CHAPTER 1

INTRODUCTION

Oral cancer is the sixth most common cancer in the world and it is estimated that up to 80% of these cancers occur in Asia. Despite the advances in treatment, the 5-year survival rates have not changed significantly over the past decades (Funk *et al.*, 2002; Stahl *et al.*, 2004). Clinical examination, biopsy and imaging for oral cancer diagnosis have shown little improvement in sensitivity and specificity, therefore a better understanding of oral carcinogenesis is needed to improve diagnosis, treatment and monitoring of the disease (Todd and Wong, 2002). With the advent of high-throughput microarray, we can now study the gene expression of thousands of genes simultaneously. Previous studies have demonstrated the use of microarrays either to sub-classify cancer, to compare the genetic changes at different stages of diseases or to identify genes that can be used as prognosticators (Golub *et al.*, 1999; van 't Veer *et al.*, 2002). The use of microarray is useful particularly in cases where clinical information alone may not be sufficient for diagnosis. This is important when applied to the clinical setting, as accurate diagnosis of the cancer will ensure that patients are given the most appropriate treatment. An understanding of the genetic alterations underlying cancer can result in the identification of possible therapeutic targets to increase the use of molecular targeted therapy in the clinics. To date, microarray experiments have mostly utilized fresh frozen samples. However, a new microarray platform from Ilumina (DASL) has developed an assay for microarray experiments using formalin-fixed paraffin embedded (FFPE) tissues thus enabling retrospective studies to be performed. FFPE samples have two key advantages: firstly, they are collected routinely in clinical practice and therefore readily available and secondly, clinical follow-up data would be

available for these archived specimens. Microarrays have previously been used to compare the genetics between normal and cancerous tissues with the aim of determining biomarkers that will contribute to diagnosis and therapeutic strategies. However, most published microarray papers have identified candidate genes based on a mixture of tumours from different sites of the oral cavity including a variety of head and neck tissues in the same experiments (Mendez *et al.*, 2002; Ginos *et al.*, 2004). This in part explains the dissimilarity in the genes that were identified from different experiments on oral cancers. Indeed, Warner *et al.*, 2004 demonstrated that the genetic profile of cancer cell lines clustered them based on the sites from which the cell lines were derived, therefore, suggesting the possibility that the genetics of oral cancers from different sites may be distinct from one another (Warner *et al.*, 2004). The identification of site-specific gene expression signatures has important implications. For example, a molecular target identified in the west where OSCC are commonly seen in the floor of the mouth and tongue, will not be directly applicable to patients in Malaysia whose main site of the disease is at the cheek mucosa. Despite the increasing numbers of biomarkers identified, few have been successfully translated into clinically relevant markers in part because of the heterogeneity of specimens used in the studies. For that reason, this study aims to establish the similarities and differences between OSCC from the different sites, and to investigate the mechanism of specific genes in driving OSCC.

CHAPTER 2

PURPOSE OF STUDY

2.1. Aim of Study

This study aims to identify the gene expression patterns of OSCC from different sites of the oral cavity using formalin-fixed paraffin-embedded (FFPE) samples, and to determine the biological significance of site-specific genetic alterations.

2.2. Specific Objectives

2.2.1. To determine the quantity and quality of RNA extracted from FFPE specimens for the use in microarray experiments.

2.2.2. To identify and validate differentially expressed pathways and genes implicated in OSCC.

2.2.3. To analyse gene expression variation in OSCC from cheek, gum and tongue.

2.2.4. To identify and validate differentially expressed genes and pathways implicated in OSCC from the three distinct sites (cheek, gum and tongue).

2.2.5. To determine cancer characteristics conferred by putative oral cancer gene using genetically engineered oral cancer cell lines.

CHAPTER 3

LITERATURE REVIEW

3.1. Oral Cancer

Oral cancer is a malignancy that arises from the oral cavity and can arise from tongue, floor of mouth, gum, cheek mucosa, palate and lip. The World Health Organization (WHO) classifies oral cancer by the International Classification of Disease, Version 10 (ICD 10) with the codes C00-C06 where C00 encodes the lip excluding the skin of the lip, C01-C02 encodes the tongue, C03 encodes the gum, C04 encodes floor of the mouth, C05 encodes palate and, C06 encodes other unspecified parts of the mouth which includes the cheek mucosa (http://www.who.int/classifications/icd/en/). Oral cancer is among the top 10 most common cancers in the world and it is estimated that up to 60% of these cancers occur in Asia (Jemal *et al.*, 2011). Despite the advances in cancer treatment, the 5-year survival rates for oral cancer have not changed in the past few decades (Funk *et al.*, 2002; Stahl *et al.*, 2004). The overall mortality rate remains high at approximately 50% and this is consistent with the advanced stage of disease presentation (McMahon and Chen, 2003; Walker *et al.*, 2003). More than 90% of oral cancers are squamous cell in origin and hence these cancers are often referred to as oral squamous cell carcinoma (OSCC) (Cawson *et al.*, 1998; Walker *et al.*, 2003).

3.2. Epidemiology

3.2.1. World Wide

GLOBOCAN is a programme of the International Agency for Research in Cancer (IARC) that estimates cancer incidence and mortality based on the most recently

available data collected at the IARC or available in routine reports from local registries. Based on GLOBOCAN 2008, cancer of the oral cavity is the 15th most common cancer worldwide and it was further ranked 10th in men and 13th in women.

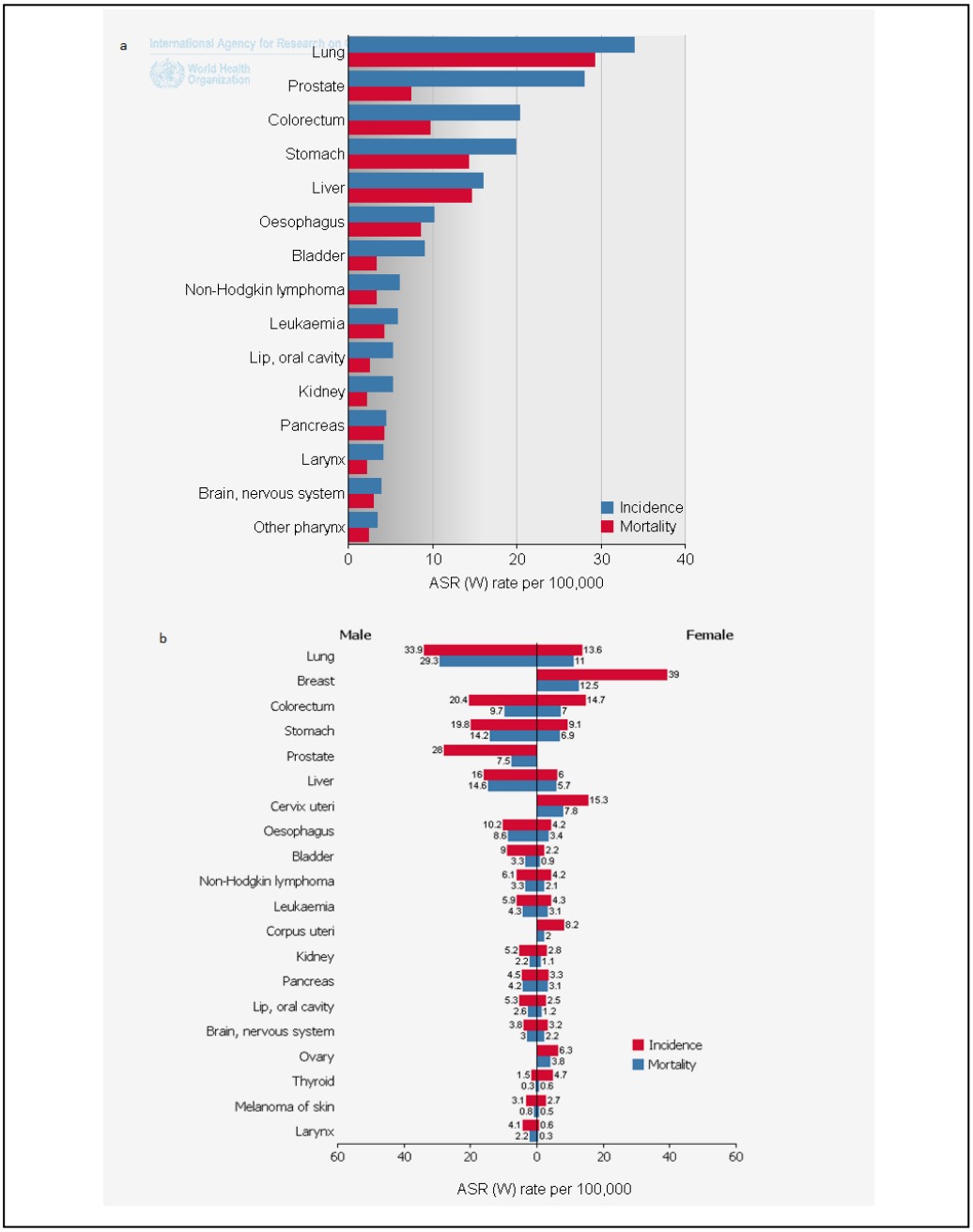


Figure 3.1: Cancer incidence worldwide based on ASR (W) rate per 100,000 estimated by GLOBOCAN 2008 in (a) both gender and (b) among males and females respectively (data obtained from GLOBOCAN.iarc.fr).

A total of 263,000 new cases of oral cancer were reported in 2008 of which 64.8% were female. Furthermore 65.4% of these cases were from the less developed countries which include China, Asia (Outer Eastern, South Eastern, South Central, and Western), Melanesia and Micronesia/Polynesia. Notably, a total of 128,000 deaths were estimated in 2008 of which almost 80% were from the less developed countries (Table 3.1).

Table 3.1: Incidence and mortality in oral cancer including the lips in 2008 (adapted from globocan.iarc.fr).

Incidence	Overall	%	Male	%	Female	%
World	263,055		92,524	35.2	170,496	64.8
More developed	91,217	34.7	62,757	68.9	28,391	31.1
Less developed	171,935	65.4	107,739	62.7	64,133	37.3
Mortality	Overall	%	Male	%	Female	%
World	127,719		83,109	65.1	44,545	34.9
More developed	30,760	24.0	21,878	71.3	8,811	28.6
Less developed	97,028	76.0	61,231	63.1	35,734	36.8

Overall, the highest number of oral cavity cancers was estimated in Melanesia, South Central Asia and Oceania with ASR of 17.8, 7.4 and 7.1 per 100,000 populations respectively. Notably, Northern America and Europe except Northern Europe have relatively high ASR ranging from 4.6-4.9 per 100,000 populations. The lowest incidence of oral cancer was reported in Africa, Central America and Eastern Asia with ASR of 2.2, 1.9 and 1.5 per 100,000 populations respectively (Table 3.2).

Table 3.2 Estimated Age-Standardized Incidence Rates of Oral Cancer Per 100,000 by World Area based on GLOBOCAN 2008 (http://globocan.iarc.fr/).

	ASR per 100,000		
	Overall	Male	Female
World	3.8	5.3	2.5
More developed regions	4.4	6.8	2.3
Less developed regions	3.6	4.6	2.6
Africa	2.5	3.0	2.0
Sub-Saharan Africa	2.7	3.3	2.1
Eastern Africa	3.2	4.2	2.4
Middle Africa	1.9	2.2	1.6
Northern Africa	1.9	2.3	1.6
Southern Africa	2.8	4.5	1.5
Western Africa	2.3	2.5	2.1
Latin America and Caribbean	3.2	4.6	1.9
Caribbean	3.7	5.2	2.4
Central America	2.2	2.7	1.7
South America	3.4	5.2	1.9
Northern America	4.9	7.1	2.9
Asia	3.7	4.7	2.7
Eastern Asia	1.5	2.1	0.9
South-Eastern Asia	3.0	3.4	2.7
South-Central Asia	7.4	9.4	5.5
Western Asia	2.2	2.9	1.6
Europe	4.6	7.4	2.2
European Union (EU-27)	4.6	7.0	2.4
Central and Eastern Europe	4.8	9.0	1.8
Northern Europe	3.8	5.1	2.5
Southern Europe	4.8	7.5	2.3
Western Europe	4.6	6.6	2.8
Oceania	7.1	9.5	4.8
Melanesia	17.8	24.0	12.0
Micronesia/Polynesia	2.5	4.0	1.0
Micronesia	1.8	2.4	1.3
Polynesia	3.2	5.3	0.9

3.2.2. Malaysia

Based on the GLOBOCAN data, oral cancer is ranked as the 13th most common cancers in Malaysia overall, with the ASR of 3.5 per 100,000 population. When gender was taken into account, oral cancer ranked 10th among males and 12th among females. As Malaysia is a multi-ethnic country consisting of 3 main ethnic groups: Malay (51.0%), Chinese (24.2%) and Indian (7.1%) (Statistics Department, 2001), it is important to examine the incidence of oral cancer in the different ethnic groups. Notably, the Indian female community in Malaysia is disproportionately affected. Based on the Malaysia National Cancer Registry Report which records cancer cases from government hospitals in Malaysia with the exception of Sabah and Sarawak, the ASR of oral cancer excluding tongue cancer for Indian females was exceptionally high with the ASR value of 14.4 per 100,000 population compared to Malay and Chinese with only 0.8 and 0.6 per 100,000 respectively (Lim *et al.*, 2008). In fact, this value is even higher than the value estimated by GLOBOCAN 2008 for oral cancer incidence in Melanesia which has the highest ASR value among females (Table 3.2) thus indicating the immense burden of oral cancer amongst the female Indian ethic group in Malaysia.

3.3. Risk Factors for Oral Cancer

Smoking, alcohol consumption, betel quid use and HPV infection are the major risk factors for oral cancer with smoking and alcohol reported to have synergistic effects (Blot *et al.*, 1988; Andre *et al.*, 1995). However the contribution of each risk factor to the burden of oral cancer varies across geographical regions (Jemal *et al.*, 2011). Furthermore, other factors such as diet and nutrition, occupational risk, poor oral health

hygiene, immune disturbances, and hereditary influences were also reported to be involved in oral cancer development (Clayman, 1997; Stewart *et al.*, 2003).

3.3.1. Smoking

The strong association between cancers of the oral cavity with smoking is well established. Epidemiological studies have shown that the risk of developing oral cancer is five to nine times greater in smokers compared to nonsmokers, and this risk may increase to as much as 17 times greater for extremely heavy smokers of 80 or more cigarettes per day (Blot *et al.*, 1988; Jovanovic *et al.*, 1993; Mashberg *et al.*, 1993; Andre *et al.*, 1995; Lewin *et al.*, 1998). Apart from smoking, habits such as tobacco chewing is reported to be associated with an increased risk to oral cancer (IARC, 1986). It has been reported that tobacco smoke contains in excess of 300 carcinogens and pro-carcinogens which will contaminate the saliva and induce DNA adducts leading to cancer (IARC, 1986). Furthermore, certain pro-carcinogens such as such as polycyclic aromatic hydrocarbons (e.g. benzo(a)pyrene), tobacco specific nitrosamine (e.g. 4-(methylnitrosamine) and aromatic amines (e.g. 4-aminobiphenyl) require metabolic activation through xenobiotic enzymes in particular the cytochrome p450 before exerting its effect (Hecht, 1999). Due to the fact that nearly all carcinogens and pro-carcinogens requires activation by xenobiotic enzymes and detoxifying enzymes to deactivate carcinogens and their intermediate by-products, there have been extensive studies linking genetic polymorphism of these xenobiotic enzymes and its ability to modify individual's response to such carcinogens (Ho *et al.*, 2007) .

3.3.2. Alcohol Consumption

The consumption of any type of alcoholic beverages are associated with an increased risk to oral cancer. Alcohol consumption has been shown to have a role in oral cancer

independent of tobacco use (IARC, 1989). In studies controlled for smoking, moderate-to-heavy drinkers have been shown to have three to nine times greater risk of developing oral cancer and among extremely heavy drinkers (greater than 100 gm of alcohol per day) the risk increased to 30 times (Blot *et al.*, 1988; Jovanovic *et al.*, 1993; Mashberg *et al.*, 1993; Andre *et al.*, 1995; Lewin *et al.*, 1998). Notably alcohol abuse also potentiates the effect of other carcinogens, particularly tobacco where the risk to oral cancer increases up to a 100 times in heavy smokers and heavy drinkers (Blot *et al.*, 1988; Andre *et al.*, 1995). Alcohol may have local effects whereby it directly acts on cell membranes and alters the mucosal permeability that would contribute to increase penetration of carcinogens across the oral mucosa (Walker *et al.*, 2003; Ogden, 2005). Moreover, during alcohol metabolism, acetaldehyde which is a cytotoxic compound is produced which lead to the production of free radicals and hydroxylation of DNA bases which further causes cellular DNA damage (Scully *et al.*, 2000).

3.3.3. Betel Quid Use

A consensus workshop held in 1996 recommended that the term 'quid' should be defined as 'a substance or mixture of substances, placed in the mouth, usually containing at least one of the two basic ingredients, tobacco or areca nut, in raw or any manufactured or processed form' (Zain *et al.*, 1999). There are many different composition of chewing substances and in many countries, ready-made, mass-produced packets are available as proprietary mixtures known as pan masala or gutka (Table 3.3) (IARC, 2004). In Malaysia, the main quid ingredients are areca nut (taken either fresh or dried), betel leaf and slaked lime, sometimes folded in betel leaves like little parcels and chewed. However, the quid from different ethnic groups could have its own quid

mixtures with additional ingredients such as tobacco and spices and may practise different chewing methods (Zain RB, 1999).

Table 3.3 Composition of different chewing substance (IARC, 2004).

	Areca nut[a]	Betel[b]			Catechu[d]	Tobacco[e]	Slaked lime
		Leaf	Inflo-rescence	Stem[c]			
Areca	X						X
Betel quid without tobacco	X	X			(X)[f]		X
Betel quid with tobacco	X	X			(X)[f]	X	X
Gutka	X				X	X	X
Pan masala	X				X		X
Khaini						X	X
Mawa	X					X	X
Mainpuri tobacco	X					X	X
Loa-hwa (Taiwan)	X[g]		X				X
Betel quid (Taiwan)	X[g]	X					X
Stem quid (Taiwan)	X[g]			X			X
Naswar						X	X
Zarda						X	X

[a] May be used unripe, raw or processed by baking, roasting or baking with sweetening, flavouring and decorative agents
[b] In place of the leaf, the inflorescence or its stem may also be used
[c] Stem of inflorescence
[d] In powdered or paste form
[e] In flaked, powdered or paste form, with or without processing, with or without sweetening
[f] ( ) means optional

Strong evidence has associated chronic betel quid chewing with oral cancer (Henderson and Aiken, 1979; Daftary , 1991). Betel quid chewing produces reactive oxygen species (ROS) which has many harmful effects on the oral mucosa. ROS is directly involved in tumour initiation process by inducing gene mutation or causing structural changes in

the oral mucosa which may facilitate the penetration of other betel quid ingredients and environmental toxicants into the mucosa (Walker *et al.*, 2003).

3.3.4. HPV Infection

Human papilloma viruses (HPVs) are epitheliotropic DNA viruses that can induce hyperplastic, papillomatous and verrucous squamous cell lesions in the stratified squamous epithelia of skin and mucosa. There are nearly 100 genotypes of HPV but of particular interest are HPV-16 and HPV-18, both of which are strongly associated with malignancy and are termed oncogenic or high risk genotypes (Scully, 1996). Recently, human papilloma virus (HPV) infection has been identified as an aetiologic agent for a subset of OSCCs, specifically those that arise from the base of the tongue and tonsil (Gillison, 2007; Gillison *et al.*, 2008). Patients with HPV DNA-positive OSCCs have been shown to be younger in age by 3 to 5 years and are less likely to have a history of tobacco or alcohol use compared to patients with HPV DNA-negative OSCCs (Gillison, 2007). Interestingly, it has been shown that HPV-positive head and neck squamous cell carcinoma (HNSCC) patients have better prognosis (Weinberger *et al.*, 2006; Ang *et al.*, 2010). However, no studies have yet been conducted in oral cancer alone.

3.3.5. Others

Epidemiological studies have reported that family history may posed as a risk factor for oral cancer (Foulkes *et al.*, 1996). In addition, there may be heritable influences such as genetic polymorphism for xenobiotic enzymes or for DNA repair which results in increased susceptibility to oral cancer (Hung *et al.*, 1997; Anantharaman *et al.*, 2007). Apart from genetic influences, diet rich in animal origin and animal fat as well as low intake of fruits and vegetables are related to increased cancer risk (Edefonti *et al.*, 2010). Furthermore, poor oral hygiene, periodontal disease, chronic candidiasis

infections have been previously linked with oral cancer but the mechanisms involved are largely unknown (Meurman, 2010). Infections may trigger cell proliferation, inhibit apoptosis, interfere with cellular signaling mechanisms and up-regulate tumour promoters. Several oral micro-organisms has been shown to metabolize alcohol to carcinogenic acetaldehyde and thereby explaining the association between poor oral hygiene, alcohol consumption and carcinogenesis (Meurman, 2010). On the other hand, excessive use of mouth wash containing alcohol is also a risk factor for OSCC. It has been reported that the use of alcoholic mouthwash twice daily increased the chance of acquiring oral cancer by over nine times (OR 9.15) for current smokers, over five times for those who drank alcohol (OR 5.12) and almost five times for those who never drank alcohol (OR 4.96) (Guha *et al.*, 2007; McCullough and Farah, 2008).

3.4. Geographical Variation in Risk Factors

According to the IARC data, it is apparent that there is geographical or regional variation in the prevalence of oral cancer. As socio-cultural lifestyle of a population also plays an important role in the aetiology and pathogenesis of oral cancer it is not surprising that the variation in the prevalence is closely related to the practice of risk habits such as chewing of betel quid, smoking and alcohol use amongst the different populations (Zain, 2001; Petti, 2009).

Betel quid chewing is commonly practised in South and South East Asia as well as in the Asia Pacific region. Traditionally betel quid was consumed due to its capacity to induce alertness and euphoria thus improving human capacity in everyday life activities especially in the adverse environmental conditions in these regions. Betel quid use then acquired a primary role in social relations, in public and private ceremonial occasions and thus became an essential part in the tradition and culture in many South and

Southeast Asia and Asia Pacific countries (Gupta and Ray, 2004; Petti, 2009). Betel quid is chewed by approximately 600-1200 million people which makes up 10-20% of the world's population (Gupta and Warnakulasuriya, 2002), making it the fourth most frequently consumed psychoactive substance after nicotine, ethanol and caffeine (Norton, 1998; Gupta and Ray, 2004). At present, betel quid use with its different and heterogeneous form has been reported in countries including Pakistan, India, Bangladesh, Nepal, Sri Lanka, Thailand, Cambodia, Malaysia, Indonesia, Myanmar, Laos, Vietnam, China, Taiwan, Papua New Guinea and several Pacific Islands (Petti, 2009). Prevalence of betel quid usage among adults in Southeast Asia have been reported to be between 25-50% depending on countries but can peak to 80-90% in some areas or among some rural ethnic groups (Gupta and Ray, 2004). In Western countries, usage among the South-Asian migrant communities have been reported to be high (Warnakulasuriya, 2002; Gupta and Ray, 2004). For instance, the prevalence of betel quid chewing among Bangladeshi communities living in London is as high as 80%, of which women are the majority (Bedi and Gilthorpe, 1995; Ahmed *et al.*, 1997). Notably, studies conducted in Taiwan, Micronesia and migrant communities in UK have also shown that 60-70% of children and teenagers have tried betel quid and many of them became regular users (George *et al.*, 1994; Prabhu *et al.*, 2001; Shah *et al.*, 2002; Tsai *et al.*, 2002; Oakley *et al.*, 2005). In terms of smoking habits, despite the fact that it has now reached a global epidemic where tobacco companies are producing cigarettes at the rate of five and a half trillion a year which is nearly 1000 cigarettes for every man, women and child in the planet (Mackay, 2002), there is still difference in smoking prevalence across the different parts of the world. According to the global data on cigarette smoking prevalence, almost 1 billion men in the world smoke of which 35% and 50% were from developing and developed countries respectively. About 250 million women in the world are daily smokers, of which 22% and 9% were from

developed and developing countries respectively. Notably, the percentage of daily smokers among adolescence seemed to be high in Eastern Europe, Latin America, US and South Africa (Petti, 2009).

As for alcohol consumption, history has shown that the consumption of alcoholic beverages is highly valued by certain communities. With the availability of cheap and strong alcoholic beverages, alcohol drinking has become a recreational activity which raises the concern of social impact due to alcohol abuse as seen in tobacco and betel quid use (Musto, 1999). According to the WHO estimates, almost 2 billion people worldwide consume alcohol, and almost 80 million have diagnosable alcohol abuse disorders. The highest consumption levels were reported in European countries such as Czech Republic, Ireland and France with values ranging from 13.5 to 17.5 liter of pure ethanol consumption per capita per year among individuals aged 15 years or greater. The prevalence of heavy drinking among adults has been reported to be high in African, Eastern European and Latin American countries (Petti, 2009). In addition, alcohol use is widespread among adolescents, where heavy episodic drinking or binge drinking (defined as non-habitual drinking occasion leading to intoxication) has been reported to be as high as 20-30% among teenagers in UK, Ireland, Malta, Sweden, Finland, Iceland, Poland and Hungary (WHO, 2004) .

The data above clearly supports the fact that the contribution of risk factors to oral cancer burden varies across different region. Furthermore, La Vecchia and colleagues showed a dramatic increase of oral cancer attributed to smoking and alcoholic drinking in developed countries alone (La Vecchia *et al.*, 1997). Worldwide, 25% of oral cancers were attributed to smoking and 7-19% to alcoholic drinking but in developed countries alone, alcohol consumption and smoking account for up to 75% of oral cancer (La

Vecchia *et al.*, 1997). Similarly Danaei and colleagues in 2005, showed that there is a difference in the number of death attributed to alcohol use and smoking in oral cancer (including the oropharynx) when comparing between low-to-middle income and high-income countries. In low-to-middle income countries, the percentage of death due to alcohol use is 14% while smoking is 37% whereas in high-income countries the percentage of death increased to 33% in alcohol use category and 71% in smoking category (Danaei *et al.*, 2005). As for betel quid chewing, more than 50% of oral cancer is attributed to betel quid chewing in areas of high chewing prevalence (Petti, 2009).

Malaysia is a multiracial country, consisting of Malays, Indians, Chinese and indigenous people, with very different cultural habits. For example betel quid chewing is a traditional custom amongst the Indians, Malays and certain indigenous tribes but not among the Chinese (Awang, 1988). Evidently, betel quid chewing was reported to be widely practiced among Indians working in plantation areas, elderly Malays living in rural areas, as well as indigenous tribes in Sabah and Sarawak with the average prevalence of approximately 7% (Zain and Ghazali, 2001). However in plantation areas alone, the prevalence was reported to be as high as 16% (Tan *et al.*, 2000). On the other hand, smoking was found to be most common among Malays followed by the indigenous tribes in Sabah and Sarawak with prevalence ranging from 22-25% (Haniza *et al.*, 1999; Abd Muttalib *et al.*, 2002). As for alcohol drinking, the prevalence in Malaysia ranged from 4.2%-8.6% with the highest seen among Indians followed by the indigenous tribes in Sabah and Sarawak and among the Chinese (Zain *et al.*, 1995; Zain *et al.*, 1997; Abd Muttalib *et al.*, 2002). When the risk habits were analysed according to gender, betel quid chewing is more common in women with a men to women ratio of 1:3 while smoking and alcohol consumption was more common in men with men to women ratios of 12:1 and 11:1 respectively (Abd Muttalib *et al.*, 2002). Furthermore,

based on a small study conducted by Ng and colleagues in 1986 on oral cancer patients and their risk habits, betel quid chewing represented the most common single habit (83%) followed by smoking (12%) and alcohol consumption (5%). Among those who have multiple risk habits, betel quid combined with alcohol consumption is the most common (59%), followed by betel quid chewing, smoking and alcohol (23%), smoking and alcohol (13%) and betel quid chewing and smoking (5%) (Ng *et al.*, 1986).

3.5. Geographical Variation in the Prevalence of Oral Cancer Subsites

Like the risk habits for oral cancer, the anatomical sites of oral cancer (ICD 10 C00-C06) vary across different parts of the world. In Western countries, tongue, floor of the mouth and lip have been reported to be the most common sites for oral cancer. For example in the United States of America, it has been reported that the tongue remains the most common site of oral cancer (30%) followed by lip (17%) and floor of the mouth (14%) (Silverman, 2001). Similarly in Hungary, these sites were again amongst the most common site of oral cancer with floor of the mouth being the most common (27.7%) followed by lip (26.9%) and tongue (22.7%). It is worth noting that Hungary tops both the morbidity as well as mortality list for both genders in Europe (Nemes *et al.*, 2008). Likewise, tongue cancer has been reported to be the most common cancer in United Kingdom where it accounted for about 40% of the total oral cancer cases (Rodrigues *et al.*, 1998). In Denmark, a small study found that tongue and floor or the mouth were the most common site as well (Pinholt *et al.*, 1997). On the other hand, in Europe and the United Kingdom cancer originating from the gum was the least common at 6.7% and 5% respectively (Rodrigues *et al.*, 1998; Nemes *et al.*, 2008). However a proper comparison for the United States cannot be done with accuracy because cancer

of the gum is commonly grouped together with other sites of oral cavity (Silverman, 2001) which is probably due to its rarity in prevalence.

In Asia, studies looking at the prevalence of different sites of oral cancer have been conducted since the 1960's. A study done in Malaysia or Malaya at that time found that the majority of the oral cancer cases were from the cheek (Marsden, 1960; Ramanathan *et al.*, 1973). Similarly Hirayama and colleagues found similar results in India and Thailand with percentage ranging from 56 to 80% depending on the country (Hirayama, 1966). Since then, similar studies were carried out, and cheek remained as the most common site for oral cancer in India and Malaysia (Chattopadhyay, 1989; Siar *et al.*, 1990). Likewise in Taiwan, Solomon's Island and Sri Lanka, OSCC from the cheek has been reported to be the most prevalent cancer site (Chen *et al.*, 1999; Lumukana and King, 2003; Warnakulasuriya, 2009). Apart from OSCC from the cheek, gum is also a common site in Asian countries as compared to the western world (Chattopadhyay, 1989; Siar *et al.*, 1990; Chen *et al.*, 1999; Chen *et al.*, 2007).

Evidently, OSCC derived from the tongue is increasing among Asian countries of which cheek OSCC was the most common previously with the exception of Taiwan. In Sri Lanka, increasing numbers of tongue cancer among those less than 35 years old have been reported. Iype and colleagues in 2001 reported that tongue is the most common site (52%) followed by cheek (26%) as compared to the previous report from the same center where cancers from the cheek outnumbered those from the tongue (49.9% and 23.9% respectively) (Iype *et al.*, 2001). In Malaysia, a similar trend is observed where the percentage of OSCC from the tongue increased over two decades and outnumbered cheek OSCC (Siar *et al.*, 1990; Lim *et al.*, 2008). Based on the cancer incidence report in 2003-2005, the most common site for oral cancer cases among government hospitals

in Peninsular Malaysia was the tongue (55.3%) followed by the cheek (25.4%) (Figure 3.2) (Lim *et al.*, 2008). The changes seen may perhaps reflect the changing lifestyle habits associated with oral cancer. It has been reported that betel quid chewing is now becoming to be an uncommon habit among those living in urban areas and only practiced only among the people of Sabah and Sarawak and some elderly Malays and Indians living in rural villages (Zain and Ghazali, 2001). Interestingly, based on the WHO Global Status Report on Alcohol, alcohol consumption in Malaysia has increased over the past twenty years (WHO, 2004).

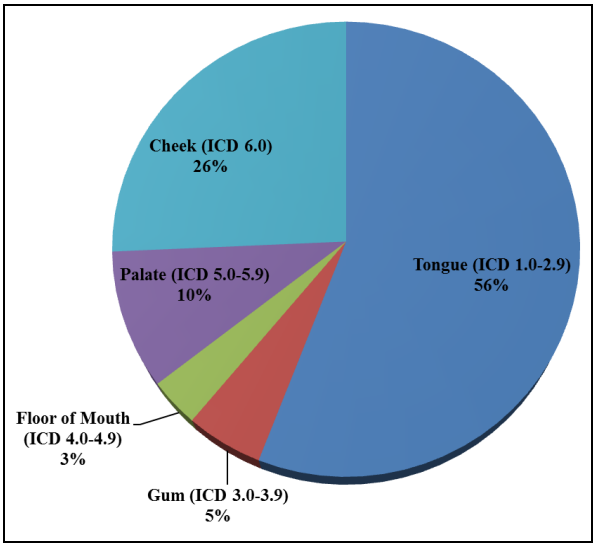


Figure 3.2. Number of oral cancer case by subsites reported in government hospital in Malaysia in 2003-2005 (Lim *et al.*, 2008).

3.6. **Challenges in Oral Cancer Management**

In oral cancer, treatment modality and prognostication relies mainly on clinical staging and histological assessment of the patient and tumour, which includes tumour stage, nodal status, metastasis, pathological grading, pattern of invasion at the invasive front, perineural invasion and excision margins. However, these factors have inherent limitations, for example in disease stage, it is well recognized that patients with similar

stages of oral cancer can have diverse clinical outcome and response to similar treatment regimens (Bankfalvi and Piffko, 2000). Furthermore, in oral potentially malignant disorder, histological features of dysplasia provide little value in terms of predicting which dysplastic lesion may be more or less aggressive over time (Kuo, 2003). Therefore, there is a strong need for the development of a more objective prognostic and predictive tool that can help clinicians define the most appropriate strategies in managing individual patients. In view of these, efforts toward the establishment of molecular markers to complement the current diagnostics and prognostic strategies have been conducted, including the use of molecular classification by gene expression studies (Mendez *et al.*, 2009). Current diagnostic strategies could benefit from a molecular insight into what is happening within the tumour cells, as has been shown with other cancers (van't Veer and Bernards, 2008).

3.7. Global Gene Expression Studies in Cancer

Cancer is caused by the accumulation of genetic and epigenetic changes resulting from altered sequence or expression of cancer related genes such as oncogenes, tumour suppressor genes and genes involved in cell cycle control, apoptosis, cell adhesion, DNA repair and angiogenesis (Hanahan and Weinberg, 2011). With the completion of the human genome project, along with the new technological advances, it is now possible to perform large-scale gene expression analysis to study the genetic complexity of human cancers. There are different methods available for such large scale gene expression analysis which includes differential display (Liang and Pardee, 1992), serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995), representation differential analysis (Diatchenko *et al.*, 1996) and DNA microarray (Golub *et al.*, 1999; Perou *et al.*, 2000). When compared across the different techniques, DNA based microarrays have been popular as they are relatively easy to use

and allow parallel quantification of thousands of genes from multiple samples (Ramaswamy and Golub, 2002; Russo *et al.*, 2003). The DNA microarrays can be defined as an ordered collection of micro spot or probes, each spot containing unique sequences representing the genes in the genome (Russo *et al.*, 2003). In general the technology is based on hybridization between targets derived from biological samples and an array of probes that are immobilized on a matrix (Southern *et al.*, 1999). The hybrization signal produced on each probe is the mRNA level of the corresponding gene in the sample and therefore the signal detected in all the probes reflects the gene expression profile, gene signature or molecular portrait for each of the sample (Russo *et al.*, 2003).

According to Russo *et al*.(2003), gene expression profiling of cancer represent the largest category of research using microarray and appears to be the most comprehensive and affordable approach to characterize cancer at a molecular level. Due to the power of this approach, microarray-based gene expression studies have been performed on a huge variety of cancers including breast, leukemia, head and neck, liver, lung, ovary, pancreas, prostate and stomach amongst others. There are various strategies in microarray based cancer profiling such as 1) tumour versus control, where tumour expression pattern is compared to the control to measure the differences and similarities 2) cancer stratification, where the gene expression of different samples of the same cancer type can be compared to find distinct subgroups 3) comparing gene expression patterns of cancer derived from different stages of progression to identify genetic differences in early and advanced stage of cancer (Russo *et al.*, 2003).

Golub *et al*.(1999) and colleagues demonstrated elegantly that it is possible to distinguish and classify acute myelogenous leukemia (AML) and acute lymphocytic

leukemia (ALL) based solely on gene expression. A panel of 50 gene predictors was shown to be able to distinguish AML and ALL. This is extremely important in acute leukemia as treatment regimens for ALL and AML are different and correct classification ensures that patients are given the right treatment (Golub *et al.,* 1999). Cancer classification based on microarray studies has also been established for solid tumours. Perou and colleagues have shown that breast cancers can be classified into various subtypes according to their gene expression profile, which they refer to as molecular portrait (Perou *et al.*, 2000; Sorlie *et al.*, 2001). Furthermore, Hedenfalk *et al*. (2001)have identified a panel of genes which could distinguish BRCA-1 mutated tumours from BRCA-2 mutated which could lead to a more precise molecular classification of breast cancer (Hedenfalk *et al.*, 2001).

Notably, Alizadeh *et al*. (2000) also demonstrated that gene expression profiling could predict disease outcome. They showed that diffuse B-cell lymphoma (DLBCL) can be separated into two distinct subtypes with different treatment outcomes and survival patterns (Alizadeh *et al.*, 2000). Subsequently, this was confirmed by another group (Shipp *et al.*, 2002). Since then, more studies have been conducted in other cancer types including prostate and medullablastoma to explore the expression-based outcome/survival prediction (Dhanasekaran *et al.*, 2001; Pomeroy *et al.*, 2002).

3.7.1. Challenges in Conducting Microarray Experiment

The quality and amount of RNA required for microarray experiments remains the main limitation of this technique in looking at global gene expression changes. Microarray analyses primarily use fresh frozen tissue samples, which are limited mainly due to cost and feasibility of collecting and storing large numbers of these samples. Processing of tissue is crucial as high quality RNA is needed for successful microarray experiments.

Ideally, tumour specimens from surgical incision should be snap frozen in liquid nitrogen in the operation theatre to prevent degradation of nucleic acid (Srinivasan *et al.*, 2002). Another challenge lies in the inherent heterogeneity of human cancers where each tumour consists of varying proportion of tumour cell, stromal elements, vasculatures and other cells such as inflammatory cells. Therefore changes in gene expression patterns when comparing two different biopsies samples are a product of all the cell types present in that samples (Russo *et al.*, 2003). Researchers have tried to control this variability by using specimens of similar composition of tumour cells with the help of laser capture microdissection, or macrodissection which ensures that cancer cells are at least 70% in all tumour specimen and similarly to the non-cancerous oral mucosa, at least 70% epithelial cells (de Bruin *et al.*, 2005; Roepman *et al.*, 2009).

Furthermore, because of costs and the rarity of certain clinical samples, performing large studies are difficult. In addition, expression based profiling should also be coupled with clinical data and outcome such as survival. To address the aforementioned challenges, the suggested solutions were 1) to link large scale expression profiling with a clinical trial since ideally, clinical studies should be coupled with clinical data as well as the understanding of other molecular changes at DNA and protein level (Ramaswamy and Golub, 2002) and 2) to set up tissue banks for cancer that would allow researchers to perform a comprehensive cancer profiling at mRNA, DNA as well as protein level with complete clinical data and follow up (Bathe, 2009). Microarray experiments generate a vast quantity of data therefore making sense of this vast amount of data poses a huge challenge (Brazma *et al.*, 2000). In fact it has been reported that the bottle neck in biological investigation has shifted from data generation to data analysis (Sherlock, 2001). In recognizing the challenge, many integrated cancer microarray database with data mining tools that have been developed and made available to the public including

Oncomine and DAVID (Rhodes *et al.*, 2004; Huang da *et al.*, 2009). In addition, due to the growing demand for a need of public repositories for microarray data, the National Center for Biotechnology Information (NCBI) has set up the Gene Expression Omnibus (GEO) which allows submission, retrieval and storage of microarray data (Edgar and Lash, 2002). However, it is still difficult to find a single analysis tool that can answer all questions and thus a mixture of analysis tools are currently used by many microarray researchers in order to find biological relevance of their data (Butte, 2002).

Nonetheless despite the challenges associated with microarray experiments, this technology has been identified as a core technology for the advancement of medicinal product development and individualized medicine by the US Food and Drug Administration (FDA) Critical Path Initiative (http://www.fda.gov/ScienceResearch/SpecialTopics/Critical Path Initiative (2009)). Moreover, the Microarray Quality Control Study (MAQC) have demonstrated inter-platform and inter-laboratory reproducibility and technical reliability of the DNA microarray based test using breast cancer as an example (Shi *et al.*, 2006). Given the power and reliability of this platform, molecular diagnostics tests are expected to become an important tool in tailoring cancer management for individual patients as well as in identifying patients who respond to experimental anticancer drugs in clinical trials (van't Veer and Bernards, 2008). In fact, the successful use of the microarray has resulted in a FDA approved prognostic test for breast cancer commercially known as MammaPrint (Agendia, Netherlands).

3.8. Types of Biological Specimens for Gene Expression Microarray Studies

As the quality of input RNA directly influences the reliability and amount of valuable data that can be obtained from microarray, snap frozen tissue continues to be the preferred source of RNA for the use in these experiments (Elkahloun *et al.*, 2002; Coudry *et al.*, 2007). Thus, the majority of the available array technologies such a cDNA spotted array, the genechip™ array and BeadArray™ are all designed to use fresh frozen tissues. Unfortunately, fresh frozen samples are not always readily available and need to be collected prospectively from patients at clinical centers unless tissue banks with readily available samples are accessible. Despite the effort in systematic banking of tissues for research, prospective collection of patient tissues particularly from rare diseases will limit their immediate use. Moreover, the use of fresh frozen tissues particularly to examine patient outcome is limited by the availability of clinical and follow-up data associated with the patient. In contrast, formalin-fixed paraffin-embedded (FFPE) tissues are abundant as they are processed and stored routinely in clinical practice. Further, information on various disease stage associated with these patients can be correlated with molecular findings immediately.

Traditionally, FFPE samples are not utilized in microarray experiments due to the chemical modification of nucleic acid by formalin resulting in poor RNA quality (Masuda *et al.*, 1999; Williams *et al.*, 1999; Karsten *et al.*, 2002). In view of this, novel microarray technologies and tissue processing protocols specifically designed to address RNA degradation issues has recently been developed (Coyle and Johnston, 2010). These includes modification in the RNA extraction processes, cDNA synthesis and microarray platforms. Among the microarray platforms adapted for FFPE samples is

the cDNA-mediated annealing, selection, extension and ligation (DASL) assay that is based on BeadArray™ technology, which opens up new opportunities to study cancer, as FFPE samples represent the largest source of archival specimens with clinical data. The DASL assay is a sensitive and flexible gene expression profiling system that does not depend on intact poly-A tail and has been shown to be able to analyse compromised RNA samples (Bibikova *et al.*, 2004a; Rentoft *et al.*, 2009; Saleh *et al.*, 2010). In addition it has been reported that as little as 50 ng of RNA is needed for RNA profiling from FFPE tissues stored from 1 to more than 10 years (Bibikova *et al.*, 2004a). Other microarrays platforms adapted for FFPE samples include CupPrint assay which is an oligonucleotide array specifically design for adenocarcinoma of unknown origin (Horlings *et al.*, 2008) and high density Disease Specific MicroArrays (DSAs™) which captures all transcripts transcribed in specific disease setting such as breast, colorectal or non-small cell lung cancer (Farragher *et al.*, 2008).

Despite the concerns of the use of FFPE specimens in microarray experiments, Haque *et al*. (2007) demonstrated that the genes found to be differentially expressed in glioblastomas compared to normal control brain between fresh frozen and FFPE were similar, and further demonstrated that although the number of differentially expressed genes was smaller in the FFPE group, the molecular sub-classification of glioblastomas was nevertheless possible in both types of specimens (Haque *et al.*, 2007). To further support the consistency between FFPE and fresh frozen samples, Srivastava *et al*. (2008) identified similar key pathways involved in prostate cancer development and progression in both fresh frozen and FFPE samples (Srivastava *et al.*, 2008). A high number of FFPE samples yielded good quality microarray data obtained from FFPE samples reflect the success of using FFPE in microarray studies (Hoshida *et al.*, 2008; Saleh *et al.*, 2010). Notably, others have also successfully utilized FFPE specimens on

microarrays to establish gene signatures that are indicative of patient prognosis (Chung *et al.*, 2004; Bibikova *et al.*, 2007) thus strongly supporting the use of FFPE specimens for the reliable identification of biomarkers for cancer.

3.9. **Gene Expression Studies in Oral Squamous Cell Carcinoma (OSCC)**

Gene expression microarrays have been widely used to study OSCC, which represents more than 90% of oral cancer. In general, microarrays have been used for two main purposes: first, to compare the gene expression profile of OSCC to normal oral mucosa to provide an insight into the molecular mechanism of OSCC; and second, to identify prospective biomarkers for early detection, prognosis as well as therapeutic targets (Leethanakul et al., 2000; Alevizos et al., 2001; Hwang et al., 2003). Likewise, metastatic tumours were also compared to non-metastatic to determine the genes that may be involved in mobility and metastasis (Nagata et al., 2003; Warner et al., 2004). Microarray studies were also conducted to determine how risk factors may modulate the gene expression in OSCC. Tsai et al. (2004) in Taiwan analysed mRNA expression patterns in oral cancer patients with betel quid chewing habit and found eighty four genes involving cell adhesion, cell shape, growth, apoptosis, angiogenesis, metastasis, and metabolism were deregulated (Tsai *et al.*, 2004). Recently,it was reported that gene expression was found to be closely influenced by exposure to different risk factors. Cheong *et al*. (2009)demonstrated that gene expression patterns of oral cancer from betel quid chewers were different from those of smokers and the authors suggested that these differences should be taken into consideration when developing biomarkers for prognosis or therapeutic application (Cheong *et al*., 2009).

Microarray derived data was also found to be associated with clinical outcome. Several groups have suggested that gene expression profiles may be used to classify OSCC patients into subgroups based on prognosis and this grouping was a useful outcome predictor (O'Donnell *et al.*, 2005; Chen *et al.*, 2008a). In recent times, Mendez *et al*. (2009)demonstrated that a combination of gene expression signature and TNM staging better predicts survival of OSCC patients compared to TNM staging alone which further warrant the use of microarray toward better management of oral cancer patients (Mendez *et al.*, 2009). Resulting from microarray experiments, potential biological relevant targets have been investigated in *in vitro* models through exogenous expression or suppression of these genes to reveal their specific roles and further provide clues to the mechanism and pathways that may be involved in tumour initiation and progression (Kim *et al.*, 2004; Miyazaki *et al.*, 2006).

Notably not all microarray studies yielded similar differentially expressed genes. This may in part be due to the use of tissues from different sites of the oral cavity and some have even included a variety of head and neck tissues in the same experiments (Ginos *et al.*, 2004; Toruner *et al.*, 2004). Several groups have shown that the expression of oral cavity samples is different from other HNSCC sites and appears to be more heterogeneous (Huang *et al.*, 2002; Chung *et al.*, 2004). In fact, Warner *et al*. (2004) also showed that based on their gene expression profiles, HNSCC cell lines could be clustered according to the sites from which they were derived thus suggesting that HNSCC from different sites may be distinct from one another (Warner *et al.*, 2004). Furthermore, Severino *et al*. (2008)demonstrated the molecular heterogeneity in OSCC from different sites of the oral cavity by comparing two sites, tongue and floor of the mouth (Severino *et al.*, 2008). Therefore, the heterogeneity of samples included in

many gene expression studies may partially explain the dissimilarity in the genes identified from different experiments using OSCC (Mendez *et al.*, 2009).

3.9.1. Evidence of Molecular Differences in OSCC Subsites

In 1996, Paterson *et al.* reviewed the spectrum of molecular changes in OSCC from Western countries in which the predominant site is the tongue and floor of mouth, and Asian countries where cheek and gum are the most common sites. It is found that p53 mutations are common in tumours from the West (47%) while tumours from Asian countries were characterized by the involvement of ras oncogenes, including mutations, loss of heterozygozity of H-ras and amplification of K-ras, H-ras, events which are uncommon in the West (Paterson *et al.*, 1996). In 2000, Schwart and colleagues used a hamster model to demonstrate that cancers from the cheek and cancers from the tongue exhibit differences in growth, oncogene expression and development of program cell death based on immunohistochemistry analysis of p53, proliferating cell nuclear antigen (PCNA), BCL-2 and nucleosome formation (Schwartz *et al.*, 2000). Similarly, Sathyan and colleagues reported that cancer of the cheek and tongue represent different biological subentities for oral cancer by demonstrating that the expression of the major cell cycle regulatory proteins including p53, Rb, p16, p21, cyclin D1, CDK4 and PCNA were different for different sites. These results further indicate that OSCC from different anatomical sub-sites are characterised by alteration of different genetic pathways (Sathyan *et al.*, 2006).

3.9.2. Clinical Differences in OSCC Subsites

At the clinical level, tumours from different sites of the oral cavity have been reported to behave differently. The biological aggressiveness of Stage I tongue cancer is noteworthy and is reflected in higher rates of occult metastasis than similarly staged lesions arising from other oral sites (Clayman, 1997). Occult node metastasis are present in 30-40% of early lesions while local/regional recurrence in patients with tongue cancer account for 60-70% cancer deaths which is higher than that of other sites of the oral cavity (Clayman, 1997). Cheek cancers are generally the least aggressive while gum cancer may invade the underlying bone, thus up grading the stage of disease (Rautava *et al.*, 2007). In addition, patients with tumours from the different sites of the oral cavity have different survival rate. Sathyan and colleagues conducted a clinico-epidemiological study in India comparing tongue cancer and cheek cancer and found that even though the size of tongue cancers were small and in early stage at the time of presentation, disease free survival and overall survival were poor in tongue cancer compared to that in cheek (Sathyan *et al.*, 2006). Similarly, Rusthoven and colleagues demonstrated that amongst patients with Stage I/II SCC of the oral cavity, oral tongue SCC is associated with the lower rates of overall and cause specific survival compare to the other oral cavity subsites (Rusthoven *et al.*, 2008). Consistently, Brandizzi and colleagues showed that in general, tongue and floor of the mouth has the lowest 5-year survival rate (27%, 19% respectively) compared to cheek and gum which has a much higher survival rate at 54% and 41% respectively (Figure 3.3) (Brandizzi *et al.*, 2008). Moreover, a study conducted in Taiwan has identified the anatomical site as an important prognostic factor of which the tongue is associated with poorer prognosis (Chen *et al.*, 2007).

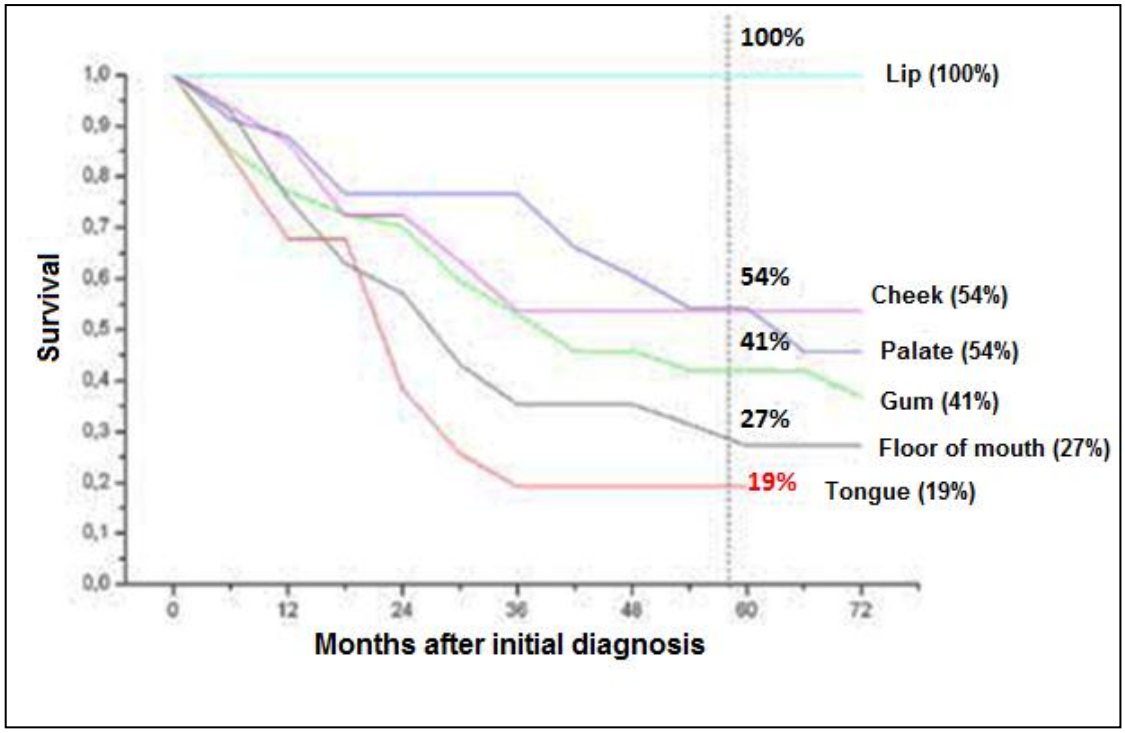


Figure 3.3 Kaplan Meier survival curve of oral cancer demonstrating the different survival rates of patients with OSCC from different sites.

Interestingly, tumours from different sites of the oral cavity have been shown to respond to treatment differently. Zelefsky *et al. (*1990) reviewed treatment results from postoperative radiotherapy of different sites of advanced OSCC and found that there were significant differences in terms of response to combined surgery and radiotherapy. With similar T stage, margin status and median radiotherapy dose, the 5-year local failure rate was higher in tongue (38%) compared to floor of mouth (11%). Furthermore the median survival after recurrence was 9 months for tongue cancer and floor of mouth was 40 months (Zelefsky *et al.*, 1990). Similarly Yao *et al.* (2007) have reported that in 55 patients who received intensity modulated radiation therapy (IMRT), tongue cancer again was associated with significantly worse 2-year locoregional recurrence free survival compared to floor of the mouth and other oral cavity subsites which includes OSCC from the cheek, alveolar ridge, retromolar area and lips (Yao *et al.*, 2007). Therefore, current evidence indicates that OSCC from different sites of the oral cavity

have diverse clinical behavior, respond differently to therapeutic regimes and have distinct survival rates. This could be due to differences in the genetic alterations already reported in specific genes, however a comparison of the global genetic alterations in the various sites is currently limited.

3.10. **Importance in Addressing the Heterogeneity in the Different Oral Subsites in OSCC**

Consistent with clinical observations, the analysis of specific genes revealed that there are molecular heterogeneity in OSCC from different anatomical subsites (Schwartz *et al.*, 2000; Sathyan *et al.*, 2006). On the other hand, some may argue that it is the anatomical position of the oral subsites and not the molecular heterogeneity that contributes to clinical differences seen in OSCC for example, the tongue is closer to the lymph nodes compared to the other sites thus facilitating the metastasis of tumour cells (Werner *et al.*, 2003). Nonetheless, regardless of the factors contributing to the clinical differences, many research groups particularly those involved in identifying prognostic and diagnostic markers in OSCC have already begin to control for the possible heterogeneity in their study design by including only biological samples from a specific subsites of oral cancer (Ye *et al.*, 2008; Estilo *et al.*, 2009; Rentoft *et al.*, 2009; Saleh *et al.*, 2010). Moreover, focusing on specific anatomical site is important as it provides accurate and clinically useful information on the biology and prognostic significance of genetic alterations in oral cancer (Sathyan *et al.*, 2006). In order to determine if the genetic heterogeneity could be associated with the clinical differences, several studies were recently conducted (Ziober *et al.*, 2006; Severino *et al.*, 2008). However, such studies were conducted using a small number of samples and focused mainly on tongue and floor of the mouth of which are amongst the most prevalent sites in the West. A gene expression comparison study that compares other subsites including cheek and

gum which are amongst the most common sites in Asian countries is much needed to explore the molecular differences of these oral subsites during carcinogenesis.

CHAPTER 4

MATERIALS & METHODS

4.1. Study Design

A cross sectional study was conducted using formalin fixed paraffin embedded (FFPE) samples to analyse the gene expression patterns of three different sites of OSCC compared to non-cancerous oral mucosal tissues. The workflow of this study is depicted in Figure 4.1.

4.2. Study Specimens

Three types of specimens were used in the different experiments in the study. A total of 116 FFPE samples were used in the gene expression experiments and 95 specimens were utilised for immunohistochemistry staining. A total of 54 fresh frozen samples obtained from the Malaysian Oral Cancer Database and Tumour Bank System (MOCDTBS) and Cancer Research Initiatives Foundation (CARIF) were used in quantitative polymerase chain reaction (qPCR) . OSCC cell lines established by CARIF were used in the development of *in vitro* model and functional assays. The sociodemographic characteristics of the specimens such as age, gender, habits as well as clinical data were obtained from MOCDTBS co-ordinated by the Oral Cancer Research and Coordinating Center (OCRCC) (Appendix A-C). Ethical approval was obtained from the Institutional Review Board of the Faculty of Dentistry, University Malaya (Ethical Clearance No: DF OP0601/0003(L)). All selected cases (test and control) were reviewed and verified by an oral pathologist.

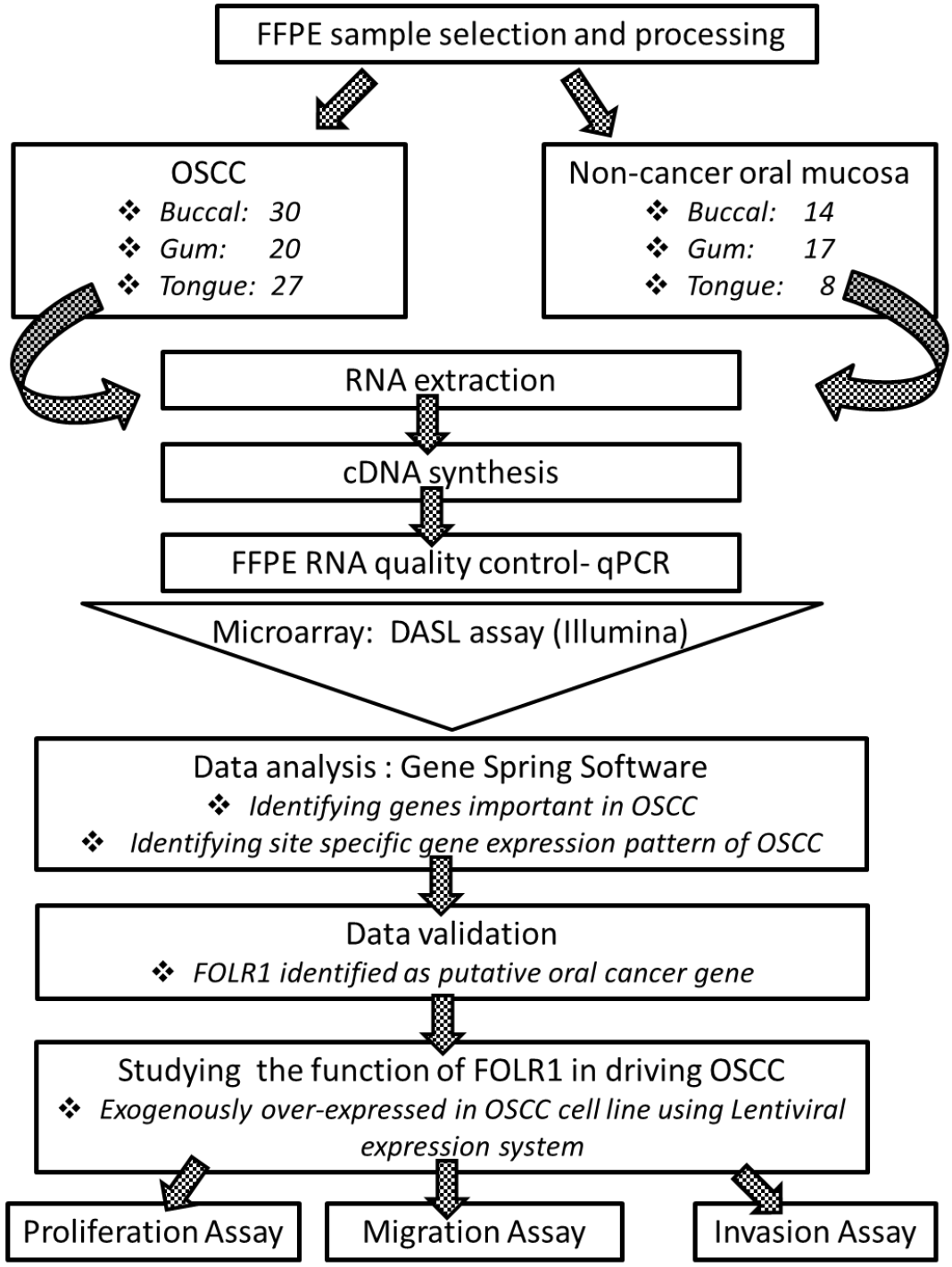


Figure 4.1 Project workflow.

4.3. Specimen Selection and Processing

Sample selection for all specimens was based on the inclusion and exclusion criteria mentioned below.

Inclusion criteria:

1. Samples histologically confirmed to be OSCC and non-cancerous oral mucosa.
2. OSCC and non-cancerous oral mucosa which are specific to only one site of the oral cavity, either cheek, tongue or gum.

Exclusion criteria:

1. Samples, which were not histologically confirmed to be OSCC or non-cancerous oral mucosa.
2. OSCC and non-cancerous oral mucosa, which are not specific to one site of the oral cavity.

Processing of specimen was done according to the types of specimens detailed below.

4.3.1. Formalin-Fixed Paraffin-Embedded (FFPE) Specimens

For microarray experiments and immunohistochemistry, OSCC, which were FFPE specimens were obtained from the Oral Pathology Diagnostic Laboratory, Faculty of Dentistry, University of Malaya. Non-cancerous oral mucosal tissues used in this study were surface epithelium of oral lesion from matching sites, mainly the fibro epithelial polyps, fibrous epulis and gingival tissues obtained during the surgical removal of impacted wisdom tooth. These non-cancerous oral mucosal tissues were obtained from the Oral Pathology Diagnostic Laboratory, Faculty of Dentistry, University of Malaya and Oral Pathology Department, Faculty of Dentistry, Universiti Kebangsaan Malaysia.

For OSCC specimens, reference slides were first made for confirmation of the diagnosis and to gauge the percentage of tumour cells. Areas that have at least 70% tumour cells were marked on the reference slide and a similar area was matched and marked on the block using a blade (Figure 4.2a). A total of 80-100 µm sections of tissue from the marked area for each of the block were placed into a 1.5 ml microcentrifuge tube filled with 1 ml of xylene to remove the paraffin before the RNA extraction procedure detailed in Appendix D was performed. Similarly, for non-cancerous oral mucosal tissues, reference slides were made from each FFPE samples for confirmation of the diagnosis and to gauge the percentage of epithelial cells. Here, areas with at least 70% epithelial cells were marked on the block. The marked area of the normal tissue block, were cored with a 1.5 mm needle using the ATA-100 Advanced Tissue Arrayer (Millipore (Chemicon), Billerica, U.S.A) (Figure 4.2 b-d). The tissue cores were placed into a 1.5 ml microcentrifuge tube filled with 1 ml of xylene to remove the paraffin before the RNA extraction procedure (Appendix D).

For immunohistochemistry experiments, an independent set of FFPE OSCC and non-cancerous oral mucosal tissues (n = 95) from cheek , tongue and gum was chosen to create an array of tissues, which is referred to as a tissue macroarray (TMaA). The tissue arrays were then sectioned at 4 µm thickness and used in immunohistochemistry experiments (Figure 4.2e).

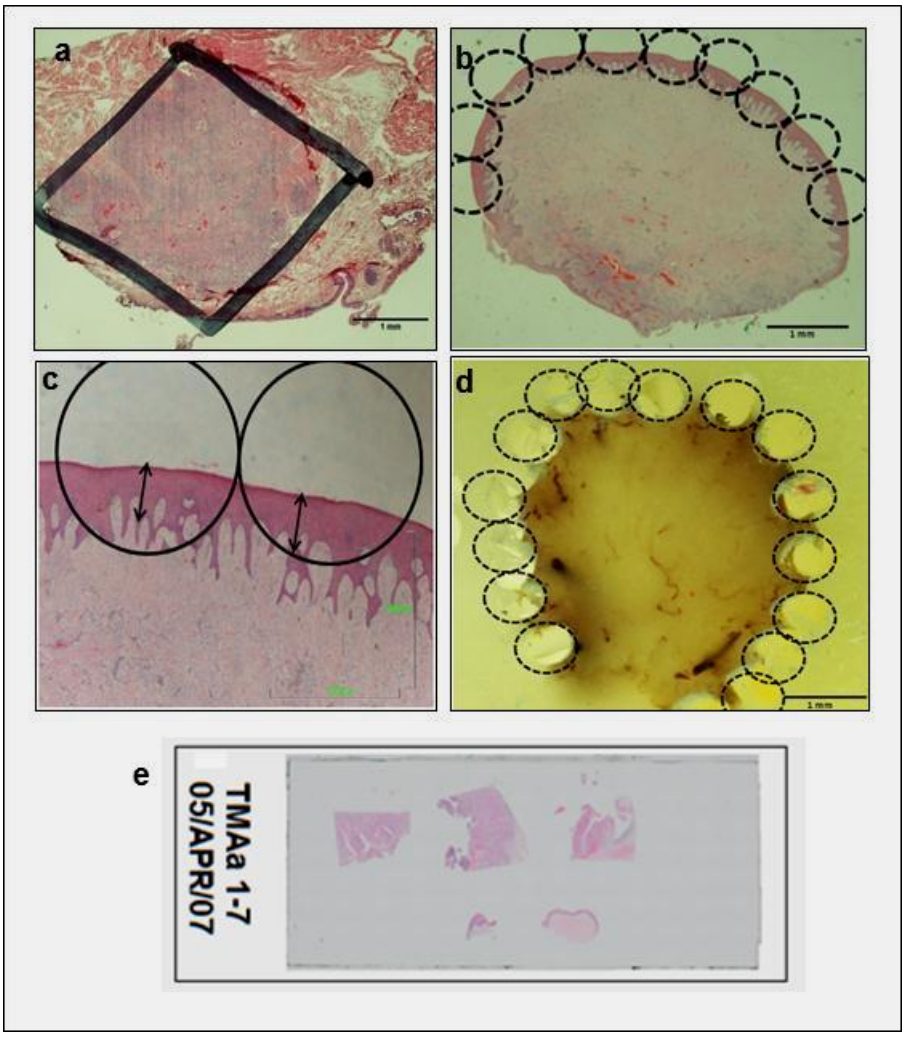


Figure 4.2 Images depicting how a specific area in a specimen was selected. (a) For OSCC samples, tumour percentage was gauged under microscope and areas with >70% tumour cells were marked. Marked areas were then macro dissected, sectioned and included in the experiment. (b) For non-cancerous oral mucosa, epithelial areas were marked and the percentage of epithelial cells were gauged under the microscope and cored with a tissue arrayer using a 1.5 mm needle (c) H&E picture of surface epithelium of non-cancerous oral mucosa at 400x magnification, indicating that the epithelial is approximately 1/3 of the diameter of the needle. (d) Tissue block after coring, the dotted circle represents the actual size of needle for b & d. (e) Representative image of a H&E stained section of a tissue macroarray (TMaA) block where sections were used during the immunohistochemistry experiments.

4.3.2. Fresh Frozen Specimens

cDNA from fresh frozen specimens were obtained from OCRCC-CARIF nucleic acid bank. Briefly, fresh frozen samples stored in liquid nitrogen were taken out and sectioned frozen in a cryostat (Leica Microsystem, Wetzlar, Germany) at about -20 to -30°C. A reference slide was made to confirm diagnosis and to gauge the percentage of tumour cells for tumour tissues and epithelial cell for non-cancerous oral mucosal tissues. Tissue samples that meet the criteria of at least 70% tumour /epithelial cells, were cryo-sectioned to obtain a total of 500 µm. RNA was extracted from these tissues, RNA which passed the quality control criteria (section 4.5.1) were converted to cDNA, which was further used in qPCR experiments to validate differentially expressed genes identified from the microarray analysis.

4.3.3. Cell Lines

OSCC cell lines and normal oral keratinocyte primary cultures derived by others at CARIF were used to study the function of specific genes in oral cancer (Hamid *et al.*, 2007). The cells were maintained as described previously (Freshney, 1987) in DMEM:F12 with 10% fetal bovine serum (FBS). Cells were grown at 37°C in a humidified atmosphere with 5% CO_2. Sub-confluent OSCC cells were routinely subcultured by treating them with 0.25% trypsin for 10-15 minutes as described previously (Freshney, 1987). DMEM:F12 with 10% FBS was added to neutralize the trypsin and the cell suspension was pelleted at 1200 g for 5 minutes in a centrifuge. For storage, cells were suspended in DMEM:F12 with 10% FBS and 10% dimethyl sulphoxide (DMSO) at a concentration 0.5-1 x 10^6 cells per ml and frozen slowly to -70°C before transferring to liquid nitrogen (-184°C) for long term storage. Cells were revived by thawing the vial in warm water. Cells were then pelleted and resuspended in

DMEM:F12 with 10% FBS before being transferred to 25 cm^2 flask for routine culture as described above.

4.4. RNA Extraction

RNA was extracted from three different types of specimens namely FFPE tissues, fresh frozen tissues and cell lines. Various methods were used to extract RNA as detailed in Appendix D.

4.5. RNA Quantitation and Quality Control

RNA extracted was quantitated using Nanodrop Spectrophotometer (Thermo Scientific, Waltham, U.S.A). RNA extracted from fresh frozen tissues and cell lines were electrophoresed on a 2% agarose gel (Appendix E) and stained with ethidium bromide. Absorbance ratio of 260/280 was measured to determine the quality of the RNA. In addition, the quality of the RNA was determined using Agilent 2100 bioanalyser (Agilent Technologies, California, U.S.A).

4.5.1. RNA Quality Control for FFPE Specimens

To determine the utility of FFPE specimens to be used in microarray experiments, a quality control assay was done using quantitative polymerase chain reaction (qPCR) as recommended by the manufacturer (Illumina, San Diego, U.S.A). This assay detected the expression of *RPL13a* gene (90bp) and the Ct value for each sample was compared to the Ct value from a commercial human reference control (Clontech, Mountain View, U.S.A) which was run in parallel (Bibikova *et al.*, 2004a; Bibikova *et al.*, 2007). Only samples which had a Ct difference of less than 17 cycles compared to the reference were included in the microarray study. qPCR was performed in a 10 µl reaction

mixture containing 0.25 µM of RPL13A primer set (Appendix E), 1 µl of cDNA in 1X concentration of Power SYBR® green dye (Applied Biosystems, Foster Drive, U.S.A) in triplicates using ABI 7000 DNA Sequence Analyzer (Applied Biosystems, Foster Drive, U.S.A). The reaction was performed using standard real time PCR cycle of 95°C for 1 minute, followed by 40 cycles of 95°C for 1 minute and 60°C for 1 minute.

4.6. cDNA Synthesis from Extracted RNA

cDNA was synthesized from RNA extracted in section 4.4 and used as templates in microarray experiments, qPCR validation as well as in the development of *in vitro* model and functional assays. RNA from FFPE specimens were converted to cDNA as mentioned in section 4.7 whereas RNA from fresh frozen specimens and cell lines were converted to cDNA using High Capacity cDNA reverse transcription kit (Applied Biosystem, Foster Drive, U.S.A). A total of 2 µg of RNA was added to the reverse transcription master mix to a total volume of 100 µl (Appendix E). Reverse transcription was performed in a thermal cycler (GeneAmp PCR System, Applied Biosystems), under these conditions: 25°C for 10 minutes, 37°C for 120 minutes and the reverse transcription product (cDNA) was either used directly or stored at -20°C until further use.

4.7. Microarray Experiments

Microarray experiments were carried out using DASL Assay (c**D**NA mediated **A**nnealing, **S**election, extension and **L**igation assay). The DASL assay from Illumina (San Diego, U.S.A) interrogates 502 genes that have been implicated in cancer initiation and progression including molecules within cancer signaling pathways. Only RNA from FFPE samples that have a concentration of at least 40 ng/µl and passed the quality

control assay described in section 4.5.1 were included in the RNA profiling. Briefly, cDNA were synthesized from 200 ng of RNA using biotinylated oligo dT and random nanomers. The query oligos, which consist of the 502 genes (http://www.illumina.com), were annealed to the biotinylated cDNA and bound to streptavidin beads to select the cDNA/oligo complexes. Non-specifically hybridized and excess oligos were removed by washing steps. Only oligos that were correctly annealed were extended and ligated to generate amplifiable products. These templates were fluorescently labeled during PCR amplification, hybridized to DASL Sentrix Beadchip array and scanned using Illumina BeadArray Reader to generate fluorescence intensity data.

4.8. **Microarray Data Analysis**

The array signals generated were uploaded to Beadstudio (Illumina, San Diego, U.S.A) for background normalization. The array data was filtered by excluding genes with detection score of > 0.01. Samples where less than 50% of the genes had significant intensities were excluded in further analysis. The data was exported to GeneSpring GX Version 7.0 (Agilent Technologies, Santa Clara, U.S.A) for analysis. Here, unsupervised analysis was carried out to determine the distribution of OSCC and non-cancerous oral mucosal samples using hierarchical clustering and Principal Component Analysis (PCA). In this analysis, the software segregates samples based on the similarity of their gene expression pattern.

4.8.1. Identification of Genes Differentially Expressed Between OSCC and Non-cancerous Oral Mucosa

In order to identify differentially expressed genes between OSCC and non-cancerous oral mucosal tissues, Welch t-test analysis (p-value < 0.05) with a False Discovery Rate (FDR) below 0.05 (Benjamin and Hochberg method) was used, where an average signal intensity from all OSCC specimens was compared to the average signal intensity from non-cancerous oral mucosal specimens for each gene hybridized to the Beadchip™ regardless of which site the cancer originated from. Genes in OSCC that were 1.5 fold changed compared to non-cancerous oral mucosa were considered differentially expressed and labeled as the T vs. N gene list.

4.8.2. Identification of Differentially Expressed Genes in OSCC and Non-cancerous Oral Mucosa of Specific Sites

To identify genes that are differentially expressed between tumour and non-cancerous oral mucosal tissues in each site, all tissues were grouped according to their sites: cheek, gum or tongue. Three separate differentially expressed gene lists were generated from comparing site specific OSCC to non-cancerous oral mucosal specimens from matching sites. The gene lists were labeled as B vs. NB (cheek OSCC compared to non-cancerous oral mucosa from cheek), T vs. NT (tongue OSCC compared to non-cancerous oral mucosa from tongue), and AM vs. NG (gum OSCC compared to non-cancerous oral mucosa from gum).

4.8.3. Identification of Significantly Altered Pathways in OSCC

In order to identify important signaling pathways involved in oral carcinogenesis, the four differentially expressed gene list between tumour versus non-cancerous oral

mucosa (T vs. N) and site specific (B vs. NB, T vs. NT and AM vs. NG) were analysed using DAVID Bioinformatics Database (Huang da *et al.*, 2009). The analysis was conducted by comparing the gene list to known cancer pathways in the database.

4.8.4. Comparison of Significant Pathways Commonly Changed and Those That Are Site Specific to OSCC

In order to find genes and pathways common and enriched in OSCC derived from cheek, gum and tongue, the site specific gene lists and the pathway lists were compared using a Venn diagram from the Overlapper Software (www.bea.ki.se/jnlp/overlapper). Genes and pathways that were overlapping among all sites were termed common whereas genes and pathways that were exclusively to each site were termed enriched in particular sites.

4.9. Validation of Microarray Results

Genes selected for validation were based on the fold change from the microarray experiments and their potential biological relevance in OSCC. The potential relevance of the genes were gauged by extensive literature review and by comparing the pattern of expression level generated from this study to previous microarray studies where data have been publically deposited in the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/sites/entrez?db=geo) and Oncomine 4.3.

4.9.1. Quantitative Polymerase Chain Reaction (qPCR)

qPCR reaction mix were prepared in a 96-well plate in a 20 µl mixture which consisted of 2 µl cDNA (Section 4.6), 1x Power SYBR® Green Master Mix (Applied Biosystems, Foster Drive, U.S.A) and 0.75 µM of primers (Sigma-Aldrich, St. Louis, U.S.A)

(Appendix F) and amplified for 40 cycles using ABI 7000 Sequence Detector (Applied Biosystems, Foster Drive, U.S.A) as described in section 4.5.1. Primers used in the qPCR were designed using Primer Express Software (Applied Biosystems, Foster Drive, U.S.A) and Primer Blast (NCBI) using the reference sequence from microarray probes that generated the highest intensity signal from the microarray experiment.

4.9.2. Immunohistochemistry (IHC)

Immunohistochemistry was carried out using two different systems depending on the antibodies that were used. Dako Envision® + Dual Link System Peroxidase (DAB+) kit (DAKO, Glostrup, Denmark) was used for antibodies generated from mouse/rabbit including ITGB4, MMP1, MMP10, FOXP3 and CD3 while DAKO LSAB™+HRP Rabbit/Mouse/Goat (DAB+) (DAKO, Glostrup, Denmark) was used for the antibody against BCL2A1, which was raised in goat. In brief, paraffin sections of 4 µm thickness were deparaffinized and rehydrated in graded ethanol. After rehydration, the tissues were subjected to heat induced antigen retrieval method using microwave using retrieval buffers tabulated in Appendix F. The tissues were further blocked with endogenous peroxidase prior to immunostaining of primary antibody for 30 minutes. The tissues were then washed with 1X PBS and incubated with a secondary antibody provided by the respective kits mentioned previously for another 30 minutes at room temperature. The bound antibody was then visualized using DAKO DAB chromogen (DAKO, Glostrup, Denmark). Negative controls were performed by omitting the respective primary antibodies. Positive controls were included where possible. (Appendix F).

BCL2A1, ITGB4 & MMP1 immunoreactivity was scored +1, +2, +3 based on relative intensity and zero was scored for negative staining. Zero and +1 were then categorized

as low expression and +2 and +3 categorized as high expression as described previously (Tomozawa *et al.,* 2000). For CD3 and FOXP3 antibodies, two consecutive sections were stained for CD3 and FOXP3 respectively. The frequency of CD3 and FOXP3 positive cells was determined as described by others (Loddenkemper *et al.*, 2006; de Boer *et al.,* 2007). Briefly, 10 randomly selected high power field (HPF) of the CD3-stained section was captured at 200X magnification and the same areas were retraced in the corresponding set of FOXP3 stained sections and digital images of these were captured for quantitation. The total numbers of positive cells were counted using "cell counter" option in Image J Software and the HPF as mentioned previously were average in each case. The percentage of FOXP3 positive cells was expressed as percentage of CD3 positive cells per HPF.

4.10. **Developing an *In Vitro* Model to Study the Function of the FOLR1 in OSCC**

Cell lines that were suitable to investigate the function of *FOLR1,* a gene identified from the study were selected by first determining the expression level of this gene in several OSCC cell lines. The detailed protocol for selecting these cell lines can be found in Appendix D. The cell line exogenously expressing FOLR1 was referred to as 188T_FOLR1 and the cell line that was transduced by the vector alone was designated 188T_pLenti from here on.

4.11. Using an *In Vitro* Model to Perform Functional Assays to Determine the Role of FOLR1 in OSCC

4.11.1. Cell Proliferation Assay

To determine if FOLR1 conferred an increase in proliferation, cell proliferation assay was conducted by seeding a specific number of cells and counting them at 24 hour intervals. A total of 2×10^4 cells per ml of 188T_FOLR1 and 188T_pLenti cells were seeded in triplicate in a 6-well tissue culture plate. Every 24 hours, the cells were trypsinized and total viable cells were counted using CASY® cell counter (Innovartis, Germany). A proliferation curve was plotted for 188T_FOLR1 and 188T_pLenti. The doubling times and cell proliferation ratios were calculated to compare the proliferation rate of 188T_FOLR1 and 188T_pLenti (Appendix G).

4.11.2. Cell Migration Assay

The role of FOLR1 in promoting cell migration was determined by monolayer wound healing assay as described by Lampugnani (Lampugnani, 1999). Prior to seeding, a grid was drawn at the bottom of the plate as a guide for making the scratch and capturing images. A total of 4×10^5 cells/ml of 188T_FOLR1 and 188T_pLenti were seeded into 6-well flat bottom tissue culture plates in triplicates. The plates were then incubated at 37°C for 24 hours to allow the cells to attach onto the plate and to form a monolayer. Following that, the media in the wells were removed and replaced with 2 ml DMEM:F12 culture media containing mitomycin C (10 µg/ml) and this was incubated at 37°C for 2 hours to inhibit cell proliferation in order to negate its effect on cell migration. After the 2 hours incubation, the media was removed and a scratch was created using a yellow p200 tip (Thermo Fisher Scientific, Waltham, U.S.A). Cell

debris generated from the scratch was washed off by rinsing twice with 2 ml of sterile PBS pH7.4. The scratch/wound created was observed under 40X objective phase contrast microscope (Olympus, Tokyo, Japan). Images were taken at 0 hour and the PBS in the wells was removed and replaced with 2 ml DMEM: F12 culture media and the cells were further incubated for another 22 hours where images were taken again. The open wound area of each image was analysed using a computer software, TScratch (Geback *et al.*, 2009).

4.11.1. Cell Invasion Assay

The role of FOLR1 in promoting cell invasion was determined by two methods, namely the transwell invasion assay and organotypic co-culture invasion assay.

4.11.1.1. Transwell Invasion Assay

In transwell invasion assay, firstly 3T3 cells were grown in 75cm^2 flask until 70% confluent with DMEM high glucose supplemented with 10% FBS. The media was replaced and the cells were further incubated for additional 2 days at 37°C in order to collect the conditioned media from 3T3 cells. The conditioned media was collected into a sterile universal bottle, centrifuged at 1600 g for 5 minutes to pellet down the debris and the media was decanted into a new universal bottle and stored for further use. The conditioned media was used as a chemoattractant in this assay. Transwell inserts (porous size 8 µm polycarbonate; BD Biosciences, New Jersey, U.S.A) were prepared by placing them into a 24-well companion plate using a sterile forceps and left at the biosafety cabinet for 15-20 minutes in order to stabilize them to room temperature. Once stabilized, the inserts were transferred into the next row of the wells containing serum free DMEM:F12 media where 500 µl of serum free DMEM:F12 media were

pipetted into the interior of inserts and subjected to rehydration for 2 hours at 37°C in 5% CO_2 incubator. During the hydration process, 188T_FOLR1 and 188T_pLenti cells were prepared for seeding at a concentration of 1.25 x 10^5 cells per ml in serum free DMEM:F12 supplemented with 0.l % BSA. Once the rehydration step was completed, media in the inserts were removed and the inserts were transferred into the next row of wells containing 700 µl of 3T3 cells conditioned media. 200 µl of well-resuspended 188T_FOLR1 or 188T_pLenti cells were pipetted into each of the inserts respectively in duplicate. The plate containing the inserts was incubated for 24 hours in 37°C in a 5% CO_2 incubator. After 24 hours, the inserts were removed, transferred into a separate 24-well companion plate containing 500 µl of 1X trypsin/EDTA and incubated for 30 minutes in 37°C in 5% CO_2 incubator to ensure that all the cells that invaded across the matrigel have detached. The detached cells were resuspended in 500µl of DMEM:F12 complete media to stop the trypsin reaction and placed in a tube and flushed again with 500 µl of PBS. The total of 1 ml cell suspension was then added to 9 ml of CASY®ton buffer to conduct a cell count. Invasion index was calculated by dividing the mean of 188T_FOLR1 cells that have invaded across the membrane over the mean of invaded 188T_pLenti cells.

4.11.1.2. Organotypic co-culture Invading Assay

Organotypic co-cultures were established as described by others (Nystrom *et al.*, 2005). 6-well inserts were placed in the recesses of 6-well culture plates (BD Biosciences, New Jersey, U.S.A) and 1.5 ml gel mixture at pH7 (type I collagen, matrigel, 10X DMEM, FBS containing 5x10^5 fibroblast cells; at 4:4:1:1 ratio) was added to the inserts. The gel mixture was left to solidify for 30-60 minutes at 37ºC and media was added to both inner and outer chamber of the inserts. The following day, the medium inside the insert

was aspirated and 5 x 10^5 188T_FOLR1 or 188T_pLenti cells re-suspended in 500 µl DMEM:F12 containing 10% FBS and 0.5 µg/ml hydrocortisone were added to the respective inner chambers. Following a second overnight incubation at 37ºC, all media in the outer and inner chambers were aspirated and 8 ml of culture medium consisting of DMEM:F12 containing 10% FBS, 0.5 µg/ml hydrocortisone and 1X Pen/Strep (100IU Pen/ml; 100µg Strep/ml) was added to the outer chamber of the insert where the bottom of the gel was just in contact with the culture medium, but the top was air exposed. The medium was changed every 2-3 days and the organotypic cultures were harvested at Day 14. At Day 14, the inner chamber containing the gel mixture was fixed in formaldehyde with saline for 24 hours at room temperature, dissected to 3 parts and embedded in a 1% agar solution (Appendix E). Once set, the agar containing the gel was placed into a labeled histology cassette, processed and paraffin-embedded. Sections of 4 µM were cut and stained with Haematoxylin and Eosin as described previously (Lilie, 1965). The invasiveness of 188T_FOLR1 and 188T_pLenti were determined using Image-J (http://rsb.info.nih.gov/ij/). From Image J, the number of particles invading and total area were obtained. The total invading area was then calculated as the number of particles multiplied by the total area with exception of the surface epithelial layer as previously described by (Gaggioli *et al.*, 2007) with slight modification. Invasion index was calculated by comparing the total invading area of 188T_FOLr1 to 188T_pLenti.

4.11.1.3. Statistical Analysis

In all functional assays performed, 3 independent experiments were conducted to confirm the findings. Significant differences observed between 188T_FOLR1 and

188T_pLenti were tested using a student t-test. A *p*-value < 0.05 is considered to be significant.

CHAPTER 5

RESULTS

5.1. Quantity and Quality of RNA extracted from FFPE Specimens

5.1.1. RNA Extraction

A total 116 FFPE blocks were collected for this study that corresponded to 77 tumours [cheek (B) 30, tongue (T) 27, gum (AM) 20] and 39 non-cancerous oral mucosa [cheek (NB) 14, tongue (NT) 8, gum (NG) 17] samples from year 2001-2007. The clinical data associated with these samples are detailed in Appendix A-C. All FFPE samples from tumour and non-cancerous oral mucosal tissues were processed accordingly for RNA extraction as mentioned in section 4.4, 91.4 % of these samples had sufficient RNA (≥ 40 ng/ µl) for further analysis and was subjected to quality control. Ten samples were excluded, as the RNA yield was less than 40 ng/µl. In tumours, the RNA yield ranged from 3.9-623.1 ng/µl with a mean of 201.1 ng/µl from tissue volumes between 0.5-4.5 mm^3 (mean, 2.6 mm^3). As for the non-cancerous oral mucosal tissues, the RNA yield ranged from 37.2-899.0 ng/µl with a mean of 350.5 ng/µl, extracted from tissue volumes between 2.4-35.3 mm^3 (mean, 10.6 mm^3).

5.1.2. RNA Quality Control

The 106 tissues that met the yield requirement (68 tumour and 38 non-cancerous oral mucosal samples) were further examined for quality by amplifying the RPL13A housekeeping gene by qPCR. A high percentage of samples (98%) passed the quality control. All tumours passed the quality control assay with Ct difference ranging from 8-

17 (mean of 13 Ct difference) except for two samples (T19, AM10) with Ct difference of 18 and 19 respectively. All non-cancerous oral mucosal samples passed the quality control with Ct difference ranging from 0-15 (mean of 11 Ct difference)

5.1.3. Factors Associated with the Quality of RNA Extracted from FFPE Specimens

Multiple linear regression was carried out to determine whether the concentration of RNA correlated with tissue volume, age of tissues and/or Ct differences. From the analysis, we found that only tissue volume had a significant association with RNA concentration (*p*-value < 0.001). For every 1 mm^3 increase in tissue volume, the RNA concentration may increase by ~13 ng/µl (Table 5.1). The age of the tissues and Ct difference were not significantly associated with RNA concentration (p > 0.05).

Table 5.1 Factors associated with RNA concentration in FFPE specimens.

Factors	**Parameter estimates**	***p*-value**	**Confidence Intervals (95%)**	
			Lower	**Upper**
Tumour volume (mm3)	13.12	.000	7.71	18.54
Age of samples (months)	-8.84	.406	-2.93	1.19
Ct difference	-.87	.105	-19.55	1.88

5.2. Quality of Microarray Data Generated from RNA Extracted from FFPE Specimens

All tumour samples (68) including those with more than 17 Ct difference (T19, AM10) and all non-cancerous oral mucosal tissues (38) were included in subsequent gene expression profiling experiments using the DASL assay. The number of genes that were significantly expressed (p-value <0.01) were examined for each sample which should reflect the quality of the samples placed on the arrays. Samples with less than 50% of genes with significant intensity were omitted from further analysis. Again a high number of samples (91.5%) had more than 50% genes with significant intensity which indicated that the quality of the RNA was adequate. In the tumours, the percentage of genes with significant intensity ranged from 14.5%- 87.3% with a mean of 74.8%. All tumour samples with exception of 5 samples had more than 50% genes with significant intensity. In the non-cancerous oral mucosal samples, the percentage of genes with significant intensity ranged from 17.5%-89% with a mean of 69.7%. All non-cancerous oral mucosa with exception of 3 samples had more than 50% genes with significant intensity. Surprisingly a lower Ct difference did not correlate with higher number of significant genes detected (Figure 5.1a). Overall, the age of tissues did not appear to be associated with the percentage of genes with significant intensity, however NB10 which was older than 5 years did have a lower number of genes compared to the majority of the more recent samples (Figure 5.1b). Based on the number of significant genes detected, 98 samples (63 tumour and 35 non-cancerous mucosal) were selected for further analysis

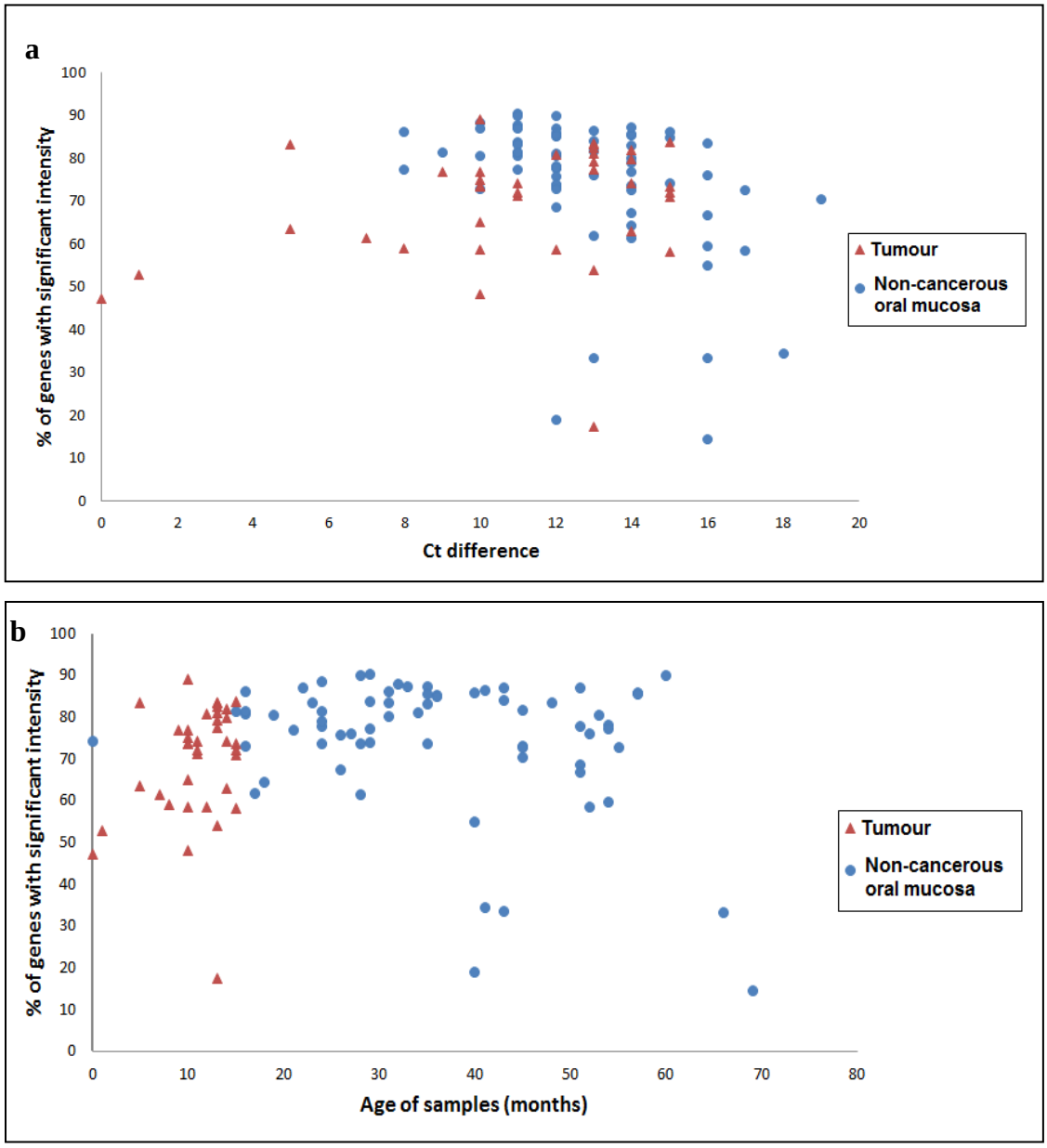


Figure 5.1 Percentage of genes with significant intensity generated from microarray compared to a) Ct difference b) age of FFPE specimens.

5.2.1. Gene Expression Pattern of OSCC and Non-cancerous Oral Mucosa

Using unsupervised hierarchical clustering (HC) and principal component analysis (PCA), the tumour samples and non-cancerous oral mucosal samples were grouped separately thus indicating that tumour and non-cancerous oral mucosa have distinct gene expression. This strongly indicates that despite the use of FFPE samples, the quality of microarray data was maintained and the identification of genes that were differentially expressed between the two groups is highly possible (Figure 5.2). It is interesting to note that one sample meant to be a non-cancerous oral mucosa labeled NT04 clustered among the tumour samples. NT04 was initially diagnosed as moderate dysplasia and this specimen was used in this study, however in 2007, following the diagnosis of dysplasia, this patient subsequently developed OSCC. A total of 14 samples were excluded in the differential analysis based on 2 criteria (1) low number of genes with significant intensity (<50%) of which 8 samples were excluded, and (2) outlier where the samples did not cluster accordingly in the unsupervised gene expression clustering of which 6 samples were excluded. With the exclusions, 92 samples were analysed further to identify differentially expressed genes (62 tumour and 30 non-cancerous oral mucosa).

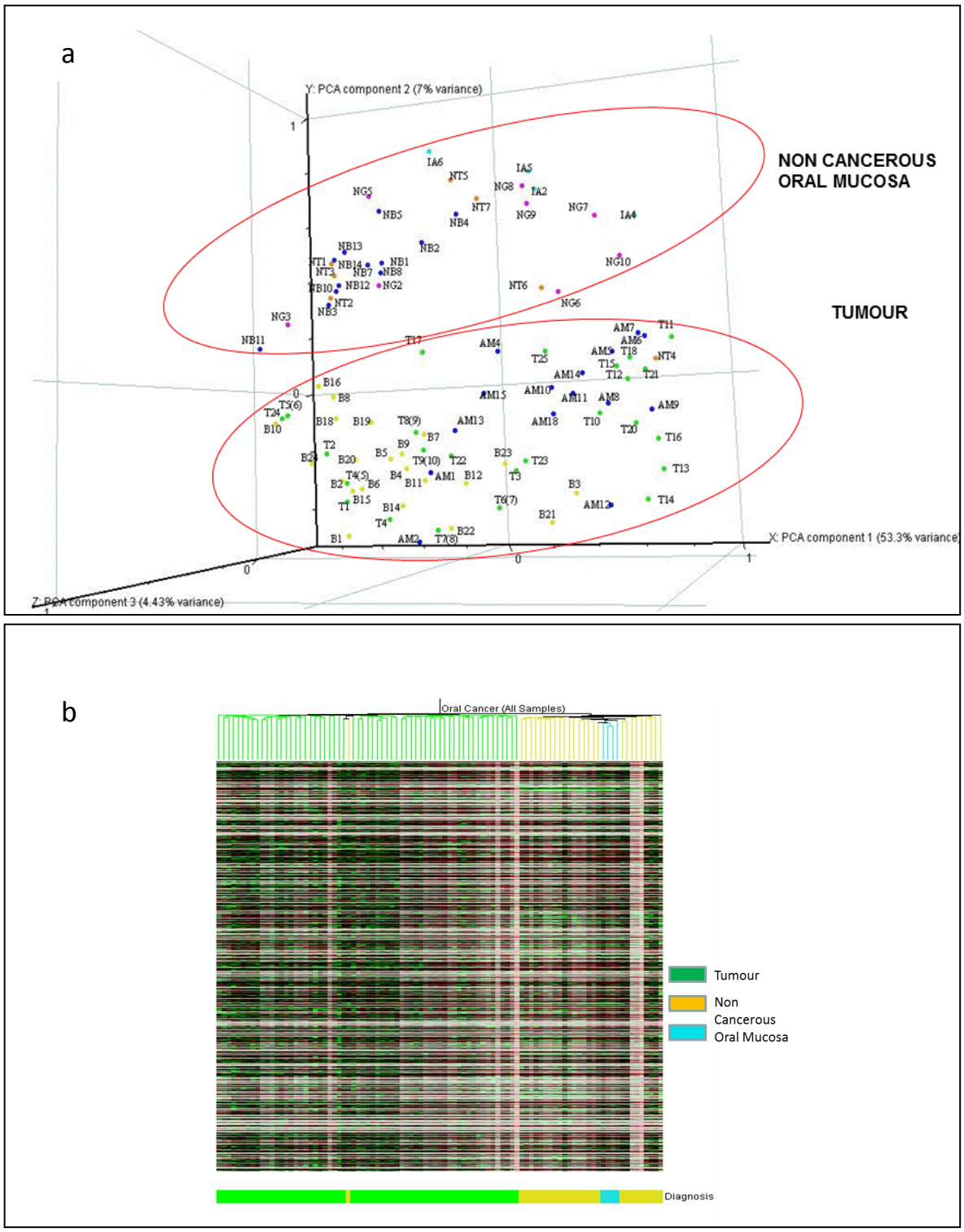


Figure 5.2 Unsupervised clustering of OSCC and non-cancerous oral mucosa through (a)Principal Component Analysis and (b) Hierarchical Clustering.

5.3. Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC

In order to find genes important for OSCC, a differential analysis was carried out to identify differentially expressed genes between tumour and non-cancerous oral mucosa using the method described in section 4.8. A total of 153 genes were found to be differentially expressed between OSCC and non-cancerous oral mucosa at a 1.5 fold change (Table 5.2). Using DAVID, the genes were assigned to the different pathways on the database and they were mainly involved in cytokine-cytokine receptor interaction, MAPK signaling pathway, focal adhesion, cell cycle, regulation of actin-cytoskeleton pathways and p53 signaling pathways amongst others (Table 5.3). When the genes were analysed in terms of Gene Ontology (GO), a consistent result was seen, where the de-regulated genes were mainly categorized as those involve in cell growth and proliferation such as *TGFB1* and *STAT1*, genes involved in collagen degradation and proteolysis such as *MMP1* and *MMP9*, genes involved in cell migration, invasion and regulation of cell adhesion such as *ITGB4* and *LAMB1* as well as genes involved in inflammatory/defense response which includes *IL8*, *IL1B* and *CXCL9*. In addition, anti-apoptotic genes including *BCL2A1* were also found to be de-regulated.

Table 5.2 Genes differentially expressed in OSCC compared to non-cancerous oral mucosa. FC = fold change (OSCC/Non-cancerous oral mucosa)

Gene	FC	Gene	FC	Gene	FC	Gene	FC
MMP1	12.61	*BAK1*	1.89	*MAPK14*	-1.59	DLG3	-2.20
IL8	12.43	*DKC1*	1.88	*BAG1*	-1.59	PPARG	-2.22
MMP9	10.38	*TRAF3*	1.88	*RBBP1*	-1.61	MLLT4	-2.23
SPP1	9.56	*TGFA*	1.86	*MSH3*	-1.61	MYCL1	-2.23
MMP3	9.51	*TGFB1*	1.84	*MADH2*	-1.61	SEMA3F	-2.24
CXCL9	9.42	*EXT2*	1.82	*CDK7*	-1.62	TIAM1	-2.46
MMP10	8.98	*HCK*	1.82	*RAN*	-1.62	AR	-2.82
SERPINE1	6.73	*LTA*	1.80	*LMO2*	-1.62	CEACAM1	-2.85
AIM2	5.90	*MET*	1.80	*BRAF*	-1.62	MXI1	-2.99
IL6	5.85	*PTK2*	1.79	*FGF2*	-1.63	IGFBP5	-3.08
TERT	5.63	*COL4A3*	1.77	*MAPK10*	-1.63	TYRO3	-3.16
MMP7	5.02	*CCND2*	1.75	*CBFA2T1*	-1.64	PNUTL1	-4.02
CSF3R	4.43	*MUC1*	1.74	*NTRK2*	-1.66	PBX1	-4.55
PTHLH	4.38	*ITGB4*	1.74	*SKI*	-1.67	HLF	-7.14
IRF1	4.32	*ETS1*	1.72	*TSG101*	-1.68	ALOX12	-19.19
IL1B	3.82	*CSF1R*	1.70	*NOTCH2*	-1.68		
IL11	3.29	*TK1*	1.68	*CDKN1B*	-1.71		
VAV2	3.13	*CDC2*	1.68	*GRB7*	-1.72		
PTGS2	2.90	*ELL*	1.67	*CEBPA*	-1.72		
PLAUR	2.87	*TFRC*	1.65	*TNFRSF6*	-1.75		
PTPRH	2.85	*OSM*	1.64	*DCN*	-1.76		
ICAM1	2.82	*KAI1*	1.62	*CTNNA1*	-1.76		
LYN	2.81	*MCAM*	1.62	*QARS*	-1.78		
LIF	2.74	*MMP14*	1.62	*MAF*	-1.83		
BCL2A1	2.62	*STAT1*	1.56	*FRZB*	-1.83		
PLA2G2A	2.58	*BCL6*	1.54	*MAD*	-1.86		
ZNFN1A1	2.57	*ENC1*	1.54	*RAD50*	-1.89		
CDKN2A	2.53	*GLI3*	1.53	*WEE1*	-1.89		
LAMB1	2.47	*PRKR*	1.52	*CDKN2C*	-1.89		
TNFRSF1B	2.46	*TNFSF10*	1.51	*XPA*	-1.90		
TRAF4	2.45	*LCN2*	-1.51	*CD34*	-1.93		
CDH11	2.42	*IGF1*	-1.51	*MYB*	-1.93		
BIRC5	2.37	*TFAP2C*	-1.51	*PIM1*	-1.94		
TNFRSF10B	2.35	*FGFR2*	-1.51	*BARD1*	-1.95		
TNFSF8	2.32	*FOSL2*	-1.51	*DLC1*	-1.97		
TNF	2.22	*ARHA*	-1.52	*EPHA1*	-1.98		
TIMP1	2.21	*WRN*	-1.52	*EPO*	-2.00		
E2F3	2.19	*DSP*	-1.52	*ERBB2*	-2.04		
PML	2.17	*IGFBP6*	-1.53	*ERBB3*	-2.05		
CTSL	2.15	*WNT5A*	-1.55	*WNT2B*	-2.07		
NFKB2	2.14	*NEO1*	-1.55	*IL1RN*	-2.08		
BIRC3	2.08	*CD9*	-1.55	*MRE11A*	-2.09		
CDC25B	2.04	*APC*	-1.55	*FGF12*	-2.15		
DTR	2.03	*CCND1*	-1.57	*IGFBP2*	-2.16		
CBLB	1.98	*FER*	-1.58	*FGFR3*	-2.16		
PTK7	1.90	*EPS8*	-1.58	*GAS7*	-2.18		

Table 5.3 Significant de-regulated pathways in OSCC generated by comparing the T vs. N list to the pathways database in DAVID.

No	KEGG Pathway (T vs. N)	Count	%	*p*-value	Genes
1	Cytokine-cytokine receptor interaction	18	1.02	4.42E-06	*IL6, TNF, IL8, MET, CXCL9, TGFB1, IL11, TNFSF8, LIF, OSM, TNFSF10, TNFRSF1B, TNFRSF10B, CSF3R, IL1B, LTA, CSF1R, EPO*
2	MAPK signaling pathway	14	0.79	1.13E-03	*FGFR2, FGFR3, TNF, BRAF, NFKB2, MAPK10, FGF12, TGFB1, CDC25B, MAPK14, NTRK2, PLA2G2A, IL1B, FGF2*
3	Focal adhesion	13	0.73	2.88E-04	*BRAF, ERBB2, MET, ITGB4, IGF1, MAPK10, VAV2, BIRC3, PTK2, CCND1, CCND2, LAMB1, SPP1*
4	Jak-STAT signaling pathway	11	0.62	5.18E-04	*OSM, LIF, CCND1, CBLB, IL6, CCND2, PIM1, CSF3R, STAT1, IL11, EPO*
5	Hematopoietic cell lineage	10	0.56	2.44E-05	*CD9, IL6, TNF, TFRC, CD34, IL1B, CSF3R, IL11, CSF1R, EPO*
6	Toll-like receptor signaling pathway	10	0.56	8.83E-05	*IL6, TNF, IL8, MAPK14, CXCL9, IL1B, MAPK10, STAT1, TRAF3, SPP1*
7	Cell cycle	10	0.56	4.49E-04	*E2F3, CCND1, CDKN1B, CDKN2A, CCND2, CDKN2C, CDK7, WEE1, TGFB1, CDC25B*
8	Regulation of actin cytoskeleton	10	0.56	1.72E-02	*FGFR2, PTK2, FGFR3, BRAF, TIAM1, ITGB4, FGF12, VAV2, FGF2, APC*
9	Chemokine signaling pathway	9	0.51	2.16E-02	*PTK2, IL8, BRAF, LYN, TIAM1, HCK, CXCL9, VAV2, STAT1*
10	ErbB signaling pathway	8	0.45	1.03E-03	*CBLB, PTK2, CDKN1B, BRAF, ERBB3, ERBB2, TGFA, MAPK10*
11	NOD-like receptor signaling pathway	7	0.40	9.08E-04	*IL6, TNF, IL8, MAPK14, IL1B, MAPK10, BIRC3*
12	Leukocyte transendothelial migration	7	0.40	2.13E-02	*ICAM1, PTK2, MAPK14, MMP9, VAV2, CTNNA1, MLLT4*
13	p53 signaling pathway	6	0.34	8.08E-03	*CCND1, CDKN2A, TNFRSF10B, CCND2, SERPINE1, IGF1*
14	Fc epsilon RI signaling pathway	6	0.34	1.42E-02	*TNF, LYN, MAPK14, PLA2G2A, MAPK10, VAV2*
15	Epithelial cell signaling in Helicobacter pylori infection	5	0.28	3.63E-02	*IL8, LYN, MAPK14, MET, MAPK10*
16	RIG-I-like receptor signaling pathway	5	0.28	4.16E-02	*TNF, IL8, MAPK14, MAPK10, TRAF3*

5.3.1. Validation of the Expression of Genes Implicated in OSCC

A total of 17 differentially expressed genes were selected for validation based on 1.5 fold change cut off value and literature supporting its role in OSCC. Using qPCR, 76 % (13/17) of the genes were validated in more than 50% of the independent samples indicating the reliability of the microarray data and confirming that these differentially expressed genes were indeed changed in oral cancer (Figure 5.3; Table 5.4). Three genes, *BCL2A1*, *ITGB4* and *MMP1* were further validated at the protein level using IHC and were found to be significantly over-expressed in oral cancer in comparison to non-cancerous oral mucosa (Figure 5.4; Table 5.5, Appendix B and C).

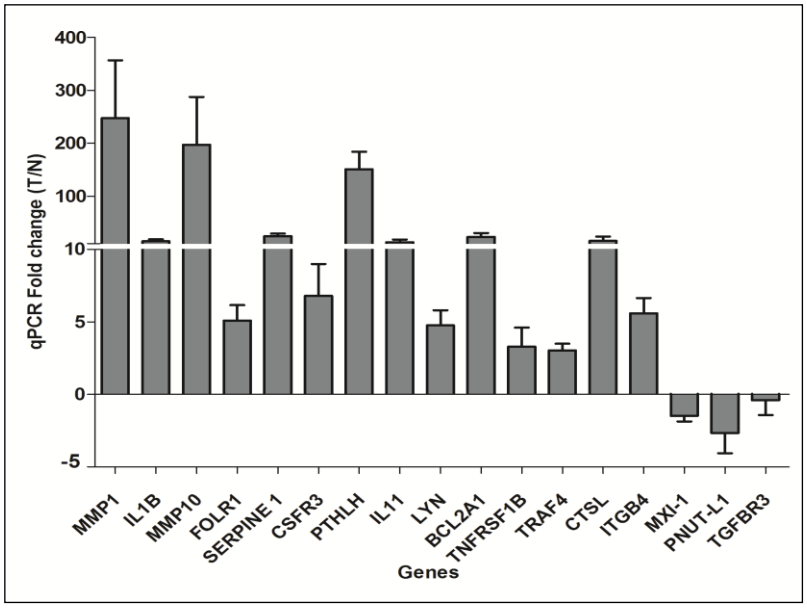


Figure 5.3 Differentially expressed genes validated through qPCR using independent samples of fresh frozen OSCC. Data shown here is an average of fold change of a minimum of 23 OSCC fresh frozen samples as compared to the non-cancerous oral mucosa.

Table 5.4 Differentially expressed genes validated at mRNA level through qPCR. Samples are considered validated when the expression pattern is similar to microarray using cut off value of 2.0 fold change.

No	Genes	Microarray expression pattern	Fold change generated from microarray	Number of samples validated in qPCR	% of samples validated in qPCR
1	*MMP1*	up-regulated	12.61	29/35	83.0
2	*IL1B*	up-regulated	12.43	26/29	90.0
3	*MMP10*	up-regulated	8.98	26/35	74.3
4	*FOLR1*	up-regulated	8.24	21/33	64.6
5	*SERPINE 1*	up-regulated	6.73	27/33	81.8
6	*CSFR3*	up-regulated	4.43	22/31	71.0
7	*PTHLH*	up-regulated	4.38	30/34	88.2
8	*IL11*	up-regulated	3.29	17/30	56.7
9	*LYN*	up-regulated	2.81	18/30	60.0
10	*BCL2A1*	up-regulated	2.62	24/24	77.4
11	*TNFRSF1B*	up-regulated	2.46	11/23	47.8
12	*TRAF4*	up-regulated	2.45	16/25	64.0
13	*CTSL*	up-regulated	2.15	29/30	96.7
14	*ITGB4*	up-regulated	1.74	27/34	79.4
15	*MXI-1*	down-regulated	-2.99	13/30	43.3
16	*PNUT-L1*	down-regulated	-4.02	10/23	43.5
17	*TGFBR3*	down-regulated	-5.15	15/35	42.9

Table 5.5 IHC scoring for BCL2A1, ITGB4 and MMP1 in OSCC and non-cancerous oral mucosa.

		Cases	Positive Staining (%)	Negative Staining (%)	*p*-value
BCL2A1					
	Cancer	31	19 (61%)	12 (39%)	< 0.001
	Non-cancerous oral mucosa	20	0 (0%)	20 (100%)	
ITGB4					
	Cancer	49	33(67%)	16 (33%)	< 0.001
	Non-cancerous oral mucosa	26	5(19%)	21 (81%)	
MMP1					
	Cancer	51	41 (80%)	10 (20%)	< 0.001
	Non-cancerous oral mucosa	26	1 (4%)	25 (96%)	

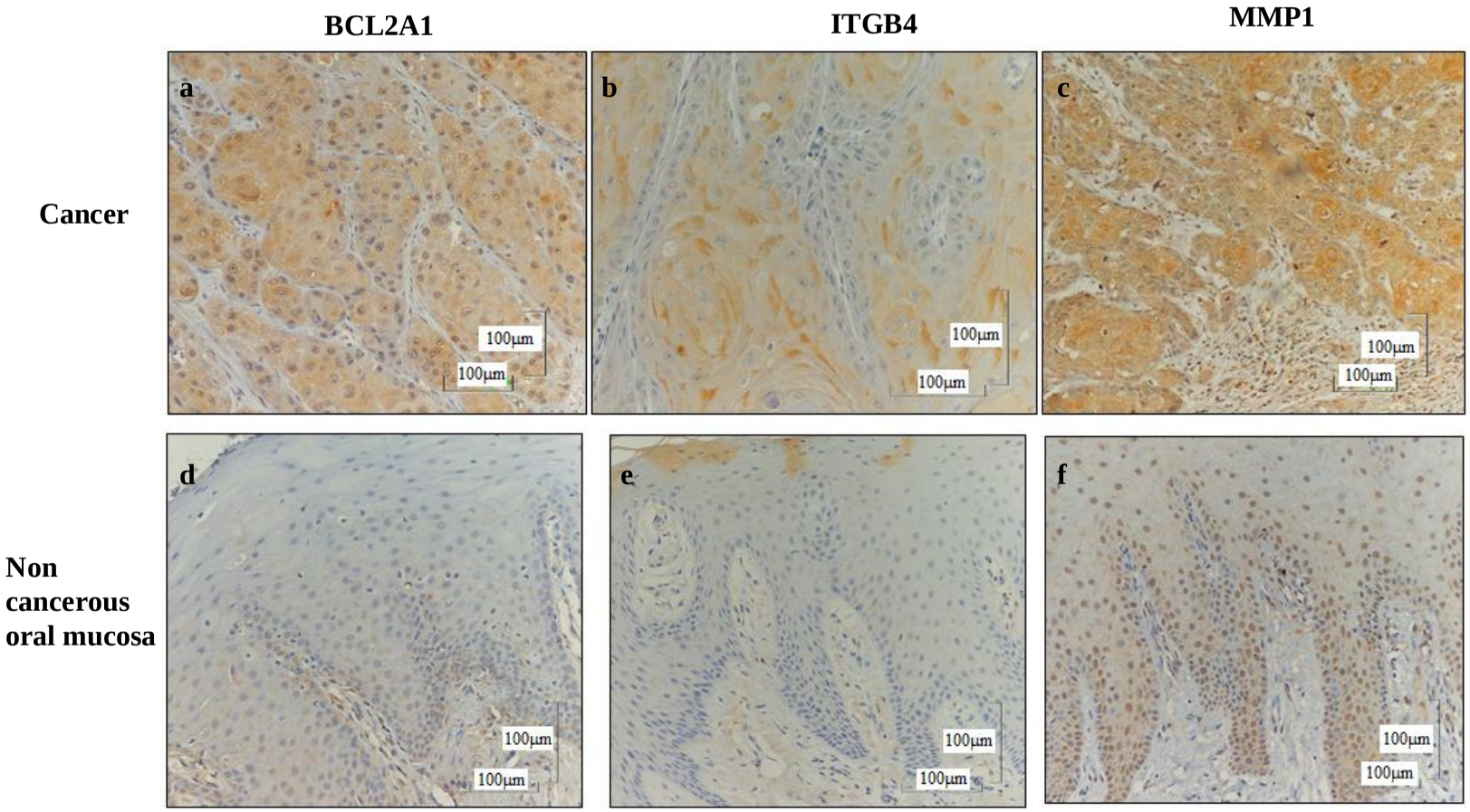


Figure 5.4 Representative images of OSCC and non-cancerous oral mucosa immunohistochemically stained with BCL2A1, ITGB4 and MMP1. OSCC (a, b, c) and non-cancerous oral mucosa (d,e,f) demonstrating the over-expression of these proteins in OSCC and low expression in non-cancerous oral mucosa. Original magnification 200X (a,-,f).

5.4. Analysis of Gene Expression Variation in OSCC from the Different Sites of the Oral Cavity

Considering the tumour cluster in the PCA, it was observed that all OSCC from the cheek clustered closely together within the yellow boundary and OSCC from the Gum grouped together within the blue boundary suggesting gene expression profile of different sites were distinct (Figure 5.5). It is interesting to note that OSCC from the tongue did not cluster together like the cheek and gum but was distributed randomly within the tumour boundary. This suggests that within the group of tongue OSCC, high degree heterogeneity in the gene expression existed.

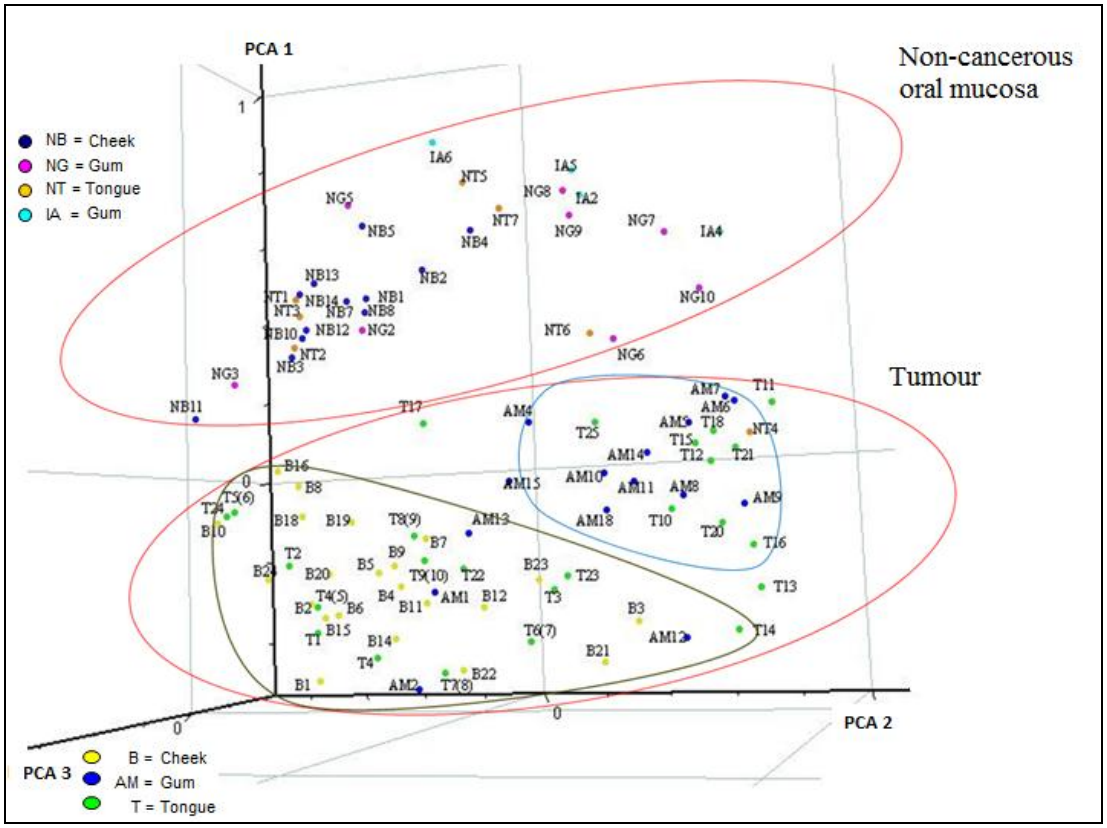


Figure 5.5 PCA demonstrating the clustering pattern of OSCC from different sites of the oral cavity (Cheek - yellow border, gum -blue border and tongue- green dots). IA = gingival tissue obtained from surgically removal of impacted wisdom tooth.

To analyse this further, 51 genes were chosen at random (Appendix D) and the standard deviation in the gene expression was calculated based on the microarray intensity data for OSCC of the cheek, gum and tongue. The higher the standard deviation within the expression of a gene indicates that the expression of this particular gene is heterogeneous in the samples. In Table 5.6, OSCC with the highest standard deviation was highlighted in bold and this demonstrated that OSCC from the tongue had the most number of genes with the highest standard deviation at 58.8% (30/51) followed by OSCC from gum with 37.3% (19/51) whereas OSCC of the cheek had the lowest percentage of genes with the highest standard deviation at 2.9% (2/51) (Figure 5.6). This data further supported that tongue OSCC has the highest variation across the 51 genes chosen at random when compared to other sites of OSCC thus suggesting heterogeneity in the gene expression within OSCC from the tongue.

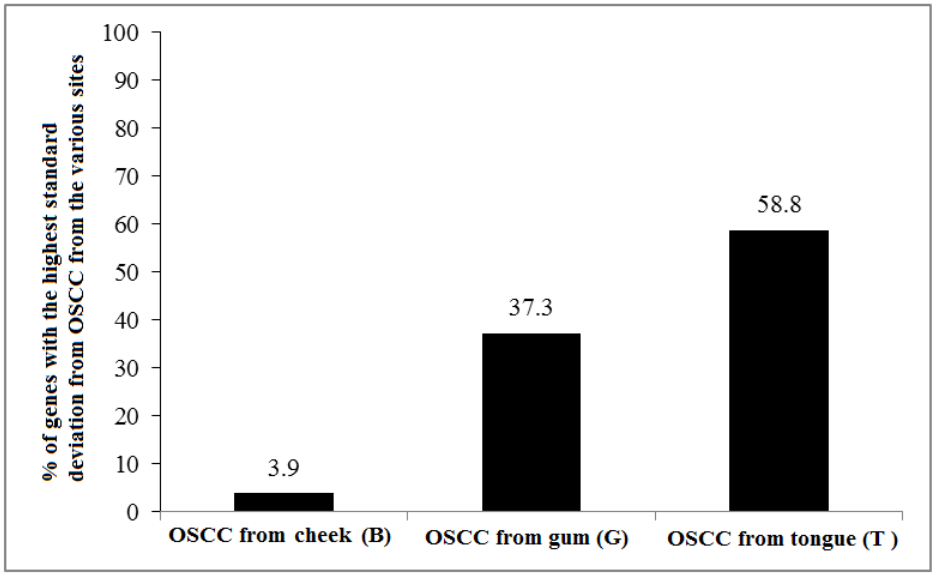


Figure 5.6: The percentage of OSCC from the different sites having the highest standard deviation across the 51 gene chosen at random.

Table 5.6 Standard deviation of 51 genes randomly chosen for OSCC from cheek (B), gum (G) and tongue (T) were calculated using SPSS 16. The highest standard deviation were higlighted in bold.

Gene	Site	Std dev	Gene	Site	Std dev	Gene	Site	Std dev	Gene	Site	Std dev	Gene	Site	Std dev	Gene	Site	Std dev
ABCB1	B	7.20E+02	*CSF2*	B	2.74E+03	*FOSL2*	B	1.84E+03	*MCAM*	B	4.12E+03	*PTHLH*	B	4.44E+03	*TNFRSF6*	B	2.00E+03
	T	8.90E+02		**T**	**4.58E+03**		**T**	**3.90E+03**		**T**	**4.93E+03**		**T**	**6.52E+03**		T	2.91E+03
	G	**1.02E+03**		G	3.18E+03		G	2.78E+03		G	3.88E+03		G	4.44E+03		**G**	**3.79E+03**
ALOX12	B	1.79E+03	*CUL2*	B	1.81E+03	*GFI1*	B	3.23E+02	*MLF2*	B	1.98E+03	*RAD50*	B	1.89E+02	*TRAF4*	B	2.10E+03
	T	4.39E+03		**T**	**2.82E+03**		**T**	**7.63E+02**		**T**	**3.79E+03**		T	4.96E+02		**T**	**3.33E+03**
	G	**4.47E+03**		G	1.83E+03		G	7.60E+02		G	2.38E+03		**G**	**6.60E+02**		G	2.73E+03
ARNT	B	2.04E+03	*DDX6*	B	2.17E+03	*HDAC1*	B	3.27E+03	*MMP7*	B	1.84E+03	*RAP2A*	B	1.59E+03	*VEGFB*	B	2.44E+03
	T	**2.70E+03**		T	2.48E+03		**T**	**3.88E+03**		T	3.09E+03		T	2.28E+03		**T**	**4.19E+03**
	G	1.58E+03		**G**	**2.99E+03**		G	2.88E+03		**G**	**4.82E+03**		**G**	**3.05E+03**		G	2.81E+03
BCL3	B	3.80E+03	*E2F1*	B	2.28E+03	*IFNGR2*	B	2.26E+03	*MTA1*	B	2.02E+03	*RECQL*	B	1.03E+03	*WRN*	B	2.65E+03
	T	**4.93E+03**		**T**	**4.05E+03**		**T**	**4.07E+03**		**T**	**4.06E+03**		**T**	**2.05E+03**		**T**	**3.01E+03**
	G	4.22E+03		G	2.77E+03		G	3.02E+03		G	2.26E+03		G	1.17E+03		G	2.63E+03
BRCA2	B	1.55E+03	*ENC1*	B	2.94E+03	*IL11*	B	5.61E+02	*NAT2*	B	1.89E+02	*SERPINE1*	B	4.04E+03	*YES1*	B	2.17E+03
	T	1.78E+03		**T**	**4.37E+03**		T	8.91E+02		G	4.96E+02		T	5.00E+03		**T**	**2.76E+03**
	G	**2.75E+03**		G	3.71E+03		**G**	**2.03E+03**		**T**	**6.60E+02**		**G**	**5.50E+03**		G	2.24E+03
CCNA2	B	3.42E+03	*ERCC2*	B	5.36E+02	*IL6*	B	2.29E+03	*NOTCH2*	B	1.59E+03	*SPP1*	B	1.85E+03	*ZNFN1A1*	B	3.71E+03
	T	**4.03E+03**		T	8.73E+02		**T**	**3.02E+03**		G	2.28E+03		T	3.60E+03			**4.80E+03**
	G	3.68E+03		**G**	**9.36E+02**		G	2.76E+03		**T**	**3.05E+03**		**G**	**7.14E+03**			3.31E+03
CD9	B	2.50E+03	*ETV6*	B	1.67E+03	*JUNB*	B	2.06E+03	*PBX1*	B	1.35E+03	*TEK*	B	1.30E+03			
	T	**5.92E+03**		**T**	**3.54E+03**		**T**	**5.13E+03**		T	2.26E+03		**T**	**1.84E+03**			
	G	4.17E+03		G	3.03E+03		G	2.48E+03		**G**	**2.98E+03**		G	9.14E+02			
CDK6	B	2.36E+03	*FGF12*	B	1.61E+03	*LCN2*	B	3.52E+03	*PIK3CA*	B	1.06E+03	*TGFB2*	B	1.15E+03			
	T	**4.53E+03**		**T**	**2.25E+03**		T	4.07E+03		T	1.04E+03		T	1.30E+03			
	G	3.54E+03		G	1.95E+03		**G**	**4.31E+03**		**G**	**1.87E+03**		**G**	**1.49E+03**			
CEBPA	B	2.59E+03	*FGFR3*	B	2.52E+03	*MADH2*	**B**	**2.55E+03**	*PPARG*	B	8.37E+02	*TIMP2*	B	1.65E+03			
	T	**3.24E+03**		T	3.92E+03		T	1.94E+03		T	7.12E+02		**T**	**3.91E+03**			
	G	3.82E+03		**G**	**4.82E+03**		G	2.34E+03		G	8.36E+02		G	3.24E+03			

5.5. Identification of Differentially Expressed Pathways and Genes Important in OSCC from Distinct Sites

Differentially expressed genes between OSCC and non-cancerous oral mucosal specimens specific to B (B vs. NB), gum (AM vs. NG) and tongue (T vs. NT) were generated as explained in section 4.8.2. In order to identify common genes and pathways amongst the different sites, Venn diagram analysis was utilized. The number of genes in the various sub-groups is shown in Figure 5.7.

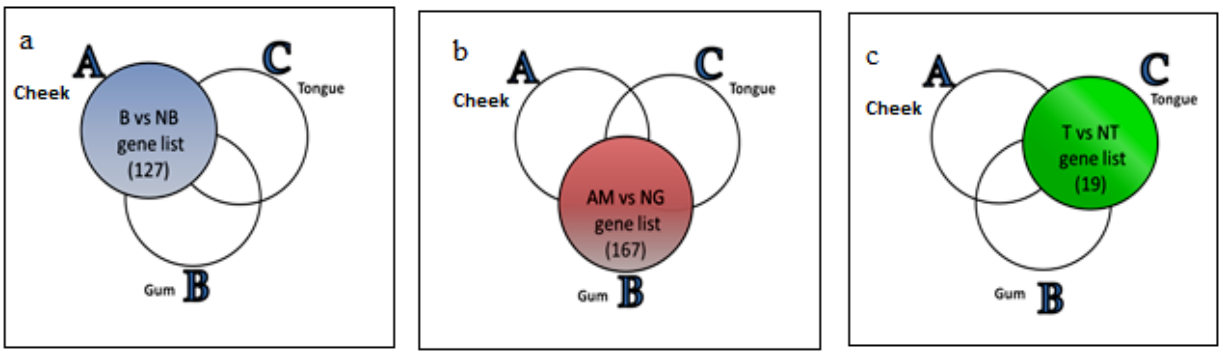


Figure 5.7 Using Venn Diagram to compare genes and pathways among the different sites of OSCC. Each gene list was generated by comparing OSCC from (a) cheek, (b) gum and (c) tongue to the non-cancerous oral mucosa from the respective sites.

In B vs. NB gene list, a total of 127 genes were differentially expressed (69 up-regulated and 58 down-regulated; Table 5.7). Among the significantly de-regulated pathways were cytokine-cytokine receptor interaction, JAK-STAT and focal adhesion (Table 5.8). As for AM vs. NG gene list, a total of 167 genes were found to be differentially expressed (55 up-regulated, 112 down-regulated; Table 5.9)

Table 5.7 B vs. NB gene list. Differentially expressed genes in cheek OSCC compared to non-cancerous oral mucosal samples from the cheek.

Gene	FC	Gene	FC	Gene	FC	Gene	FC
CXCL9	34.49	*OSM*	3.21	*ETV6*	1.52	*CEBPA*	-2.27
IL8	19.41	*LTA*	3.21	*TGFBR2*	-1.51	*ERBB2*	-2.28
MMP3	18.63	*LIF*	3.19	*MAF*	-1.52	*AKT2*	-2.3
SPP1	16.36	*BIRC3*	3.16	*EGFR*	-1.54	*EPHA1*	-2.3
MMP10	14.58	*ICAM1*	3.13	*APC*	-1.55	*MLLT4*	-2.39
MMP9	14.41	*ING1*	2.95	*CDKN1B*	-1.56	*FGF2*	-2.41
MMP1	14.37	*GFI1*	2.81	*BRAF*	-1.57	*ERBB3*	-2.42
AIM2	12.93	*FGR*	2.7	*FER*	-1.57	*NQO1*	-2.51
IL6	9.49	*PLAUR*	2.67	*TFG*	-1.58	*DLG3*	-2.65
CSF3	8.93	*ENC1*	2.67	*IGFBP6*	-1.59	*TIAM1*	-2.89
PTHLH	8.38	*MATK*	2.51	*WEE1*	-1.6	*IGFBP2*	-2.91
MMP7	7.6	*BIRC5*	2.49	*MRE11A*	-1.61	*CD34*	-2.92
CSF2	7.56	*TNFSF10*	2.49	*CDK7*	-1.61	*EPO*	-2.94
SERPINE1	6.92	*PRKR*	2.45	*CD9*	-1.62	*GAS7*	-3.05
IL1B	6.37	*LIG4*	2.44	*NOTCH2*	-1.63	*WNT2B*	-3.13
TERT	5.33	*VEGF*	2.35	*CTNNA1*	-1.64	*MXI1*	-3.5
BCL2A1	5.23	*STAT1*	2.2	*CBFA2T1*	-1.72	*AR*	-3.65
IFNG	5.12	*MYBL2*	2.2	*DDB2*	-1.73	*CEACAM1*	-3.89
HOXA9	5.09	*SH3BP2*	2.17	*CCND1*	-1.73	*PBX1*	-4.08
IRF1	4.75	*BAK1*	2.16	*RBBP1*	-1.77	*IGFBP5*	-4.22
HMMR	4.72	*MCAM*	2.13	*DCN*	-1.8	*TYRO3*	-4.41
RARB	4.52	*PTK7*	2.05	*FOSL2*	-1.8	*TGFBR3*	-5.15
VAV2	4.45	*MET*	1.97	*MAD*	-1.82	*PNUTL1*	-7.41
IL11	4.43	*GLI3*	1.97	*XPA*	-1.82	*HLF*	-9.8
LYN	4.31	*CCNA2*	1.94	*IL1RN*	-1.83	*ALOX12*	-23.81
ARHH	4.19	*RET*	1.94	*QARS*	-1.88		
CTSL	4.07	*GADD45A*	1.89	*SEMA3F*	-1.92		
PDGFRA	4.04	*BTK*	1.88	*JUND*	-1.92		
CSF3R	3.99	*ITGB4*	1.82	*MYB*	-1.93		
CDC2	3.59	*CDC25B*	1.78	*FGFR3*	-1.96		
CDKN2A	3.51	*LCK*	1.72	*PIM1*	-2		
LAMB1	3.48	*IFNGR1*	1.61	*TIMP2*	-2.02		
CDH11	3.27	*PML*	1.59	*EVI1*	-2.24		
EXT2	3.22	*IL2*	1.54	*EPS8*	-2.26		

Table 5.8 De-regulated pathways in cheek OSCC generated by comparing the B vs. NB list to the pathways database in DAVID.

No	KEGG Pathway (BT vs. BN)	Count	%	*p*-value	Genes
1	Cytokine-cytokine receptor interaction	19	16.24	5.50E-07	*IL2, IL1B, IL11, IFNGR1, CSF3R, CXCL9, LTA, TNFSF10, CSF3, IFNG TGFBR2, IL8, OSM, PDGFRA, EGFR, CSF2, MET, LIF, EPO,*
2	Jak-STAT signaling pathway	14	11.97	3.04E-06	*AKT2, IL2, PIM1, IL11, IFNGR1, CSF3R, CSF3, STAT1, IFNG, OSM, CSF2, EPO, LIF, CCND1*
3	Focal adhesion	12	10.26	9.49E-04	*VAV2, LAMB1, ITGB4, AKT2, ERBB2, BRAF, BIRC3, PDGFRA, SPP1, EGFR, MET, CCND1*
4	MAPK signaling pathway	11	9.40	2.00E-02	*GADD45A, AKT2, TGFBR2, BRAF, JUND, IL1B, PDGFRA, FGF2, EGFR, FGFR3, CDC25B*
5	Cell cycle	9	7.69	1.08E-03	*GADD45A, CDC2, CDKN2A, CDK7, WEE1, CDKN1B, CCND1, CCNA2, CDC25B*
6	Natural killer cell mediated cytotoxicity	9	7.69	2.40E-03	*LCK, VAV2, ICAM1, IFNG, BRAF, IFNGR1, SH3BP2, TNFSF10, CSF2*
7	Regulation of actin cytoskeleton	9	7.69	4.27E-02	*VAV2, ITGB4, APC, BRAF, PDGFRA, FGF2, EGFR, TIAM1, FGFR3*
8	Hematopoietic cell lineage	8	6.84	8.79E-04	*CSF3, CD34, IL1B, IL11, CSF3R, CD9, CSF2, EPO*
9	Adherens junction	7	5.98	2.26E-03	*TGFBR2, ERBB2, EGFR, CTNNA1, MLLT4, FER, MET*
10	p53 signaling pathway	6	5.13	8.02E-03	*SERPINE1, GADD45A, CDC2, CDKN2A, DDB2, CCND1*
11	ErbB signaling pathway	6	5.13	1.98E-02	*AKT2, ERBB2, BRAF, CDKN1B, EGFR, ERBB3*
12	T cell receptor signaling pathway	6	5.13	2.92E-02	*LCK, VAV2, IL2, IFNG, AKT2, CSF2*
13	Toll-like receptor signaling pathway	6	5.13	3.95E-02	*AKT2, STAT1, IL8, IL1B, SPP1, CXCL9*
14	Dorso-ventral axis formation	5	4.27	1.59E-03	*ERBB2, BRAF, ETV6, EGFR, NOTCH2*

Table 5.9 AM vs. NG gene list. Differentially expressed genes in gum OSCC compared to non-cancerous oral mucosal samples from gum.

Gene	FC	Gene	FC	Gene	FC	Gene	FC
SPP1	12.79	*ETS1*	1.97	*NF1*	-1.62	MSH3	-1.98
MMP1	8.35	*PTK2*	1.94	*RLF*	-1.62	TYRO3	-2.03
MMP9	7.1	*CTSL*	1.93	*WEE1*	-1.62	FER	-2.07
SERPINE1	6.26	*CDK2*	1.92	*RAD50*	-1.62	CDKN2B	-2.08
IL8	6.12	*PTK7*	1.82	*CDK7*	-1.63	MAD	-2.12
MMP3	5.46	*FYN*	1.82	*CD9*	-1.63	ERCC5	-2.13
TERT	5.39	*PML*	1.79	*CDKN1B*	-1.63	CTNNA1	-2.14
MMP7	4.54	*CCND3*	1.78	*CDH1*	-1.64	FLT3	-2.18
WNT10B	4.34	*ICAM1*	1.76	*IGFBP6*	-1.64	MLLT4	-2.19
HOXA9	3.99	*TFRC*	1.67	*LMO2*	-1.65	ARHI	-2.22
IL6	3.96	*FZD7*	1.64	*ING1*	-1.66	MYCL1	-2.28
CSF3R	3.79	*PDGFB*	1.59	*FGF2*	-1.67	FGFR3	-2.31
IRF1	3.68	*SPI1*	1.59	*FGF9*	-1.67	RAD52	-2.32
LIF	3.33	*MDM4*	-1.50	*VEGFB*	-1.68	PIM1	-2.34
MMP10	3.19	*MCC*	-1.50	*WNT2B*	-1.69	PPARG	-2.36
CXCL9	3.06	*PLAT*	-1.50	*MLL*	-1.69	MAF	-2.38
E2F3	2.95	*BRAF*	-1.51	*SOD1*	-1.69	PNUTL1	-2.38
CDK4	2.93	*TSC2*	-1.51	*RAP2A*	-1.70	IL1RN	-2.39
MYBL2	2.71	*TP53*	-1.51	*RBBP6*	-1.70	TSG101	-2.4
ARHH	2.71	*PDGFRB*	-1.51	*ARHA*	-1.71	MRE11A	-2.41
TIMP1	2.69	*CD59*	-1.52	*DLG3*	-1.72	IGF1	-2.43
TNFRSF1B	2.68	*RBBP1*	-1.52	*EPS15*	-1.72	XPA	-2.43
TGFB1	2.57	*ZNFN1A1*	-1.52	*S100A4*	-1.73	FGF7	-2.46
IL11	2.54	*WNT5A*	-1.52	*NEO1*	-1.74	DLC1	-2.48
TNFSF8	2.53	*VIL2*	-1.52	*MALT1*	-1.74	ERBB3	-2.51
RECQL	2.51	*EGR1*	-1.53	*APC*	-1.75	IGFBP2	-2.58
AIM2	2.42	*SRC*	-1.54	*RBL2*	-1.75	MAPK10	-2.6
PLAUR	2.35	*CCNH*	-1.55	*IL1A*	-1.76	FRZB	-2.62
BIRC5	2.3	*MADH4*	-1.55	*TOP1*	-1.76	MXI1	-2.72
HCK	2.29	*DCN*	-1.56	*YY1*	-1.77	FGF12	-2.75
PTPRH	2.21	*HDAC1*	-1.56	*OGG1*	-1.78	FHIT	-2.78
E2F1	2.18	*ETS2*	-1.57	*TNFRSF10A*	-1.82	SEMA3F	-2.8
CDC2	2.15	*GRB7*	-1.57	*DDX6*	-1.82	LAF4	-2.89
PTHLH	2.14	*ERCC4*	-1.58	*IGFBP5*	-1.83	BMP4	-3
LYN	2.11	*SPARC*	-1.58	*EPS8*	-1.84	BARD1	-3.08
BAK1	2.09	*NOTCH1*	-1.58	*MYCN*	-1.84	NGFR	-3.3
CDH11	2.09	*RB1*	-1.58	*DSP*	-1.87	TIAM1	-3.73
CDC25B	2.05	*BCL3*	-1.58	*MAPK14*	-1.87	PBX1	-4.29
FANCG	2.04	*FOSL2*	-1.59	*EPHA1*	-1.89	TGFBR3	-4.69
LAMB1	2.01	*LCN2*	-1.59	*BAG1*	-1.89	HLF	-7.09
CSF1R	1.98	*TPR*	-1.59	*ERBB2*	-1.91	ALOX12	-19.38
PCNA	1.97	*SMARCA4*	-1.6	*GAS7*	-1.91		

Table 5.10 De-regulated pathways in gum OSCC, generated by comparing AM vs. NG list to the pathways database in DAVID

	KEGG Pathway (GT vs. GN)	Count	%	***p*-value**	Genes
1	Cell cycle	25	11.6	3.1E-14	*TFDP1, CDC2, E2F3, CREBBP, TGFB3, GADD45A, E2F1, TGFB1, CDK2, CDKN1A, CDK7, CDKN2B, TP53, WEE1, CDKN1B, HDAC1, CCNH, RB1, RBL2, CCND3, PCNA, CDK6, ATM, CDC25B, CDK4,*
2	MAPK signaling pathway	24	11.2	4.2E-06	*AKT2, TGFB1, BRAF, IL1A, MAPK10, TP53, FGF12, PDGFRB, FGF7, NF1, TGFB3, GADD45A, JUND, PDGFB, PDGFRA, FGF2, EGFR, TNFRSF1A, RAF1, MAPK14, FGF9, CRKL, FGFR3, CDC25B,*
3	Cytokine-cytokine receptor interaction	23	10.7	1.2E-05	*TGFB1, IL12B, TNFSF8, IL1A, IL11, CSF3R, CXCL9, PDGFRB, VEGFB, CSF1R, TGFB3, FLT3, IL8, NGFR, PDGFB, PDGFRA, IFNGR2, TNFRSF1B, EGFR, TNFRSF1A, IL4, LIF, TNFRSF10A,*
4	Focal adhesion	20	9.3	1.1E-05	*LAMB1, AKT2, CTNNB1, BRAF, MAPK10, PTEN, IGF1, SRC, PTK2, PDGFRB, VEGFB, CCND3, ERBB2, PDGFB, PDGFRA, FYN, SPP1, EGFR, RAF1, CRKL,*
5	p53 signaling pathway	17	7.9	1.4E-10	*CDK2, CDKN1A, PTEN, CDC2, IGF1, TP53, MDM4, DDB2, CASP8, GADD45A, SERPINE1, CCND3, TSC2, APAF1, CDK6, ATM, CDK4,*
6	Regulation of actin cytoskeleton	15	7.0	7.0E-03	*BRAF, FGF12, PTK2, TIAM1, FGF7, PDGFRB, APC, PDGFB, PDGFRA, FGF2, RAF1, EGFR, FGF9, CRKL, FGFR3,*
7	ErbB signaling pathway	13	6.0	1.1E-05	*AKT2, BRAF, MAPK10, CDKN1A, FRAP1, SRC, CDKN1B, PTK2, ERBB2, RAF1, EGFR, ERBB3, CRKL,*
8	Jak-STAT signaling pathway	13	6.0	2.8E-03	*AKT2, IL12B, PIM1, IL11, CSF3R, JAK2, CREBBP, STAT3, STAT1, CCND3, IFNGR2, IL4, LIF,*

Table 5.10, continued

	KEGG Pathway (GT vs. GN)	Count	%	***p*-value**	Genes
9	Adherens junction	12	5.6	1.6E-05	*PTPRF, CTNNB1, ERBB2, FYN, SRC, CDH1, EGFR, CTNNA1, MLLT4, FER, TCF7L2, CREBBP,*
10	Wnt signaling pathway	12	5.6	7.6E-03	*APC, CTNNB1, CCND3, MAPK10, TP53, WNT2B, WNT10B, MMP7, FZD7, TCF7L2, WNT5A, CREBBP,*
11	Apoptosis	11	5.1	2.6E-04	*CASP8, AKT2, IL1A, APAF1, RIPK1, TP53, TNFRSF1A, NFKBIA, ATM, TNFRSF10A, PRKAR1A,*
12	Hematopoietic cell lineage	11	5.1	2.9E-04	*CD34, FLT3, CD59, IL1A, IL11, CSF3R, CD9, IL4, TFRC, CD44, CSF1R,*
13	Toll-like receptor signaling pathway	11	5.1	1.3E-03	*CASP8, AKT2, STAT1, IL12B, MAPK10, IL8, RIPK1, SPP1, CXCL9, NFKBIA, MAPK14,*
14	Dorso-ventral axis formation	8	3.7	1.8E-05	*ERBB2, BRAF, ETS2, EGFR, NOTCH1, RAF1, ETS1, NOTCH2,*
15	Adipocytokine signaling pathway	8	3.7	7.7E-03	*STAT3, AKT2, MAPK10, FRAP1, TNFRSF1B, TNFRSF1A, NFKBIA, JAK2,*
16	TGF-beta signaling pathway	8	3.7	2.3E-02	*TGFB3, TGFB1, RBL2, TFDP1, CDKN2B, DCN, BMP4, CREBBP,*
17	Epithelial cell signaling in H. pylori infection	7	3.3	2.0E-02	*MAPK10, IL8, LYN, SRC, EGFR, NFKBIA, MAPK14,*
18	Fc epsilon RI signaling pathway	7	3.3	2.8E-02	*AKT2, MAPK10, FYN, LYN, RAF1, IL4, MAPK14,*
19	mTOR signaling pathway	6	2.8	2.1E-02	*AKT2, TSC2, BRAF, FRAP1, IGF1, VEGFB,*

In OSCC from the gum, among the significant de-regulated pathways were those involved in cell cycle, MAPK signaling pathway and cytokine-cytokine receptor (Table 5.10). While in the T vs. NT gene list, only 19 genes were found to be differentially expressed (10 up-regulated, 9 down-regulated; Table 5.11). There were only 2 significant de-regulated pathways involved which were the hematopoietic cell lineage and cytokine-cytokine interaction receptor interaction pathways (Table 5.12).

Table 5.11 T vs. NT gene list. Differentially expressed genes in tongue OSCC compared to non-cancerous oral mucosal samples.

Gene	FC	Gene	FC
IL8	15.2	*MAPK14*	-1.9
MMP9	13.9	*DLC1*	-2.0
MMP10	10.6	*BAG1*	-2.0
FOLR1	8.2	*FGF2*	-2.3
SERPINE1	7.00	*FLT3*	-3.3
CSF3	6.4	*TYRO3*	-4.8
TERT	5.4	*PBX1*	-6.4
CSF2	4.2	*HLF*	-7.9
CTSL	2.3	*ALOX12*	-18.4
CDK2	1.7		

Table 5.12 De-regulated pathways in tongue OSCC, generated by comparing the T vs. NT list to the pathways database in DAVID.

No	KEGG PATHWAY (TT vs. TN)	Count	%	*p*-value	Genes
1	Hematopoietic cell lineage	3	16.7	0.01	*CSF3, FLT3, CSF2,*
2	Cytokine-cytokine receptor interaction	4	22.2	0.01	*CSF3, FLT3, IL8, CSF2,*

Through the Venn diagram, genes and KEGG pathways, which are common and site enriched were identified. Analysis of the genes revealed that there were more overlapping differentially expressed genes between OSCC originating from cheek and gum as compared to OSCC from cheek and tongue or OSCC from the gum and tongue (Figure 5.8; Table 5.13-Table 5.15).

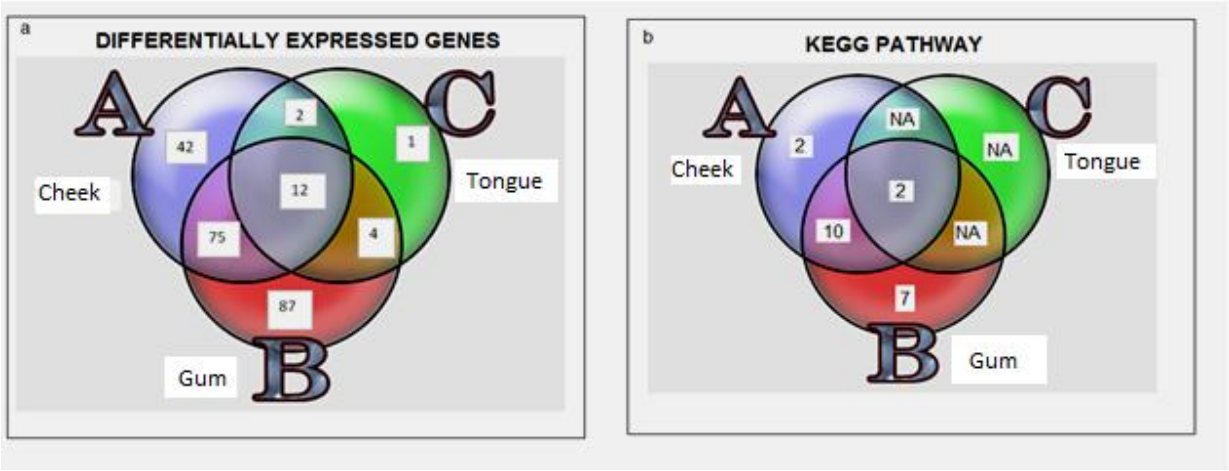


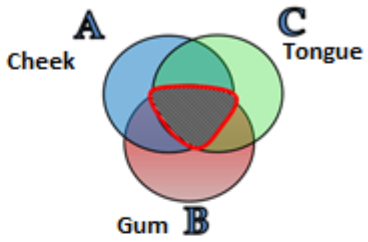
Figure 5.8 Venn diagram showing (a) number of genes and (b) number of de-regulated pathways which were site enriched and common in different sites of OSCC.

As for signaling pathways, cytokine-cytokine receptor interaction and hematopoietic cell lineage pathways were common amongst the 3 sites of the OSCC (Table 5.13b) while JAK-STAT pathways, MAPK pathways, cell cycle and regulation of actin-cytoskeleton were enriched in both cheek and gum OSCC (Table 5.14). Genes and pathways enriched in only a specific site of OSCC were identified (Table 5.16-5.18). Pathways involved in immune system regulation i.e. T-cell receptor regulatory pathway and Natural Killer (NK) cell pathway were enriched in only cheek OSCC whereas in gum OSCC, pathways specifically enriched were mainly those involved in apoptosis, WNT and TGF signaling pathway (Table 5.16; Table 5.17b;Table 5.17b). As for tongue OSCC, no signaling pathway was identified since there was only one gene which was enriched (Table 5.18).

Table 5.13 Differentially expressed genes and pathways which were enriched all 3 sites of OSCC – cheek, gum and tongue. Area shaded in black denotes the overlapping region in the Venn Diagram.

a. Genes enriched in Buccal, Gum & Tongue OSCC

	Gene	FC Buccal	FC Gum	FC Tongue
1	*IL8*	19.41	6.12	15.19
2	*MMP10*	14.58	3.19	10.61
3	*MMP9*	14.41	7.1	13.86
4	*SERPINE1*	6.92	6.26	6.99
5	*TERT*	5.33	5.39	5.35
6	*CTSL*	4.07	1.93	2.25
7	*MAPKI4*	-1.31	-1.87	-1.91
8	*FGF2*	-2.41	-1.67	-2.28
9	*PBX1*	-4.08	-4.29	-6.37
10	*TYRO3*	-4.41	-2.03	-4.83
11	*HLF*	-9.8	-7.09	-7.87
12	*ALOX12*	-23.81	-19.38	-18.42



b. Pathways enriched in Buccal, Gum and Tongue OSCC

	KEGG Pathway	Count	%	*p*-value	Genes
1	Cytokine-cytokine receptor interaction				
	Buccal	19	16.2	5.50E-07	*IL2, IL1B, IL11, IFNGR1, CSF3R, CXCL9, LTA, TNFSF10, CSF3, IFNG, TGFBR2, IL8, OSM, PDGFRA, EGFR, CSF2, MET, LIF, EPO,*
	Gum	23	10.7	1.20E-05	*TGFB1, IL12B, TNFSF8, IL1A, IL11, CSF3R, CXCL9, PDGFRB, VEGFB, CSF1R, TGFB3, FLT3, IL8, NGFR, PDGFB, PDGFRA, IFNGR2, TNFRSF1B, EGFR, TNFRSF1A, IL4, LIF, TNFRSF10A,*
	Tongue	4	22.2	1.40E-02	*CSF3, FLT3, IL8, CSF2*
2	Hematopoietic cell lineage				
	Buccal	8	6.8	8.80E-04	*CSF3, CD34, IL1B, IL11, CSF3R, CD9, CSF2, EPO,*
	Gum	11	5.1	2.90E-04	*CD34, FLT3, CD59, IL1A, IL11, CSF3R, CD9, IL4, TFRC, CD44, CSF1R,*
	Tongue	3	16.7	1.30E-02	*CSF3, FLT3, CSF2,*

Table 5.14 De-regulated genes and pathways that were enriched in cheek and gum OSCC. Area shaded in black denotes the overlapping region in the Venn diagram.

a. Genes Enriched in Buccal and Gum OSCC

	Genes	FC Buccal	FC Gum		Genes	FC Buccal	FC Gum		Genes	FC Buccal	FC Gum		Genes	FC Buccal	FC Gum
1	*CXCL9*	34.49	3.06	25	*LIF*	3.19	3.33	50	*MRE11A*	-1.61	-2.41	74	*DLG3*	-2.65	-1.72
2	*IL8*	19.41	6.12	26	*ICAM1*	3.13	1.76	51	*CDK7*	-1.61	-1.63	75	*TIAM1*	-2.89	-3.73
3	*MMP3*	18.63	5.46	27	*ING1*	2.95	-1.66	52	*CD9*	-1.62	-1.63	76	*IGFBP2*	-2.91	-2.58
4	*SPP1*	16.36	12.79	28	*PLAUR*	2.67	2.35	53	*NOTCH2*	-1.63	-1.35	77	*CD34*	-2.92	-1.44
5	*MMP10*	14.58	3.19	30	*STAT1*	2.20	-1.31	54	*CTNNA1*	-1.64	-2.14	78	*GAS7*	-3.05	-1.91
6	*MMP9*	14.41	7.10	31	*MYBL2*	2.20	2.71	55	*DDB2*	-1.73	-1.29	79	*WNT2B*	-3.13	-1.69
7	*MMP1*	14.37	8.35	32	*BAK1*	2.16	2.09	56	*RBBP1*	-1.77	-1.52	80	*MXI1*	-3.50	-2.72
8	*AIM2*	12.93	2.42	33	*PTK7*	2.05	1.82	57	*DCN*	-1.80	-1.56	81	*PBX1*	-4.08	-4.29
9	*IL6*	9.49	3.96	34	*GADD45A*	1.89	-1.30	58	*FOSL2*	-1.80	-1.59	82	*IGFBP5*	-4.22	-1.83
10	*PTHLH*	8.38	2.14	35	*CDC25B*	1.78	2.05	59	*MAD*	-1.82	-2.12	83	*TYRO3*	-4.41	-2.03
11	*MMP7*	7.60	4.54	36	*PML*	1.59	1.79	60	*XPA*	-1.82	-2.43	84	*TGFBR3*	-5.15	-4.69
12	*SERPINE1*	6.92	6.26	37	*ARHA*	-1.29	-1.71	61	*IL1RN*	-1.83	-2.39	85	*PNUTL1*	-7.41	-2.38
13	*TERT*	5.33	5.39	38	*MAPK14*	-1.31	-1.87	62	*QARS*	-1.88	-1.48	86	*HLF*	-9.80	-7.09
14	*HOXA9*	5.09	3.99	39	*DSP*	-1.35	-1.87	63	*SEMA3F*	-1.92	-2.80	87	*ALOX12*	-23.81	-20.10
15	*IRF1*	4.75	3.68	40	*TSC2*	-1.37	-1.51	64	*JUND*	-1.92	-1.44				
16	*IL11*	4.43	2.54	41	*MAF*	-1.52	-2.38	65	*FGFR3*	-1.96	-2.31				
17	*LYN*	4.31	2.11	42	*EGFR*	-1.54	-1.34	66	*PIM1*	-2.00	-2.34				
18	*ARHH*	4.19	2.71	43	*APC*	-1.55	-1.75	67	*EPS8*	-2.26	-1.84				
19	*CTSL*	4.07	1.93	44	*CDKN1B*	-1.56	-1.63	68	*ERBB2*	-2.28	-1.91				
20	*PDGFRA*	4.04	-1.38	45	*BRAF*	-1.57	-1.51	69	*AKT2*	-2.30	-1.37				
21	*CSF3R*	3.99	3.79	46	*FER*	-1.57	-2.07	70	*EPHA1*	-2.30	-1.89				
22	*CDC2*	3.59	2.15	47	*TFG*	-1.58	-1.37	71	*MLLT4*	-2.39	-2.19				
23	*LAMB1*	3.48	2.01	48	*IGFBP6*	-1.59	-1.64	72	*FGF2*	-2.41	-1.67				
24	*CDH11*	3.27	2.09	49	*WEE1*	-1.60	-1.62	73	*ERBB3*	-2.42	-2.51				

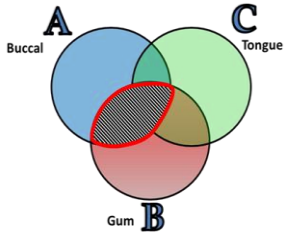


Table 5.14 , continued

	b. Pathways enriched in Buccal and Gum OSCC				
	KEGG Pathway	**Count**	**%**	***p*-value**	**Genes**
1	Jak-STAT signaling pathway				
	Buccal	14	12.0	3.0E-06	*AKT2, IL2, PIM1, IL11, IFNGR1, CSF3R, CSF3, STAT1, IFNG, OSM, CSF2, EPO, LIF, CCND1,*
	Gum	13	6.1	2.8E-03	*AKT2, IL12B, PIM1, IL11, CSF3R, JAK2, CREBBP, STAT3, STAT1, CCND3, IFNGR2, IL4, LIF,*
2	Focal adhesion				
	Buccal	12	10.3	9.5E-04	*VAV2, LAMB1, ITGB4, AKT2, ERBB2, BRAF, BIRC3, PDGFRA, SPP1, EGFR, MET, CCND1,*
	Gum	20	9.3	1.1E-05	*LAMB1, AKT2, CTNNB1, BRAF, MAPK10, PTEN, IGF1, SRC, PTK2, PDGFRB, VEGFB, CCND3, ERBB2, PDGFB, PDGFRA, FYN, SPP1, EGFR, RAF1, CRKL,*
3	MAPK signaling pathway				
	Buccal	11	9.4	2.0E-02	*GADD45A, AKT2, TGFBR2, BRAF, JUND, IL1B, PDGFRA, FGF2, EGFR, FGFR3, CDC25B,*
	Gum	24	11.2	4.2E-06	*AKT2, TGFB1, BRAF, IL1A, MAPK10, TP53, FGF12, PDGFRB, FGF7, NF1, TGFB3, GADD45A, JUND, PDGFB, PDGFRA, FGF2, EGFR, TNFRSF1A, RAF1, MAPK14, FGF9, CRKL, FGFR3, CDC25B,*
4	Cell cycle				
	Buccal	9	7.7	1.1E-03	*GADD45A, CDC2, CDKN2A, CDK7, WEE1, CDKN1B, CCND1, CCNA2, CDC25B,*
	Gum	25	11.6	3.1E-14	*TFDP1, CDC2, E2F3, CREBBP, TGFB3, GADD45A, E2F1, TGFB1, CDK2, CDKN1A, CDK7, CDKN2B, TP53, WEE1, CDKN1B, HDAC1, CCNH, RB1, RBL2, CCND3, PCNA, CDK6, ATM, CDC25B, CDK4*
5	Regulation of actin cytoskeleton				
	Buccal	9	7.7	4.3E-02	*VAV2, ITGB4, APC, BRAF, PDGFRA, FGF2, EGFR, TIAM1, FGFR3,*
	Gum	15	7.0	7.0E-03	*BRAF, FGF12, PTK2, TIAM1, FGF7, PDGFRB, APC, PDGFB, PDGFRA, FGF2, RAF1, EGFR, FGF9, CRKL, FGFR3,*

Table 5.14, continued

6	Adherens junction				
	Buccal	7	6.0	2.3E-03	*TGFBR2, ERBB2, EGFR, CTNNA1, MLLT4, FER, MET,*
	Gum	12	6.0	1.6E-05	*PTPRF, CTNNB1, ERBB2, FYN, SRC, CDH1, EGFR, CTNNA1, MLLT4, FER, TCF7L2, CREBBP,*
7	p53 signaling pathway				
	Buccal	6	5.1	8.0E-03	*SERPINE1, GADD45A, CDC2, CDKN2A, DDB2, CCND1,*
	Gum	17	7.9	1.4E-10	*CDK2, CDKN1A, PTEN, CDC2, IGF1, TP53, MDM4, DDB2, CASP8, GADD45A, SERPINE1, CCND3, TSC2, APAF1, CDK6, ATM, CDK4,*
8	ErbB signaling pathway				
	Buccal	6	5.1	2.0E-02	*AKT2, ERBB2, BRAF, CDKN1B, EGFR, ERBB3,*
	Gum	13	6.1	1.1E-05	*AKT2, BRAF, MAPK10, CDKN1A, FRAP1, SRC, CDKN1B, PTK2, ERBB2, RAF1, EGFR, ERBB3, CRKL,*
9	Toll-like receptor signaling pathway				
	Buccal	6	5.1	4.0E-02	*AKT2, STAT1, IL8, IL1B, SPP1, CXCL9,*
	Gum	11	5.1	1.3E-03	*CASP8, AKT2, STAT1, IL12B, MAPK10, IL8, RIPK1, SPP1, CXCL9, NFKBIA, MAPK14,*
10	Dorso-ventral axis formation				
	Buccal	5	4.3	1.6E-03	*ERBB2, BRAF, ETV6, EGFR, NOTCH2,*
	Gum	8	3.7	1.8E-05	*ERBB2, BRAF, ETS2, EGFR, NOTCH1, RAF1, ETS1, NOTCH2,*

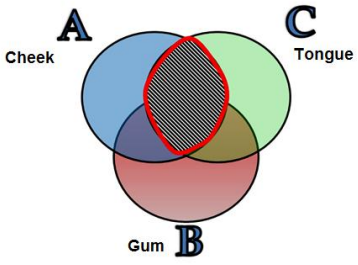
Table 5.15 Differentially expressed genes which were enriched in a. OSCC from the cheek and tongue b. OSCC from gum and tongue. Area shaded in black denotes the overlapping region in the Venn Diagram.

a. Genes enriched in Cheek & Tongue OSCC

Gene	FC Buccal	FC Tongue
IL8	19.41	15.19
MMP10	14.58	10.61
MMP9	14.41	13.86
CSF3	8.93	6.35
CSF2	7.56	4.19
SERPINE1	6.92	6.99
TERT	5.33	5.35
CTSL	4.07	2.25
MAPK14	-1.31	-1.91
FGF2	-2.41	-2.28
PBX1	-4.08	-6.37
TYRO3	-4.41	-4.83
HLF	-9.80	-7.87
ALOX12	-23.81	-18.42

b. Genes enriched in Gum & Tongue OSCC

	Gene	FC Gum	FC Tongue
1	*MMP9*	7.10	13.86
2	*SERPINE1*	6.26	6.99
3	*IL8*	6.12	15.19
4	*TERT*	5.39	5.35
5	*MMP10*	3.19	10.61
6	*CTSL*	1.93	2.25
7	*CDK2*	1.92	1.71
8	*FGF2*	-1.67	-2.28
9	*MAPK14*	-1.87	-1.91
10	*BAG1*	-1.89	-2.00
11	*TYRO3*	-2.03	-4.83
12	*FLT3*	-2.18	-3.33
13	*DLC1*	-2.48	-1.98
14	*PBX1*	-4.29	-6.37
15	*HLF*	-7.09	-7.87
16	*ALOX12*	-19.38	-18.42



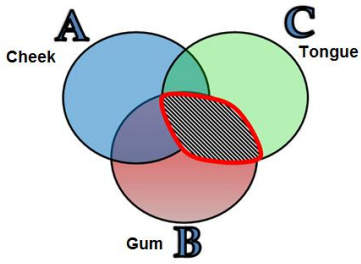
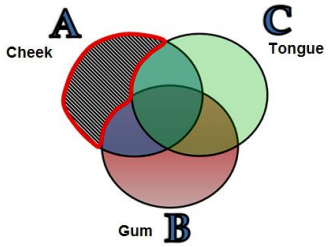


Table 5.16 Differentially expressed genes and pathways which were enriched in OSCC from cheek only. Area shaded in black denotes the site specific region in the Venn diagram.

a. Genes Enriched in only Cheek OSCC

	Genes	FC Cheek		Genes	FC Cheek		Genes	FC Cheek		Genes	FC Cheek
1	*IL1B*	6.37	12	*GFI1*	2.81	23	*GLI3*	1.97	34	*CCND1*	-1.73
2	*BCL2A1*	5.23	13	*FGR*	2.70	24	*CCNA2*	1.94	35	*MYB*	-1.93
3	*IFNG*	5.12	14	*ENC1*	2.67	25	*RET*	1.94	36	*TIMP2*	-2.02
4	*HMMR*	4.72	15	*MATK*	2.51	26	*BTK*	1.88	37	*EVI1*	-2.24
5	*RARB*	4.52	16	*TNFSF 10*	2.49	27	*ITGB4*	1.82	38	*CEBPA*	-2.27
6	*VAV2*	4.45	17	*PRKR*	2.45	28	*LCK*	1.72	39	*NQO1*	-2.51
7	*CDKN 2A*	3.51	18	*LIG4*	2.44	29	*IFNGR1*	1.61	40	*EPO*	-2.94
8	*EXT2*	3.22	19	*VEGF*	2.35	30	*IL2*	1.54	41	*AR*	-3.65
9	*OSM*	3.21	20	*SH3BP2*	2.17	31	*ETV6*	1.52	42	*CEACAM 1*	-3.89
10	*LTA*	3.21	21	*MCAM*	2.13	32	*TGFBR2*	-1.51			
11	*BIRC3*	3.16	22	*MET*	1.97	33	*CBFA2T1*	-1.72			



b. Pathways enriched in Cheek OSCC

No	KEGG Pathway	Count	%	*p*-value	Genes
1	Natural killer cell mediated cytotoxicity	9	7.69%	2.40E-03	*LCK, VAV2, ICAM1, IFNG, BRAF, IFNGR1, SH3BP, TNFSF10, CSF2,*
2	T cell receptor signaling pathway	6	5.13%	2.92E-02	*LCK, VAV2, IL2, IFNG, AKT2, CSF2,*

Table 5.17 Differentially expressed genes and pathways which were enriched in OSCC from the gum only. Area shaded in black denotes the site specific region in the Venn diagram.

a. Genes Enriched in only Gum OSCC

	Genes	FC Gum		Genes	FC Gum		Genes	FC Gum		Genes	FC Gum		Genes	FC Gum
1	*WNT10B*	4.34	19	*TFRC*	1.67	37	*ETS2*	-1.57	55	*SOD1*	-1.69	73	*ARHI*	-2.22
2	*E2F3*	2.95	20	*FZD7*	1.64	38	*GRB7*	-1.57	56	*RAP2A*	-1.70	74	*MYCL1*	-2.28
3	*CDK4*	2.93	21	*PDGFB*	1.59	39	*ERCC4*	-1.58	57	*RBBP6*	-1.70	75	*RAD52*	-2.32
4	*TIMP1*	2.69	22	*SPI1*	1.59	40	*SPARC*	-1.58	58	*EPS15*	-1.72	76	*PPARG*	-2.36
5	*TNFRSF 1B*	2.68	23	*MDM4*	-1.50	41	*NOTCH1*	-1.58	59	*S100A4*	-1.73	77	*TSG101*	-2.40
6	*TGFB1*	2.57	24	*MCC*	-1.50	42	*RB1*	-1.58	60	*NEO1*	-1.74	78	*IGF1*	-2.43
7	*TNFSF8*	2.53	25	*PLAT*	-1.50	43	*BCL3*	-1.58	61	*MALT1*	-1.74	79	*FGF7*	-2.46
8	*RECQL*	2.51	26	*TP53*	-1.51	44	*LCN2*	-1.59	62	*RBL2*	-1.75	80	*MAPK 10*	-2.60
9	*HCK*	2.29	27	*PDGFR B*	-1.51	45	*TPR*	-1.59	63	*IL1A*	-1.76	81	*FRZB*	-2.62
10	*PTPRH*	2.21	28	*CD59*	-1.52	46	*SMARCA4*	-1.60	64	*TOP1*	-1.76	82	*FGF12*	-2.75
11	*E2F1*	2.18	29	*ZNFN1 A1*	-1.52	47	*NF1*	-1.62	65	*YY1*	-1.77	83	*FHIT*	-2.78
12	*FANCG*	2.04	30	*WNT5A*	-1.52	48	*RLF*	-1.62	66	*OGG1*	-1.78	84	*LAF4*	-2.89
13	*CSF1R*	1.98	31	*VIL2*	-1.52	49	*RAD50*	-1.62	67	*TNFRSF 10A*	-1.82	85	*BMP4*	-3.00
14	*PCNA*	1.97	32	*EGR1*	-1.53	50	*CDH1*	-1.64	68	*DDX6*	-1.82	86	*BARD1*	-3.08
15	*ETS1*	1.97	33	*SRC*	-1.54	51	*LMO2*	-1.65	69	*MYCN*	-1.84	87	*NGFR*	-3.30
16	*PTK2*	1.94	34	*CCNH*	-1.55	52	*FGF9*	-1.67	70	*MSH3*	-1.98			
17	*FYN*	1.82	35	*MADH4*	-1.55	53	*VEGFB*	-1.68	71	*CDKN2B*	-2.08			
18	*CCND3*	1.78	36	*HDAC1*	-1.56	54	*MLL*	-1.69	72	*ERCC5*	-2.13			

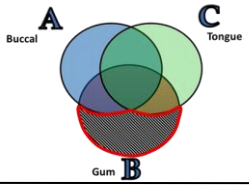


Table 5.17, continued

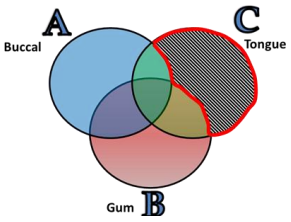
b. Pathways enriched in Gum OSCC

No	KEGG Pathway	Count	%	*p*-value	Genes
1	Wnt signaling pathway	12	5.58%	7.58E-03	*APC, CTNNB1, CCND3, MAPK10, TP53, WNT2B, WNT10B, MMP7, FZD7, TCF7L2, WNT5A, CREBBP,*
2	Apoptosis	11	5.12%	2.64E-04	*CASP8, AKT2, IL1A, APAF1, RIPK1, TP53, TNFRSF1A, NFKBIA, ATM, TNFRSF10A, PRKAR1A,*
3	Adipocytokine signaling pathway	8	3.72%	7.66E-03	*STAT3, AKT2, MAPK10, FRAP1, TNFRSF1B, TNFRSF1A, NFKBIA, JAK2,*
4	TGF-beta signaling pathway	8	3.72%	2.26E-02	*TGFB3, TGFB1, RBL2, TFDP1, CDKN2B, DCN, BMP4, CREBBP,*
5	Epithelial cell signaling in Helicobacter pylori infection	7	3.26%	1.95E-02	*MAPK10, IL8, LYN, SRC, EGFR, NFKBIA, MAPK14,*
6	Fc epsilon RI signaling pathway	7	3.26%	2.83E-02	*AKT2, MAPK10, FYN, LYN, RAF1, IL4, MAPK14,*
7	mTOR signaling pathway	6	2.79%	2.13E-02	*AKT2, TSC2, BRAF, FRAP1, IGF1, VEGFB,*

Table 5. 18 Differentially expressed gene which were enriched in OSCC from the tongue only. Area shaded in black denotes the site specific region in the Venn diagram.

Gene Enriched in only Tongue OSCC

Gene	FC Tongue
FOLR1	8.25



5.5.1. Validation of Site Enriched Pathways and Genes Identified among the Different Sites of OSCC

5.5.1.1. De-regulation of Immune System Pathway in OSCC

Through pathway analysis (Table 5.16-5.18), it was evident that pathways regulating the immune system were enriched in cheek OSCC compared to gum and tongue OSCC. In order to validate this finding, T-cell infiltration was measured by performing immunohistochemistry on CD3, which is the marker for T-cells. However, an increase in T-cell infiltration does not necessarily provide protection and increase in immune surveillance but it may suppress the immune system by another type of T-cell known as regulatory T-cells, which can be identified by the expression of FOXP3. Using 18 OSCC from 3 different sites (Cheek 9, Gum 2 and Tongue 7) and 9 non-cancerous oral mucosal tissues, it was demonstrated that T-lymphocytes infiltration was significantly higher in cheek, gum and tongue OSCC compared to non-cancerous oral mucosa (*p*-value < 0.05) (Figure 5.9). The mean of T-lymphocytes infiltration per 20X bright field was 95 in OSCC from cheek, 110 in gum OSCC, 116 in tongue OSCC and only 19 in non-cancerous oral mucosal tissues (Table 5.19). Again, FOXP3 positive T-cells were significantly higher in OSCC from the cheek, gum and tongue compared to non-cancerous oral mucosal samples (*p*-value < 0.05). The mean of FOXP3 (+)ve T-lymphocytes per 20X bright field was 29 in cheek OSCC, 28 in gum OSCC, 35 in tongue OSCC and none were seen in all non-cancerous oral mucosal tissue. However, no significant difference was detected in the presence of FOXP3 positive T-lymphocytes between OSCC from the cheek (31.4%) compared to OSCC from the gum (28.5%) and tongue (29.4%) (*p*-value > 0.05) (Table 5.21).

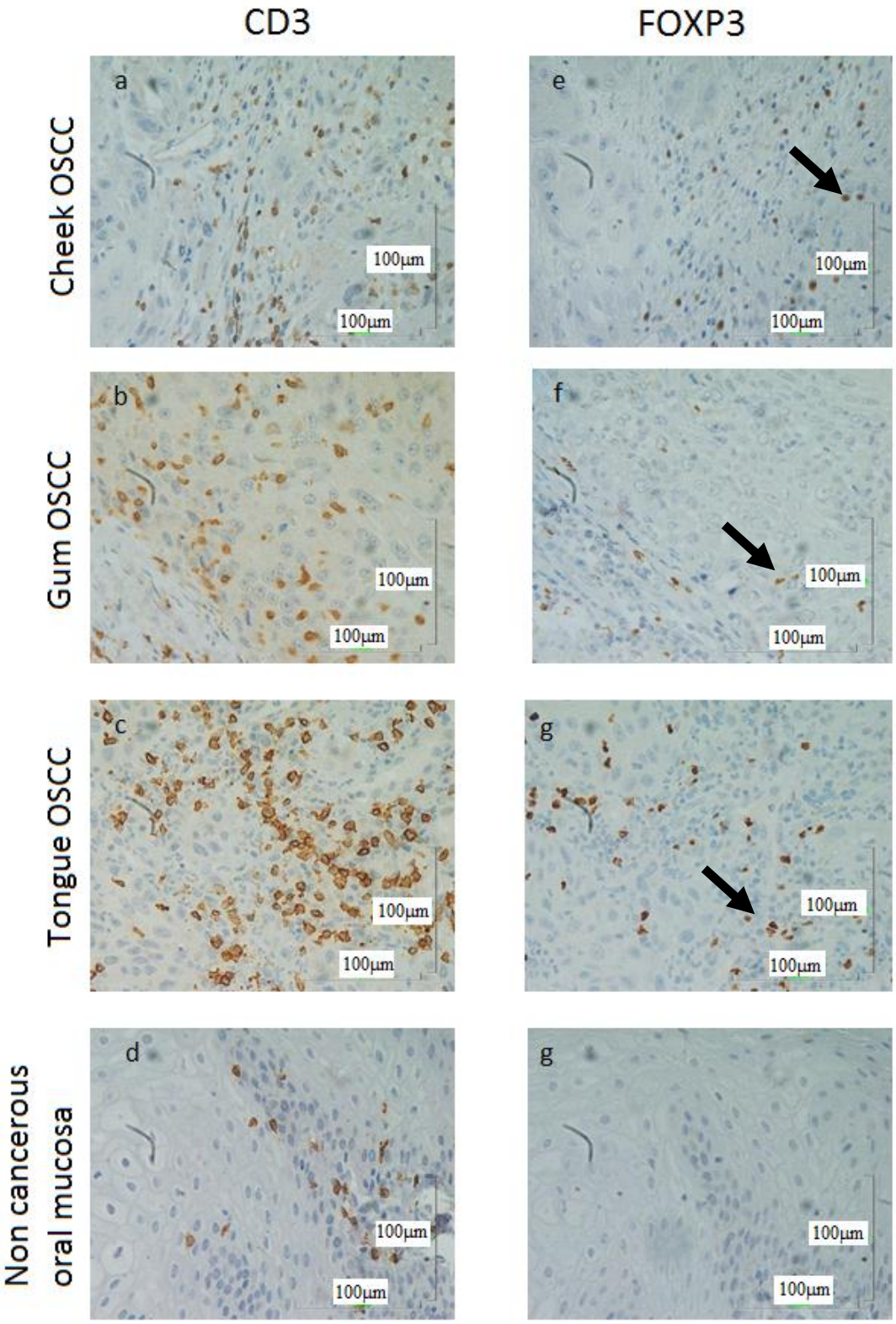


Figure 5.9 Representative images of OSCC and non-cancerous oral mucosa immunohistochemically stained with CD3 and FOXP3. OSCC from cheek (a,e), gum (b,f), tongue (c,g) and non-cancerous oral mucosal tissues from gum (d, h) at 400x magnification. Arrow showing FOXP3 positive T-lymphocytes.

Table 5.19 Mean per high power field of CD3, FOXP3 and percentage of FOXP3 in OSCC from cheek, gum and tongue and non-cancerous oral mucosa with standard deviation.

	CD3 (Mean ± SD)	FOXP3 (Mean ± SD)	%FOXP3 (Mean ± SD)
Cheek OSCC	95 ± 47	30 ± 17	31.4 ± 7.5
Gum OSCC	110 ± 76	29 ± 13	28.5 ± 7.7
Tongue OSCC	116 ± 56	35 ± 24	29.4 ± 5.8
Non- cancerous mucosa	19 ± 8	0	0

5.5.1.1. Expression of Folate Receptor Alpha (*FOLR1*) in OSCC.

Based on literature review and the fact that folate receptor alpha (*FOLR1*) gene is the only gene that wass enriched in tongue cancer, the expression of this gene was validated in OSCC. qPCR was conducted on 31 fresh frozen OSCC specimens from the different sites (Cheek 10, Gum 5 and Tongue 16) to measure the mRNA levels of *FOLR1*. The gene over-expression was found to be correlated to the site of oral cancer with tongue OSCC having the highest over-expression with a mean of 5.8 fold change followed by cheek OSCC with a mean of 3.17 fold change and gum OSCC with a mean of 2.95 fold change compared to non-cancerous oral mucosal (Figure 5.10). Immunohistochemistry was carried out to determine FOLR1 protein in OSCC tissues. However no positive staining was seen despite several attempts using several different antibodies. Nonetheless through qPCR, indeed FOLR1 was validated to be over-expressed in OSCC and found to be highly up-regulated in OSCC from the tongue in comparison to other sites.

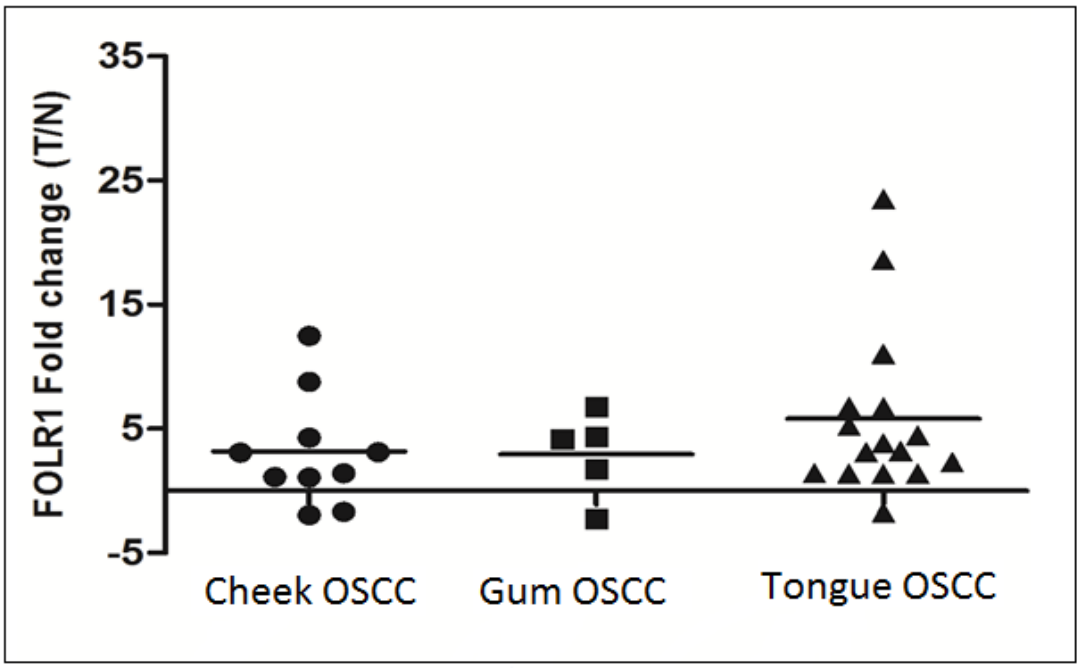


Figure 5.10 Validation of FOLR1 expression was determined by qPCR in 31 fresh frozen OSCC samples from cheek, gum and tongue.

5.6. Development of Oral Cancer Cell Lines That Over-Expresses Putative Oral Cancer Gene *FOLR1*

Based on qPCR site specific validation, *FOLR1* gene was selected for cloning into lentiviral plasmids. An oral cancer cell line with lowest *FOLR1* expression was identified using qPCR (Figure 5.11a). The *FOLR1* gene was cloned and exogenously over-expressed in 188T cell line. qPCR analysis confirmed the over-expression of 188T_FOLR1 where *FOLR1* was over-expressed by 48.5 fold change compared to 188T_pLenti (Figure 5.11b).

5.7. Determining the Role of Putative Oral Cancer Gene *FOLR1* in OSCC using Genetically Modified Cell Lines

The role of the FOLR1 in cell proliferation, cell migration and cell invasion in oral carcinogenesis was determined using the genetically modified cell line 188T_FOLR1 and 188T_pLenti.

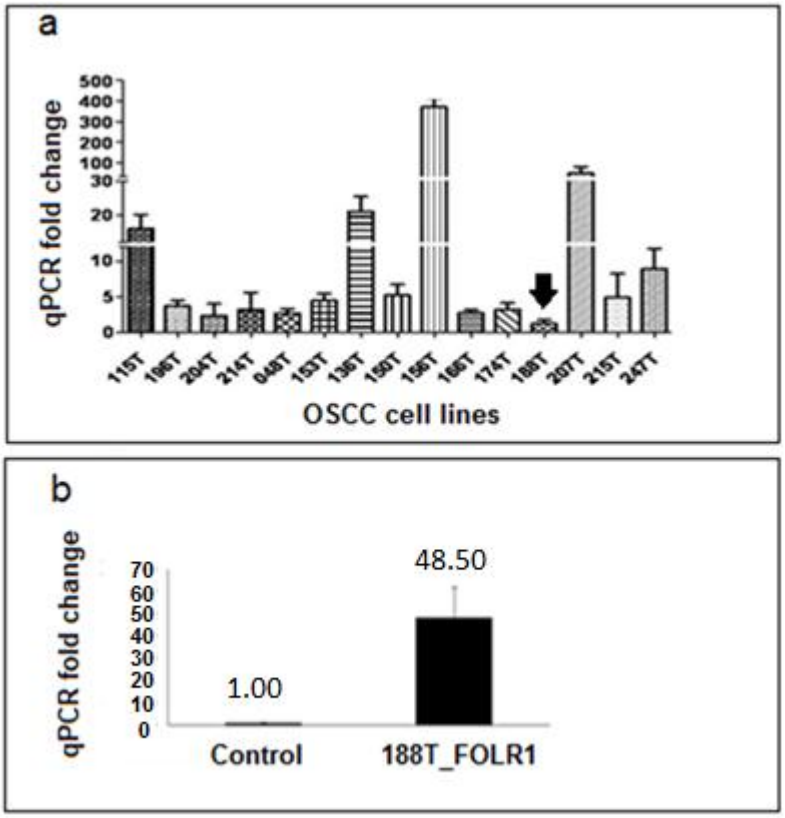


Figure 5.11 Development of OSCC cell line overexpressing FOLR1. (a) FOLR1 mRNA level in OSCC cell lines (arrow showing the lowest FOLR1 expression) (b) 188T_FOLR1 was over-expressed 48.5 fold compared to the 188T_pLenti.

5.7.1. The Role of FOLR1 in Cell Proliferation

The role of FOLR1 in promoting cell proliferation was determined using proliferation assay on 188T_FOLR1 and 188T_pLenti (control) cell lines as described in section 4.11.1. The assay was conducted three times in triplicates to confirm the results. The findings were consistent whereby no significant difference (*p*-value = 0.902) in cell

proliferation was seen between 188T_FOLR1 and 188T_pLenti, with both cell lines having the cell doubling rate of 27.1 hours (Figure 5.12a).

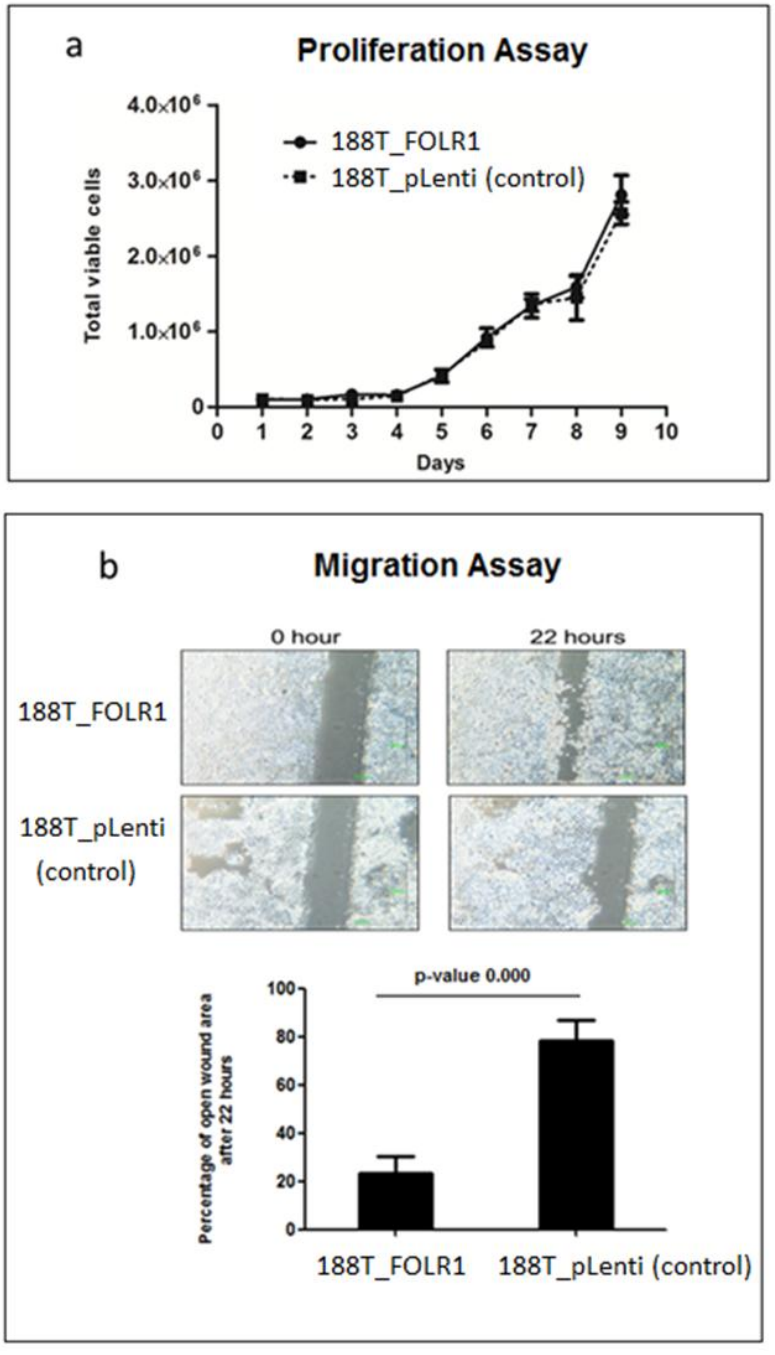


Figure 5.12 Determining the role of FOLR1 in cell proliferation and migration. (a) Graph showing cell proliferation of 188T_FOLR1 compared to 188T_pLenti where no significant difference was seen (*p*-value = 0.902) (b) Representative of the migration assay, taken at 0 and 22 hours and bar chart showing that 188T_FOLR1 cells are able to close the wound/scratch much faster than 188T_pLenti.

5.7.2. The Role of FOLR1 in Cell Migration

The role of FOLR1 in cell migration was determined using monolayer wound healing assay on 188T_FOLR1 and 188T_pLenti cell lines as described in section 4.11.2. Using the T-Scratch software, the open wound areas of both 188T_FOLR1 and 188T_pLenti at 0 hour and 22 hours were analysed. From the analysis, it was found that the percentage of open wound area in the 188T_FOLR1 cells was significantly lower compared to the 188T_pLenti cells (Figure 5.12b), indicating that the cell line with over-expression of FOLR1 migrates faster compared to 188T_pLenti.

5.7.3. The Role of FOLR1 in Cell Invasion

The role of FOLR1 in cell invasion was determined using 2 methods i) transwell invasion assay and ii) organotypic co-culture invasion assay as described in section 4.11.1. In transwell invasion assay, results showed no significant increase in invasion in 188T_FOLR1 compared to its control. Invasion index from 188T_FOLR1 compared to 188T_pLenti was 1.06 (Figure 5.13a). However, using organotypic co-culture invasion assay, 188T_FOLR1 have more invading cells compared to the 188T_pLenti. More cells broke free from the surface epithelium and there were more tumour islands compared to 188T_pLenti (Figure 5.13b). Using Image J software, the number of invading tumour island and the area of invading island was calculated as mentioned in section 4.11.1.2. 188T_FOLR1 has significantly more invading area compared to the 188T_pLenti with invasion index of 2.5 as shown in Figure 5.13b.

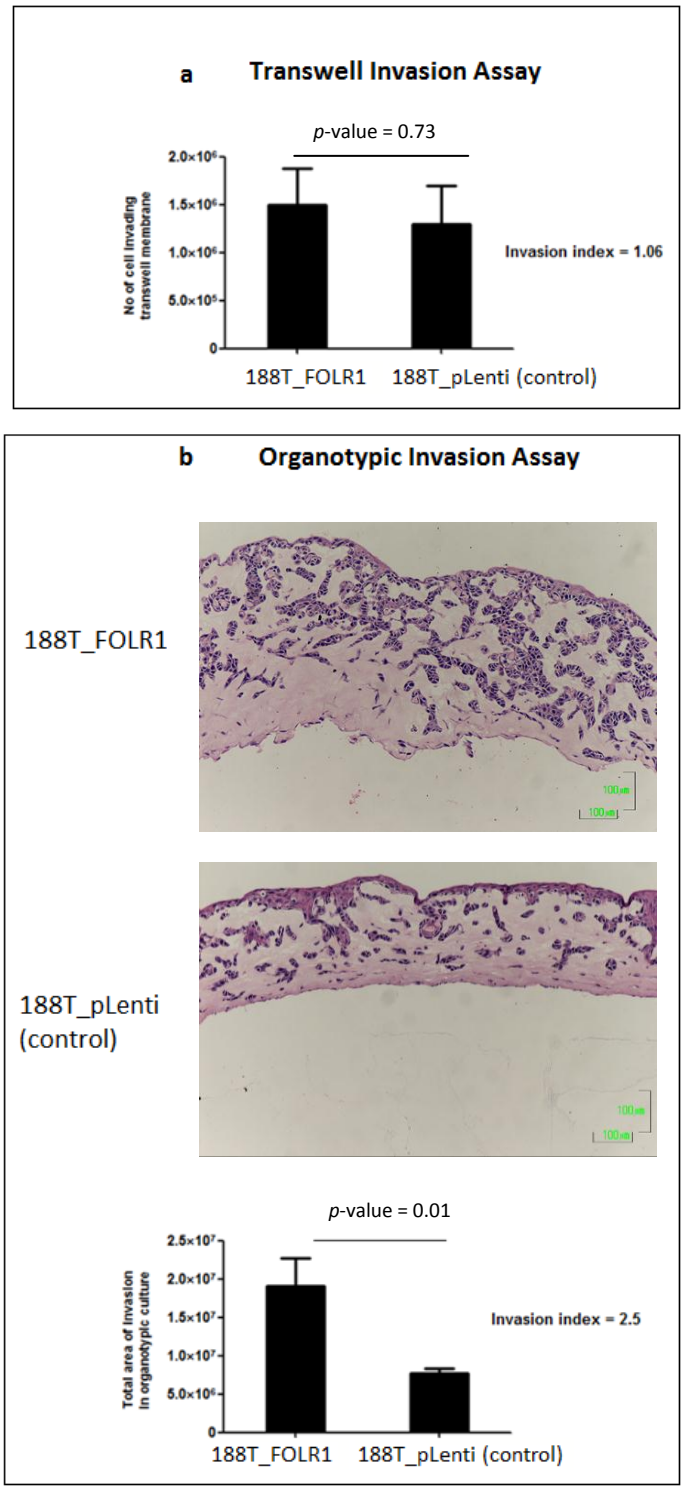


Figure 5.13 Determining the role of FOLR1 in cell invasion. (a) Using transwell (Boyden chamber) cell invasion assay, no significant difference was seen in 188T_FOLR1 compare to 188T_pLenti. (b) Organotypic invasion assay, representative of H&E staining where significant difference was seen in the invasion pattern of 188T_FOLR1 with invasion index of 2.5 compared to the 188T_pLenti.

CHAPTER 6

DISCUSSION

6.1. Quantity of RNA Obtained from FFPE Specimens

FFPE specimens represent the most abundant form of archived tissues and with complete clinico-pathological data, they are a good resource for studying the molecular mechanisms of diseases. In this study, we used FFPE specimens on a microarray platform to identify genes that were differentially expressed between OSCC and normal non-cancerous mucosa. As it has been established that specific cell types have distinct expression profiles (Elkahloun *et al.*, 2002; Lechpammer and Sgroi, 2004), we macro-dissected specimens before sectioning to ensure that more than 70% of tissues contained either OSCC cells or epithelium of non-cancerous oral mucosa. Using tissues stored between 1-7 years, a high percentage of the FFPE samples (91.4%) had sufficient RNA for the DASL gene expression assay. Consistent with other reports (Penland *et al.*, 2007), the RNA yield did not correlate directly with the age of the paraffin blocks but depended solely on the volume of tissues used for the extraction (Table 5.1). Notably, we managed to get a high yield of RNA from the normal epithelium by coring the epithelium using a 1.5 mm needle as reported by others (Prince *et al.*, 2007; Schobesberger *et al.*, 2008). Perhaps not surprisingly, the concentration of RNA obtained from OSCC tissues varied considerably more than those obtained from normal tissues due to a higher degree of variability in keratinization and necrosis in OSCC tissues compared to normal epithelium, as have also been observed by others (Malhotra *et al.*, 2004; Ravo *et al.*, 2008).

Studies conducted by other groups have found that tissue processing and long term storage of FFPE tissues compromises their use in gene expression as valuable information is lost due to RNA degradation (Paik *et al.*, 2004; Paik *et al.*, 2005; Lassmann *et al.*, 2009). Although the DASL assay takes into account some extent of degradation, a quality check was performed by looking at the Ct difference between the FFPE samples and a commercially available reference RNA when amplifying RPL13A gene. Although a somewhat arbitrary unit of 12 was recommended by others previously (Bibikova *et al.*, 2004a; Bibikova *et al.*, 2004b), a Ct difference of up to 17 was used in this study. The Ct difference with the number of genes that has significant intensity generated from the Beadchip was compared for each sample and it was found that the Ct value did not correlate directly with the number of significant genes on the array. In fact, samples with a Ct difference of 16 (T15) had more significant genes compared to a sample that had a smaller Ct difference (NB9, N10). This is perhaps consistent with previous reports demonstrating that the use of a single house-keeping gene may not represent the sample quantity or quality accurately (Bustin, 2000; Warrington *et al.*, 2000; Vandesompele *et al.*, 2002). Therefore, in view of this, perhaps looking at the number of genes with significant intensity for each sample is also important in selecting samples for further analysis.

6.2. Quality of Microarray Data Generated from RNA Extracted from FFPE Specimens

Using unsupervised hierarchical clustering and PCA, the tumours samples were clearly separated from the non-cancerous oral mucosal specimens indicating that indeed the gene signatures identified from our microarray experiments were not random and therefore, relevant genes implicated in OSCC development could be identified.

The expression of a subset of genes thought to be of biological relevance were further confirmed using qPCR. The expression of genes both by microarray and by qPCR were in concordance, again indicating that the microarray data was reliable. Indeed, by comparing genes that were identified using fresh frozen and FFPE tissues, Haque et al. (2007) reported that the gene lists were similar, and further demonstrated that although the number of differentially expressed genes was smaller in the FFPE group, the molecular sub-classification of glioblastomas was nevertheless possible in both types of specimens (Haque *et al.*, 2007). Notably, others have also successfully utilized FFPE specimens on microarrays to establish gene signatures that are indicative of patient prognosis (Chung *et al.*, 2004; Bibikova *et al.*, 2007) strongly supporting the use of FFPE specimens for the reliable identification of biomarkers.

6.3. Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC

In this study, pathways and genes that are associated with OSCC development were identified (Table 5.2) and gene expression of selected genes were further validated in a high percentage of independent samples (Figure 5.4; Table 5.4). Notably, matrix metalloproteinases such as MMP1, MMP9 and MMP10 which are key proteases involved in extracellular matrix (ECM) degradation [42]. were consistently seen to be up-regulated in the tumour samples of this study as have been reported by others in oral cancer and HNSCC (Mercurio and Rabinovitz, 2001; Jordan et al., 2004; Vachani et al., 2007; Cheong et al., 2009; Estilo et al., 2009; Yen et al., 2009). Since the extracellular matrix undergoes constant remodeling; metalloproteinase proteins can degrade the extracellular matrix proteins and lead to invasion and metastasis of cancer cells (Egeblad and Werb, 2002; Tanzer, 2006). MMP1 in particular, is consistently up-

regulated in invasive oral cancer (Vachani et al., 2007; Belbin et al., 2008; Estilo et al., 2009; Yen et al., 2009) and reported to be an important marker of malignant transformation in oral premalignant lesions (Jordan et al., 2004).

Interestingly pro-inflammation genes were also found to be up-regulated including IL8, IL1B and CXCL9. Similarly, Roa et al. (2010) have also shown that genes involved in inflammation and wound healing were up-regulated in OSCC and further indicate a strong link between inflammation and OSCC development and revealed that IL-8 as a potential mediator (Rao *et al.*, 2010). In addition, integrin which are heterodimeric, cation-dependent transmembrane glycoproteins that mediates cell-cell and cell-matrix interactions as well as played a role in maintaining tissue integrity, differentiation and migration (Thomas and Speight, 2001) was seen to be de-regulated. In OSCC, ITGB4 has been shown to be associated with early recurrence and metastasis and poor prognosis (Cortesina *et al.*, 1995; Eriksen *et al.*, 2004; Kurokawa *et al.*, 2008). Consistently, laminin B1 (LAMB1) which is one of the ligands for ITGB4 was also found to be de-regulated in this study as was reported previously (Cheong *et al.*, 2009). Notably, genes that have not previously received much attention in OSCC have also emerged in our study. For example, BCL2A1, an anti-apoptotic gene that acts by blocking the activation of the pro-apoptotic protein BAX and BAK (D'Sa-Eipper *et al.*, 1996; Vogler *et al.*, 2009) has been found to be over-expressed in 88.9% of our samples. Over-expression of BCL2A1 has also been reported in stomach cancers and breast cancers (Choi *et al.*, 1995; Yoon *et al.*, 2003). Importantly, BCL2A1 has been shown to suppress chemotherapy induced apoptosis (Wang *et al.*, 1999; Morales *et al.*, 2005) suggesting that modulating BCL2A1 expression may increase response to chemotherapy.

Despite the numerous challenges that are associated with the use of FFPE specimens, our study demonstrated that it is possible to identify genes that are important for OSCC development. The abundance of FFPE specimens associated with available clinical data coupled with the advancements in molecular technologies and statistical methods (Bibikova *et al.*, 2004a; Chung *et al.*, 2006) would certainly improve the identification of clinically relevant biomarkers for cancer. Intriguingly, one sample originally included in the non-cancerous oral mucosa group was found to segregate with the tumour samples upon PCA analysis suggesting that the gene expression signature of this sample (NT04) was more similar to that of the tumour specimens. Upon close inspection of this specimen, it was found that there was moderate dysplasia within this sample and more importantly, this lesion progressed to OSCC, 2 years following the biopsy of the dysplastic lesion. This reiterates the reliability of the use of FFPE samples in microarray, and strongly indicate that molecular changes can precede obvious histological changes (WHO, 2004)

6.4. **Gene Expression Variation in OSCC from Cheek, Gum and Tongue**

Many published microarray papers have identified genes that may play a role driving oral squamous cell carcinoma (OSCC) based on experiments that were conducted on a mixture of tumours from different sites of the oral cavity and some have even included a variety of head and neck tissues in the same experiments (Mendez *et al.*, 2002; Toruner *et al.*, 2004).

This partially explains the dissimilarity in the genes that were identified from different experiments that were conducted on OSCC. At the molecular level, Warner and

colleagues have demonstrated that the genetic profile of head and neck cancer cell lines differed based on the sites from which the cell lines were derived, suggesting that the genetics of head and neck cancers from different sites may be distinct from one another (Warner *et al.*, 2004). Furthermore, in 2008, Belbin et al. (2008) compared the gene expression pattern across three different sites of HNSCC: oral cavity, oropharynx and larynx/hypopharynx where each of the sites were then further analysed based on aggressive phenotype (patients whose disease progressed in 24 months' time). They further found through gene expression profiling that differences in gene expression of HNSCC disease are highly site-specific where only a fraction of identified genes were shared between any two sites. They further postulated that specific biological mechanisms underlying tumor aggressiveness are heavily influenced by the anatomic site of the primary tumor, such that different mechanisms offer advantage only within the specific environment of a single anatomic site (Belbin *et al.*, 2008).

At the clinical level, tumour of different sites of OSCC have distinct properties with regards to their clinical behavior and responsiveness to therapy (Zelefsky *et al.*, 1990; Chen *et al.*, 2007; Rautava *et al.*, 2007; de Araujo *et al.*, 2008). Patients with tumour from different sites of OSCC have been shown to have different survival rates with patients with tongue cancer having the worst prognosis due to high potential of local invasiveness and high tendency to metastasis to the nodes (Clayman, 1997; O-charoenrat *et al.*, 2003; Brandizzi *et al.*, 2008; Rusthoven *et al.*, 2008). Consistently at the molecular level, there is a growing body of evidence through experimental oral carcinogenesis studies using animal model and the analysis of major cell cycle regulatory proteins indicating that OSCC of different sub-sites are characterised by alterations in different genetic pathways (Schwartz *et al.*, 2000; Sathyan *et al.*, 2006). Chung et al. (2004) conducted a gene expression profiling on the various site of

HNSCC and found that compared to the rest of HNSCC, oral cavity derived tumours were the most heterogeneous. From then on, a number of microarray gene expression studies were then conducted to analyze the expression pattern of the different subsites of OSCC. However such studies were conducted on a small sample size and mainly consist of OSCC from the tongue and floor of the mouth which are the common OSCC sites in the West (Ziober *et al.*, 2006; Severino *et al.*, 2008). By comparing the gene expression pattern of OSCC from the tongue to the other sites of OSCC (cheek, floor of mouth, mandible and gum) in relation to matched non-cancerous oral tissues, Ziober et al. (2006) demonstrated through unsupervised clustering that OSCC from the tongue clustered differently compared to the non-tongue site. This further indicates and supports the fact that molecular heterogeneity exist within the oral subsites (Ziober *et al.*, 2006). In addition, Severino et al. (2008) also showed similar findings by comparing gene expression of OSCC of the tongue to OSCC of the floor of the mouth in relation to matched non-cancerous oral tissues where the OSCC samples do not group according to their pathological stages but in fact clustered mostly according to the anatomic subsites (Ziober *et al.*, 2006; Severino *et al.*, 2008).

Using FFPE specimens, the gene expression profiles of OSCC from three different sites: cheek, gum and tongue were examined. This is by far the largest study to conduct a comparison of gene expression from the different sites of the oral cavity and the first gene expression to study all 3 subsites of non-cancerous oral mucosal tissues simultaneously. In our study, through unsupervised hierarchical and PCA analysis (Figure 5.2), we observed that OSCC and non-cancerous oral mucosa clustered accordingly. Upon closer inspection in the PCA analysis, we observed that cheek and gum OSCC clustered accordingly to their anatomic subsites thus suggesting that the expression pattern of the different sites of OSCC are different (Figure 5.5). This finding

is parallel to the published data discussed above. However with tongue cancer specimens, this study demonstrated that the OSCC from the tongue did not have any specific clustering within the tumour boundary, as compared to the other two studies which reported that the gene expression pattern of tongue is distinct from the rest of the sites of OSCC (Ziober *et al.*, 2006). In fact, we observed that based on the molecular profile, tongue cancer specimens were distributed all over within the cheek and gum clusters thus suggesting that tongue cancers are heterogeneous. As mentioned previously, cancer from the oral cavity has also been reported to be heterogeneous (Chung *et al.*, 2004), although these reports do not mention the exact sites in the oral cavity where the cancer occurs. It should be noted that tongue cancer constitutes between 30-40% of oral cancer in the West (Rodrigues *et al.*, 1998; Silverman, 2001; Nemes *et al.*, 2008) and therefore it is probable that the majority of oral cancer samples included in the study were taken from the tongue. If so, these findings further strengthen our hypothesis that oral cancer, particularly from the tongue is heterogeneous.

To support these findings, 51 genes were examined by calculating the standard deviation of each gene for OSCC from the three different sites. Standard deviation here measured the variability or diversity of a particular gene within a group of samples showing how much variation or dispersion from the average mean, reflecting the variability and heterogeneity of the gene expression within a sample population. A higher standard deviation here denotes higher variability and heterogeneity of the expression amongst the samples. With that, the percentage of the genes with the highest standard deviation was seen in tongue OSCC (68%) compared to other subsites thus reinforcing heterogeneity in the gene expression pattern of tongue OSCC. Then again, a direct comparison of gene expression pattern of the tongue generated from our data and

the two mentioned studies cannot be done directly since all 3 studies used different microarray platforms.

In an attempt to understand major molecular differences between the sites and whether the differences seen at the gene expression level may in part contribute to the clinical difference seen in cheek, gum and tongue, the genes were further functionally annotated using KEGG pathways and further analysed using a Venn diagram to find the common and site enriched gene function and pathways. By a simple observation of the number of genes in the Venn diagram, there were two interesting findings. Firstly more genes were found to be overlapping between cheek and gum (75) as compared to cheek and tongue (2) and gum and tongue (4) (Figure 5.8). This may indicate that cheek and gum has more genetic similarities as compared to tongue. Most of the overlapping genes between cheek and gum OSCC were mainly clustered in pathways pertaining to cell communication pathways, cell growth and signal transduction pathways (p53, JAK-STAT and MAPK) (Table 5.14).

The second interesting observation that was seen in the differentially expressed gene list derived from tongue OSCC. In contrast with cheek and gum OSCC gene list, there were only 19 differentially expressed genes identified in tongue OSCC which were mainly those involved in cytokine-cytokine receptor interaction and hematopoietic cell lineage pathways (Table 5.11; Table 5.12). The differentially expressed genes constituted only 3.8% (19/502) of the total genes on the cancer panel. This is probably due to the heterogeneous gene expression pattern discussed previously and the fact that the gene expression analysis was based on only 502 cancer related genes. On the other hand, it may also indicate that, perhaps fewer gene expression changes are required for the progression of tongue cancer. To the same extent, Estilo et al. (2009) also identified

a relatively small number of differentially expressed probe sets in tongue OSCC. A total of 77 probe sets were identified from 12,625 probe sets in Affymetrix HG_U95Av2 array which represents only 0.6% of the total probe sets (Estilo *et al.*, 2009). Furthermore, a proteomic study also showed a significantly lower differentially expressed protein spots detected in tongue OSCC compared to the cheek OSCC (Chen *et al.*, 2004; He *et al.*, 2004), which further supports the hypothesis of fewer gene expression changes required for tongue cancer development and that OSCC from the tongue has a heterogeneous expression pattern.

Gene expression has been shown to be influenced by risk habits (Cheong et al. 2009). Therefore, the contribution of risk habit exposure could have contributed to the observations discussed above. Of note, patients with cheek OSCC and gum OSCC has a similar pattern of risk habits which is mainly betel quid chewing either as a single habits or in combination with other risk factors therefore, perhaps it is not too surprising to observe closer similarities between the gene expression of OSCC from the cheek and gum as compared to the gene expression from the cheek and tongue. Furthermore, the heterogeneity in the gene expression observed in OSCC of the tongue could also reflect the fact that betel quid and smoking were present in equal proportions in these patients. followed by alcohol and smoking (66.7%) and having all three risk habits (50%) (Apendix H). However, the analysis incorporating the data on risk habits was not possible at this time, as complete data on all patients included in this study was not available.

6.5. Identification and Validation of Differentially Expressed Pathways and Genes Implicated in OSCC from Distinct Sites

6.5.1. Common Pathways and Genes in OSCC Derived from Cheek, Gum and Tongue.

Through Venn diagram analysis, hemapoietic cell lineage and cytokine-cytokine interaction pathways were found to be common in OSCC derived from cheek, gum and tongue. Hemapoietic cell lineage pathway has been shown to have a role in the regulation of tumour angiogenesis as well as tumour progression and metastasis (Pollard, 2004; Kopp *et al.*, 2006). As for cytokine interaction pathways, there is increasing evidence suggesting the role of cytokines in tumour progression either by recruiting immune cells such as tumour associated macrophages to specific sites that stimulate tumour progression or acting directly, as their receptors are often found on tumour cells where they can increase tumour growth and migration (Pollard, 2004). In addition, certain cytokines/chemokines have been reported to induce signaling that activates multiple upstream and downstream signaling pathways which will lead to cancer progression (Waugh and Wilson, 2008). In our study two chemokines were found to be highly over-expressed, CXCL9 and IL8 at 34.5 and 19.4 fold increased respectively. With regards to CXCL9, Amatschek and group has recently shown that a high concentration of CXCL9 induces spontaneous migration and disruption of the endothelial barrier in melanoma cells and suggested that this may be a novel mechanism by which the cells leave the tumour and metastasize to distant sites (Amatschek *et al.*, 2011). CXCL9 has also been suggested to be a potential immunotherapeutic target where immune-competent mice injected with CXCL9-expressing tumor cells developed smaller local tumors and fewer lung metastases, and survived longer than mice injected with vector-control tumor where NK cells are crucial in the mechanism by which

CXCL9 limits metastasis (Walser *et al.*, 2007). IL-8 or alternatively known as CXCL8 is an NFκB dependent cytokine which has been associated with cancer progression in many cancers including OSCC. Up-regulation of IL8, triggers activation of NFκB pathways of which has been associated with high risk HNSCC (Chung *et al.*, 2006). In addition targeting IL-8 expression by siRNA significantly reduced the survival of OSCC cells, indicating that it plays an important role in OSCC development and/or progression (Rao *et al.*, 2010). Furthermore, because the majority of clinical studies confirmed that the over-expression of this chemokine in advanced disease, IL-8 or its associated CXC-chemokines may have important implications for the systemic treatment of aggressive and metastatic disease (Waugh and Wilson, 2008).

6.5.2. Site Specific Enriched Pathways and Genes in OSCC Derived from the Cheek, Gum and Tongue

Apart from identifying pathways and genes, which are common in OSCC derived from the cheek, gum and tongue, Venn diagram analysis could also identify those enriched in a specific site of OSCC (Figure 5.7; Table 5.16-Table 5. 18).

6.5.2.1. OSCC of the Cheek

Pathways that were found to be enriched in cheek OSCC were those involve in the immune system such as Natural Killer (NK) cell mediated cytotoxicity and T-cell receptor signaling pathways (Table 5.16). Both NK cells and cytotoxic T cells have been reported as important mediators of anti-tumour immunity as they are ultimately responsible for the destruction of malignant cells. Besides, NK cells which constitute the innate immune system were also able to influence the development of adaptive T-cells and B cell immune responses which represent specific immunity and

immunological memory to tumour and pathogen (Smyth *et al.*, 2002). Our finding is in line with Chen et al. (2004) where proteomic analysis demonstrated that genes enriched in cheek were mainly those involve in stimulating cell defense system in response to malignant transformation (Chen *et al.*, 2004) .

As previous studies indicated that patients with cheek OSCC has better prognosis and survival compared to those with OSCC of the gum and tongue (Brandizzi *et al.*, 2008), it is tempting to speculate that these immune pathways may be responsible for the reported clinical behaviour. In order to validate this speculation, the tumour infiltrating lymphocytes (TIL) were examined in OSCC from the 3 different sites (cheek, gum and tongue) by immunostaining with CD3 which is the marker for T-lymphocytes (Chetty and Gatter, 1994). The sub-population of T-lymphocytes known as regulatory T-cells (Tregs) that plays a critical role in regulating the balance between tolerance, attack of self and tumour associated antigens (Loddenkemper *et al.*, 2006) was also examined. Tregs have been implicated in the development of autoimmunity, allergy, and rejection of organ transplant as well as the suppression of immune responses in cancer. Furthermore an increased presence of Tregs has been reported in a wide variety of cancers such as lung, ovarian, gastrointestinal as well as head and neck cancer (Nizar *et al.*, 2009). A high percentage of Tregs has been associated with poor prognosis in a number of cancers including ovarian and breast cancer (Curiel *et al.*, 2004; Bohling and Allison, 2008). Interestingly, a high density of Tregs have also been reported to be correlated with improved overall survival in cancer patients in follicular B cells, Hodgkin lymphoma, colorectal cancer and head and neck cancer (Alvaro *et al.*, 2005; Badoual *et al.*, 2006; Carreras *et al.*, 2006; Salama *et al.*, 2009) through a proposed mechanism that Tregs down-regulate harmful chronic inflammation which would cause tumour progression (Maloy *et al.*, 2003; Schottelius and Dinter, 2006; Badoual *et al.*,

2009). Furthermore, it has been proposed that Tregs may also express cytotoxic molecules such as granzyme and perforin which could further induce the cell death of tumour cells (Grossman *et al.*, 2004; Erdman *et al.*, 2005).

Based on a study done by Loddenkemper et al. (2006), Tregs can be detected by co-staining CD3 with FOXP3 which is the master regulator in the development and function of regulatory T cells (Sakaguchi, 2000; Schubert *et al.*, 2001; Fontenot *et al.*, 2003; Loddenkemper *et al.*, 2006). T-cell infiltration was indeed significantly higher in OSCC compared to non-cancerous oral mucosal tissues (Table 5.19) which suggests heavy infiltration of immune cells, is in line with published data (Badoual *et al.*, 2008). However, no significant difference was seen between the different sites of OSCC. Since the networks of immune cells are known to be complex the staining of only CD3 may not be sufficient to validate the pathways which were found to be enriched in cheek OSCC. Perhaps a more detailed analysis of the immune system pathways that take into consideration the humoral and adaptive component of the systems needs to be carried out. When Tregs were analysed based on FOXP3 marker, again the percentage of FOXP3 positive T-lymphocytes was significantly higher compared to non-cancerous oral mucosa, which is consistent with published data (Loddenkemper *et al.*, 2006). Again, no significant differences were seen among the different sites of OSCC. Due to the complexity and intricateness of the immune system particularly the T-lymphocyte pathways (Cosmi *et al.*, 2003; Walker *et al.*, 2005), the immunostaining of CD3 and FOXP3 were not sufficient to validate this pathway. Thus, it would be interesting to further test this hypothesis by a detailed analysis with additional immunological markers such as CD4 and CD25.

6.5.2.2. OSCC of the Gum

Interestingly, signal transduction pathways which include TGFβ, Wnt and mTOR were found to be enriched in OSCC from the gum along with pathways that governs apoptosis. Generally, OSCC from the gum is known to metastasize to the bones (O'Brien *et al.*, 2003; Ishikuro *et al.*, 2008). Moreover, maxillary bone invasion of gum OSCC has been found be an indicator of cervical lymph node metastasis (Ogura *et al.*, 2003). Indeed, the 5-year survival of patients with tumour not invading the mandible was reported to be higher compared to those that invade to the mandible (53% and 25% respectively; p-value < 0.02) (Jones *et al.*, 1997). Intriguingly consistent with the phenotype of cancer cells in the gum, TGF-β and Wnt pathways has been previously shown to be associated with bone invasion. The TGF-β pathway has been implicated in bone formation and remodeling and also reported to be the central element in bone metastasis in breast and prostate cancer (Guise and Mundy, 1998; Sato *et al.*, 2008; Futakuchi *et al.*, 2009). A number of recent studies have suggested that the Wnt signaling pathway plays a central role in the regulation of bone development and homeostasis (Westendorf *et al.*, 2004; Glass *et al.*, 2005) and deregulation in this pathway has also been reported to be involved in bone metastasis (Hall *et al.*, 2006; Yee *et al.*, 2010). In our data, genes such as *WNT10b* and Frizzled protein 7 (*FZD7*) a receptor for Wnt pathway were found to be up-regulated (Table 5.17). Over-expression of WNT10b has been reported in osteosarcoma which was shown to cause stabilization and activation of B-catenin, leading to the deregulation of Wnt pathway which leads to deregulation of the Wnt pathway (Chen *et al.*, 2008b). Besides these genes, the *APC* gene which is part of the regulatory mechanism of the Wnt pathway was found to be down-regulated in this dataset (Table 5.17). In 2010, Svetlund and group found that the loss of *APC* gene led to aberrant activation of WNT pathway through accumulation of

B-catenin (Svedlund *et al.*, 2010). Therefore in a nutshell, perhaps the deregulation of TGFβ and Wnt signaling pathway may contribute to the clinical behaviour of gum OSCC to metastasize to the bones. .

6.5.2.3. OSCC of the Tongue

FOLR1, a folate receptor gene that plays a role in DNA synthesis was the only gene found to be enriched in OSCC of the tongue. *FOLR1* was found to be over-expressed by 8.2 fold change in tongue cancer (Table 5.11). Although it was previously found to be differentially expressed in OSCC from the tongue, the role of this gene has not been studied extensively in OSCC. In ovarian cancer, over-expression of FOLR1 has been associated with tumour progression and grade, resistance to therapy and decreased survival (Toffoli *et al.*, 1997; Toffoli *et al.*, 1998; Allard *et al.*, 2007; Kalli *et al.*, 2008). Over-expression of FOLR1 was seen in uterine serous carcinoma that is the most aggressive type of uterine cancers which are associated with shortened progression free survival (Allard *et al.*, 2007; Dainty *et al.*, 2007). Interestingly, tongue cancers which has the highest expression of FOLR1 have been reported be more aggressive compare to the OSCC from other sites (Rusthoven *et al.*, 2008) suggesting that expression of this gene could at least in part lead to the aggressive behavior of tongue OSCC. However, the role of folate in cancer development is still unclear. Epidemiologic studies have showed that folate intake above basal requirements reduces risk of developing various cancer by about 30-50% in breast cancer, colon adenomas and colon cancer. In contrast, mouse models have shown that folate levels above the physiological need (4-20 times more) can lead to progressive worsening of cancer (Kelemen, 2006). This led to the suggestion that increased growth and folate accumulation with elevated FOLR1

may be one mechanism by which folate could drive carcinoma development in humans (Kane *et al.*, 1988; Antony, 1996).

On the other hand, it has been proposed that the role of FOLR1 in cancer may be independent of its involvement in folate internalization (Antony, 1996). Bagnoli et al. (2000) analysed the interaction between FOLR1 and Caveolin (CAV-1) . Using ovarian cell line models, they found a reciprocal interaction between FOLR1 and CAV-1 where in CAV-1 transfected cells, FOLR1 expression was decreased and in FOLR1 transfected cells, CAV-1 expression was down-regulated (Bagnoli *et al.*, 2000). Both FOLR1 and CAV-1 are membrane proteins found to bind to the caveolar structure, which is a specialized membrane invagination, involved in vesicular trafficking. It has recently been reported that CAV-1 was able to regulate signal transduction pathways through the caveolar structure and thus has been considered as a putative tumour suppressor gene (Koleske *et al.*, 1995; Engelman *et al.*, 1997; Engelman *et al.*, 1998). It was further suggested that an increase in FOLR1 expression may be a new mechanism to silence CAV-1 which further leads to tumour transformation (Bagnoli *et al.*, 2000). Further, using animal models, loss of CAV-1 was reported to accelerate the onset of mammary tumors and enhance lung metastasis ((Williams *et al.*, 2004). In addition, Miotti et al. (2000) also demonstrated that the interaction between FOLR1 and other signaling molecules such as LYN and heterotrimeric G proteins of which has been proposed to be part of the macromolecular complex in which FOLR1, can generate intracellular signaling to promote cancer development (Miotti *et al.*, 2000). As FOLR1 is expressed at high levels in different cancers, the development of immunotherapy including peptide vaccine are being explored (Clifton *et al.*, 2011). Furthermore, the overexpression of folate receptors have been capitalized for drug delivery (Elnakat and Ratnam, 2004) for the following reasons: 1) folate receptor can bind to small molecules that are amenable

to chemical conjugation with other molecules without decreasing the binding affinity, 2) it can shuttle between cell surface and intracellular compartments effectively, 3) folate receptors expression in most normal proliferating cells is restricted to the luminal surface where it is inaccessible to circulation while it is consistently expressed in cancers and accessible via the circulation. Many innovative strategies for targeting folate receptor have been reported including radionucleotide conjugates of folic acid for whole body imaging of ovarian cancer (Leamon and Low, 2001) and folate targeted nanoparticle (Quintana *et al.*, 2002).

Saba et al. (2009), have examined the expression of folate receptors in HNSCC through immunohistochemistry using a polyclonal antibody that recognises FOLR1, FOLR3 and FOLRγ (Santa Cruz Biotechnology, Santa Cruz, U.S.A). Folate receptors were detected in 45% of primary tumors and 40% of corresponding lymph node metastases which affords an opportunity for the development of folate mediated nanotherapeutic drug delivery in HNSCC (Saba *et al.*, 2009). Unlike Saba et al. (2009), this study used an antibody that recognises FOLR1 specifically. Surprisingly FOLR1 expression was not detected on FFPE tissue despite the high mRNA expression found through qPCR. Two different antibodies were used, a mouse monoclonal antibody targeting residue 189 amino acid of FOLR1 (Novocastra Laboratories Ltd., Newcastle upon Tyne, United Kingdom) and a mouse polyclonal antibody raised using whole FOLR1 protein (Abnova, Taipei City, Taiwan). The lack of detection could be due to several possibilities. Firstly, the gene may be transcribed but not expressed at the protein level in OSCC or FOLR1 itself goes through a post translational modification which is specific to OSCC and thus the antibodies were not able to detect this modification. Secondly, it may be due to technical reason where the antibodies were not able to detect the epitope. Technical limitation in the use of FOLR1 antibodies in

immunohistochemistry analysis has been reported, where heterogeneous FOLR1 expression was observed in tumour from different patients and even within the same tumour tissues (Stein *et al.*, 1991; Toffoli *et al.*, 1997; Wu *et al.*, 1999). Furthermore, reproducibility and reliability of commercially available antibodies have recently been questioned (Bordeaux *et al.*, 2010). Lastly, it may be that the membrane bound FOLR1 was converted to a soluble form by metalloproteinase and other proteolytic enzymes (Kane *et al.*, 1988; Campbell *et al.*, 1991; Brown and Waneck, 1992), hence secreted but not detected on tissue specimens. Elwood et al. (1991) has identified three endoproteolytic cleavage site within the FOLR1 amino acid sequence and when cleaved the carboxyl terminus from residue 227-257 remain intact on the membrane bound FOLR1 (Elwood *et al.*, 1991). Nonetheless, as FOLR1 was shown to be expressed at a high levels at the mRNA level (Figure 5.10) its possible role in oral carcinogenesis was examined in OSCC cell lines.

6.6. Determining the Role of FOLR1 in OSCC using Genetically Modified Cell Line

With successful over-expression of *FOLR1* open reading frame into oral cancer 188T (Figure 5.11), three different experiments were carried out to determine the role of FOLR1 in driving oral cancer.

6.6.1. The Role of FOLR1 in Cell Proliferation

Since folate has been frequently reported to be involved in DNA synthesis, replication and cell division (Kelemen, 2006), the first assay that was conducted was the proliferation assay. From the results, there was no significant difference found between 188T_FOLR1 and 188T_pLenti where the doubling time for both cell lines were the

same (27.1 hours) (Figure 5.12a). The uptake of folate for cell proliferation is not a straight forward process. Two different systems exist for the cellular uptake of folate, the first, which is dependent on membrane bound folate receptors (FOLR1 and FOLR3) internalizes folate by receptor mediated endocytosis and second is through reduced folate carrier (RFC) which uses bi-directional anion exchange mechanism to transport folate into the cytoplasm. This mechanism of action is highly dependent on the folate concentration. RFC will bind folate with high affinity when folic acid levels are very high above physiologic level (μmolars) which is equivalent to the level of folate in cell culture media, while the receptors have a high affinity for folate at physiological levels (nanomolars) (Kelemen, 2006). Bottero et al. (1993) have shown that in standard medium containing high level of folate, the growth potential for cell lines with over-expressed FOLR1 and the control is the same thus indicating that at high folate concentration found in cell culture media, both cell lines can access folate though RFC.

In contrast, a decrease in growth curve was seen in the control cell when the cells were grown in folate-depleted media compared to the over-expressed FOLR1. Thus suggesting that in folate-depleted environment, FOLR1 are more efficient in folate uptake compared to RFC and allows the cell to proliferate even in a folate depleted situation (Bottero *et al.*, 1993). In this experiment, cells were grown and cultured in complete media that has a very high levels of folic (6.0 μM) therefore this may contribute to the insignificant difference seen in the proliferating rate between 188T_FOLR1 compared to 188T_pLenti. Thus, the effect of FOLR1 on proliferation requires further investigation.

6.6.2. The Role of FOLR1 in Cell Migration and Invasion

This study showed that 188T_FOLR1 migrates and invades significantly faster than 188T_pLenti grown and cultured in a complete culture medium. The ability of FOLR1 to promote migration was tested using a monolayer wound healing assay. This assay is a classic and commonly used method to study cell migration and the underlying biology (Lampugnani, 1999; Liang *et al.*, 2007). It is based on the observation that, upon creation of an artificial gap (scratch/wound) on a confluent cell monolayer, the cells at the edge of the newly created gap will move towards the opening to close the "scratch" or "wound" until new cell-cell contacts are established again (Liang *et al.*, 2007). One of the major advantage of this method is that it mimics to some extent the migration of cell in vivo (Haudenschild and Schwartz, 1979) along with the fact that it is straight forward and economical method (Liang *et al.*, 2007). On the other hand, there are some limitations to this assay of which need to be considered when designing migration assay using this method as compared to other popular assay such as Boyden chamber assay. Firstly, the wound healing assay only monitors cell movement but not chemotaxis ability as no chemical gradient can be established in this assay. Secondly, it takes relatively longer to perform the assay since one to two days are needed for the formation of monolayer prior to the migration assay and 8-24 hours for cell migration to close the wound/scratch. Thirdly, it requires a large number of cells since it is usually performed in a tissue culture dish and thus may not be a suitable assay to use if the availability of cell is limited (Liang *et al.*, 2007).

As for invasion assay, two different invasion assays were carried out in this study, which includes transwell invasion assay (Boyden chamber) and organotypic culture invasion assay. The transwell invasion assay is a 2-D assay, which only measures

number of invaded cells through a microporous (8 μm) membrane coated with matrigel in response to the chemoattractant underneath, without considering the interaction with the stroma or factors in the tumour microenvironment. The organotypic culture invasion assay however, is a 3-D assay where cancer cells are cultured over a period of 10-14 days on top of a matrix embedded with fibroblasts, which mimics the stroma of the tumour microenvironment and thus takes into consideration the interaction between tumour cells and its microenvironment. Interestingly, cells over-expressing FOLR1 showed no significant difference in transwell invasion assay with invasion rate of 1.1 compared to the 188T_pLenti but on the other hand, using the 3-D organotypic culture invasion assay, cells over-expressing FOLR1 showed significant increase in invasion rate of 2.5 compared to control cells. This may indicate that perhaps the tumour microenvironment plays an important role in FOLR1-mediated cell invasion process. This could be because the tumour microenvironment can create folate deficient areas in which the FOLR1 positive cells could have a growth advantage (Kalli *et al.*, 2008).

It is intriguing to find that FOLR1 up-regulation correspond to site of OSCC with tongue. Folate deficiency has been reported to be a common sign of chronic alcohol consumption. Studies conducted in U.S.A and Britain on alcoholics showed that the majority (up to 80%) has low folate levels (Halsted *et al.*, 2002). Moreover, experimental studies have shown not only that the levels of folate fall dramatically with acute alcohol ingestion but it also decreases the supply of folate to the tissue (Steinberg *et al.*, 1982). Among individuals with a low intake of folate and a high intake of ethanol, the risks of colorectal adenoma and cancer were increased >2-fold (Giovannucci *et al.*, 1995; Baron *et al.*, 1998). In cancer patients with chronic excessive consumption of ethanol, folate deficiency may occur because of inadequate dietary

intake, malabsorption, or defects in folate-binding proteins or in the enterohepatic absorption and recirculation of folate (Schottenfeld and Beebe-Dimmer, 2006).

Interestingly, most Western countries, tongue and floor of the mouth is the most common site of oral cancer and frequently associated with alcohol and tobacco consumption (La Vecchia *et al.*, 1992; Oliver *et al.*, 1996; La Vecchia *et al.*, 1997). Even in our study cohort, 35% of tongue cancer patient were associated with either smoking or drinking alcohol or combination of both (Appendix A) whereas a lower percentage of similar risk habits was seen in OSCC derived from the gum and cheek at 25% and 5% respectively. Perhaps alcohol consumption and smoking risk habits could be among the factors that may lead to folate deficiency, which lead to the up-regulation of FOLR1 as a "survival of the fittest" mechanism since FOLR1 can bind folate more efficiently at a very low levels and the fact that folate is critical for cell viability (Antony, 1992). This may lead to cancer progression and invasive nature in tongue cancer.

6.7. **Limitation of Study**

The purpose of this study was to establish the gene expression patterns of OSCC from different sites of the oral cavity (cheek, gum and tongue) using formalin-fixed paraffin-embedded samples by microarray. In addition, this study compared the gene expression pattern generated from each site of the oral cavity to determine the similarities and differences in the pathways that are activated in the 3 different sites of the OSCC. Whilst this study was the first to identify similarities and differences in terms of gene expression of OSCC from the cheek, gum and tongue, a number of limitations were recognised.

In this study, the DASL assay containing 502 cancer related genes (DASL Cancer Panel) was utilized. Here, the limitation encountered was related to the presence of a limited number of genes on the cancer panel. Nonetheless, despite the limited number of genes, the tumour and non-cancerous oral mucosal tissues were able to be clustered accordingly thus indicating the reliability of the data. Using similar platforms, other groups have shown the successful use of FFPE samples in identifying gene signatures indicative of patient prognosis (Bibikova *et al.*, 2004a; Bibikova *et al.*, 2007; Haque *et al.*, 2007; Hoshida *et al.*, 2008). In addition, the DASL Cancer Panel assay is a good starting point to examine site specific differences in OSCC as all of the genes within the panel are those within molecular pathways that are associated with cancer thus is suitable for a cancer-focused study. More recently, a DASL assay containing probes for the whole genome designed for the use of FFPE samples have been developed (Illumina Inc., U.S.A). Notably, by comparing gene expression data from the DASL 502 gene and that with the whole genome DASL assay, Reinholz et al.(2010) demonstrated biological consistency between these two platforms in spite of the different densities of genes in each platform (Reinholz *et al.*, 2010). The second challenge with the current cancer panel is that the small number of genes present on the panel may limit the pathway analysis, where only a limited number of pathways could be identified. Coupled with the fact that the expression pattern in tongue cancer was heterogeneous, the challenge was more apparent in tongue cancer where pathway analysis could not be performed since only one gene was found to be enriched. It would be interesting to analyse the data using the Whole Genome DASL platform, where a larger gene set will provide us granularity i.e. an in-depth knowledge and understanding in pattern of oral carcinogenesis.

The other limiting factor within this project was the ability to use normal tissues for controls. The selection of non-cancerous oral mucosal tissues as controls for microarray experiments have been heavily debated (Choi and Chen, 2005). In this study epithelial surface from fibroepithelial polyps or fibrous epulis were selected as control specimens to match OSCC derived from a specific site (cheek, gum and tongue) as they can be found more commonly and ethically easier to obtain as they are generally excised for clinical examination. Furthermore, even though these tissues have reactive stroma,the epithelial compartment can be considered fairly normal (Scully *et al.*, 2010) and perhaps it may be the closest tissue type to normal oral mucosa. Fibrous polyps or also known as fibroepithelial polyps (FEP) are the most common hyperplastic lesions in the oral cavity and mostly found in trauma prone sites such as the cheek mucosa and tongue. When the fibroepithelial polyp appears in gum, it is usually referred to as fibrous epulis (Prabhu, 2008).

In order to address the reliability of the choice of control in this study, another set of FFPE controls which were gingival tissues obtained during the surgical removal of impacted wisdom tooth were included. Interestingly, through unsupervised expression analysis, all the non-cancerous oral mucosal samples consisting of FEP, fibrous epulis and gingival tissues were found to segregated together and apart from the tumours indicating that the gene expression patterns of all non-cancerous oral mucosal tissues are more similar to one another and distinct from that of the tumours, thus suggesting that these non-cancerous oral mucosal tissue were indeed non-cancerous and suitable to be used as control in gene expression analysis.

It has been reported that risk habits can influence the gene expression pattern in OSCC (Cheong *et al.*, 2009). In this study, it was assumed that the differentially expressed genes found here are unique to each sites, however it cannot be negated that risk factors may also influence gene expression changes. This was not taken into consideration during statistical analysis as complete data on all patients included in this study was not available.

CHAPTER 7

CONCLUSION AND RECOMMENDATIONS

7.1. Conclusion

Through unsupervised hierarchical clustering and PCA, this study showed that the gene expression pattern of tumour and non-cancerous oral mucosal tissues could be segregated accordingly. Genes that were previously associated with OSCC were identified in this study and successfully validated in independent fresh frozen samples. This further indicates that good reliable data were generated from RNA extracted from FFPE specimens and that microarray experiments can be conducted using FFPE tissues. In this study, 153/502 differentially expressed genes were identified in OSCC compared to the non-cancerous oral mucosal tissues where selected genes were validated in a high percentage of independent fresh frozen samples at mRNA and protein level. Interestingly, genes that have not received much attention in OSCC were also identified and validated in this study. Importantly, Principal Component Analysis demonstrated that gene expression variation exists between OSCC from the different sites. The gene expression signature of OSCC derived from cheek was different from those derived from the gum whereas OSCC of the tongue had a more heterogeneous gene expression pattern. From this data, it was evident that only a small number of genes (19) were differentially expressed between OSCC from the tongue and non-cancerous tongue mucosa, which may be due to the heterogeneity in the gene expression of OSCC from the tongue or that the number of genes needed for the initiation and development of tongue cancer is inherently small. This study also identified cytokine-cytokine

interaction and hematopoietic cell lineage pathways to be common in all three sites and further demonstrated that the gene expression of the cheek is more similar to the gum as compared to the tongue. The overlapping genes between OSCC from the cheek and gums were mainly those involved in cell communication, cell growth and signal transduction. Pathways enriched in specific sites were identified where for OSCC from the cheek, immune system pathways were enriched. In OSCC of the gum, TGF-β and WNT signaling pathways were enriched, while in OSCC from the tongue, FOLR1, a single gene which has not been studied in OSCC was found to be enriched. With the genetically modified *FOLR1* over-expressing cell line, functional assays to test the ability of FOLR1 to promote cell proliferation, cell migration and cell invasion were performed and it was found that *FOLR1* plays a role in migration and invasion but did not confer a growth advantage.

7.2. Recommendations for Future Work

Below are several recommendations for future research in order to broaden and improve on the findings of this study.

7.2.1. Conducting gene expression analysis using a whole genome platform

In this study, gene expression analysis was conducted on the DASL assay containing 502 cancer related genes (DASL Cancer Panel). Although genes and pathways which are differentially expressed in oral cancer as compared to the non-cancerous oral mucosa were able to be identified, this did not provide an in-depth and comprehensive gene expression pattern. More recently, a DASL assay containing probes for the whole genome designed for the use of FFPE samples have been developed (Illumina Inc.,

U.S.A). Therefore it is highly recommended that gene expression analysis should be conducted using such platform in future studies.

7.2.2. Controlling for Confounding Factors

Risk habits has been found to influence the gene expression of OSCC (Cheong *et al.*, 2009). In addition, risk habits or lifestyle habits may influence the site of OSCC occurrence (Zain, 2001; Petti, 2009). Therefore in future studies to further cluster the genes according to the sites, it is highly recommended that confounding factors such as risk habits (including the exposure length), stage of disease and patient survival status should be controlled in the data analysis.

7.2.3. Conducting Premalignant Gene Expression Analysis Based on Preliminary Data Generated from this Study

In addition to understanding the gene expression pattern of different sites of OSCC, this study has generated interesting data on genes that maybe involved in malignant transformation. Therefore, it would be interesting to confirm this finding by comparing premalignant lesion samples that did not progress to OSCC to premalignant lesion those that have progressed to OSCC and to conduct these experiments on a higher density array such as the DASL Whole Genome platform in order to further substantiate this finding.

7.2.4. Up-Regulation of Immune Pathways in OSCC from the Cheek

This study was not able to validate the up-regulation of the immune systems pathways and did not find any significant difference in tumour infiltrating lympohocytes (TIL) in

cheek OSCC compared to gum and tongue. However, the validation was merely based on the expression of CD3 and FOXP3. Therefore, in future studies identifying the different components of the immune system should be carried out. Certainly, further understanding of the immune system is crucial since their activity is most likely to affect immunotherapy, which is one of the promising new cancer treatments.

7.2.5. The Role of in Promoting Migration and Invasion in OSCC

From this study, *FOLR1* over-expression was shown to promote migration and invasion. This indicates that *FOLR1* may stimulate aggressive nature of the cancer. It has been proposed that overexpression of *FOLR1* is one of the mechanism that caused the loss of *CAV-1* which a putative tumour suppressor gene. This hypothesis now opens up a new research areas to further investigate the role of *FOLR1* in aggressive nature in OSCC, in particular, tongue cancer.