi-disciplinary meetings at each of the representative institutes. Patients with any grade of dysplasia were excluded from further analysis. Specimens in the control group consisted of patients with upper aerodigestive symptoms with no clinical, endoscopic, radiological, or histological evidence of malignancy or atypia.

The control group was recruited from multiple patient populations a) patients presenting to gastroenterology services with dyspepsia, reflux symptoms or follow up from prior gastroduodenal ulceration and b) patients presenting to otorhinolaryngology-head and neck services with oropharyngeal symptoms (pain, globus and dysphonia).

#### 3.2.1.2 Blood collection and storage

All serum specimens were collected prior to commencing any cancer treatments. Samples were collected in multiple settings a) at the time of clinical consultation or b) prior to diagnostic endoscopy/biopsy procedure (prior to administration of intravenous, oral, or inhaled anaesthetic agents). Venous blood was collected into an 8 millilitre Z serum separator clot activator tubes (Vacuette, cat#455078). They were incubated in room temperature for 16 - 25 hours before further processing. Serum was separated via centrifugation of whole blood samples at 650 g for 15 minutes and stored in 1 millilitre aliquots at -80 degrees Celsius for later use.

### 3.2.2 MicroRNA extraction and profiling

#### 3.2.2.1 Small extracellular vesicle isolation

To exclude larger microparticles, 1mL aliquots of serum samples were retrieved, thawed, and centrifuged at 16,000g for 30 minutes at 4 C. 250 microlitre supernatant from each of the samples were then processed with an ExoQuick kit (System Biosciences, CA, United States) according to manufacturer instructions. In brief, the samples were incubated with ExoQuick at 4 C for 16 hours. The isolated pellet was resuspended with 50 microlitre of phosphate buffered saline (PBS). This method was previous reported to isolate small extracellular vesicles in the sizes ranging from 30 – 150 nm (Chiam *et al*. 2020).

#### 3.2.2.2 MicroRNA extraction

The microRNA from small extracellular vesicles were extracted using a Commercial miRNeasy serum/Plasma kit (QIAGEN) according to manufacturer instructions. A spike in procedure was conducted by adding 5 microlitre (0.1pmol)

of each of the synthetic RNA molecules ath-miR-159a and cel-miR-54 (Shanghai, Genepharma Co Ltd) to the 50 microlitre of QIAzol vesicle lysate. The final RNA was eluted into 24 microlitre of RNase-free ultrapure water.

#### 3.2.2.3 TaqMan OpenArray microRNA profiling

MicroRNA profiling utilised custom-made high throughput QuantStudio 12K Flex OpenArrays. These arrays comprised of 112 microRNA probes selected due to previously reported high relative abundance. 3.35 microlitre of RNA was reverse transcribed for each sample using a matching custom OpenArray microRNA RT pool (Life Technologies) and TaqMan microRNA reverse transcription kit (Life Technologies). Pre-amplifications were carried out on 7.5 microlitre of cDNA (per sample for each pool) with a matching custom OpenArray PreAmp pool (Life Technologies) and TaqMan PreAmp master mix (Life Technologies). As per manufacturer recommendations, the pre-amplified products were diluted 1:40 ratio with 156 microlitre of RNase free ultra-pure water and then mixed with TaqMan OpenArray RT PCR master mix (Life Technologies). These were subsequently loaded in to a 384 well TaqMan OpenArray loading plate. PCR was conducted using a QuantStudio 12K Flex RT PCR system. OpenArrays were run with standardised methodology, in years 2018 and 2020. The individual data points were labelled appropriately to distinguish samples run in 2018 from samples run in 2020. This allowed for identification of run-to-run variability between OpenArray plates that were run in 2018 and the OpenArray plates that were run in 2020.

### 3.2.3 MicroRNA data analysis methods

Data analysis for this chapter was conducted using Microsoft Excel (Version 2006, 64bit), RStudio (version 1.3.959, 64bit) and R (version 4.0, 64bit). The qpcR

package (version 1.4) was used to calculate cycle threshold (Ct) value for each PCR assay. microRNAs with Cts in at least 50% of the samples were considered for further analysis. The relative expression of each microRNA was calculated based on $2^{(40-Ct)}$.

Housekeeping genes for data normalisation were selected based on a) they were expressed in all samples, b) they were not statistically different between groups, c) they were not highly variable (no outliers), and d) they correlated well ($r>0.7$) with each of the other housekeeping genes. The relative expression levels of microRNA were normalised to the geometric mean of these housekeeping genes (Table). The samples that were not amplified or had concerns of quality assurance were excluded from further analysis.

#### 3.2.3.1 Data preparation for classification

MicroRNAs and patients were deidentified and indexed. Zero value analysis was conducted and microRNA variables with more than 5% empty/zero data were excluded from classification. Algorithms that required no missing values had imputed. Multiple missing data imputation methods were explored. The normalised microRNA dataset was variably distributed based on the microRNA of interest. Therefore, non-parametric and distribution-free imputation methods were considered. Missing data were imputed based on non-linear iterative partial least squares (NIPALS) algorithm (Wold 1966). However, for SPLSDA, the analyses were conducted with and without data imputation to assess its sensitivity. The data was then centred and scaled prior to randomly selecting the training (70%) and testing (30%) cohorts. Unscaled data was used for random forest modelling.

### 3.2.3.2 Classification methods

One unsupervised classification method, two supervised classification methods and two machine learning algorithms were utilised to assess the robustness of the dataset and appropriateness of the algorithm. In each case, except for logistic regression, a training dataset was used to develop the model and tune the model parameters. Subsequently, the optimum model in each case was tested on the independent testing cohort.

#### *3.2.3.2.1 Binary logistic regression*

A binary logistic regression model that had a nested cross-validation with lasso regression developed by Mayne *et al* using small extracellular vesicle microRNAs was recently published using the data from year 2018 (Mayne *et al*. 2020). Data specific methodology is not further discussed in detail within this thesis due to a pending patent related to this particular model.

#### *3.2.3.2.2 Principal component analysis*

The R (version 4.0.1) package 'mixOmics' (version 6.12.1) with 'pca' function on RStudio (version 1.3959) software is used for this analysis. The principal components were selected based on the highest eigenvalues (Figure 18). The first two principal components of the analysis were used to classify the training dataset (Figure 20).

#### *3.2.3.2.3 Sparse partial least squares discriminant analysis*

The R (version 4.0.1) package 'mixOmics' (version 6.12.1) with 'plsda', 'splsda', 'tune.splsda' functions on RStudio (version 1.3959) software was used for this analysis. Here the dataset with and without imputed data was assessed for

sensitivity of the imputed data. The initial SPLSDA model is tuned using k-fold cross validation with 15-folds and 50-repeats of the analysis to identify the ideal number of components to include in the model and individual variables to use for the classification (Rodriguez-Perez *et al*. 2018). Performance of the model was analysed by comparing the classification error rate and the number of components included in the model. Subsequently, the optimal number of microRNA variables contributing to each component was tuned by selecting one microRNA at a time for each component, then calculating the balanced error rate. Following the tuning process, optimum number of components and number of microRNA variables in each component (Figure 26) were used to develop the final SPLSDA model prior to testing the model in an independent testing cohort.

The 2018 dataset contained control samples from two distinct cohorts of patients a) patient group with a prior diagnosis of gastro-oesophageal reflux disease without dysplasia or malignancy and b) patients who presented to gastroenterology services without any clinical findings of malignancy or other inflammatory disease. Another SPLSDA model was generated to reflect above dependant variable information (i.e., three dependent categories: cancer, benign with GORD and benign without GORD) to assess its classification accuracy. The p-values indicated with area under the curve (AUC) refers to statistical significance compared to AUC of 0.5 after a z-test.

#### *3.2.3.2.4 Support vector machines*

The R (version 4.0.1) package ‘e1071’ (version 1.7-3) with ‘svm’ function on RStudio (version 1.3959) software was used for this analysis.

#### *3.2.3.2.5 Random forest*

The R (version 4.0.1) package ‘randomForest’ (version 4.6-14) with ‘randomForest’ function on RStudio (version 1.3959) software was used for this analysis.

### 3.2.3.3 Descriptive statistical analysis

Wilcoxon rank sum (non-parametric) univariate analysis was conducted to identify any significant differences between comparison groups for individual microRNA variables. Tables of these results including test statistic, initial and adjusted significance values are displayed in Appendix 1.

## 3.3 Results

Univariate analysis with Wilcoxon signed rank test of all microRNA variables (with less than 25% zero/missing values) indicated miR1, miR6, miR7, miR8, miR19, miR31, miR38, miR59, miR60, miR62 and miR68 to be significantly different (unadjusted p-value < 0.05) between cancer and control groups (Appendix 1). Spearman correlation between microRNA variables indicated clusters of microRNAs that correlate well with each other (Figure 16). The highest correlation coefficients were between miR44, miR46, miR49, miR50, miR51, mir60, miR63 and mir68.

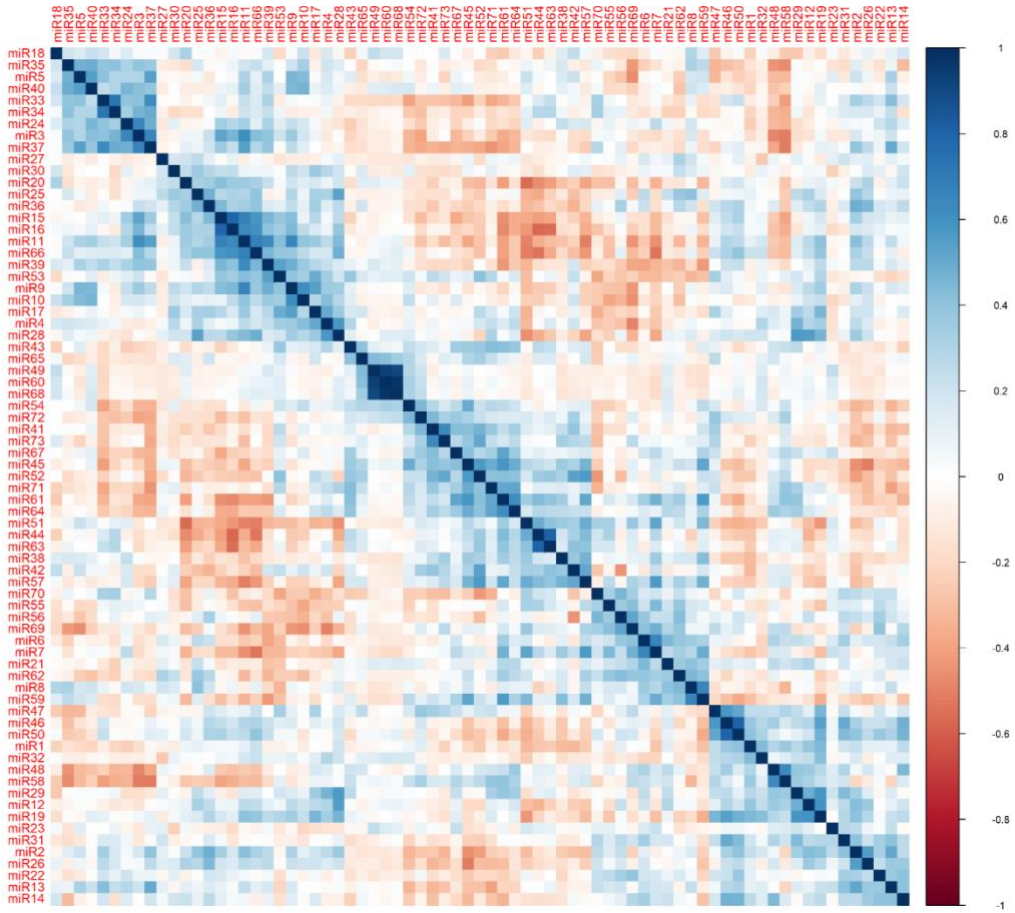



***Figure 16. Correlation matrix between individual microRNA normalised relative expression levels after data imputation. Darker blues indicate positive correlation values and darker reds indicated negative correlation values.***

Some of the classification methods required imputation of missing data. Therefore, data is imputed using the Random Forest algorithm. Figure 17 displays a sample of variables with the density function of the original data (blue) and the multiple imputations of the new data (red) indicate an appropriate fit.

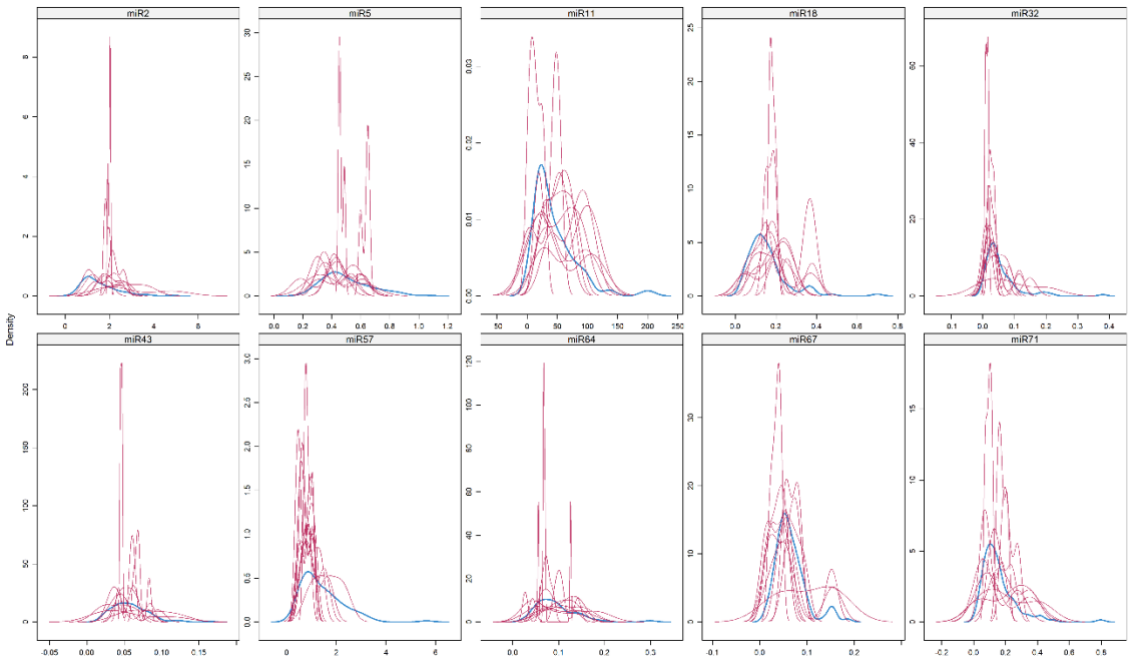



***Figure 17. Density function of imputed data using random forest method indicating close approximation to observed data distribution. X - axis indicates the range of values reported for individual microRNAs and the Y-axis indicates the density. The blue line indicates the original data distribution, each of the red lines indicates an iteration of imputed density function. The closest approximation to the original data function is used for missing data imputation.***

### 3.3.1 Principal component analysis

The first two components of the PCA (excluding variables with any missing or zero values) could only describe just over 35% of the data variance when data from the 2018 and 2020 OpenArrays were combined (Figure 18). This poor unsupervised separation indicated that algorithms such as SVM or RF would not be suitable for further classification of this dataset.

Only 14 microRNAs had no missing or zero values, hence only these were used for PCA analysis without data imputation. Individual microRNA levels that contributed to the first two components of the PCA were further investigated based on their interdependent correlations (Figure 19). As expected, widely spread

MicroRNAs displayed in Figure 19 indicates the lack of significant correlations between the variables used to construct the first two components of this PCA model. Figure 20 indicates the labelling of samples based on diagnosis (cancer or control) following the PCA classification, indicating a poor separation based purely on diagnosis.

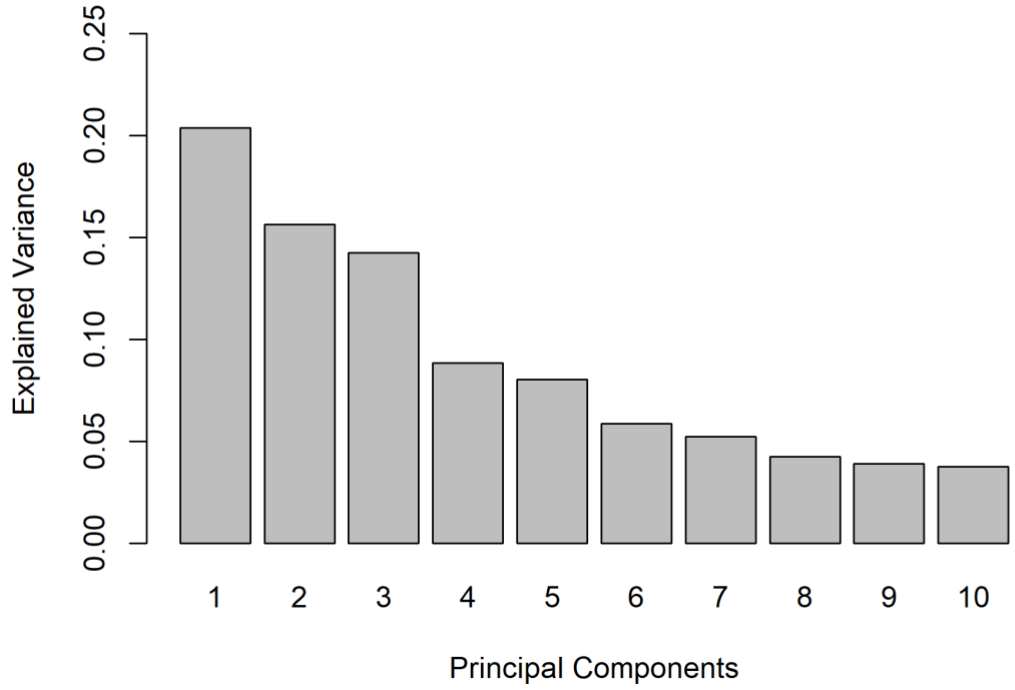



***Figure 18. Proportion of explained variance by principal components.***

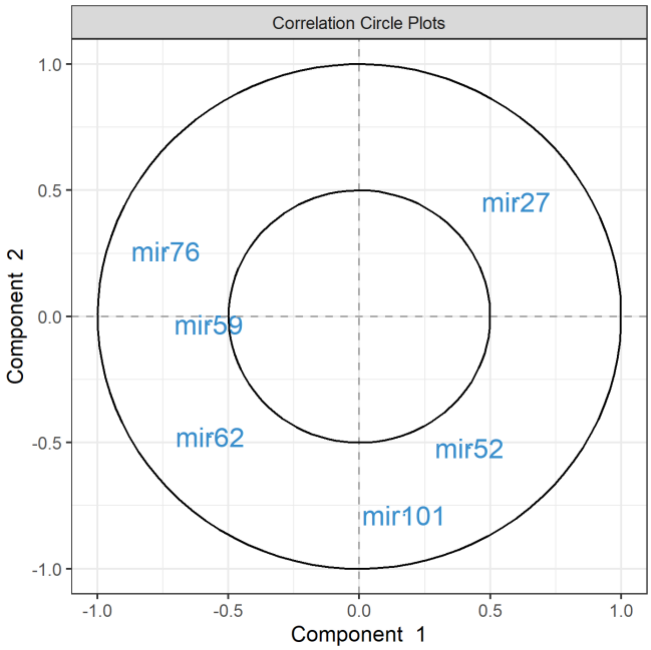



***Figure 19. Correlation between microRNAs included in first two principal components. MicroRNA variables with coefficients above ± 0.5 displayed.***

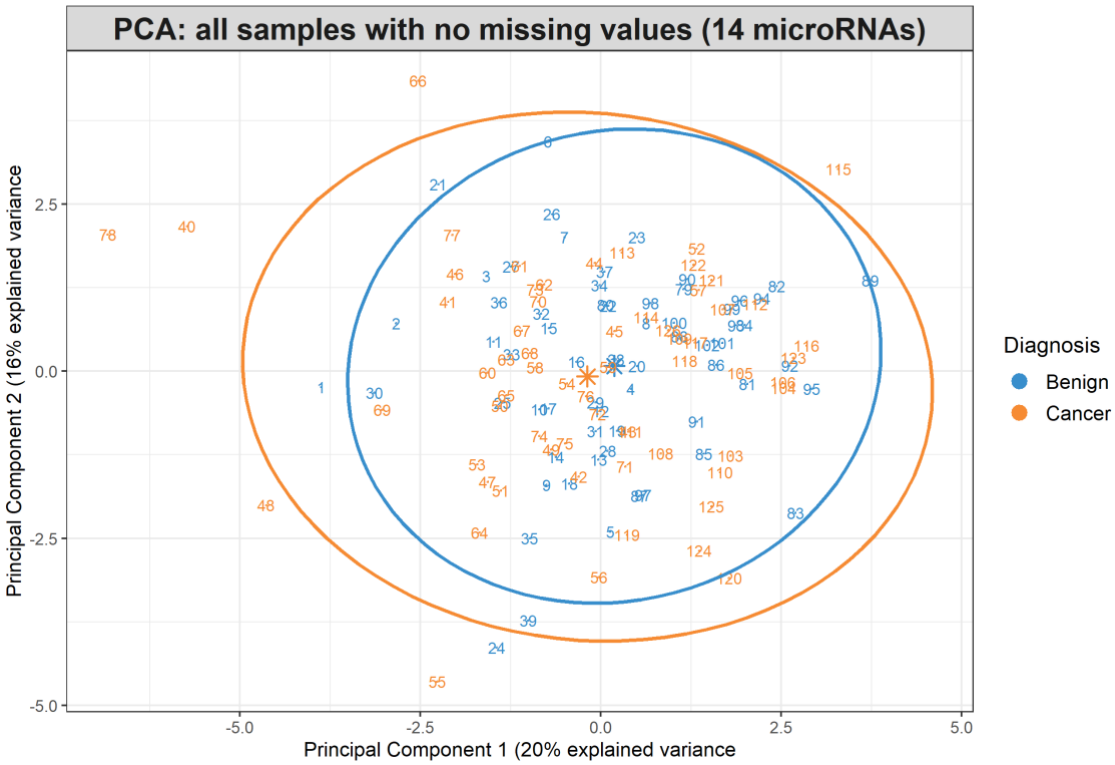



***Figure 20. Unsupervised classification of combined 2018 and 2020 data based on the first two components of PCA. Individual numbers on the plot represent sample identifiers. Cancer and control samples labelled as per the legend.***

In order to identify any OpenArray dependent features in the dataset, further labelling was conducted based on the OpenArray information. Where the 2018 data labelled as OpenArray 1 and 2020 data labelled as OpenArray 2. The samples contained in each of these OpenArrays were clustered into distinct groups with some overlap (Figure 21). Suggesting an inherent feature based on the OpenArray causing a separation, therefore, these patient cohorts were not combined for future analyses. There were no clear classifications based on PCA for gender, smoking status, primary tumour stage, neck node presence or p16 status (Appendix 1, Figures 1-9).

The reason for unsupervised separation of data between year 2018 and 2020 was identified as a result of R-code used for generation of Ct values. Analysis of qPCR fluorescence data from the OpenArray slides indicated that the 2020 data had aberrant baselines. Upon further inspection it appeared that some microRNAs were given inaccurate Cts by both the Quant Studio software and by the second derivative method using R software, relative to other samples of the same tissue type. Therefore, linear baselines from raw data were calculated by generating new R-code. New normalised relative microRNA levels were generated using the new R-code. Therefore, R-code used to generate 2018 dataset is inherently different to the 2020 dataset. Hence, unable to pool datasets for further analysis.

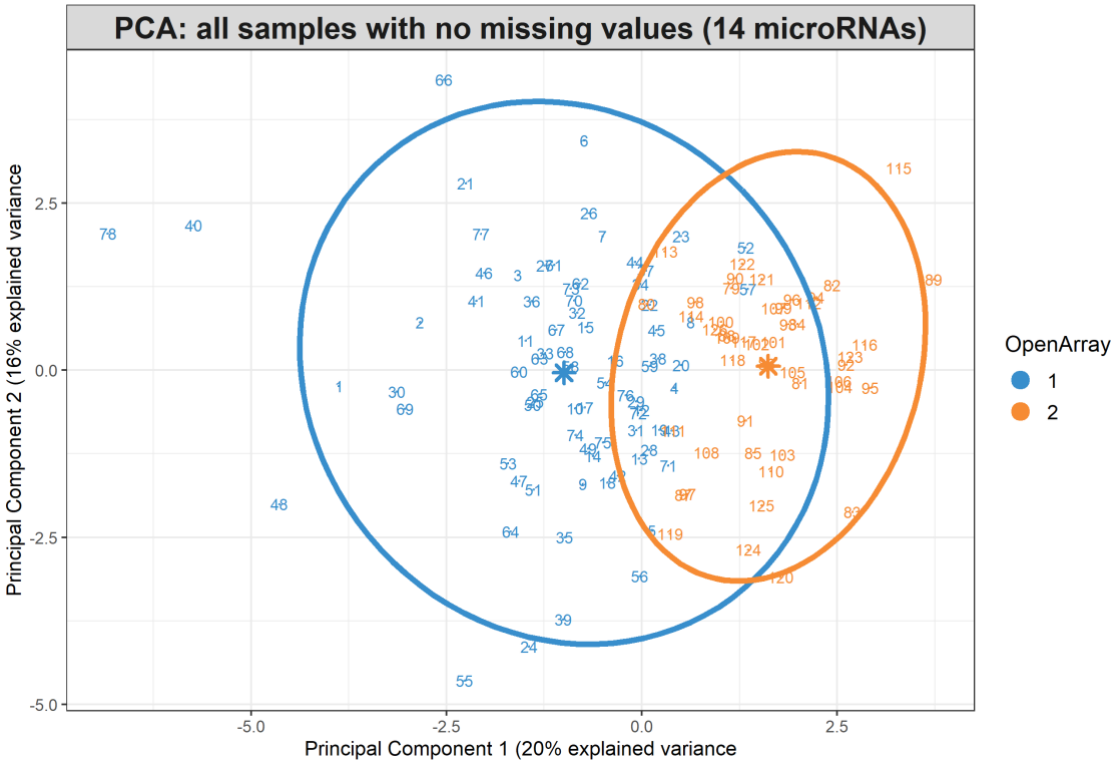
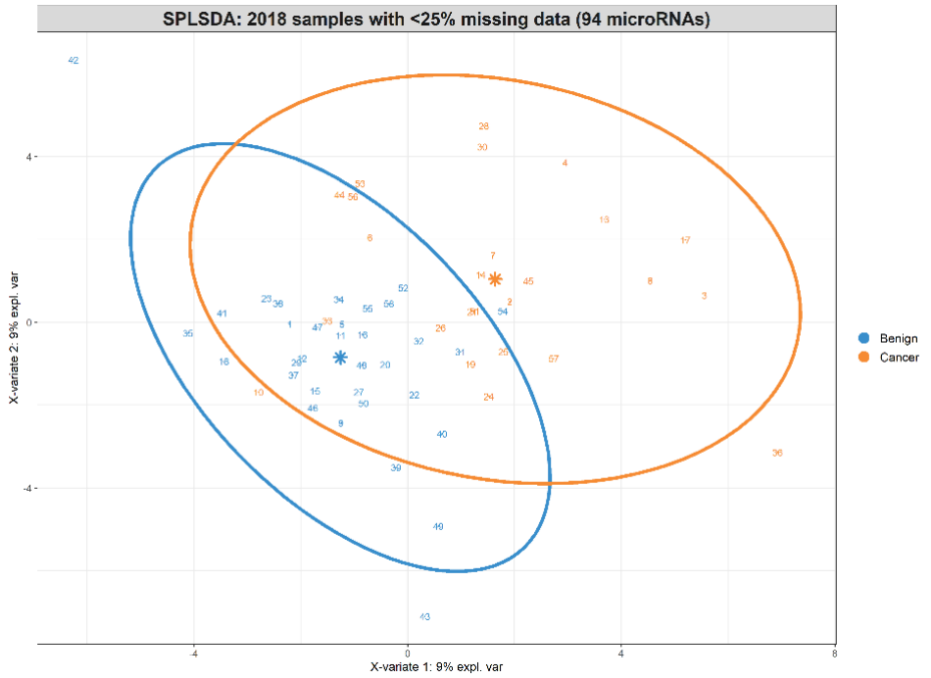



***Figure 21. Unsupervised classification of the whole dataset based on the first two components of PCA. Individual numbers on the plot represent sample identifiers. Samples were labelled based on the OpenArray each sample is run.***

### 3.3.2 Sparse partial least squares discriminant analysis

The exploratory analysis (PCA) above indicated that 2018 and 2020 datasets were inherently different due to a technical concern described earlier. Therefore, in SPLSDA of this data, separate analyses (for 2018 and 2020 datasets) were conducted for clarity and valid analysis. Both datasets were randomly split into training and testing datasets as described earlier. The classification plots show the separation of the training datasets based on diagnosis (Figure 22). However, each of the initial components only explained 8-9% of the variance based on primary diagnosis as the dependent variable.

A




B

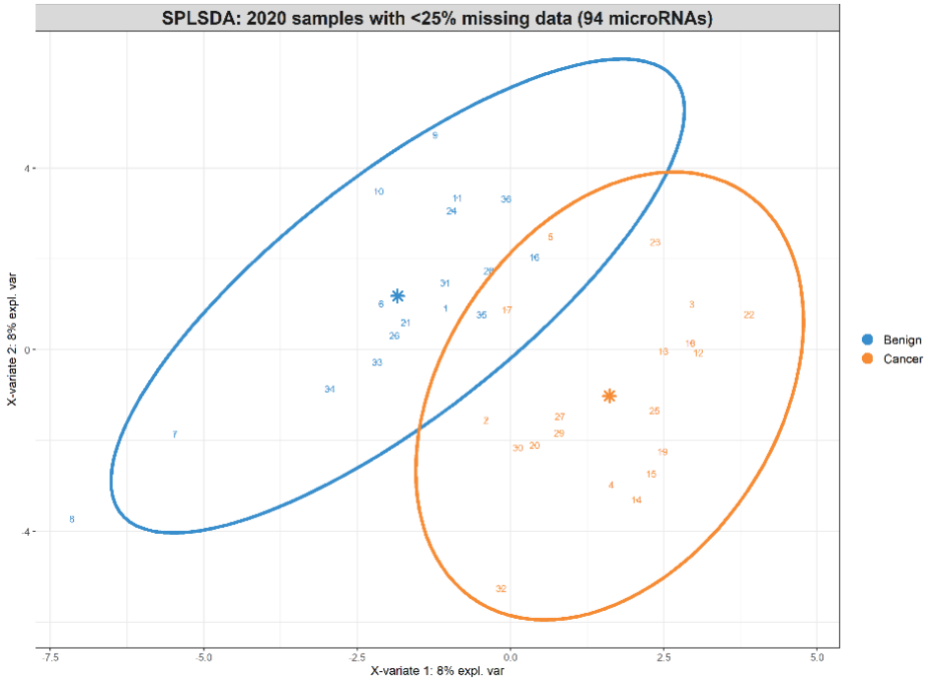



***Figure 22. Sparse partial least squares discriminant analysis of both 2018 (A, n=58) and 2020 (B, n=36) datasets classified based on primary diagnosis (Training datasets only). The centroid value of the dependent variable is displayed as a star and ellipses represent a 95% confidence interval from the respective centroid. Expl. Var = explained variance specific to the represented component.***

The correlation intensity maps did not display a clear pattern of microRNAs that were contributing to the classification of cancer and control (benign) patients.

Therefore, further tuning of microRNA variables contributing to each component was conducted with balanced error rates displayed below (Figure 26).

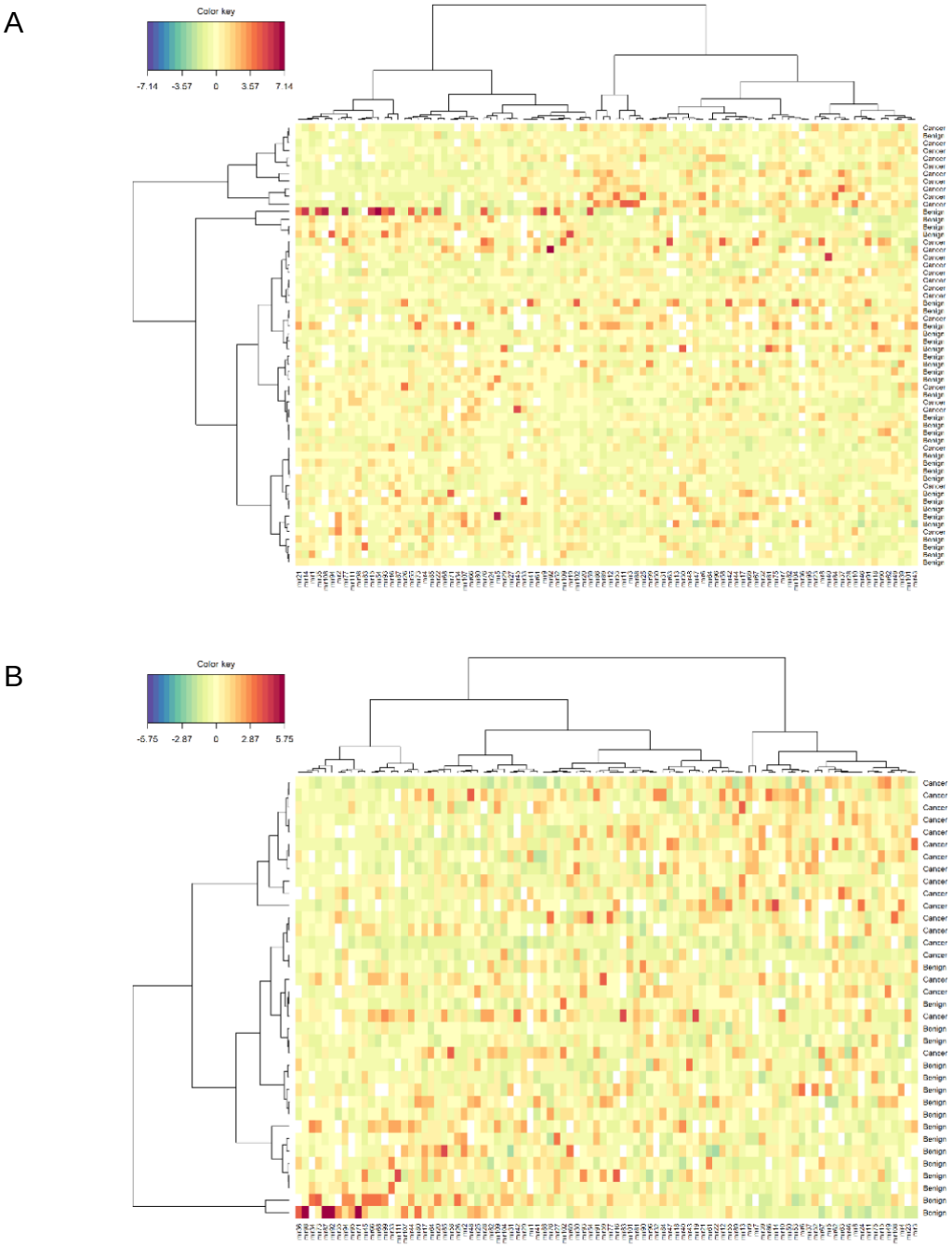



***Figure 23. Correlation intensity map describing the first component of the sparse partial least squares discriminant analysis for both 2018 (A, n=58) and 2020 (B, n=36) datasets (training set only) based on the primary diagnosis.***

These correlation plots (Figure 24) displays individually normalised microRNA levels that correlated (correlation coefficient above 0.6) with the first two components generated for each dataset indicating the greatest contribution to the respective components. However, there were no common microRNAs with a correlation coefficient higher than 0.6 between the two datasets (2018 and 2020). Further evidence indicating a fundamental difference in data between the two OpenArrays.

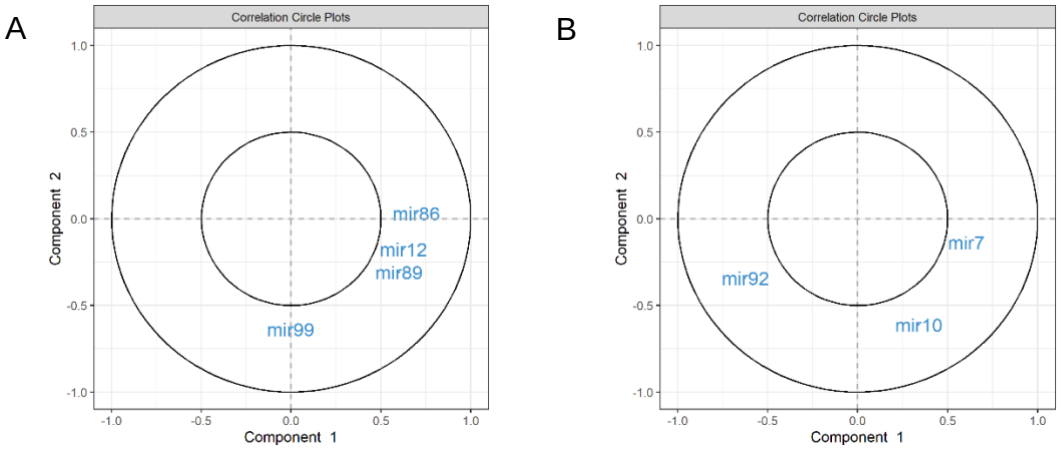



***Figure 24. Correlation plot of microRNA variables and first two components of 2018 (A) and 2020 (B) training datasets. microRNAs with a correlation coefficient higher than 0.6 were displayed here.***

In order to identify how many components should be included in the final predictive model, the classification error rate was calculated for the training dataset when each component data was sequentially added to the model (Figure 25). The 2018 dataset (Figure 25-A) exhibited an unexpected pattern of classification error. Where the incorporation of additional components to the model, increased the overall error. The expected pattern was observed in the 2020 dataset (Figure 25-B), where incorporation of additional components reduced the classification error rate until a plateau was reached. The lowest classification error rate for the 2018

and 2020 training datasets were reached with the inclusion of the first two components and the first three components respectively (Figure 25).

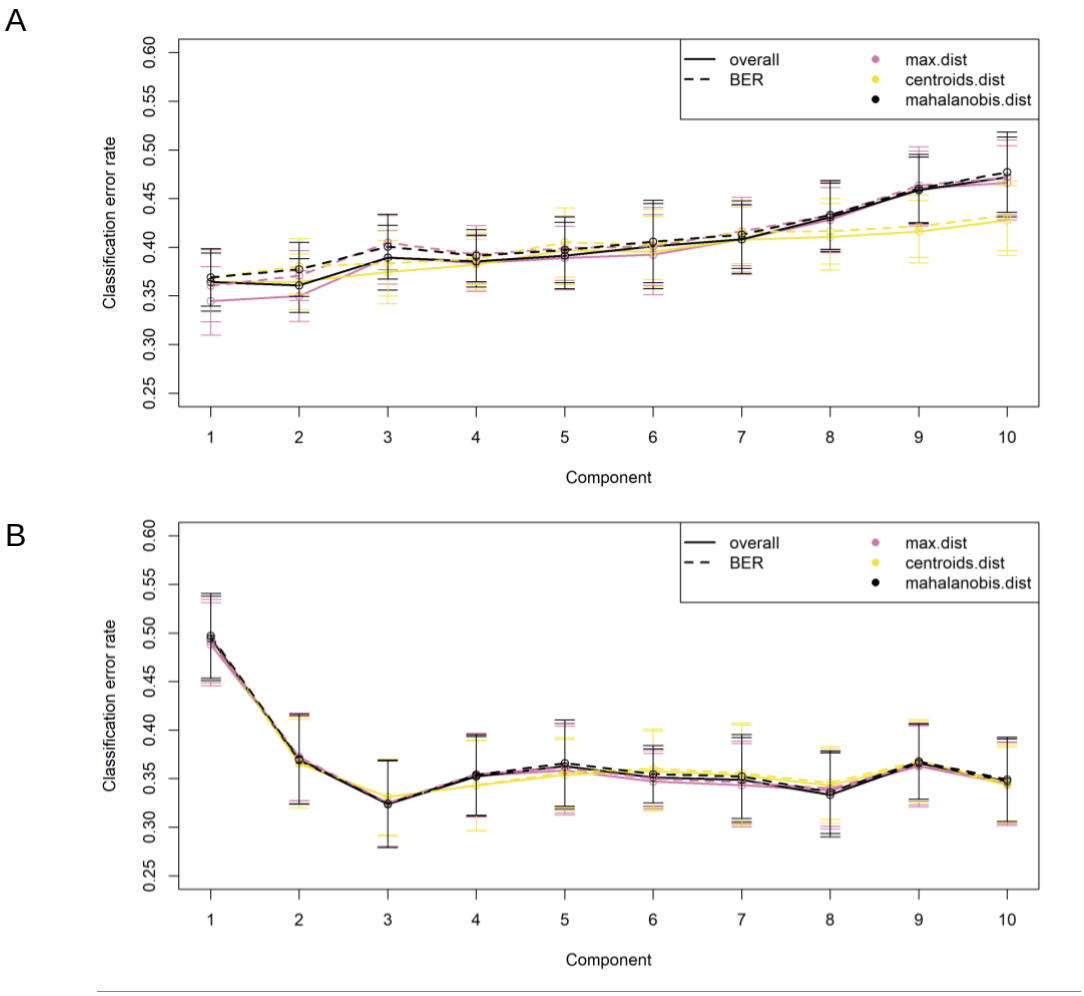



***Figure 25. Classification error rate of SPLSDA of the training datasets (A, 2018. B, 2020) relative to the number of components included in the analysis. Error bars display standard deviation of error rate. BER, balanced error rate. dist, distribution.***

Based on the results of the overall classification error rates, the models were tuned further to identify which normalised microRNA levels should be included in each of the components (This was limited to two components to save computing power). The lowest balanced error rate (BER) for the 2018 training dataset was achieved at 93 and 39 independent normalised microRNA levels for the first and second component respectively (Figure 26-A). The number of microRNA

variables that achieved the lowest BER for the 2020 dataset was 2, 90 and 1 for the first three components, respectively. However, the first component with only 2 variables generated a larger number of misclassifications during further testing. Therefore, for the final model using the 2020 data, only second and third components were used (two components) (Figure 27).

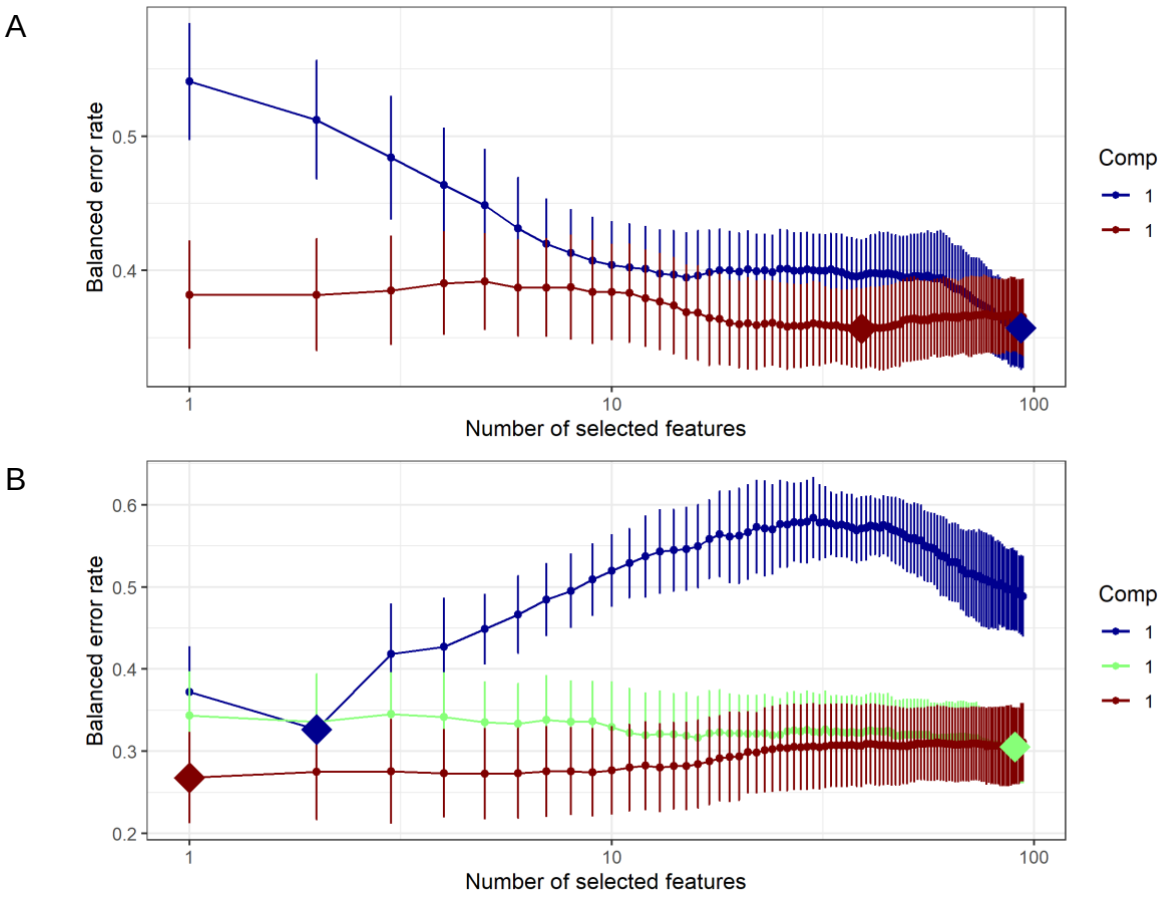



***Figure 26. Balanced error rate of first three components of SPLSDA with increasing number of feature (individual microRNA level) selection per component. A, 2018 training dataset. B, 2020 training dataset. Comp, component. Error bars, standard deviation from mean balanced error rate. Diamond shape indicates the number of features selected for each component with the lowest balanced error rate.***

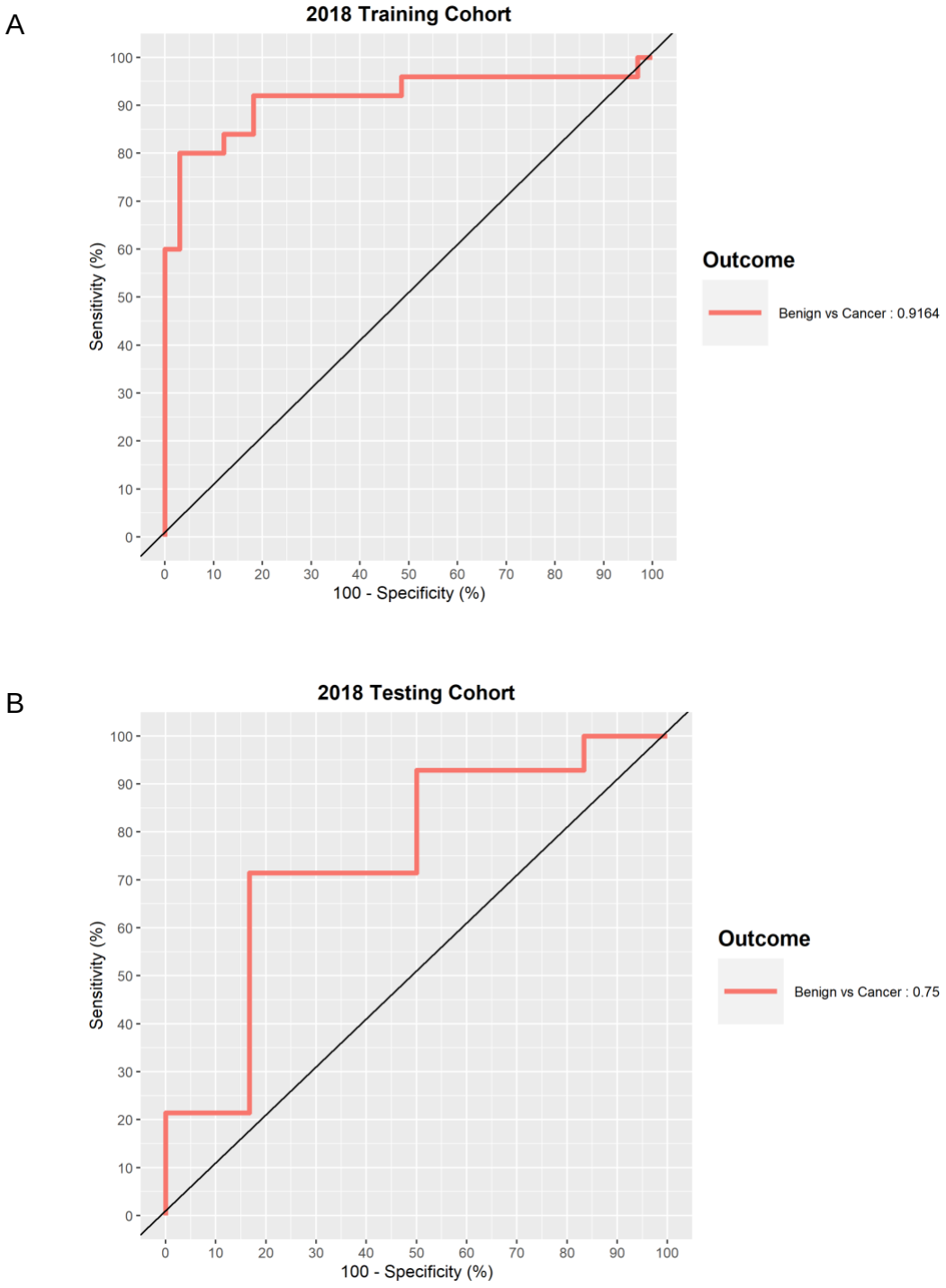



***Figure 27. Receiver operating characteristic curve for classification accuracy for the 2018 training (A – AUC = 0.9164) and testing (B – AUC = 0.7500) datasets. AUC, area under the curve.***

The Figure 27 and Table 25 displays the classification accuracy of the 2018 dataset. However, as described earlier the 2018 dataset contains two distinct control groups, therefore, further classification was conducted with incorporation of

this dependent variable information (Figure 28). This analysis indicated a clear separation of the GORD control group when compared with the cancer and rest of the control patients. The accuracy of this classification is further evident on ROC analysis on the independent testing cohort (Figure 29) as well as in the classification table (Table 26). Suggesting that the panel of microRNAs included in this particular model were likely to predict the presence of GORD when compared to benign samples instead of the presence of oropharyngeal cancer.

***Table 25. Classification table of 2018 testing dataset classification based on two SPLSDA components.***

| | | Clinical Diagnosis | |
|---|---|---|---|
| | | Cancer | Control |
| Predicted | Cancer | 6 | 0 |
| | Control | 8 | 6 |

Control, control patients with gastro-oesophageal reflux disease with no evidence of malignancy or patients who presented to gastroenterology clinical services with no evidence of malignancy or other inflammatory disease. Cancer, patients with oropharyngeal squamous cell carcinoma. Model sensitivity and specificity for detecting cancer is 42% and 100% respectively.

***Table 26. Classification table of SPLSDA of 2018 testing cohort classification based on three diagnosis.***

| | | Clinical Diagnosis | | |
|---|---|---|---|---|
| | | Cancer | Gastro-Benign | GORD-Benign |
| Predicted | Cancer | 10 | 3 | 0 |
| | Gastro-Benign | 0 | 0 | 0 |
| | GORD-Benign | 4 | 1 | 2 |

GORD-Benign, control patients with gastro-oesophageal reflux disease with no evidence of malignancy. Gastro-Benign, control patients who presented to gastroenterology clinical services with no evidence of malignancy or other inflammatory disease. Cancer, patients with oropharyngeal squamous cell carcinoma. Model sensitivity for detecting: Cancer = 71%, GORD = 100%

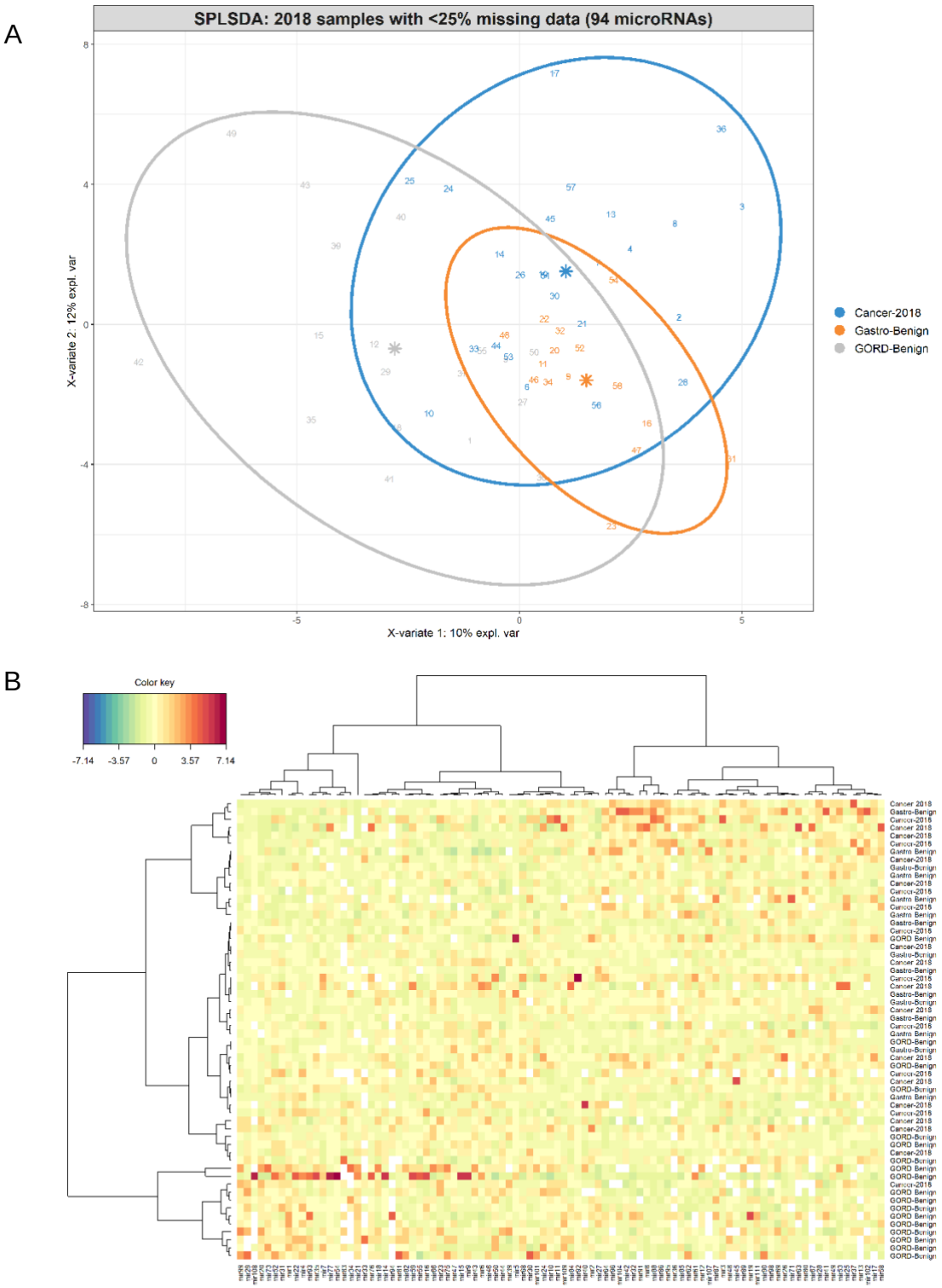



***Figure 28. SPLSDA classification based on three groups (Cancer, Gastro-benign and GORD-benign). A, classification plot using the first two components of SPLSDA with ellipses representing 95% confidence interval from the centroid. B, correlation intensity map with normalised microRNA levels on x-axis and the three categories on the y-axis.***

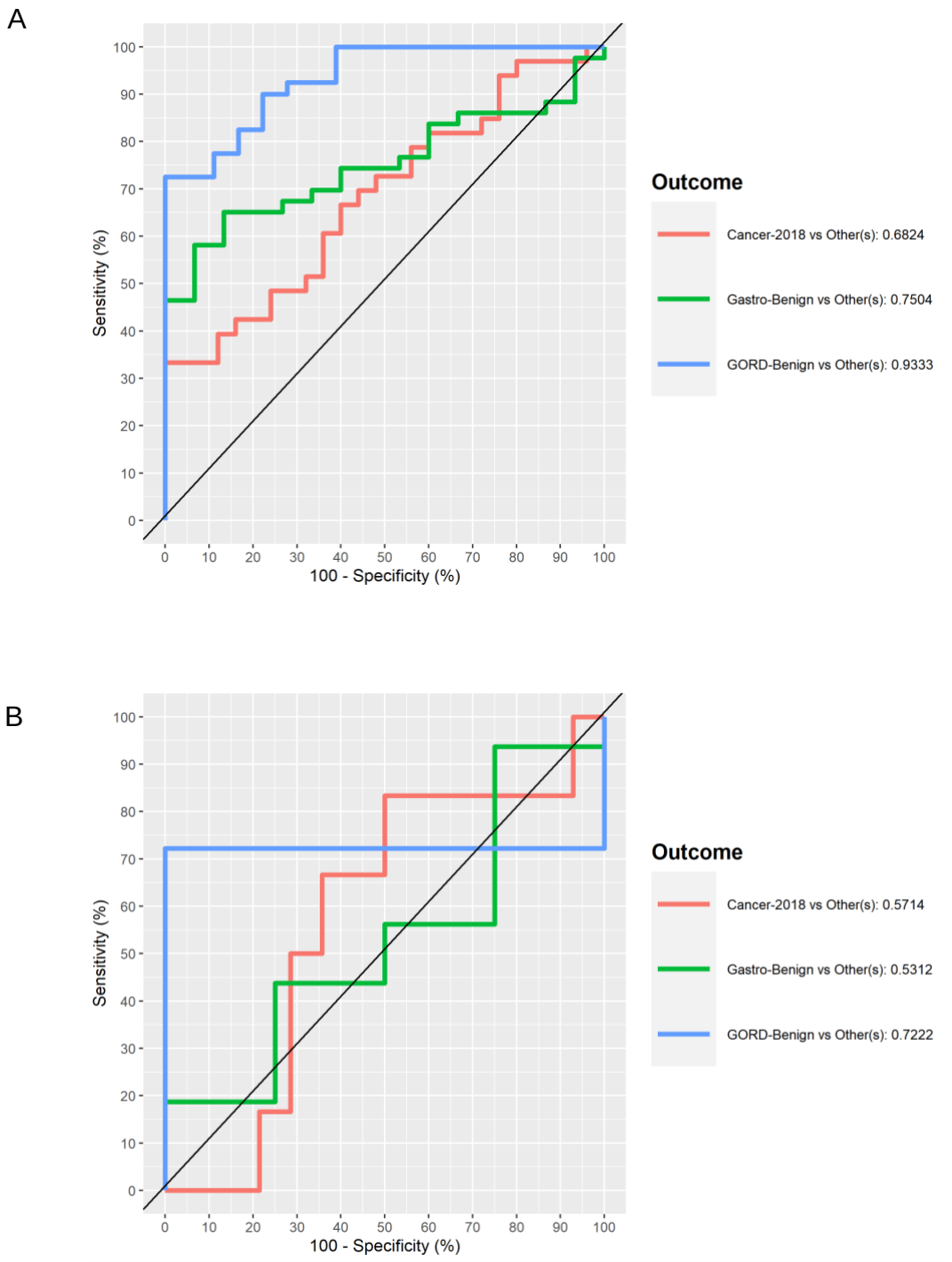



***Figure 29. Receiver operating characteristic curve for classification accuracy based on three groups (Cancer, Gastro-benign and GORD-benign) for the training (A) and testing (B) datasets. Using the first component of SPLSDA with contributions from 94 normalised microRNA levels.***

The SPLSDA model based on the 2020 datas*et al*so performed well with an AUC of 0.9783 ($p < 0.0001$) and 0.8 ($p = 0.08816$) for the training and testing cohorts

respectively (Figure 30). Figure 31 displays the correlation matrix of microRNAs that contributed to the SPLSDA model based on the 2020 dataset. Maximum distance-based classification of the testing cohort (Table 27) indicated a sensitivity of 60% and specificity of 100% with an overall balanced error rate of 0.2. Finally, the model generated using the complete 2020 dataset (as the training cohort) was used to test if the 2018 dataset (independent testing cohort) can be classified to cancer and control groups.

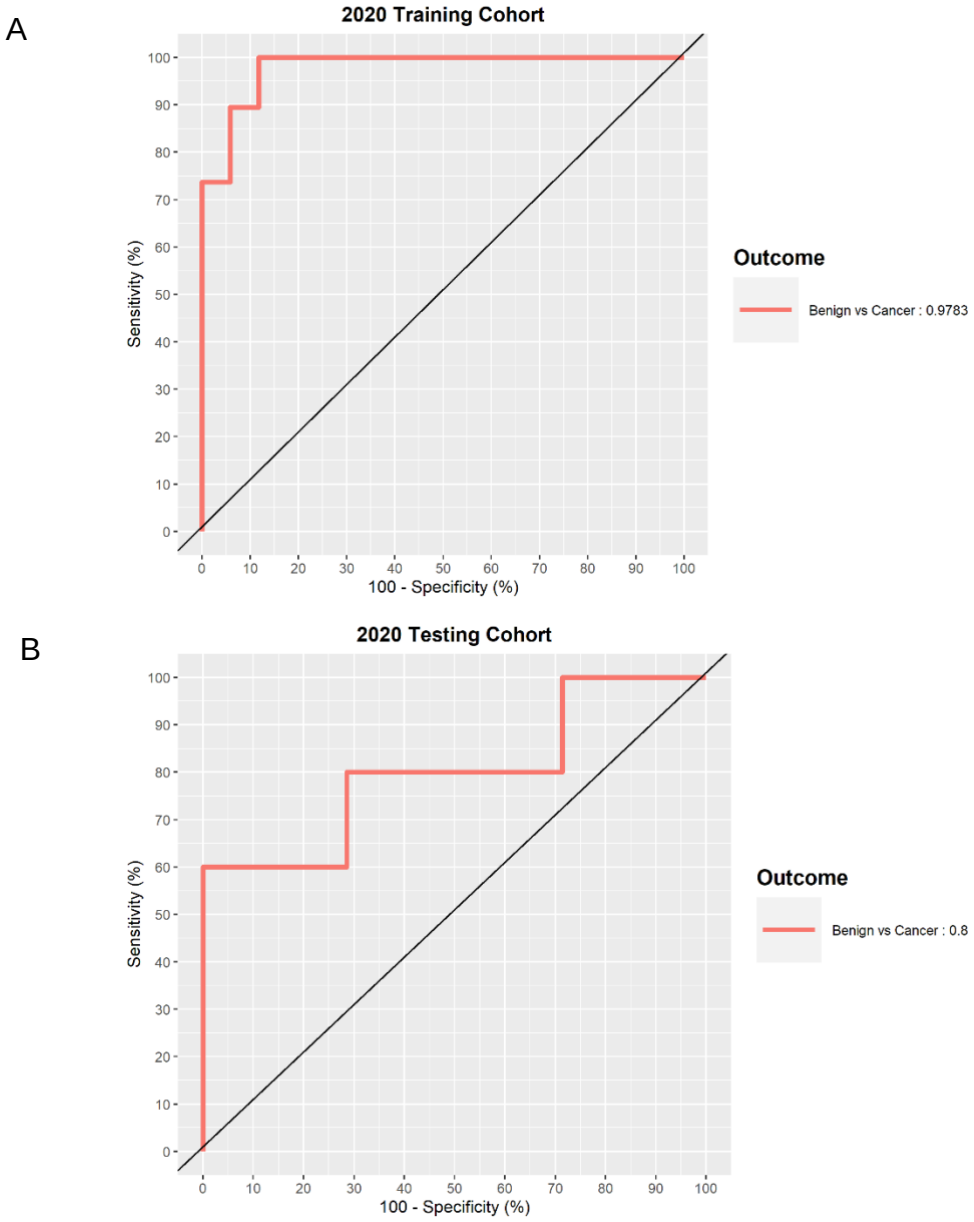



***Figure 30. Receiver operating characteristic curve for classification accuracy of the 2020 training (A) and testing (B) datasets.***

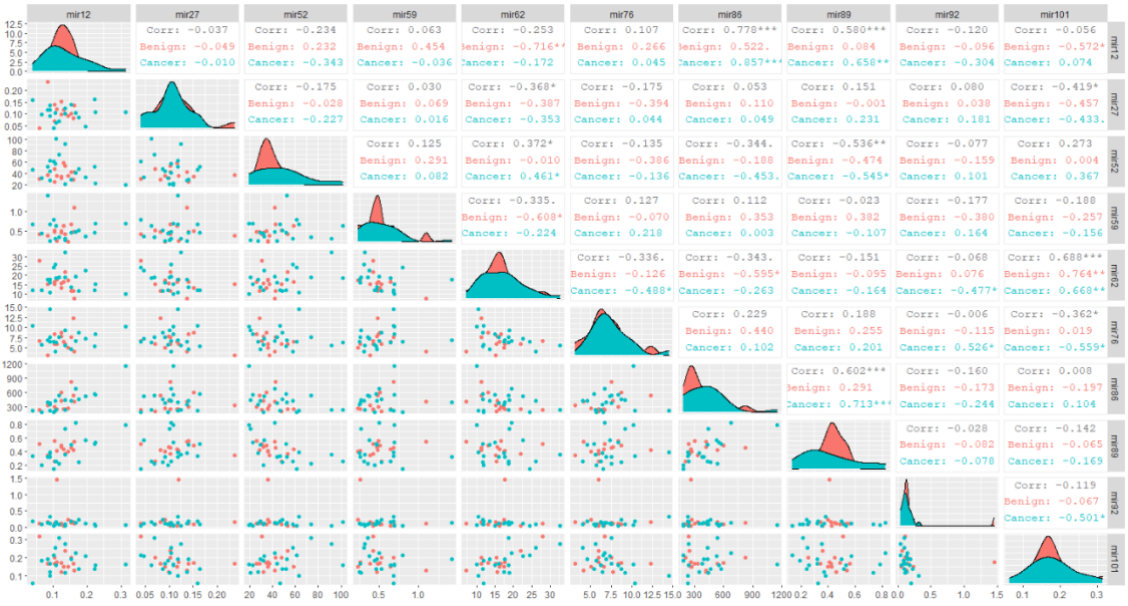



***Figure 31. Spearman correlation between microRNAs contributing the classification between cancer and control (benign) patients. *, p < 0.05. **, p< 0.01. ***, p < 0.001. Deidentified microRNAs are presented here.***

***Table 27. Classification table of 2020 testing dataset classification based on two SPLSDA components.***

| | | Clinical Diagnosis | |
|---|---|---|---|
| | | Cancer | Control |
| Predicted | Cancer | 3 | 0 |
| | Control | 2 | 7 |

Control, control patients with gastro-oesophageal reflux disease with no evidence of malignancy or patients who presented to gastroenterology clinical services with no evidence of malignancy or other inflammatory disease. Cancer, patients with oropharyngeal squamous cell carcinoma. Model sensitivity and specificity for detecting cancer is 60% and 100% respectively.

The model generated using the 2020 samples classified the 2018 dataset with a sensitivity of 35% and a specificity of 80% for cancer diagnosis. When these samples were substituted with each other (2018 – training and 2020 – testing), the resulting sensitivity and specificity for cancer detection were 21% and 96% respectively (Table 29).

***Table 28. Classification table of SPLSDA model generated from the 2020 dataset tested on the 2018 data as an independent testing cohort.***

| | | Clinical Diagnosis | |
|---|---|---|---|
| | | Cancer | Control |
| Predicted | Cancer | 14 | 6 |
| | Control | 25 | 33 |

Control, control patients with gastro-oesophageal reflux disease with no evidence of malignancy or patients who presented to gastroenterology clinical services with no evidence of malignancy or other inflammatory disease. Cancer, patients with oropharyngeal squamous cell carcinoma. Model sensitivity and specificity for detecting cancer is 35% and 85% respectively. Three component were a under the curve for training set = 0.9826 ($p < 0.0001$), testing set = 0.5970 ($p = 0.1405$)

***Table 29. Classification table of SPLSDA model generated from the 2018 dataset tested on the 2020 data as an independent testing cohort.***

| | | Clinical Diagnosis | |
|---|---|---|---|
| | | Cancer | Control |
| Predicted | Cancer | 5 | 1 |
| | Control | 19 | 23 |

Control, control patients with gastro-oesophageal reflux disease with no evidence of malignancy or patients who presented to gastroenterology clinical services with no evidence of malignancy or other inflammatory disease. Cancer, patients with oropharyngeal squamous cell carcinoma. Model sensitivity and specificity for detecting cancer is 21% and 96% respectively. Single component area under the curve for training set = 0.8297 ($p < 0.0001$), for testing set = 0.6007 ($p = 0.2317$).

### 3.3.3 SVM

Unsupervised classification above indicated poor separation of these datasets. Therefore, SVM is unlikely to be useful in this study for predictive analysis as poor unsupervised classification is considered a contraindication for using SVM methods for subsequent modelling. As seen in the figure below Figure 32, the confidence intervals were significantly wide and often crosses the 0.5 AUC. Therefore, no further direction comparisons to other modelling methods were made using SVM.

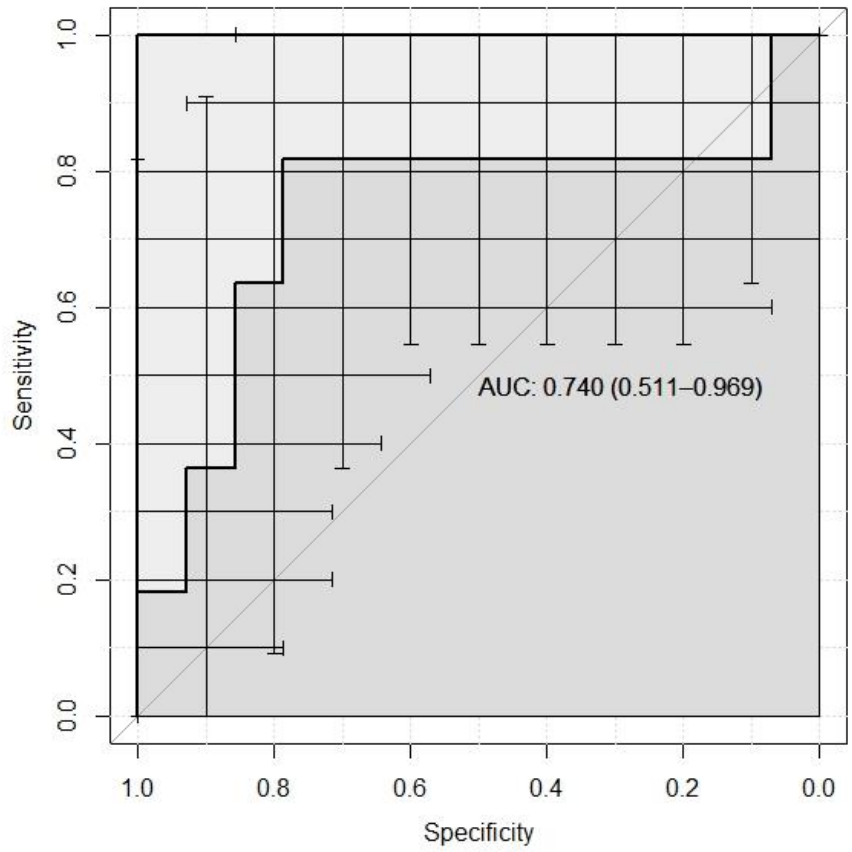



***Figure 32. Receiver operating characteristic curve of the 2020 dataset following SVM C-classification using radial bias function. AUC with confidence intervals displayed on the plot. Error bars indicate the 95% confidence intervals. AUC, area under the curve.***

### 3.3.4 Random Forest

Similar fashion to SVM, random forest modelling failed to classify these datasets accurately based on diagnosis. Figure 32 indicates that the accuracy of prediction based on a RF model plateaus at 0.714 with an n-tree of 50. Also, the incorporation of additional mtry values reduced the overall accuracy of the model (Figure 34), indicating poor fitting RF algorithm for predictive analysis of this data. Therefore, no further direct comparisons between other modelling algorithms were made in this thesis.

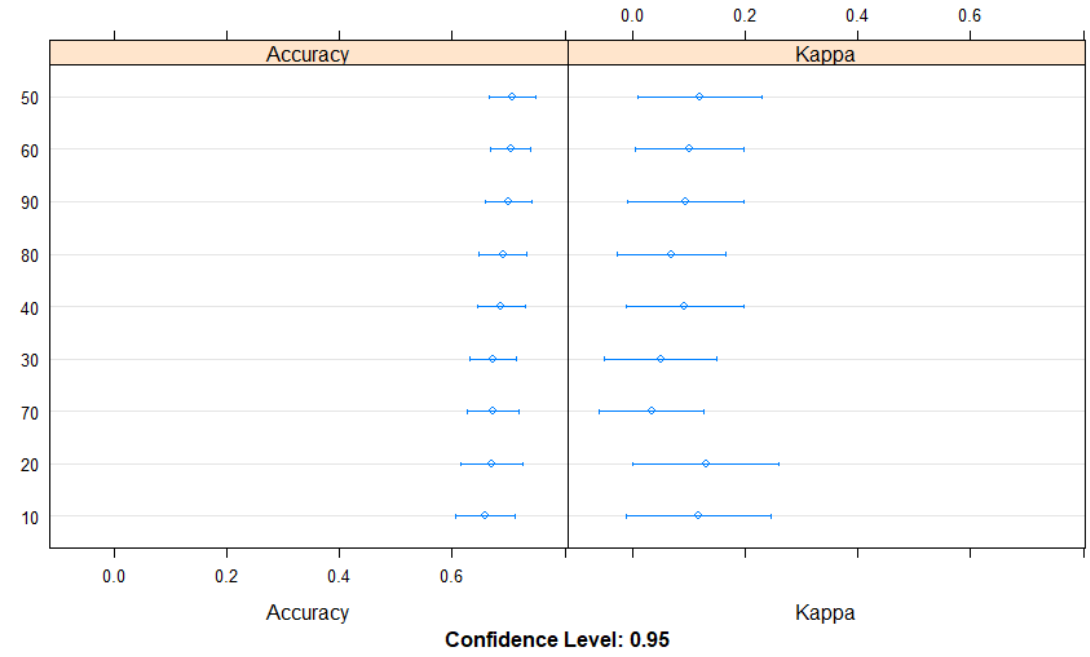



***Figure 33. Tuning random forest model to identify the best ntree value. This plot indicates a ntree=50 to be the most accurate with a median accuracy of 0.714 (range 0.429 – 1.000).***

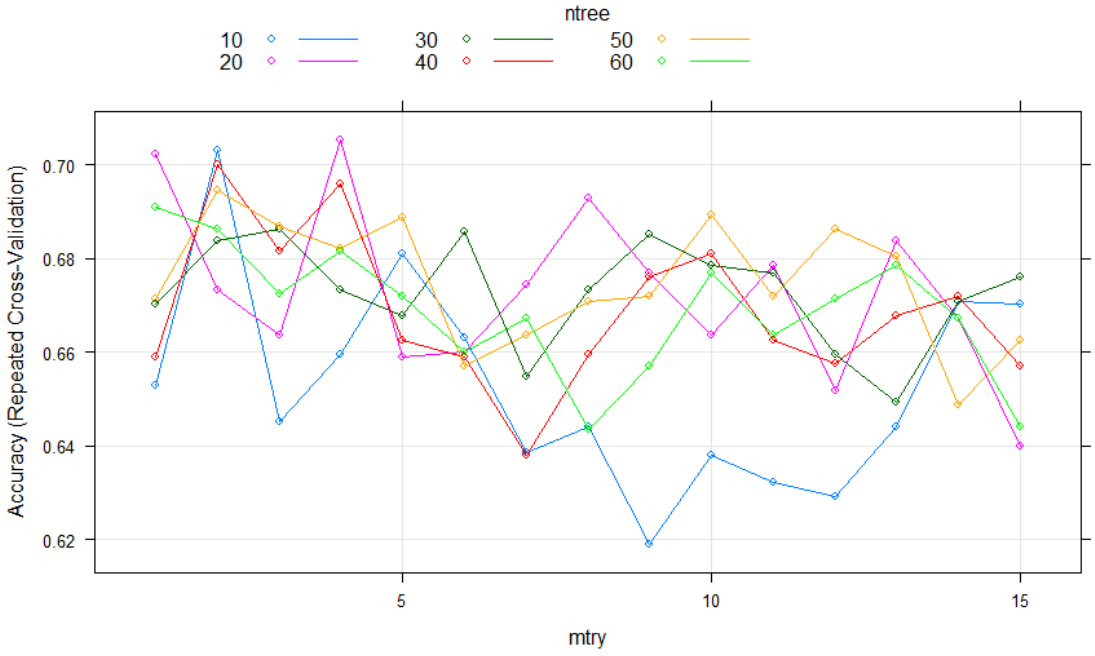



***Figure 34. Tuning random forest model to identify the best ntree and mtry values. This plot indicates that the highest accuracy of the model is achieved with a ntree = 20 and mtry = 4.***

### 3.3.5 Important microRNAs and their targets to consider for detecting HNSCC based on above classification methods.

The microRNA data integration portal (mirDIP) was utilised to generate a table of all microRNAs identified by this study as predictive of HNSCC. The UniProt accession numbers for the previous search was used to identify relevant genes, proteins, and relevant functions (Table 30). The top 10 genes/proteins with the highest number of overlapping microRNAs that target them (Table 31) as well as the bottom 10 with the least number of microRNAs that target them (Table 32) are reported here.

***Table 30. Summary of microRNAs considered to contribute to predictive modelling based on various algorithms tested here.***

| Name | miRBase # | Chromosome | Roles | HNSCC evidence |
|---|---|---|---|---|
| hsa-miR-18a-5p | MIMAT0000072 | 13 | OC | Yes |
| hsa-miR-21-5p | MIMAT0000076 | 17 | OC | Yes |
| hsa-miR-26a-5p | MIMAT0000082 | 3 and 12 | TS | Yes |
| hsa-miR-27a-3p | MIMAT0000084 | 19 | OC | Yes |
| hsa-miR-93-5p | MIMAT0000093 | 7 | MX | Yes |
| hsa-miR-106a-5p | MIMAT0000103 | X | TS | Yes |
| hsa-miR-125a-5p | MIMAT0000443 | 19 | MX | Yes |
| hsa-miR-150-5p | MIMAT0000451 | 19 | MX | Yes |
| hsa-miR-152-3p | MIMAT0000438 | 17 | TS | Yes |
| hsa-miR-193a-3p | MIMAT0000459 | 17 | TS | Yes |
| hsa-miR-193b-3p | MIMAT0002819 | 16 | TS | Yes |
| hsa-miR-199a-3p | MIMAT0000232 | 1, 9 and 19 | TS | Yes |
| hsa-miR-206 | MIMAT0000462 | 6 | TS | Yes |
| hsa-miR-222-3p | MIMAT0000279 | X | OC | Yes |
| hsa-miR-223-3p | MIMAT0000280 | X | OC | Yes |
| hsa-miR-375 | MI0000783 | 2 | TS | Yes |
| hsa-miR-483-5p | MIMAT0004761 | 11 | OC | Yes |
| hsa-miR-494-3p | MIMAT0002816 | 14 | TS | Yes |
| hsa-miR-574 | MI0003581 | 4 | MX | Yes |
| hsa-miR-590-5p | MIMAT0003258 | 7 | NA | Yes |
| hsa-let-7e-5p | MIMAT0000066 | 19 | TS | Yes |

hsa, homo sapiens. miR, microRNA. 3p, three prime. 5p, five prime. miRBase, mirbase.org accession number. OC, oncologic. TS, tumour suppressive. MX, mixed. NA, not

applicable/unknown. HNSCC, head and neck squamous cell carcinoma. Chromosome location identified using rnacentral.org.

***Table 31. Protein and their corresponding genes with the most overlapping target microRNAs of interest***

| Rank | UniProt # | Protein | Gene | Relevant Functions |
|---|---|---|---|---|
| 1 | Q96LI5 | CCR4-NOT transcription complex subunit 6-like | CNOT6L | mediate cell proliferation and survival |
| 2 | P54253 | Ataxin-1 | ATXN1 | Repress NOTCH singalling pathway |
| 3 | P51608 | Methyl-CpG-binding protein 2 | MECP2 | Binds to methylated DNA |
| 4 | P61962 | CUL4-associated factor 7 | DCAF7 | Craniofacial development and diseased skin development |
| 5 | Q6ZU65 | Ubinuclein-2 | UBN2 | Nil known |
| 6 | Q9UPQ9 | Trinucleotide repeat-containing gene 6B protein | TNRC6B | Wnt signaling pathway and calcium modulating pathway |
| 7 | P08047 | Transcription factor Sp1 | SP1 | regulates cell growth, apoptosis, differentiation and immune responses |
| 8 | Q96QR8 | Transcriptional activator protein Pur-beta | PURB | regulates cell differentiation and proliferation |
| 9 | O00712 | Nuclear factor 1 B-type | FNIB | Present in viral and cellular promoters. Essential for brain and crainal nerve development |
| 10 | O94916 | Nuclear factor of activated T-cells 5 | NFAT5 | Transcriptionally regulate cytokine production |

Ranked according to highest to lowest overlap between microRNAs of interest. UniProt, Universal Protein Resource (UniProt 2019).

***Table 32. Protein and their corresponding genes with the least overlapping target microRNAs of interest***

| Rank | UniProt # | Protein | Gene | Relavant Functions |
|---|---|---|---|---|
| 1 | Q8N5Z0 | Kynurenine | AADAT | mitochondrial protein |
| 2 | O75027 | ATP binding cassette sub-family B member 7 | ABCB7 | mitochondrial protein |
| 3 | Q9NUT2 | Mitochondrial potassium channel ATP-binding subunit | ABCB8 | mitochondrial protein |
| 4 | Q5T3U5 | Multidrug resistance-associated protein 7 | ABCC10 | transmembrane transport |
| 5 | Q96J65 | Multidrug resistance-associated protein 9 | ABCC12 | transmembrane transport |
| 6 | O15438 | Canalicular multi-specific organic anion transporter 2 | ABCC3 | transmembrane transport |
| 7 | Q9UNQ0 | Broad substrate specificity ATP-binding cassette transporter | ABCG2 | cellular detoxification, transmembrane transport |
| 8 | Q9NUJ1 | Mycophenolic acid acyl-glucuronide esterase | ABHD10 | mitochondrial protein |
| 9 | Q7Z5M8 | Protein ABHD12B | ABHD12B | nil known |
| 10 | Q5VST6 | Alpha/beta hydrolase domain-containing protein 17B | ABHD17B | regulate protein localisation |

Ranked according to low - high overlap between microRNAs of interest. UniProt, Universal Protein Resource (UniProt 2019).

## 3.4 Discussion

### 3.4.1 Classification methods

Novel genomic techniques such as OpenArrays used in this study to investigate circulating small extracellular vesicle microRNAs in OPSCC generates large quantities of data, that require careful deciphering. In the past, data generation was the bottleneck, whereas today the ability to extract the biologically relevant information from large datasets has become a limitation. Analysis of this form of data brings researchers to the crossroads of classical statistics and machine learning. In statistical methods, inferences were made based on the distribution of

data and related assumptions. Whereas machine learning aims to use all available data to generate algorithms to recognise patterns, classify based on those patterns and prediction of new data using the algorithms generated by learning from previously available data (Van Calster *et al*. 2019). Although, Bayesian statistics (that has no set distribution assumptions) falls in the middle with inference methods of Frequentist statistics at one end and machine learning algorithms at the other.

This study has demonstrated that classical statistical comparison of microRNA levels between cancer and control groups yielded no statistically significant results when adjusted for multiple comparisons. However, when the same data was used for classification based on a supervised algorithm (SPLSDA), there is clear separation of the dataset with good disease predictive ability (AUC > 0.7) for both training testing datasets. Interestingly, the major difference is that classical statistical testing penalises datasets with large variance, whereas machine learning embraces variance. For example, the generation of principal components are based on how well a single component in the algorithm can explain the variance of a given dataset. Here the sample dependent variability of a single microRNA level is 'loaded' to principal component as an arbitrary value. However, this assumes that all microRNA levels provided were normalised and scaled. Hence, continual dependence of classical statistics where scaling of individual variables can be explained as mathematically transforming all variables to the same distribution (Bzdok *et al*. 2018).

As introduced earlier, there are two primary types of machine learning a) supervised and b) unsupervised. Majority of the complex algorithms that can accurately predict the dependent variable are supervised methods (Xu and

Jackson 2019). In this thesis, one unsupervised method (PCA) and three supervised machine learning methods (SPLSDA, SVM and Random Forest) were explored. The unsupervised PCA classification of the dataset was instrumental in identifying inherent differences between the 2018 and 2020 datasets that were run on separate custom OpenArrays. However, the same unsupervised classification was unable to differentiate between patients with and without cancer.

Current shortcomings in the understanding of molecular heterogeneity within HNSCC limits the use of completely unsupervised classification methods in genomic data. To explain further, while clinically it would be ideal to be able identify a patient with cancer from a pool of patients, the molecular signature of individual patients have a spectrum of molecular level heterogeneity (i.e. not dichotomous) (Pusztai *et al*. 2006; Zhang *et al*. 2014). Therefore, expecting a completely unsupervised machine learning algorithm to classify data into two molecularly arbitrary states results in poor classification as found in this study. In contrast, supervised machine learning techniques such as SPLSDA overcome such misclassification by incorporating the biologically relevant information to the dependent variable (in this case cancer vs control) of the classification algorithm. Although, it is paramount that an independent test cohort is used to validate the model accuracy. In this study, the SPLSDA model is able to accurately classify the independent test cohort of patients to cancer or control.

Each of the classification methods used in this study have their limitations. The main limitation of PCA is the need for a complete dataset, where complete datasets are rare in health sciences. Imputation of missing data is accepted in many machine learning spheres but not so in health sciences. Valid outliers exist in real biological data and imputation of missing data may over-correct such data

by normalising to a centroid value. SPLSDA is found to be the most robust machine learning technique with the ability to handle missing data, while reducing the dimensions similar to a PCA analysis. Although, careful pre-processing of data is required a) exclusion of variables with a certain percentage of missing or zero values, b) transforming variables to a single arbitrary scale and c) randomly splitting the dataset to a training and a testing cohort. It is important to note that all above steps has the potential to add bias to the analysis. The sample size ultimately becomes the key issue for reducing bias and gaining accuracy of the model as described earlier. A systematic search of PubMed found no HNSCC related microRNA studies that utilised SPLSDA or any form of partial least square discriminant analysis. However, it has been utilised successfully in the setting of breast cancer (Jiang *et al*. 2018), colon cancer (Hofsli *et al*. 2013) and neuroblastoma (Utnes *et al*. 2019).

### 3.4.2 Roles of microRNAs with a significant association with disease prediction

The microRNA levels used for predictive modelling above were coded for legal reasons associated with a patent pending. However, the circulating microRNA markers that had a significant contribution to the predictive modelling are discussed below. The roles of such microRNAs were divided into previously described tumour suppressor, oncogenic or mixed roles in HNSCC as well as roles in cancer pathways with no clear evidence of involvement in the pathogenesis of HNSCC. The following databases were used to confirm the updated nomenclature of the microRNAs of interest, their targets, their molecular roles, and related publications: mirbase.org and ncbi.nl.nih.gov/gene. While the microRNAs detailed below were identified based on predictive modelling, further

individual laboratory-based analyses are required prior to definite claims of clear mechanistic action of these particular microRNAs in OPSCC can be made (these steps were out of scope for this thesis).

#### 3.4.2.1 MicroRNAs with tumour suppressor roles

MicroRNA-26a is considered to have a protective role in HNSCC and located in chromosome 3p22.2 with a single exon. Tissue microRNA studies exhibit a downregulation of microRNA-26a in oral SCC (Fukumoto *et al*. 2015) and in addition, overexpression of microRNA-26a in cell culture inhibits cell proliferation (Zhu *et al*. 2012) and malignant invasion. Both *in vitro* and *in vivo* studies have demonstrated a tumour suppressive role for microRNA-26a in HNSCC. While there are no current studies examining the role of microRNA-26a as a minimally invasive biomarker in HNSCC, other cancer studies have demonstrated a down regulation of it in serum leads to poor prognosis (Cho *et al*. 2017; Ghanbari *et al*. 2015; Hauser *et al*. 2012; Qiu *et al*. 2016). Mechanistically, in HNSCC microRNA-26a has associations with multiple molecular pathways a) the activation of pRb to down regulate the G1 to S transition of the cell cycle (Zhu *et al*. 2012), b) down regulation of matrix metalloproteinase-2 (MMP-2) and MMP-9 proteins by directly inhibiting cyclin-dependent kinases regulatory subunit-2 (CKS2) leading to a reduction of cell proliferation, migration, and invasion (Wu *et al*. 2018b). In contrast, to the evidence above an *in vitro* model suggested an oncogenic role in ovarian cancer cells (Shen *et al*. 2014). Although, no further validation of this work is reported. As a biomarker of HNSCC, this microRNA is unlikely to be specific to HNSCC given the current evidence. However, further studies are warranted to examine the accuracy of diagnosing or prognosticating HNSCC with microRNA-26a.

MicroRNA-106a-5p is located in chromosome Xq26.2 and associated with oral SCC that inhibits cell proliferation and epithelial-mesenchymal transition that prevents metastasis (Shi *et al*. 2019). The ability to inhibit epithelial-mesenchymal transition is also demonstrated in nasopharyngeal carcinoma (Zheng *et al*. 2019). Other cancers have also demonstrated a tumour suppressive role for microRNA-106a-5p (Chen and Pan 2019; Zhu *et al*. 2019). Although, there is a large body of literature associating microRNA106a-5p to inflammatory diseases such as coronary artery disease, myasthenia gravis (Xu *et al*. 2020), osteoarthritis (Zhang *et al*. 2019) and pre-eclampsia (Zhao *et al*. 2020). While it is possible that microRNA-106a-5p has tumour suppressive role, the wide association with inflammatory diseases suggests a lack of specificity to HNSCC.

MicroRNA-152 is considered a therapeutic target in HNSCC, that is downregulated in oral SCC (Ma *et al*. 2018a), laryngeal SCC (Song *et al*. 2014) and NPC (Li *et al*. 2017b). Clinically, it is associated with lymph node metastasis and poor survival (Ma *et al*. 2018a). Over-expression of microRNA-152 appears to directly down-regulate cellular-mesenchymal to epithelial transition factor (cMET) as well as related PI3K/AKT signalling pathway (Ma *et al*. 2018a). Therefore, cMET/PI3K/AKT pathway is thought to be the tumour suppressive mechanism of microRNA-152 in oral SCC. However, conflicting evidence suggests microRNA-152 inhibits apoptosis and promote cell proliferation in NPC via down-regulation of phosphatase and tensin homolog (PTEN) (Huang *et al*. 2016).

MicroRNA-193a is located in chromosome 17q11.2 and appears to be silenced by DNA hypermethylation in head and neck SCC (Chen *et al*. 2013). In oral SCC cell

lines, over-expression of microRNA-193a down-regulates E2F transcription factor 6 and reduces cell growth (Kozaki *et al*. 2008).

MicroRNA-199a-3p and -5p has been reported to down-regulate in HPV positive HNSCC (Miller *et al*. 2015). MicroRNA-199a-3p is located in chromosome 19p13.2 and has a tumour suppressive role in various cancers including thyroid carcinoma (Liu *et al*. 2017a; Minna *et al*. 2014) and cutaneous SCC (Wang *et al*. 2014a). Uniquely, microRNA-199a-3p is also reported to have a role in regulating viral replications in hepatitis C (Murakami *et al*. 2009), hence, having a potential role in HPV replication. In addition, microRNA-199a-5p is also reported as a tumour suppressor in oral SCC (Wei *et al*. 2019a; Wei *et al*. 2019b).

MicroRNA-206 is a tumour suppressor located in chromosome 6p12.2, that is implicated in cutaneous (Singh *et al*. 2016) and mucosal squamous cell carcinoma (SCC) (Liu *et al*. 2017b). In laryngeal SCC, microRNA-206 is demonstrated to down regulate CyclinD2, a regulator of cell cycle at G1 to S phase that is often over-expressed in SCC tissues (Yu *et al*. 2015b). VEGF and EGFR were over-expressed in HNSCC and their expression is down-regulated by up-regulation of microRNA-206 activity (Bonner *et al*. 2006; Koshizuka *et al*. 2017a; Zhang *et al*. 2011). The tumour suppressor activity of microRNA-206 is postulated to occur by down regulating HDAC6, p-ERK1/2 and p-ATK expression via the PTEN/AKT/mTOR pathway (Liu *et al*. 2017b) as well as directly affecting the expression of VEGF and EGFR.

MicroRNA-375 is located in chromosome 2q35 and is extensively described in head and neck literature with respect to oral (Wu *et al*. 2017; Zhang *et al*. 2017), laryngeal (Guo *et al*. 2016; Wang *et al*. 2016) and sinus (Kinoshita *et al*. 2012)

SCC. Expression of microRNA-375 is down-regulated in tumour tissue and is thought to act via LAMC1, EDIL3, FN1, VEGFA, IGF2BP2 and IGF2BP3 in head and neck SCC (Cen *et al*. 2018). Janus kinase 2 (JAK2) is a targeted by microRNA-375 to inhibit IFN-gamma induced programmed death ligand 1 (PDL-1) expression. Hence, an immunomodulatory function of microRNA-375 (Wu *et al*. 2018a).

Low expression of microRNA-494 in whole blood is associated with recurrence of disease following treatment of oral SCC (Ries *et al*. 2017). Its levels were also lower in advanced stage oral SCC patients compared to patients with early or no oral SCC (Al-Khanbashi *et al*. 2016; Ries *et al*. 2014a). It is located in chromosome 14q32.31 and associated with microRNA-154 family (Nie *et al*. 2015). Mechanisms of tumour suppression has been demonstrated to involve down-regulation of Bmi1 to induce cellular senescence (Chang *et al*. 2015). A similar effect on cells is seen after radiation therapy (Weng *et al*. 2016). Furthermore, HOXA10 has an inversely proportional expression pattern compared to microRNA-154 in oral SCC. Where HOXA10 is found to be up-regulated in oral SCC cells (Liborio-Kimura *et al*. 2015). In nasopharyngeal carcinoma GALNT7 is carcinogenic causing structural changes in cell proteins that promotes tumour invasion properties, and the up-regulation of microRNA-494 inhibits the growth of nasopharyngeal carcinoma via the inhibition of GALNT7 (Duan *et al*. 2016; Nie *et al*. 2015).

MicroRNA-let-7e-5p (previously known as microRNA-let-7e) is located in 19q13.41 with a single exon has been reported to inhibit the proliferation and metastasis of HNSCC (Wang *et al*. 2019). It is also reported in various other malignancies to have a tumour suppressive role (Liu *et al*. 2020b; Park and Kim 2019; Tsai *et al*.

2015; Xiao *et al*. 2017; Zhu *et al*. 2014). Although, it appears to have a strong correlation with active inflammation (Beyer *et al*. 2015; Lin *et al*. 2017; Zwiers *et al*. 2012), suggesting that it may be a predictor of inflammation rather than malignancy.

#### 3.4.2.2 MicroRNAs with oncogenic roles

MicroRNA-18a-5p (previously known as microRNA-18a) is located in chromosome 13q31.3 is reported to have an oncogenic role *in vitro* oral SCC cell line. Where they found a higher expression level of microRNA-18a-3p in oral SCC cell and promoted cell viability and invasion while inhibiting apoptosis upon overexpression (Huang *et al*. 2019). This oncogenic role is reported to be facilitated by inhibiting the expression of SMAD2 (Huang *et al*. 2019). Where SMAD2 is signal transducer protein previously reported to suppress tumour progression (Liu *et al*. 2015).

MicroRNA-21-5p is 22 nucleotides long and located in chromosome 17 of the human genome. It is well described in HNSCC, predominantly as regulator of tumour progression in oral cavity SCC (Zeng *et al*. 2019). It is also described in association with multiple other neoplasms including brain tumours (He *et al*. 2016b), lung cancer (Ma *et al*. 2012), breast cancer (Fu *et al*. 2011), oesophageal cancer, gastric cancer (Wang *et al*. 2014b), colorectal cancer (Xia *et al*. 2013), hepatocellular carcinoma (Kim *et al*. 2018b), pancreatic cancer (Liu *et al*. 2018), cervical cancer (Zhang *et al*. 2018b), leukemias (Ma *et al*. 2018b; Zhou *et al*. 2017) and lymphomas (Fu *et al*. 2011). In all types of cancers where microRNA-21-5p is associated with, the higher level of its expression is directly proportional to poor overall prognosis. However, the reported thresholds for microRNA-21-5p in each of the cancers vary significantly (Fu *et al*. 2011). Therefore, pooled

analysis of threshold levels to identify a cut-off threshold for disease diagnosis is only possible per specific tumour subsite (Yang *et al*. 2014). In oral cancer, microRNA-21-5p dependent positive regulation of cytokine release (Mehdipour *et al*. 2018b), cell proliferation (Mahmoud *et al*. 2018), cell infiltration (Kawakita *et al*. 2014; Yu *et al*. 2017), angiogenesis, as well as post-transcriptional gene silencing of tumour suppressor pathways (Tseng *et al*. 2017) have been previously reported. As a biomarker, microRNA-21-5p is unlikely to be HNSCC specific given its association with multiple neoplasms. However, its significant predictive ability proves useful as a screening or a prognostic tool for cancer, especially given its detectable in serum (Ren *et al*. 2014; Singh *et al*. 2018), saliva (Mehdipour *et al*. 2018a) and cytological (He *et al*. 2016a) specimens for minimally invasive sampling. However, in oral SCC a common clinical question is whether a lesion is inflammatory or malignant, where further investigations or treatments can be tailored. While microRNA-21-5p has the ability to detect pre-malignant conditions such as sub mucosal fibrosis (Singh *et al*. 2018), further research is warranted to investigate its ability to differentiate between chronic inflammatory lesions and early carcinomas.

MicroRNA-27a is located in chromosome 19p13.12. In oral SCC microRNA-27a induces epidermal-mesenchymal transition (EMT) via Wnt/beta-catenin pathway with directly targeting the downregulation of secreted frizzled-related protein 1 (SFRP1) (Qiao *et al*. 2017) and Yes-associated protein-1 (YAP1) (Zeng *et al*. 2016). Oncogenic role of microRNA-27a is also action by down regulation go f microcephalin 1 (MCPH1) gene expression, a previously described tumour suppressor gene in oral SCC. MicroRNA-27a is also overexpressed in laryngeal cancer tissues and has a strong negative correlation with level of cell differentiation with high expression levels in poorly differentiation laryngeal SCC

(Chen *et al*. 2017). Targets of microRNA-27a in laryngeal SCC is identified as APAF-1 and PLK2 (Tian *et al*. 2014; Wang *et al*. 2016). Over expression of microRNA27a may also cause resistance to treatment by up regulating intracellular heat shock proteins (Kariya *et al*. 2014). While microRNA-27a is extensively investigated *in vitro* and *in vivo* cancer models, its ability in diagnosing or prognosticating HNSCC is not currently established.

MicroRNA-193b is located in chromosome 16p13.12 and is shown to be over-expressed in tumour tissues compared to control inflammatory tissues irrespective of p16 status associated with poor survival (Bersani *et al*. 2018; Wong *et al*. 2016). There is a strong negative correlation between microRNA-193b and CD8+ T-cell infiltration in HPV positive tonsil and base of tongue SCC likely to be due to up-regulation of IL-10 expression by microRNA-193b (Bersani *et al*. 2018). It is also implicated in the pathogenesis of oral SCC (Li *et al*. 2017a). Further head and neck SCC tissue studies have found a lower disease-free survival associated with high expression of microRNA-193b that is thought to drive tumour progression by down-regulating neurofibromin 1 (NF1) and activating ERK pathway (Lenarduzzi *et al*. 2013).

MicroRNA-222-5p is located in chromosome Xp11.3 and closely associated with microRNA-221-5p. In oral cavity SCC, it has an oncogenic role with promotion of cell proliferation and migration (Liu *et al*. 2009). It is upregulated in tongue SCC (Zhao *et al*. 2015) and related downregulation of p53 protein is reported (Jiang *et al*. 2014). Its broad association with thyroid cancers (Huang *et al*. 2018b), their metastasis (Liang *et al*. 2018), their recurrence (Lee *et al*. 2013) and in anaplastic differentiation is also noted (Xiang *et al*. 2019). Other cancers have also reported an oncogenic role associated with microRNA-222-5p (Wei *et al*. 2016b). Due to

the clear oncogenic role of microRNA-222-5p, it would be an ideal candidate to be included in a panel of microRNAs capable of diagnosing HNSCC.

MicroRNA-223-3p (previously known as microRNA-223) is located in human chromosome Xq12 and described in the setting of inflammatory disease and cancer including HNSCC (Jeffries *et al*. 2019). In laryngeal SCC, microRNA-223-3p is thought to have an oncogenic role where *in vitro* competitive inhibition of microRNA-223-3p lead to regulation of TGF-beta-R3 and limit laryngeal cell proliferation (Wei *et al*. 2019c). Has.circ.0042666 a novel circulating microRNA is reported competitively inhibit microRNA-223-3p and limit progression laryngeal cancer cells *in vitro* (Wei *et al*. 2019c). MicroRNA-223-3p has been previously detected in serum samples and its diagnostic potential has been described in oral SCC (Tachibana *et al*. 2016), nasopharyngeal carcinoma (Zeng *et al*. 2012) as well the ability to monitor treatment efficacy (Hou *et al*. 2015a). While, expression of this microRNA is not limited to HNSCC, clear potential as biomarker in HNSCC has been demonstrated and would form an ideal candidate for a microRNA panel to detect HNSCC.

MicroRNA-483 is upregulated in oral SCC and positively correlates with increasing stage of disease (Xu *et al*. 2016a). MicroRNA-483 mediated inhibition of DCD25A expression in-turn inhibits the proliferation of keratinocytes. This action is thought to be via the phosphorylation of CDK4/6 kinase activation to sequester cells in early G1 phase (Bertero *et al*. 2013). MicroRNA-483 over-expression is also noted to sensitize tumour cells to chemotherapeutic and other pro-apoptotic drugs such as cisplatin (Fan *et al*. 2015). This pro-apoptotic effect is via inhibiting expression of ‘survivin’ (BIRC5) and RAN, where these are known to be overexpressed in SCC and other carcinomas (Bertero *et al*. 2013). MicroRNA-483

also downregulates FIS1 and affect the mitochondrial apoptotic pathways and therefore cisplatin sensitivity to tumour cells (Fan *et al*. 2016).

#### 3.4.2.3 MicroRNAs with mixed roles

MicroRNA-93 appears to have varying effects on different cell lines with some oncogenic and tumour suppressive effects. In laryngeal SCC, microRNA-93 plays an oncogenic role via regulating cyclin G2 expression (Xiao *et al*. 2015). In NPC, microRNA-93 is also found to be oncogenic by down-regulation of TGF-Beta-R2, a tumour suppressor gene in NPC (Lyu *et al*. 2014). Clinically obtained tissue samples indicate microRNA-93 expression correlates with poor overall prognosis, tumour staging and nodal staging (Li *et al*. 2015). However, in saliva of oral SCC patient's microRNA-93 has significantly lower expression compared to healthy patients indicating a tumour suppressive role (Greither *et al*. 2017).

MicroRNA-125a is located in chromosome 19q13.41. In oral SCC and NPC microRNA-125a is shown to directly interact with anti-sense noncoding RNA in the INK4 locus (ANRIL) and estrogen-related receptor alpha (ESRRA) (Chai *et al*. 2018; Hu *et al*. 2017; Tiwari *et al*. 2014). Overexpression of ANRIL and decreased expression of microRNA-125a in oral SCC tissues and serum samples indicate a tumour suppressive role in oral SCC (Chai *et al*. 2018). This tumour suppressor effect is also demonstrated in laryngeal SCC via targeting Hexokinase 2, a major catalyst in aerobic glycolysis (Sun *et al*. 2017). In contrast, in-vitro oesophageal SCC model indicates promotion of cell migration and invasion by microRNA-125a via the zinc finger E-box binding homeobox 1 (ZEB1) pathway (Ma *et al*. 2017). A contradictory role is also described in NPC where microRNA-125a is upregulated

along with microRNA-125b in chemo-resistant NPC via down-regulation of p53 (Chen *et al*. 2015).

MicroRNA-150 regulates SP1 and TGF-beta pathways that is implicated in tumour progression in head and neck SCC (Citron *et al*. 2017). Both strands of pre-microRNA-150 is downregulated in cultured HNSCC cells and SPOCK1 is thought to be a target of microRNA-150 that positively correlate with tumour aggressiveness (Koshizuka *et al*. 2017b). However, this micro-RNA is overexpressed in HPV-induced tissues (Santos *et al*. 2017) and in plasma samples of oral SCC patients (Chang *et al*. 2018). In NPC cell lines, inhibition of microRNA-150 reversed the radio-resistance that is thought to be via the downstream upregulation of GSK3-beta protein (Huang *et al*. 2018a).

MicroRNA-574 is located in chromosome 4p14 and identified as a circulating biomarker that is a marker of prognosis and therapeutic monitoring (Summerer *et al*. 2015). In HNSCC microRNA-574 is down-regulated, this tumour suppressor gene delayed cell growth *in vitro* (Veit *et al*. 2015). However, in papillary thyroid cancer, it is over-expressed promoting cell proliferation via Wnt/beta-catenin signalling pathway by affecting the downstream expression of 'Quaking', an RNA-binding protein and a suppressor of cancer cell invasion (Wang *et al*. 2017) (Zhang *et al*. 2018a). This overexpression is also noted in adenoid-cystic carcinomas of the head and neck (Veit *et al*. 2015).

#### 3.4.2.4 MicroRNAs with no clearly defined role in head and neck squamous cell carcinoma but differentially expressed in this dataset.

MicroRNA-590-5p is located in chromosome 7q11.23 and was previously reported to differentially express in a microarray study when plasma levels were compared

between before and after chemoradiotherapy treatment of patients with HNSCC (Summerer *et al*. 2013). However, this particular microRNA was below the detection levels when individual qRT-PCR studies were conducted, hence difficult to associate to the development of HNSCC. While microRNA-590-3p is reported to down regulate *in vitro* cell proliferation, migration and clonogenicity of HNSCC cell lines and other human cancer tissues (Di Agostino *et al*. 2018). The non-coding gene mir205hg appears to deplete microRNA-590-3p resulting in an upregulation of cyclin B, cdk1 and YAP proteins. Subsequently leading to tissues with p53 mutation to proliferate (Di Agostino *et al*. 2018). Although, MicroRNA-590-5p is associated with multiple other cancers (Chen *et al*. 2018; Gao *et al*. 2019; Kim *et al*. 2018a; Xiao *et al*. 2013) with a mixed role (oncogenic and tumour suppressive) and detected in tissue (Xu *et al*. 2018a) as well as serum samples (Khandelwal *et al*. 2020).

#### 3.4.2.5 Molecular level function of microRNAs of interest

Overall molecular function of the microRNAs of interest were further explored using previously curated databases. This explorative analysis elucidated molecular pathways that were previously implicated in HNSCC. Interestingly, when target proteins were regulated by multiple microRNAs of interest, cancer relevant pathways were identified. In contrast, when proteins were regulated by a single or a few microRNAs, the related molecular pathways were generic. However, metabolic pathways of cancer are very broad, and all of the pathways mentioned or described earlier have some role in cancer biology. This further supports the need for a panel of microRNAs that are related at a mechanistic level

to be included as biomarkers of HNSCC instead of a single microRNA marker that is likely to be non-specific given their many regulatory roles in health and disease.

### 3.4.3 Summary

This chapter highlights the ability of circulating microRNAs contained in small extracellular vesicles to detect oropharyngeal SCC. Importantly, it has demonstrated the need for careful selection of control groups to elucidate the research question. This study can accurately differentiate patients between cancer and potentially other inflammatory states (i.e. compared to the GORD patient cohort). However, further studies are required to identify whether the proposed microRNA signature is able to differentiate between OPSCC and oropharyngeal inflammation such as tonsillitis. Because often the clinical question is whether a patient's throat pain is an infection or cancer.

The high specificity demonstrated in SPLSDA classification of the independent test cohorts, suggest such a test could be used as a "rule in" test, where a negative result is likely to be clinically meaningless (low sensitivity), a positive test suggests that the probability of having cancer is high (85-100% specificity). Therefore, in a primary care setting a patient presenting with a sore throat (in the absence of overt features of cancer or risk factors) where the physician or the patient is concern of the diagnosis a serum test could be administered to triage the patient, with a positive test leading to immediate referral for specialist investigations compared to close monitoring with a negative test result.

While supervised classification methods were able to accurately classify datasets based on a known diagnosis, the value of unsupervised classification such as PCA

is highlighted here. Unsupervised classification allows for the identification of factors that may be contributing to the overall variance of the dataset without the researcher's knowledge and therefore should be recommended as a mandatory data explorative method prior to any supervised classification.

The microRNAs considered to be important in this study had previously demonstrated molecular roles in HNSCC (Table 30). None of the markers had a role specific to oropharyngeal cancer alone and such a molecular marker is unlikely to be present. A thorough systematic review for each microRNA of interest would be required to confirm such subsite dependent bias. Anecdotally, during the review of literature for this thesis, there were a large amount of microRNA related studies for oral cavity SCC and nasopharyngeal carcinoma compared to other subsites of HNSCC. Epidemiology of the disease may explain this finding as, majority of the publications were geographically from East Asian nations and the predominant subsites affected were nasopharyngeal and oral cavity (Cohen *et al*. 2018b).

Finally, while there is a clear potential for detecting HNSCC using a panel of circulating microRNAs, large scale studies with standardised methodology and analytical techniques need to be conducted along with concurrent assessment of health economic impact prior to routine use of this technology to improve patient outcomes.

# 4 Sources of variability and breath analytical techniques in head and neck cancer

## 4.1 Introduction

Analysis of the human volatilome provides a unique opportunity to examine the bio-signature of head and neck cancer patients (Opitz and Herbarth 2018). Volatile compounds present in breath can be used as a non-invasive method of sampling their volatilome and has the potential to be further utilised as a biomarker of screening, diagnosis, or therapeutic monitoring (Dharmawardana *et al*. 2020c). However, there is marked variability in methodology of breath analysis. In this chapter, the utility of breath analysis in head and neck cancer detection including methods of human breath collection, processing, data analysis and compounds implicated in HNSCC are described.

## 4.2 Methods

A scoping review was conducted using PubMed, Embase, Scopus, ProQuest, Web of Science (Core collection) databases. The inclusion criteria for the search strategy were a) head and neck cancer (any pathology), b) exhaled breath analysis and c) peer reviewed published studies. The search was conducted initially in February 2017 and repeated in June 2018 and June 2019 with no limitations to publication date or language to identify all relevant studies. The duplicates were removed, then title and abstract review was conducted using EndNote software (version X9.2, Clarivate Analytics, US). Twelve studies meeting the inclusion criteria were identified and described in detail. The search strategy is detailed in Appendix II. Results of this scoping review was published in Oral Oncology (Dharmawardana *et al*. 2020c).

## 4.3 Results

### 4.3.1 Methods of Breath Collection

Human breath largely consists of nitrogen, oxygen, carbon dioxide, argon, and water vapour. Exhaled VOCs are also detectable in trace concentrations up to ppt (Lourenco and Turner 2014) and various collection methods have been described to maximise the yield of exhaled VOC detection (Lawal *et al*. 2017). A complete collection of exhaled breath would constitute mixed alveolar gas concentrations, contributed by inhaled, exhaled and dead space gases. The exhaled breath can be divided into three specific phases based on the carbon dioxide concentrations, namely: dead-space, transition and alveolar, with each phase associated with increasing carbon dioxide concentrations respectively (Miekisch *et al*. 2008). VOC concentrations are relatively high in pure alveolar breath samples compared to mixed alveolar breath due to lack of dilution by dead space gases and is the preferred sample source for breath analysis. However, mixed alveolar samples are practically easier to collect, and most reported studies have described this sampling technique (Cao and Duan 2007).

#### 4.3.1.1 Direct

Direct breath sampling in its simplest form is best compared to a road-side alcohol breathalyser. These high-performance breathalysers are fuel cell testers, where the alcohol in the breath sample is oxidised by a porous platinum electrode into acetic acid, protons, and electrons (Swift 2003). Therefore, the level of electrons produced is dependent on the breath alcohol content. This generates an electrical signal that can be calibrated and measured (Swift 2003). However, more

sophisticated mass spectrometers also allow for direct sampling of breath without pre-processing.

Direct measurement is advantageous because the sample is less likely to be contaminated by storage media and allows for sampling and analysing specific phases of exhaled breath simultaneously. In a clinical setting this would be ideal as a point of care test, allowing for real-time information to be provided for clinical decision making. However, there is potential for the breath sample to be contaminated by volatile compounds released by the surrounding environment such as cleaning products, perfumes, clothing, or smoking fumes. This is applicable for any breath sampling method and careful consideration to environmental factors described in Table 33 should be made when collecting breath samples. At present, apart from e-nose devices and handheld breathalysers, there are no 'direct' sampling devices for breath analysis in HNSCC with the exception of instruments such as SIFT-MS and PTR-MS, which are large and bulky, and currently not suitable in clinical settings. Low volatile emitting bags and various sorbent canisters have been used for breath collection in clinical settings, with samples transported to a central laboratory for subsequent processing and analysis.

***Table 33. Considerations for breath sample collection***

| | Considerations |
|---|---|
| Patient preparation | Fasting, Tooth brushing, Mouth ishing Smoking status, Alcohol intake |
| Filtered inhaled air | Low VOC environment, Charcoal filter |
| Exhaled Gas fractionation | Volume based fractionation, CO2 sensor-based fractionation |
| Container | Adsorbent materials, tubes, bags |
| Room air | Measurement, exclusion, inclusion, statistical incorporation |

VOC, volatile organic compound. CO2, carbon dioxide.

#### 4.3.1.2 Temporary storage

Multiple methods for temporary storage of exhaled breath samples have been described using 1) gas-tight syringes, 2) glass containers, 3) various bags and 4) other commercial devices (Lawal *et al*. 2017). In earlier studies gas-tight glass products were considered ideal due to the low VOC profile of the container compared to all other products. However, the cost, size and durability of these glass products are prohibitive for anything outside of research applications. Therefore, recent studies have migrated to more versatile products such as low VOC bags.

Numerous commercially available breath collection bags have been described in the literature. The bags are generally constructed from low volatile emitting materials and have low reactivity with volatile compounds. They may be made using materials such as polypropylene (Mylar), aluminium, polytetrafluoroethylene ethylene (Teflon©), polyvinyl fluoride (Tedlar©), polyethylene terephthalate (Nalophan©) and stainless steel (Mochalski *et al*. 2009). Polymer bags have been popular for breath analysis due to their relative low cost, storability, and reusability. However, storage of VOCs using bags is time restricted as, the most high-performance product sees a significantly decrease in VOC concentrations after 72 hours (Mochalski *et al*. 2009). Temperature and compound dependent degradation of VOC levels in collection bags are also reported causing potential bias in biomarker studies (Ghosh *et al*. 2011). Hence, VOC analysis potentially can be affected by delay, after 72 hours, in VOC analysis after collection.

Bio-VOC is a product capable of short-term storage of exhaled breath with the ability to capture alveolar component and subsequently concentrate into solid phase microextraction (SPME) fibres or sorbent tubes (Kwak *et al*. 2014). However, the major limitation of this product has been the small sample volume it can collect. The water volume of this device is a maximum of 175mL, and air volume is reported to be less and often unpredictable (van den Velde *et al*. 2007). This can exclude the use of analytical techniques such as SIFT-MS or other time-of-flight techniques, where much larger volume in one sitting is required for analysis. An advantage of the Bio-VOC device is the ability to collect the alveolar component of exhaled breath. Exhalation of specific phases into a VOC collection bag has been described before, however, Bio-VOC does this with ease and without the need for special instructions to the participant. Overall, the Bio-VOC device would be ideal in clinical applications for measurement of the alveolar component of previously known, highly abundant VOCs. However, it is unlikely to be useful for breath VOC discovery research given the low collection volume directly proportional to low VOC intensity.

Rtube is primarily an EBC collection device that can be modified to measure exhaled VOCs along with EBC (Martin *et al*. 2010). This is a commercially available device that can collect airway lining fluid along with genomic, protein and metabolites. It is a reliable method for collecting EBC with excellent long-term data reproducibility (Huttmann *et al*. 2011; Huttmann *et al*. 2016). Use of a SPME fibre within the Rtube apparatus appears to collect similar concentrations of VOC when compared to breath collected in a bag and subsequently concentrated into SPME fibre (Martin *et al*. 2010). Hence, there is potential to use this commercial product for VOC discovery experiments as well as measurement of known

compounds with the added benefit of measuring genomic, proteomic, and other metabolomic content of exhaled breath.

#### 4.3.1.3 Long term storage

Various other canister devices have been used to collect human breath samples for VOC quantification. These are often containers made of low VOC emitting material with or without absorptive material. Glass vials and gas tight syringes are one of the simplest protocols for collection and storage of VOCs. However, due to limitations associated with absolute gas volume, the overall VOC concentration is often limited, restricting the analytical options. Therefore, canisters that contain an adsorptive material are preferred for VOC discovery research.

Thermal desorption (TD) tubes are perhaps the most commonly used vesicle for pre-concentration and storage of human breath samples prior to analysis. Breath samples collected into bags in clinical settings can be pumped into a TD tube, allowing for long-term storage. TD tubes are often constructed from stainless steel or glass with various sorbent materials used to fill the tube (Table 34). These sorbent products are largely composed of graphitised carbon black (GCB) or carbon molecular sieve (CMS) (Brown and Shirey 2001). GCB is a non-porous carbon product that can absorb a large molecular range of VOCs. However, CMS is a porous carbon skeletal framework (a molecular sieve) used to effectively trap compounds with very small molecular size. The absorption and desorption properties of CMS is dependent on the size and the shape of the molecule (Brown and Shirey 2001).

***Table 34. Sorbent materials and their applications***

| Sorbent name | Material | Application |
|---|---|---|
| Tenax TA® | 2,6-diphenylene-oxide porous polymer resin | C7 -C26 |
| Tenax GR® | Composite between graphite (30%) and 2,6-diphenylene-oxide (70%) | C7-C26 |
| Carbotrap® | Graphitised carbon black | |
| C | | C12 – C20 |
| B | | C5 – C12 |
| F | | >C20 |
| X | | C3-C9 |
| Y | | C12-C20 |
| Carboxen® | Carbon molecular sieve | C2-C5 |
| 1016 | | C3-C9 |
| 569 | | C2-C5 |
| 1000 | | C2-C5 |
| 1021 | | C2-C5 |
| 1021 | | C2-C5 |
| 1018 | | C2-C5 |
| 1003 | | C2-C5 |
| 1012 | | C2-C5 |
| Carbosieve® | Carbon molecular sieve | C2-C5 |
| G | | C2-C5 |
| S-II | | C2-C5 |
| S-III | | C2-C5 |
| Glass beads | Glass | Very large molecular weight compounds |
| XAD® | Hydrophobic organic porous polymer | Soluble organic compounds |
| Silica Gel | Silica Gel | Hydrocarbons, LMW mercaptans, methanol, amines, inorganic acids |
| Anasorb® | Synthetic carbon with low ash content similar to Carbotrap® | Polar and non-polar organic vapours |
| Charcoal | Coconut Charcoal | Non-polar compounds |
| Chemsorb® | Porous polymer | Low-boiling hydrocarbons, benzene, labile compounds, volatile oxygenated compounds |

C#, molecular carbon-size. GR, graphitised.

### 4.3.2 Methods of VOC Detection and Analysis

The current literature has described a multitude of novel technologies that has the potential to detect VOCs in human breath. These technologies fall into two main

categories. Firstly, pattern recognition devices based on arrays of particle or gaseous compound sensors. Secondly, mass spectrometry techniques that range from gas chromatography mass spectrometry (GC-MS) to real-time gas analysers such as proton transfer reaction mass spectrometry (PTR-MS) and selected-ion flow tube mass spectrometry (SIFT-MS). Studies investigating breath biomarkers for head and neck have described the use of SIFT-MS, GC-MS, PTR-MS, and E-nose technologies (Table 15 and Table 16). The use of secondary electrospray ionisation-mass spectrometry (SESI-MS) that can also conduct real-time breath analysis with relevant clinical applications is also described here.

#### 4.3.2.1 Electronic-Nose

Electronic nose (E-Nose) technology was first introduced in 1982 by Persaud and Dodd (Persaud and Dodd 1982) who described a system that emulated the human olfactory system by using an array of gas sensors and pattern recognition software (Figure 35). Since then, various modifications and improvements have been made with a range of commercialised E-nose devices available today (Rocco 2018). This technology can recognise patterns and compare between patient groups to classify patients according to disease. Five studies (Hakim *et al*. 2011; Lang *et al*. 2016; Leunis *et al*. 2014; van de Goor *et al*. 2017; van Hooren *et al*. 2016) have used E-nose to differentiate HNSCC patients from a pool of patients with and without cancer.

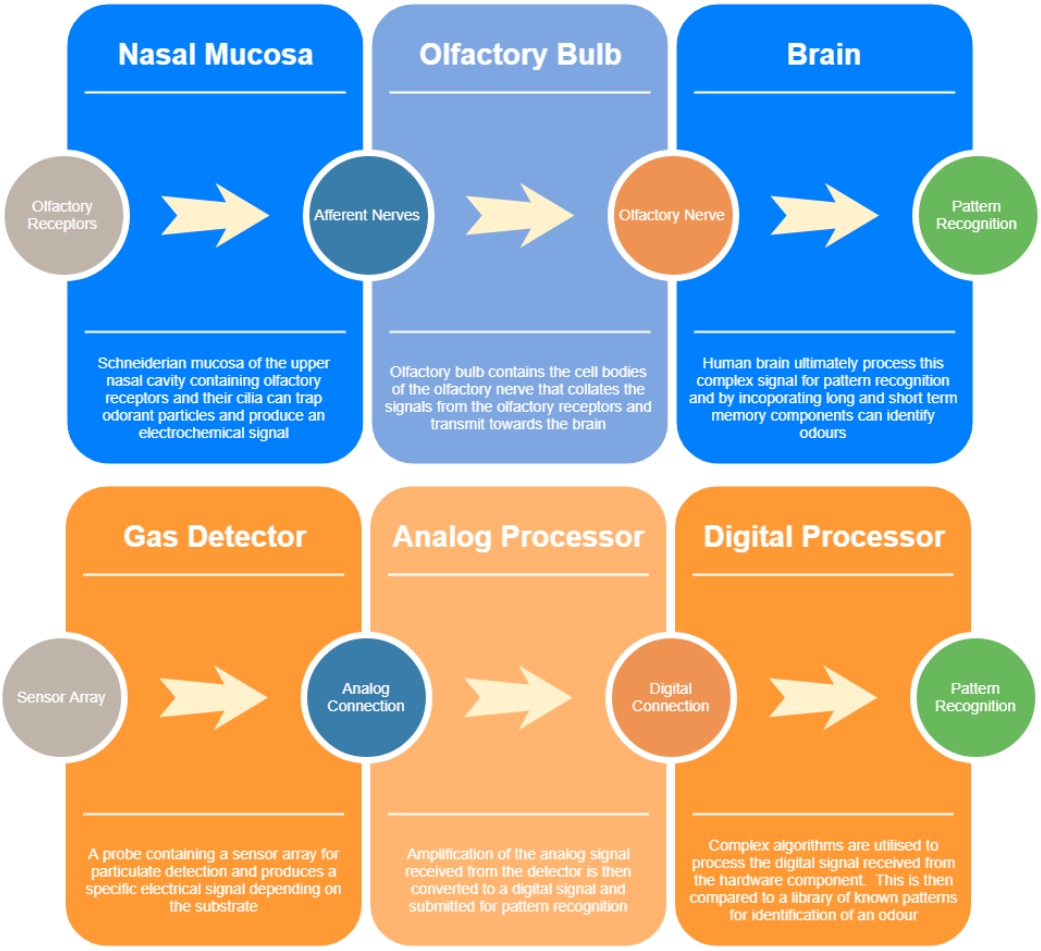



***Figure 35. Comparison of human olfactory system to E-Nose.***

Benefits of this technology includes the simplicity of administration and the lack of data processing required to identify compounds. However, a significant limitation is the difficulty of ascertaining which specific VOCs were contributing to the detection capability without concurrent analysis of the same sample using other techniques. From a research perspective utilising multiple modalities such as paring E-nose analysis with other mass spectroscopy techniques could be used to elucidate potential volatile biomarkers of interest as described by Hakim *et al* (Hakim *et al*. 2011).

van de Goor *et al* (van de Goor *et al*. 2017) and van Hooren *et al* (van Hooren *et al*. 2016) have described the use of Aeonose®. This is a type of e-nose technology

that uses reduction and oxidation reaction of VOCs with metal oxide to detect electrical conduction differences to differentiate between gas profiles. The most recent study (van de Goor *et al*. 2017) compared breath samples between patients with HNSCC (n = 100), bladder cancer (n = 40) and colon cancer (n = 28). This study utilised neural network data modelling to analyse the ability of this technology to differentiate between the groups. They were able to discriminate between HNSCC, colon cancer and bladder cancer with 81%, 84% and 84% respective overall accuracies. However, the baseline demographic characteristics of the comparison groups were significantly different and therefore, cast doubt on the validity of the overall findings. van Hooren *et al* (van Hooren *et al*. 2016) in 2016 described the comparison between HNSCC and lung cancer (LC) patients using the same protocol as above. Here, they were able to differentiate these two patient groups with 93% overall accuracy. However, as with the previous dataset, the baseline patient characteristics were significantly different between comparison groups. Additionally, the fasting protocol for this study was markedly different from other studies included in this review, where the total fasting time is 4 hours to solids and liquids, then participants were allowed to consume 'clear liquids' up to 2 hours prior to breath sample collection. Water consumption alone may not have interfered with the sampled volatiles. However, if the undefined clear liquid mentioned here contained sugar or other transparent dissolved compounds, this may alter the sampled VOC profile. The volume of water consumed prior to breath sampling should be considered as the water volume may affect gastrointestinal motility, that could displace VOC concentrations in breath (Urita *et al*. 2002). Further research is required to ascertain the fasting requirements during breath sampling. There is a lack of human studies exploring the implications of fasting on VOC biomarker discovery. Although, animal studies

indicate a significant variability in VOC concentrations as a result of food intake (Fischer *et al*. 2015).

Lang *et al* (Lang *et al*. 2016) has described the use of membrane surface stress sensors to differentiate between pre- and post-surgical HNSCC patients with PCA. They compared various polymer membranes with variable reactivity to VOCs. The interaction of a volatile compound with the polymer membrane would cause varying levels of swelling, hence affecting the membrane's ability to bend. This bending pattern is then recognised by attached piezo resistors that can output a unique electrical signal based on the VOC profile exposed to the membrane. Their analysis clearly separated samples pre-surgery (n = 3), post-surgery (n = 3) as well as a cohort of healthy controls (n = 4). However, there was large intra-sample variability in HNSCC patients compared to the control group. While the technology described in this study is novel and interesting as a point of care testing device with potential clinical utility, there are major drawbacks from a clinical study design perspective. None of the following common considerations were adequately described: a) fasting status, b) ambient VOC levels, c) collection environment or d) demographic comparison between groups. The sample size is also inadequate to draw further conclusions about this technology based on this study alone.

Leunis *et al* (Leunis *et al*. 2014) has described the use of the DiagNose® e-nose device to differentiate between HNSCC (n = 36) and healthy controls (n = 23). They used a logistic regression model to differentiate between the groups with a sensitivity of 90% and specificity of 80% (AUC = 0.89). They utilised a bootstrapping procedure (undefined) to validate the model that maintained the AUC at 0.85. While the patient groups were matched for smoking, no formal

statistical comparison was reported to compare baseline patient characteristics between the groups.

Hakim *et al* (Hakim *et al*. 2011) demonstrated the use of a nanoscale artificial nose (NA-NOSE) to differentiate between patients with HNSCC (n = 22), lung cancer (n = 25) and healthy controls (n = 40). This technology uses spherical gold nanoparticles with five sensors based on the type of ligand (2-mercaptobenzoazole, 1-butanethiol, hexanethiol, tert-dodecanethiol and 3-methyld-1-butanethiol) (Peng *et al*. 2010). PCA was successful at differentiating the dataset between HNSCC and controls. They also re-analysed the data using a support-vector machine, a non-linear machine learning model that differentiated between study groups with up to 100% sensitivity and 92% specificity. This study also conducted concurrent GC-MS analysis of breath samples to identify specific VOCs that is further described below. Interestingly, NA-NOSE data analysis had superior discriminative power compared to GC-MS. Overall, this is a well-designed study with useful findings for further research. Although, it would be ideal to conduct the breath sampling prior to any biopsy procedure as acute inflammation is likely to interfere with exhaled VOC profiles (van de Kant *et al*. 2012).

#### 4.3.2.2 Gas Chromatography Mass Spectrometry (GC-MS)

Gas chromatography (GC) allows for the separation of individual compounds based on their physical behaviour in mobile phase and stationary phase from a gas mixture (Gebicki *et al*. 2016). It achieves this by carrying the sample gas into a temperature controlled chromatographic column (packed with solid or liquid film) by an inert gas, where unique compounds move at different velocities based on

the interaction with the column. Substances that have a strong interaction with the packed column move slower compared to the rest, hence separating the compounds. As the unique compounds are eluted from the column, the mass spectrometer (MS) can identify these based on type of ions formed following analyte decomposition. While this is a well-established and reproducible technique, a large amount of sample pre-processing is required. SPME is another addon technique to extract the compounds separated by the GC column without introducing solvents that may contaminate volatile gas samples (Spietelun *et al*. 2010).

*Hakim et al* (Hakim *et al*. 2011), used a GC-MS with SPME to identify six volatile compounds (Table 15) that could distinguish HNSCC from benign patients. These compound masses ranged between 43 to 95 m/z, although, ion fragmentation products were not reported. Interestingly, they conducted a concurrent analysis with an E-nose concept that is superior at separating cancer patients from healthy patients with SVM modelling compared to GC-MS. While GC-MS would be considered gold standard for discovering volatile compounds present in breath, its ability to quantify relative amounts based on disease status is limited.

*Garcia et al* (García *et al*. 2013) identified seven unique VOCs that was able to discriminate between patients with cancer (n = 11) and healthy controls (n = 10) (Table 15). They were also able to identify four compounds that differentiated between smokers and non-smokers in healthy controls. However, the study methods did not describe the environmental setting where the samples were collected, nor did they report room air measurements of these compounds of interest. Our experience and that of others has shown that compounds such as ethanol are present in significantly high concentrations in room air from clinical

settings compared to what is produced in human breath (Appendix VI. Measurement of room air volatile organic compounds and statistical analyses). Therefore, it is important to measure and report room air concentrations of VOCs of interest for standardisation of results. Importantly, Garcia et al also demonstrated the need of standardisation of methodology for exhaled breath studies. They experimented with four different type of SPME fibres that significantly affected the volatile compounds that was able to be detected (Garcia *et al.* 2013). They also reported the potential contamination of breath samples by the compounds emitted from sampling bag material (Garcia *et al.* 2013).

*Hartwig et al*, reported a prospective cohort study (n = 10) comparing breath VOCs before and after curative surgery for oral cavity cancer (Hartwig *et al*. 2017). These results were also compared to healthy controls (n = 4). A clear standardised collection protocol was described, where they collected samples from fasted patients who abstained from smoking and collected matched room air samples. A charcoal filter was used for the inhalation phase and the exhaled breath was collected in Mylar® bags. These samples were then transferred to TD tubes prior to analysis using SPME and GC-MS. This study detected 125 VOCs in human breath samples and calculated differences pre- and post-surgery for ten compounds (Table 15) that were not present in room air samples. This was a very well designed, methodologically sound study that establishes the methodological standard required for breath analysis research. Although, the sample size is too small to report any valid statistically significant differences,

*Gruber et al* also demonstrated the feasibility of differentiating patients with HNSCC (n = 22), patients with benign tumours (n = 21) and healthy patients (n = 19) (Gruber *et al*. 2014) based on breath biomarkers. They identified the VOCs

ethanol, 2-propanenitrile and undecane (Table 15) as uniquely discriminant between these groups. This study also instructed participants to breathe through an air filter (undefined) for 3-5 minutes prior to collection of samples to "flush-out" the lung gases of environmental contaminants. They also described an undefined device to separate the dead space gases from the alveolar component, only analysing the alveolar gas mix. Duplicate samples from the participants were collected in Mylar© bags. These gas samples in bags were pumped in to ORBO™ 420 Tenax® TA sorption tubes. Room air samples were directly pumped into Tenax© tubes without using a collection bag. Prior to GC-MS analysis the sorbent material from the Tenax© tubes were again transferred to glass TD tubes. Interestingly, this study also analysed the same breath samples using a nanomaterial-based sensor array composed of spherical gold nanoparticles and carbon nanotubes. These arrays were pre-calibrated to known volatile gases and data analysed using a discriminant function analysis for predictive modelling. The methodological aspects to this study were very well described. However, it would have been ideal to describe the filter details as well as the details of the gas fractionation device for completeness. The room air comparison data may not be valid as the collection method used was not identical to the patient breath. Furthermore, Mylar© bags are known to have a certain VOC profile that would be omitted in this dataset of room air samples as the air is pumped directly into Tenax© tubes.

#### 4.3.2.3 Proton Transfer Reaction Mass Spectrometry (PTR-MS)

PTR-MS takes advantage of real-time chemical ionisation reactions to measure VOCs. $H_3O^+$ reagent ions are generated from water via an emissions source such as hollow cathode or radioactive source (Moser *et al*. 2005). This reagent ions are

injected into a chamber (with low pressure) along with the sample gas. This allows for an exothermic proton transfer reaction (PTR) to occur between the reagent ion and the sample compounds forming further product ions that can be detected using a mass spectrometer. The initial iterations of this technology were limited, mainly because of the inability to detect ion fragmentation and related isomers. The recent advent of time-of-flight (ToF) mass spectrometry coupled to a PTR source allows for distinguishing a higher range of mass resolutions. This allows for distinct identification of compounds with similar integer mass without a chromatographic step (Pleil *et al*. 2019). PTR-MS can also use other reagent ions such as NO+, O2+ and NH4+ to identify distinct isomeric species for more accurate quantification of VOCs.

Schmutzhard *et al* (Schmutzhard *et al*. 2008) is the only study to have used PTR-MS to differentiate between HNSCC and benign patients. They found it is feasible to collect samples Teflon-bags in a clinical setting and analyse using the heated 'breath arm' of PTR-MS. A range of product masses were identified as significantly different between cancer and benign patient groups (Table 15).

#### 4.3.2.4 Selected Ion Flow-Tube Mass Spectrometry (SIFT-MS)

SIFT-MS was initially described in 1976 by Adams and Smith (Smith *et al*. 1989). This technique has become popular in the real-time analysis of VOCs present in the environment as well as in biological specimens including exhaled breath (Spanel *et al*. 2019). SIFT-MS relies on previously described kinetic behaviours of ion-molecule reactions under specific thermal conditions to detect and quantify VOC concentrations. It uses three different precursor ions (originally $H_3O^+$, $O_2^+$ and $NO^+$) that are selected using a mass filter according to their mass-to-charge

ratio, allowing for the differential identification of isomers of the same compound with the same integer mass (Spanel *et al*. 1996) (Figure 36). The analytical process allows for the calculation of concentrations based on a single reagent ion such as ($H_3O^+$) or multiple reagent ions, providing the ability to directly compare between PTR-MS or GC-MS findings. The main advantage of SIFT-MS is its ability to process real-time clinical samples directly via the 'breath-arm' of the equipment as well as its ability to analyse stored samples from sealed bags (Chandran *et al*. 2019). Furthermore, SIFT-MS has been used in for *in vitro* studies, where VOCs in the headspace of containers with cell culture or other biological specimens can be directly analysed, allowing for mechanistic inferences about the biogenesis of VOCs (Gibson *et al*. 2015; Shestivska *et al*. 2018).

Although, SIFT-MS has been described in oesophago-gastric cancer (Kumar *et al*. 2013, Marker *et al.* 2018) and colorectal cancer (Markar *et al*. 2019), only one study to date has described the use of SIFT-MS to detect patients with HNSCC (Chandran *et al*. 2019). In this study, hydrogen cyanide was identified as significantly higher in patients with HNSCC (Chandran *et al*. 2019). Subsequent to this scoping review Chapter 5 of this thesis was published as further evidence for using SIFT-MS to detect HNSCC.

#### 4.3.2.5 Secondary Electrospray Ionisation Mass Spectrometry (SESI-MS)

SESI-MS is a relatively new technology that uses an electrospray solution (chosen based on molecular characteristics of the target compounds) delivered at a 5 $nLs^{-1}$ on to a carrier gas (usually carbon dioxide) via a needle (various shapes and positions, usually silica capillary tip of approximately 40 µm in internal diameter) towards the detector of a mass spectrometer (Zhu *et al*. 2013a). The carrier gas

'carries' the volatile compounds from the sample chamber into the SESI reaction chamber where the volatiles are ionised within the cloud of electrospray (Martínez-Lozano and de la Mora 2007). These ionised volatiles are detected via the mass spectrometer (Bean *et al*. 2011). Unique to this technology is that the ionisation is conducted in atmospheric pressure, meaning there is no need for cumbersome and expensive vacuum pumps, reducing the workspace required for such equipment. Although its application for biomarker discovery is limited, the SESI-MS represents an ideal pathway forward for small, benchtop point of care testing devices that is attractive for clinical translation of breath analysis.

The use of SESI-MS is not reported in HNSCC literature. However, pilot data in breast cancer suggests its ability to differentiate cancer patients from benign patients with sensitivity and specificity above 90% using exhaled breath (Martinez-Lozano Sinues *et al*. 2015).

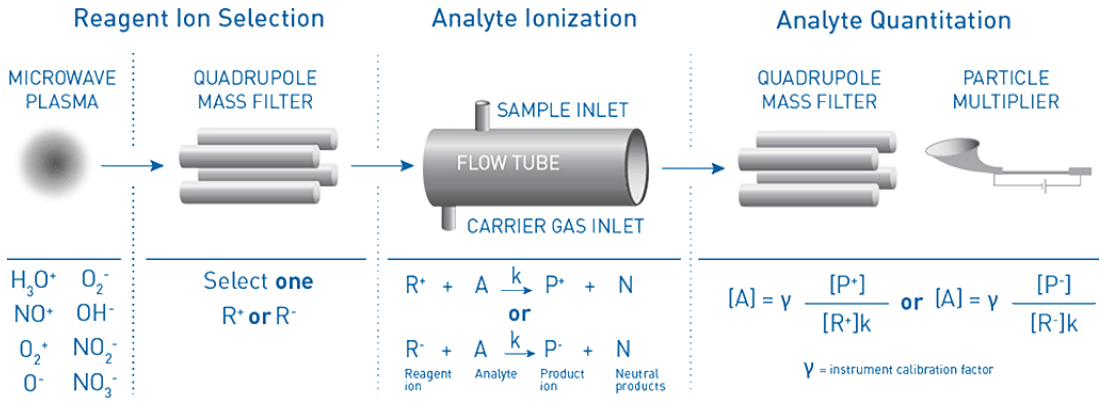



***Figure 36. Schematic of SIFT-MS instrument. Showing the general ion chemistry that occur between reagent ions, the sample gas and resulting products. (Syft Technologies, Christchurch, New Zealand instrument manual)***

## 4.4 Discussion

### 4.4.1 Cost-effectiveness

In general, there is a sparsity of cost-effectiveness studies published in relation to head and neck cancer care (Caulley *et al*. 2019; Patel *et al*. 2016), without direct evaluations of breath analysis as a cost-effective method for screening or diagnosing HNSCC. Several studies reviewed here refer to cost-effectiveness of breath analysis in HNSCC without formal analysis (Chandran *et al*. 2019; Gruber *et al*. 2014; Hakim *et al*. 2011; Hartwig *et al*. 2017; Leunis *et al*. 2014). However, head and neck cancer literature does report the yield of routine diagnostic procedures, their cost-effectiveness (Benninger *et al*. 1993; Naidu *et al*. 2012) and differences between treatment modalities (Morton 1997) in the context of patient reported quality of life outcomes.

In our experience, in the research phase, the establishment costs of a breath analysis service are significant. Cost of a mass spectrometer can range between AU$300000 to AU$1000000, with instrument maintenance costs approximately AU$30000 per annum and personnel cost at a minimum of AU$120000 per annum. Bioinformatics costs are also upwards of AU$5000 per annum. However, once the bioinformatics pipeline has been established there are no significant ongoing costs. The current cost associated with sample collection is approximately AU$50 per sample (including the cost price of the collection bag and technician time required for processing). Although, additional costs may be indicated if clinician or nursing time for collection was incorporated into the overall cost. An online tool was used to convert Australian currency to date adjusted US

dollars as per Shemilt *et al* (Shemilt *et al*. 2010). If a ten-year instrument depreciation (OECD adjusted US$44621/year), maintenance (OECD adjusted US$20916/year) and personnel costs (OECD adjusted US$83666/year) were incorporated to derive an ongoing cost per sample, it would be AU$1145 (OECD adjusted US$798) per sample provided a minimum 100 samples were processed each year.

In Australia, a day procedure such as a panendoscopy, costs the Australian government AU$1753 (OECD adjusted US$1294; Independent-Hospital-Pricing-Authority 2014) and in the United States of America this cost can be as high as US$9024 (Naidu *et al*. 2012). Patients treated for advanced HNSCC including microvascular reconstruction can cost up wards of US$22733 (Dautremont *et al*. 2013). Therefore, a breath test that could detect HNSCC earlier or diagnose HNSCC without the need for a panendoscopy has the potential to significantly reduce associated costs. Even as a supplementary test the cost of a breath test is much lower than currently used diagnostic tests described above. However, a direct cost-effectiveness analysis is required to support this alternative approach.

### 4.4.2 Research Implications

The standardisation of breath analysis is paramount to progress VOC research beyond pilot studies. Meticulous reporting of the breath sample collection method would be the first step (Table 33). It is largely agreed that breath samples should be collected from fasted patients with or without gas fractionation. Additionally, whether the sample is directly analysed or stored in a container depends largely on access to certain technologies and therefore would be difficult to explicitly standardise. However, if research groups provide detailed reporting of their

methods, it would allow for valid replication of experimental results and support a realistic pathway for clinical translation of these breath biomarker models.

Data analytical methods also required further consideration in this field of research. Appropriate reporting of descriptive statistics is as important as the complex machine learning modelling and machine learning algorithms. Ideally a citable data repository with open access to normalised breath data would allow for benchmarking novel experiments and improve reporting accuracy. In particular, HNSCC that has a relatively low incidence of disease compared to other malignancies, would largely benefit from such a repository. Many of the studies reviewed in this chapter did not provide adequate statistical parameters for direct comparison of study technique, or patient groups. Only two studies provided sensitivity and specificity data (Hakim *et al.* 2011, Chandran *et al.* 2019), while only one study provided positive and negative predictive values (Chandran *et al.* 2019). None of the studies reported here had considered appropriate sample sizes, population prevalence to estimate pre-test probability, nor balancing the comparison groups to artificially set the pre-test probability to 0.5. Due to the low population prevalence of HNSCC, a significantly large sample size (n > 4000) would be required to emulate the real pre-test probability with a statistical power > 80% (Bujang *et al.* 2016).

### 4.4.3 Summary

Early detection is a priority in the overall management of HNSCC (). In the era of rising health care costs, cost-effectiveness of early detection needs to be carefully considered. Based on the available literature, breath analysis as a tool for early detection of cancer is still at its early stages with no clear evidence for cost-

effectiveness. However, recent studies using breath analysis to detect upper gastrointestinal cancers have shown promising results (Markar *et al*. 2018). Therefore, well-designed studies in HNSCC using this technology are required to demonstrate clinical utility.

In summary, standardisation of exhaled breath collection, analysis and reporting would improve the overall quality of published research and allow the discovery of markers that are clinically translatable. There is clear evidence of exhaled volatile compound's potential for differentiating a patient with cancer from one without. However, small patient cohort studies published to date limit the translational potential of exhaled breath analysis for the detection of HNSCC. Given HNSCC is relatively rare compared to other solid cancers, it is important to consider standardised data repositories for future pooled analysis. The next chapter describes the analysis of raw data from exhaled breath analysis with a moderate sample size with good predictive values.

# 5 Exhaled breath as a predictor of head and neck squamous cell carcinoma

## 5.1 Introduction

The human body emits a variety of VOCs with previous studies indicating the potential to discriminate patients with cancer from healthy controls (Hanna *et al*. 2019). The objective of this multicentre study was to establish the diagnostic accuracy of using specific raw mass to charge spectra for detecting patients with HNSCC.

## 5.2 Methods

### 5.2.1 Ethics

Initial ethics approval was obtained from the Southern Adelaide Local Health Network Human Ethics Committee (SALHN HREC) on the 29th of April 2016. (HREC reference number: HREC/16/SAC/70 66.16 and Site-Specific Application (SSA) reference number: SSA/16/SAC/76). This was subsequently amended to include patients from the Flinders Medical Centre and the Royal Adelaide Hospital (approved on 31$^{st}$ of March 2017, reference number: OFR # 66.16 HREC/16/SAC/70).

### 5.2.2 Patients

Patients referred to a head and neck cancer clinic at Flinders Medical Centre and Royal Adelaide Hospital were assessed by an experienced Otolaryngologist for clinical suspicion of HNSCC were recruited for this study. As part of their routine assessment, these patients were listed for an elective panendoscopy procedure to thoroughly examine the upper aerodigestive tract under general anaesthesia. A tissue biopsy was taken from the suspected area for cancer.

All patients with HNSCC had a histologically confirmed diagnosis of and SCC arising from an upper aerodigestive subsite. The control group consisted of healthy patients with either a benign tissue biopsy or patients with no lesions identified during the panendoscopy.

Exclusion criteria included patients with a) mucosal high-grade dysplasia, b) concurrent malignancy, c) a history of cancer, d) cutaneous malignancies, e) paediatric patients (age < 18 years) and f) patients with an active inflammatory condition (or infection with or without antibiotic use). Patients with SCC in neck lymph node with no identifiable primary mucosal tumour (unknown primary) were also excluded from further analysis.

Patients with oropharyngeal primary lesions had immunohistochemical characterisation of p16 marker and all patients with cancer were staged using the 8$^{th}$ Edition of the AJCC Cancer Staging Manual (Lydiatt *et al*. 2017). Patient demographics, comorbidities, risk factors and medication intake were recorded and categorised based on American Society of Anaesthesiologists (ASA) grade (Table 35)

***Table 35. Comparison of demographics, comorbidities, and medications between patient groups***

| Variable | | Control Group No. (%) | Cancer Group No. (%) | P Value |
|---|---|---|---|---|
| Count | | 50 (50) | 50 (50) | |
| Age - Years, median (Range) | | 56 (31-86) | 58 (33-88) | 0.104 |
| Sex | | | | |
| | Female | 25 (50) | 7 (14) | <0.001* |
| | Male | 25 (50) | 43 (86) | |
| Smoking | | | | |
| | Never smoked | 19 (38) | 11 (22) | 0.311 |
| | Ex-smoker | 15 (30) | 22 (44) | |
| | Current smoker | 16 (32) | 17 (34) | |
| Smoking Pack Years (Range) | | 23.6 (0-104) | 24.0 (0-198) | |
| Alcohol | | | | |
| | No alcohol intake | 18 (36) | 11 (22) | 0.264 [a] |
| | Intake ≤ 2 days per week | 19 (38) | 17 (34) | |
| | Intake ≥ 3 days per week | 13 (26) | 22 (44) | |
| BMI | | 28.5 (17.9-45.4) | 25.6 (17.3-38.2) | 0.002 [b] * |
| BMI Class [d] | | | | |
| | Under Weight | 1 (2) | 3 (6) | |
| | Normal Weight | 11 (22) | 20 (40) | |
| | Pre-Obesity | 18 (36) | 17 (34) | 0.118 [a] |
| | Obesity: Class I | 14 (28) | 8 (16) | |
| | Class II | 2 (4) | 2 (4) | |
| | Class III | 4 (8) | 0 | |
| ASA grade | | | | |
| | 1 | 34 (68) | 33 (66) | 0.328 [a] |
| | 2 | 12 (24) | 17 (34) | |
| | 3 | 4 (8) | 0 | |
| Comorbidities | | | | |
| | Ischemic Heart Disease | 3 (6) | 5 (1) | |
| | Chronic Respiratory Disease | 7 (14) | 8 (16) | 0.623 [a] |
| | Chronic Renal Disease | 0 | 0 | |
| | Chronic Liver Disease | 1 (2) | 1 (2) | |
| | Diabetes | 6 (12) | 6 (12) | |
| Medications [c] | | | | |
| | Anti-reflux | 20 (40) | 9 (18) | |
| | Anti-hypertensives | 12 (24) | 18 (36) | 0.126 [a] |
| | Antibiotics | 0 | 0 | |
| | Anti-platelet/coagulant | 2 (4) | 5 (1) | |

a, Chi-square test. b, Mann-Whitney-U test. * statistical significance ($p < 0.05$). c, medications taken in the last seven days are reported. d, BMI class reported as per World Health Organization classification

### 5.2.3 Breath Collection

Patients who were planned for elective panendoscopy were contacted the day before sample collection. To reduce the effect of any confounding contamination

to the breath sample, they were instructed not to wear perfume or deodorant sprays, not to brush their teeth, not to smoke and not to use any mouth wash on the morning of sample collection. Patients adhered to the instructions, except for 73% of patients who reported brushing their teeth prior to sample collection. The median time between tooth brushing and breath collection was 3 hours (Range 2 – 6 hours). Patients fasted overnight for a minimum of six hours as per anaesthetic requirements for the planned surgical procedures. All exhaled breath samples were collected prospectively in the peri-operative setting prior to any surgical procedures or anaesthetic administration.

On arrival to hospital, patients were asked to rest in a bed (in a four to six bay ward) for at least fifteen minutes prior to sample collection. During the rest period, a structured patient history was collected using a questionnaire (Appendix V. Patient dependent factors associated with breath, blood, and microbiome markers.) prior to sample collection. This included smoking status, smoking pack years, alcohol intake (days per week), comorbidities, height, weight, fasting time, tooth brushing, mouth wash use and chewing gum use. They were then asked to take a deep breath in through the nose, followed by a single continuous forced exhalation through the mouth (while pinching their nostrils closed) into a sealed three litre FlexFoil® PLUS® bag (SKC Ltd, Pennsylvania, USA) resulting in a mixed alveolar gas sample (mixture of alveolar air and respiratory dead space air). While the capacity of samples bags were 3 litres, the exhaled breath volume varied among patients (not measured). A room air sample was collected into another FlexFoil® PLUS® bag immediately after the breath sample collection for comparison and quality assurance. Samples were then transported in a temperature stable bag (Esky®, Coleman Brands, New South Wales, Australia) maintained at 37°C (Deltaphase® isothermal pads, Braintree Scientific Inc,

Massachusetts, USA) to the laboratory for immediate analysis. All FlexFoil® PLUS® bags were preconditioned by flushing with 99% nitrogen gas five times prior to breath sample collection.

### 5.2.4 Breath Analysis using the SIFT-MS Instrument

On arrival at the lab, samples were stored in a dedicated 37ºC incubator and were analysed within three hours of collection by selected ion flow-tube mass spectrometry (SIFT-MS; Voice 200®, Syft® Technologies, Christchurch, New Zealand). The SIFT-MS was calibrated twice daily (morning and afternoon) using a standard gas mix (1,2,3,4-Tetrafluoro benzene, benzene, ethylene, isobutane, octafluorotoluene, p xylene, perfluorobenzene, toluene and nitrogen – Scotty® specialty gases, Pennsylvania, USA).

For sample analysis, the septum of the SKC® FlexFoil® PLUS® bag was pierced with a non-coring needle and attached directly to the SIFT-MS 'breath-head' (Voice 200®, Syft® Technologies, Christchurch, New Zealand). All samples (including room air) were scanned using the full mass scan mode (15m/z to 250m/z) with three reagent ions ($H_3O^+$, $NO^+$, $O_2^+$) for 30 cycles. Approximately 1 litre of the gas sample was used for SIFT-MS analysis; however, SIFT-MS analysis sensitivity is not volume dependent. The instrument corrected intensities for all mass to charge ratios were extracted from the instrument generated files. 15mL of the exhaled breath samples were also analysed by Isotope Ratio Mass-Spectrometry (Sercon®, Crewe, United Kingdom) for carbon dioxide quantification. Samples found to have less than 3% carbon dioxide in breath were excluded from further analysis, as this carbon dioxide level would be less than expected from a mixed alveolar breath sample.

### 5.2.5 Statistical and data analysis

The sample size required for statistical modelling was calculated based on the prevalence of disease and fixed type I error with at least 80% power. The point prevalence of HNSCC in our dataset was artificially set to 50% by balancing the cancer and control groups. Sample size calculation tables published by *Bujang et al* were used to estimate a minimum total sample size of 62, with a minimum of 31 patients with HNSCC, to test for 70% sensitivity and 70% specificity with 80% power (Bujang and Adnan 2016).

All statistical analyses were performed using IBM® SPSS® (Version 26. Chicago, United States, with R© version 3.5 - extensions). Normality of data distribution was tested using Kolmogorov-Smirnov test. Fixed patient factors were compared between cancer and control groups using Mann-Whitney-U (MWU) test for continuous fixed patient factors (age, smoking pack years and body mass index [BMI]) and Chi-squared test for categorical fixed patient factors (gender, smoking status, alcohol intake, BMI class, ASA, comorbidities, medications, and tooth brushing status. The associations between categorical patient factors and SIFT-MS derived variables were assessed using the Mann Whitney U test (with Bonferroni correction). Spearman's rho test was used to determine associations between continuous patient factors with SIFT-MS derived variables. Statistically significant association was considered if $p < 0.05$.

A zero-value analysis was conducted, where variables with more than 70% zero value readings were excluded as it was difficult to ascertain if the zero reading was due to limit of detection of the instrument or it was a true zero reading for that

sample. Masses known to conflict with reagent ions (reported by the instrument manufacturer) were also removed from further analysis. Variables that indicated a significant association to fixed patient factors were also excluded. Following the data optimisation process, 440 variables were available for predictive analysis. The dataset was then randomly split into 70% (n=76) training and 30% (n=24) testing datasets using the SPSS "Select Cases" function (Appendix III) (Figure 37).

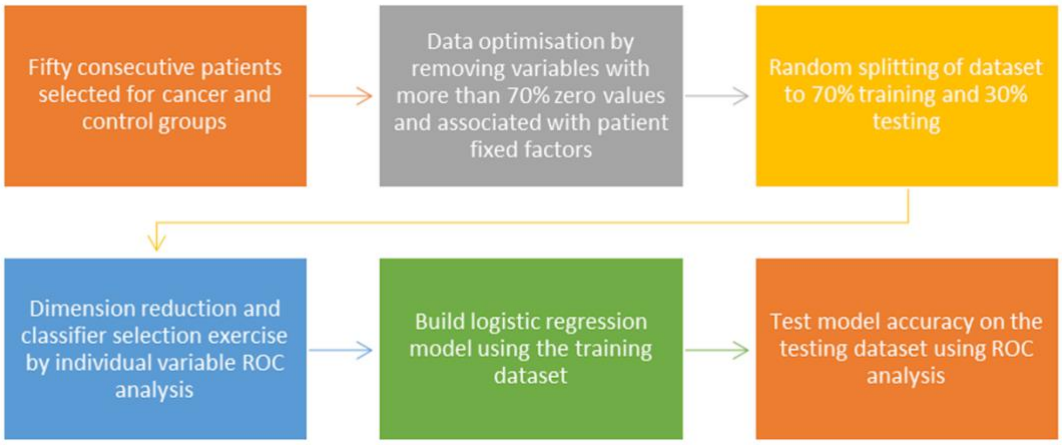



***Figure 37. Data analysis pipeline***

***Table 36. Variables selected and ranked based on ROC-AUC per reagent***

| Rank | Variable Name | | |
|---|---|---|---|
| | R19 ($H_3O^+$) | R30 ($NO^+$) | R32 ($O2^+$) |
| 1 | R19P49 | R30P147 | R32P135 |
| 2 | R19P125 | R30P143 | R32P109 |
| 3 | R19P30 | R30P187 | R32P82 |
| 4 | R19P189 | R30P211 | R32P138 |
| 5 | R19P59 | R30P240 | R32P31 |
| 6 | R19P18 | R30P64 | R32P47 |
| 7 | R19P81 | R30P188 | R32P184 |
| 8 | R19P48 | R30P80 | R32P40 |
| 9 | R19P133 | R30P95 | R32P103 |
| 10 | R19P77 | R30P159 | R32P95 |

***Table 37. Variables selected based on overall ROC-AUC.***

| Rank | Variable Name |
|---|---|
| 1 | R30P147 |
| 2 | R19P49 |
| 3 | R30P143 |
| 4 | R30P187 |
| 5 | R19P125 |
| 6 | R30P211 |
| 7 | R30P240 |
| 8 | R32P135 |
| 9 | R19P30 |
| 10 | R32P109 |
| 11 | R30P64 |
| 12 | R30P188 |
| 13 | R30P80 |
| 14 | R32P82 |
| 15 | R30P95 |
| 16 | R30P159 |
| 17 | R32P138 |
| 18 | R32P31 |

## 5.3 Results

A total of 181 patients were recruited. 74 patients were excluded due to previous history of cancer (n=23), high grade dysplasia (n=7), unknown primary HNSCC (n=3), active infection or poor carbon dioxide concentration in the breath sample (n=41). Consecutive control (n=50) and cancer (n=50) patients were selected from the remaining 107 patients for further analysis to maintain pre-test probability at 0.5 (50%-artificial point prevalence). Breath sampling and processing did not result in any adverse events. Total time to collect a patient breath sample after instruction and demonstration was approximately one minute.

Control group patients had a normal upper aerodigestive tract examination, or a benign biopsy result following histological analysis. The HNSCC group were predominantly male ($p = 0.001$) with a lower BMI than the control group ($p = 0.002$). There was no significant difference in age, smoking status, smoking pack years and alcohol intake between groups ($p > 0.05$). Most patients in the control group had an ASA grade of 2 or lower with no incapacitating comorbidities. However, four control patients were classified as ASA 3 due to morbid obesity. There was a significant difference between tooth brushing status between cancer and control patient groups ($p = 0.015$). Cancer-specific factors including head and neck subsite and tumour stage are reported in Table 2. HNSCC patients largely comprised of oropharyngeal SCC (46%), followed by oral cavity SCC (32%), and laryngeal SCC (30%). Nineteen (83%) of the oropharyngeal SCCs were p16 positive. Regional node metastasis was present in 58% of patients.

HNSCC patients in this cohort largely comprised of oropharyngeal SCC (46%), then oral cavity SCC (32%), and laryngeal SCC (30%). Nineteen (83%) of the oropharyngeal SCCs were p16 positive. The Majority (68%) of the HNSCC patients were early T-stage (T1 or T2) while 58% of these patients had regional node metastasis (Table 38).

***Table 38. Description of cancer specific factors***

| Cancer Related Factor | Patients, No. (%) |
|---|---|
| Head and neck subsite | |
| Oral Cavity | 16 (32) |
| Oropharynx | 23 (46) |
| p16 positive | 19 |
| p16 negative | 4 |
| Larynx | 15 (30) |
| Clinical T stage | |
| 1 | 16 (32) |
| 2 | 17 (34) |
| 3 | 7 (14) |
| 4 | 10 (20) |
| Neck node status | |
| Negative | 21 (42) |
| Positive | 29 (58) |
| Overall prognostic stage | |
| I | 18 (36) |
| II | 11 (22) |
| III | 9 (18) |
| IV | 12 (24) |

## 5.3.1 Predictive models

### 5.3.1.1 Logistic regression

Binomial logistic regression models, generated on mass spectrometry derived variables from exhaled breath, were able to classify patients to cancer vs. control groups. The accuracy of this classification was maintained when tested on the independent cohort with an AUC up to 0.9 with fifteen variables (Supplementary Table 5). However, maintaining the model sensitivity and specificity and reducing over-fitting of the data resulted in an optimal model with two variables (Model 1) with an overall classification accuracy of 83% (AUC of 0.814 95% CI 0.611-1.000),

sensitivity of 80% and specificity of 86% (Figure 38). This model used the variables R30P147 and R19P49. These variables belong to two distinct reagent ions and adding a third variable (Model 2) from another reagent ion (R32P135) marginally improves the overall model (AUC 0.821 95% CI 0.625-1.000, Figure 39).

***Table 39. Predication based on gender and BMI.***

| | Train | n=76 | | | Test | n=24 | | |
|---|---|---|---|---|---|---|---|---|
| VAR | SEN | SPC | AUC | 95% CI | SEN | SPC | AUC | 95% CI |
| G | 85 | 44.4 | 0.647 | 0.521-0.773 | 90 | 64.3 | 0.771 | 0.577-0.966 |
| BMI | 41 | 80.6 | 0.654 | 0.530-0.777 | 60 | 64.3 | 0.693 | 0.482-0.904 |
| G + BMI | 87.2 | 47.2 | 0.725 | 0.610-0.841 | 90 | 57.1 | 0.764 | 0.575-0.954 |

G, gender. BMI, body mass index. SEN, sensitivity. SPC, specificity. VAR, variable. AUC, area under the curve.

***Table 40. Linear regression between included model variables and BMI on testing cohort***

| Variables | Adjusted R2 | F statistic | P value |
|---|---|---|---|
| R32P135 + R30P147+R19P49 | 0.041 | 0.696 | 0.565 |
| R32P135 + R19P49 | 0.014 | 0.838 | 0.446 |
| R30P147 + R19P49 | 0.016 | 0.82 | 0.454 |
| R30P147 + R32P135 | 0.03 | 0.67 | 0.523 |
| R19P49 | 0.016 | 0.643 | 0.431 |
| R30P147 | 0.01 | 1.242 | 0.277 |
| R32P135 | 0.024 | 0.455 | 0.507 |

***Table 41. Variables selected for modelling based on AUC ranking per reagent ion.***

| | Train | n=76 | | | | Test | n=24 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI |
| 30 | 100.00% | 100.00% | 1 | 1 | 1 | 40.00% | 64.30% | 0.529 | 0.284 | 0.773 |
| 27 | 100.00% | 100.00% | 1 | 1 | 1 | 60.00% | 64.30% | 0.579 | 0.332 | 0.826 |
| 24 | 95.00% | 88.90% | 0.978 | 0.953 | 1 | 70.00% | 71.40% | 0.621 | 0.387 | 0.856 |
| 21 | 82.50% | 80.60% | 0.929 | 0.876 | 0.983 | 70.00% | 50.00% | 0.493 | 0.243 | 0.743 |
| 18 | 80.00% | 69.40% | 0.874 | 0.799 | 0.949 | 80.00% | 57.10% | 0.814 | 0.625 | 1 |
| 15 | 75.00% | 72.20% | 0.862 | 0.783 | 0.94 | 90.00% | 85.70% | 0.9 | 0.748 | 1 |
| 12 | 67.50% | 69.40% | 0.815 | 0.722 | 0.908 | 70.00% | 78.60% | 0.779 | 0.587 | 0.97 |
| 9 | 75.00% | 77.80% | 0.806 | 0.705 | 0.908 | 80.00% | 71.40% | 0.786 | 0.585 | 0.987 |
| 6 | 75.00% | 69.40% | 0.774 | 0.668 | 0.881 | 40.00% | 64.30% | 0.586 | 0.342 | 0.83 |
| 3 | 75.00% | 66.70% | 0.754 | 0.645 | 0.863 | 80.00% | 85.70% | 0.821 | 0.625 | 1 |

***Table 42. Variables selected for modelling based on overall AUC ranking.***

| | Train | n=76 | | | | Test | n=24 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI |
| 10 | 72.50% | 75.00% | 0.817 | 0.723 | 0.912 | 80.00% | 71.40% | 0.707 | 0.482 | 0.932 |
| 9 | 75.00% | 72.20% | 0.808 | 0.711 | 0.904 | 80.00% | 71.40% | 0.729 | 0.507 | 0.95 |
| 8 | 70.00% | 63.90% | 0.795 | 0.697 | 0.893 | 70.00% | 64.30% | 0.671 | 0.441 | 0.902 |
| 7 | 72.50% | 63.90% | 0.797 | 0.699 | 0.894 | 60.00% | 64.30% | 0.657 | 0.42 | 0.894 |
| 6 | 77.50% | 63.90% | 0.79 | 0.687 | 0.892 | 50.00% | 71.40% | 0.629 | 0.384 | 0.874 |
| 5 | 72.50% | 69.40% | 0.785 | 0.683 | 0.888 | 50.00% | 71.40% | 0.664 | 0.422 | 0.907 |
| 4 | 70.00% | 66.70% | 0.752 | 0.643 | 0.861 | 80.00% | 64.30% | 0.857 | 0.688 | 1 |
| 3 | 72.50% | 66.70% | 0.737 | 0.624 | 0.851 | 70.00% | 78.60% | 0.779 | 0.571 | 0.986 |
| 2 | 72.50% | 66.70% | 0.747 | 0.633 | 0.86 | 80.00% | 85.70% | 0.814 | 0.611 | 1 |
| 1 | 80.00% | 36.10% | 0.702 | 0.583 | 0.822 | 80.00% | 71.40% | 0.725 | 0.485 | 0.965 |

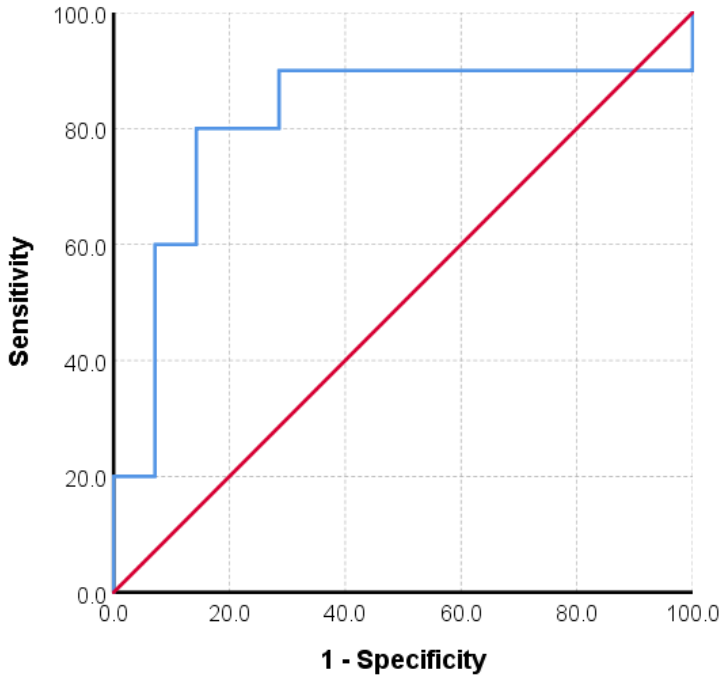



***Figure 38. Model 1 - Receiver operating curve of the testing cohort based on logistic regression model using the two variables from distinct reagent ions.***

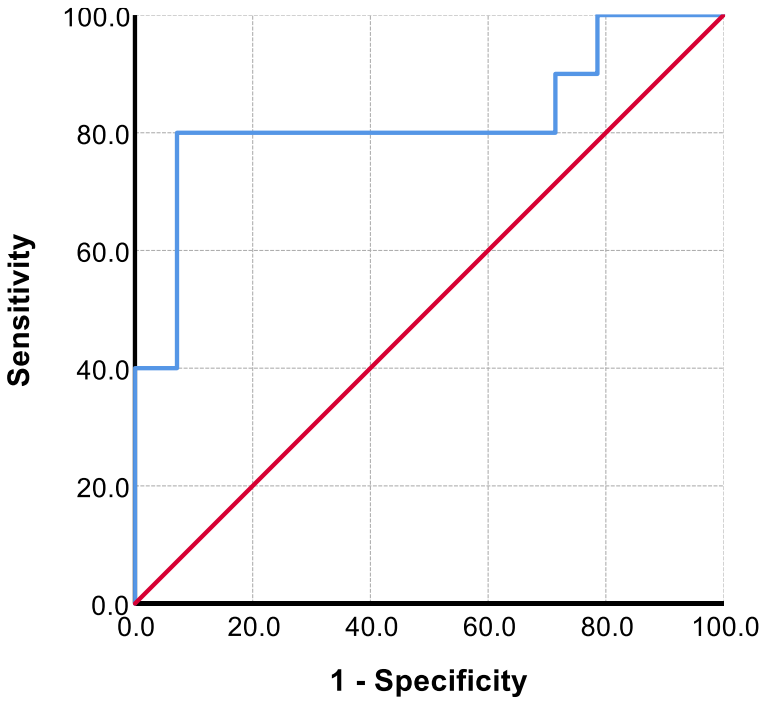



***Figure 39. Model 2 - Receiver operating curve of the testing cohort based on logistic regression model with one variable from each reagent ion (three variables in total).***

Gender specific classification analysis of Model 1 using the testing cohort indicated a 79% and 90% overall accuracy for the male and female cohorts, respectively. The same variables (R30P147 and R19P49) were used to build a model to predict gender in the training and testing cohorts with poor overall accuracy of 58% in the test cohort (AUC 0.721 95% CI 0.499-0.944). There was no linear relationship (based on linear regression) between BMI and the variables used in Model 1 or Model 2 (Appendix III). The addition of gender and BMI to Model 1 improved the AUC (0.979 95%CI 0.929-1.000) and sensitivity (100%) while reducing the specificity (71.4%).

***Table 43. Comparison of significantly different patient factors between training and testing patient cohorts***

| Variable | | Training | | P Value | Testing | | P Value |
|---|---|---|---|---|---|---|---|
| | | Control | Cancer | | Control | Cancer | |
| n | | 36 | 40 | | 14 | 10 | |
| Sex | | | | | | | |
| | Female | 16 | 6 | 0.005 a * | 9 | 1 | 0.008 a * |
| | Male | 20 | 34 | | 5 | 9 | |
| BMI | | 30.22 | 26.51 | 0.005 b * | 27.5 | 23.44 | 0.026 b * |
| Toothbrushing | | | | | | | |
| | Yes | 29 | 26 | 0.130 a | 1 | 5 | 0.022 a * |
| | No | 7 | 14 | | 12 | 5 | |

BMI, body mass index. n, sample size. a, Chi-square test. b, Mann-Whitney-U test. *, statistical significance ($p < 0.05$).

The independently curated Syft® compound library was interrogated using LabSyft® (version 1.5.1, Syft®, Christchurch, New Zealand) software to identify the VOCs contributing to the variables (product mass ions) included in modelling described above. VOCs with poor branching ratios (less than 0.2), poor reaction rates (less than 1.0 $e^{-11}$), and isomers with the same primary chemistry were

excluded. Two of the ten compounds that corresponded to the underlying chemistry of the SIFT-MS derived variables have been previously described as predictive of lung (Wehinger *et al*. 2007) and oral cavity cancer (Szabó *et al*. 2015).

***Table 44. Volatiles organic compounds corresponding to product mass ions of interest.***

| Variable | VOCs | Product Formulae | Chem | Carcinogenicity/Toxicity |
|---|---|---|---|---|
| R19P49 Reagent ion H3O+ Product m/z 49 | Formaldehyde* | $H_2CO.H^+.H_2O$ | S | Known Carcinogen |
| | Methyl mercaptan* | $CH_4S.H^+$ | P | Acute toxicity |
| | Sevoflurane | $CH_2FO^+$ | P | Irritant |
| R30P147 Reagent ion NO+ Product m/z 147 | Benzyl cyanide | $C_8H_7N.NO^+$ | P | Acute toxicity |
| | Coumarin | $C_9H_6O_2^+.H^+$ | S | Suspected Carcinogen |
| | Cuminal | $C_{10}H_{11}O^+$ | P | Irritant |
| R32P135 Reagent ion O2+ Product m/z 135 | 2-aminoacetophenone | $C_8H_9NO^+$ | P | Non-toxic |
| | Amphetamine | $C_9H_{13}N^+$ | P | Acute toxicity |
| | Benzothiazole | $C_7H_5NS^+$ | P | Acute toxicity |
| | Carvacrol | $C_9H_{11}O^+$ | P | Irritant |

P, primary. S, secondary. R, reagent ion. P, product ion. Indicated carcinogenicity and toxicity is based on details from pubchem.ncbi.nlm.nih.gov. Previously reported predictive nature of cancer indicated with * and relevant citation.

## 5.4 Discussion

This study describes a novel breath test for the detection of HNSCC, with a sensitivity of 80% and specificity of 86% in the optimal model. The robustness of this breath test was demonstrated by the consistently high performance noted in the independent testing cohort. The predictive ability of this test is significantly

higher than that reported for clinical symptoms and examination alone (Tikka *et al*. 2016), suggesting that raw mass spectra breath analysis for HNSCC has the potential to improve current clinical practice. Additionally, this study presents a novel method of analysing SIFT-MS derived raw data.

Patients with suspected head and neck cancer often present to a primary care setting with non-specific upper aerodigestive symptoms; however, complete examination by the general practitioner is restricted by technical limitations. Whilst the oral cavity, part of the oropharynx and the external neck can be examined thoroughly by general practitioners, the nasopharynx, base of tongue and the larynx can only be examined in specialist clinics with the use of flexible endoscopes. In addition, early stage HNSCC are unlikely to show overt macroscopic clinical signs and may not be visible in cross sectional imaging. Therefore, when a patient presents to a primary care physician with upper aerodigestive symptoms without a clear risk profile associated with HNSCC, instead of considering antibiotics or deciding to watch and wait, an exhaled breath test might be considered. Given the high sensitivity and specificity of our breath test, the hypothesis is that a positive result would trigger an urgent specialist referral for further evaluation. A negative result could support a short period of observation, or medical treatment with further investigations instigated should symptoms persist. The intention of such a breath test would not be to replace clinical expertise or diagnostic procedures such as a panendoscopy and biopsy under general anaesthesia. Instead, a breath test might provide a triaging system to risk stratify patients and potentially alleviate some pressure from healthcare resources (Park and Look 2019).

An acknowledged limitation of our study was that gender and BMI were significantly different between comparison groups. From an epidemiological perspective, HNSCC predominantly affects the male population (Benchetrit *et al*. 2019) and weight loss is a well-documented hallmark sign of malignancy and a marker of poor prognosis (Langius *et al*. 2013). Univariate statistical comparisons were used prior to classifier selection to ensure the variables used in model generation were not associated with gender or BMI. The variables selected for these models did not classify patients based on gender and BMI, indicating the chosen classifiers are not solely dependent on these patient factors (Appendix III). Furthermore, exhaled breath compounds reported to be gender (Das *et al*. 2014) and BMI (Alkhouri *et al*. 2015) specific were not associated with the SIFT-MS derived variables included in our final model. Future studies are indicated using larger cohorts to validate models based on gender or BMI specific groups.

There was a significant disparity with regards to tooth brushing status prior to sampling between comparison groups. In this study the association between tooth brushing status and classifiers were addressed using univariate statistical comparison and excluded prior to model generation. Vadhwana *et al* have reported the effects of various oral cleansing methods on sampled breath VOCs that were sampled immediately after oral cleansing, finding differences based on the type of cleanser, compound of interest and the time between oral cleansing and breath sampling (Vadhwana *et al*. 2020). Exhaled menthone, decane, dodecane and p-cresol concentrations increased while ammonia, butanoic acid and dimethyl disulphide concentrations decreased, with no reported changes to aldehydes, alcohol, and some hydrocarbon levels after toothbrushing (Vadhwana *et al*. 2020). In our study, participants that had brushed their teeth had done so at least three hours prior to breath sampling, minimising any potential effects on the

breath VOCs reported. Furthermore, there were no associations between the compounds prone to variate based on toothbrushing and the variables selected for model generation in this study. Nonetheless, future validation studies in larger cohorts should seek to determine an optimum and consistent oral cleansing method immediately prior to breath collection.

SIFT-MS allows for full mass scan of a sample as well as calculating concentrations of individual VOCs and has been comprehensively described by Spanel *et al* (Spanel *et al*. 1996). By utilizing the SIFT-MS product-ion mass-spectra, the complete breath profile of the participant can be interrogated, removing any bias associated with VOC concentration calculations (Spanel *et al*. 1996). This method potentially allows for cross-platform comparison of datasets, facilitating capacity for multi-centre studies. Schumtzhard *et al* have reported the analysis of raw mass spectra of breath samples for HNSCC using proton transfer reaction mass spectrometry (Schmutzhard *et al*. 2008), comparing the $H_3O^+$ product masses between HNSCC (n=22) and control patients (benign, high-risk, and post-treatment). They found 42 masses that were significantly different between the comparison groups but did not report sensitivity or specificity for the direct comparison. However, none of these raw mass intensities (or the corresponding VOCs identified using the Syft® library) were reported in the recent meta-analysis by Hanna *et al* as predictive of cancer (Hanna *et al*. 2019). Formaldehyde and Methyl mercaptan were identified as potential VOCs related to cancer with a product mass of 49 based on reagent $H_3O^+$. Exhaled formaldehyde has been reported to be predictive of primary lung cancer (Wehinger *et al*. 2007) and it is thought to be from methylotrophic bacteria present in the aerodigestive tract (Anesti *et al*. 2005; Kovaleva *et al*. 2019). Microbial dysbiosis has been reported in cancer, which could contribute to a distinctive VOC profile. Similarly,

methyl mercaptan is a volatile sulphur compound detected in the human oral cavity and is widely implicated in periodontal disease and halitosis (Szabó *et al*. 2015). Clinically, patients with oral cavity malignancies present with concurrent halitosis. While poor oral hygiene or halitosis is not a direct risk factor for cancer, objective measurement of halitosis with volatile sulphur compounds such as methyl mercaptan may assist in risk stratifying patients with poor oral hygiene.

Four prior studies using breath analysis to detect HNSCC have reported sensitivities ranging from 77% to 100% and specificities ranging from 76% to 92% (Chandran *et al*. 2019; Gruber *et al*. 2014; Hakim *et al*. 2011; Leunis *et al*. 2014). The total sample size in these studies ranged between 41 to 62 with unequal groups. While these results are promising, the statistical power based on sample size alone is low. In comparison, our study had a larger sample size, with 50 cancer and 50 control patients, providing 80% statistical power for determining sensitivity and specificity above 70%. Much larger samples would be required for independent validation (Bujang and Adnan 2016). However, only 30% of HNSCC patients in our current study were T1 and T2 stage with no lymph node metastasis, thus limiting statistical power to determine the accuracy of this model for identifying early stage HNSCC. Hence, prior to recommending this breath test for routine clinical use, a large-scale diagnostic accuracy study should be performed in the primary care setting. Given the relatively low incidence of HNSCC, to adequately power (>80%) a primary care study, more than 5000 participants would be required. The large-scale multi-site trial (PAN Study, ClinicalTrials.gov Identifier: NCT03756597) to investigate breath analysis for the detection of early cancer in patients across the UK represents an example of the study design necessary to validate this new technology (Abderrahman 2019b).

Clinically, the most convenient use of a breath test would be as a point of care device, providing an immediate result for clinical decision making. Advanced smart gas sensors (electronic noses) are routinely used in defence and environmental protection agencies for environmental toxic gas detection (Feng *et al*. 2019). Various nanomaterial-based sensors have also been developed for specific VOC detection properties for development of point of care devices (Broza *et al*. 2018). These arrays of very sensitive gas detectors can profile complex gas samples, such as human breath using computerized pattern recognition to discriminate cancer from healthy patients (Hakim *et al*. 2011; Leunis *et al*. 2014). Therefore, the product ion masses identified in this study could potentially be used to calibrate these devices to specifically detect HNSCC, thereby presenting a pathway for clinical translation.

In summary, this chapter provides further evidence for the feasibility of using breath analysis as a method of detecting HNSCC. This non-invasive technology has the potential for early detection of this disease and as a result the ability to improve patient outcomes.

# 6 Influence of human microbiome on breath volatile organic compounds in the context of head and neck squamous cell carcinoma

## 6.1 Introduction

There is an abundance of microbiota including viruses, bacteria, protozoa, archaea, and fungi that cohabitate the human body and alterations to this microbiome has been implicated as a risk factor for cancer development (Nauts 1989). In head and neck cancer, virally mediated mechanisms of oncogenesis in the context of HPV (Hebner and Laimins 2006) and EBV (Tay *et al*. 2020) have been identified. However, the association between other microorganisms such as bacterial species and human oncogenesis is unclear (Astudillo-de la Vega *et al*. 2019). Microbe dependent changes to the tissue microenvironment (Chang and Parsonnet 2010), microbe-host dynamic immune response (Elinav *et al*. 2013), microbe dependent epigenetic changes (Hattori and Ushijima 2016) and tumour dependent changes to the resident microbiome (Bhatt *et al*. 2017) have been suggested as potential mechanisms; however, supporting evidence is still lacking.

Preferential colonisation of microbiota on specific tissue surfaces has been previously described (Mager 2006). Therefore, characterisation of site dependent microbiome changes is important to understand their contribution to oncogenesis. In the context of HNSCC, oral cavity SCC is the most commonly affected subsite in the world and its microbiome has been widely studied, partly due to the ease of access to the oral cavity for sampling (Li *et al*. 2020). However, this analysis of oral cavity microbiome has not been correlated to exhaled breath compounds. Specific microbiota present in the upper aerodigestive mucosal surfaces (Elmassry and Piechulla 2020) as well as bacterial fermentation in the lower digestive tract (Arasaradnam *et al*. 2016) are thought to contribute to the exhaled breath VOC profile.

### 6.1.1 Oral cavity microbiome and dysbiosis

The oral cavity is a unique anatomical subsite that is exposed to numerous bacteria through food and fluid intake, as well as from the air inspired into the oral cavity and expired from the lungs. Its continuous relationship with the oropharynx, nasopharynx, laryngopharynx, and the upper oesophagus contaminates the transient microbiota from each of those anatomical subsites via air, mucous and saliva flow. However, there are well described resident microbiota that are subsite specific in the oral cavity.

Up to 90% of the human microbiome is uncultivable by traditional methods. (Dewhirst *et al*. 2010). The advent of advanced sequencing techniques has meant that culture independent methods of characterising the oral microbiome has improved the identification of types of resident bacteria. However, bacterial abundance calculated based on 16S rRNA sequencing methods does not directly explain the activity level of these microorganisms, nor does it identify viral or fungal organisms. Hence, causative links of microorganisms to carcinogenesis needs to be carefully considered. However, the ability to accurately quantify relative abundance of bacteria between cancer and healthy patients represents an avenue for developing minimally invasive diagnostic tests. Especially as the oral cavity microbiome samples can be collected in a non-invasive manner with bacterial swabs (Zhao *et al*. 2017), saliva (Lee *et al*. 2017) or oral rinses (Bornigen *et al*. 2017). A recent systematic review that included fifteen studies investigating the healthy oral and oropharyngeal bacteriome using culture independent methods found seven abundant phyla in the oral cavity (Gopinath and Menon 2018). Firmicutes was reported as the most common phylum followed by Bacteroidetes

and Actinobacteria. However, the cancer dependent changes to these phyla were mixed between studies (Gopinath and Menon 2018).

#### 6.1.1.1 Salivary microbiome and head and neck squamous cell carcinoma

In the setting of HNSCC, at a genus level Dialister, Selenomonas, Streptococcus and Treponema have been associated with oral SCC (Guerrero-Preston *et al*. 2016; Guerrero-Preston *et al*. 2017; Wolf *et al*. 2017). At a species level, *Streptoccocus anginosus*, *Porphyromonas gingivalis* and *candida albicans* were implicated in HNSCC (Ha *et al*. 2015; Katz *et al*. 2011). The interaction between the oral microbiota, the host immune system and resulting inflammatory response appears to play a key role in carcinogenesis. While the exact mechanisms are unclear, increased concentrations of inflammatory cytokines have been reported to correlate with the salivary microbiome (Vesty *et al*. 2018). However, the use of salivary microbiome as a biomarker of detecting HNSCC is questionable due to poor differentiation from patients with benign dental pathology (Vesty *et al*. 2018).

### 6.1.2 Intestinal microbiome and dysbiosis

Microbial dysbiosis is defined as any changes to the resident commensal bacterial community in an individual compared to healthy individuals (Petersen and Round 2014). Dysbiosis in the human intestine has been associated with the initiation and pathogenesis of cancer (Engen *et al*. 2015; Savin *et al*. 2018; Zitvogel *et al*. 2015). Alterations to the composition and function of the intestinal microbiome could potentially lead to changes in host immunity, driving a systemic inflammatory process and pre-disposition towards cancer (Turnbaugh *et al*. 2007).

Changes in by-products of carbohydrate fermentation in the large bowel could be utilised as a surrogate marker of gut microbial dysbiosis. The major by-products of microbial carbohydrate fermentation in the large bowel include water, hydrogen ($H_2$), methane ($CH_4$) and short chain fatty acids (SCFA) amongst various other gases (Tadesse *et al.* 1980). Measurement of microbial fermentation products in the large bowel have principally focussed on faecal SCFAs as static markers of microbial function and composition (Pimentel *et al*. 2013; Wilkens *et al*. 1994). However, this approach is cumbersome and cannot provide real-time information on changes to microbial activity. Breath hydrogen and methane have been previously investigated as a non-invasive, surrogate markers of the fermentation status of the large bowel (Peled *et al*. 1987), and changes in breath hydrogen and methane have been linked to numerous pathological states including cystic fibrosis (Bujanover *et al*. 1987; Lisowska *et al*. 2009), vascular ischaemia (Varga *et al*. 2012), colorectal cancer (Karlin *et al*. 1985), and functional gastrointestinal disorders (Perman *et al*. 1984; Tadesse and Eastwood 1978; Tadesse *et al*. 1980). The hydrogen breath test has significant clinical utility for the diagnosis of irritable bowel syndrome (IBS) (Wilder-Smith *et al*. 2018), small intestinal bacterial overgrowth (SIBO) (Kerlin and Wong 1988; Suri *et al*. 2018) and other carbohydrate malabsorption syndromes (Yao *et al*. 2018). The clinical relevance of breath methane is becoming increasingly apparent, with up to 77% of healthy adults producing a methane positive breath sample. These ‘methane producers’ were considered Archaea (a dominant methanogenic anaerobic organism in the human colon) positive, and the presence of Archaea species influences SCFA metabolism. SCFAs are a group of organic compounds (Rombeau and Kripke 1990) (1 – 6 length of carbon atoms) with acetic acid, propanoic acid and butanoic acid accounting for more than 85% of these volatile fatty acids (Demigne and Remesy 1985). An accumulating body of evidence has implicated SCFA

production in several aspects of human physiology, and maintenance of intestinal homeostasis (Chambers *et al*. 2015).

Despite the growing number of studies connecting the microbiome to disease pathology, in oncological research, much of the focus has been directed towards colorectal cancer (Casanova *et al*. 2018; Le Marchand *et al*. 1993; Ma *et al*. 2019). The by-products of microbial fermentation in the large bowel have been described *in vitro* and *in vivo* as protective against colorectal cancer, breast cancer (Thirunavukkarasan *et al*. 2017) and HNSCC (Krishna *et al*. 2002). Recent studies have described a broader link between cancer risk and the gut microbiome (Candela *et al*. 2011; Dzutsev *et al*. 2015; Holma *et al*. 2013), where the effectiveness and adverse event profiles of some chemotherapeutic agents appear to be dependent on the intestinal microbiome (Gopalakrishnan *et al*. 2018; Holma *et al*. 2013). For example, the alpha diversity and relative abundance of known gut microbes were significantly higher in patients responding to immunotherapy compared to non-responders (Gopalakrishnan *et al.* 2018). Therefore, emergent, non-invasive, technologies that can accurately measure real-time fermentation capacity of the large bowel would likely influence clinical decision making in the future.

Breath VOCs have demonstrated promise as non-invasive biomarkers of HNSCC (Hakim *et al*. 2011; Hartwig *et al*. 2017; Schmutzhard *et al*. 2008). Amongst the list of breath VOCs identified as discriminatory for HNSCC are SCFAs (Shoffel-Havakuk *et al*. 2016), suggesting changes to the microbial fermentation patterns in the large bowel of HNSCC patients, which are reflected in breath. Furthermore, none of these studies (Gruber *et al*. 2014; Hakim *et al*. 2011; Hartwig *et al*. 2017; Shoffel-Havakuk *et al*. 2016) have determined how breath hydrogen or methane

concentrations are modified in HNSCC. In this chapter, I aimed to characterise hydrogen, methane and SCFA in the breath of patients with HNSCC and their relationship to patient characteristics using a large cohort of newly diagnosed cancer patients and demographically matched benign patients.

## 6.2 Methods

### 6.2.1 Ethical Approval

Ethical approval for this study was obtained and detailed in Section 5.2.1

### 6.2.2 Patients

The patient groups were defined as per the previous chapter detailed in Section 5.2.2

### 6.2.3 Breath analysis

Exhaled breath samples were collected and analysed as described in Section 5.2. The methodology specific to analysis of exhaled breath hydrogen and methane levels is further described below.

#### 6.2.3.1 Exhaled breath hydrogen and methane analysis

30mL of gas was transferred from the SKC® FlexFoil® PLUS bag using a QuinTron low VOC Luer-Lock syringe with a needle that pierced the septum of the SKC® FlexFoil® PLUS bag maintaining the isolation from room air. The extracted gas was analysed using the QuinTron® BreathTracker® (Milwaukee, WI, USA) for hydrogen, methane, and carbon dioxide levels. The instrument was calibrated

with standard gas mix prior to each use and tested with the standard gas mix periodically for quality assurance.

### 6.2.4 Saliva microbiome analysis

Microbiome analysis of stimulated saliva samples were conducted based on previously validated methodology (Choo *et al*. 2015) with additional modifications described below. The DNA extraction, 16S rRNA gene sequencing and bioinformatics analysis was conducted by the South Australian Health and Medical Research Institute's microbiome research group in the David R Gunn sequencing facilities.

#### 6.2.4.1 Saliva sample collection

Stimulated saliva samples were collected from a subset of patients that provided a breath sample. 10mL of sterile water was provided to the patient to wash the oral cavity and subsequently discarded. 2mL stimulated saliva was then collected using a Norgen saliva DNA collection and preservation device (Norgen Biotek Corp., ON, Canada) as per manufacturer instructions. The samples were collected within a 10-minute period and the preservative (Norgen 589320, Norgen Biotek Corp., ON, Canada) added before storage in -80˚C prior to further analysis.

#### 6.2.4.2 DNA extraction and quantification

3mL of thawed saliva samples (total sample containing 2mL of saliva and 2mL of preservative) were centrifuged at 13,000 x g for 10 minutes and cell pellets were subjected to bead beating (silica/zirconium 1:1 ratio beads, Daintree Scientific, Tasmania, Australia) at 6.5m/s for 1 minute (FastPrep-24 bead beater, MP

Biomedicals, Santa Ana, USA). Samples were then incubated for 1 hour at 37°C in 2.9 mg/mL of lysozyme and 0.14 mg/mL of lysostaphin (Sigma-Aldrich, St. Louis, MO, USA). Extraction of DNA from saliva samples were performed with GenElute Bacterial Genomic DNA kit as per the manufacturer instructions (Sigma-Aldrich, St. Louis, MO, USA). The total DNA was eluted into 100µL of sterile water. The final DNA concentration was quantified fluorometrically with Qubit dsDNA HS Assay kit (Life Technologies).

#### 6.2.4.3 16S rRNA gene sequencing

The highly conserved 16S rRNA sequence was amplified using the V4 hypervariable region from the extracted saliva DNA, with Illumina adapter overhang sequences. The amplicons generated were, cleaned, indexed, and sequenced according to the Illumnia miSeq 16S Metagenomic Sequencing Library Preparation protocol with minor modifications. The initial PCR reaction contained a total volume of 25uL with 12.5ng of DNA (minimum), 1uM of forward primer (5uL), 1uM of reverse primer (5uL) and 12.5mL of KAPA HiFi Hotstart ReadyMix (KAPA Biosystems, Wilmington, MA, USA). The PCR reactions were performed on a Veriti 96-well Thermal Cycler (Life Technologies) using the following settings: 95˚C for 3 minutes, 25 cycles of 95˚C for 30 seconds, 50˚C for 30 seconds and 72˚C for 30 seconds and a final extension step at 72˚C for 5 minutes. Sample multiplexing was conducted using a dual-index approach using the Nextera XT Index kit (Illumina Inc., San Diego, CA, USA) as per the manufacturer instructions. The final library was paired end sequenced at 2 x 300 bp using a miSeq Reagent kit version 3 on the Illumina MiSeq platform.

#### 6.2.4.4 Bioinformatics analysis

The Quantitative Insights into Microbial Ecology (QIIME, version 2) software was used to analyse the 16S rRNA sequences generated from paired-end amplicon sequences. Barcoded forward and reverse sequencing reads were quality filtered and merged using Paired-End read merger (PEAR v0.9.6). Chimeras were detected and filtered from the paired end reads using USEARCH (v6.1) against the 97% clustered representative sequences from the Greengenes database (v13.8). Operational taxonomic units (OTUs) were assigned using an open reference approach with UCLUST algorithm (v1.2.2q) against the SILVA database release 132 (on 11/05/2020), clustered at 97% identity. During the OUT assignment, sequences were pre-clustered at 80% against the reference prior to *de novo* clustering. A minimum subsampling depth of 10,409 reads is selected based on the lowest read depth of samples. QIIME is used to generate an alpha rarefaction curve to confirm that the relationship between read depth and new taxon detection approached an asymptote in all samples at this depth. No samples were eliminated following subsampling.

#### 6.2.4.5 Statistical analyses

QIIME 2 was used to calculate taxa abundance, Shannon's diversity index and Simpson's evenness index for each sample. Beta diversity (Bray-Curtis similarity) was determined using PRIMER (version 7). The similarity matrices based on Bray-Curtis distance were calculated using sample-normalised, square root transformed relative OUT abundance. Principal coordinates analysis was used to visualise clustering of samples based on their similarity matrices. The two-factor permutational multivariate analysis of variance (PERMANOVA) on the Bray-Curtis matrices was used to test the null hypothesis of no difference amongst a priori-defined groups using PERMANOVA. The test was computed using unrestricted

permutation of raw data with 9999 random permutations and at significance level set at 0.05.

Statistical analysis for hydrogen and methane levels were performed using IBM® SPSS® (Version 25. Chicago, IL). Gas concentrations were expressed as median +/- interquartile range (IQR) in parts per million (ppm). Distribution of data was tested using Kolmogorov-Smirnov test. Mann-Whitney-U (MWU) and Kruskal-Wallis tests were used for two-tailed statistical significance testing. Chi-Square statistics were calculated for contingency tables with categorical data. Non-parametric bivariate correlation is conducted using Spearman's rho correlation coefficients. Statistical significance is considered if $p < 0.05$.

## 6.3 Results

### 6.3.1 Distribution of hydrogen, methane and SCFA in breath across cancer and benign patients

Patient groups were matched for age, smoking pack years (SPY) and fasting time (FT). A statistically significant difference in body mass index (BMI) was noted (Table 45), but there were no BMI dependent changes in analysed breath compounds. Median levels of hydrogen, methane and SCFA showed no statistically significant differences between cancer (n=67) and benign (n=50) patients (Table 46, Figure 40).

***Table 45. Relative differences in patient factors between cancer and control patient groups.***

| | Controls (n = 61) | | Cancer (n = 74) | | |
|---|---|---|---|---|---|
| **Patient Factors** | **Median** | **Range** | **Median** | **Range** | **p value** |
| Age (Years) | 57.00 | 31-86 | 56.50 | 33-88 | 0.340 |
| BMI ($kgm^{-2}$) | 28.26 | 20-45 | 26.37 | 17-38 | 0.003* |
| Gender (M : F) | 32 : 29 | | 59 : 15 | | 0.001* |
| Smoking Pack Years | 18.00 | 0-104 | 29.70 | 0-295 | 0.125 |
| Fasting Time (Hours) | 13.00 | 6-18 | 12.00 | 6-26 | 0.533 |

***Table 46. Relative differences in compounds of interest between cancer and benign patient groups.***

| | Benign (n = 61) | | Cancer (n = 74) | | MWU |
|---|---|---|---|---|---|
| | Median | Range | Median | Range | p value |
| CH4 / H2 Ratio | 0.39 | 0-6.71 | 0.67 | 0-17.67 | 0.044* |
| Hydrogen (ppm) | 5.00 | 0-97 | 5.00 | 0-33 | 0.616 |
| Methane (ppm) | 3.00 | 0-50 | 5.00 | 0-58 | 0.697 |
| Acetic Acid (ppm) | 0.31 | 0.05-6.66 | 0.28 | 0.05-4.79 | 0.620 |
| Butanoic Acid (ppm) | 0.01 | 0-0.08 | 0.01 | 0-0.13 | 0.763 |
| Formic Acid (ppm) | 0.11 | 0.04-13.40 | 0.13 | 0.04-3.14 | 0.620 |
| Propanoic Acid (ppm) | 0.02 | 0-0.12 | 0.02 | 0-0.04 | 0.887 |

MWU = Mann Whitney-U test, BMI = Body Mass Index, ppm = parts per million

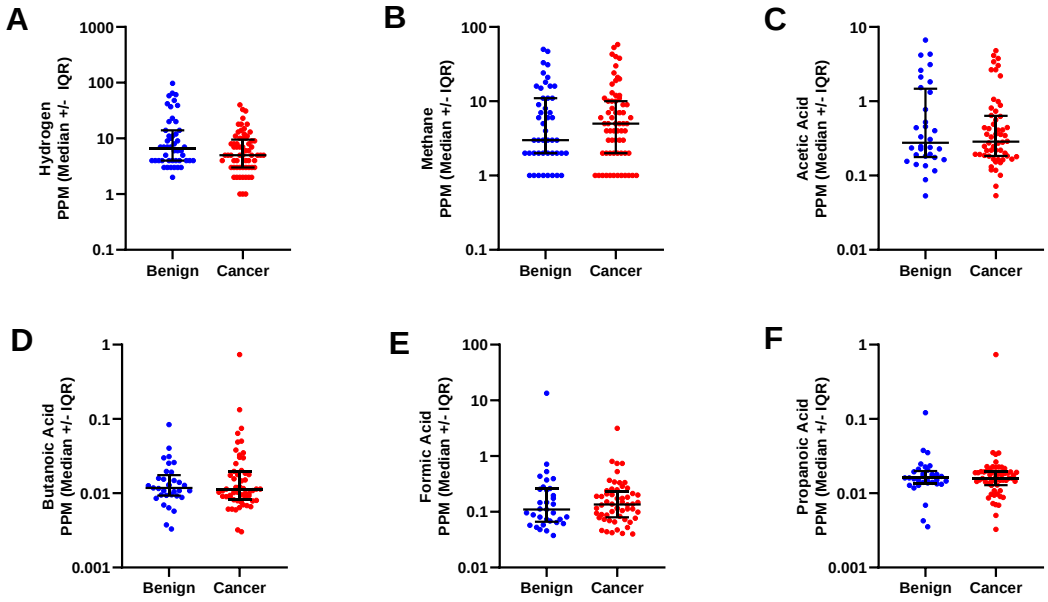



***Figure 40. Relative difference in hydrogen (A), methane (B) and volatile SCFA (C, D, E, F) levels in breath of participants with and without cancer. Y-axis in Log$_{10}$ scale to demonstrate data spread. Individual data points are reported with lines indicating median and interquartile range (IQR). PPM = parts per million***

Intestinal methane, hydrogen and carbon dioxide have a dynamic relationship, where high concentrations of methane can significantly affect the local hydrogen concentration (Jang *et al*. 2010). Therefore, the methane to hydrogen concentration ratio was utilised to explore this dynamic system. The methane to hydrogen ratio was significantly different between patients with (*n = 74)* and without (*n = 61*) cancer, with a lower ratio found in the control group compared to cancer (*p = 0.0440*) (Figure 41-A). The CH4:H2 ratio increased with increasing cancer T-stage and was significantly higher in T4 stage (*Median: 1.6, Range: 0.18 – 17.48*) cancers compared to T1 (*Median: 0.25, Range: 0 – 14.5*) (p = 0.0259; Figure 41-B).

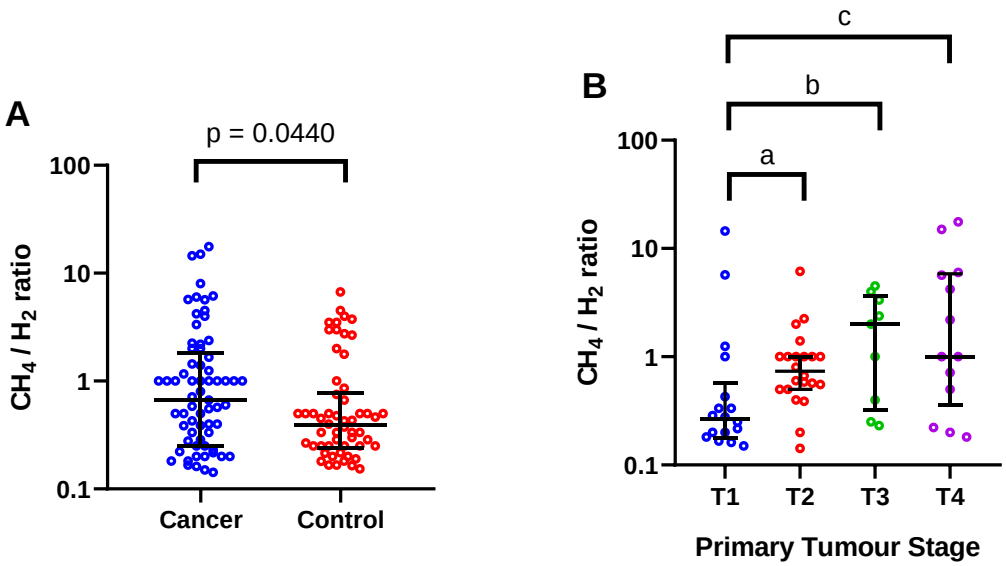



***Figure 41. (A) Cancer patients compared to control patients (MWU, p = 0.0440), (B) T-stage dependent variability in methane to hydrogen ratio. (KW, overall p = 0.0121, Dunn's correction for multiple comparisons between tumour stage p values; a = 0.1449, b = 0.0687, c = 0.0259). Solid lines describe the median and interquartile ranges. Individual data points indicate the value for each patient.***

The relative abundance of methane and various SCFA for the whole cohort appeared to decrease as the number of carbon atoms in these molecules increased. There was also a strong correlation between compounds with similar number of carbon atoms (Table 47, *Figure 42*).

***Table 47. Statistically significant correlations between tested breath compounds***

| Comparison | Carbon Atoms | Correlation Coefficient | P-value |
|---|---|---|---|
| H – M | NA – C1 | 0.305 | 0.004 |
| AA – FA | C2 – C1 | *0.717 | 0.0001 |
| AA – BA | C2 – C4 | 0.323 | 0.002 |
| AA – PA | C2 – C3 | 0.255 | 0.016 |
| BA – PA | C4 – C3 | *0.719 | 0.0001 |

H = hydrogen, M = methane, AA = acetic acid, FA = formic acid, BA = butanoic acid, PA = propanoic acid, NA = not applicable, * indicates high correlation coefficients

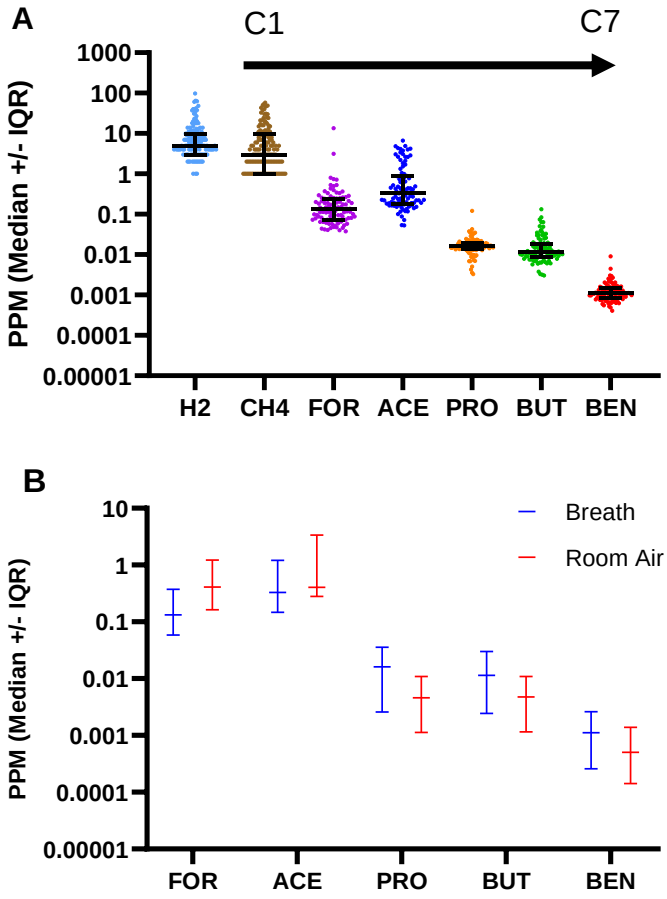



***Figure 42. Relative abundance of hydrogen, methane, and volatile short chain fatty acids in this patient population (Panel A). Panel B demonstrates relative differences between breath and room air for volatile SCFA. H2 = hydrogen, CH4 = methane, ACE = acetic acid, BEN = benzoic acid, BUT = butanoic acid, FOR = formic acid, PRO = propanoic acid, C1 = one carbon atom, C7 = seven carbon atoms, PPM = parts per million, IQR = interquartile range. Solid lines describe the median and interquartile ranges. Individual data points indicate the value for each patient.***

There have been inconsistencies in the literature on the absolute cut-off values to define hydrogen or methane producers. Therefore, a range of cut-off values were used to separate high and low hydrogen or methane producers. These values were then used to test if they can be used to differentiate between benign and cancer patients (*Figure 43*). There appeared to be a separation between cancer and benign patients at a methane concentration of 4 ppm, with more methane producers in the cancer group compared to benign; however, this was not statistically significant ($p < 0.05$). Hydrogen, cut-off values between 2 ppm to 5

ppm could differentiate between cancer and benign patients ($p < 0.05$), with low hydrogen concentration detected in the cancer group compared to benign.

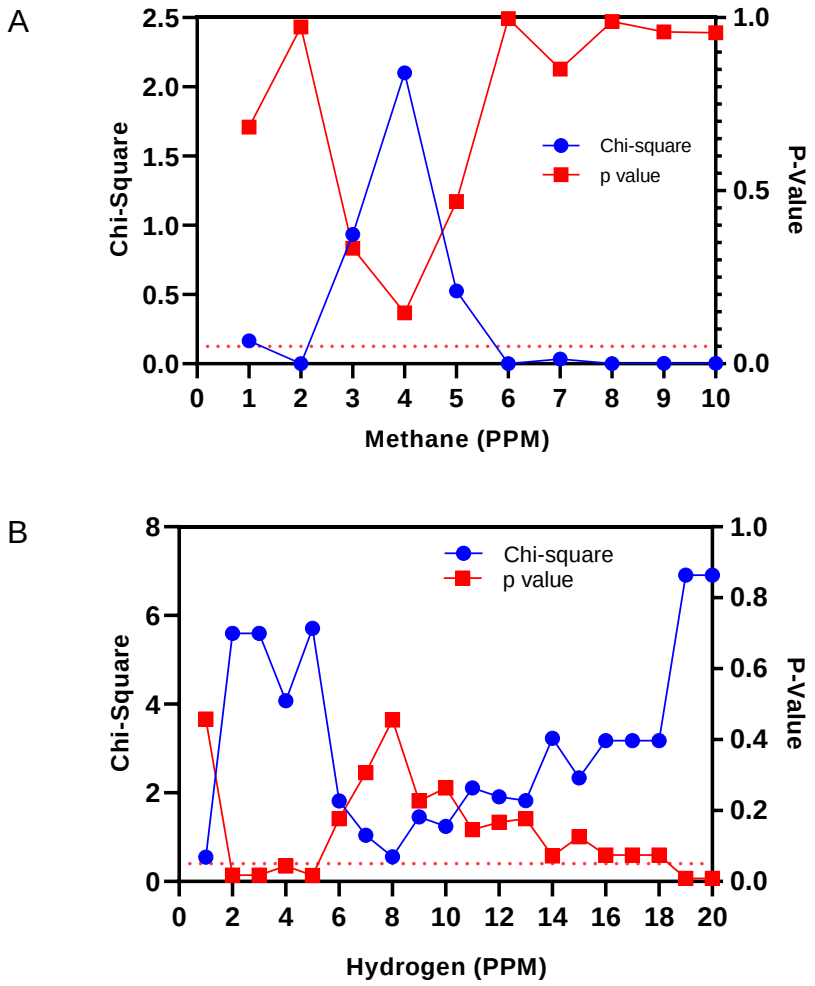



***Figure 43. Range of cut off values for separating methane (A) and hydrogen (B) producers from non-producers and ability to differentiate between cancer and benign patients. Dotted line indicates p-value at 0.05.***

There was a weak, positive correlation between breath methane concentrations and tumour staging ($p = 0.0323$, $R^2 = 0.1747$) (*Figure 44*); however, no such relationship was seen between breath hydrogen concentration and tumour stage.

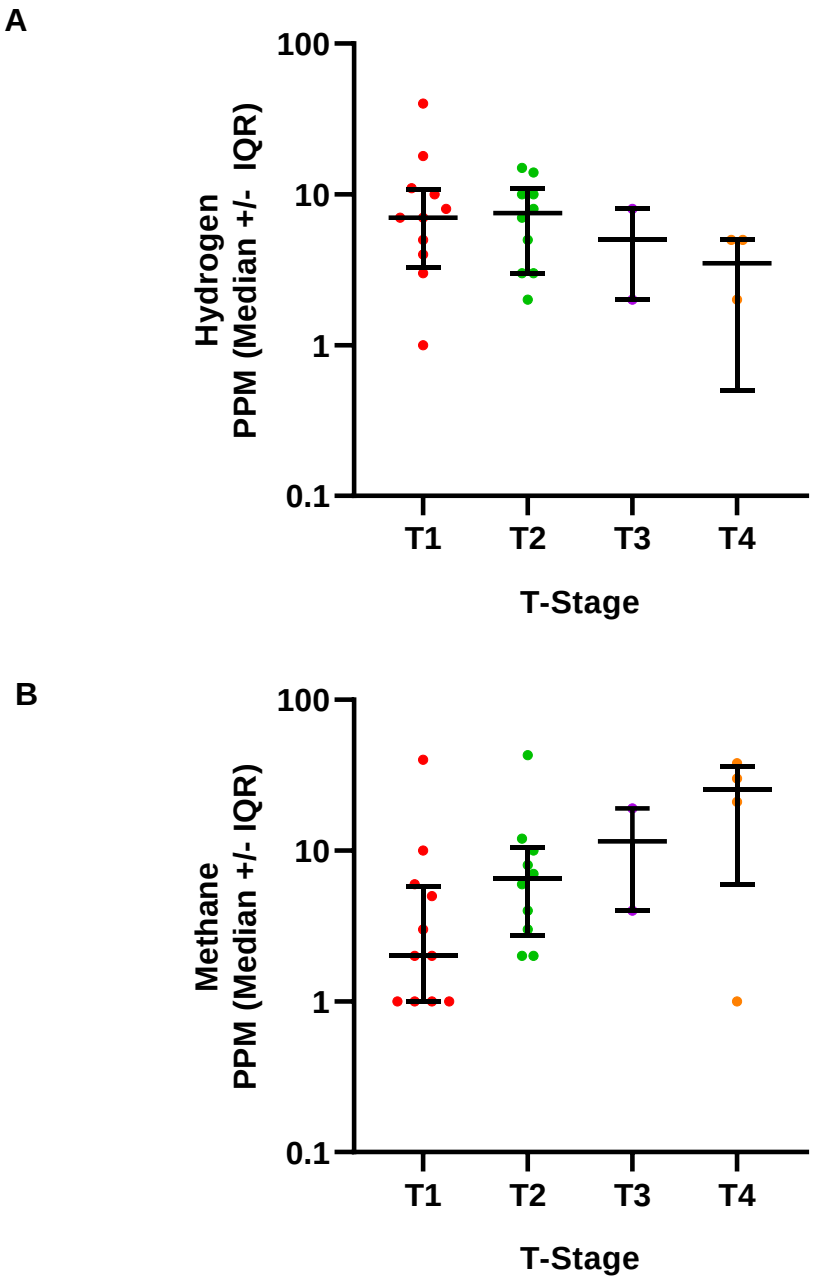



***Figure 44. Relative differences in hydrogen (A) and methane (B) levels based on primary tumour stage of oropharyngeal head and neck cancers. Statistically significant median increase based on T-stage is observed for methane levels ($p = 0.0323$, $R^2 = 0.1747$).***

### 6.3.2 Saliva microbiome relative abundances

A small cohort of the overall group of patients provided a saliva sample for microbiome analysis (HNSCC = 30, Control = 14). There were no significant differences based on patient demographics (Table 48). The total bacterial load was not significantly different between patients with cancer compared to the control group (Figure 45). Additionally, there were no significant differences in Alpha diversity analysis of these patient groups (Figure 46).

***Table 48. Relative differences in saliva microbiome based on patient factors.***

| Comparison | Df | F model | R2 | P value |
|---|---|---|---|---|
| **Disease (Cancer/Control)** | 1 | 0.87208 | 0.02575 | 0.4845 |
| **Smoking Status (NS/ES/CS)** | 2 | 0.67382 | 0.02575 | 0.8072 |
| **Tooth Brushing (Y/N)** | 1 | 1.0758 | 0.03157 | 0.3516 |
| **Gender (M/F)** | 1 | 0.51013 | 0.01522 | 0.8472 |
| **T Stage (1-4)** | 5 | 1.4447 | 0.19942 | 0.0759 |
| **Tumour Subsite** | 4 | 0.65821 | 0.08068 | 0.9021 |
| **Node Positive (Y/N)** | 2 | 0.59329 | 0.03575 | 0.8731 |
| **HPV Status (Y/N)** | 2 | 0.52702 | 0.03189 | 0.9361 |
| **Age Category** | 3 | 1.2315 | 0.10649 | 0.2118 |

HPV, human papillomavirus. Df, degrees of freedom. T, tumour. NS, non-smoker. ES, ex-smoker. CS, current smoker. M, male. F, female.

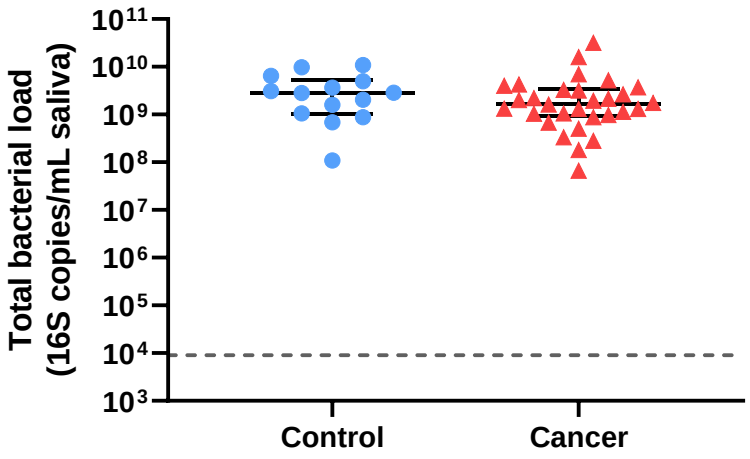
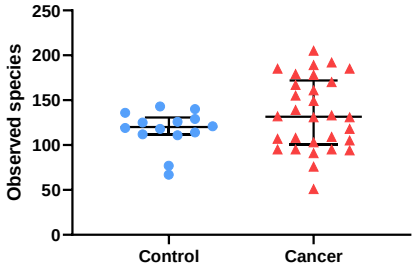
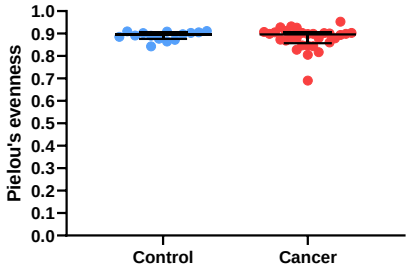



***Figure 45. Total bacterial load comparison between patients with cancer and control groups.***

A




B




C

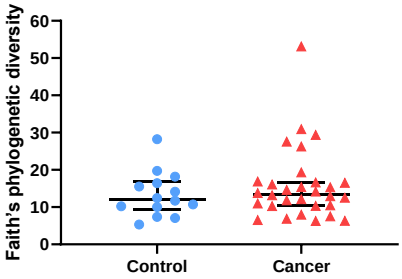



***Figure 46. Alpha diversity analysis of saliva microbiome between patients with cancer and control group. A – microbial richness, B – microbial evenness, C – microbial diversity.***

The relative abundance analysis of species between patients with cancer and the control group indicated no clear differences (Figure 47). Further analysis of compositional data at a genus level was also not significant ($p = 0.7097$) (Figure 48). While the overall relative abundances were not significantly different, four bacterial taxa had a higher relative abundance in patients with HNSCC ($p < 0.05$ without correcting for multiple comparisons) (Figure 49). However, these differences were not statistically significant when a false discovery rate was applied due to multiple comparisons. Figure 50 displays the taxa with high Cliff's delta estimates indicative of a variable distribution between the patient groups.

Here the patients with cancer had a larger variation in these taxa compared to the control group.

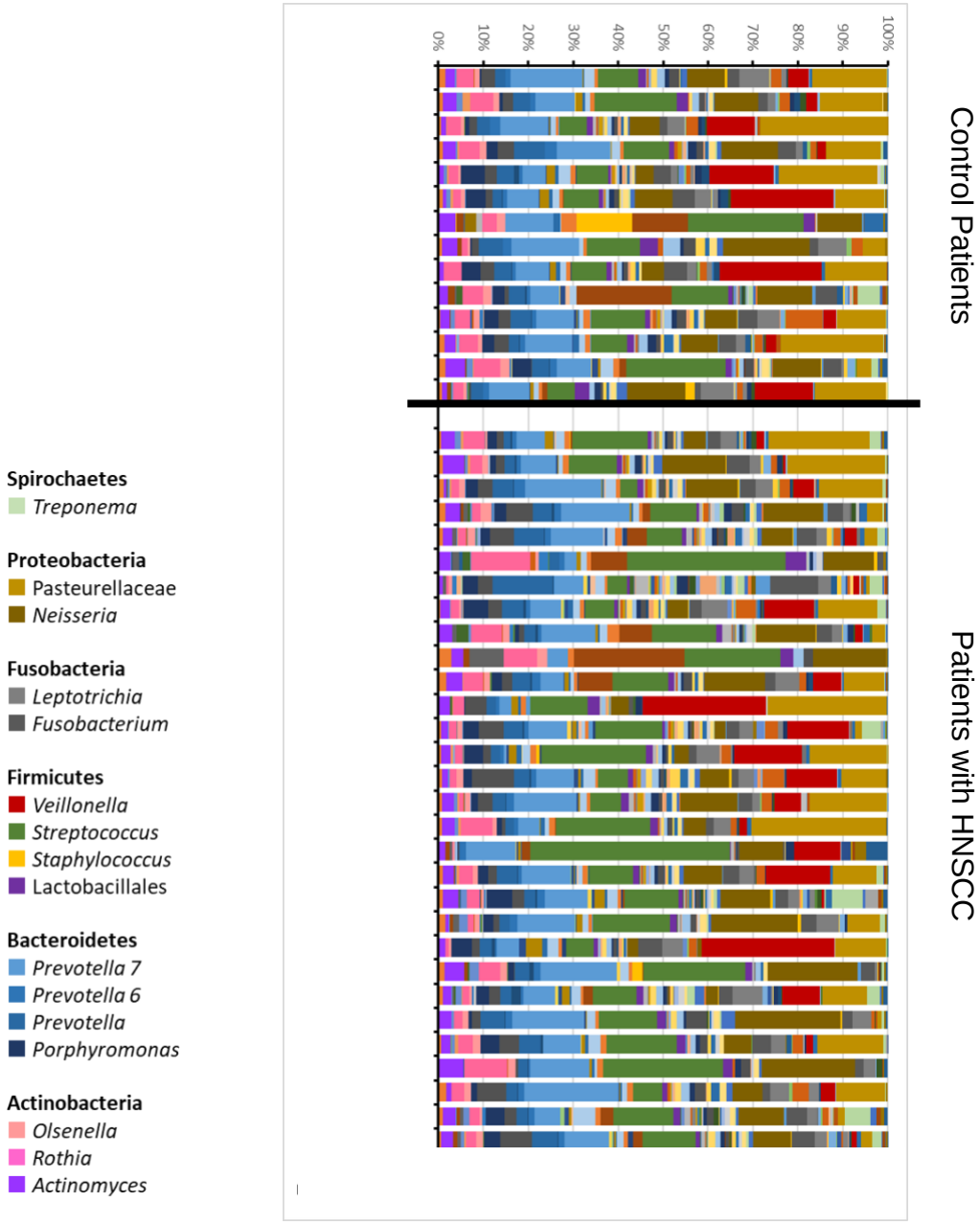



***Figure 47. Bacterial relative abundance plot based on subsampling depth of 7501 reads. Common representation of species is listed above with corresponding colour. HNSCC, head and neck squamous cell carcinoma.***

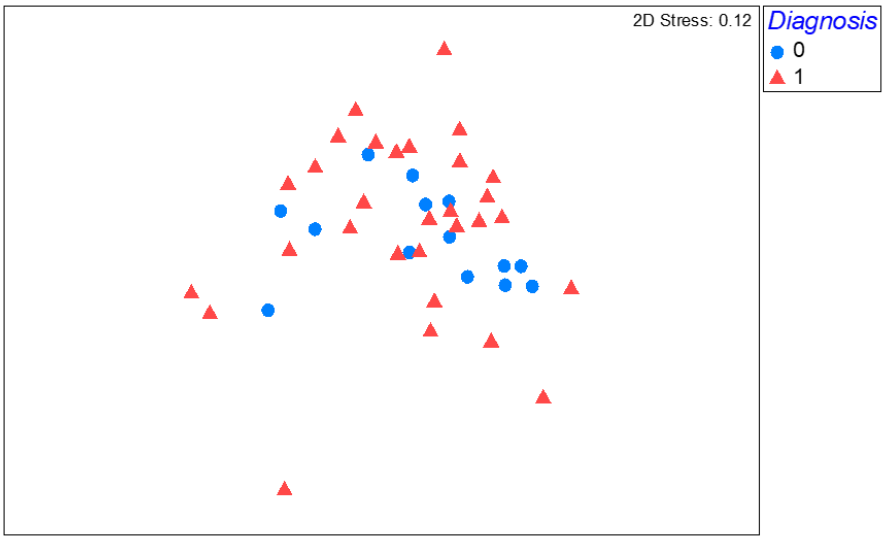



***Figure 48. Ordination of samples on a non-metric multidimensional scaling plot visualisation of compositional changes between patient groups. 0 – control, 1 – cancer.***

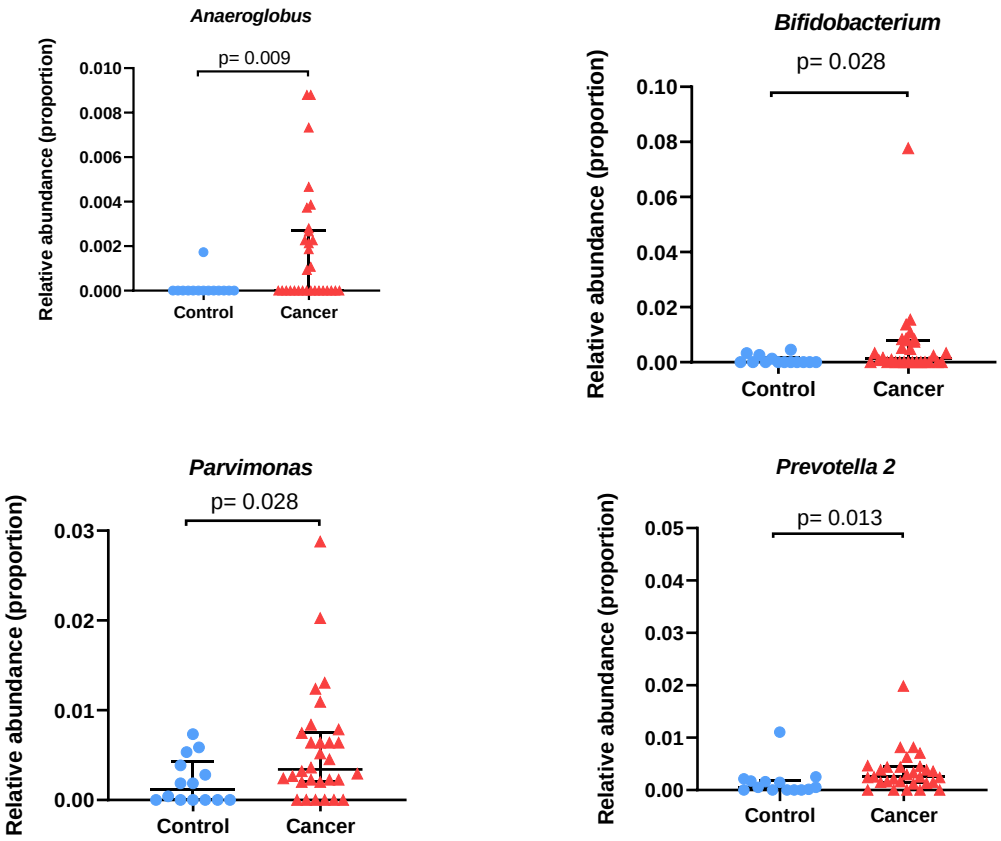



***Figure 49. Comparison of relative abundances and the effect size of bacterial taxa between patient groups.***

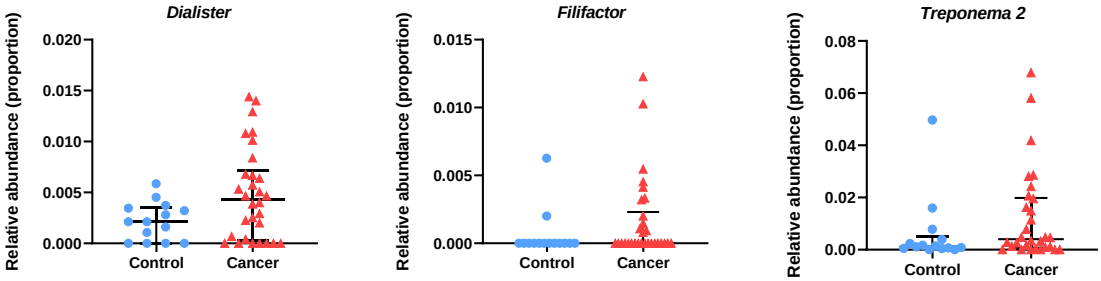



***Figure 50. Taxa with high Cliff's delta estimate values.***

## 6.4 Discussion

This is the first study to characterise breath hydrogen, methane and SCFA in a cohort of non-gastrointestinal cancer patients. Using modern breath analysis techniques, these findings suggest significant changes to exhaled hydrogen and methane concentrations in patients with cancer that could indicate a modified fermentation profile in the large bowel of HNSCC patients. Furthermore, an independent relationship between tumour stage and gastrointestinal gases exhaled in breath are noted suggestive of an intestinal dysbiosis in HNSCC. In addition, oral cavity microbiome was analysed in a subgroup of patients with no statistically significant difference in patients with HNSCC compared to the control group.

Breath hydrogen and methane are derived almost exclusively from anaerobic microbial fermentation of carbohydrates in the large bowel (Figure 51), and have been used as a non-invasive, surrogate marker of fermentation activity in the colon (Rezaie *et al*. 2017). Sloan *et al* have reported a decrease in breath hydrogen in humans consuming a low Fermentable Oligo-, Di-, Mono-saccharides and Polyols (FODMAP) diet supplemented with oligofructose, finding associations between microbial abundance and hydrogen production (Sloan *et al*. 2018). Additionally,

the authors reported lower microbial diversity and increased abundance of methanogenic bacteria to be associated with high methane production (Sloan *et al*. 2018). However, longitudinal studies are needed to quantify exhaled gases of intestinal origin in parallel to quantifying dietary factors, and the composition of the intestinal microbiome, to determine specific systemic effects of intestinal dysbiosis and cancer risk.

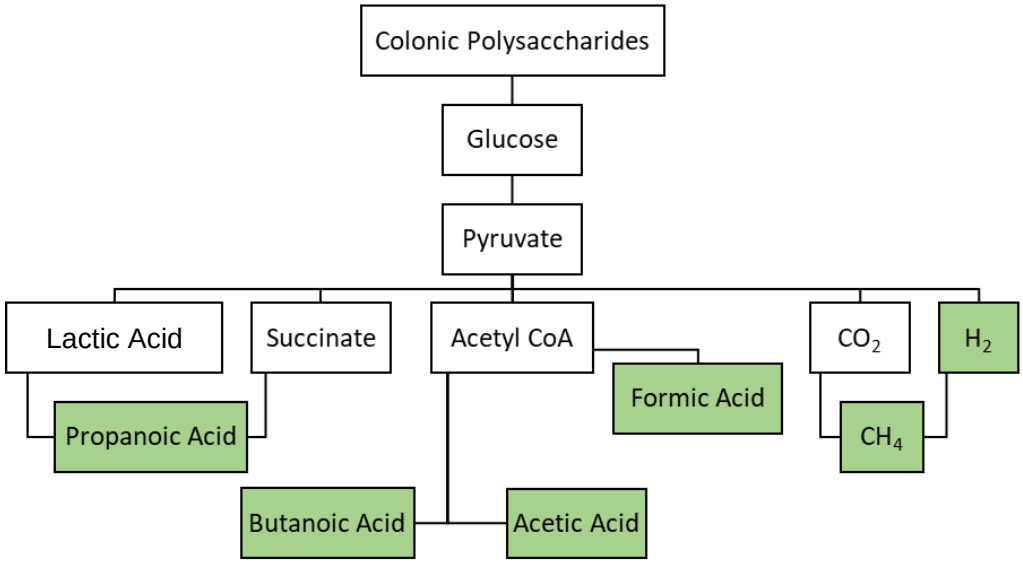



***Figure 51. Main volatile products during bacterial fermentation of colonic polysaccharides detectable in breath (highlighted in green), $CO_2$, carbon dioxide. $H_2$, hydrogen. $CH_4$, methane.***

Elevated breath methane has been associated with delayed gastrointestinal transit, chronic constipation, and diverticulitis (Suri *et al*. 2018). Intestinal methane is produced when $CO_2$ is reduced to $CH_4$ by methanogenic bacteria using $H_2$ as an electron donor. Only three phylotypes of the methanogens have been detected in humans, with the dominant intestinal methanogens being Methanobrevibacter smithii (M. smithii), which has a dependent relationship with hydrogen produced by other organisms to reduce colonic carbon dioxide to methane (Gaci *et al*. 2014). *Methanobreribacter oralis* has also been reported in the oral cavity; however, it's

contribution to overall breath methane is likely negligible (Carbonero *et al*. 2012). The population distribution of M. smithii is variable and can be dependent on various patient factors such as their weight, diet, and presence of disease (Gaci *et al*. 2014).

Definition of a methane producer ranges between 1 to 10 ppm in breath. Approximately 28% of patients in this study were high methane producers (conservatively defining high methane production as 10 ppm), likely due to a methanogen dominant intestinal microbiome. Baseline population studies using a methane production cut-off of 1 ppm have suggested that 34% of adults produced methane in their breath up to 80 ppm (median 11.8 ppm, mean 15.2 ppm) (Levitt *et al*. 2006). When a 1ppm methane cut-off was applied to this dataset, 95% of our cohort could be classified as methane producers; however, it is likely that 1ppm likely falls within the inherent error of measurement equipment and such a low cut-off should be reconsidered. The North American Consensus Guidelines recommends a 10 ppm breath methane to define a methane producer, the remaining literature has been less clear on this definition, and previous studies have used cut-off values ranging from 1 ppm and higher (Gottlieb *et al*. 2017). If breath markers of intestinal gases are to have any utility as non-invasive measurement tool of the functional capacity of the intestinal microbiome, then reliable cut-off thresholds must be established to clearly identify hydrogen and methane producers.

In the context of a biomarker for detecting cancer, using 4 ppm of methane as a cut-off was able to differentiate patients with cancer from the control patients. Gottlieb *et al* has also recommended a cut-off value between 4 and 5 ppm for

accurately identifying a methane producer (Gottlieb *et al*. 2017). However, further large-scale studies are needed to ascertain if there is geographical or ethnicity dependent variability, as these factors are likely to be associated with differing dietary habits that could affect abundance of methanogens. The concentration of exhaled methane in our patient cohort ranged between 0 to 58 ppm (median: 4 ppm). This was significantly lower compared to previously reported baseline population levels (median: 12 ppm, Range: 0 - 80) (Peled *et al*. 1987). These exhaled methane levels in isolation were not significantly different between cancer and benign patients ($p > 0.05$). Therefore, lower concentrations of exhaled methane observed in our dataset could potentially reflect dietary variations and environmental exposures associated with the Australian population. Overnight fasting (6-12 hours) has been described to decrease gastrointestinal motility (Di Lorenzo *et al*. 1991) and therefore, could affect exhaled methane levels (Triantafyllou *et al*. 2014). Patients in our study fasted between 6 to 12 hours prior to providing the breath sample as a requirement for general anaesthesia for an elective procedure. While the median exhaled levels of methane were lower in this study compared to literature, a fasting time-dependent relationship was difficult to demonstrate based on the available data. Therefore, it is difficult to elucidate a causative mechanism to explain the low exhaled methane levels. The participants were not controlled for the types of carbohydrates they consumed prior to the fasting period. While there is potential for a participants' pre-sample food intake to affect exhaled methane concentrations, prior literature only reports minimal changes in breath methane levels as a result of ingesting specific carbohydrates (Rumessen *et al*. 1994).

Exhaled hydrogen concentrations are more sensitive to dietary influence compared to methane (Perman *et al*. 1984). Therefore, the variability in breath hydrogen could have been directly influenced by food intake prior to breath sampling. Exhaled hydrogen levels in our cohort of patients ranged between 0 and 98ppm (median: 5 ppm). When a range of cut-off values were used for breath hydrogen levels, an association between 'low' and 'mid-range' hydrogen production and HNSCC was observed. While the patients in this study were fasted for a minimum of six hours, the food consumed prior to this could have continued to influence breath hydrogen levels due to varying gut transit times for individual patients (Di Lorenzo *et al*. 1991; Triantafyllou *et al*. 2014). As a future direction, an external measure of gut transit time (such as radio-opaque gel capsules) could be used to normalise this data.

Additionally, due to the dynamic relationship between hydrogen, methane and carbon dioxide in the human intestine, isolated analysis of individual gases may not be an accurate representation, and concurrent reporting of the methane to hydrogen ratio to represent the 'hydrogen economy' may have greater utility. The hydrogen economy has been described previously as the transfer of electrons between metabolic reactions involved in cellular respiration and colonic fermentation involving redox reactions (Carbonero *et al*. 2012). As the majority of exhaled hydrogen and methane is secondary to fermentation in the large bowel, any detectable changes to these compounds are likely a representation of intestinal microbial dysbiosis (Rezaie *et al*. 2017). Whether this dysbiosis is causative or because of the malignancy is difficult to elucidate in the current study.

HNSCC generally affects the structure and function of the upper aerodigestive tract hence the ability to smell, taste and swallow food are affected (Hammerlid *et al*. 1998). Therefore, it is likely that the profile and amount of food ingested by the cancer patient group could have been different to the control group. In most head and neck subsites, as the tumour size increases (increased T-stage), the patients' ability to swallow certain types of food becomes limited due to structural obstruction and functional loss (Kubrak *et al*. 2019). Therefore, this could explain the significant differences in breath markers of intestinal health. The ability to detect such differences using non-invasive means provides an avenue for monitoring nutritional status in this patient group. Furthermore, real-time monitoring of changes in the intestinal microbiome with breath analysis may also prove to be a useful in light of recent reports on the influence of the gut microbiome on cancer therapy (Gopalakrishnan *et al*. 2018).

Tobacco smoking is a commonly cited carcinogen to humans and a significant risk factor for HNSCC (Buszewski *et al*. 2009). Variations in breath methane production has been previously reported in smoking populations (Bolin *et al*. 1996). Bolin *et al* reported smokers were more prevalent methane producers compared to non-smokers (using a cut-off value of 2 ppm) (Bolin *et al*. 1996). Breath hydrogen levels are reported to increase up to 137-fold above baseline levels if breath sample is collected within five minutes of cigarette smoking and returns to baseline within 10 to 15 minutes (Rosenthal and Solomons 1983) likely due to volatile compounds in cigarette smoke with similar retention-times affecting the gas-chromatographic analysis (Tadesse and Eastwood 1977). However, this cohort of patients abstained from smoking for at least six hours prior to the collection of the breath sample, therefore, it is unlikely to have directly contributed

to the observed differences. The changes in breath hydrogen could possibly reflect complex changes to the intestinal microbiota in the large bowel of HNSCC patients that are likely driven by diet and smoking. Interestingly, a lower breath hydrogen concentration was observed in smokers compared to non-smokers. (Appendix V - Figure 19).

The volatility of organic compounds is dependent on their molecular weight, polar groups, and hydrogen bonding (Martinez-Lozano and Fernandez de la Mora 2008). The volatile SCFAs measured in this chapter demonstrated an inverse relationship between the number of carbon atoms (and increasing molecular weight) and their abundance (Figure 42. Relative abundance of hydrogen, methane, and volatile short chain fatty acids in this patient population (Panel A). Panel B demonstrates relative differences between breath and room air for volatile SCFA. H2 = hydrogen, CH4 = methane, ACE = acetic acid, BEN = benzoic acid, BUT = butanoic acid, FOR = formic acid, PRO = propanoic acid, C1 = one carbon atom, C7 = seven carbon atoms, PPM = parts per million, IQR = interquartile range. Solid lines describe the median and interquartile ranges. Individual data points indicate the value for each patient. The VOCs with higher abundance in breath were also present at high level in the ambient air samples. Therefore, these highly volatile compounds potentially influenced the measured levels of volatile SCFAs in breath. The highly volatile SCFA that are present in the room air are likely to equilibrate faster with mixed alveolar breath samples similar to volatile agents used in general anaesthesia (Khan *et al*. 2014). Hence, it may not represent an accurate measure of endogenously produced volatile SCFA as some of these compounds are present in cleaning products routinely used in the hospital environment. The SCFA where the room air levels are lower than the breath

samples are likely to represent endogenous volatile compounds. However, these high molecular weight SCFA are less likely to be absorbed from the large intestines and limited detection using mass spectrometry in order to be quantified using exhaled breath (Martinez-Lozano and Fernandez de la Mora 2008). Therefore, at this stage reliable measurement of SCFA as a marker of intestinal microbial function is limited.

There is a rapid emergence of literature describing an interdependent relationship between the gut microbiome, human immune system, and cancer (Gopalakrishnan *et al*. 2018; Li *et al*. 2019). Intestinal dysbiosis has been implicated in malignant transformation (Russo *et al*. 2016), therapeutic response (Shigematsu and Inamura 2018) and long-term survival (Gopalakrishnan *et al*. 2018). Therefore, effective assessment of intestinal dysbiosis could provide positive outcomes to these patients. Breath testing represents a non-invasive, real-time, and low-cost microbial profile of patients. This report further establishes the utility of human breath as a non-invasive indicator of changes in the gut microbiome.

Supporting the argument for novel techniques to non-invasively assess the intestinal microbiome, initial attempts at collecting stool microbiome samples from HNSCC patients were largely unsuccessful due to logistical considerations and patient disinclination not further described here. However, a sub sample of oral microbiome data was analysed to characterise the saliva microbiome and its relationship to HNSCC, and breath biomarkers measured here. Overall, there was no statistically significant differences between the patient groups. However, specific taxa were highly variable in patients with HNSCC compared to the control group. This could indicate potential differences in the oral cavity microbiome of

HNSCC patients that is undetected here due to the small sample size. Although, it appears unlikely that oral microbiome is contributing to the exhaled breath biomarkers measured here (Appendix V). The unequal comparison groups with a heterogeneous HNSCC group that included patients with cancers from various subsites precludes further statistical analysis.

This is the first report characterising exhaled hydrogen and methane in patients with HNSCC. While there are demonstrated associations with the presence of HNSCC, T-stage and smoking status it is difficult to elucidate a causal relationship. Further research into use of breath markers of intestinal homeostasis is warranted and may play a role in personalised treatment plans for this heterogeneous group of patients. The limited oral cavity microbiome analysis conducted here did not indicate any statistically significant differences between patients with HNSCC and the control group. However, the variability observed in some taxa, implies the need for further controlled studies to delineate a true difference in oral cavity microbiome unique to patients with HNSCC. These changes are likely to be cancer subsite dependent.

# 7 Overall discussion and conclusions

## 7.1 Summary of findings

HNSCC is a devastating disease with an obvious need for earlier detection tools. The novel findings in this thesis have demonstrated that a breath test or liquid biopsy is a potential minimally invasive method for detecting HNSCC. To date, there are a paucity of biomarkers translated from bench to bedside due to various methodological shortcomings during the discovery and validation phases of research. This thesis provides evidence that both circulating extracellular microRNAs and exhaled breath volatile compounds can accurately detect HNSCC (Mayne *et al.* 2020, Dharmawardana *et al.* 2020b). These minimally invasive biomarkers have a complex relationship with the host and disease factors that require careful consideration.

## 7.2 Clinical implications

Measurement of circulating microRNAs and exhaled breath compounds represent minimally invasive techniques for detecting HNSCCs. In the context of early detection and precision medicine, these modalities are likely to play a significant role in the diagnosis of human disease.

The prevalence of HNSCC in the general population is relatively low (Mehanna *et al*. 2013). Therefore, screening at a population level is unlikely to have a high yield. Screening programs based on clinical symptoms and signs can detect HNSCC, although, signs and symptoms with high specificity is congruent with advanced disease (Dolan *et al*. 1998). Therefore, minimally invasive biomarkers capable of detecting early disease could have a role in the primary care setting to expedite early referral and diagnosis. For example, when a patient with mild signs

or symptoms of HNSCC (with low risk factors) presents to a primary care physician, a breath or a blood test could be offered to prioritise the urgency of referral or further investigation. The findings in this thesis demonstrate that these tests carry a relatively high specificity to rule in HNSCC meaning that a positive test could instigate further investigations or specialist referral (Figure 52). Their high sensitivity also indicates potential as a 'rule out' test for cancer and reassurance for patients. However, as it stands now, the sensitivity is not adequate to use as a rule out test in the context of heterogeneity associated with HNSCC.

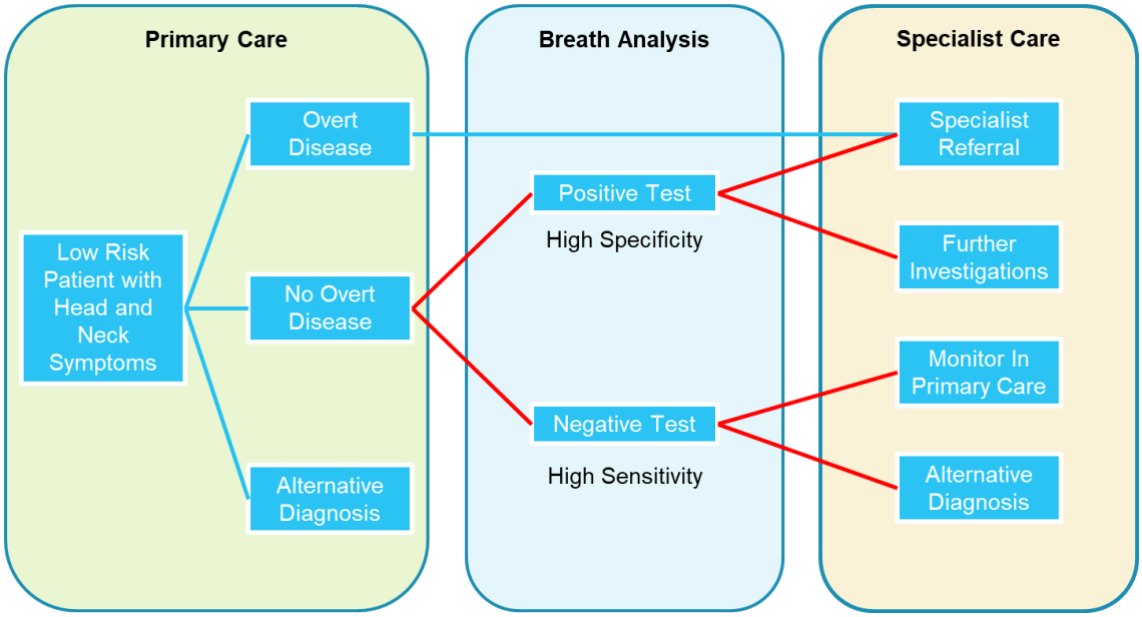



***Figure 52. Proposed implementation of a breath test at a primary care setting to expedite specialist referral or further investigations.***

Metabolomic tests such as exhaled breath tests are unique; in that it also opens the possibility to assign an alternative diagnosis based on their volatile breath profile. While further clinical research is required, novel evidence indicate the presence of disease specific breath profiles (Ferrandino *et al*. 2020; Kort *et al*.

2020). High powered clinical trials along with high quality methods are required for the potential clinical translation of this technology.

Future direction of this technology has the potential to determine a patient's biomarker profile and individualise their treatment paradigm and predict their treatment outcomes (Figure 53). However, large scale studies are required prior to implementing such technology. This form of precision medicine would have the potential to provide superior care for these patients and save valuable health dollars to the health sector (Gavan *et al*. 2018).

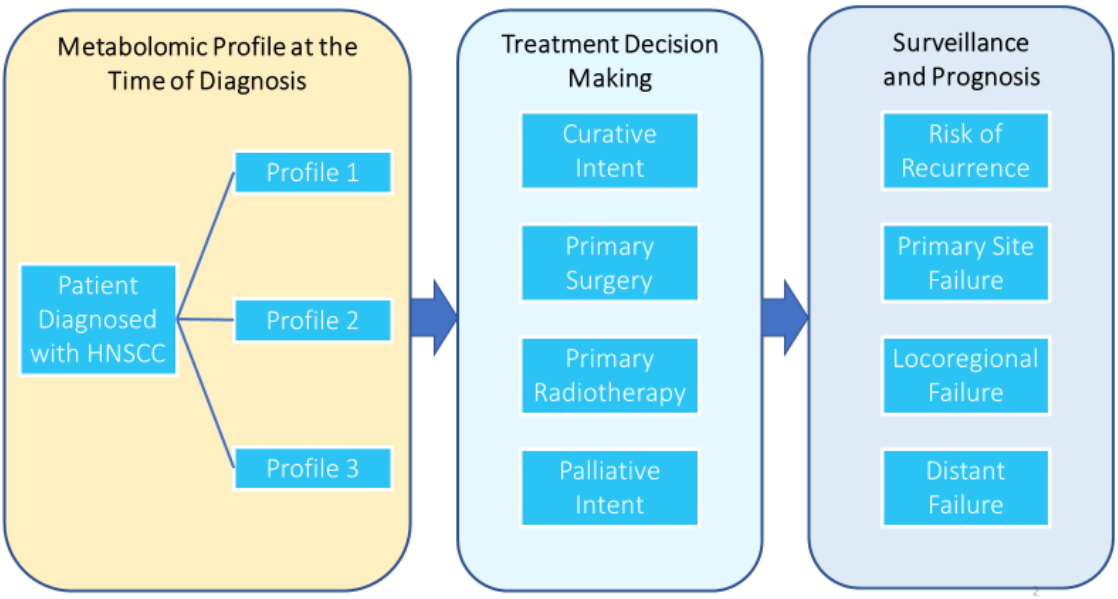



***Figure 53. Potential clinical applications of metabolomic profiles of patients with head and neck cancer. HNSCC, head and neck squamous cell carcinoma. Profile 1-3 are arbitrary representations. Treatment options do not depict all available options. Prognostic endpoints do not represent all available options.***

The changes observed in hydrogen and methane levels in relation to HNSCC and novel research in this thesis describe the intricate relationship between the intestinal microbiome and the host immune system. Many clinical trials are proving

the utility of immunotherapy in HNSCC (Subramaniam *et al*. 2019), although, individual patient outcomes appear to vary (Napolitano *et al*. 2019). Potentially the individual patient responses to response to immunotherapy is determined by pathological and molecular factors including the host microbiome. Research to categorise the host microbiome profile can potentially lead to better selection of patients who would benefit from immunotherapy (Li *et al*. 2019). The breath Methane to hydrogen ratio provides a non-invasive tool to categorise patients based on their intestinal microbiome (Figure 54).

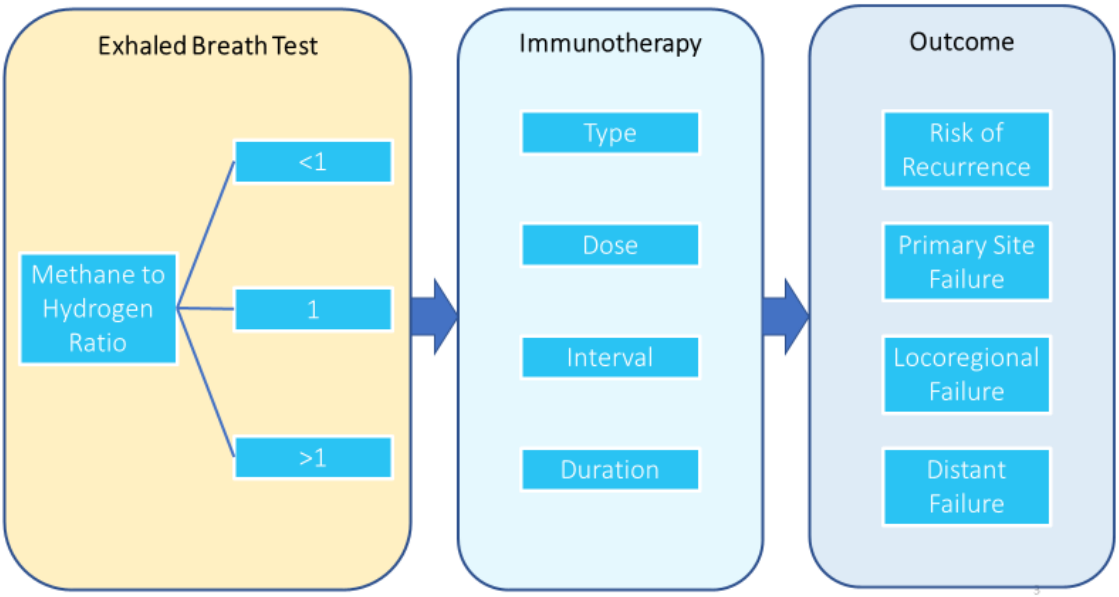



***Figure 54. Potential clinical application of methane to hydrogen ratio as surrogate marker of intestinal microbiome for immunotherapy decision making. Arbitrary ratios are indicated as an example.***

Whilst the potential clinical applications of current technology are exciting, it is important to consider physiological, pathological, molecular, and metabolic mechanisms that contribute to such individual patient profiles.

## 7.3 Research implications and limitations

The experimental methods used in this thesis demonstrates the complexities associated with standardisation of clinical research methodology and subsequent data analysis. While these studies provide insight towards the use of minimally invasive biomarkers for detection of HNSCC, there were significant potential confounders including the host microbiome, fixed patient factors and cancer risk factors that require ongoing research.

Numerous machine learning algorithms are available to classify or predict human disease, and careful consideration must be given to the appropriateness of individual methods and the dataset. The majority of past biomarker studies reviewed in this thesis only used frequentist statistical methods to compare certain markers or patient groups. Of the studies that did utilise advanced modelling, they did not consider various aspects (sample size, pre-test probability and confounders) that have the potential to introduce bias. Therefore, the outcome of such modelling studies has resulted so far in low translational utility. Overall, there is a need for analytical guidance and standardisation for large biological datasets.

Discovery of biomarkers for HNSCC are exemplified by many large-scale studies published in recent literature (Konings *et al*. 2020). However, validation studies for these markers are lacking (Kim *et al*. 2014). The reasons are multifactorial as discussed in previous chapters. However, in an ideal setting, a standard collection, processing, and analytical method needs to be agreed upon. Subsequently, a large enough sample size needs to be recruited to generate an 80% or more statistical power while representing the true prevalence of disease in the community. Once these criteria have been met, the plethora of previously

‘discovered’ biomarkers can be validated and translated for clinical use. However, this would require multi-institute collaborations with centralised data processing and data analysis.

In breath research, in an ideal setting, patient samples should be analysed using multiple instruments. This form of cross validation using multiple instruments adds further confidence to the overall result. Also, various equipment platforms allow specialist functions that are not available in all methods. For example, e-nose devices are clearly capable of pattern recognition of various breath compounds associated with disease. However, without mass-spectrometry examination of these samples, mechanisms that generate the gas pattern may never be elucidated. Although, the lack of established best practises in gas sampling and analysis precludes direct inference between technologies (Herbig and Beauchamp 2014). Therefore, comparative studies provided by Cauchi *et al* that directly compare the performance and outcomes of multiple gas analytical modalities using the same samples should be encouraged for the progress of this field (Cauchi *et al*. 2015).

Similarly, in circulating microRNA research, discovery and validation processes require cross validation of available techniques. Poel *et al* provides an example of this process whereby comparing the pre-analytical processes and the normalisation strategies using the same raw serum samples allows for identification of optimal methodology. While it is unrealistic to expect all experiments to have this strategy (mainly due to associated cost and lack of samples), it is important to encourage similar studies in the field to establish future best practice.

The host microbiome likely has a complex relationship with the biomarkers studied in this thesis. Ideally, this concept is further explored in future research with a subgroup of patients with HNSCC compared to patients with upper aerodigestive tract inflammatory disease (e.g. tonsillitis). The variability in the microbiome of the upper aerodigestive tract as well as the lower intestinal microbiome should be concurrently investigated. Until intricate study of the host microbiome along-side the biomarker of interest is studied, it would be difficult to consider the direct influence of these micro-organisms to the individual variability observed in biomarkers. However, the recruitment of an age matched, relatively homogeneous group of patients in a short period of time is difficult. Therefore, this highlights the need for Biobanks of samples and data for future analysis.

A simple method to resolve this issue is the publication of accurate (and complete) methodology with raw data. The advent of cloud computing and online research repositories with minimal access fees provides an excellent platform. Therefore, journals should request for compulsory deposition of such data for future use. Freely available data repositories would provide an avenue for researchers to confirm accuracy of their data and identify any methodological shortfalls. These large repositories would have the potential to generate pooled high-level evidence regarding the utility of these biomarkers in clinical practice.

Research groups around the globe have made significant progress in HNSCC molecular level research. However, careful interrogation of these studies indicated small sample sizes, variable methodologies and non-standardised analyses and reporting methods. Therefore, it is impractical to pool such data to generate more

meaningful results. In the field of HNSCC biomarker discovery and detection, there is real scope to provide governance to standardise methodology.

## 7.4 Concluding remarks

The findings described in this thesis provide evidence that circulating extracellular vesicle microRNAs and raw mass spectra of exhaled breath analysis can accurately detect patients with HNSCC with high sensitivity and specificity comparable to existing biomarker studies. The relationship between exhaled hydrogen and methane and HNSCC as surrogate markers of intestinal dysbiosis are described for the first time. Collectively, these findings have potential significant clinical implications for translating into minimally invasive tools for detection of cancer, for improved molecular selection of patients for various treatments, and for further methodological standardisation in cancer research.

# Bibliography

Abderrahman B (2019a) Exhaled breath biopsy: a new cancer detection paradigm Future Oncology 15:1679-1682 doi:10.2217/fon-2019-0091

Abderrahman B (2019b) Exhaled breath biopsy: a new cancer detection paradigm Future Oncol 15:1679-1682 doi:10.2217/fon-2019-0091

Agrawal N et al. (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1 Science 333:1154-1157 doi:10.1126/science.1206923

Ai L et al. (2003) The p16 (CDKN2a/INK4a) tumor-suppressor gene in head and neck squamous cell carcinoma: a promoter methylation and protein expression study in 100 cases Mod Pathol 16:944-950 doi:10.1097/01.MP.0000085760.74313.DD

AIHW (2013) Cancer in Australia. Australian Institute of Health and Welfare.

AIHW (2014) Head and Neck Cancers in Australia. Australian Institute of Health and Welfare.

AIHW (2017) Cancer in Australia. Australian Institute of Health and Welfare.

Al-Khanbashi M, Caramuta S, Alajmi AM, Al-Haddabi I, Al-Riyami M, Lui WO, Al-Moundhri MS (2016) Tissue and Serum miRNA Profile in Locally Advanced Breast Cancer (LABC) in Response to Neo-Adjuvant Chemotherapy (NAC) Treatment PLoS One 11:e0152032 doi:10.1371/journal.pone.0152032

Alkhouri N et al. (2014) Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children Eur J Gastroenterol Hepatol 26:82-87 doi:10.1097/MEG.0b013e3283650669

Alkhouri N et al. (2015) Breathprints of childhood obesity: changes in volatile organic compounds in obese children compared with lean controls Pediatr Obes 10:23-29 doi:10.1111/j.2047-6310.2014.221.x

Allaband C et al. (2019) Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians Clin Gastroenterol Hepatol 17:218-230 doi:10.1016/j.cgh.2018.09.017

Alsarraf M, Chu CYT, Lai L, Carey MK, Drelichman A (1981) Multiple Tumor-Markers in Monitoring Patients with Epidermoid Cancer of the Head and Neck Proceedings of the American Association for Cancer Research 22:285-285

Anesti V, McDonald IR, Ramaswamy M, Wade WG, Kelly DP, Wood AP (2005) Isolation and molecular detection of methylotrophic bacteria occurring in the human mouth Environ Microbiol 7:1227-1238 doi:10.1111/j.1462-2920.2005.00805.x

Arasaradnam RP et al. (2016) Non-invasive exhaled volatile organic biomarker analysis to detect inflammatory bowel disease (IBD) Dig Liver Dis 48:148-153 doi:10.1016/j.dld.2015.10.013

Arcangeli G, Mauro F, Morelli D, Nervi C (1979) Multiple daily fractionation in radiotherapy: Biological rationale and preliminary clinical experiences European Journal of Cancer (1965) 15:1077-1083 doi:https://doi.org/10.1016/0014-2964(79)90123-3

Argiris A, Karamouzis MV, Raben D, Ferris RL (2008) Head and neck cancer Lancet 371:1695-1709 doi:10.1016/S0140-6736(08)60728-X

Astudillo-de la Vega H, Alonso-Luna O, Ali-Perez J, Lopez-Camarillo C, Ruiz-Garcia E (2019) Oncobiome at the Forefront of a Novel Molecular Mechanism to Understand the Microbiome and Cancer Adv Exp Med Biol 1168:147-156 doi:10.1007/978-3-030-24100-1_10

Bannister M, Ah-See KW (2014) Is oropharyngeal cancer being misdiagnosed as acute tonsillitis? The British journal of general practice : the journal of the Royal College of General Practitioners 64:e742-e744 doi:10.3399/bjgp14X682537

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function Cell 116:281-297

Basheeth N, Patil N (2019) Biomarkers in Head and Neck Cancer an Update Indian J Otolaryngol Head Neck Surg 71:1002-1011 doi:10.1007/s12070-019-01683-1

Baskar R, Dai J, Wenlong N, Yeo R, Yeoh K-W (2014) Biological response of cancer cells to radiation treatment Frontiers in molecular biosciences 1:24-24 doi:10.3389/fmolb.2014.00024

Bayrakli I, Turkmen A, Akman H, Sezer MT, Kutluhan S (2016) Applications of external cavity diode laser-based technique to noninvasive clinical diagnosis using expired breath ammonia analysis: chronic kidney disease, epilepsy J Biomed Opt 21:87004 doi:10.1117/1.JBO.21.8.087004

Bean HD, Zhu J, Hill JE (2011) Characterizing bacterial volatiles using secondary electrospray ionization mass spectrometry (SESI-MS) J Vis Exp doi:10.3791/2664

Bellairs JA, Hasina R, Agrawal N (2017) Tumor DNA: an emerging biomarker in head and neck cancer Cancer Metastasis Rev 36:515-523 doi:10.1007/s10555-017-9685-x

Bellmunt À M, López-Puerto L, Lorente J, Closa D (2019) Involvement of extracellular vesicles in the macrophage-tumor cell communication in head and neck squamous cell carcinoma PLoS One 14:e0224710 doi:10.1371/journal.pone.0224710

Benchetrit L et al. (2019) Gender disparities in head and neck cancer chemotherapy clinical trials participation and treatment Oral Oncol 94:32-40 doi:10.1016/j.oraloncology.2019.05.009

Benninger MS, Enrique RR, Nichols RD (1993) Symptom-directed selective endoscopy and cost containment for evaluation of head and neck cancer Head Neck 15:532-536 doi:10.1002/hed.2880150610

Bentzen SM, Saunders MI, Dische S, Bond SJ (2001) Radiotherapy-related early morbidity in head and neck cancer: quantitative clinical radiobiology as deduced from the CHART trial Radiother Oncol 60:123-135 doi:10.1016/s0167-8140(01)00358-9

Bernier J, Bentzen SM (2006) Radiotherapy for head and neck cancer: latest developments and future perspectives Curr Opin Oncol 18:240-246 doi:10.1097/01.cco.0000219252.45467.88

Bersani C et al. (2018) MicroRNA-155, -185 and -193b as biomarkers in human papillomavirus positive and negative tonsillar and base of tongue squamous cell carcinoma Oral Oncol 82:8-16 doi:10.1016/j.oraloncology.2018.04.021

Berta E, Atallah I, Reyt E, Boyer E, Karkas A, Righini CA (2014) The role of tonsillectomy in the initial diagnostic work-up of head and neck squamous cell carcinoma of unknown primary Eur Ann Otorhinolaryngol Head Neck Dis 131:305-308 doi:10.1016/j.anorl.2014.03.005

Bertero T et al. (2013) Tumor suppressor function of miR-483-3p on squamous cell carcinomas due to its pro-apoptotic properties Cell Cycle 12:2183-2193 doi:10.4161/cc.25330

Beyer C et al. (2015) Signature of circulating microRNAs in osteoarthritis Ann Rheum Dis 74:e18 doi:10.1136/annrheumdis-2013-204698

Bhatt AP, Redinbo MR, Bultman SJ (2017) The role of the microbiome in cancer development and therapy CA Cancer J Clin 67:326-344 doi:10.3322/caac.21398

Bianchi F (2015) Molecular profile of liquid biopsies: next generation biomarkers to improve lung cancer treatment Ecancermedicalscience 9:598 doi:10.3332/ecancer.2015.598

Bik EM et al. (2010) Bacterial diversity in the oral cavity of 10 healthy individuals ISME J 4:962-974 doi:10.1038/ismej.2010.30

Binderup HG, Madsen JS, Heegaard NHH, Houlind K, Andersen RF, Brasen CL (2018) Quantification of microRNA levels in plasma - Impact of preanalytical and analytical conditions PLoS One 13:e0201069 doi:10.1371/journal.pone.0201069

Biomarkers Definitions Working G (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework Clinical pharmacology and therapeutics 69:89-95 doi:10.1067/mcp.2001.113989

Blanchard P et al. (2011) Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site Radiother Oncol 100:33-40 doi:10.1016/j.radonc.2011.05.036

Bock JB, Matern HT, Peden AA, Scheller RH (2001) A genomic perspective on membrane compartment organization Nature 409:839-841 doi:10.1038/35057024

Bolin TD, Genge JR, Duncombe VM, Soe A, Myo K (1996) Patterns of methane production in a Burmese (Myanmar) population J Gastroenterol Hepatol 11:71-76

Bonner JA et al. (2006) Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck N Engl J Med 354:567-578 doi:10.1056/NEJMoa053422

Bonney GE (1987) Logistic regression for dependent binary observations Biometrics 43:951-973

Bornigen D et al. (2017) Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer Sci Rep 7:17686 doi:10.1038/s41598-017-17795-z

Bourhis J et al. (2007) Individual patients' data meta-analyses in head and neck cancer Curr Opin Oncol 19:188-194 doi:10.1097/CCO.0b013e3280f01010

Braakhuis BJ, Leemans CR, Brakenhoff RH (2004) A genetic progression model of oral cancer: current evidence and clinical implications J Oral Pathol Med 33:317-322 doi:10.1111/j.1600-0714.2004.00225.x

Brennan JA et al. (1995) Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck N Engl J Med 332:712-717 doi:10.1056/NEJM199503163321104

Brown J, Shirey B (2001) A Tool for Selecting an Adsorbent for Thermal Desorption Applications - A Technical Report. Sigma-Aldrich-Supelco. https://www.sigmaaldrich.com/technical-documents/articles/analytical/purification/carbon-adsorbents-physical-characteristics.html.

Brown ML, Glanzmann C, Huber G, Bredell M, Rordorf T, Studer G (2016) IMRT/VMAT for malignancies in the head-and-neck region : Outcome in patients aged 80 Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft [et al] 192:526-536 doi:10.1007/s00066-016-0986-8

Broza YY, Vishinkin R, Barash O, Nakhleh MK, Haick H (2018) Synergy between nanomaterials and volatile organic compounds for non-invasive medical evaluation Chemical Society Reviews 47:4781-4859 doi:10.1039/C8CS00317C

Bujang MA, Adnan TH (2016) Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis J Clin Diagn Res 10:YE01-YE06 doi:10.7860/JCDR/2016/18129.8744

Bujanover Y, Peled Y, Blau H, Yahav J, Katzenelson D, Gilat T (1987) Methane production in patients with cystic fibrosis J Pediatr Gastroenterol Nutr 6:377-380

Bull SB (1993) Sample size and power determination for a binary outcome and an ordinal exposure when logistic regression analysis is planned Am J Epidemiol 137:676-684 doi:10.1093/oxfordjournals.aje.a116725

Bungum A, Jensen JS, Jakobsen KK, Christensen A, Grønhøj C, von Buchwald C (2020) Impact of surgical resection margins less than 5 mm in oral cavity squamous cell carcinoma: a systematic review Acta Otolaryngol:1-7 doi:10.1080/00016489.2020.1773532

Burd EM, Dean CL (2016) Human Papillomavirus Microbiology Spectrum 4 doi:UNSP DMIH2-0001-2015

10.1128/microbiolspec.DMIH2-0001-2015

Buszewski B, Ulanowska A, Ligor T, Denderz N, Amann A (2009) Analysis of exhaled breath from smokers, passive smokers and non-smokers by solid-phase microextraction gas chromatography/mass spectrometry Biomed Chromatogr 23:551-556 doi:10.1002/bmc.1141

Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning Nat Methods 15:233-234 doi:10.1038/nmeth.4642

Cadoni G et al. (2017) Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy Acta Otorhinolaryngol Ital 37:458-466 doi:10.14639/0392-100X-1246

Candela M, Guidotti M, Fabbri A, Brigidi P, Franceschi C, Fiorentini C (2011) Human intestinal microbiota: cross-talk with the host and its potential role in colorectal cancer Crit Rev Microbiol 37:1-14 doi:10.3109/1040841X.2010.501760

Cao WQ, Duan YX (2007) Current status of methods and techniques for breath analysis Crit Rev Anal Chem 37:3-13 doi:10.1080/10408340600976499

Cao, W., Liu, Z., Gokavarapu, S., Chen, Y., Yang, R., & Ji, T. (2016). Reformed smokers have survival benefits after head and neck cancer. *The British journal of oral & maxillofacial surgery*, *54*(7), 818–825. https://doi.org/10.1016/j.bjoms.2016.06.013

Carbonero F, Benefiel AC, Gaskins HR (2012) Contributions of the microbial hydrogen economy to colonic homeostasis Nat Rev Gastroenterol Hepatol 9:504-518 doi:10.1038/nrgastro.2012.85

Casanova MR, Azevedo-Silva J, Rodrigues LR, Preto A (2018) Colorectal Cancer Cells Increase the Production of Short Chain Fatty Acids by Propionibacterium freudenreichii Impacting on Cancer Cells Survival Front Nutr 5:44 doi:10.3389/fnut.2018.00044

Castellsague X et al. (2004) The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis Int J Cancer 108:741-749 doi:10.1002/ijc.11627

Cauchi M et al. (2015) Comparison of GC-MS, HPLC-MS and SIFT-MS in conjunction with multivariate classification for the diagnosis of Crohn's disease in urine Analytical Methods 7:8379-8385

Caulley L, Rodin D, Kilty S, Randolph G, Hunink MG, Shin JJ (2019) Evidence-Based Medicine in Otolaryngology Part 10: Cost-Effectiveness Analyses in Otolaryngology Otolaryngol Head Neck Surg 161:375-387 doi:10.1177/0194599819852104

Cen WN et al. (2018) The expression and biological information analysis of miR-375-3p in head and neck squamous cell carcinoma based on 1825 samples from GEO, TCGA, and peer-reviewed publications Pathol Res Pract 214:1835-1847 doi:10.1016/j.prp.2018.09.010

Chai L, Yuan Y, Chen C, Zhou J, Wu Y (2018) The role of long non-coding RNA ANRIL in the carcinogenesis of oral cancer by targeting miR-125a Biomed Pharmacother 103:38-45 doi:10.1016/j.biopha.2018.01.105

Chambers ES et al. (2015) Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults Gut 64:1744-1754 doi:10.1136/gutjnl-2014-307913

Chandran D, Ooi EH, Watson DI, Kholmurodova F, Jaenisch S, Yazbeck R (2019) The Use of Selected Ion Flow Tube-Mass Spectrometry Technology to Identify Breath Volatile Organic Compounds for the Detection of Head and Neck Squamous Cell Carcinoma: A Pilot Study Medicina (Kaunas) 55 doi:10.3390/medicina55060306

Chang AH, Parsonnet J (2010) Role of bacteria in oncogenesis Clin Microbiol Rev 23:837-857 doi:10.1128/CMR.00012-10

Chang AI, Schwertschkow AH, Nolta JA, Wu J (2015) Involvement of mesenchymal stem cells in cancer progression and metastases Current Cancer Drug Targets 15:88-98

Chang M-C et al. (2014) Areca Nut Components Affect COX-2, Cyclin B1/cdc25C and Keratin Expression, PGE2 Production in Keratinocyte Is Related to Reactive Oxygen Species, CYP1A1, Src, EGFR and Ras Signaling PLOS ONE 9:e101959 doi:10.1371/journal.pone.0101959

Chang YA et al. (2018) A Three-MicroRNA Signature as a Potential Biomarker for the Early Detection of Oral Cancer Int J Mol Sci 19 doi:10.3390/ijms19030758

Chemi F et al. (2019) Pulmonary venous circulating tumor cell dissemination before tumor resection and disease relapse Nat Med 25:1534-1539 doi:10.1038/s41591-019-0593-1

Chen AM et al. (2011) Tobacco smoking during radiation therapy for head-and-neck cancer is associated with unfavorable outcome International journal of radiation oncology, biology, physics 79:414-419 doi:10.1016/j.ijrobp.2009.10.050

Chen D, Cabay RJ, Jin Y, Wang A, Lu Y, Shah-Khan M, Zhou X (2013) MicroRNA Deregulations in Head and Neck Squamous Cell Carcinomas J Oral Maxillofac Res 4:e2 doi:10.5037/jomr.2013.4102

Chen JCT, Girvigian MR (2006) Stereotactic radiosurgery: indications and results - part 2 The Permanente journal 10:9-15 doi:10.7812/tpp/04-076

Chen JJ, Liu SX, Chen MZ, Zhao ZY (2015) HasmiR125a and 125b are induced by treatment with cisplatin in nasopharyngeal carcinoma and inhibit apoptosis in a p53dependent manner by targeting p53 mRNA Mol Med Rep 12:3569-3574 doi:10.3892/mmr.2015.3863

Chen K, Pan G (2019) Dysregulation of microRNA-106a-5p-RUNX1 axis associates with clinical progression and prognosis of osteosarcoma patients Pathol Res Pract 215:152686 doi:10.1016/j.prp.2019.152686

Chen M et al. (2018) miR-590-5p suppresses hepatocellular carcinoma chemoresistance by targeting YAP1 expression EBioMedicine 35:142-154 doi:10.1016/j.ebiom.2018.08.010

Chen S, Sun YY, Zhang ZX, Li YH, Xu ZM, Fu WN (2017) Transcriptional suppression of microRNA-27a contributes to laryngeal cancer differentiation via GSK-3beta-involved Wnt/beta-catenin pathway Oncotarget 8:14708-14718 doi:10.18632/oncotarget.14769

Chen X, Ishwaran H (2012) Random forests for genomic data analysis Genomics 99:323-329 doi:10.1016/j.ygeno.2012.04.003

Cheol Park G et al. (2017) 18 F-FDG PET/CT vs. human papillomavirus, p16 and Epstein-Barr virus detection in cervical metastatic lymph nodes for identifying primary tumors Int J Cancer 140:1405-1412 doi:10.1002/ijc.30550

Chiam K et al. (2020) Serum outperforms plasma in small extracellular vesicle microRNA biomarker studies of adenocarcinoma of the esophagus World J Gastroenterol 26:2570-2583 doi:10.3748/wjg.v26.i20.2570

Cho HJ et al. (2017) Low levels of circulating microRNA-26a/29a as poor prognostic markers in patients with hepatocellular carcinoma who underwent curative treatment Clin Res Hepatol Gastroenterol 41:181-189 doi:10.1016/j.clinre.2016.09.011

Choi SY, Kahyo H (1991) Effect of cigarette smoking and alcohol consumption in the aetiology of cancer of the oral cavity, pharynx and larynx Int J Epidemiol 20:878-885 doi:10.1093/ije/20.4.878

Choo JM, Leong LE, Rogers GB (2015) Sample storage conditions significantly influence faecal microbiome profiles Sci Rep 5:16350 doi:10.1038/srep16350

Chow LQM (2020) Head and Neck Cancer N Engl J Med 382:60-72 doi:10.1056/NEJMra1715715

Chung D, Keles S (2010) Sparse partial least squares classification for high dimensional data Stat Appl Genet Mol Biol 9:Article17 doi:10.2202/1544-6115.1492

Cianchetti M, Mancuso AA, Amdur RJ, Werning JW, Kirwan J, Morris CG, Mendenhall WM (2009) Diagnostic evaluation of squamous cell carcinoma metastatic to cervical lymph nodes from an unknown head and neck primary site Laryngoscope 119:2348-2354 doi:10.1002/lary.20638

Cinpolat O, Unal ZN, Ismi O, Gorur A, Unal M (2017) Comparison of microRNA profiles between benign and malignant salivary gland tumors in tissue, blood and saliva samples: a prospective, case-control study Braz J Otorhinolaryngol 83:276-284 doi:10.1016/j.bjorl.2016.03.013

Citron F et al. (2017) An Integrated Approach Identifies Mediators of Local Recurrence in Head and Neck Squamous Carcinoma Clin Cancer Res 23:3769-3780 doi:10.1158/1078-0432.CCR-16-2814

Cognetti DM, Weber RS, Lai SY (2008) Head and neck cancer: an evolving treatment paradigm Cancer 113:1911-1932 doi:10.1002/cncr.23654

Cohen LE, Morrison KA, Taylor E, Jin J, Spector JA, Caruana S, Rohde CH (2018a) Functional and Aesthetic Outcomes in Free Flap Reconstruction of Intraoral Defects With Lip-Split Versus Non-Lip-Split Incisions Ann Plast Surg 80:S150-S155 doi:10.1097/SAP.0000000000001373

Cohen N, Fedewa S, Chen AY (2018b) Epidemiology and Demographics of the Head and Neck Cancer Population Oral Maxillofac Surg Clin North Am 30:381-395 doi:10.1016/j.coms.2018.06.001

Cohen, E., Bell, R. B., Bifulco, C. B., Burtness, B., Gillison, M. L., Harrington, K. J., Le, Q. T., Lee, N. Y., Leidner, R., Lewis, R. L., Licitra, L., Mehanna, H., Mell, L. K., Raben, A., Sikora, A. G., Uppaluri, R., Whitworth, F., Zandberg, D. P., & Ferris, R. L. (2019). The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC). Journal for immunotherapy of cancer, 7(1), 184. https://doi.org/10.1186/s40425-019-0662-5

Coordes A, Lenz K, Qian X, Lenarz M, Kaufmann AM, Albers AE (2016) Meta-analysis of survival in patients with HNSCC discriminates risk depending on combined HPV and p16 status Eur Arch Otorhinolaryngol 273:2157-2169 doi:10.1007/s00405-015-3728-0

Curtis B (2015) “I can tell by the way you smell” The Senses and Society 3:5-22 doi:10.2752/174589308x266434

D'Cruz AK et al. (2009) Elective neck dissection for the management of the N0 neck in early cancer of the oral tongue: need for a randomized controlled trial Head Neck 31:618-624 doi:10.1002/hed.20988

Dahlstrom KR, Li G, Hussey CS, Vo JT, Wei Q, Zhao C, Sturgis EM (2015) Circulating human papillomavirus DNA as a marker for disease extent and recurrence among patients with oropharyngeal cancer Cancer 121:3455-3464 doi:10.1002/cncr.29538

Das MK, Bishwal SC, Das A, Dabral D, Varshney A, Badireddy VK, Nanda R (2014) Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS Anal Chem 86:1229-1237 doi:10.1021/ac403541a

Dassi E et al. (2014) Enhanced microbial diversity in the saliva microbiome induced by short-term probiotic intake revealed by 16S rRNA sequencing on the IonTorrent PGM platform J Biotechnol 190:30-39 doi:10.1016/j.jbiotec.2014.03.024

Dautremont JF et al. (2013) Cost-effectiveness analysis of a postoperative clinical care pathway in head and neck surgery with microvascular reconstruction J Otolaryngol Head Neck Surg 42:59 doi:10.1186/1916-0216-42-59

Dauzier E et al. (2019) Role of chemotherapy in 5000 patients with head and neck cancer treated by curative surgery: A subgroup analysis of the meta-analysis of chemotherapy in head and neck cancer Oral Oncol 95:106-114 doi:10.1016/j.oraloncology.2019.06.001

De Felice F et al. (2018) Radiotherapy Controversies and Prospective in Head and Neck Cancer: A Literature-Based Critical Review Neoplasia 20:227-232 doi:https://doi.org/10.1016/j.neo.2018.01.002

de Lacy Costello B et al. (2014) A review of the volatiles from the healthy human body J Breath Res 8:014001 doi:10.1088/1752-7155/8/1/014001

De Virgilio A et al. (2020) Transoral robotic surgery and intensity-modulated radiotherapy in the treatment of the oropharyngeal carcinoma: a systematic review and meta-analysis Eur Arch Otorhinolaryngol doi:10.1007/s00405-020-06224-z

Demigne C, Remesy C (1985) Stimulation of absorption of volatile fatty acids and minerals in the cecum of rats adapted to a very high fiber diet J Nutr 115:53-60 doi:10.1093/jn/115.1.53

Deschler DG, Richmon JD, Khariwala SS, Ferris RL, Wang MB (2014) The "new" head and neck cancer patient-young, nonsmoker, nondrinker, and HPV positive: evaluation Otolaryngol Head Neck Surg 151:375-380 doi:10.1177/0194599814538605

Dewhirst FE et al. (2010) The human oral microbiome J Bacteriol 192:5002-5017 doi:10.1128/JB.00542-10

Dharmawardana N, Goddard T, Woods C, Watson DI, Butler R, Ooi EH, Yazbeck R (2020a) Breath methane to hydrogen ratio as a surrogate marker of intestinal dysbiosis in head and neck cancer Sci Rep 10:1-8

Dharmawardana N, Goddard T, Woods C, Watson DI, Ooi EH, Yazbeck R (2020b) Development of a non-invasive exhaled breath test for the diagnosis of head and neck cancer British Journal of Cancer:1-7

Dharmawardana N, Ooi EH, Woods C, Hussey D (2019) Circulating microRNAs in head and neck cancer: a scoping review of methods Clin Exp Metastasis 36:291-302 doi:10.1007/s10585-019-09961-6

Dharmawardana N, Woods C, Watson DI, Yazbeck R, Ooi EH (2020c) A review of breath analysis techniques in head and neck cancer Oral Oncol 104:104654 doi:10.1016/j.oraloncology.2020.104654

Di Agostino S et al. (2018) Long Non-coding MIR205HG Depletes Hsa-miR-590-3p Leading to Unrestrained Proliferation in Head and Neck Squamous Cell Carcinoma Theranostics 8:1850-1868 doi:10.7150/thno.22167

Di Ciaccio A, Coli M, Angulo Ibañez JM (2012) Advanced statistical methods for the analysis of large data-sets. Studies in theoretical and applied statistics. Springer, Berlin ; New York

Di Lorenzo C, Dooley CP, Valenzuela JE (1991) Role of fasting gastrointestinal motility in the variability of gastrointestinal transit time assessed by hydrogen breath test Gut 32:1127-1130 doi:10.1136/gut.32.10.1127

Dias R, Torkamani A (2019) Artificial intelligence in clinical and genomic diagnostics Genome Med 11:70 doi:10.1186/s13073-019-0689-8

Dolan RW, Vaughan CW, Fuleihan N (1998) Symptoms in early head and neck cancer: an inadequate indicator Otolaryngol Head Neck Surg 119:463-467 doi:10.1016/s0194-5998(98)70102-0

Dong Y et al. (2013) Microfluidics and circulating tumor cells J Mol Diagn 15:149-157 doi:10.1016/j.jmoldx.2012.09.004

Duan ZQ, Shi JD, Wu MN, Hu NZ, Hu YZ (2016) Influence of miR-30b regulating humoral immune response by genetic difference Immunol Res 64:181-190 doi:10.1007/s12026-015-8736-z

Duray A, Demoulin S, Hubert P, Delvenne P, Saussez S (2010) Immune suppression in head and neck cancers: a review Clin Dev Immunol 2010:701657-701657 doi:10.1155/2010/701657

Dzutsev A, Goldszmid RS, Viaud S, Zitvogel L, Trinchieri G (2015) The role of the microbiota in inflammation, carcinogenesis, and cancer therapy Eur J Immunol 45:17-31 doi:10.1002/eji.201444972

Economopoulou P, Kotsantis I, Kyrodimos E, Lianidou ES, Psyrri A (2017) Liquid biopsy: An emerging prognostic and predictive tool in Head and Neck Squamous Cell Carcinoma (HNSCC). Focus on Circulating Tumor Cells (CTCs) Oral Oncology 74:83-89 doi:https://doi.org/10.1016/j.oraloncology.2017.09.012

Economopoulou P, Kotsantis I, Psyrri A (2020) Special Issue about Head and Neck Cancers: HPV Positive Cancers International journal of molecular sciences 21:3388 doi:10.3390/ijms21093388

Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM Ann Surg Oncol 17:1471-1474 doi:10.1245/s10434-010-0985-4

Egyud M et al. (2019) Plasma circulating tumor DNA as a potential tool for disease monitoring in head and neck cancer Head & neck 41:1351-1358 doi:https://dx.doi.org/10.1002/hed.25563

Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA (2013) Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms Nat Rev Cancer 13:759-771 doi:10.1038/nrc3611

Elmassry MM, Piechulla B (2020) Volatilomes of Bacterial Infections in Humans Front Neurosci 14:257 doi:10.3389/fnins.2020.00257

Elrefaey S, Massaro MA, Chiocca S, Chiesa F, Ansarin M (2014) HPV in oropharyngeal cancer: the basics to know in clinical practice Acta Otorhinolaryngol Ital 34:299-309

Engen PA, Green SJ, Voigt RM, Forsyth CB, Keshavarzian A (2015) The Gastrointestinal Microbiome: Alcohol Effects on the Composition of Intestinal Microbiota Alcohol Res 37:223-236

Fabris F, Doherty A, Palmer D, de Magalhães JP, Freitas AA (2018) A new approach for interpreting Random Forest models and its application to the biology of ageing Bioinformatics 34:2449-2456 doi:10.1093/bioinformatics/bty087

Faisal M et al. (2018) Depth of invasion (DOI) as a predictor of cervical nodal metastasis and local recurrence in early stage squamous cell carcinoma of oral tongue (ESSCOT) PLoS One 13:e0202632 doi:10.1371/journal.pone.0202632

Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR (2008) Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood Proc Natl Acad Sci U S A 105:16266-16271 doi:10.1073/pnas.0808319105

Fan L, Qi H, Teng J, Su B, Chen H, Wang C, Xia Q (2016) Identification of serum miRNAs by nano-quantum dots microarray as diagnostic biomarkers for early detection of non-small cell lung cancer Tumour Biol 37:7777-7784 doi:10.1007/s13277-015-4608-3

Fan S et al. (2015) miR-483-5p determines mitochondrial fission and cisplatin sensitivity in tongue squamous cell carcinoma by targeting FIS1 Cancer Lett 362:183-191 doi:10.1016/j.canlet.2015.03.045

Fang CH, Friedman R, White PE, Mady LJ, Kalyoussef E (2015) Emergent Awake tracheostomy--The five-year experience at an urban tertiary care center Laryngoscope 125:2476-2479 doi:10.1002/lary.25348

Feng S, Farha F, Li Q, Wan Y, Xu Y, Zhang T, Ning H (2019) Review on Smart Gas Sensing Technology Sensors (Basel) 19:3760 doi:10.3390/s19173760

Ferlay J et al. (2018) Global Cancer Observatory: Cancer Today. . https://gco.iarc.fr/today. Accessed 24/01/2020 2020

Ferrandino G et al. (2020) Breath Biopsy Assessment of Liver Disease Using an Exogenous Volatile Organic Compound-Toward Improved Detection of Liver Impairment Clin Transl Gastroenterol 11:e00239-e00239 doi:10.14309/ctg.0000000000000239

Fischbach W, Meyer T, Barthel K (1990) Squamous cell carcinoma antigen in the diagnosis and treatment follow-up of oral and facial squamous cell carcinoma Cancer 65:1321-1324 doi:10.1002/1097-0142(19900315)65:6<1321::aid-cncr2820650612>3.0.co;2-y

Fischer, S., Bergmann, A., Steffens, M., Trefz, P., Ziller, M., Miekisch, W., Schubert, J. S., Köhler, H., & Reinhold, P. (2015). Impact of food intake on in vivo VOC concentrations in exhaled breath assessed in a caprine animal model. *Journal of breath research*, *9*(4), 047113. https://doi.org/10.1088/1752-7155/9/4/047113

Fu X et al. (2011) Prognostic role of microRNA-21 in various carcinomas: a systematic review and meta-analysis Eur J Clin Invest 41:1245-1253 doi:10.1111/j.1365-2362.2011.02535.x

Fujiwara K et al. (2005) Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer Clin Cancer Res 11:1219-1225

Fukumoto I et al. (2015) MicroRNA expression signature of oral squamous cell carcinoma: functional role of microRNA-26a/b in the modulation of novel cancer pathways Br J Cancer 112:891-900 doi:10.1038/bjc.2015.19

Furquim CP et al. (2017a) The Salivary Microbiome and Oral Cancer Risk: a Pilot Study in Fanconi Anemia J Dent Res 96:292-299 doi:10.1177/0022034516678169

Furquim CP et al. (2017b) The Salivary Microbiome and Oral Cancer Risk: A Pilot Study in Fanconi Anemia Journal of Dental Research 96:292-299 doi:10.1177/0022034516678169

Gaci N, Borrel G, Tottey W, O'Toole PW, Brugere JF (2014) Archaea and the human gut: new beginning of an old story World J Gastroenterol 20:16062-16078 doi:10.3748/wjg.v20.i43.16062

Galot R, Machiels JH (2020) Current applications and challenges of circulating tumor DNA (ctDNA) in squamous cell carcinoma of the head and neck (SCCHN) Cancer Treat Rev 85:101992 doi:10.1016/j.ctrv.2020.101992

Gao J, Yu SR, Yuan Y, Zhang LL, Lu JW, Feng JF, Hu SN (2019) MicroRNA-590-5p functions as a tumor suppressor in breast cancer conferring inhibitory effects on cell migration, invasion, and epithelial-mesenchymal transition by

downregulating the Wnt-β-catenin signaling pathway J Cell Physiol 234:1827-1841 doi:10.1002/jcp.27056

García RA, Morales V, Martín S, Vilches E, Toledano A (2013) Volatile Organic Compounds Analysis in Breath Air in Healthy Volunteers and Patients Suffering Epidermoid Laryngeal Carcinomas Chromatographia 77:501-509 doi:10.1007/s10337-013-2611-7

Gavan SP, Thompson AJ, Payne K (2018) The economic case for precision medicine Expert Rev Precis Med Drug Dev 3:1-9 doi:10.1080/23808993.2018.1421858

Gebicki J, Bylinski H, Namiesnik J (2016) Measurement techniques for assessing the olfactory impact of municipal sewage treatment plants Environ Monit Assess 188:32 doi:10.1007/s10661-015-5024-2

Ghanbari R et al. (2015) Downregulation of Plasma MiR-142-3p and MiR-26a-5p in Patients With Colorectal Carcinoma Iran J Cancer Prev 8:e2329 doi:10.17795/ijcp2329

Gholizadeh P, Eslami H, Yousefi M, Asgharzadeh M, Aghazadeh M, Kafil HS (2016) Role of oral microbiome on oral cancers, a review Biomedicine & Pharmacotherapy 84:552-558 doi:10.1016/j.biopha.2016.09.082

Ghosh S, Kim K-H, Sohn JR (2011) Some insights into analytical bias involved in the application of grab sampling for volatile organic compounds: a case study against used Tedlar bags ScientificWorldJournal 11:2160-2177 doi:10.1100/2011/529532

Gibson A, Malek L, Dekker RFH, Ross B (2015) Detecting volatile compounds from Kraft lignin degradation in the headspace of microbial cultures by selected ion flow tube mass spectrometry (SIFT-MS) Journal of microbiological methods 112:40-45 doi:10.1016/j.mimet.2015.03.008

Gilbert ES (2009) Ionising radiation and cancer risks: what have we learned from epidemiology? Int J Radiat Biol 85:467-482 doi:10.1080/09553000902883836

Gillison ML (2004) Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity

Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C (2015) Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma J Clin Oncol 33:3235-3242 doi:10.1200/JCO.2015.61.6995

Glinge C et al. (2017) Stability of Circulating Blood-Based MicroRNAs - Pre-Analytic Methodological Considerations PLoS One 12:e0167969 doi:10.1371/journal.pone.0167969

Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA (2018) The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy Cancer Cell 33:570-580 doi:10.1016/j.ccell.2018.03.015

Gopinath D, Menon RK (2018) Comments on "Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer" by Yang et al Oral Oncol 78:216-217 doi:10.1016/j.oraloncology.2018.02.007

Gordon RT et al. (2008) The use of canines in the detection of human cancers J Altern Complement Med 14:61-67 doi:10.1089/acm.2006.6408

Gorges TM, Pantel K (2013) Circulating tumor cells as therapy-related biomarkers in cancer patients Cancer Immunol Immunother 62:931-939 doi:10.1007/s00262-012-1387-1

Gottlieb K, Le C, Wacher V, Sliman J, Cruz C, Porter T, Carter S (2017) Selection of a cut-off for high- and low-methane producers using a spot-methane breath test: results from a large north American dataset of hydrogen, methane and carbon dioxide measurements in breath Gastroenterol Rep (Oxf) 5:193-199 doi:10.1093/gastro/gow048

Gourzones C et al. (2013) Consistent high concentration of the viral microRNA BART17 in plasma samples from nasopharyngeal carcinoma patients - Evidence of non-exosomal transport Virology Journal 10 doi:10.1186/1743-422X-10-119

Gourzones C et al. (2010) Extra-cellular release and blood diffusion of BART viral micro-RNAs produced by EBV-infected nasopharyngeal carcinoma cells Virol J 7:271 doi:10.1186/1743-422x-7-271

Gray V (2017) Principal component analysis : methods, applications, and technology. Mathematics research developments. Novinka, New York

Gregoire V et al. (2014) Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines Radiother Oncol 110:172-181 doi:10.1016/j.radonc.2013.10.010

Greither T et al. (2017) Salivary miR-93 and miR-200a as post-radiotherapy biomarkers in head and neck squamous cell carcinoma Oncol Rep 38:1268-1275 doi:10.3892/or.2017.5764

Gromadzińska J, Reszka E, Bruzelius K, Wąsowicz W, Åkesson B (2008) Selenium and cancer: biomarkers of selenium status and molecular action of selenium supplements European Journal of Nutrition 47:29-50 doi:10.1007/s00394-008-2005-z

Grønhøj C et al. (2019) Impact on survival of tobacco smoking for cases with oropharyngeal squamous cell carcinoma and known human papillomavirus and p16-status: a multicenter retrospective study Oncotarget 10:4655-4663 doi:10.18632/oncotarget.27079

Gronhoj Larsen C et al. (2014) Correlation between human papillomavirus and p16 overexpression in oropharyngeal tumours: a systematic review Br J Cancer 110:1587-1594 doi:10.1038/bjc.2014.42

Grove D, Braun P, Paschke K, Dweik RA (2012) A comparison of volatile organic compounds from blood headspace to exhaled breath American Journal of Respiratory and Critical Care Medicine 185

Gruber M et al. (2014) Analysis of exhaled breath for diagnosing head and neck squamous cell carcinoma: a feasibility study Br J Cancer 111:790-798 doi:10.1038/bjc.2014.361

Guerrero-Preston R et al. (2016) 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment Oncotarget 7:51320-51334 doi:10.18632/oncotarget.9710

Guerrero-Preston R et al. (2017) High-resolution microbiome profiling uncovers Fusobacterium nucleatum, Lactobacillus gasseri/johnsonii, and Lactobacillus vaginalis associated to oral and oropharyngeal cancer in saliva from HPV positive and HPV negative patients treated with surgery and chemo-radiation Oncotarget 8:110931-110948 doi:10.18632/oncotarget.20677

Guo Y, An R, Zhao R, Sun Y, Liu M, Tian L (2016) miR-375 exhibits a more effective tumor-suppressor function in laryngeal squamous carcinoma cells by regulating KLF4 expression compared with simple co-transfection of miR-375 and miR-206 Oncol Rep 36:952-960 doi:10.3892/or.2016.4852

Gupta PC, Warnakulasuriya S (2002) Global epidemiology of areca nut usage Addict Biol 7:77-83 doi:10.1080/13556210020091437

Gustafsson H, Franzén L, Grankvist K, Anniko M, Henriksson R (1988) Glycoprotein tumour markers in head and neck neoplasms--a consecutive study on CA-50, CA 19-9, and CEA Journal of cancer research and clinical oncology 114:394-398 doi:10.1007/bf02128184

Ha NH et al. (2015) Prolonged and repetitive exposure to Porphyromonas gingivalis increases aggressiveness of oral cancer cells by promoting acquisition of cancer stem cell properties Tumour Biol 36:9947-9960 doi:10.1007/s13277-015-3764-9

Haddad RI, Shin DM (2008) Recent advances in head and neck cancer N Engl J Med 359:1143-1154 doi:10.1056/NEJMra0707975

Haines A, Metz G, Dilawari J, Blendis L, Wiggins H (1977) Breath-methane in patients with cancer of the large bowel Lancet 2:481-483

Hakim M et al. (2011) Diagnosis of head-and-neck cancer from exhaled breath Br J Cancer 104:1649-1655 doi:10.1038/bjc.2011.128

Hammerlid E et al. (1998) Malnutrition and food intake in relation to quality of life in head and neck cancer patients Head Neck 20:540-548

Hanna GB, Boshier PR, Markar SR, Romano A (2019) Accuracy and Methodologic Challenges of Volatile Organic Compound-Based Exhaled Breath Tests for Cancer Diagnosis A Systematic Review and Meta-analysis Jama Oncology 5 doi:ARTN e182815 10.1001/jamaoncol.2018.2815

Harréus U (2013) Surgical errors and risks - the head and neck cancer patient GMS Curr Top Otorhinolaryngol Head Neck Surg 12:Doc04-Doc04 doi:10.3205/cto000096

Harris MS, Phillips DR, Sayer JL, Moore MG (2013) A comparison of community-based and hospital-based head and neck cancer screening campaigns: identifying high-risk individuals and early disease JAMA Otolaryngol Head Neck Surg 139:568-573 doi:10.1001/jamaoto.2013.3153

Hartwig S, Raguse JD, Pfitzner D, Preissner R, Paris S, Preissner S (2017) Volatile Organic Compounds in the Breath of Oral Squamous Cell Carcinoma Patients: A Pilot Study Otolaryngol Head Neck Surg 157:981-987 doi:10.1177/0194599817711411

Hashim D et al. (2016) The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium Ann Oncol 27:1619-1625 doi:10.1093/annonc/mdw224

Hattori N, Ushijima T (2016) Epigenetic impact of infection on carcinogenesis: mechanisms and applications Genome Med 8:10 doi:10.1186/s13073-016-0267-2

Hauser S et al. (2012) Analysis of serum microRNAs (miR-26a-2*, miR-191, miR-337-3p and miR-378) as potential biomarkers in renal cell carcinoma Cancer Epidemiol 36:391-394 doi:10.1016/j.canep.2012.04.001

He Q et al. (2016a) microRNA-21 and microRNA-375 from oral cytology as biomarkers for oral tongue cancer detection Oral Oncol 57:15-20 doi:10.1016/j.oraloncology.2016.03.017

He XY, Liao YD, Guo XQ, Wang R, Xiao ZY, Wang YG (2016b) Prognostic Role of microRNA-21 Expression in Brain Tumors: a Meta-analysis Mol Neurobiol 53:1856-1861 doi:10.1007/s12035-015-9140-3

He Y et al. (2015) Current State of Circulating MicroRNAs as Cancer Biomarkers Clin Chem 61:1138-1155 doi:10.1373/clinchem.2015.241190

Hebner CM, Laimins LA (2006) Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity Rev Med Virol 16:83-97 doi:10.1002/rmv.488

Hellings P, Jorissen M, Ceuppens JL (2000) The Waldeyer's ring Acta oto-rhino-laryngologica Belgica 54:237-241

Henken FE et al. (2011) PIK3CA-mediated PI3-kinase signalling is essential for HPV-induced transformation in vitro Mol Cancer 10:71 doi:10.1186/1476-4598-10-71

Herbig J, Beauchamp J (2014) Towards standardization in the analysis of breath gas volatiles J Breath Res 8:037101 doi:10.1088/1752-7155/8/3/037101

Ho WJ, Mehra R (2019) Pembrolizumab for the treatment of head and neck squamous cell cancer Expert Opin Biol Ther 19:879-885 doi:10.1080/14712598.2019.1644315

Hobbs CG, Sterne JA, Bailey M, Heyderman RS, Birchall MA, Thomas SJ (2006) Human papillomavirus and head and neck cancer: a systematic review and meta-analysis Clin Otolaryngol 31:259-266 doi:10.1111/j.1749-4486.2006.01246.x

Hocking J, Mithraprabhu S, Kalff A, Spencer A (2016) Liquid biopsies for liquid tumors: emerging potential of circulating free nucleic acid evaluation for the management of hematologic malignancies Cancer Biol Med 13:215-225 doi:10.20892/j.issn.2095-3941.2016.0025

Hofsli E et al. (2013) Identification of serum microRNA profiles in colon cancer Br J Cancer 108:1712-1719 doi:10.1038/bjc.2013.121

Holdenrieder S (2016) Liquid Profiling of Circulating Nucleic Acids as a Novel Tool for the Management of Cancer Patients Adv Exp Med Biol 924:53-60 doi:10.1007/978-3-319-42044-8_11

Holma R et al. (2013) Colonic methane production modifies gastrointestinal toxicity associated with adjuvant 5-fluorouracil chemotherapy for colorectal cancer J Clin Gastroenterol 47:45-51 doi:10.1097/MCG.0b013e3182680201

Holsti LR (1995) Development Of Clinical Radiotherapy Since 1896 Acta Oncologica 34:995-1003 doi:10.3109/02841869509127225

Homer JJ, Fardy MJ (2016) Surgery in head and neck cancer: United Kingdom National Multidisciplinary Guidelines J Laryngol Otol 130:S68-S70 doi:10.1017/S0022215116000475

Horváth, I., Hunt, J., Barnes, P. J., Alving, K., Antczak, A., Baraldi, E., Becher, G., van Beurden, W. J., Corradi, M., Dekhuijzen, R., Dweik, R. A., Dwyer, T., Effros,

R., Erzurum, S., Gaston, B., Gessner, C., Greening, A., Ho, L. P., Hohlfeld, J., Jöbsis, Q., … ATS/ERS Task Force on Exhaled Breath Condensate (2005). Exhaled breath condensate: methodological recommendations and unresolved questions. *The European respiratory journal*, *26*(3), 523–548. https://doi.org/10.1183/09031936.05.00029705

Horváth I et al. (2017) A European Respiratory Society technical standard: exhaled biomarkers in lung disease European Respiratory Journal 49:1600965 doi:10.1183/13993003.00965-2016

Hou B et al. (2015a) Circulating microRNAs as novel prognosis biomarkers for head and neck squamous cell carcinoma Cancer biology & therapy 16:1042-1046 doi:10.1080/15384047.2015.1045692

Hou B et al. (2015b) Circulating microRNAs as novel prognosis biomarkers for head and neck squamous cell carcinoma Cancer Biology and Therapy 16:1042-1046 doi:10.1080/15384047.2015.1045692

House R et al. (2018) Smoking-induced control of miR-133a-3p alters the expression of EGFR and HuR in HPV-infected oropharyngeal cancer PLoS One 13:e0205077 doi:10.1371/journal.pone.0205077

Howaldt HP, Kainz M, Euler B, Vorast H (1999) Proposal for modification of the TNM staging classification for cancer of the oral cavity. DOSAK J Craniomaxillofac Surg 27:275-288 doi:10.1054/jcms.1999.0070

Hsu CM, Lin PM, Wang YM, Chen ZJ, Lin SF, Yang MY (2012) Circulating miRNA is a novel marker for head and neck squamous cell carcinoma Tumour Biol 33:1933-1942 doi:10.1007/s13277-012-0454-8

Hu X, Jiang H, Jiang X (2017) Downregulation of lncRNA ANRIL inhibits proliferation, induces apoptosis, and enhances radiosensitivity in nasopharyngeal carcinoma cells through regulating miR-125a Cancer Biol Ther 18:331-338 doi:10.1080/15384047.2017.1310348

Huang C-C et al. (2017) Investigating the Association between Alcohol and Risk of Head and Neck Cancer in Taiwan Sci Rep 7:9701-9701 doi:10.1038/s41598-017-08802-4

Huang C, Song H, Lai L (2019) The role and mechanism of microRNA-18a-5p in oral squamous cell carcinoma Mol Med Rep 20:1637-1644 doi:10.3892/mmr.2019.10403

Huang S, Li X, Zhu H (2016) MicroRNA-152 Targets Phosphatase and Tensin Homolog to Inhibit Apoptosis and Promote Cell Migration of Nasopharyngeal Carcinoma Cells Med Sci Monit 22:4330-4337 doi:10.12659/MSM.898110

Huang SH, O'Sullivan B (2017) Overview of the 8th Edition TNM Classification for Head and Neck Cancer Curr Treat Options Oncol 18:40 doi:10.1007/s11864-017-0484-y

Huang SK, Hoon DS (2016) Liquid biopsy utility for the surveillance of cutaneous malignant melanoma patients Mol Oncol 10:450-463 doi:10.1016/j.molonc.2015.12.008

Huang Y, Tan D, Xiao J, Li Q, Zhang X, Luo Z (2018a) miR-150 contributes to the radioresistance in nasopharyngeal carcinoma cells by targeting glycogen synthase kinase-3beta J Cancer Res Ther 14:111-118 doi:10.4103/jcrt.JCRT_682_17

Huang Y et al. (2018b) MicroRNA-222 Promotes Invasion and Metastasis of Papillary Thyroid Cancer Through Targeting Protein Phosphatase 2 Regulatory Subunit B Alpha Expression Thyroid 28:1162-1173 doi:10.1089/thy.2017.0665

Husson Fo, Lê Sb, Pagès Jrm (2011) Exploratory multivariate analysis by example using R. Chapman & Hall/CRC computer science and data analysis. CRC Press, Boca Raton

Huttmann EM et al. (2011) Comparison of two devices and two breathing patterns for exhaled breath condensate sampling PLoS One 6:e27467 doi:10.1371/journal.pone.0027467

Huttmann EM et al. (2016) Correction: Comparison of Two Devices and Two Breathing Patterns for Exhaled Breath Condensate Sampling PLoS One 11:e0152620 doi:10.1371/journal.pone.0152620

Imai R et al. (2015) Prognostic significance of serum squamous cell carcinoma antigen in patients with head and neck cancer Acta oto-laryngologica 135:295-301 doi:10.3109/00016489.2014.951454

Independent-Hospital-Pricing-Authority (2014) National Hospital Cost Data Collection vol 18. https://www.ihpa.gov.au/

International Consortium for Outcome Research in H et al. (2014) Primary tumor staging for oral cancer and a proposed modification incorporating depth of invasion: an international multicenter retrospective study JAMA Otolaryngol Head Neck Surg 140:1138-1148 doi:10.1001/jamaoto.2014.1548

Iqbal H, Pan Q (2016) Capecitabine for treating head and neck cancer Expert Opin Investig Drugs 25:851-859 doi:10.1080/13543784.2016.1181747

Ivaldi E et al. (2019) Postoperative radiotherapy (PORT) for early oral cavity cancer (pT1-2,N0-1): A review Crit Rev Oncol Hematol 143:67-75 doi:10.1016/j.critrevonc.2019.08.003

Jang DW, Teng MS, Ojo B, Genden EM (2013) Palliative surgery for head and neck cancer with extensive skin involvement Laryngoscope 123:1173-1177 doi:10.1002/lary.23657

Jang SI, Kim JH, Youn YH, Park H, Lee SI, Conklin JL (2010) Relationship between intestinal gas and the development of right colonic diverticula J Neurogastroenterol Motil 16:418-423 doi:10.5056/jnm.2010.16.4.418

Jeffries J, Zhou W, Hsu AY, Deng Q (2019) miRNA-223 at the crossroads of inflammation and cancer Cancer Lett 451:136-141 doi:10.1016/j.canlet.2019.02.051

Jethwa AR, Khariwala SS (2017) Tobacco-related carcinogenesis in head and neck cancer Cancer Metastasis Rev 36:411-423 doi:10.1007/s10555-017-9689-6

Jiang F, Zhao W, Zhou L, Zhang L, Liu Z, Yu D (2014) miR-222 regulates the cell biological behavior of oral squamous cell carcinoma by targeting PUMA Oncol Rep 31:1255-1262 doi:10.3892/or.2014.2985

Jiang Q, Wang G, Jin S, Li Y, Wang Y (2013) Predicting human microRNA-disease associations based on support vector machine International Journal of Data Mining and Bioinformatics 8:282-293 doi:10.1504/IJDMB.2013.056078

Jiang R, Zhao C, Gao B, Xu J, Song W, Shi P (2018) Mixomics analysis of breast cancer: Long non-coding RNA linc01561 acts as ceRNA involved in the progression of breast cancer Int J Biochem Cell Biol 102:1-9 doi:10.1016/j.biocel.2018.06.003

Jiang WW et al. (2005) Increased mitochondrial DNA content in saliva associated with head and neck cancer Clinical Cancer Research 11:2486-2491 doi:Doi 10.1158/1078-0432.Ccr-04-2147

Jorge K et al. (2017) Characterization of MicroRNA Expression Profiles and Identification of Potential Biomarkers in Leprosy J Clin Microbiol 55:1516-1525 doi:10.1128/JCM.02408-16

Kao SS-T, Marshall-Webb M, Dharmawardana N, Foreman A, Ooi EH (2018) Gastrostomy tube insertion outcomes in South Australian head and neck cancer patients Australian Journal of Otolaryngology 1

Kariya A, Furusawa Y, Yunoki T, Kondo T, Tabuchi Y (2014) A microRNA-27a mimic sensitizes human oral squamous cell carcinoma HSC-4 cells to hyperthermia through downregulation of Hsp110 and Hsp90 Int J Mol Med 34:334-340 doi:10.3892/ijmm.2014.1758

Karlin DA, Mastromarino AJ, Jones RD, Stroehlein JR, Lorentz O (1985) Fecal skatole and indole and breath methane and hydrogen in patients with large bowel polyps or cancer J Cancer Res Clin Oncol 109:135-141

Kato H, Miyauchi F, Morioka H, Fujino T, Torigoe T (1979) Tumor antigen of human cervical squamous cell carcinoma: correlation of circulating levels with disease progress Cancer 43:585-590 doi:10.1002/1097-0142(197902)43:2<585::aid-cncr2820430227>3.0.co;2-0

Katz J, Onate MD, Pauley KM, Bhattacharyya I, Cha S (2011) Presence of Porphyromonas gingivalis in gingival squamous cell carcinoma Int J Oral Sci 3:209-215 doi:10.4248/IJOS11075

Kawakita A, Yanamoto S, Yamada S, Naruse T, Takahashi H, Kawasaki G, Umeda M (2014) MicroRNA-21 promotes oral cancer invasion via the Wnt/beta-catenin pathway by targeting DKK2 Pathol Oncol Res 20:253-261 doi:10.1007/s12253-013-9689-y

Kerlin P, Wong L (1988) Breath hydrogen testing in bacterial overgrowth of the small intestine Gastroenterology 95:982-988

Khan KS, Hayes I, Buggy DJ (2014) Pharmacology of anaesthetic agents II: inhalation anaesthetic agents Continuing Education in Anaesthesia Critical Care & Pain 14:106-111 doi:10.1093/bjaceaccp/mkt038

Khan N, Tomar SL (2020) The incidence of oropharyngeal cancer and rate of human papillomavirus vaccination coverage in Florida, 2011 through 2015 J Am Dent Assoc 151:51-58 doi:10.1016/j.adaj.2019.08.022

Khandelwal A, Seam RK, Gupta M, Rana MK, Prakash H, Vasquez KM, Jain A (2020) Circulating microRNA-590-5p functions as a liquid biopsy marker in non-small cell lung cancer Cancer Sci 111:826-839 doi:10.1111/cas.14199

Kim CW et al. (2018a) Hypoxia-induced microRNA-590-5p promotes colorectal cancer progression by modulating matrix metalloproteinase activity Cancer Lett 416:31-41 doi:10.1016/j.canlet.2017.12.018

Kim KY, McShane LM, Conley BA (2014) Designing biomarker studies for head and neck cancer Head & neck 36:1069-1075 doi:10.1002/hed.23444

Kim L, King T, Agulnik M (2010) Head and neck cancer: changing epidemiology and public health implications Oncology (Williston Park) 24:915-919, 924

Kim SS et al. (2018b) Plasma MicroRNA-21, 26a, and 29a-3p as Predictive Markers for Treatment Response Following Transarterial Chemoembolization in Patients with Hepatocellular Carcinoma J Korean Med Sci 33:e6 doi:10.3346/jkms.2018.33.e6

Kinoshita T et al. (2012) Tumor suppressive microRNA-133a regulates novel targets: moesin contributes to cancer cell proliferation and invasion in head and neck squamous cell carcinoma Biochem Biophys Res Commun 418:378-383 doi:10.1016/j.bbrc.2012.01.030

Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, Reid G (2011) Haemolysis during sample preparation alters microRNA content of plasma PLoS One 6:e24145 doi:10.1371/journal.pone.0024145

Konings H et al. (2020) A Literature Review of the Potential Diagnostic Biomarkers of Head and Neck Neoplasms Front Oncol 10:1020-1020 doi:10.3389/fonc.2020.01020

Konstantinidi, E. M., Lappas, A. S., Tzortzi, A. S., & Behrakis, P. K. (2015). Exhaled Breath Condensate: Technical and Diagnostic Aspects. *TheScientificWorldJournal*, *2015*, 435160. https://doi.org/10.1155/2015/435160

Kort S et al. (2020) Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters ERJ Open Res 6:00221-02019 doi:10.1183/23120541.00221-2019

Koshizuka K et al. (2017a) Dual-receptor (EGFR and c-MET) inhibition by tumor-suppressive miR-1 and miR-206 in head and neck squamous cell carcinoma J Hum Genet 62:113-121 doi:10.1038/jhg.2016.47

Koshizuka K et al. (2017b) Deep sequencing-based microRNA expression signatures in head and neck squamous cell carcinoma: dual strands of pre-miR-150 as antitumor miRNAs Oncotarget 8:30288-30304 doi:10.18632/oncotarget.16327

Kovaleva OV, Romashin D, Zborovskaya IB, Davydov MM, Shogenov MS, Gratchev A (2019) Human Lung Microbiome on the Way to Cancer Journal of Immunology Research 2019:1394191 doi:10.1155/2019/1394191

Kozaki K, Imoto I, Mogi S, Omura K, Inazawa J (2008) Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer Cancer Res 68:2094-2105 doi:10.1158/0008-5472.CAN-07-5194

Krishna S, Brown N, Faller DV, Spanjaard RA (2002) Differential effects of short-chain fatty acids on head and neck squamous carcinoma cells Laryngoscope 112:645-650 doi:10.1097/00005537-200204000-00010

Kubrak C et al. (2019) Prevalence and prognostic significance of malnutrition in patients with cancers of the head and neck Clin Nutr doi:10.1016/j.clnu.2019.03.030

Kulasinghe A, Hughes BGM, Kenny L, Punyadeera C (2019) An update: circulating tumor cells in head and neck cancer Expert Review of Molecular Diagnostics 19:1109-1115 doi:10.1080/14737159.2020.1688145

Kulasinghe A, Perry C, Jovanovic L, Nelson C, Punyadeera C (2015) Circulating tumour cells in metastatic head and neck cancers Int J Cancer 136:2515-2523 doi:10.1002/ijc.29108

Kumar S, Huang J, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB (2013) Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer Anal Chem 85:6121-6128 doi:10.1021/ac4010309

Kuper SW, Bignall JR (1966) Survival after resection of bronchial carcinomas. Significance of tumour cells in the blood Lancet 1:10-11 doi:10.1016/s0140-6736(66)90003-1

Kwak J et al. (2014) Evaluation of Bio-VOC Sampler for Analysis of Volatile Organic Compounds in Exhaled Breath Metabolites 4:879-888 doi:10.3390/metabo4040879

Lampignano R et al. (2019) Multicenter Evaluation of Circulating Cell-Free DNA Extraction and Downstream Analyses for the Development of Standardized (Pre)analytical Work Flows Clin Chem doi:10.1373/clinchem.2019.306837

Lampri ES et al. (2015) Biomarkers of head and neck cancer, tools or a gordian knot? Int J Clin Exp Med 8:10340-10357

Lang HP et al. (2016) Piezoresistive Membrane Surface Stress Sensors for Characterization of Breath Samples of Head and Neck Cancer Patients Sensors (Basel) 16 doi:10.3390/s16071149

Lange J, Leidinger P, Oehler T, Keller A, Meese E (2010) miRNA biomarkers from blood - A promising approach for minimally invasive diagnostic testing Geburtshilfe und Frauenheilkunde 70:137-141 doi:10.1055/s-0029-1240793

Langevin SM et al. (2013) Gastric reflux is an independent risk factor for laryngopharyngeal carcinoma Cancer Epidemiol Biomarkers Prev 22:1061-1068 doi:10.1158/1055-9965.EPI-13-0183

Langius JA, Bakker S, Rietveld DH, Kruizenga HM, Langendijk JA, Weijs PJ, Leemans CR (2013) Critical weight loss is a major prognostic indicator for disease-specific survival in patients with head and neck cancer receiving radiotherapy Br J Cancer 109:1093-1099 doi:10.1038/bjc.2013.458

Lassen P et al. (2014) Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer Radiother Oncol 113:310-316 doi:10.1016/j.radonc.2014.11.032

Lawal O, Ahmed WM, Nijsen TME, Goodacre R, Fowler SJ (2017) Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis Metabolomics 13:110 doi:10.1007/s11306-017-1241-8

Le Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems BMC Bioinformatics 12:253 doi:10.1186/1471-2105-12-253

Le Marchand L, Wilkens LR, Harwood P, Cooney RV (1993) Breath hydrogen and methane in populations at different risk for colon cancer Int J Cancer 55:887-890

Leary RJ et al. (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing Sci Transl Med 4:162ra154 doi:10.1126/scitranslmed.3004742

Lee JC et al. (2013) MicroRNA-222 and microRNA-146b are tissue and circulating biomarkers of recurrent papillary thyroid cancer Cancer 119:4358-4365 doi:10.1002/cncr.28254

Lee LC, Liong CY, Jemain AA (2018) Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps Analyst 143:3526-3539 doi:10.1039/c8an00599k

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14 Cell 75:843-854

Lee WH et al. (2017) Bacterial alterations in salivary microbiota and their association in oral cancer Sci Rep 7:16540 doi:10.1038/s41598-017-16418-x

Leemans CR, Braakhuis BJ, Brakenhoff RH (2011) The molecular biology of head and neck cancer Nat Rev Cancer 11:9-22 doi:10.1038/nrc2982

Lenarduzzi M et al. (2013) MicroRNA-193b enhances tumor progression via down regulation of neurofibromin 1 PLoS One 8:e53765 doi:10.1371/journal.pone.0053765

Leon SA, Shapiro B, Sklaroff DM, Yaros MJ (1977) Free DNA in the serum of cancer patients and the effect of therapy Cancer Res 37:646-650

Lerner C et al. (2016) Characterization of miR-146a and miR-155 in blood, tissue and cell lines of head and neck squamous cell carcinoma patients and their impact on cell proliferation and migration J Cancer Res Clin Oncol 142:757-766 doi:10.1007/s00432-015-2087-y

Leunis N, Boumans ML, Kremer B, Din S, Stobberingh E, Kessels AG, Kross KW (2014) Application of an electronic nose in the diagnosis of head and neck cancer Laryngoscope 124:1377-1381 doi:10.1002/lary.24463

Levitt MD, Furne JK, Kuskowski M, Ruddy J (2006) Stability of human methanogenic flora over 35 years and a review of insights obtained from breath methane measurements Clin Gastroenterol Hepatol 4:123-129 doi:10.1016/j.cgh.2005.11.006

Li G et al. (2015) Increased expression of miR-93 is associated with poor prognosis in head and neck squamous cell carcinoma Tumour Biol 36:3949-3956 doi:10.1007/s13277-015-3038-6

Li J et al. (2013) The saliva microbiome of Pan and Homo BMC Microbiol 13:204 doi:10.1186/1471-2180-13-204

Li S et al. (2017a) Complex integrated analysis of lncRNAs-miRNAs-mRNAs in oral squamous cell carcinoma Oral Oncol 73:1-9 doi:10.1016/j.oraloncology.2017.07.026

Li W, Deng Y, Chu Q, Zhang P (2019) Gut microbiome and cancer immunotherapy Cancer Lett 447:41-47 doi:10.1016/j.canlet.2019.01.015

Li W, Duan Y (2015) Human exhaled breath analysis: trends in techniques and its potential applications in non-invasive clinical diagnosis Progress in Chemistry 27:321-335 doi:10.7536/PC141021

Li Y, Min D, Wang K, Yin S, Zheng H, Liu L (2017b) MicroRNA152 inhibits cell proliferation, migration and invasion by directly targeting MAFB in nasopharyngeal carcinoma Mol Med Rep 15:948-956 doi:10.3892/mmr.2016.6059

Li W, Deng Y, Chu Q, Zhang P (2019) Gut microbiome and cancer immunotherapy Cancer Lett 447:41-47 doi:10.1016/j.canlet.2019.01.015

Li Y et al. (2020) Composition and function of oral microbiota between gingival squamous cell carcinoma and periodontitis Oral Oncol 107:104710 doi:10.1016/j.oraloncology.2020.104710

Liang L, Zheng X, Hu M, Cui Y, Zhong Q, Wang S, Huang F (2018) MiRNA-221/222 in thyroid cancer: A meta-analysis Clinica chimica acta; international journal of clinical chemistry 484:284-292 doi:10.1016/j.cca.2018.06.012

Liang S et al. (2016) Increased Serum Level of MicroRNA-663 Is Correlated with Poor Prognosis of Patients with Nasopharyngeal Carcinoma Dis Markers 2016:7648215 doi:10.1155/2016/7648215

Liborio-Kimura TN, Jung HM, Chan EK (2015) miR-494 represses HOXA10 expression and inhibits cell proliferation in oral cancer Oral Oncol 51:151-157 doi:10.1016/j.oraloncology.2014.11.019

Lim Y, Fukuma N, Totsika M, Kenny L, Morrison M, Punyadeera C (2018) The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers Frontiers in Cellular and Infection Microbiology 8 doi:ARTN 267 10.3389/fcimb.2018.00267

Lin Y-M et al. (2015) High PD-L1 Expression Correlates with Metastasis and Poor Prognosis in Oral Squamous Cell Carcinoma PLOS ONE 10:e0142656 doi:10.1371/journal.pone.0142656

Lin Z et al. (2017) Let-7e modulates the inflammatory response in vascular endothelial cells through ceRNA crosstalk Sci Rep 7:42498 doi:10.1038/srep42498

Lisowska A, Wojtowicz J, Walkowiak J (2009) Small intestine bacterial overgrowth is frequent in cystic fibrosis: combined hydrogen and methane measurements are required for its detection Acta Biochim Pol 56:631-634

Liu C et al. (2020a) Cannabinoids Promote Progression of HPV-Positive Head and Neck Squamous Cell Carcinoma via p38 MAPK Activation Clinical Cancer Research 26:2693 doi:10.1158/1078-0432.CCR-18-3301

Liu C, Xing M, Wang L, Zhang K (2017a) miR-199a-3p downregulation in thyroid tissues is associated with invasion and metastasis of papillary thyroid carcinoma Br J Biomed Sci 74:90-94 doi:10.1080/09674845.2016.1264705

Liu CH, Huang Q, Jin ZY, Xie F, Zhu CL, Liu Z, Wang C (2018) Circulating microRNA-21 as a prognostic, biological marker in cholangiocarcinoma J Cancer Res Ther 14:220-225 doi:10.4103/0973-1482.193125

Liu F et al. (2016) Human papillomavirus DNA positivity and seropositivity in rural Chinese men and women: a population-based cross-sectional study Sci Rep 6:26343 doi:10.1038/srep26343

Liu F, Zhao X, Qian Y, Zhang J, Zhang Y, Yin R (2017b) MiR-206 inhibits Head and neck squamous cell carcinoma cell progression by targeting HDAC6 via PTEN/AKT/mTOR pathway Biomed Pharmacother 96:229-237 doi:10.1016/j.biopha.2017.08.145

Liu L, Zhao Z, Zhou W, Fan X, Zhan Q, Song Y (2015) Enhanced Expression of miR-425 Promotes Esophageal Squamous Cell Carcinoma Tumorigenesis by Targeting SMAD2 J Genet Genomics 42:601-611 doi:10.1016/j.jgg.2015.09.010

Liu N et al. (2014) A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma Int J Cancer 134:1359-1368 doi:10.1002/ijc.28468

Liu X, Yu J, Jiang L, Wang A, Shi F, Ye H, Zhou X (2009) MicroRNA-222 regulates cell invasion by targeting matrix metalloproteinase 1 (MMP1) and manganese superoxide dismutase 2 (SOD2) in tongue squamous cell carcinoma cell lines Cancer Genomics Proteomics 6:131-139

Liu Y, Zhang Y, Chi Q, Wang Z, Sun B (2020b) Methyltransferase-like 1 (METTL1) served as a tumor suppressor in colon cancer by activating 7-methyguanosine (m7G) regulated let-7e miRNA/HMGA2 axis Life Sci 249:117480 doi:10.1016/j.lfs.2020.117480

Lo Nigro C, Denaro N, Merlotti A, Merlano M (2017) Head and neck cancer: improving outcomes with a multidisciplinary approach Cancer Manag Res 9:363-371 doi:10.2147/CMAR.S115761

Lopez S, Bermudez B, Montserrat-de la Paz S, Abia R, Muriana FJG (2018) A microRNA expression signature of the postprandial state in response to a high-saturated-fat challenge J Nutr Biochem 57:45-55 doi:10.1016/j.jnutbio.2018.03.010

Lourenco C, Turner C (2014) Breath analysis in disease diagnosis: methodological considerations and applications Metabolites 4:465-498 doi:10.3390/metabo4020465

Lydiatt WM et al. (2017) Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual CA Cancer J Clin 67:122-137 doi:10.3322/caac.21389

Lyu X et al. (2014) TGFbetaR2 is a major target of miR-93 in nasopharyngeal carcinoma aggressiveness Mol Cancer 13:51 doi:10.1186/1476-4598-13-51

Ma C et al. (2018a) A comprehensive meta-analysis of circulation miRNAs in glioma as potential diagnostic biomarker PLoS One 13:e0189452 doi:10.1371/journal.pone.0189452

Ma H et al. (2019) Correlation between microbes and colorectal cancer: tumor apoptosis is induced by sitosterols through promoting gut microbiota to produce short-chain fatty acids Apoptosis 24:168-183 doi:10.1007/s10495-018-1500-9

Ma J et al. (2018b) Ibrutinib targets microRNA-21 in multiple myeloma cells by inhibiting NF-κB and STAT3 Tumor Biology 40 doi:10.1177/1010428317731369

Ma J et al. (2017) ZEB1 induced miR-99b/let-7e/miR-125a cluster promotes invasion and metastasis in esophageal squamous cell carcinoma Cancer Lett 398:37-45 doi:10.1016/j.canlet.2017.04.006

Ma XL et al. (2012) Prognostic role of microRNA-21 in non-small cell lung cancer: a meta-analysis Asian Pac J Cancer Prev 13:2329-2334 doi:10.7314/apjcp.2012.13.5.2329

Machiels JP et al. (2015) Afatinib versus methotrexate as second-line treatment in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck progressing on or after platinum-based therapy (LUX-Head & Neck 1): an open-label, randomised phase 3 trial Lancet Oncol 16:583-594 doi:10.1016/S1470-2045(15)70124-5

Maclellan SA, Lawson J, Baik J, Guillaud M, Poh CF, Garnis C (2012) Differential expression of miRNAs in the serum of patients with high-risk oral lesions Cancer Med 1:268-274 doi:10.1002/cam4.17

Mager DL (2006) Bacteria and cancer: cause, coincidence or cure? A review J Transl Med 4:14 doi:10.1186/1479-5876-4-14

Mahmoud EH, Fawzy A, RA AE (2018) Serum MicroRNA-21 Negatively Relates to Expression of Programmed Cell Death-4 in Patients with Epithelial Ovarian Cancer Asian Pac J Cancer Prev 19:33-38 doi:10.22034/apjcp.2018.19.1.33

Mäkitie, A. A., Almangush, A., Youssef, O., Metsälä, M., Silén, S., Nixon, I. J., Haigentz, M., Jr, Rodrigo, J. P., Saba, N. F., Vander Poorten, V., & Ferlito, A. (2020). Exhaled breath analysis in the diagnosis of head and neck cancer. *Head & neck*, *42*(4), 787–793. https://doi.org/10.1002/hed.26043

Malicki J (2015) Medical physics in radiotherapy: The importance of preserving clinical responsibilities and expanding the profession's role in research, education, and quality control Reports of practical oncology and radiotherapy : journal of Greatpoland Cancer Center in Poznan and Polish Society of Radiation Oncology 20:161-169 doi:10.1016/j.rpor.2015.01.001

Markar SR et al. (2019) Breath Volatile Organic Compound Profiling of Colorectal Cancer Using Selected Ion Flow-tube Mass Spectrometry Ann Surg 269:903-910 doi:10.1097/SLA.0000000000002539

Markar SR et al. (2018) Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer JAMA Oncol 4:970-976 doi:10.1001/jamaoncol.2018.0991

Martin AN, Farquar GR, Jones AD, Frank M (2010) Human breath analysis: methods for sample collection and reduction of localized background effects Anal Bioanal Chem 396:739-750 doi:10.1007/s00216-009-3217-7

Martínez-Lozano P, de la Mora JF (2007) Electrospray ionization of volatiles in breath Int J Mass Spectrom 265:68-72 doi:10.1016/j.ijms.2007.05.008

Martinez-Lozano P, Fernandez de la Mora J (2008) Direct analysis of fatty acid vapors in breath by electrospray ionization and atmospheric pressure ionization-mass spectrometry Anal Chem 80:8210-8215 doi:10.1021/ac801185e

Martinez-Lozano Sinues P et al. (2015) Secondary electrospray ionization-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study J Breath Res 9:031001 doi:10.1088/1752-7155/9/3/031001

Martinez BV et al. (2015) Circulating small non coding RNA signature in head and neck squamous cell carcinoma Oncotarget 6:19246-19263

Mayne G et al. (2020) Cross validated serum small extracellular vesicle microRNAs for the detection of oropharyngeal squamous cell carcinoma Journal of translational medicine 18:1-12

Mazurek AM, Rutkowski T, Śnietura M, Pigłowski W, Suwiński R, Składowski K (2019) Detection of circulating HPV16 DNA as a biomarker in the blood of patients with human papillomavirus-positive oropharyngeal squamous cell carcinoma Head & Neck 41:632-641 doi:10.1002/hed.25368

McKelvey KJ, Powell KL, Ashton AW, Morris JM, McCracken SA (2015) Exosomes: Mechanisms of Uptake J Circ Biomark 4:7 doi:10.5772/61186

Meccariello G et al. (2019) Trans-oral robotic surgery for the management of oropharyngeal carcinomas: a 9-year institutional experience Acta Otorhinolaryngol Ital 39:75-83 doi:10.14639/0392-100X-2199

Mehanna H, Beech T, Nicholson T, El-Hariry I, McConkey C, Paleri V, Roberts S (2013) Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region Head Neck 35:747-755 doi:10.1002/hed.22015

Mehanna, H., Kong, A., & Ahmed, S. K. (2016). Recurrent head and neck cancer: United Kingdom National Multidisciplinary Guidelines. The Journal of laryngology and otology, 130(S2), S181–S190. https://doi.org/10.1017/S002221511600061X

Mehanna H et al. (2020) De-Escalation After DE-ESCALATE and RTOG 1016: A Head and Neck Cancer InterGroup Framework for Future De-Escalation Studies J Clin Oncol:JCO2000056 doi:10.1200/JCO.20.00056

Mehdipour M et al. (2018a) Diagnostic and prognostic relevance of salivary microRNA-21, -125a, -31 and -200a levels in patients with oral lichen planus - a short report Cellular Oncology:1-6 doi:10.1007/s13402-018-0372-x

Mehdipour M et al. (2018b) Diagnostic and prognostic relevance of salivary microRNA-21, -125a, -31 and -200a levels in patients with oral lichen planus - a short report Cellular oncology (Dordrecht) 41:329-334 doi:10.1007/s13402-018-0372-x

Mehrtash H et al. (2017) Defining a global research and policy agenda for betel quid and areca nut Lancet Oncol 18:e767-e775 doi:10.1016/S1470-2045(17)30460-6

Merlano MC et al. (2020) Phase III Randomized Study of Induction Chemotherapy Followed by Definitive Radiotherapy + Cetuximab Versus Chemoradiotherapy in Squamous Cell Carcinoma of Head and Neck: The INTERCEPTOR-GONO Study (NCT00999700) Oncology:1-8 doi:10.1159/000507733

Miekisch W, Kischkel S, Sawacki A, Liebau T, Mieth M, Schubert JK (2008) Impact of sampling procedures on the results of breath analysis J Breath Res 2:026007 doi:10.1088/1752-7155/2/2/026007

Miller DL et al. (2015) 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 185:679-692 doi:10.1016/j.ajpath.2014.11.018

Minna E et al. (2014) miR-199a-3p displays tumor suppressor functions in papillary thyroid carcinoma Oncotarget 5:2513-2528 doi:10.18632/oncotarget.1830

Mitchell PS et al. (2008) Circulating microRNAs as stable blood-based markers for cancer detection Proc Natl Acad Sci U S A 105:10513-10518 doi:10.1073/pnas.0804549105

Mochalski P, Wzorek B, Sliwka I, Amann A (2009) Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis J Chromatogr B Analyt Technol Biomed Life Sci 877:189-196 doi:10.1016/j.jchromb.2008.12.003

Montero PH, Patel SG (2015) Cancer of the oral cavity Surg Oncol Clin N Am 24:491-508 doi:10.1016/j.soc.2015.03.006

Moons KG, Altman DG, Reitsma JB, Collins GS, Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Development I (2015) New Guideline for the Reporting of Studies Developing, Validating, or Updating a Multivariable Clinical Prediction Model: The TRIPOD Statement Adv Anat Pathol 22:303-305 doi:10.1097/PAP.0000000000000072

Morton RP (1997) Laryngeal cancer: quality-of-life and cost-effectiveness Head Neck 19:243-250 doi:10.1002/(sici)1097-0347(199707)19:4<243::aid-hed1>3.0.co;2-0

Moser B, Bodrogi F, Eibl G, Lechner M, Rieder J, Lirk P (2005) Mass spectrometric profile of exhaled breath--field study by PTR-MS Respir Physiol Neurobiol 145:295-300 doi:10.1016/j.resp.2004.02.002

Mukherjee A, Idigo AJ, Ye Y, Wiener HW, Paluri R, Nabell LM, Shrestha S (2020) Geographical and Racial Disparities in Head and Neck Cancer Diagnosis in South-Eastern United States: Using Real-World Electronic Medical Records Data Health Equity 4:43-51 doi:10.1089/heq.2019.0092

Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach BMC Med Res Methodol 18:143 doi:10.1186/s12874-018-0611-x

Murakami Y, Aly HH, Tajima A, Inoue I, Shimotohno K (2009) Regulation of the hepatitis C virus genome replication by miR-199a J Hepatol 50:453-460 doi:10.1016/j.jhep.2008.06.010

Nahum AM, Mullally W, Marmor L (1961) A syndrome resulting from radical neck dissection Arch Otolaryngol 74:424-428 doi:10.1001/archotol.1961.00740030433011

Naidu H, Noordzij JP, Samim A, Jalisi S, Grillone GA (2012) Comparison of efficacy, safety, and cost-effectiveness of in-office cup forcep biopsies versus operating room biopsies for laryngopharyngeal tumors J Voice 26:604-606 doi:10.1016/j.jvoice.2011.10.003

Napolitano M, Schipilliti FM, Trudu L, Bertolini F (2019) Immunotherapy in head and neck cancer: The great challenge of patient selection Critical Reviews in Oncology/Hematology 144 doi:10.1016/j.critrevonc.2019.102829

National Center for Biotechnology Information (2020). PubChem Compound Summary for CID 14257, Undecane. Retrieved November 29, 2020 from https://pubchem.ncbi.nlm.nih.gov/compound/Undecane.

Nauts HC (1989) Bacteria and cancer--antagonisms and benefits Cancer Surv 8:713-723

Nees M et al. (1993) Expression of mutated p53 occurs in tumor-distant epithelia of head and neck cancer patients: a possible molecular basis for the development of multiple tumors Cancer Res 53:4189-4196

Neu U, Mainou BA (2020) Virus interactions with bacteria: Partners in the infectious dance PLoS Pathog 16:e1008234 doi:10.1371/journal.ppat.1008234

Nie CL, Ren WH, Ma Y, Xi JS, Han B (2015) Circulating miR-125b as a biomarker of Ewing's sarcoma in Chinese children Genetics and Molecular Research 14:19049-19056 doi:10.4238/2015.December.29.12

Nutting CM et al. (2011) Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial Lancet Oncol 12:127-136 doi:10.1016/S1470-2045(10)70290-4

O'Malley DP, Yang Y, Boisot S, Sudarsanam S, Wang JF, Chizhevsky V, Zhao G, Arain S, Weiss LM (2019) Immunohistochemical detection of PD-L1 among diverse human neoplasms in a reference laboratory: observations based upon 62,896 cases. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 32(7), 929–942. https://doi.org/10.1038/s41379-019-0210-3

Opitz P, Herbarth O (2018) The volatilome - investigation of volatile organic metabolites (VOM) as potential tumor markers in patients with head and neck squamous cell carcinoma (HNSCC) J Otolaryngol Head Neck Surg 47:42 doi:10.1186/s40463-018-0288-5

Park J, Look KA (2019) Health Care Expenditure Burden of Cancer Care in the United States Inquiry 56:46958019880696 doi:10.1177/0046958019880696

Park Y, Kim J (2019) Regulation of IL-6 signaling by miR-125a and let-7e in endothelial cells controls vasculogenic mimicry formation of breast cancer cells BMB Rep 52:214-219 doi:10.5483/BMBRep.2019.52.3.308

Patel KR et al. (2016) Value within otolaryngology: Assessment of the cost-utility analysis literature World J Otorhinolaryngol Head Neck Surg 2:28-37 doi:10.1016/j.wjorl.2016.01.001

Patterson JM (2019) Late Effects of Organ Preservation Treatment on Swallowing and Voice; Presentation, Assessment, and Screening Front Oncol 9:401 doi:10.3389/fonc.2019.00401

Peacock B, Rigby A, Bradford J, Pink R, Hunter K, Lambert D, Hunt S (2018) Extracellular vesicle microRNA cargo is correlated with HPV status in oropharyngeal carcinoma J Oral Pathol Med 47:954-963 doi:10.1111/jop.12781

Peled Y, Weinberg D, Hallak A, Gilat T (1987) Factors affecting methane production in humans. Gastrointestinal diseases and alterations of colonic flora Dig Dis Sci 32:267-271

Peng G et al. (2010) Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors Br J Cancer 103:542-551 doi:10.1038/sj.bjc.6605810

Pereira J et al. (2015) Breath analysis as a potential and non-invasive frontier in disease diagnosis: An overview Metabolites 5:3-55 doi:10.3390/metabo5010003

Perman JA, Modler S, Barr RG, Rosenthal P (1984) Fasting breath hydrogen concentration: normal values and clinical application Gastroenterology 87:1358-1363

Persaud K, Dodd G (1982) Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose Nature 299:352-355 doi:10.1038/299352a0

Peters BA et al. (2017) Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers Cancer Res 77:6777-6787 doi:10.1158/0008-5472.CAN-17-1296

Petersen C, Round JL (2014) Defining dysbiosis and its influence on host immunity and disease Cell Microbiol 16:1024-1033 doi:10.1111/cmi.12308

Pimentel M, Mathur R, Chang C (2013) Gas and the microbiome Curr Gastroenterol Rep 15:356 doi:10.1007/s11894-013-0356-y

Pleil JD, Hansel A, Beauchamp J (2019) Advances in proton transfer reaction mass spectrometry (PTR-MS): applications in exhaled breath analysis, food science, and atmospheric chemistry J Breath Res 13:039002 doi:10.1088/1752-7163/ab21a7

Plieskatt JL, Feng Y, Rinaldi G, Mulvenna JP, Bethony JM, Brindley PJ (2014a) Circumventing qPCR inhibition to amplify miRNAs in plasma Biomark Res 2:13 doi:10.1186/2050-7771-2-13

Plieskatt JL et al. (2014b) Methods and matrices: approaches to identifying miRNAs for nasopharyngeal carcinoma J Transl Med 12:3 doi:10.1186/1479-5876-12-3

Poel D, Buffart TE, Oosterling-Jansen J, Verheul HM, Voortman J (2018) Evaluation of several methodological challenges in circulating miRNA qPCR

studies in patients with head and neck cancer Exp Mol Med 50:e454 doi:10.1038/emm.2017.288

Poeta ML et al. (2007) TP53 mutations and survival in squamous-cell carcinoma of the head and neck N Engl J Med 357:2552-2561 doi:10.1056/NEJMoa073770

Proctor GB (2016) The physiology of salivary secretion Periodontol 2000 70:11-25 doi:10.1111/prd.12116

Pusztai L, Mazouni C, Anderson K, Wu Y, Symmans WF (2006) Molecular classification of breast cancer: limitations and potential Oncologist 11:868-877 doi:10.1634/theoncologist.11-8-868

Pyeon D et al. (2007) Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers Cancer Res 67:4605-4619 doi:10.1158/0008-5472.CAN-06-3619

Qiao DD et al. (2017) Expression of microRNA-122 and microRNA-22 in HBV-related liver cancer and the correlation with clinical features Eur Rev Med Pharmacol Sci 21:742-747

Qiu X et al. (2016) Circulating MicroRNA-26a in Plasma and Its Potential Diagnostic Value in Gastric Cancer PLoS One 11:e0151345 doi:10.1371/journal.pone.0151345

Rabinowits G et al. (2017) Comparative Analysis of MicroRNA Expression among Benign and Malignant Tongue Tissue and Plasma of Patients with Tongue Cancer Front Oncol 7:191 doi:10.3389/fonc.2017.00191

Rapado-González Ó et al. (2018) Human salivary microRNAs in Cancer Journal of Cancer 9:638-649 doi:10.7150/jca.21180

Rapisuwon S, Vietsch EE, Wellstein A (2016) Circulating biomarkers to monitor cancer progression and treatment Comput Struct Biotechnol J 14:211-222 doi:10.1016/j.csbj.2016.05.004

Reichart PA, Phillipsen HP (1998) Betel chewer's mucosa--a review J Oral Pathol Med 27:239-242 doi:10.1111/j.1600-0714.1998.tb01949.x

Reinhart BJ et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans Nature 403:901-906 doi:10.1038/35002607

Ren W et al. (2014) Circulating microRNA-21 (MIR-21) and phosphatase and tensin homolog (PTEN) are promising novel biomarkers for detection of oral squamous cell carcinoma Biomarkers 19:590-596 doi:10.3109/1354750x.2014.955059

Rezaie A et al. (2017) Hydrogen and Methane-Based Breath Testing in Gastrointestinal Disorders: The North American Consensus Am J Gastroenterol 112:775-784 doi:10.1038/ajg.2017.46

Ricieri Brito JA et al. (2010) Reduced expression of mir15a in the blood of patients with oral squamous cell carcinoma is associated with tumor staging Exp Ther Med 1:217-221 doi:10.3892/etm_00000035

Ries J, Baran C, Wehrhan F, Weber M, Neukam FW, Krautheim-Zenk A, Nkenke E (2017) Prognostic significance of altered miRNA expression in whole blood of OSCC patients Oncol Rep 37:3467-3474 doi:10.3892/or.2017.5639

Ries J, Vairaktaris E, Agaimy A, Kintopp R, Baran C, Neukam FW, Nkenke E (2014a) miR-186, miR-3651 and miR-494: potential biomarkers for oral squamous cell carcinoma extracted from whole blood Oncol Rep 31:1429-1436 doi:10.3892/or.2014.2983

Ries J, Vairaktaris E, Kintopp R, Baran C, Neukam FW, Nkenke E (2014b) Alterations in miRNA expression patterns in whole blood of OSCC patients In Vivo 28:851-861

Rischin D et al. (2019) Protocol-specified final analysis of the phase 3 KEYNOTE-048 trial of pembrolizumab (pembro) as first-line therapy for recurrent/metastatic head and neck squamous cell carcinoma (R/M HNSCC) Journal of Clinical Oncology 37:6000-6000 doi:10.1200/JCO.2019.37.15_suppl.6000

Robbins KT et al. (2002) Neck dissection classification update: revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology-Head and Neck Surgery Arch Otolaryngol Head Neck Surg 128:751-758 doi:10.1001/archotol.128.7.751

Rocco G (2018) Every breath you take: The value of the electronic nose (e-nose) technology in the early detection of lung cancer J Thorac Cardiovasc Surg 155:2622-2625 doi:10.1016/j.jtcvs.2017.12.155

Rodriguez-Perez R, Fernandez L, Marco S (2018) Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study Anal Bioanal Chem 410:5981-5992 doi:10.1007/s00216-018-1217-1

Roesch R et al. (2020) Cell-Free DNA in Plasma or Saliva for Minimally Invasive Monitoring of Head and Neck Cancer Patients Oncology Research and Treatment 43:99-99

Rombeau JL, Kripke SA (1990) Metabolic and intestinal effects of short-chain fatty acids JPEN J Parenter Enteral Nutr 14:181S-185S doi:10.1177/014860719001400507

Rosenberg AJ, Vokes EE (2020) Optimizing Treatment De-Escalation in Head and Neck Cancer: Current and Future Perspectives Oncologist doi:10.1634/theoncologist.2020-0303

Rosenthal A, Solomons NW (1983) Time-course of cigarette smoke contamination of clinical hydrogen breath-analysis tests Clin Chem 29:1980-1981

Rothenberg SM, Ellisen LW (2012) The molecular pathogenesis of head and neck squamous cell carcinoma J Clin Invest 122:1951-1957 doi:10.1172/jci59889

Rudmik L, Lau HY, Matthews TW, Bosch JD, Kloiber R, Molnar CP, Dort JC (2011) Clinical utility of PET/CT in the evaluation of head and neck squamous cell carcinoma with an unknown primary: a prospective clinical trial Head Neck 33:935-940 doi:10.1002/hed.21566

Rumessen JJ, Nordgaard-Andersen I, Gudmand-Hoyer E (1994) Carbohydrate malabsorption: quantification by methane and hydrogen breath tests Scand J Gastroenterol 29:826-832

Russo E, Taddei A, Ringressi MN, Ricci F, Amedei A (2016) The interplay between the microbiome and the adaptive immune response in cancer development Therap Adv Gastroenterol 9:594-605 doi:10.1177/1756283X16635082

Santos JMO et al. (2017) Dysregulated expression of microRNA-150 in human papillomavirus-induced lesions of K14-HPV16 transgenic mice Life Sci 175:31-36 doi:10.1016/j.lfs.2017.03.008

Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG (2018) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration J Natl Cancer Inst 110:803-811 doi:10.1093/jnci/djy088

Savin Z, Kivity S, Yonath H, Yehuda S (2018) Smoking and the intestinal microbiome Arch Microbiol 200:677-684 doi:10.1007/s00203-018-1506-2

Schache AG et al. (2013) Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma Br J Cancer 108:1332-1339 doi:10.1038/bjc.2013.63

Schmutzhard J et al. (2008) Pilot study: volatile organic compounds as a diagnostic marker for head and neck tumors Head Neck 30:743-749 doi:10.1002/hed.20779

Schneider A et al. (2018) Tissue and serum microRNA profile of oral squamous cell carcinoma patients Sci Rep 8:675 doi:10.1038/s41598-017-18945-z

Schutte HW et al. (2020) Impact of Time to Diagnosis and Treatment in Head and Neck Cancer: A Systematic Review Otolaryngol Head Neck Surg 162:446-457 doi:10.1177/0194599820906387

Seitz HK, Stickel F (2007) Molecular mechanisms of alcohol-mediated carcinogenesis Nat Rev Cancer 7:599-612 doi:10.1038/nrc2191

Sell SL, Widen SG, Prough DS, Hellmich HL (2020) Principal component analysis of blood microRNA datasets facilitates diagnosis of diverse diseases PLoS One 15:e0234185 doi:10.1371/journal.pone.0234185

Shariat SF, Lotan Y, Vickers A, Karakiewicz PI, Schmitz-Dräger BJ, Goebell PJ, Malats N (2010) Statistical consideration for clinical biomarker research in bladder cancer Urologic oncology 28:389-400 doi:10.1016/j.urolonc.2010.02.011

Sharma DC (2003) Betel quid and areca nut are carcinogenic without tobacco Lancet Oncol 4:587 doi:10.1016/s1470-2045(03)01229-4

Shaw JA et al. (2000) Microsatellite alterations plasma DNA of primary breast cancer patients Clin Cancer Res 6:1119-1124

Shaw R, Beasley N (2016) Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines J Laryngol Otol 130:S9-S12 doi:10.1017/S0022215116000360

Shemilt I, Thomas J, Morciano M (2010) A web-based tool for adjusting costs to a specific target currency and price year Evid Policy 6:51-59 doi:Doi 10.1332/174426410x482999

Shen W, Song M, Liu J, Qiu G, Li T, Hu Y, Liu H (2014) MiR-26a promotes ovarian cancer proliferation and tumorigenesis PLoS One 9:e86871 doi:10.1371/journal.pone.0086871

Shestivska V, Olsinova M, Sovova K, Kubista J, Smith D, Cebecauer M, Spanel P (2018) Evaluation of lipid peroxidation by the analysis of volatile aldehydes in the headspace of synthetic membranes using selected ion flow tube mass spectrometry Rapid Commun Mass Spectrom 32:1617-1628 doi:10.1002/rcm.8212

Shi B, Ma C, Liu G, Guo Y (2019) MiR-106a directly targets LIMK1 to inhibit proliferation and EMT of oral carcinoma cells Cell Mol Biol Lett 24:1 doi:10.1186/s11658-018-0127-8

Shigematsu Y, Inamura K (2018) Gut microbiome: a key player in cancer immunotherapy Hepatobiliary Surg Nutr 7:479-480 doi:10.21037/hbsn.2018.10.02

Shiota S, Yamaoka Y (2014) Biomarkers for Helicobacter pylori infection and gastroduodenal diseases Biomark Med 8:1127-1137 doi:10.2217/bmm.14.72

Shoffel-Havakuk H, Frumin I, Lahav Y, Haviv L, Sobel N, Halperin D (2016) Increased number of volatile organic compounds over malignant glottic lesions Laryngoscope 126:1606-1611 doi:10.1002/lary.25733

Shukla M, Forghani R, Agarwal M (2020) Patient-Centric Head and Neck Cancer Radiation Therapy: Role of Advanced Imaging Neuroimaging Clin N Am 30:341-357 doi:10.1016/j.nic.2020.04.005

Singh A, Willems E, Singh A, Hafeez BB, Ong IM, Mehta SL, Verma AK (2016) Ultraviolet radiation-induced tumor necrosis factor alpha, which is linked to the development of cutaneous SCC, modulates differential epidermal microRNAs expression Oncotarget 7:17945-17956 doi:10.18632/oncotarget.7595

Singh P, Srivastava AN, Sharma R, Mateen S, Shukla B, Singh A, Chandel S (2018) Circulating MicroRNA-21 Expression as a Novel Serum Biomarker for Oral Sub-Mucous Fibrosis and Oral Squamous Cell Carcinoma Asian Pac J Cancer Prev 19:1053-1057 doi:10.22034/APJCP.2018.19.4.1053

Sivertsen SM, Bjorneklett A, Gullestad HP, Nygaard K (1992) Breath methane and colorectal cancer Scand J Gastroenterol 27:25-28

Sloan TJ et al. (2018) A low FODMAP diet is associated with changes in the microbiota and reduction in breath hydrogen but not colonic volume in healthy subjects PLoS One 13:e0201410 doi:10.1371/journal.pone.0201410

Smith D, Mcintosh BJ, Adams NG (1989) A Selected Ion Flow Tube Study of the Reactions of the Ph0-4n+ Ions with Several Molecular Gases at 300-K J Chem Phys 90:6213-6219 doi:Doi 10.1063/1.456337

Smith JE (2000) The obstructed airway in head and neck surgery Anaesthesia 55:291 doi:10.1046/j.1365-2044.2000.01333.x

Song Y, Tian Y, Bai WL, Ma XL (2014) Expression and clinical significance of microRNA-152 in supragalottic laryngeal carcinoma Tumour Biol 35:11075-11079 doi:10.1007/s13277-014-2406-y

Spanel P, Rolfe P, Rajan B, Smith D (1996) The selected ion flow tube (SIFT) - A novel technique for biological monitoring Annals of Occupational Hygiene 40:615-626 doi:DOI 10.1093/annhyg/40.6.615

Spanel P, Spesyvyi A, Smith D (2019) Electrostatic Switching and Selection of H3O(+), NO(+), and O2(+*) Reagent Ions for Selected Ion Flow-Drift Tube Mass Spectrometric Analyses of Air and Breath Anal Chem 91:5380-5388 doi:10.1021/acs.analchem.9b00530

Spietelun A, Pilarczyk M, Kloskowski A, Namiesnik J (2010) Current trends in solid-phase microextraction (SPME) fibre coatings Chem Soc Rev 39:4524-4537 doi:10.1039/c003335a

Statnikov A, Wang L, Aliferis CF (2008) A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification BMC Bioinformatics 9:319 doi:10.1186/1471-2105-9-319

Steeb WH (2008) The nonlinear workbook : chaos, fractals, cellular automata, neural networks, genetic algorithms, gene expression programming, support vector machine, wavelets, hidden Markov models, Fuzzy logic with C++, Java and SymbolicC++ programs. 4th edn. World Scientific, New Jersey

Stransky N et al. (2011) The mutational landscape of head and neck squamous cell carcinoma Science 333:1157-1160 doi:10.1126/science.1208130

Strimbu K, Tavel JA (2010) What are biomarkers? Current Opinion in Hiv and Aids 5:463-466 doi:10.1097/COH.0b013e32833ed177

Stroun M, Anker P (2005) Circulating DNA in higher organisms cancer detection brings back to life an ignored phenomenon Cell Mol Biol (Noisy-le-grand) 51:767-774

Su Z, Hou XK, Wen QP (2014) Propofol induces apoptosis of epithelial ovarian cancer cells by upregulation of microRNA let-7i expression Eur J Gynaecol Oncol 35:688-691

Subramaniam SS, Paterson C, McCaul JA (2019) Immunotherapy in the management of squamous cell carcinoma of the head and neck Br J Oral Maxillofac Surg 57:957-966 doi:10.1016/j.bjoms.2019.08.002

Summerer I et al. (2013) Changes in circulating microRNAs after radiochemotherapy in head and neck cancer patients Radiat Oncol 8:296 doi:10.1186/1748-717x-8-296

Summerer I et al. (2015) Circulating microRNAs as prognostic therapy biomarkers in head and neck cancer patients Br J Cancer 113:76-82 doi:10.1038/bjc.2015.111

Sun G, Cao Y, Wang P, Song H, Bie T, Li M, Huai (2018a) miR-200b-3p in plasma is a potential diagnostic biomarker in oral squamous cell carcinoma Biomarkers 23:137-141 doi:10.1080/1354750x.2017.1289241

Sun G, Cao Y, Wang P, Song H, Bie T, Li M, Huai D (2018b) miR-200b-3p in plasma is a potential diagnostic biomarker in oral squamous cell carcinoma Biomarkers 23:137-141 doi:10.1080/1354750X.2017.1289241

Sun L, Liang J, Wang Q, Li Z, Du Y, Xu X (2016a) MicroRNA-137 suppresses tongue squamous carcinoma cell proliferation, migration and invasion Cell Proliferation doi:10.1111/cpr.12287

Sun L, Liu L, Fu H, Wang Q, Shi Y (2016b) Association of Decreased Expression of Serum miR-9 with Poor Prognosis of Oral Squamous Cell Carcinoma Patients Med Sci Monit 22:289-294

Sun W et al. (2014) Activation of the NOTCH pathway in head and neck cancer Cancer Res 74:1091-1104 doi:10.1158/0008-5472.CAN-13-1259

Sun Z, Zhang W, Li Q (2017) miR-125a suppresses viability and glycolysis and induces apoptosis by targeting Hexokinase 2 in laryngeal squamous cell carcinoma Cell Biosci 7:51 doi:10.1186/s13578-017-0178-y

Suri J, Kataria R, Malik Z, Parkman HP, Schey R (2018) Elevated methane levels in small intestinal bacterial overgrowth suggests delayed small bowel and colonic transit Medicine (Baltimore) 97:e10554 doi:10.1097/MD.0000000000010554

Sutton M, Thiebaut R, Liquet B (2018) Sparse partial least squares with group and subgroup structure Stat Med 37:3338-3356 doi:10.1002/sim.7821

Swiecicki PL, Brennan JR, Mierzwa M, Spector ME, Brenner JC (2019) Head and Neck Squamous Cell Carcinoma Detection and Surveillance: Advances of Liquid Biomarkers Laryngoscope 129:1836-1843 doi:10.1002/lary.27725

Swift R (2003) Direct measurement of alcohol and its metabolites Addiction 98 Suppl 2:73-80 doi:10.1046/j.1359-6357.2003.00605.x

Szabó A et al. (2015) Volatile sulphur compound measurement with OralChroma(TM): a methodological improvement J Breath Res 9:016001 doi:10.1088/1752-7155/9/1/016001

Tachibana H et al. (2016) Circulating miR-223 in Oral Cancer: Its Potential as a Novel Diagnostic Biomarker and Therapeutic Target PLoS One 11:e0159693 doi:10.1371/journal.pone.0159693

Tadesse K, Eastwood M (1977) Breath-hydrogen test and smoking Lancet 2:91

Tadesse K, Eastwood MA (1978) Metabolism of dietary fibre components in man assessed by breath hydrogen and methane Br J Nutr 40:393-396

Tadesse K, Smith D, Eastwod MA (1980) Breath hydrogen (H2) and methane (CH4) excretion patterns in normal man and in clinical practice Q J Exp Physiol Cogn Med Sci 65:85-97

Taguchi YH (2016) Identification of More Feasible MicroRNA-mRNA Interactions within Multiple Cancers Using Principal Component Analysis Based Unsupervised Feature Extraction Int J Mol Sci 17 doi:10.3390/ijms17050696

Talamini R et al. (2002) Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study Cancer Causes Control 13:957-964 doi:10.1023/a:1021944123914

Tanaka S, Ishikawa M, Arai M, Genda Y, Sakamoto A (2012) Changes in microRNA expression in rat lungs caused by sevoflurane anesthesia: a TaqMan(R) low-density array study Biomed Res 33:255-263

Tang F, Ishwaran H (2017) Random Forest Missing Data Algorithms Stat Anal Data Min 10:363-377 doi:10.1002/sam.11348

Tay JK, Siow CH, Goh HL, Lim CM, Hsu PP, Chan SH, Loh KS (2020) A comparison of EBV serology and serum cell-free DNA as screening tools for nasopharyngeal cancer: Results of the Singapore NPC screening cohort Int J Cancer 146:2923-2931 doi:10.1002/ijc.32774

Thery C et al. (2018) Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines J Extracell Vesicles 7:1535750 doi:10.1080/20013078.2018.1535750

Thirunavukkarasan M et al. (2017) Short-chain fatty acid receptors inhibit invasive phenotypes in breast cancer cells PLoS One 12:e0186334 doi:10.1371/journal.pone.0186334

Thorsson V et al. (2018) The Immune Landscape of Cancer Immunity 48:812-830.e814 doi:10.1016/j.immuni.2018.03.023

Tian Y et al. (2014) MicroRNA-27a promotes proliferation and suppresses apoptosis by targeting PLK2 in laryngeal carcinoma BMC Cancer 14 doi:10.1186/1471-2407-14-678

Tiberio P, Callari M, Angeloni V, Daidone MG, Appierto V (2015) Challenges in using circulating miRNAs as cancer biomarkers Biomed Res Int 2015:731479 doi:10.1155/2015/731479

Tikka T, Pracy P, Paleri V (2016) Refining the head and neck cancer referral guidelines: a two-centre analysis of 4715 referrals Clinical Otolaryngology 41:66-75 doi:10.1111/coa.12597

Tiwari A, Shivananda S, Gopinath KS, Kumar A (2014) MicroRNA-125a reduces proliferation and invasion of oral squamous cell carcinoma cells by targeting estrogen-related receptor alpha: implications for cancer therapeutics J Biol Chem 289:32276-32290 doi:10.1074/jbc.M114.584136

Trang P, Weidhaas JB, Slack FJ (2009) MicroRNAs as potential cancer therapeutics Oncogene 27:S52 doi:10.1038/onc.2009.353

Triantafyllou K, Chang C, Pimentel M (2014) Methanogens, methane and gastrointestinal motility J Neurogastroenterol Motil 20:31-40 doi:10.5056/jnm.2014.20.1.31

Trotti A et al. (2007) TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group Lancet Oncol 8:613-624 doi:10.1016/S1470-2045(07)70144-4

Tsai CH et al. (2015) Over-expression of cofilin-1 suppressed growth and invasion of cancer cells is associated with up-regulation of let-7 microRNA Biochim Biophys Acta 1852:851-861 doi:10.1016/j.bbadis.2015.01.007

Tseng HH et al. (2017) Next-generation Sequencing for microRNA Profiling: MicroRNA-21-3p Promotes Oral Cancer Metastasis Anticancer Res 37:1059-1066 doi:10.21873/anticanres.11417

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project Nature 449:804-810 doi:10.1038/nature06244

Turner C, Walton C, Hoashi S, Evans M (2009) Breath acetone concentration decreases with blood glucose concentration in type I diabetes mellitus patients during hypoglycaemic clamps J Breath Res 3:046004 doi:10.1088/1752-7155/3/4/046004

UniProt C (2019) UniProt: a worldwide hub of protein knowledge Nucleic Acids Res 47:D506-D515 doi:10.1093/nar/gky1049

Uppal N, Singh P (2016) Oral cancer: Breath of death Br Dent J 221:212 doi:10.1038/sj.bdj.2016.619

Urabe F, Kosaka N, Yoshioka Y, Egawa S, Ochiya T (2017) The small vesicular culprits: the investigation of extracellular vesicles as new targets for cancer treatment Clin Transl Med 6:45 doi:10.1186/s40169-017-0176-z

Urita Y, Hike K, Torii N, Kikuchi Y, Sasajima M, Miki K (2002) Efficacy of lactulose plus 13C-acetate breath test in the diagnosis of gastrointestinal motility disorders J Gastroenterol 37:442-448 doi:10.1007/s005350200064

Utnes P, Løkke C, Flægstad T, Einvik C (2019) Clinically Relevant Biomarker Discovery in High-Risk Recurrent Neuroblastoma Cancer Inform 18:1176935119832910 doi:10.1177/1176935119832910

Vadhwana B, Belluomo I, Boshier PR, Pavlou C, Španěl P, Hanna GB (2020) Impact of oral cleansing strategies on exhaled volatile organic compound levels Rapid Commun Mass Spectrom 34:e8706 doi:10.1002/rcm.8706

Van Calster B, Verbakel JY, Christodoulou E, Steyerberg EW, Collins GS (2019) Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important) J Clin Epidemiol 116:137-138 doi:10.1016/j.jclinepi.2019.08.002

van de Goor RM, Leunis N, van Hooren MR, Francisca E, Masclee A, Kremer B, Kross KW (2017) Feasibility of electronic nose technology for discriminating between head and neck, bladder, and colon carcinomas Eur Arch Otorhinolaryngol 274:1053-1060 doi:10.1007/s00405-016-4320-y

van de Kant KD, van der Sande LJ, Jobsis Q, van Schayck OC, Dompeling E (2012) Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review Respir Res 13:117 doi:10.1186/1465-9921-13-117

van den Velde S, Quirynen M, van Hee P, van Steenberghe D (2007) Halitosis associated volatiles in breath of healthy subjects J Chromatogr B Analyt Technol Biomed Life Sci 853:54-61 doi:10.1016/j.jchromb.2007.02.048

van Hooren MR, Leunis N, Brandsma DS, Dingemans AC, Kremer B, Kross KW (2016) Differentiating head and neck carcinoma from lung carcinoma with an electronic nose: a proof of concept study Eur Arch Otorhinolaryngol 273:3897-3903 doi:10.1007/s00405-016-4038-x

van Weert S, Leemans CR (2020) Salvage Surgery in Head and Neck Cancer Oral Dis doi:10.1111/odi.13582

Varga G, Erces D, Tuboly E, Kaszaki J, Ghyczy M, Boros M (2012) [Characterization of the antiinflammatory properties of methane inhalation during ischaemia-reperfusion] Magy Seb 65:205-211 doi:10.1556/MaSeb.65.2012.4.6

Vassilakopoulou M et al. (2016) Evaluation of PD-L1 Expression and Associated Tumor-Infiltrating Lymphocytes in Laryngeal Squamous Cell Carcinoma Clinical Cancer Research 22:704 doi:10.1158/1078-0432.CCR-15-1543

Veit JA et al. (2015) MicroRNA expression in differentially metastasizing tumors of the head and neck: adenoid cystic versus squamous cell carcinoma Anticancer Res 35:1271-1277

Vesty A, Gear K, Biswas K, Radcliff FJ, Taylor MW, Douglas RG (2018) Microbial and inflammatory-based salivary biomarkers of head and neck squamous cell carcinoma Clin Exp Dent Res 4:255-262 doi:10.1002/cre2.139

Võ U-UT, Morris MP (2014) Nonvolatile, semivolatile, or volatile: Redefining volatile for volatile organic compounds Journal of the Air & Waste Management Association 64:661-669 doi:10.1080/10962247.2013.873746

Wang S et al. (2019) Hsa-let-7e-5p Inhibits the Proliferation and Metastasis of Head and Neck Squamous Cell Carcinoma Cells by Targeting Chemokine Receptor 7 Journal of Cancer 10:1941-1948 doi:10.7150/jca.29536

Wang SH, Zhou JD, He QY, Yin ZQ, Cao K, Luo CQ (2014a) MiR-199a inhibits the ability of proliferation and migration by regulating CD44-Ezrin signaling in cutaneous squamous cell carcinoma cells Int J Clin Exp Pathol 7:7131-7141

Wang X, Lu X, Geng Z, Yang G, Shi Y (2017) LncRNA PTCSC3/miR-574-5p Governs Cell Proliferation and Migration of Papillary Thyroid Carcinoma via Wnt/beta-Catenin Signaling J Cell Biochem 118:4745-4752 doi:10.1002/jcb.26142

Wang Y et al. (2020) Nanodetection of Head and Neck Cancer on Titanium Oxide Sensing Surface Nanoscale research letters 15:33 doi:https://dx.doi.org/10.1186/s11671-020-3262-x

Wang Y, Zhang ZX, Chen S, Qiu GB, Xu ZM, Fu WN (2016) Methylation Status of SP1 Sites within miR-23a-27a-24-2 Promoter Region Influences Laryngeal Cancer Cell Proliferation and Apoptosis Biomed Res Int 2016:2061248 doi:10.1155/2016/2061248

Wang Z, Cai Q, Jiang Z, Liu B, Zhu Z, Li C (2014b) Prognostic role of microRNA-21 in gastric cancer: a meta-analysis Med Sci Monit 20:1668-1674 doi:10.12659/MSM.892096

Watanabe S (1954) The Metastasizability of Tumor Cells Cancer 7:215-223 doi:Doi 10.1002/1097-0142(195403)7:2<215::Aid-Cncr2820070203>3.0.Co;2-6

Weber A, Schmoz S, Bootz F (2001) CUP (carcinoma of unknown primary) syndrome in head and neck: clinic, diagnostic, and therapy Onkologie 24:38-43

Wehinger A, Schmid A, Mechtcheriakov S, Ledochowski M, Grabmer C, Gastl GA, Amann A (2007) Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas International Journal of Mass Spectrometry 265:49-59 doi:https://doi.org/10.1016/j.ijms.2007.05.012

Wei D et al. (2019a) MicroRNA-199a-5p functions as a tumor suppressor in oral squamous cell carcinoma via targeting the IKKβ/NF-κB signaling pathway Int J Mol Med 43:1585-1596 doi:10.3892/ijmm.2019.4083

Wei D et al. (2019b) MicroRNA-199a-5p suppresses migration and invasion in oral squamous cell carcinoma through inhibiting the EMT-related transcription factor SOX4 Int J Mol Med 44:185-195 doi:10.3892/ijmm.2019.4174

Wei L, Mao M, Liu H (2016a) Droplet digital PCR and qRT-PCR to detect circulating miR-21 in laryngeal squamous cell carcinoma and pre-malignant laryngeal lesions Acta Otolaryngol 136:923-932 doi:10.3109/00016489.2016.1165862

Wei T, Ye P, Peng X, Wu LL, Yu GY (2016b) Prognostic Value of miR-222 in Various Cancers: a Systematic Review and Meta-Analysis Clin Lab 62:1387-1395 doi:10.7754/Clin.Lab.2016.160102

Wei Z, Chang K, Fan C (2019c) Hsa_circ_0042666 inhibits proliferation and invasion via regulating miR-223/TGFBR3 axis in laryngeal squamous cell carcinoma Biomed Pharmacother 119:109365 doi:10.1016/j.biopha.2019.109365

Wells LA, Junor EJ, Conn B, Pattle S, Cuschieri K (2015) Population-based p16 and HPV positivity rates in oropharyngeal cancer in Southeast Scotland J Clin Pathol 68:849-852 doi:10.1136/jclinpath-2015-202947

Weng JH, Yu CC, Lee YC, Lin CW, Chang WW, Kuo YL (2016) miR-494-3p Induces Cellular Senescence and Enhances Radiosensitivity in Human Oral Squamous Carcinoma Cells Int J Mol Sci 17 doi:10.3390/ijms17071092

Whiteside TL (2018) Exosome and mesenchymal stem cell cross-talk in the tumor microenvironment Semin Immunol 35:69-79 doi:10.1016/j.smim.2017.12.003

Wilder-Smith CH, Olesen SS, Materna A, Drewes AM (2018) Fermentable Sugar Ingestion, Gas Production, and Gastrointestinal and Central Nervous System Symptoms in Patients With Functional Disorders Gastroenterology 155:1034-1044 e1036 doi:10.1053/j.gastro.2018.07.013

Wilkens LR, Le Marchand L, Harwood P, Cooney RV (1994) Use of breath hydrogen and methane as markers of colonic fermentation in epidemiological studies: variability in excretion Cancer Epidemiol Biomarkers Prev 3:149-153

Willers H, Dahm-Daphi J, Powell SN (2004) Repair of radiation damage to DNA Br J Cancer 90:1297-1301 doi:10.1038/sj.bjc.6601729

Willers H, Gheorghiu L, Liu Q, Efstathiou JA, Wirth LJ, Krause M, von Neubeck C (2015) DNA Damage Response Assessments in Human Tumor Samples Provide Functional Biomarkers of Radiosensitivity Seminars in Radiation Oncology 25:237-250 doi:10.1016/j.semradonc.2015.05.007

Wilson PF, Freeman CG, McEwan MJ, Milligan DB, Allardyce RA, Shaw GM (2001) Alcohol in breath and blood: a selected ion flow tube mass spectrometric study Rapid communications in mass spectrometry : RCM 15:413-417 doi:10.1002/rcm.244

Wilson PF, Freeman CG, McEwan MJ, Milligan DB, Allardyce RA, Shaw GM (2002) In situ analysis of solvents on breath and blood: a selected ion flow tube mass spectrometric study Rapid communications in mass spectrometry : RCM 16:427-432 doi:10.1002/rcm.594

Wold H (1966) Estimation of principal components and related models by iterative least squares. In: Krishnaiah PR (ed) Multivariate Analysis. . Academic Press, N.Y., pp 391-420

Wolf A, Moissl-Eichinger C, Perras A, Koskinen K, Tomazic PV, Thurnher D (2017) The salivary microbiome as an indicator of carcinogenesis in patients with oropharyngeal squamous cell carcinoma: A pilot study Sci Rep 7:5867 doi:10.1038/s41598-017-06361-2

Wong N et al. (2016) Prognostic microRNA signatures derived from The Cancer Genome Atlas for head and neck squamous cell carcinomas Cancer Med 5:1619-1628 doi:10.1002/cam4.718

Wu Q, Zhao Y, Sun Y, Yan X, Wang P (2018a) miR-375 inhibits IFN-gamma-induced programmed death 1 ligand 1 surface expression in head and neck squamous cell carcinoma cells by blocking JAK2/STAT1 signaling Oncol Rep 39:1461-1468 doi:10.3892/or.2018.6177

Wu Y, Sun X, Song B, Qiu X, Zhao J (2017) MiR-375/SLC7A11 axis regulates oral squamous cell carcinoma proliferation and invasion Cancer Med 6:1686-1697 doi:10.1002/cam4.1110

Wu Z, Lu B, Li X, Miao W, Li J, Shi Y, Yu W (2018b) MicroRNA-26a inhibits proliferation and tumorigenesis via targeting CKS2 in laryngeal squamous cell carcinoma Clin Exp Pharmacol Physiol 45:444-451 doi:10.1111/1440-1681.12890

Wynder EL, Bross IJ (1957) Aetiological factors in mouth cancer; an approach to its prevention Br Med J 1:1137-1143 doi:10.1136/bmj.1.5028.1137

Xanthopoulos P, Pardalos PM, Trafalis TB (2013) Robust data mining. SpringerBriefs in optimization,. Springer, New York ; London

Xia X, Yang B, Zhai X, Liu X, Shen K, Wu Z, Cai J (2013) Prognostic role of microRNA-21 in colorectal cancer: a meta-analysis PLoS One 8:e80426 doi:10.1371/journal.pone.0080426

Xiang D, Tian B, Yang T, Li Z (2019) miR-222 expression is correlated with the ATA risk stratifications in papillary thyroid carcinomas Medicine (Baltimore) 98:e16050 doi:10.1097/md.0000000000016050

Xiao M et al. (2017) Let-7e sensitizes epithelial ovarian cancer to cisplatin through repressing DNA double strand break repair J Ovarian Res 10:24 doi:10.1186/s13048-017-0321-8

Xiao X, Tang C, Xiao S, Fu C, Yu P (2013) Enhancement of proliferation and invasion by MicroRNA-590-5p via targeting PBRM1 in clear cell renal carcinoma cells Oncol Res 20:537-544 doi:10.3727/096504013x13775486749335

Xiao X, Zhou L, Cao P, Gong H, Zhang Y (2015) MicroRNA-93 regulates cyclin G2 expression and plays an oncogenic role in laryngeal squamous cell carcinoma Int J Oncol 46:161-174 doi:10.3892/ijo.2014.2704

Xu BB, Gu ZF, Ma M, Wang JY, Wang HN (2018a) MicroRNA-590-5p suppresses the proliferation and invasion of non-small cell lung cancer by regulating GAB1 Eur Rev Med Pharmacol Sci 22:5954-5963 doi:10.26355/eurrev_201809_15926

Xu C, Jackson SA (2019) Machine learning and complex biological data Genome Biol 20:76 doi:10.1186/s13059-019-1689-0

Xu H et al. (2020) Plasma exosomal miR-106a-5p expression in myasthenia gravis Muscle Nerve 61:401-407 doi:10.1002/mus.26785

Xu H et al. (2016a) Serum miR-483-5p: a novel diagnostic and prognostic biomarker for patients with oral squamous cell carcinoma Tumour Biol 37:447-453 doi:10.1007/s13277-015-3514-z

Xu X et al. (2018b) Dynamic Changes in Plasma MicroRNAs Have Potential Predictive Values in Monitoring Recurrence and Metastasis of Nasopharyngeal Carcinoma Biomed Res Int 2018:7329195 doi:10.1155/2018/7329195

Xu Y, Lin YP, Yang D, Zhang G, Zhou HF (2016b) Expression of serum microRNA-378 and its clinical significance in laryngeal squamous cell carcinoma Eur Rev Med Pharmacol Sci 20:5137-5142

Yan Y et al. (2017) Circulating miRNAs as biomarkers for oral squamous cell carcinoma recurrence in operated patients Oncotarget 8:8206-8214 doi:10.18632/oncotarget.14143

Yang F, Wang H-z, Mi H, Lin C-d, Cai W-w (2009) Using random forest for reliable classification and cost-sensitive learning for medical diagnosis BMC Bioinformatics 10 Suppl 1:S22-S22 doi:10.1186/1471-2105-10-S1-S22

Yang S-H et al. (2018) Exposure to nicotine-derived nitrosamine ketone and arecoline synergistically facilitates tumor aggressiveness via overexpression of epidermal growth factor receptor and its downstream signaling in head and neck squamous cell carcinoma PloS one 13:e0201267-e0201267 doi:10.1371/journal.pone.0201267

Yang Y, Qian J, Chen Y, Pan Y (2014) Prognostic role of circulating microRNA-21 in cancers: evidence from a meta-analysis Tumour Biol 35:6365-6371 doi:10.1007/s13277-014-1846-8

Yang ZR (2004) Biological applications of support vector machines Brief Bioinform 5:328-338 doi:10.1093/bib/5.4.328

Yao CK, Chu NHS, Tan VPY (2018) Breath Hydrogen Testing in East and Southeast Asia J Clin Gastroenterol 52:185-193 doi:10.1097/MCG.0000000000000943

Yap T, Pruthi N, Seers C, Belobrov S, McCullough M, Celentano A (2020) Extracellular Vesicles in Oral Squamous Cell Carcinoma and Oral Potentially Malignant Disorders: A Systematic Review Int J Mol Sci 21 doi:10.3390/ijms21041197

Yen CJ et al. (2020) Two-year follow-up of a randomized phase III clinical trial of nivolumab vs. the investigator's choice of therapy in the Asian population for recurrent or metastatic squamous cell carcinoma of the head and neck (CheckMate 141) Head Neck doi:10.1002/hed.26331

Yi W, Li D, Guo Y, Zhang Y, Huang B, Li X (2016) Sevoflurane inhibits the migration and invasion of glioma cells by upregulating microRNA-637 Int J Mol Med 38:1857-1863 doi:10.3892/ijmm.2016.2797

You Q, Jing X, Fan S, Wang Y, Yang Z (2020) Comparison of functional outcomes and health-related quality of life one year after treatment in patients with oral and

oropharyngeal cancer treated with three different reconstruction methods Br J Oral Maxillofac Surg doi:10.1016/j.bjoms.2020.01.010

Yu EH, Tu HF, Wu CH, Yang CC, Chang KW (2017) MicroRNA-21 promotes perineural invasion and impacts survival in patients with oral carcinoma J Chin Med Assoc 80:383-388 doi:10.1016/j.jcma.2017.01.003

Yu Q, Zhang F, Du Z, Xiang Y (2015a) Up-regulation of serum miR-744 predicts poor prognosis in patients with nasopharyngeal carcinoma Int J Clin Exp Med 8:13296-13302

Yu WF, Wang HM, Lu BC, Zhang GZ, Ma HM, Wu ZY (2015b) miR-206 inhibits human laryngeal squamous cell carcinoma cell growth by regulation of cyclinD2 Eur Rev Med Pharmacol Sci 19:2697-2702

Yuen AP, Lam KY, Wei WI, Lam KY, Ho CM, Chow TL, Yuen WF (2000) A comparison of the prognostic significance of tumor diameter, length, width, thickness, area, volume, and clinicopathological features of oral tongue carcinoma Am J Surg 180:139-143 doi:10.1016/s0002-9610(00)00433-5

Zeng B et al. (2019) LncRNA GAS5 suppresses proliferation, migration, invasion, and epithelial-mesenchymal transition in oral squamous cell carcinoma by regulating the miR-21/PTEN axis Exp Cell Res 374:365-373 doi:10.1016/j.yexcr.2018.12.014

Zeng G, Xun W, Wei K, Yang Y, Shen H (2016) MicroRNA-27a-3p regulates epithelial to mesenchymal transition via targeting YAP1 in oral squamous cell carcinoma cells Oncol Rep 36:1475-1482 doi:10.3892/or.2016.4916

Zeng X et al. (2012) Circulating miR-17, miR-20a, miR-29c, and miR-223 combined as non-invasive biomarkers in nasopharyngeal carcinoma PLoS One 7:e46367 doi:10.1371/journal.pone.0046367

Zhang B, Li Y, Hou D, Shi Q, Yang S, Li Q (2017) MicroRNA-375 Inhibits Growth and Enhances Radiosensitivity in Oral Squamous Cell Carcinoma by Targeting Insulin Like Growth Factor 1 Receptor Cell Physiol Biochem 42:2105-2117 doi:10.1159/000479913

Zhang J, Wu GQ, Zhang Y, Feng ZY, Zhu SM (2013) Propofol induces apoptosis of hepatocellular carcinoma cells by upregulation of microRNA-199a expression Cell Biol Int 37:227-232 doi:10.1002/cbin.10034

Zhang P, Mirani N, Baisre A, Fernandes H (2014) Molecular heterogeneity of head and neck squamous cell carcinoma defined by next-generation sequencing Am J Pathol 184:1323-1330 doi:10.1016/j.ajpath.2014.01.028

Zhang S et al. (2016) Roles of microRNA-124a and microRNA-30d in breast cancer patients with type 2 diabetes mellitus Tumour Biol 37:11057-11063 doi:10.1007/s13277-016-4981-6

Zhang T, Liu M, Wang C, Lin C, Sun Y, Jin D (2011) Down-regulation of MiR-206 promotes proliferation and invasion of laryngeal cancer by regulating VEGF expression Anticancer Res 31:3859-3863

Zhang X, Liu X, Ni X, Feng P, Wang YU (2019) Long non-coding RNA H19 modulates proliferation and apoptosis in osteoarthritis via regulating miR-106a-5p J Biosci 44

Zhang Z, Li X, Xiao Q, Wang Z (2018a) MiR-574-5p mediates the cell cycle and apoptosis in thyroid cancer cells via Wnt/beta-catenin signaling by repressing the expression of Quaking proteins Oncol Lett 15:5841-5848 doi:10.3892/ol.2018.8067

Zhang Z, Wang J, Wang X, Song W, Shi Y, Zhang L (2018b) MicroRNA-21 promotes proliferation, migration, and invasion of cervical cancer through targeting TIMP3 Arch Gynecol Obstet 297:433-442 doi:10.1007/s00404-017-4598-z

Zhao H et al. (2017) Variations in oral microbiota associated with oral cancer Sci Rep 7:11773 doi:10.1038/s41598-017-11779-9

Zhao H et al. (2014) Effects of preanalytic variables on circulating microRNAs in whole blood Cancer Epidemiol Biomarkers Prev 23:2643-2648 doi:10.1158/1055-9965.Epi-14-0550

Zhao L et al. (2015) Deregulation of the miR-222-ABCG2 regulatory module in tongue squamous cell carcinoma contributes to chemoresistance and enhanced migratory/invasive potential Oncotarget 6:44538-44550 doi:10.18632/oncotarget.6253

Zhao YH, Liu YL, Fei KL, Li P (2020) Long non-coding RNA HOTAIR modulates the progression of preeclampsia through inhibiting miR-106 in an EZH2-dependent manner Life Sci 253:117668 doi:10.1016/j.lfs.2020.117668

Zheng YJ, Zhao JY, Liang TS, Wang P, Wang J, Yang DK, Liu ZS (2019) Long noncoding RNA SMAD5-AS1 acts as a microRNA-106a-5p sponge to promote epithelial mesenchymal transition in nasopharyngeal carcinoma Faseb j 33:12915-12928 doi:10.1096/fj.201900803R

Zhou X, Yuan P, Liu Q, Liu Z (2017) LncRNA MEG3 regulates imatinib resistance in chronic myeloid leukemia via suppressing MicroRNA-21 Biomolecules and Therapeutics 25:490-496 doi:10.4062/biomolther.2016.162

Zhu J, Bean HD, Jimenez-Diaz J, Hill JE (2013a) Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study Journal of applied physiology (Bethesda, Md : 1985) 114:1544-1549 doi:10.1152/japplphysiol.00099.2013

Zhu J, Jimenez-Diaz J, Bean HD, Daphtary NA, Aliyeva MI, Lundblad LK, Hill JE (2013b) Robust detection of P. aeruginosa and S. aureus acute lung infections by secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting: from initial infection to clearance J Breath Res 7:037106 doi:10.1088/1752-7155/7/3/037106

Zhu M, Zhang N, He S (2019) Transcription factor KLF4 modulates microRNA-106a that targets Smad7 in gastric cancer Pathol Res Pract 215:152467 doi:10.1016/j.prp.2019.152467

Zhu WY et al. (2014) Differential expression of miR-125a-5p and let-7e predicts the progression and prognosis of non-small cell lung cancer Cancer Invest 32:394-401 doi:10.3109/07357907.2014.922569

Zhu Y et al. (2012) MicroRNA-26a/b and their host genes cooperate to inhibit the G1/S transition by activating the pRb protein Nucleic Acids Res 40:4615-4625 doi:10.1093/nar/gkr1278

Zitvogel L, Galluzzi L, Viaud S, Vetizou M, Daillere R, Merad M, Kroemer G (2015) Cancer and the gut microbiota: an unexpected link Sci Transl Med 7:271ps271 doi:10.1126/scitranslmed.3010473

zur Hausen H (1996) Papillomavirus infections--a major cause of human cancers. Biochimica et biophysica acta, 1288(2), F55–F78. https://doi.org/10.1016/0304-419x(96)00020-0

zur Hausen H (2009) Papillomaviruses in the causation of human cancers - a brief historical account Virology 384:260-265 doi:10.1016/j.virol.2008.11.046

Zwiers A, Kraal L, van de Pouw Kraan TC, Wurdinger T, Bouma G, Kraal G (2012) Cutting edge: a variant of the IL-23R gene associated with inflammatory bowel disease induces loss of microRNA regulation and enhanced protein production J Immunol 188:1573-1577 doi:10.4049/jimmunol.1101494

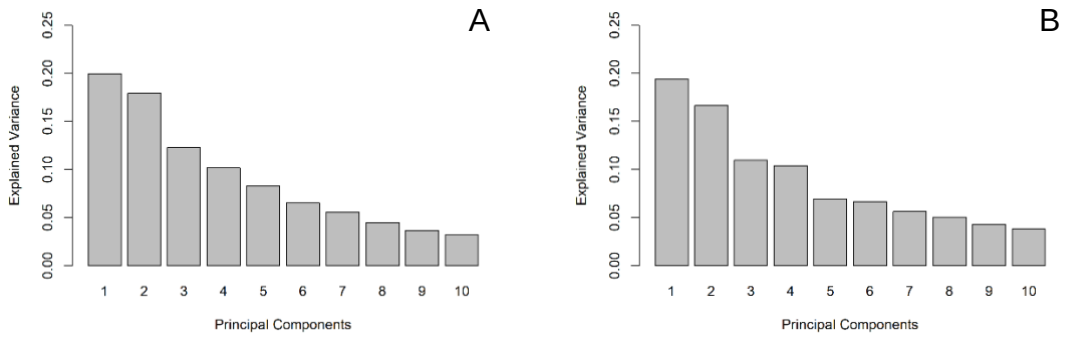
# Appendices

# Appendix I. Detailed data tables and statistical analyses for microRNA datasets

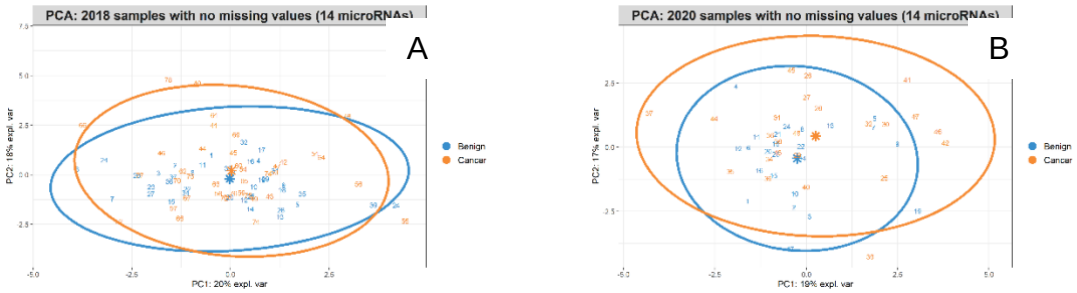
***Appendix I - Table 1. Comparison of microRNA variables between cancer and control patient groups***

| | variables | statistic | p | p.adj | p.adj.signif |
|---|---|---|---|---|---|
| 57 | miR60 | 1441.0 | 0.0105 | 0.2423600 | ns |
| 56 | miR6 | 740.5 | 0.0129 | 0.2423600 | ns |
| 55 | miR59 | 775.0 | 0.0160 | 0.2423600 | ns |
| 67 | miR7 | 790.0 | 0.0165 | 0.2423600 | ns |
| 11 | miR19 | 1414.0 | 0.0185 | 0.2423600 | ns |
| 32 | miR38 | 801.5 | 0.0207 | 0.2423600 | ns |
| 72 | miR8 | 785.0 | 0.0245 | 0.2423600 | ns |
| 59 | miR62 | 808.0 | 0.0306 | 0.2423600 | ns |
| 65 | miR68 | 1366.0 | 0.0307 | 0.2423600 | ns |
| 25 | miR31 | 1405.5 | 0.0332 | 0.2423600 | ns |
| 1 | miR1 | 1356.0 | 0.0371 | 0.2462091 | ns |

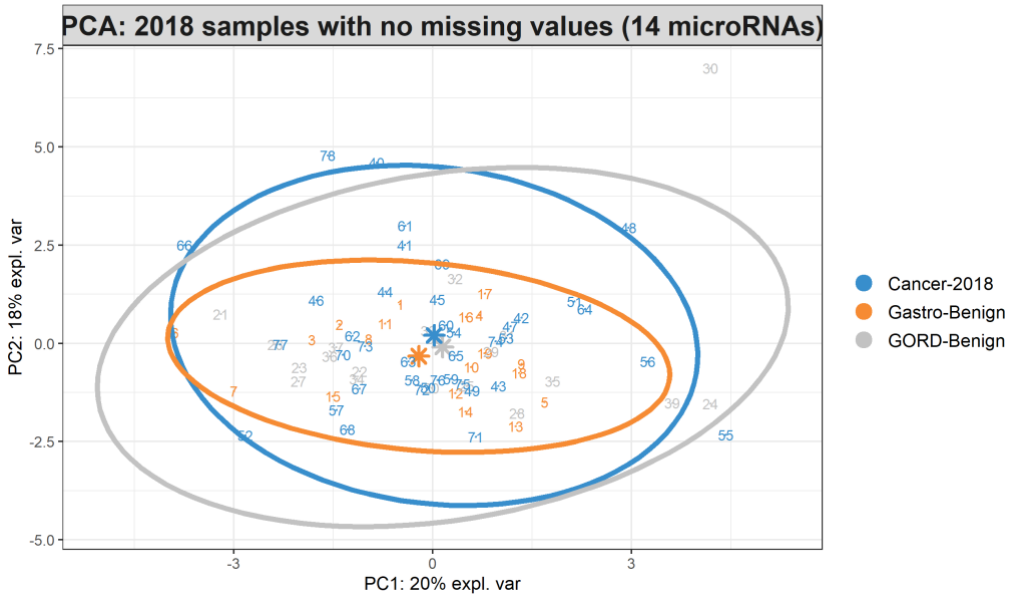
Wilcoxon signed rank test statistic, p value adjusted using Benjamini Hochberg method for multiple t-tests. cancer group (n = 62), control group (n = 36). Only statistically significant results are displayed here.



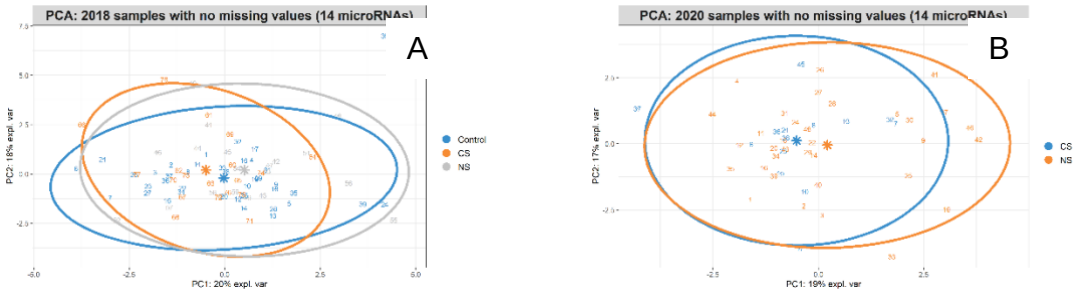

**Appendix I - Figure 1. Comparison of explained variance between 2018 (A) and 2020 (B) datasets.**



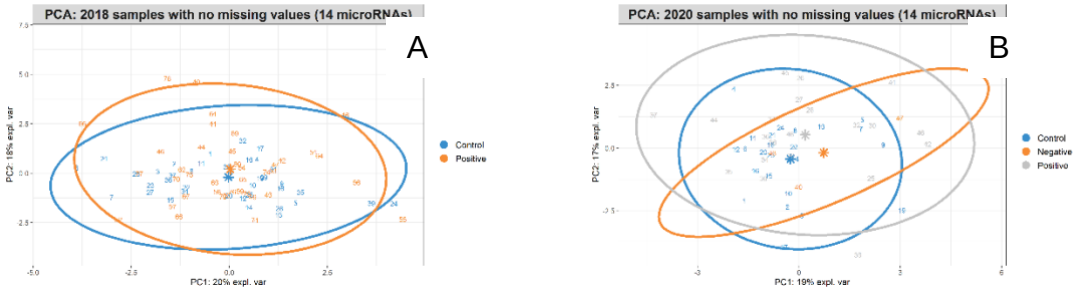

**Appendix I - Figure 2. Comparison of classification based on diagnosis on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.**



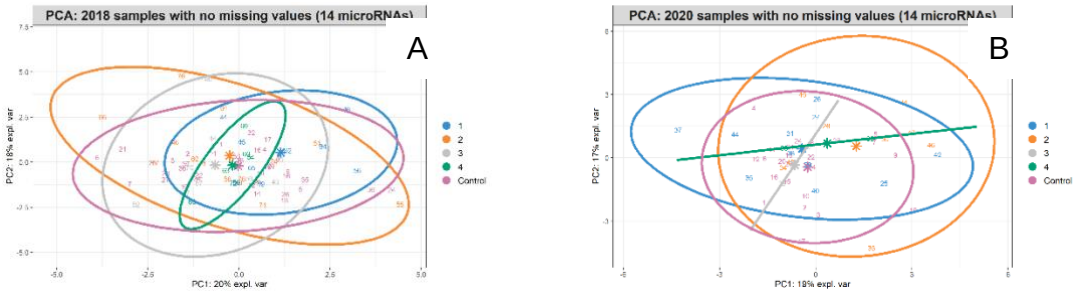

***Appendix I - Figure 3PCA of the 2018 dataset labelled according to the sample cohort. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***



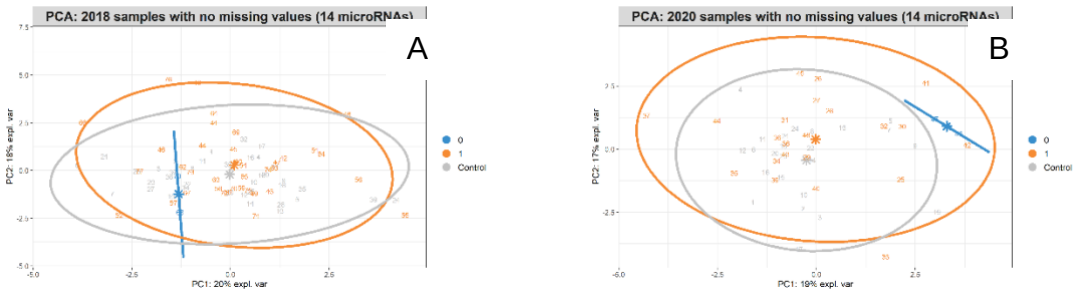

***Appendix I - Figure 4. Comparison of classification based on smoking status on PCA between 2018 (A) and 2020 (B) datasets. CS, current smokers. NS, non-smokers. Smoking status data for the control group is not available in the 2018 dataset. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***



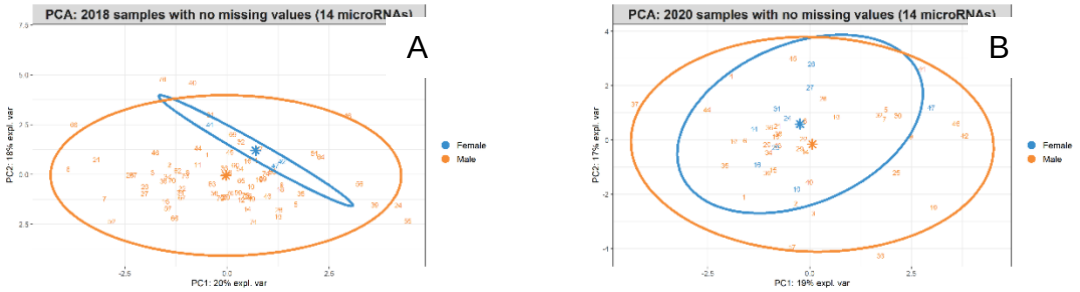

***Appendix I - Figure 5. Comparison of classification based on p16 status on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***



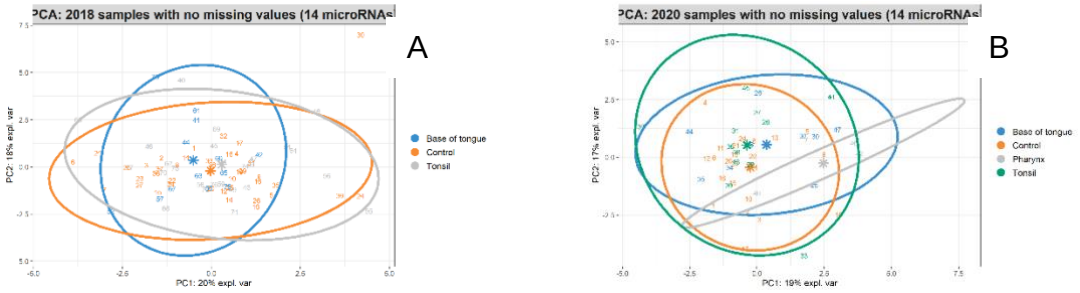

***Appendix I - Figure 6. Comparison of classification based on T-stage on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***




***Appendix I - Figure 7. Comparison of classification based on N-status on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid. 0, node negative. 1, node positive.***




***Appendix I - Figure 8. Comparison of classification based on gender on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***




***Appendix I - Figure 9. Comparison of classification based on primary tumour subsite on PCA between 2018 (A) and 2020 (B) datasets. Star (*) represents the centroid value of the dependent variable and ellipses represent a 95% confidence interval from the respective centroid.***

# Appendix II. Search strategies and keywords for systematic review of literature

## Search Strategies for Chapter 2

**PubMed Search Strategy:**
("microRNA"[tiab] OR "miRNA"[tiab] OR "micro RNA"[tiab] OR "mi RNA"[tiab] OR

"mi-RNA"[tiab] OR "micro-RNA"[tiab] OR "miR"[tiab] OR "MicroRNAs"[mesh] OR

"Circulating MicroRNA"[mesh])

AND

("blood"[tiab] OR "serum"[tiab] OR "plasma"[tiab] OR "Blood"[mesh] OR

"Serum"[mesh] OR "Plasma"[mesh])

AND

(cancer*[tiab] OR carcinoma*[tiab] OR malign*[tiab] OR "Carcinoma"[mesh])

**Embase Search Strategy:**
('microRNA':ti,ab OR 'miRNA':ti,ab OR 'micro RNA':ti,ab OR 'mi RNA':ti,ab OR

'mi-RNA':ti,ab OR 'micro-RNA':ti,ab OR 'miR':ti,ab OR 'microRNA'/exp OR

'circulating microRNA'/exp)

AND

('blood':ti,ab OR 'serum':ti,ab OR 'plasma':ti,ab OR 'blood'/exp)

AND

(cancer*:ti,ab OR carcinoma*:ti,ab OR malign*:ti,ab OR 'malignant neoplasm'/exp)

**Search Strategies for Chapter 4**

**PubMed Search Strategy:**
"cancer"[Title/Abstract] OR "carcinoma"[Title/Abstract]

AND

(((((((("head neck"[Journal] OR (("head "[All Fields]) AND "neck"[All Fields])) OR "head and neck"[All Fields]) OR ("neck"[MeSH Terms] OR "neck"[All Fields])) OR (("mouth"[MeSH Terms] OR "mouth"[All Fields]) OR "oral"[All Fields])) OR "orophary*"[All Fields]) OR "lary*"[All Fields]) OR "phary*"[All Fields]) OR "glot*"[All Fields]) OR ((("tongue"[MeSH Terms] OR "tongue"[All Fields]) OR "tongues"[All Fields]) OR "tongue s"[All Fields])

AND

"breath"[Title/Abstract] OR "exhaled"[Title/Abstract] OR "condensate"[Title/Abstract] OR "volatile"[Title/Abstract]

**Embase (All) /Scopus (All) /ProQuest (All) /Web of Science (Core Collection) Search Strategy:**
'exhaled breath condensate' OR 'volatile organic compound' OR 'breath analysis' or ‘breath’ OR ‘exhaled’ OR ‘condensate’ OR ‘volatile’

AND

'head and neck tumor' OR 'mouth tumor' OR 'oropharynx cancer' OR 'larynx cancer' OR 'pharynx cancer' OR 'tongue tumor' OR ‘head’ OR ‘neck’ OR ‘mouth’ OR ‘oral’ OR ‘oropharynx’ OR ‘larynx’ OR ‘pharynx’ OR ‘glottis’ OR ‘glottic’ OR ‘tongue’

AND

'malignant neoplasm'/exp OR 'malignant neoplasm' OR 'carcinoma'/exp OR 'carcinoma' OR cancer

# Appendix III. Detailed data tables and statistical analyses for breath biomarker datasets

***Appendix III - Table 1. Comparison of significantly different patient factors between training and testing patient cohorts***

| Variable | Training | | P Value | Testing | | P Value |
|---|---|---|---|---|---|---|
| | Control | Cancer | | Control | Cancer | |
| n | 36 | 40 | | 14 | 10 | |
| Sex | | | | | | |
| Female | 16 | 6 | 0.005 [a] * | 9 | 1 | 0.008 [a] * |
| Male | 20 | 34 | | 5 | 9 | |
| BMI | 30.22 | 26.51 | 0.005 [b] * | 27.5 | 23.44 | 0.026 [b] * |
| Brush Teeth | | | | | | |
| Yes | 29 | 26 | 0.130 [a] | 1 | 5 | 0.022 [a] * |
| No | 7 | 14 | | 12 | 5 | |

a, Chi-square test. b, Mann-Whitney-U test. *, statistical significance ($p < 0.05$).

***Appendix III - Table 2. Variables selected and ranked based on ROC-AUC per reagent***

| | Variable Name | | |
|---|---|---|---|
| Rank | R19 ($H_3O^+$) | R30 ($NO^+$) | R32 ($O2^+$) |
| 1 | R19P49 | R30P147 | R32P135 |
| 2 | R19P125 | R30P143 | R32P109 |
| 3 | R19P30 | R30P187 | R32P82 |
| 4 | R19P189 | R30P211 | R32P138 |
| 5 | R19P59 | R30P240 | R32P31 |
| 6 | R19P18 | R30P64 | R32P47 |
| 7 | R19P81 | R30P188 | R32P184 |
| 8 | R19P48 | R30P80 | R32P40 |
| 9 | R19P133 | R30P95 | R32P103 |
| 10 | R19P77 | R30P159 | R32P95 |

## *Appendix III - Table 3. Variables selected based on overall ROC-AUC.*

| Rank | Variable Name |
|---|---|
| 1 | R30P147 |
| 2 | R19P49 |
| 3 | R30P143 |
| 4 | R30P187 |
| 5 | R19P125 |
| 6 | R30P211 |
| 7 | R30P240 |
| 8 | R32P135 |
| 9 | R19P30 |
| 10 | R32P109 |
| 11 | R30P64 |
| 12 | R30P188 |
| 13 | R30P80 |
| 14 | R32P82 |
| 15 | R30P95 |
| 16 | R30P159 |
| 17 | R32P138 |
| 18 | R32P31 |

***Appendix III - Table 4. Prediction based on gender and BMI.***

| | Train | n=76 | | | Test | n=24 | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | 95% CI | Sensitivity | Specificity | AUC | 95% CI |
| G | 85 | 44.4 | 0.647 | 0.521-0.773 | 90 | 64.3 | 0.771 | 0.577-0.966 |
| BMI | 41 | 80.6 | 0.654 | 0.530-0.777 | 60 | 64.3 | 0.693 | 0.482-0.904 |
| G + BMI | 87.2 | 47.2 | 0.725 | 0.610-0.841 | 90 | 57.1 | 0.764 | 0.575-0.954 |

G, gender. BMI, body mass index.

***Appendix III - Table 5. Variables selected based on AUC ranking per reagent ion.***

| | Train | n=76 | | | | Test | n=24 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI |
| 30 | 100.00% | 100.00% | 1 | 1 | 1 | 40.00% | 64.30% | 0.529 | 0.284 | 0.773 |
| 27 | 100.00% | 100.00% | 1 | 1 | 1 | 60.00% | 64.30% | 0.579 | 0.332 | 0.826 |
| 24 | 95.00% | 88.90% | 0.978 | 0.953 | 1 | 70.00% | 71.40% | 0.621 | 0.387 | 0.856 |
| 21 | 82.50% | 80.60% | 0.929 | 0.876 | 0.983 | 70.00% | 50.00% | 0.493 | 0.243 | 0.743 |
| 18 | 80.00% | 69.40% | 0.874 | 0.799 | 0.949 | 80.00% | 57.10% | 0.814 | 0.625 | 1 |
| 15 | 75.00% | 72.20% | 0.862 | 0.783 | 0.94 | 90.00% | 85.70% | 0.9 | 0.748 | 1 |
| 12 | 67.50% | 69.40% | 0.815 | 0.722 | 0.908 | 70.00% | 78.60% | 0.779 | 0.587 | 0.97 |
| 9 | 75.00% | 77.80% | 0.806 | 0.705 | 0.908 | 80.00% | 71.40% | 0.786 | 0.585 | 0.987 |
| 6 | 75.00% | 69.40% | 0.774 | 0.668 | 0.881 | 40.00% | 64.30% | 0.586 | 0.342 | 0.83 |
| 3 | 75.00% | 66.70% | 0.754 | 0.645 | 0.863 | 80.00% | 85.70% | 0.821 | 0.625 | 1 |

***Appendix III - Table 6. Variables selected based on overall AUC ranking.***

| | Train | n=76 | | | | Test | n=24 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI | Sensitivity | Specificity | AUC | Lower 95% CI | Upper 95% CI |
| 10 | 72.50% | 75.00% | 0.817 | 0.723 | 0.912 | 80.00% | 71.40% | 0.707 | 0.482 | 0.932 |
| 9 | 75.00% | 72.20% | 0.808 | 0.711 | 0.904 | 80.00% | 71.40% | 0.729 | 0.507 | 0.95 |
| 8 | 70.00% | 63.90% | 0.795 | 0.697 | 0.893 | 70.00% | 64.30% | 0.671 | 0.441 | 0.902 |
| 7 | 72.50% | 63.90% | 0.797 | 0.699 | 0.894 | 60.00% | 64.30% | 0.657 | 0.42 | 0.894 |
| 6 | 77.50% | 63.90% | 0.79 | 0.687 | 0.892 | 50.00% | 71.40% | 0.629 | 0.384 | 0.874 |
| 5 | 72.50% | 69.40% | 0.785 | 0.683 | 0.888 | 50.00% | 71.40% | 0.664 | 0.422 | 0.907 |
| 4 | 70.00% | 66.70% | 0.752 | 0.643 | 0.861 | 80.00% | 64.30% | 0.857 | 0.688 | 1 |
| 3 | 72.50% | 66.70% | 0.737 | 0.624 | 0.851 | 70.00% | 78.60% | 0.779 | 0.571 | 0.986 |
| 2 | 72.50% | 66.70% | 0.747 | 0.633 | 0.86 | 80.00% | 85.70% | 0.814 | 0.611 | 1 |
| 1 | 80.00% | 36.10% | 0.702 | 0.583 | 0.822 | 80.00% | 71.40% | 0.725 | 0.485 | 0.965 |

***Appendix III - Table 7. Linear regression between included model variables and BMI on testing cohort***

| Variables | Adjusted $R^2$ | F statistic | P value |
|---|---|---|---|
| R32P135 + R30P147+R19P49 | 0.041 | 0.696 | 0.565 |
| R32P135 + R19P49 | 0.014 | 0.838 | 0.446 |
| R30P147 + R19P49 | 0.016 | 0.82 | 0.454 |
| R30P147 + R32P135 | 0.03 | 0.67 | 0.523 |
| R19P49 | 0.016 | 0.643 | 0.431 |
| R30P147 | 0.01 | 1.242 | 0.277 |
| R32P135 | 0.024 | 0.455 | 0.507 |

***Appendix III - Table 8. Predict gender (male) using final model variables (R30P147 and R19P49)***

| | Train | n=76 | | | Test | n=24 | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Sensitivity | Specificity | AUC | 95% CI | Sensitivity | Specificity | AUC | 95% CI |
| R30P147 | 100% | 0% | 0.582 | 0.449-0.715 | 100% | 0% | 0.5 | 0.261-0.739 |
| R19P49 | 100% | 0% | 0.581 | 0.439-0.724 | 100% | 0% | 0.5 | 0.261-0.739 |
| R19P49 + R30P147 | 100% | 0% | 0.591 | 0.458-0.723 | 100% | 0% | 0.5 | 0.261-0.739 |

# Appendix VI. Measurement of room air volatile organic compounds and statistical analyses

## Introduction

Environmental contamination of endogenous VOCs of human breath samples is an ongoing concern. While it is important to collect an ambient air sample with standardised methodology to compare to the human samples, there are no agreed upon methods of incorporating these ambient levels during data analysis. Exclusion of ambient VOCs present in high concentrations from further endogenous analysis is a common method and has been utilised in this thesis. Although, for mechanistic understanding of how VOCs contribute to disease process requires better methods of incorporating ambient VOCs into data analysis pipeline.

## Methods

As part of ongoing research into the exhaled VOCs produced in HNSCC, patients were recruited from a number of tertiary hospitals in Adelaide, South Australia. More specifically from Flinders Medical Centre, Women's and Children's Hospital and the Royal Adelaide Hospital. Prior to obtaining a patient breath sample as described in a previous chapter, a room air sample was collected for concurrent analysis. In this section, comparison of room air to patient breath for key compounds of interest and site and date dependent changes to room air concentrations were explored.

Prior to collecting breath or sampling room air, each of the FlexFoil PLUS 3 litre bags were flushed with pure (min. 99.999%) nitrogen ($N_2$) gas to ensure that there were no contaminating VOCs within the bag at baseline. This was performed in accordance with the manufacturer's instructions. Room air was sampled concurrently from the same location via a 3-way valve connected to a passivated

glass gas-tight 100mL syringe and a separate N2 flushed FlexFoil PLUS 3 litre bag. Each bag was filled with a minimum of 2 litres of breath or room air sample. The bags were then stored in an insulated bag pre-heated to 37°C using Deltaphase® isothermal pads and then transferred to an incubator until the samples were able to be processed. Maintaining a temperature of 37°C prevents the condensation of VOCs out of the gas samples. Each sample was then processed using the Mass Scan function (Molecular mass range 15-250) with a Syft Voice 200® mass spectrometer. Full mass scan mode was utilised, and VOCs calculated based on the curated library provided by Syft.

**Results and Discussion**

Standardisation of breath collection methodology is crucial to ensure comparable, and reproducible data that can underpin translation. The European Respiratory Society guidelines for exhaled breath analysis was an attempt to standardise breath research methodology and analysis (Horváth *et al*. 2017). One of the recommendations was that ambient VOC concentrations should be recorded when sampling exhaled breath to account for environmental VOCs that may confound the collected breath sample.

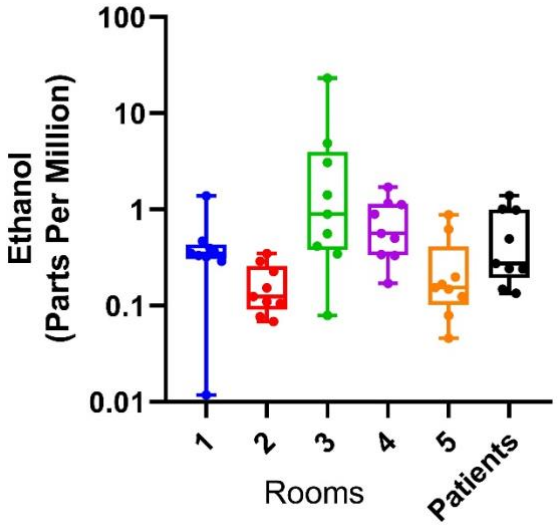
Ethanol is one of the most commonly reported breath VOC markers whiles concurrently being abundant in ambient air. For example, the comparison of ambient ethanol levels to exhaled patient breath ethanol levels shows there is a large variability between patient rooms and at time the ambient ethanol levels are elevated compared to human breath (Appendix IV - Figure 1). Statistical analysis indicated no significant differences between the human breath and individual room air levels of ethanol ($p > 0.050$). However, the ethanol levels in the RAH day

surgery area (Room 3) was significantly elevated compared to the RAH critical care area (Room 2, p = 0.0100) and the WCH outpatient area (Room 5, p = 0.0030). A highly volatile compound such as ethanol present at or above the exhaled breath level suggests potential for contamination of the exhaled breath sample.

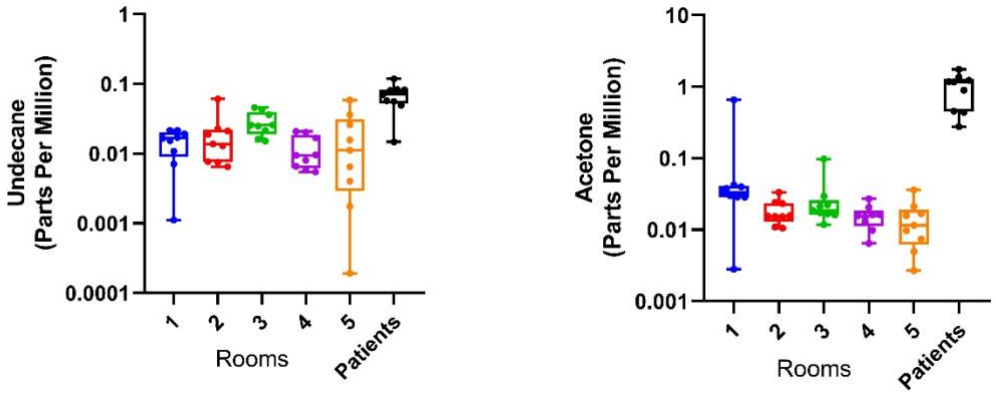
In comparison, ambient undecane levels were significantly lower compared to exhaled breath (Appendix IV - Figure 2A, p = 0.0003). Undecane is primarily detected in crude oil and plants, and importantly identified as a component of cigarette smoke and exposure to fuel fumes (National Centre for Biotechnology Information, 2020). The ambient undecane levels in the RAH day surgery area (Room 3) was not significantly different from patients' breath, while the undecane levels in all other rooms were significantly lower. This again highlights the variability of previously reported biomarkers in ambient air.

Acetone levels were also significantly lower in the ambient samples compared to human breath (Appendix IV - Figure 2B, p = 0.0001). Acetone is a known product of animal fat metabolism, found in trace concentrations in human blood, breath, and urine (National Centre for Biotechnology Information, 2020). However, acetone is also a powerful solvent present in many adhesives and cleaning product and therefore, anticipated to fluctuate in ambient air levels (National Centre for Biotechnology Information, 2020). Although, when exhaled breath levels of acetone is significantly high (Appendix IV – Figure 2), it has less concerning influence on the collected breath sample, despite the observed variability in the room air.

These gases illustrate the point that ambient levels of VOCs vary across clinical locations and in comparison, to endogenous exhaled breath levels. Therefore, highlighting the importance of measuring room air levels of VOCs during breath collection. It is important to consider the ambient VOC levels in most cases, especially if their ambient levels are higher than exhaled breath.

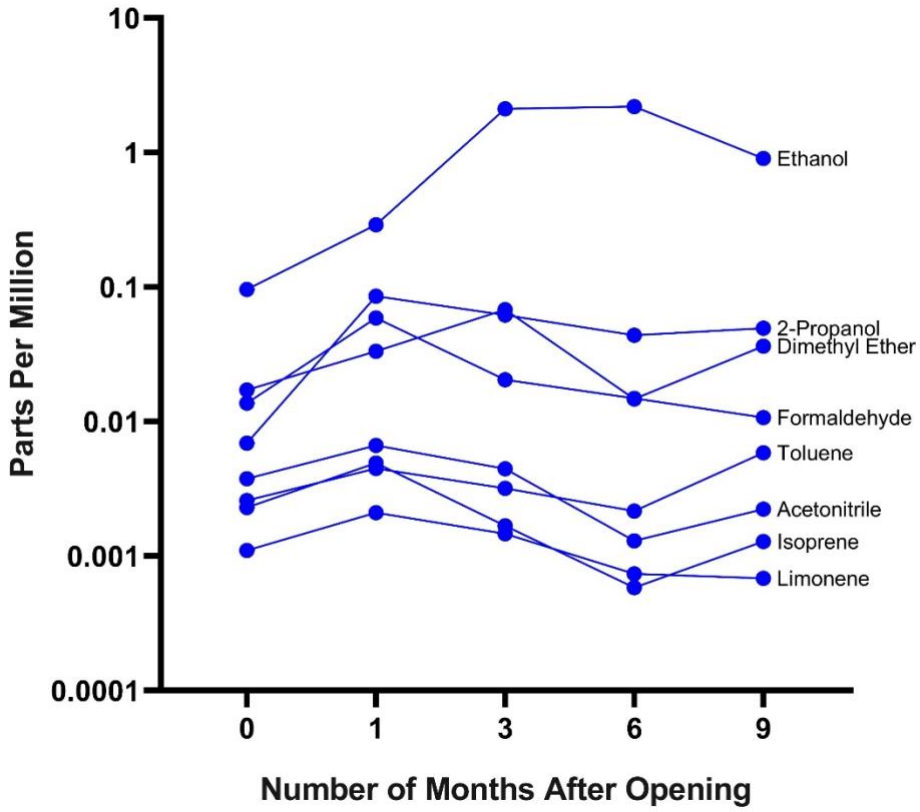



***Appendix IV - Figure 1. Room air ethanol levels compared to exhaled patient breath. Rooms 1 – 5 were 1. FMC day surgery, 2. RAH critical care, 3. RAH day surgery, 4. WCH inpatient ward and 5. WCH outpatient ward. Nine samples from each room at various times of the year and nine healthy adult control patients' data were used here. Median, interquartile range and the range are displayed with individual dots corresponding to each data point. FMC, Flinders Medical Centre. RAH, Royal Adelaide Hospital. WCH, Women's and Children's Hospital.***




***Appendix IV - Figure 2. Room air undecane (A) and acetone (B) levels compared to exhaled patient breath. Rooms 1 – 5 were 1. FMC day surgery, 2. RAH critical care, 3. RAH day surgery, 4. WCH inpatient ward and 5. WCH outpatient ward. Nine samples from each room at various times of the year and nine healthy adult control patients' data were used here. Median, interquartile range and the range are displayed with individual dots corresponding to each data point. FMC, Flinders Medical Centre. RAH, Royal Adelaide Hospital. WCH, Women's and Children's Hospital.***

During the data collection period a newly constructed public hospital was opened in South Australia. Therefore, ambient air samples were collected from specific rooms over a nine-month period as part of this project for quality control. The figure below (Appendix IV - Figure 3) indicates the ambient levels of various volatile compounds from an intensive care patient room. The ambient air levels reported here are comparable to previously reported values based on hospital measurements and further highlights the importance of measuring ambient air levels, especially if alcohols, ethers, and aldehydes are compounds of interest as these compounds appear to vary the most and present in relatively high levels in ambient air.




***Appendix IV - Figure 3. Variability in VOCs measured over a nine-month period in a newly constructed hospital. X-axis indicates the number of months after the hospital was opened to the public.***

Different approaches to correct for ambient VOC concentrations have been previously reported. Subtraction of the ambient VOC concentration from that in exhaled breath has been proposed to determine if the VOC is exogenous or endogenous in origin (Phillips 1997). However, a major limitation of this approach is that it does not account for any interactions the VOC might have with the individuals metabolism as they inhale and exhale, and it assumes that the patient has equilibrated with the room air. Other approaches using inhalation/exhalation filters have been investigated to reduce the ambient VOC concentration being inhaled by the patient (Miekisch, Schubert *et al*. 2004). The use of a filter device can be standardised across experiments without affecting trace VOCs, however, the subtraction method is limited to use with VOCs that are lower in ambient

samples compared to exhaled breath. As there are no high-level evidence directly comparing these methods, it is difficult to recommend a superior approach. However, from practical perspective breathing through an inhalation filter device prior to sample collection is the most attractive.

# Appendix V. Patient dependent factors associated with breath, blood, and microbiome markers.

## Introduction

In clinical research, patient factors often play a key role in interpreting experimental data. Since, sample size is inversely proportional to the influence of fixed patient factors, it is pertinent to understand and explore significant associations. I aimed to determine any relationships between the breath biomarkers measured in our patient cohort and fixed patient factors including cancer specific risk factors for HNSCC.

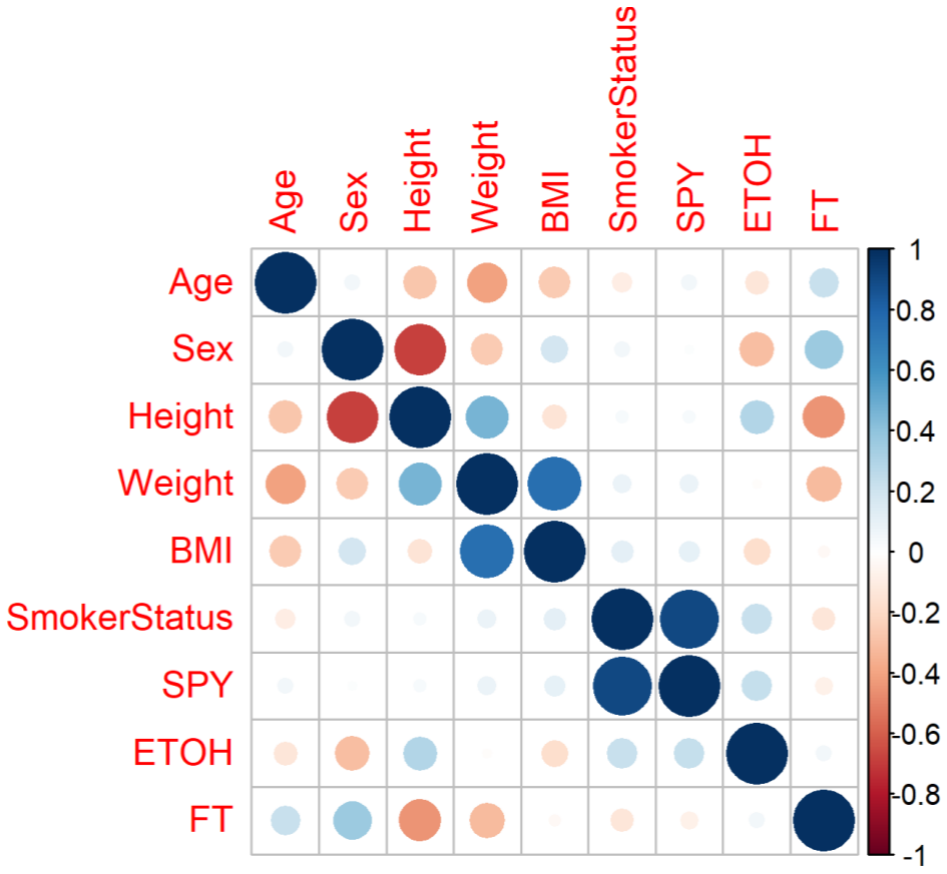
## Methods

The biomarker datasets described in this appendix were analysed using R open-source statistical package (version 3.6.1) and R-Studio interface (version 1.2.5019). Non-parametric correlation coefficients were calculated (Spearman's rho), expressed as a percentage for graphical presentation. The top 20 correlates are displayed unless more than 20 variables had a correlation coefficient larger than 0.5 (50%). Bootstrapped variables (dummy) were created as part of the analysis to test R-coding consistency. Any moderate (coefficient > 0.5) to strong (coefficient > 0.7) associations based on the initial analysis were further explored with further statistical tests as described below.

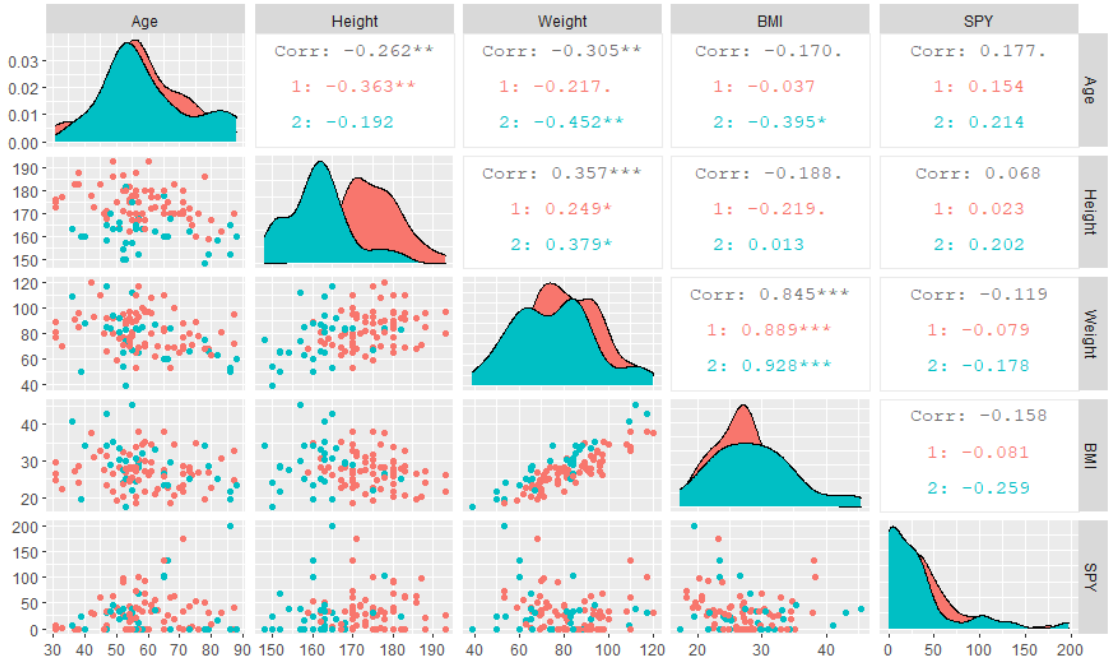
## Results and Discussion

Spearman cross correlation between patient fixed factors is displayed in Appendix V - Figure 1 where most fixed factors did not correlate with each other. However, a strong negative correlation was found between height and their gender, where females were shorter in this study population. Interestingly, body mass index

(BMI) correlated well with weight but not with height, despite being a core component of its calculation (Appendix V - Figure 2).




***Appendix V - Figure 1. Spearman correlation between fixed patient factors. BMI, body mass index. SPY, smoking pack years. ETOH, alcohol intake index. FT, fasting time. Dark blue indicates a strong positive correlation and dark red indicates a strong negative correlation. Lighter shades of colour blue of red indicate diminishing correlation coefficients with white indicating a correlation coefficient of zero. Size of the dot also indicates the magnitude of the correlation.***

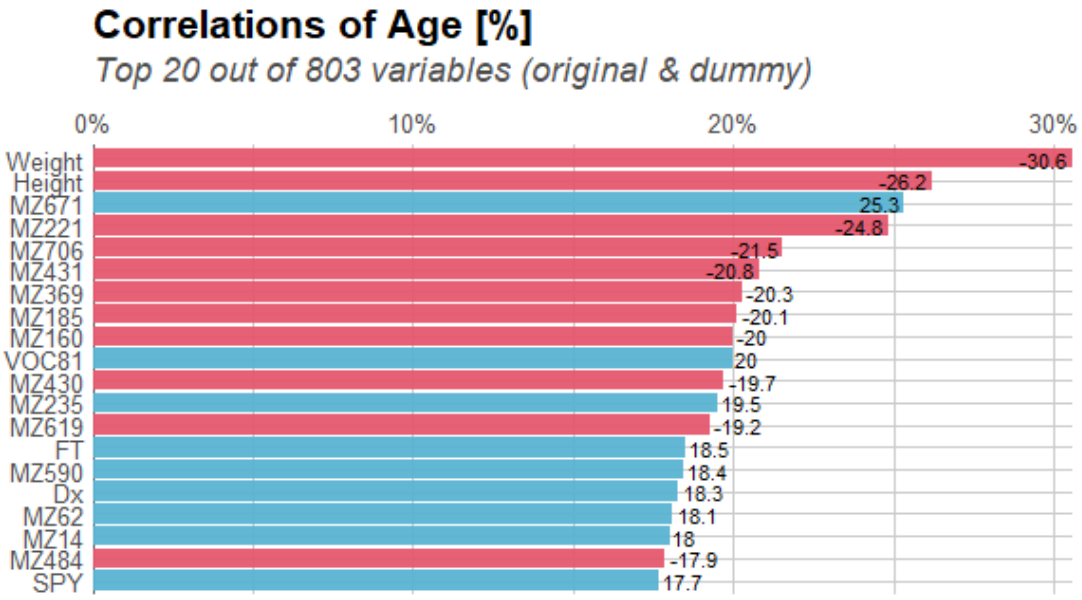



***Appendix V - Figure 2. Spearman correlation coefficients between continuous fixed factors of control (benign) patients. Colour based on gender (1 = male, 2 = female) *, p < 0.05. **, p < 0.01. ***, p < 0.001***

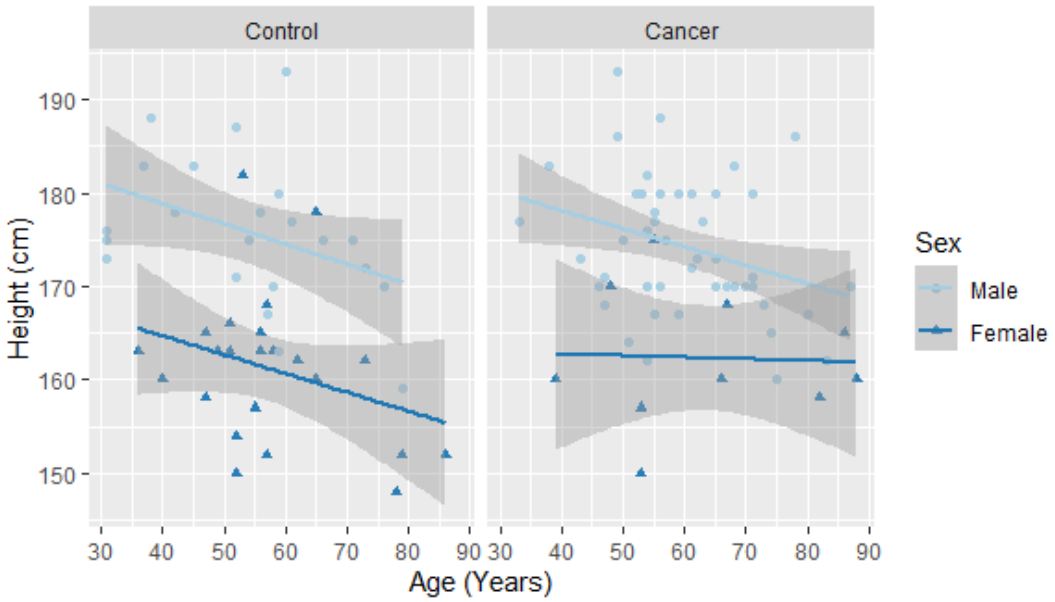
## Age

Patient age did not have any strong correlation to any of the other patient factors or biological markers of interest (Appendix V - Figure 3). However, there was a clear decrease in height (Appendix V - Figure 4) and weight (Appendix V - Figure 5) with advanced age. This was true for both male and female patients in the control group, however, the female patients with cancer had a lower height and weight regardless of the age compared to control patients and male patients with cancer. However, this may be due to sampling bias as there was only a small number of patients in the female cancer cohort of this study. The BMI was consistent with height and weight changes noted earlier (Appendix V - Figure 6). The age distribution between cancer and control groups as well as between males and females were not significantly different (Appendix V - Figure 7). One of the

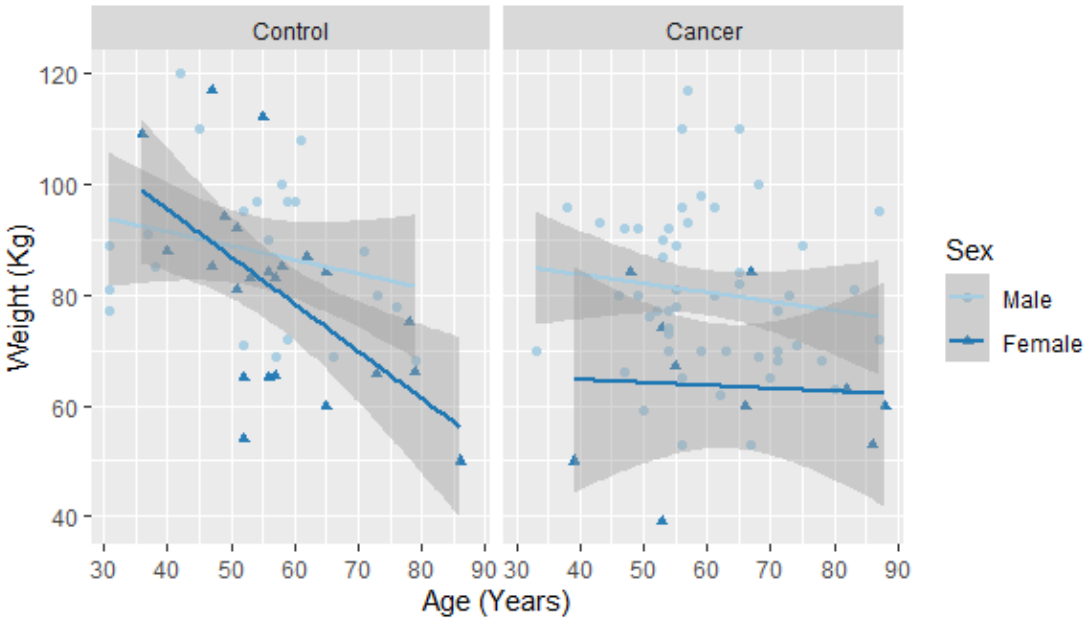
microRNAs of interest from Chapter 3 correlated well with patient age (Appendix V - Figure 8).



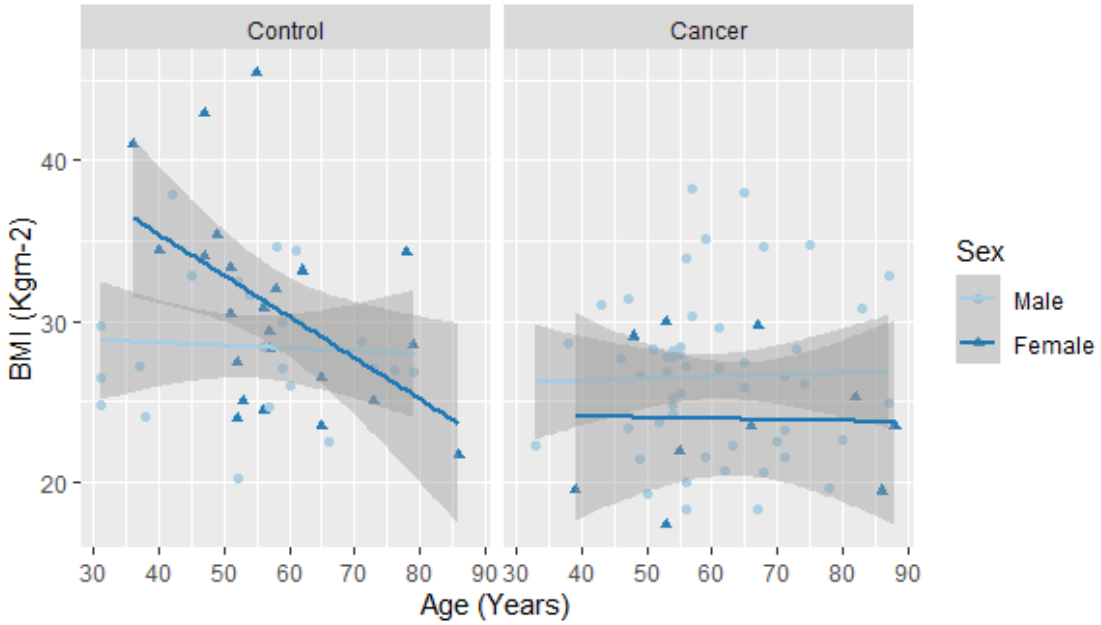

**Appendix V - Figure 3. Spearman correlation of age to other variables. Correlation coefficients expressed as a percentage for display purposes. FT, fasting time. Dx, diagnosis (cancer vs control). SPY, smoking pack years. Variables starting with MZ corresponds to a product mass to charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions.**



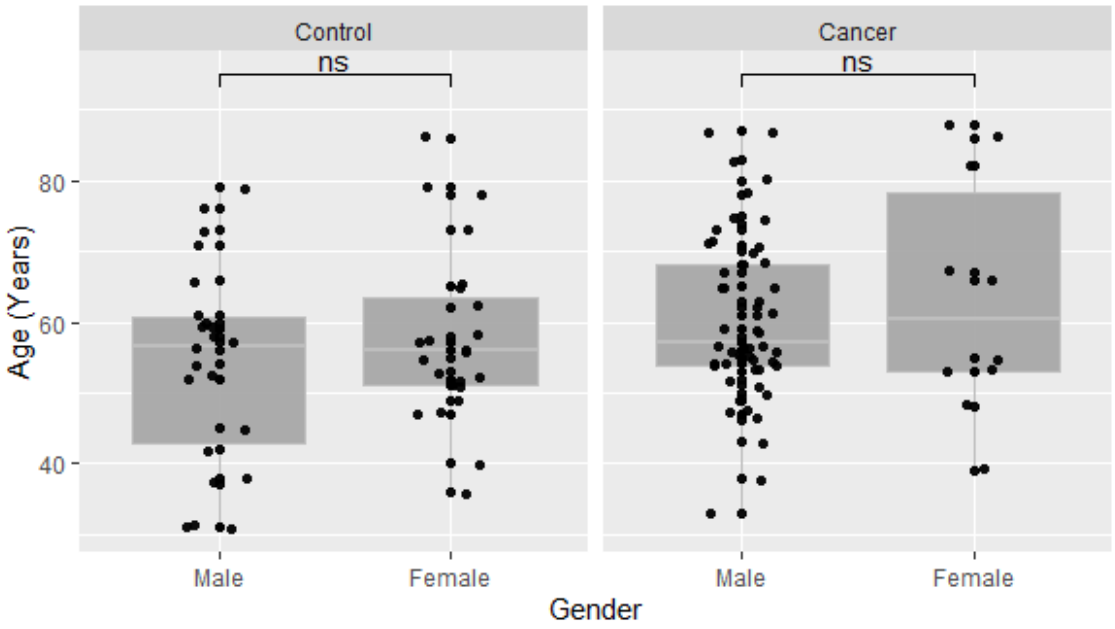

***Appendix V - Figure 4. Age dependent changes to height in cancer and control groups related to their gender.***




***Appendix V - Figure 5. Age dependent changes to weight in cancer and control groups related to their gender.***




***Appendix V - Figure 6. Age dependent changes to BMI in cancer and control groups related to their gender.***

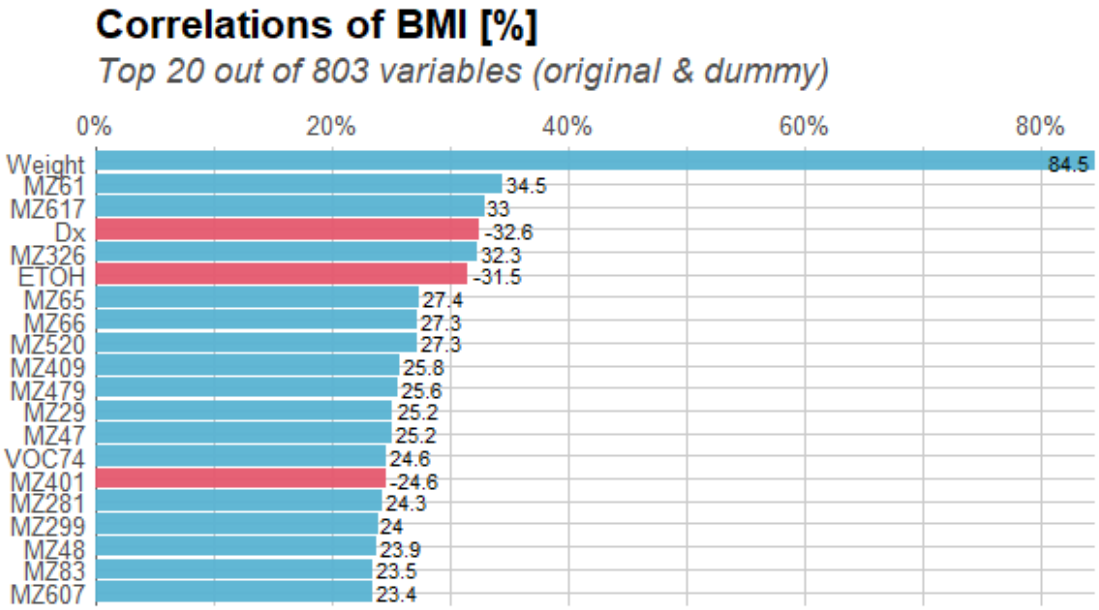



***Appendix V - Figure 7. Gender dependent age distribution between cancer and control groups.***

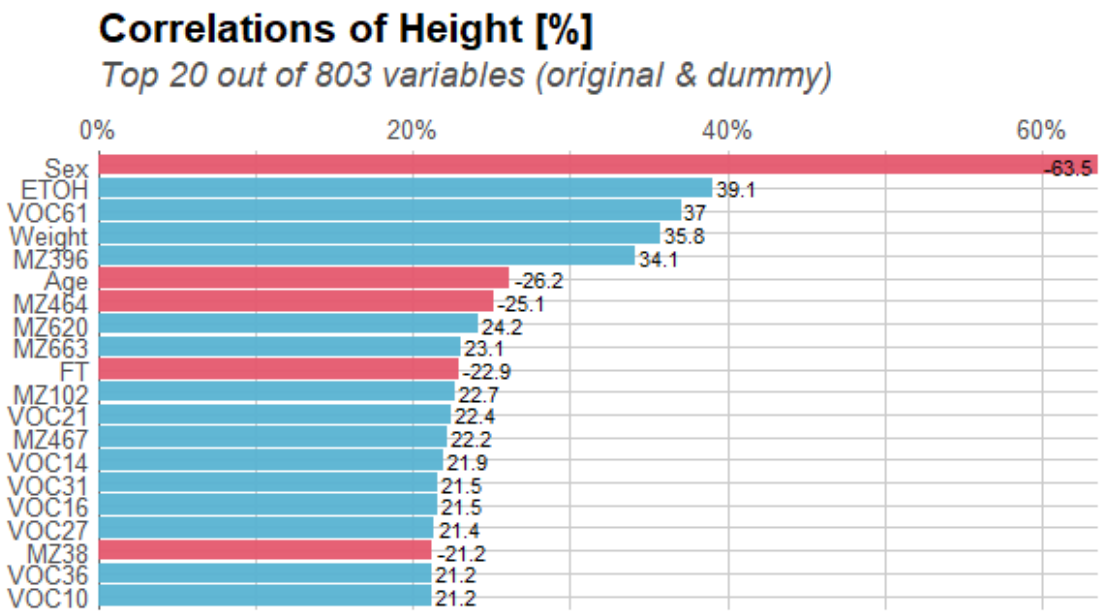

***Appendix V - Figure 8. Correlation of patient age with microRNAs of interest. CS, Current smokers. ES, Ex-smokers. NS, none smokers.***

**Body mass index, height, and weight**

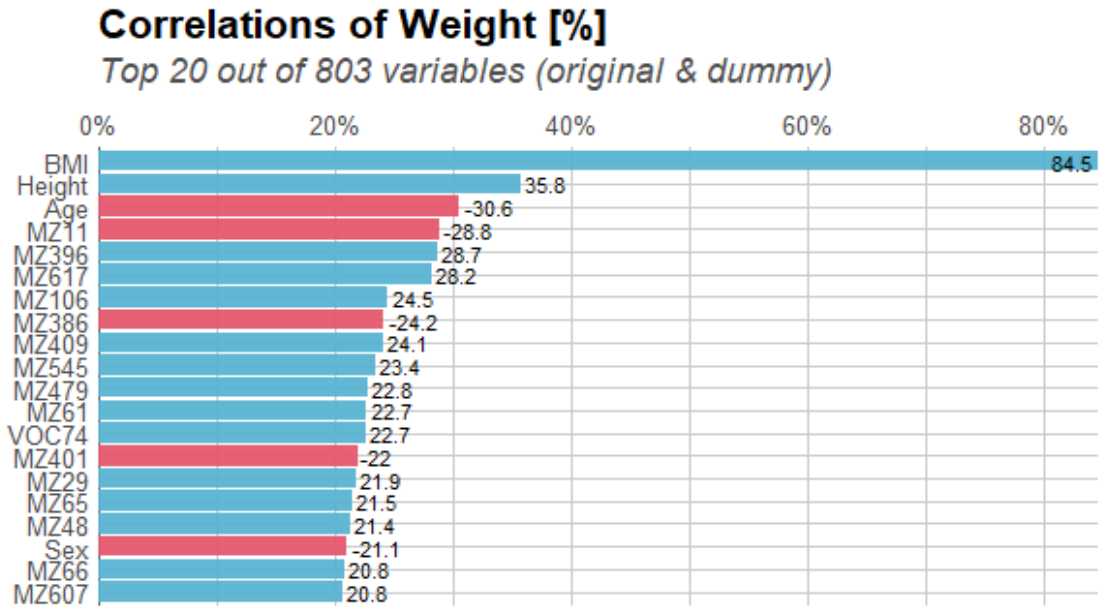
Patient BMI was only highly correlated with patient weight (Appendix V - Figure 9). None of the blood or breath biomarkers considered in this thesis had a strong correlation to BMI Appendix V - Figure 11. Patient height correlated well with their gender but not with any other factors investigated here (Appendix V - Figure 10).




***Appendix V - Figure 9. Spearman correlation of body mass index to other variables. Correlation coefficients expressed as a percentage for display purposes – see appendix for coefficient tables. Dx, diagnosis (cancer vs control). Variables starting with MZ corresponds to a product mass to charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions.***



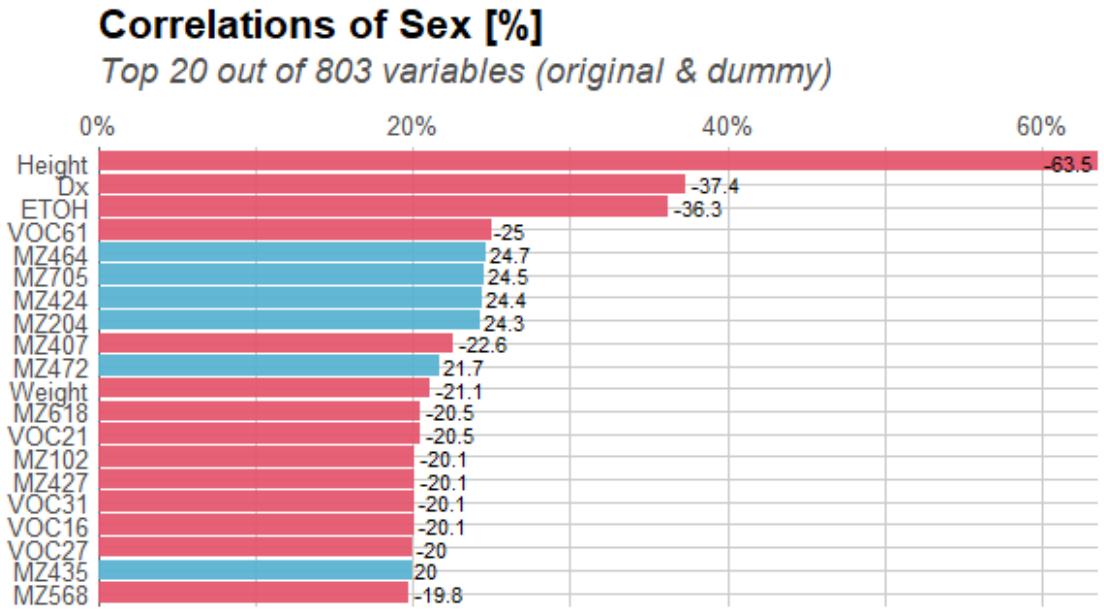

***Appendix V - Figure 10. Spearman correlation of height to other variables. Correlation coefficients expressed as a percentage for display purposes – see appendix for coefficient tables. Variables starting with MZ corresponds to a product mass to charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions.***




***Appendix V - Figure 11. Spearman correlation of weight to other variables. Correlation coefficients expressed as a percentage for display purposes – see appendix for coefficient tables. Variables starting with MZ corresponds to a product mass to charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions.***

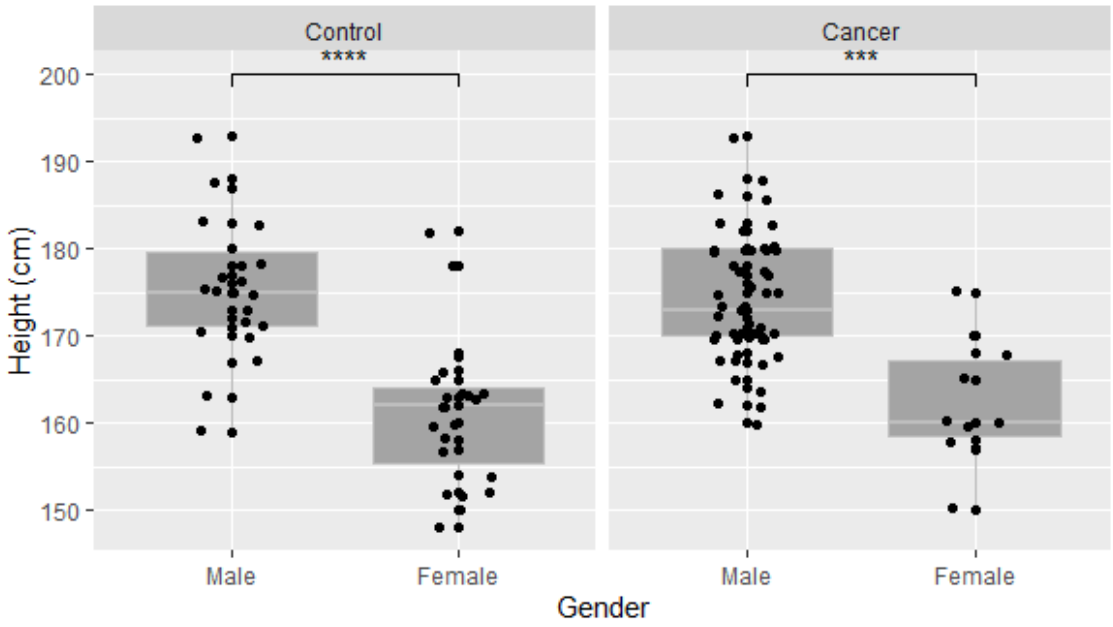
**Gender**

HNSCC is a male dominated disease, however, no significant correlations with male gender was found in this dataset (Appendix V - Figure 4, Appendix V - Figure 12) While the female gender had a clear association with height that was observed in both cancer and control groups (Appendix V - Figure 13). The patient height is a key predictor of total lung volume. Therefore, the maximal exhaled breath volume is dependent on patient height. Hence, VOCs described in literature that is thought to be predictive of disease or patient factors needs to be carefully exclude height as a confounding factor. Also, the significant difference in patient height observed based on their gender needs to be carefully considered when BMI is measured as an outcome in cancer studies where the patient cohorts are not gender matched.




***Appendix V - Figure 12. Spearman correlation of age to other variables. . Correlation coefficients expressed as percentage for display purposes – see appendix for coefficient tables. Dx, diagnosis (cancer vs control). ETOH, alcohol intake index. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions. Variables starting with MZ corresponds to a product mass to***

***charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions.***



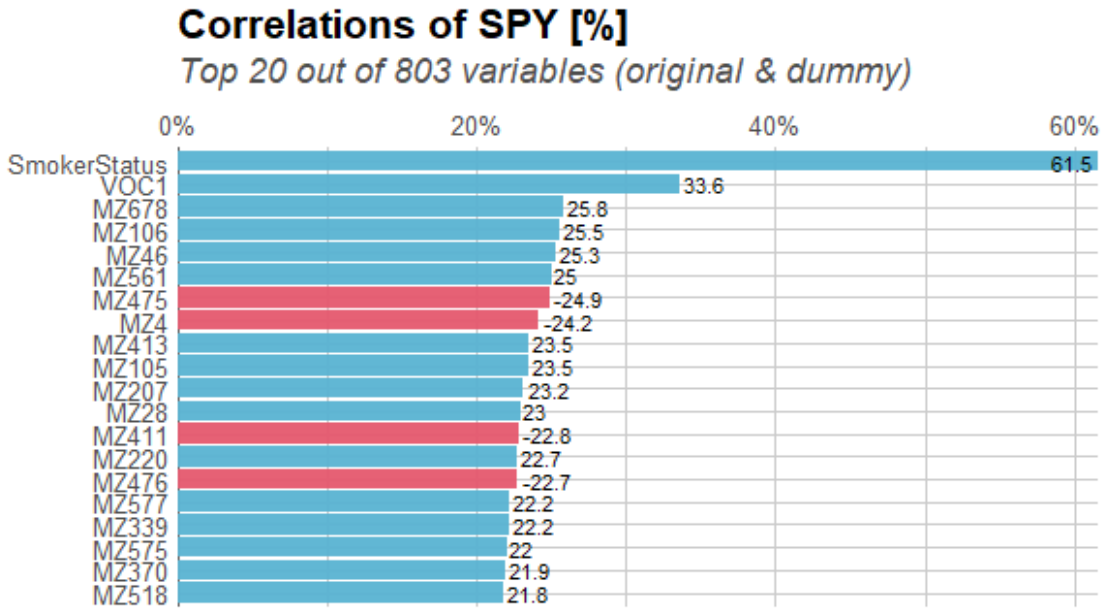

***Appendix V - Figure 13. Boxplot comparing patient height with their gender between cancer and control groups. Dots = individual patient data. ***, p<0.001. ****, p<0.0001.***

**Smoking**

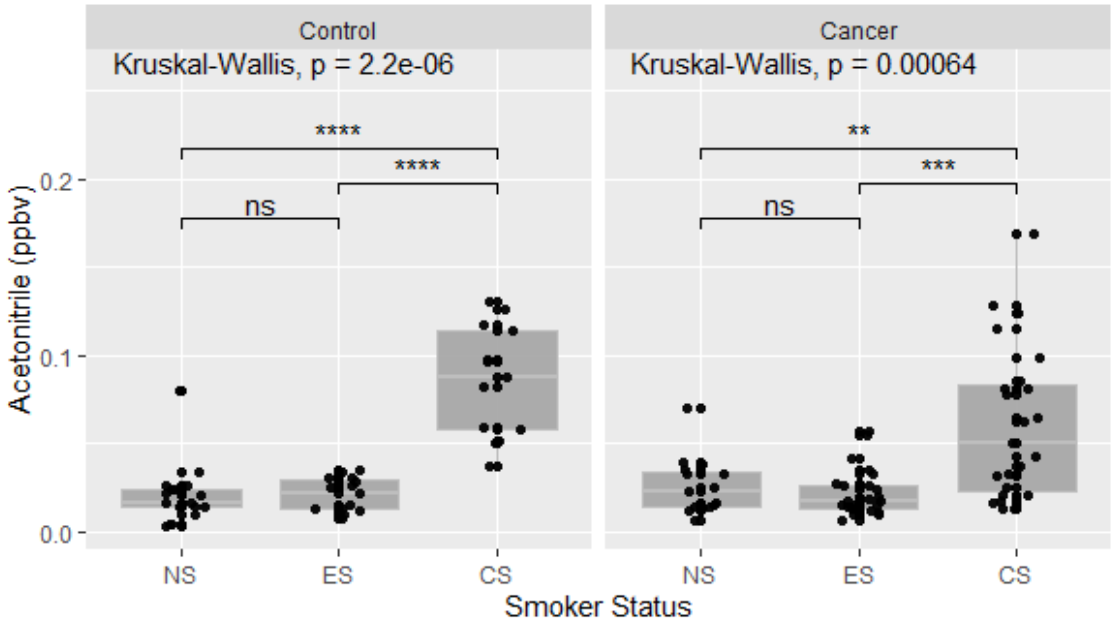
As expected, smoking pack years correlated well with smoking status of the patient (none-smokers, ex-smokers and current smokers) (Appendix V - Figure 14). VOC1 (acetonitrile) was found to have a weak correlation with smoking pack years. Acetonitrile has been previously associated with cigarette smoking and exposure to second-hand smoking. Further comparison of acetonitrile with smoking status (Appendix V - Figure 15), indicated a significant difference in the current smoking group compared to those who never smoked or ex-smokers. This pattern is the same between cancer and control groups (Appendix V - Figure 15). Interestingly, when gender was introduced as a variable, there was no difference in acetonitrile concentrations within female participants, with the only significant

difference between male smokers with and without cancer (Appendix V - Figure 16). This is potentially an artifact of maldistributed sample sizes, as majority of the cancer patients in this study were male and the majority of patients in the control group were female.

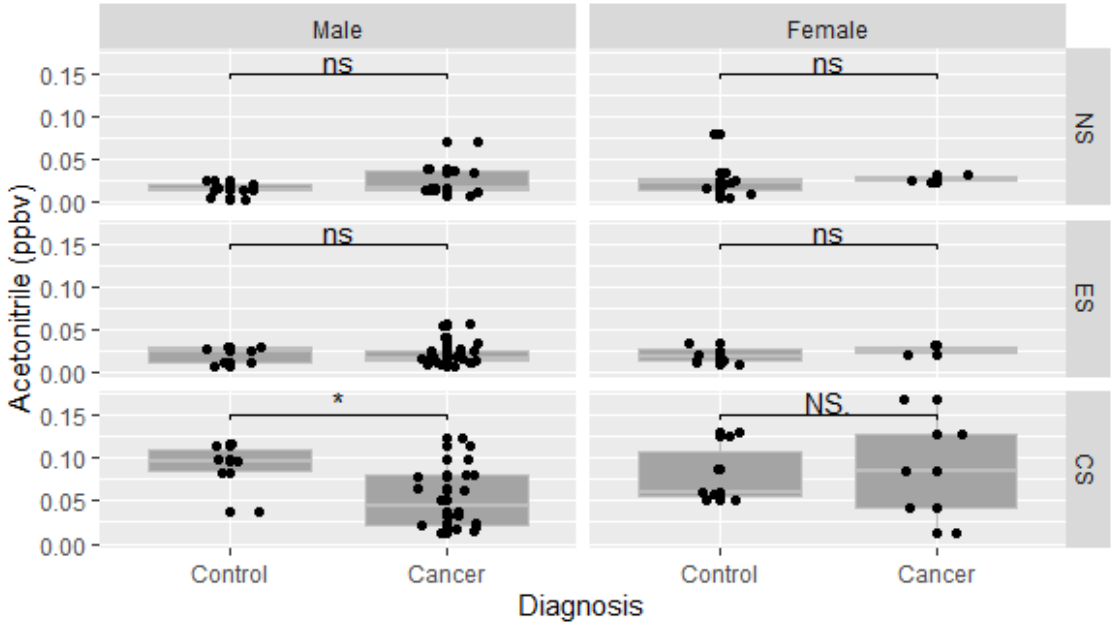
Overall, this highlights the need for even distribution of patients based on patient related fixed factors among comparison groups. In the absence of that, it is important to report on the fixed factor dependent variance of biological markers of interest as they may potentially identify patient factors instead of the disease of interest.



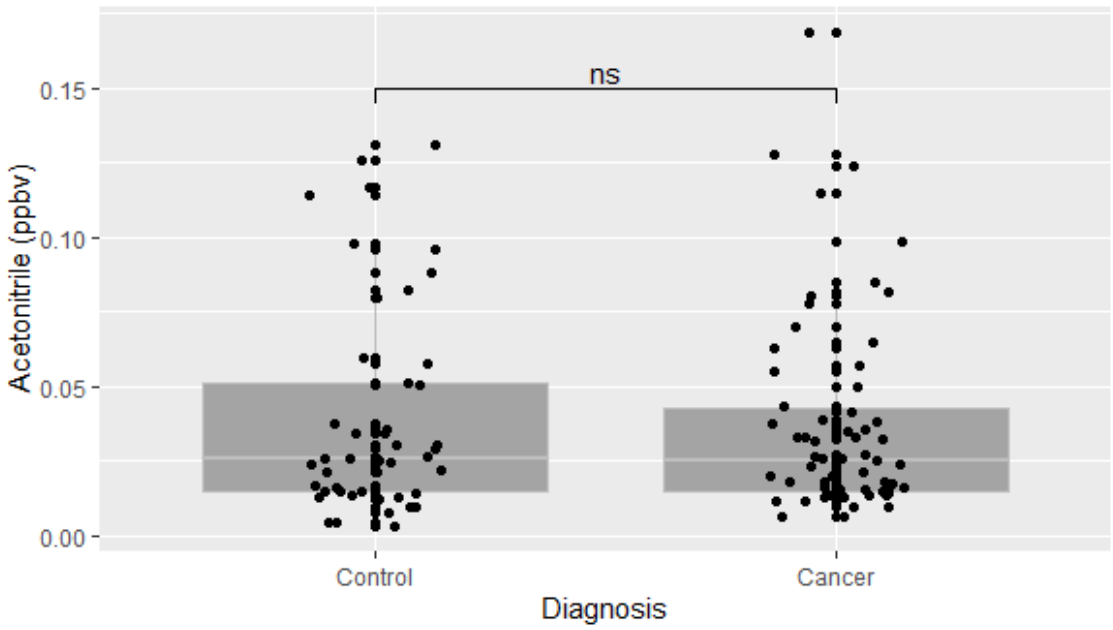

***Appendix V - Figure 14. Spearman correlation of smoking pack years to other variables. . Correlation coefficients expressed as percentage for display purposes – see appendix for coefficient tables. Dx, diagnosis (cancer vs control). ETOH, alcohol intake index. Variables starting with VOC corresponds to a specific volatile organic compound – see appendix for corresponding mass and reagent ions. Variables starting with MZ corresponds to a product mass to charge ratio for individual reagent ions – see appendix for corresponding mass and reagent ions.***



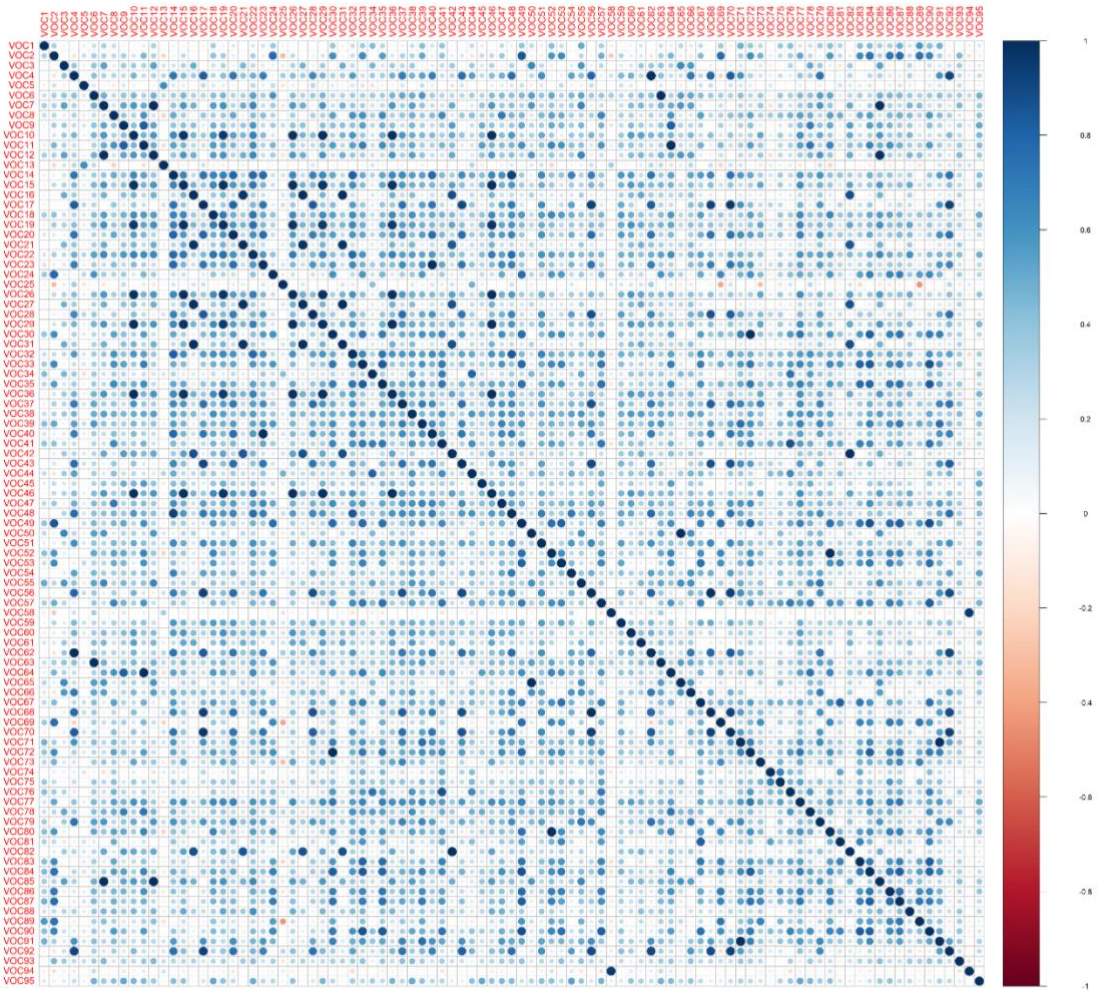

***Appendix V - Figure 15. Boxplots comparing acetonitrile (VOC1) with smoking status between cancer and control groups. Dots = individual patient data. NS, never smoked. ES, ex-smoker. CS, current smoker. ns, not significant *, p < 0.05, **, p <0.01. ***, p<0.001. ****, p<0.0001.***




***Appendix V - Figure 16. Boxplots comparing acetonitrile (VOC1) with smoking status, gender and diagnosis. Dots = individual patient data. NS, never smoked. ES, ex-smoker. CS, current smoker. ns, not significant *, p < 0.05.***



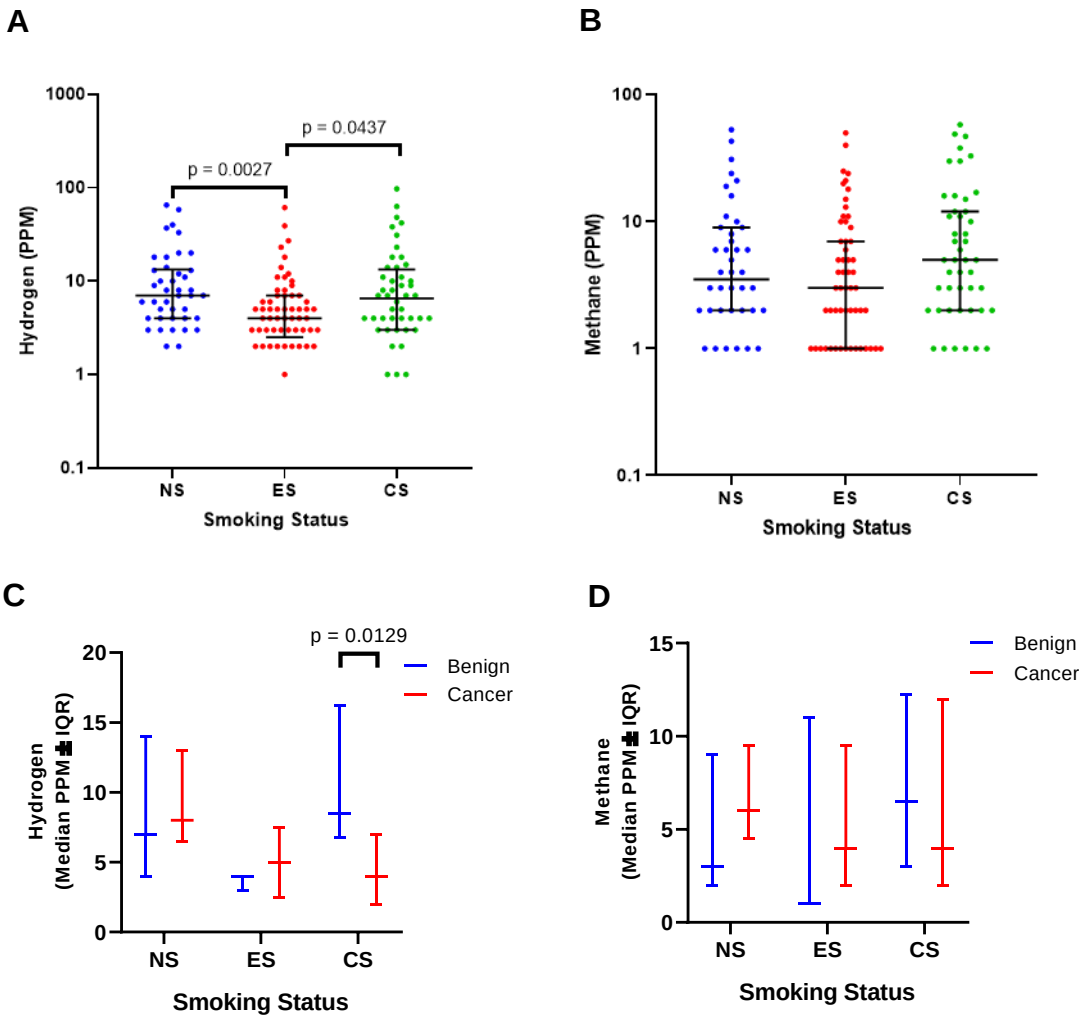

***Appendix V - Figure 17. Boxplot comparing acetonitrile (VOC1) between cancer and control patients. Dots = individual patient data. ns, not significant.***



***Appendix V - Figure 18. Spearman correlation between 95 VOCs previously described in cancer literature.***

**Distribution of hydrogen and methane based on HNSCC risk factors and other patient factors.**

Environmental factors such as tobacco smoking have been previously described as modifiers to the gut microbiome (Savin *et al*. 2018). Therefore, the effect of smoking on gastrointestinal gases exhaled in the breath were analysed (Appendix V - Figure 19). Smoking pack years (SPY) had no statistically significant relationship to levels of hydrogen or methane in breath. However, when smoking status is categorised to non-smokers (NS), ex-smokers (ES) and current smokers (CS), there is a significant difference (Kruskal-Wallis test, $p = 0.0079$) in breath hydrogen levels between the three groups (Appendix V - Figure 19A). Breath hydrogen levels were median 3ppm higher in NS ($p = 0.0027$) and CS ($p = 0.0437$) compared to ES. However, there is no significant difference in breath hydrogen between NS and CS. When these findings were further explored between cancer and benign patients, breath hydrogen is significantly lower in the cancer group who were CS compared to the respective benign control (Appendix V - Figure 19C; $p = 0.0129$). Alcohol intake had no clear effect on breath hydrogen or methane levels. Other patient factors such as their height, weight and fasting time did not have any correlation to breath hydrogen or methane levels in this cohort of patients.




***Appendix V - Figure 19. Relative differences in hydrogen (A) and methane (B) levels based on participant smoking status, smoking pack years and cancer status. Panel C and D comparing benign and cancer patients within each smoking category for hydrogen and methane, respectively. NS = non-smoker, ES = ex-smoker, CS = current smoker. P values are indicated where statistical significance is noted. PPM = parts per million, IQR = interquartile range***

In summary, patient fixed factors and disease dependent risk factors have the potential to confound biomarker discovery results. Therefore, considered recruitment approach to represent the diversity observed in the population of interest and thorough statistical exploration of these factors in comparison to the marker of interest is a must.

## Appendix VI. Participant information sheet for recruitment

**Participant Information Sheet/Consent Form**

**Non-Interventional Study** - *Adult providing own consent.*

**Title:**
***Breathing New Life into Cancer Detection: A project looking at new ways to detect cancer on the breath.***

**Principal Investigator:**
**Dr Roger Yazbek**
**Associate Investigator(s):**
**Professor David Watson**
**Associate Professor Eng Hooi Ooi**
**Dr Nuwan Dharmawardana**

**Locations:**
**Flinders Medical Centre**
**The Royal Adelaide Hospital**

## Part 1 What does my participation involve?

### 1. Introduction

You are invited to take part in this research project, *Breathing New Life into Cancer Detection: A project looking at new ways to detect cancer on the breath*. You have been asked to participate in this study because you are being investigated for either: Oesophageal (gullet) cancer, gastric (stomach) cancer, or a head and neck cancer. The research project is aiming to see whether gasses being produced by abnormal cells, such as those found in cancer, can be detected in the breath.

Please note that you are participating in a research study and acknowledge that participation does not exclude any future diagnosis of cancer unrelated to the current study.

This Participant Information Sheet/Consent Form tells you about the research project. It explains the tests and research involved. Knowing what is involved will help you decide if you want to take part in the research.

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not to take part, you might want to talk about it with a relative, friend or local doctor.

Participation in this research is voluntary. If you don't wish to take part, you don't have to. You will receive the best possible care whether or not you take part.

If you decide you want to take part in the research project, you will be asked to sign the consent section. By signing it you are telling us that you:

- Understand what you have read.
- Consent to take part in the research project.
- Consent to the tests and research that are described.
- Consent to the use of your personal and health information as described.

You will be given a copy of this Participant Information and Consent Form to keep.

## 2. What is the purpose of this research?

We are particularly interested in trying to find new ways to find oesophageal (gullet), gastric (stomach), or head and neck cancer at their earliest stage. To do this we plan to compare the breath from individuals with cancer, with individuals with Barrett's oesophagus and also with individuals who are not known to have

either of these problems. Barrett's oesophagus is a condition where the cells that line the lower gullet are abnormal.

You may have seen news stories about dogs that can ‘smell’ cancer. What the dogs actually smell, using their highly sensitive noses, are gases that are being produced by the cancer cells. We are aiming to identify these gases in the breath using highly specialised equipment. If we can identify unique ‘cancer gases’, then we hope to use this information to develop a new, non-invasive breath test for earlier detection of these cancers.

Breath testing is an experimental diagnostic or screening method. This means that it is not an approved method of screening or detecting cancer in Australia. The results of this research will be used by the study doctor, Dr Nuwan Dharmawardana to obtain a Doctor of Philosophy degree.

This research has been initiated by Professor David Watson.

This research has been funded by the Katherine Marie Enright Kelly Research fund, Garnett Passe Rodney Williams Memorial Foundation and Australia and New Zealand Head and Neck Cancer Society Foundation.

**3. What does participation in this research involve?**

The aim of the study is to see whether gasses being produced by abnormal cells, such as those found in cancer, can be detected in the breath. We are asking you to initially donate two breath samples. Should you also return to the hospital or

clinic during the next 3-18 months, we would like to ask that you donate two further breath samples at each visit.

If you agree to participate in this study, you will be first be interviewed by a member of the research team, to determine any medications or lifestyle factors that may be relevant to the study. This should only take 3-5 minutes. After this interview we will ask you to breathe into two 3 Litre bags until the bag is full. This will be a bit like blowing up a balloon. We will only collect 1-2 breath samples at a time, per patient, at two time points (pre-treatment and then 3 - 18 months after treatment).

All information collected in this study will be de-identified to protect your identity, and no one who is not directly involved in this study will be able to identify you, or link information to you.

This research project has been designed to make sure the researchers interpret the results in a fair and appropriate way and avoids study doctors or participants jumping to conclusions.

There are no costs associated with participating in this research project, nor will you be paid. You will not be reimbursed for any reasonable travel, parking, meals and other expenses associated with the research project visit.

**4. What do I have to do?**

Please be informed that participation in this study does not require:

- Lifestyle restrictions e.g. physical restrictions, participation in sport

- Dietary restrictions

Participation will not restrict you from:

- Taking regular medications
- Donating blood

**5. Do I have to take part in this research project?**

Participation in any research project is voluntary. If you do not wish to take part, you do not have to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage.

Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your routine treatment, your relationship with those treating you or your relationship with Flinders Medical Centre.

**6. What are the alternatives to participation?**

You do not have to take part in this research project to receive treatment at this hospital.

**7. What are the possible benefits of taking part?**

There are unlikely to be any direct benefits to you from participating in this study, but your help may assist with the development of new ways to find cancer earlier. The study results will not influence your medical treatment and participating will have no impact on your medical care.

## 8. What are the possible risks and disadvantages of taking part?

There is no risk or discomfort associated with breath collection. We simply require you to breath out into a plastic bag. This is similar to blowing up a balloon.

## 9. What will happen to my test samples?

Breath samples will be analysed in a laboratory at Flinders Medical Centre to determine which gases are being produced in the breath, and which relate specifically to cancer. We would like to collect as many samples as possible to generate a data bank of analysed breath samples for ongoing research.

Any excess samples will be destroyed following the completion of the study. The information gained from the study will be kept for 15 years and then destroyed as per guidelines for human research.

## 10. What if new information arises during this research project?

Sometimes during the course of a research project, new information becomes available about the treatment that is being studied. If this happens, your study doctor will tell you about it and discuss with you whether you want to continue in the research project. If you decide to withdraw, your study doctor will make arrangements for your regular health care to continue. If you decide to continue in the research project you will be asked to sign an updated consent form.

Also, on receiving new information, your study doctor might consider it to be in your best interests to withdraw you from the research project. If this happens, he/ she will explain the reasons and arrange for your regular health care to continue.

## 11. Could this research project be stopped unexpectedly?

It is unlikely that this project may stop unexpectedly. However, it can happen for a variety of reasons such as sudden illness affecting all researchers or malfunction of machine used to analyse breath samples.

## 12. What happens when the research project ends?

Your current treatment will not be affected when this project ends. Please notify a member of the research team if you would like to be contacted about results and the success of the project.

## 13. Can I have other treatments during this research project?

Participation in this study will not interfere with current or future treatments.

## 14. What if I withdraw from this research project?

If you decide to withdraw from the project, please notify a member of the research team before you withdraw.

If you do withdraw your consent during the research project, the study doctor and relevant study staff will not collect additional personal information from you, although personal information already collected will be retained to ensure that the results of the research project can be measured properly and to comply with law. You should be aware that data collected by the sponsor up to the time you

withdraw will form part of the research project results. If you do not want them to do this, you must tell them before you join the research project.

**Part 2 How is the research project being conducted?**

**15. What will happen to information about me?**

By signing the consent form, you consent to the study doctor and relevant research staff collecting and using personal information about you for the research project. Any information obtained in connection with this research project that can identify you will remain confidential. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

Breath samples will be analysed immediately after collection, and data will be electronically stored on an electronic database on a password-protected computer. This database will only be accessible to researchers involved in this study. Any paper records will be stored at Flinders Medical Centre in a locked cabinet when not in use.

Data will be retained and stored for 15 years. This is required by National Research Practice Guidelines.

Part of the breath sample will be stored on a specialised “column” i.e. a container, for future analysis by gas analysis chromatography. These “columns” will be stored in a locked laboratory and will not have your name on them. They will be identified by a number. Samples and data will not be identifiable to persons not

directly involved in the study. All records containing any personal information will remain confidential and no information that could lead to your identification will be released unless required by law. It will be disclosed only with your permission, or as required by law.

It is anticipated that the results of this research project will be published and/or presented in a variety of forums. In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except with your permission.

In accordance with relevant Australian privacy and other relevant laws, you have the right to request access to your information collected and stored by the research team. You also have the right to request that any information with which you disagree be corrected. Please contact the study team member named at the end of this document if you would like to access your information.

**16. Complaints and compensation**

If you suffer any injuries or complications as a result of this research project, you should contact the study team as soon as possible and you will be assisted with arranging appropriate medical treatment. If you are eligible for Medicare, you can receive any medical treatment required to treat the injury or complication, free of charge, as a public patient in any Australian public hospital.

**17. Who is organising and funding the research?**

This research project is being conducted by the Department of Surgery at Flinders University and there is no commercial sponsorship.

The results of this research will be used by Dr Nuwan Dharmawardana to obtain a Doctor of Philosophy degree.

Flinders University may benefit financially from this research project if, for example, the project assists approval of a new drug/ device.

By taking part in this research project you agree that samples of your breath (or data generated from analysis of these materials) may be provided to Flinders University. Flinders University may directly or indirectly benefit financially from your samples or from knowledge acquired through analysis of your samples.

You will not benefit financially from your involvement in this research project even if, for example, your samples (or knowledge acquired from analysis of your samples) prove to be of commercial value.

In addition, if knowledge acquired through this research leads to discoveries that are of commercial value to the university, the study doctors or their institutions, there will be no financial benefit to you or your family from these discoveries.

No member of the research team will receive a personal financial benefit from your involvement in this research project (other than their ordinary wages).

**18. Who has reviewed the research project?**

All research in Australia involving humans is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this research project have been approved by the HREC of Southern Adelaide Clinical.

This project will be carried out according to the National Statement on Ethical Conduct in Human Research (2007). This statement has been developed to protect the interests of people who agree to participate in human research studies.

## 19. Further information and who to contact.

The person you may need to contact will depend on the nature of your query.

If you want any further information concerning this project or if you have any medical problems that may be related to your involvement in the project (for example, any side effects), you can contact Dr Nuwan Dharmawardana on 0425822378 or any of the following people:

### Clinical contact person

| | |
|---|---|
| Name | Dr Roger Yazbek |
| Position | Coordinating Principal Investigator |
| Telephone | 0438 650 722 |
| Email | roger.yazbek@flinders.edu.au |

For matters relating to research at the site at which you are participating, the details of the local site complaints person are:

### Complaints contact person.

| | |
|---|---|
| Name | *Villis Marshall* |
| Position | *Director, Office for Research* |
| Telephone | *8204 6061* |
| Email | *Health:SALHNofficeforresearch@sa.gov.au* |

If you have any complaints about any aspect of the project, the way it is being conducted or any questions about being a research participant in general, then you may contact:

**Reviewing HREC approving this research and HREC Executive Officer**

| | |
|---|---|
| Reviewing HREC name | *Southern Adelaide Clinical* |
| HREC Executive Officer | *Damian Creaser* |
| Telephone | *8204 6453* |
| Email | *Health.SALHNofficeforresearch@sa.gov.au* |

**details**

**Local HREC Office contact (Single Site -Research Governance Officer)**

| | |
|---|---|
| Position | *Research Governance Officer* |
| Telephone | *8204 6139* |
| Email | *Health.SALHNofficeforresearch@sa.gov.au* |

## Appendix VII. Participant consent and withdrawal forms

# **Consent Form -** *Adult providing own consent.*

**Title:**
***Breathing New Life into Cancer Detection: A project looking at new ways to detect cancer on the breath.***

**Principal Investigator:**
**Dr Roger Yazbek**
**Associate Investigator(s):**
**Professor David Watson**
**Associate Professor Eng Hooi Ooi**
**Dr Nuwan Dharmawardana**

**Locations:**
**Flinders Medical Centre**
**The Royal Adelaide Hospital**

---

Declaration by Participant

I have read the Participant Information Sheet, or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research described in the project.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time during the project without affecting my future health care.

I understand that I will be given a signed copy of this document to keep.

Name of Participant
(please print) _____

Signature _____ Date _____

Name of Witness* to
Participant's Signature
(please print)

_____

Signature _____ Date _____

* Witness is not to be the investigator, a member of the study team or their delegate. In the event that an interpreter is used, the interpreter may not act as a witness to the consent process. Witness must be 18 years or older.

## Declaration by Study Doctor/Senior Researcher†

I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant has understood that explanation.

Name of Study Doctor/
Senior Researcher†
(please print)

_____

Signature _____ Date _____

† A senior member of the research team must provide the explanation of, and information concerning, the research project.

# Form for Withdrawal of Participation - *Adult providing own consent.*

**Title:**
***Breathing New Life into Cancer Detection: A project looking at new ways to detect cancer on the breath.***

**Principal Investigator:**
**Dr Roger Yazbek**
**Associate Investigator(s):**
**Professor David Watson**
**Associate Professor Eng Hooi Ooi**
**Dr Nuwan Dharmawardana**

**Locations:**
**Flinders Medical Centre**
**The Royal Adelaide Hospital**

---

Declaration by Participant

I wish to withdraw from participation in the above research project and understand that such withdrawal will not affect my routine treatment, my relationship with those treating me or my relationship with *Flinders Medical Centre or The Royal Adelaide Hospital.*

Name of Participant
(please print) _____

Signature _____ Date _____

In the event that the participant's decision to withdraw is communicated verbally, the Study Doctor/Senior Researcher will need to provide a description of the circumstances below.

**Declaration by Study Doctor/Senior Researcher†**

I have given a verbal explanation of the implications of withdrawal from the research project, and I believe that the participant has understood that explanation.

| | |
|---|---|
| Name of Study Doctor/ Senior Researcher† (please print) | |

## Appendix VIII. Patient demographic and risk factor questionnaire

**Patient questionnaire sheet**

**Breathing New Life into Cancer Detection: A project looking at new ways to detect cancer on the breath.**

Name:_____

Age: _____ MRN: _____

Gender: _____ Occupation: _____

Please answer the following questions concerning your past medical history and present health status.

GENERAL MEDICAL BACKGROUND

1. Have you ever been diagnosed with any of the following disease? Please circle all that apply.

   None

   Diabetes, Stroke, Hypertension, Myocardial infarction, Angina pectoris, Asthma, Chronic, Bronchitis, Allergy, Kidney disease, Cystitis, Hepatitis, Liver Cirrhosis, Peptic, Gastric/duodenal ulcer, Biliary stone, Chronic reflux
   Others (_____)

2. Have your parents/immediate family had any type of cancer?

   _____

   _____

_____

_____

3. What is your height and weight?

Height: _____cm Weight: _____kg

SMOKING AND DRINKING QUESTIONS

Please answer the following questions concerning smoking and drinking.

4. Have you ever been a smoker?

No Yes

If yes, how old were you when you began smoking? _____years old

Do you currently smoke cigarettes?

No If no, how old were you when you quit smoking?

_____years old

Yes If yes, how many cigarettes a day do you smoke?

_____cigarettes/ day

5. Have you ever lived with a smoker for more than 10 years?

No Yes

6. How often are you exposed to passive smoking (more than 1 hour/day) outside home, such as work?

0. Almost Never
1. 1-3 days a month
2. 1-4 days a week

3. Almost every day

7. Have you ever consumed alcohol?

No Yes

If yes, how often do you consume alcoholic beverages?

0. Almost never
1. 1-3 days a month
2. 1-2 days a week
3. 3-4 days a week
4. 5-6 days a week
5. Everyday

DIETARY AND LIFESTYLE QUESTIONS

8. What are foods you have consumed in the last 24 hours?

_____

_____

9. What was your last meal prior to providing your breath sample?

_____

_____

10. Did you brush your teeth or use a mouthwash (e.g. Listerine) before today's breath test?

_____

_____

11. Have you chewed gum or had a breath mint in the last 3 hours?

_____

_____

Thank for your time and patience.