

1.0 Introduction:

Oral cavity cancer is 12th most common type of cancer related death worldwide (Parkin *et al.*, 1999; Scully and Bedi, 2000). In 2008, Global Burden of Cancer (GLOBACAN) estimated the incidence of oral cancer (including lip cancer) as 263,900 cases and reported 128,000 of oral cancer related deaths (Jemal *et al.* 2011). In oral cancer, more than 90% is reported to be as epithelial neoplasia with majority of them is oral squamous cell carcinoma (OSCC) (Nagpal *et al.*, 2003). OSCC is derived from the surface epithelium and then expanding upward from the basement membranes which replaces the normal epithelium. Since the basement membranes are usually being penetrated, the carcinoma will invade the underlying connective tissue (Nagpal *et al.*, 2003; Silverman, 2003). In the present scenario, two-thirds of the new oral cancer cases are reported from the developing regions of the world (Parkin *et al.*, 1999). These regions include parts of Africa, Central and South America, Caribbean, China, Asia, Melanesia and Micronesia/Polynesia (Parkin *et al.*, 1999). Depending on the geographic locations, it has been noticed that there are variations in prevalence of oral cancers (Moore *et al.*, 2000). However, reports suggest that these geographic variations are reflected mainly due to the prevalence of specific environmental and habitual influences rather than any genetically determined/ ethnic/ risk factors (Moore *et al.*, 2000). Examples for these geographic variations in terms of their specific environmental or habitual conditions are; tobacco and alcohol consumption in Western and South Europe and also other Southern African nations in relation to mouth and tongue cancer (Moore *et al.*, 2000), betel quid chewing in Melanesia and South Central Asia (mainly India and Taiwan) in relation to cheek cancer (Werning and Mendenhall, 2007; Kao *et al.*, 2009) and high recurrence of solar irritation in Australia and New Zealand in relation to lip cancer (Parkin *et al.*, 1999; Parkin *et al.*, 2005). Globally, OSCC is more frequent among males than in females with a ratio of 2:1 (Parkin *et al.*, 1999). Mostly, this male

preponderance is observed in the West (Moor *et al.*, 2000), whereas in Malaysia, interestingly the female preponderance is reported (Ng *et al.*, 1985; Siar *et al.*, 1990). Despite the great therapeutic and diagnostic advancement in the past decade, the survival rate of oral cancer patient has not been improved much (Silverman, 2001).

Carcinogenesis of oral cancer is a complex process which results from a multistep pathway involving the accumulation of genetic and chromosomal instability (Scully *et al.*, 2000). This would lead to the activation of proto-oncogenes and inactivation of tumor suppressor genes (Choi and Myers, 2008). Considering the mere fact that huge numbers of genes are involved in oral carcinogenesis and tumor progression, as well as the cellular and molecular heterogeneity of OSCC, comprehensive studies are required to study the multiple gene alterations on a global scale (Estilo *et al.*, 2009).

There is an increasing demand over the incisional biopsy of the primary tumor to be used for refined initial characterization of the tumor. But, at present incisional biopsy is used only for diagnosis of the disease and semi-quantitative estimate of malignancy by tumor grading (Bockmuhl *et al.*, 2000). It becomes a necessity for the pathologist to answer whether or not the OSCC carries the potentiality for metastasis, if it is resistant to chemotherapy and/or radiotherapy and the prognosis of the patient (Bockmuhl *et al.*, 2000). Therefore, additional predictors and biomarkers are being extensively investigated with newer technologies with an aim for better patient management and for the provision of other treatment modalities such as radiotherapy and chemotherapy (Bockmuhl *et al.*, 2000; Ludwig and Weinstein, 2005; Mishra and Verma, 2010).

The initial draft human genome sequencing was completed in 2001, where a total of 30,000-35,000 genes were identified (International Human genome sequencing consortium, 2001). Further in 2003, a total of 24,500 coding genes have been estimated in the human genome by International Human Genome Sequencing Consortium (2004). This has prompted the use of molecular biology advancement to a whole new level. It has also initiated the development of high-throughput technologies like array-based comparative genomic hybridization (aCGH), which further enhanced the studies to screen entire genome in a single rapid assay (Marquis-Nicholson *et al.*, 2010). Indeed, high resolution array CGH is an important tool to discover genes that are involved with carcinogenesis and to filter out known alterations in the consensus (well known) regions (Davies *et al.*, 2005). Infact, this tool (aCGH) has allowed researchers to focus on smaller areas in the genome (Davies *et al.*, 2005). Apart from that, high resolution array CGH has increased the chances to detect small novel alterations as compared to lower resolution array CGH where small novel alteration may not be detected (Przybytkowski *et al.*, 2011). This is because of the fact that when resolution is higher, the ability to discover genes associated with cancer are also increased (Davies *et al.*, 2005).

Oral squamous cell carcinoma (OSCC) is known to develop at various anatomical subsites within the oral cavity and therefore it forms heterogeneous tumor groups (Timar *et al.*, 2005; Jarvinen *et al.*, 2006). Due to its complexity, it has been suggested that different anatomical locations of tumor would portray different biological behaviour in terms of invasion and metastasis (Woolgar, 2005). Globally, OSCC of the cheek is more common in South East Asian countries, whereas tongue OSCC is common in Europe and other Western countries (Landis *et al.*, 1998; Diaz *et al.*, 2003; Sathyan *et al.*, 2006). The biological difference between these 2 site of cancer might be due to variation in reflecting the association between risk habits practiced such as betel

quid chewing in India and South East Asia (Werning and Mendenhall, 2007), whereas tobacco and alcohol consumption in the West has been related to mouth and tongue cancers (MacFarlane *et al.*, 1996; Paterson *et al.*, 1996; Moore *et al.*, 2000; Batsakis, 2003; Parkin *et al.*, 2005; Ridge *et al.*, 2007). In Malaysia, majority of the oral cancers are reported to be of tongue and cheek carcinomas (Lim *et al.*, 2008). Both of these cancers are aggressive, but act differently. Tongue SCC is more aggressive and has the propensity to invade, leading to metastasis to the regional lymph nodes (Ridge *et al.*, 2007). But in cheek SCC, the aggressive behaviour is described with higher recurrence rate and spread of tumor that is facilitated by lack of an effective anatomic barrier (Strome *et al.*, 1999; Lee *et al.*, 2005; Lin *et al.*, 2006; Huang *et al.*, 2007). Distinct behaviours between these 2 subtype of OSCC has been further explained in proteomic level by He *et al.* (2004) and Chen *et al.* (2004a) showed that tongue and cheek SCC are involved in different pathways.

Though cancer is considered generally to have only losses of co-operative cell behaviours that normally facilitate the multicellularity (formation of tissues and organs), but they are also characterized by multiple dys-regulated pathways that control elementary cellular processes such as cell growth and cell fate (Kreeger and Lauffenburger, 2010). These dys-regulation of cell signaling pathways are usually resulted from the accumulation of genetic alteration in cancer cells particularly by the disruption of oncogenes and tumor suppressor genes in normal cell regulation (Bild *et al.*, 2006a). In fact, studying a large number of biological pathways may allow the notion to recognize a series of oncogenic pathway signatures that are resulted from genetic instability (Bild *et al.*, 2006b). Therefore, a better understanding of the pathways of various genes involved in carcinogenesis will facilitate the improvement of diagnosis,

therapeutic management and anticancer drug discovery in the war against cancer (Zhao *et al.*, 2009; Liu *et al.*, 2010; Nambiar *et al.*, 2010).

Rationale of the current study:

Unlike Asian countries, implementation of molecular cytogenetic technique in field of cancer research is widely established in Western nations (Varella-Garcia, 2003; Gasparini *et al.*, 2007). Indeed in Malaysia too, these attempts were rarely reported for oral cancer. Embarking on the applications of genome wide screening using higher resolution array CGH, we could expect the identification of highly accurate localization of specific genetic alteration that is associated with tumor progression (Kallioniemi, 2008). This would offer a better understanding on cancer development and may be regarded as an improved tool for clinical management of cancer in the area of developing diagnostic and therapeutic targets (Shinawi and Cheung, 2008).

Previous studies have shown that oral cancer like tongue and cheek can occur in different subsites and can behave differently (Chen *et al.*, 2004a; He *et al.*, 2004). In this perspective, there is a need to perform genome wide screening on exploring the causes of these two different subsites (tongue and cheek). By conducting this attempt, a new set of gene with copy number variations in OSCC could be discovered and as such development of suitable biomarkers for improving the diagnostics can be pursued. Nevertheless, the findings from the present research could be made possible to compare the available databases so far to target the genes that are involved in multiple pathways that are significant for oral cancer. Therefore, in order to fully understand the OSCC behaviour, there is a necessity to look into the chromosomal alterations and gene pathways involved with oral cancers at different sites.

Aim:

The basic aim of the present study is to determine the chromosomal aberrations and gene pathways involved in tongue and cheek SCC using ultra-high resolution array CGH.

Specific objectives of the current research:

- i. To determine the DNA copy number aberrations in tongue and cheek SCC using array CGH technology.
- ii. To identify the genes involved in tongue and cheek SCC using array CGH technology.
- iii. To determine the significant pathways involved in tongue and cheek SCC using pathway analysis software.

2.0 Literature Review

According to World Health Organization (WHO), cancer is defined as a group of disease in which the cells grow abnormally, uncontrollably and enable invasion to others part of the tissue leading to metastasis (<http://www.who.int/en/>). Any cancer occurring within the oral cavity is defined as oral cancer (<http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002030/>).

2.1 Epidemiology of oral cancer:

2.1.1 Incidence

Oral cavity cancer is ranked as the 12th most common malignancy in the world (Parkin *et al.*, 1999; Scully and Bedi, 2000). Oral cavity cancers are one of the most common malignancies found worldwide. There were 263,900 new cases of oral cancer and 128,000 deaths worldwide in 2008 according to Global Burden of Cancer (GLOBACAN) (Jemal *et al.* 2011). In United States of America, oral cancers are accounted for 30,100 cases and 7,800 deaths, representing almost 3% of all cancer (Greenlee *et al.*, 2001). In 2004, around 67,000 new cases of oral cancers were reported in countries of European Union and ranked as 7th most common malignancy (Boyle and Ferlay, 2005). Oral cancer incidence and mortality rates have either been stable or increasing in the last four decades. A sharp increase in oral cancer incidences were seen in countries like Germany, Denmark, Scotland, Central and Eastern Europe, Japan, Australia, New Zealand and also among non-white populations in the United States (Stewart & Kleihues, 2003).

The developing countries account for almost two-thirds of total oral cancer cases and accounts for an average of 200,000 deaths annually (Stewart & Kleihues, 2003). Among the Asian countries, India has the highest incidence of oral cancer with more than 100,000 new cases annually. In Malaysia, the first incidence of oral cancer was reported in 1966 by Hirayama (1966) where the incidence rate of oral and pharyngeal cancer was 3.1 per 100,000 populations. The Malaysian National Cancer Registry (MNCR) has categorized oral cavity cancers into three distinct groups; namely mouth, tongue and lip cancers (Lim *et al.*, 2008). If the incidences of these three cancers (mouth, tongue and lip cancer) were counted as one, the rank of oral cavity cancers will be higher and definitely comparable with incidence rate reported in other studies worldwide (Lim *et al.*, 2008).

2.1.2 Gender, Ethnic and Age distribution

Warnakulasuriya (2009b) reported that oral cancer is more common in men than women with a ratio of 1.5:1 in most countries of the world. Similarly, in South and Southeast Asian countries such as India, Sri Lanka, Pakistan and Taiwan, the incidence of oral cancer is higher in men. In the year 2003 and 2005, the mouth cancer was ranked as the 22nd and 15th most common cancer for males and females, respectively in Malaysia (Lim *et al.*, 2008). During these years, tongue cancer was ranked at 17th and 21st among males and females, respectively.

In USA, oral cancers are reported more frequently in blacks than in whites, ranking 6th most common among blacks and 11th among whites (Day *et al.*, 1993). There were a sharp incidence and mortality rate increment of oral cancer in Germany, Denmark, Scotland, Central and Eastern Europe, Japan, Australia, New Zealand and USA among the non whites (See review; Stewart and Kleihues, 2003). Warnakulasuriya and Johnson

(1996) pointed out that oral cancer existed to be most prevalent in areas with high Asian population such as in India and Taiwan. Interestingly in Malaysia, a marked variation in the incidence of oral cancer was observed among the different ethnic groups that make up the Malaysian population. According to Malaysia National Cancer Registry (MNCR) in the year 2003, oral cancer was ranked at 6th position and 3rd most common cancers for Indian males (ASR=7.2) and females (ASR=16.5). Meanwhile, tongue cancer was ranked as the 9th most common cancer for Indian males (ASR=6.4) and females (ASR=6.8) (Lim *et al.*, 2004). These findings confirm the results of study by Zain and Ghazali (2001), which indicated that ethnic Indian group, has the highest risk for oral cancer among the Malaysian population. This is because of the fact that Indian populations were highly involved in practice such as betel quid chewing as compared to other ethnic groups (Zain and Ghazali, 2001).

Incidence of oral cancer is more common with increasing age in all countries (Warnakulasuriya, 2009b). Approximately 95% of oral cancer cases are reported in people who are older than 40 years old (Warnakulasuriya, 2009b). Based on the database of National Cancer Institute's Surveillance, Epidemiology and the End Results (SEER) from the year 2000-2004, the median age of getting oral cancer was 62 in United States (SEER Cancer Statistics Review, 2009). However, there is a gradual increasing trend in incidence of oral cancer cases and mortality rate in young adults in European countries and United States (Macfarlane *et al.*, 1994; Shiboski *et al.*, 2005). According to the study done by British Dental Association (2000), approximately 6% of oral cancers were diagnosed in younger patients less than 45 years of old. In Indian subcontinent which has one of the highest oral cancer prevalence showed that the oral cancer usually occurs prior to the age of 35 mainly due to the tobacco chewing practice (Johnson, 1991). Besides that, in Sri Lanka, almost 5% of oral cancer is diagnosed in

young patients (Siriwardena *et al.*, 2006). In Malaysia, MNCR reported that there was an exponential increase of oral cancer incidence after the age of 40 years old for both sexes (Lim *et al.*, 2004).

Most of the studies on oral cancer uses the guideline provided by the International Classification of Disease (ICD) which normally include cancer of the lip (C00), tongue (C01-02), gum (C03), floor of the mouth (C04), palate (C-05) and other and unspecified parts of the mouth (C06) (Sugerman and Savage, 1999). But several other studies have excluded lip, salivary glands and other pharyngeal site in their classification of oral cancer (Moore *et al.*, 2000; Zain and Ghazali, 2001). In some other cases, researchers have included the sub sites within the oral cavity such as salivary glands and other pharyngeal sites (ICD-10:C11-C13) (Parkin *et al.*, 1999; Warnakulasuriya, 2009b; Ferlay *et al.*, 2010).

2.2 Clinical and Histological characteristics of oral cancer

2.2.1 Subsites of oral cancer (ICD-10)

Oral cancer is classified according to anatomical subsites of the International Classification of Disease (ICD-10), a coding system that was developed by World Health Organization (WHO) (Johnson, 2003). Accordingly the intra-oral sites corresponding to the ICD-10 code are; C00 (lip), C01 (base of tongue), C02 (other and unspecific parts of tongue), C03 (gum), C04 (floor of mouth), C05 (palates) and C06 (other and unspecific parts of the mouth) (Johnson, 2003).

2.2.2 Clinical appearance

At the initial stages, carcinomas may be asymptomatic (Scully *et al.*, 1986). They may present as indolent ulcers that fail to heal (Scully *et al.*, 1986). In other cases, it may present as an erythroleukoplastic lesion consisting of red and white areas with slight roughness that is well demarcated (Mashberg *et al.*, 1989). The adjacent soft tissues may also show induration (Bagan *et al.*, 2010). In the advanced stages, it may be manifested as characteristic features of malignancy including ulceration, nodularity and fixation to underlying tissues (Scully and Bagan, 2009a). The late stages may present as ulceration with irregular floor and margins. The patients may experience severe pain radiating to ipsilateral ear in late stages (Bagan *et al.*, 2010). Bagan *et al.* (2010) also described in the advanced stages the oral cancer may present as exophytic tumor with warty surfaces and poorly defined boundaries.

2.2.3 Histological appearance

More than 90% of oral cancer is oral squamous cell carcinoma (OSCC) where it is derived from the surface epithelium and extends from the basement membrane which would replace the normal epithelium. Since basement membranes are being penetrated, the carcinoma will invade the underlying connective tissues (Nagpal and Das, 2003; Silverman, 2003). Pindborg *et al.* (1997) defined OSCC as a malignant epithelial neoplasm that exhibits squamous differentiation as characterized by the formation of keratin and/or the presence of intercellular bridges. OSCC can be graded histopathologically into well, moderate and poorly differentiated lesions. These grading is based on the method originally as described by Broder's. This method takes into account the subjective assessment of the degree of keratinization cellular and nuclear pleomorphism and mitotic activities (Pindborg *et al.*, 1997).

2.3 Etiological factors:

Oral cancer is usually related to the chronic exposure of oral mucosa to a range of mutagens like chemical, physical and microbial agents which results in DNA mutation of oral keratinocytes (Zain and Ghazali, 2001; Scully and Bagan 2009b). Well documented etiological factors like tobacco smoking, excessive alcohol consumption and betel quid chewing are known to cause oral cancers either independently or synergistically (Blot *et al.*, 1988; Zain and Ghazali, 2001; Warnakulasuriya, 2009b).

2.3.1 Tobacco smoking

Cigarette smoking has been identified as an independent risk factor for oral cancer and all forms of tobacco including smokeless tobacco have strongly been associated with oral cancers (International Agency for Research on Cancer, 1986; Rodu and Jansson, 2004). Worldwide, the risk of getting oral cancer in smokers is 7-10 times higher than that for non-smokers (Warnakulasuriya *et al.*, 2005). Case control study of Rodriguez *et al.* (2004) showed that heavy smokers have an odd ratio of 20.7 of getting oral cancers. Besides that, another study conducted by Neville and Day (2002) showed that smokers have five to nine times higher chances to develop oral cancer compared to non smokers.

The reasons to cause oral cancer by tobacco smoking can be explained by the fact that there are more than 300 carcinogens found in tobacco smoke. Some of the most important carcinogens are polycyclic aromatic hydrocarbons (PAH), benzo- α -pyrene, tobacco specific nitrosamines including nitroso-nor-nicotine (NNN) and 4-(methylnitrosoamino)-1-(3-pyridyl)-1 butanone (NNK) that exert as DNA adducts and attack oral mucosa epithelium leading to chromosomal damage and DNA mutation. For example: change of guanidine to thymidine transversions (Hecht, 2003; IARC, 2004).

Recently, few studies performed by using genome wide screening was conducted on oral cancer patients. One study reported frequent deletion of CDKN2A (p16^{INK4}) in smokers for both lung and oral cancers (O'Regan *et al.*, 2006) when compared to non-smokers which suggested that this tumor suppressor gene play a crucial role in cancer formation to smokers. Besides that, Chattopadhyay *et al.* (2010) reported the synergistic effect of the tobacco and betel quid chewing resulted in chromosomal aberrations in esophageal squamous cell carcinoma (ESCC) such as amplification of chromosomes in 1p, 1q, 2q, 3q, 5p, 6p, 8q, 9q, 11p, 11q, 15q and deletion of chromosome in 3p, 8p, 9p, 13q, 18q.

2.3.2 Excessive alcohol consumption

Regular and excessive alcohol consumption proved to increase the risk of oral cancer as it accounts for 7-19% of cases worldwide (Petti, 2009). A review paper by Petti (2009) showed that moderate alcohol drinker have 2-3 fold higher for getting oral cancer compared to non-drinkers. The association of alcohol intake and tobacco smoking for the causage of cancer has been documented for more than 50 years ago (Rothman and Keller, 1972; Blot *et al.*, 1988). The risk is higher by 48 fold with synergistic effect between these 2 independent risk factors on oral cancer development, which accounts for more than 75% of the oral cancer cases reported in developed countries (IARC, 1990; Rodriguez *et al.*, 2004). This can be explained biologically by the ability of alcohol to remove lipid content on the membrane and causing it to be more permeable for tobacco carcinogen which exert as DNA adduct in oral mucosa and promote oral carcinogenesis (Ogden and Wight, 1998; Wight and Ogden, 1998).

Studies of Garro *et al.* (1986) and Mufti (1992) have demonstrated the mutagenic effect of excessive alcohol consumption that has shown to increase the frequency of chromatid breakage in DNA through an *in vivo* study. This suggests that alcohol have the ability to dys-regulate the DNA repair mechanism when DNA is mutated. Besides that, it was also reported that alcoholic drinker have 5-9 fold tendency to get tongue SCC (Herity *et al.*, 1981). In fact, alcoholic drinkers in the Asian populations have a higher tendency to get oral cancers. This may be due to the fact that the alcoholic drinkers have inactive aldehyde dehydrogenase enzyme. This enzyme may have the ability to detoxify the acetaldehyde, a known carcinogenic agent (DNA adducts) for oral mucosa (Petti 2009).

2.3.3 Betel quid chewing

Apart from the two well documented risk habits which are tobacco smoking and excessive alcohol consumption, studies from South Asian countries especially from India and Taiwan also showed that betel quid chewing is an important risk factor to increase the incidence of oral cancers, since it is widely practiced in these regions (Lin *et al.*, 2000; Balaram *et al.*, 2002; Gupta and Ray, 2004; Wen *et al.*, 2005).

Studies done by Daftary *et al.* (1991) and Henderson and Aiken (1979) reported that the risk of getting oral cancer associated with betel quid containing areca nut and tobacco was 8-15 times higher than using quid without tobacco which is lesser by 1-4 times. The betel quid chewing exposes the oral mucosa to a highly reactive oxygen species (ROS) (Nair *et al.*, 1987). The carcinogenic ROS and DNA adduct induce genetic instability and initiate tumorigenesis by causing the structural changes in oral mucosa which favour other betel quid compounds to penetrate the oral mucosa (Nair *et al.*, 1985).

The development of OSCC due to chromosomal instability associated with betel quid and tobacco is largely determined by the gains in chromosomes 8q, 9q, 11q, 17q, and 20q and most frequent losses are reported in chromosome arms at 3p which normally involves the inactivation of tumor suppressor genes like FHIT, RAR β and VHL. Other deleted regions detected from the CGH analysis on oral cancer patients with betel quid chewing are 4q, 5q, 9p21-23, and 18q (Lin *et al.*, 2001).

2.3.4 Human Papillomaviruses Virus (HPV)

Previously the occurrence of human papillomaviruses (HPV) in OSCC biopsy was detected using Southern Blot analysis (Villiers *et al.*, 1985). According to zur Hausen (1996), HPV is the most common virus studied in relation to the head and neck tumor. Studies on HPV incidence and oral cancer risk exhibited a wide range which varied between 0 and 100% (Kozomara *et al.*, 2005). The most common and high risk type HPVs found are HPV16 and 18 which encodes oncoproteins E6 and E7. These oncoproteins will bind to p53 and pRb, thus inactivating the tumor suppressor genes. These genes are normally involved in turning off the cell division of those cells with DNA damage. This condition could lead to genomic instability and accumulation of genetic changes, thus contributing to the malignant progression (Wilczynski *et al.*, 1998).

A case control study by International Agency for Research on Cancer (IACR) have identified DNA of HPV in 3.9% of oral cavity and 18.3% of oral premalignant cancer (Herrero *et al.*, 2003). In a Japanese population of 46 OSCC patients, 73.9% of the samples were found to have infected with HPV (Shima *et al.*, 2000). Among them 26.5% were of HPV type-16 and 73.5% were HPV type-18. Study done by Klusmann *et al.* (2009) using CGH analysis on 28 HPV related OSCC revealed that only one third

of them were showing gain in chromosome 3q26.3 region. Moreover, this study also showed that there was a high occurrence of chromosomal 16q deletion with HPV positive OSCC, which suggested that the FRA16D gene harboured in this region may increase the risk of malignancy in oral cancer (Klussmann *et al.*, 2009).

2.3.5 Genetic Susceptibility

The genetic risk for head and neck cancer has always been reported to be associated with polymorphisms of drug-metabolizing enzymes (Hahn *et al.*, 2002). This might influence an individual's susceptibility to chemical carcinogenesis such as cytochromes P450 (CYPs) which metabolizes polycyclic aromatic hydrocarbons (PAHs) and glutathione S-transferases (GSTs) that are involved in the detoxification of activated metabolites of carcinogens (Hahn *et al.*, 2002). Nevertheless, the three most studied genes for polymorphism in relation to oral cancers are GSTM1, GSTT1 and CYP1A1.

In a study conducted among Indian patients with GSTM1 null (deletion) genotype were found to have an OR of 1.3 (95% CI 0.37-4.82) (Sreelekha *et al.*, 2001). Similarly among the Japanese patients, a significant 2.2-fold increased risk (95%CI 1.4-3.6) was found for individuals with null genotype (Sato *et al.*, 1999). Another study performed among the Thais also found a 2.6-fold higher risk (95% CI 1.04-6.5) (Kietthubthew *et al.*, 2001). However, studies among the Western population found the other way around. A study conducted at France by Jourenkova-Mironova *et al.* (1999) found that there was no association between GSTM1 null genotype with oropharyngeal cancer risk (OR 0.9, 95% CI 0.5-1.5).

GSTT1 null were also found to confer an increased risk of 2.5 (95%CI 0.28-21.71) among the Indians (Sreelekha *et al.*, 2001). However, Katoh *et al.* (1999) in their study among the Japanese found no association among the null genotype with oral cancer (OR 0.68, 95% CI 0.38-1.22) and this finding is supported by Kietthubthew *et al.* (2001) in their study among oral cancer patients in Thailand.

CYP1A1 polymorphism were found to confer an OR of 5.3 (95%CI 1.03-26.28) among the Indians (Sreelekha *et al.*, 2001) which was supported by Park *et al.* (1997) in their study among Caucasian oral cancer patients where a significant 2.6-fold increased risk (95% CI 1.2-5.7) was found. This finding is further supported by Sato *et al.* (1999) who found a significant 2.3-fold increased risk (95% CI 1.1-4.7) in their study among Japanese oral cancer patients.

However, a preliminary study on 81 Jakarta oral cancer patients showed a lack of evidence to support any association between polymorphisms of GSTM1, GSTT1 or CYP1A1 with oral cancer occurrence (Amtha *et al.*, 2009).

2.3.6 Diet and Nutrition

Researchers have attempted to investigate the relationship between dietary intake and the occurrence of oral cancers since 1977 (Graham *et al.*, 1977). In mid of 1990's, Winn (1995) reported that dietary and nutritional factors have implication in oral carcinogenesis. Studies by La Vecchia *et al.* (1997) estimated that approximately 15% of oral cancer cases were caused by imbalances or deficiencies dietary in European population. According to the study done by Petridou *et al.* (2002), the vegetables, fruits, micronutrient, dairy products and olive oil can play a vital role in protecting against oral cancer through high consumption of riboflavin, iron and magnesium. Berger *et al.*

(1991) has suggested that Vitamin-A could act as a potential protective source from carcinogenesis and lack of them in the diet can lead to metastasis. According to these Berger and his team, Vitamin-A has the ability to inhibit DNA synthesis and can influence the epidermal growth factor receptors which are over expressed and can cause tumor cell proliferation through the effect of protein kinase C. Another study by Garewal (1995) mentioned that antioxidant nutrients such as β -carotene and vitamin-E could play a vital role against oral cancers based on the evidence from animal model systems and laboratory studies.

2.3.7 Mouthwash

Earlier studies showed that there are some controversies arising from the use of alcohol containing mouthwashes and oral cancer (See review Warnakulasuriya, 2009a). For example, a study conducted by Winn *et al.* (1991) showed that frequent use of mouthwash containing alcohol with concentration of greater than 25% had a higher risk to get oral and pharyngeal cancer. Similarly, McCullough and Farah (2008) showed that alcohol containing mouthwashes contributes to oral carcinogenesis. However, a recent meta-analysis conducted by La Vecchia (2009) confirmed that there are no excess risks involved in getting oral cancer either using mouthwash with or without ethanol.

2.4 Genetic Alteration

Genetic instability is a hallmark of most cancers. Chromosomal instability is characterized by the changes in chromosomal structure (inversions, point mutations and translocations) (Albertson *et al.*, 2003; Perera and Bapat, 2007) and numerical (amplification and deletions) where these events are able to cause uncontrolled cell proliferation, altered cell morphology, and tumor progression (Lengauer *et al.*, 1998; Gollin, 2004). Though there are various forms of genomic instabilities, chromosomal instability (CIN) and microsatellite instability (MIN) are the common ones to be encountered in cancer (Loeb, 2001; Negrini *et al.*, 2010).

There are two categories of structural aberrations, namely balanced (reciprocal) and unbalanced (non-reciprocal). Reciprocal alteration is the exchange of chromosome parts between non-homologous chromosomes and no genetic content is lost or gained (Albertson *et al.*, 2003). However, in non-reciprocal alteration, the exchange is unequal, resulting in extra or missing copies of genes and chromosomes regions (Albertson *et al.*, 2003). Another type of genomic instability is known as MIN, which refers to the base pair mutations that are caused by the defection of the mismatch, base excision and nucleotide excision repair genes at the nucleotide level. These alterations happening would enable to cause expansion or contraction of the number of oligonucleotide repeats at the microsatellite sequences (Lengauer *et al.*, 1998).

Epigenetic alteration involves modification of DNA transcription *via* DNA methylation and chromatin components such as histones (Baylin and Ohm, 2006). These factors do not change the DNA code, but they make accessibility of DNA for transcription which leads to the transcriptional silencing of tumor suppressor genes (Esteller and Herman, 2002). Nevertheless, the gene expression aberrations that arise due to genetic and epigenetic changes allow cells to have selective growth advantage and results in uncontrolled tumor growth (Baylin and Ohm, 2006; Jones and Baylin, 2007).

2.5 Chromosomal Instability (CIN)

Chromosomal instability (CIN) is regarded as “mutator phenotype” and is demonstrated by the cells that have increased amount of unstable chromosome content which arise from the abnormal mitosis as compared to their normal counterparts (Loeb, 2001; Chi and Jeang, 2008). The features of the unstable content include DNA translocation, aneuploidy (loss/gain of whole or portions of chromosome), changes in gene copy number and chromosomal rearrangement (Lobo, 2008).

All human cancer types undergo chromosomal and genetic aberrations, including OSCC (Reshmi and Gollin, 2005). It was established that the role of chromosomal instability contribute to tumor initiation and progression (Michoret *et al.* 2005). After his observation on chromosomal aberrations in tumor cells, Michor *et al.* (2005) suggested CIN as the central issue in cancer biology. The chromosomal aberrations will hijack the normal cellular processes; such as cell signaling, replication, and apoptosis, causing an uncontrollable cell proliferations and defects in DNA repair mechanisms which are responsible for tumorigenesis (Loeb *et al.*, 2003). Studies have been reported that over-expression of aurora kinase A-(AURKA) gene in cancer may lead to failure in maintaining the stable chromosomal contents due to the defective centrosome

maturation, bipolar spindle assembly and mitotic entry (Katayama *et al.*, 2003; Vader and Lens, 2008).

Some of the defective functioning of mitotic checkpoint machinery such as centrosomes, microtubules, kinetochores, loss of spindle check point, abnormalities of double strand break repair and telomere dysfunction in mitosis have been shown to alter chromosome number and its structure which contributes to chromosomal instability (Reshmi and Gollin, 2005; Chi and Jeang, 2008; Schwartzman *et al.*, 2010). This defective mitotic machinery is mainly due to the failure in separation of sister chromatids, before microtubule attachment during cell division (Schvartzman *et al.*, 2010).

The involvement of extrinsic cytoskeletal aberrations such as multipolar spindles (Saunders *et al.*, 2000) and alterations in centrosome number (Gisselsson *et al.*, 2002) which enabled to induce CIN were detected in few oral cancer studies. For example, Saunders *et al.* (2000) observed various degrees of multipolar spindles from OSCC cell line such as different level of chromosomal capture and alignment which enabled to induce CIN. Another study by Gisselsson *et al.* (2002) has shown the alterations of centrosome number in oral cancer where they suggested that the prevention of cytokinesis caused by the internuclear connection has lead to duplication of both chromosome and centrosome number.

Another factor that promotes CIN is increased expression of centrosomal protein during cell divisions which are actually regulated by checkpoint kinase such as CHK1 and CHK2. This phenomenon is caused by the over duplication of centrosomes during a single cell cycle and the failure of cells to undergo proper cytokinesis will lead to excessive number of centrosomes (Chi and Jeang, 2008). These genetic changes could contribute to different phenotypes that can result in increased malignancy and more aggressive behaviour (Weinstein, 2002; Chi and Jeang, 2008).

2.6 Oncogene and Tumor Suppressor Gene

In tumorigenesis, gene amplification is considered as an important step involved in oncogene activation (Coleman and Tsongalis, 2002) and loss of chromosomal materials will most likely harbour the inactivated tumor suppressor gene (TSG) (Frohling and Dohner, 2008). A plethora of genetic events leading to activation and inactivation of oncogenes and TSG, respectively are involved in the pathogenesis of OSCC (Argiris *et al.*, 2008).

2.6.1 Oncogenes

In general, oncogenes are derived from gain of functions in cellular proto-oncogenes. This is occurred through the alteration of few common mechanisms such as mutation, chromosomal rearrangement (translocations and inversions), and gene amplifications (Croce, 2008). Oncogenes are broadly classified into six functional groups such as transcription factors, growth factors, growth factor receptors, chromatin remodelers, signal transducers and apoptosis regulator (anti-apoptotic regulator) (Croce, 2008). Amplification of oncogenes are usually observed during tumor progression which is enhanced by certain encoded proteins through checkpoints dys-regulation of the cell cycle (Field, 1995). Several oncogenes have been implicated for oral carcinogenesis

such as c-myc, int-2, hst-1, cyclin D1 and EGFR (Todd *et al.*, 1997). Some studies have indicated that the epidermal growth factor receptor gene (EGFR), were over-expressed and amplified in oral cancers (Ishitoya *et al.*, 1989; Chen *et al.*, 2003). Epidermal growth factor receptor genes are involved in transmembrane tyrosine-specific phosphokinase activity in normal condition. The binding of EGF receptors to their ligands lead to phosphorylation of tyrosine residues EGFR to dimerization and activation. These would then activate two important kinase pathways and favours growth, invasion, metastasis and angiogenesis in oral cancer (Ciardiello *et al.*, 2004; Yano *et al.*, 2003; Kalyankrishna and Grandis, 2006).

Another oncogene that is implicated in oral cancer is ERBB2 gene (Xia *et al.*, 1997). This gene is activated through heterodimerization with other receptors for activating the mitogen activated protein kinase (MAPK) or phosphatidylinositol 3-kinase (PI3K)-AKT activated pathways. This gene will enhance tumor invasion, cell proliferation, differentiation, adhesion and cell migration in cancer (Olayioye *et al.*, 2000). Callender *et al.* (1994) studies reported an amplification of Cyclin D1 which was associated with tumor aggressiveness and is a late event for head and neck carcinoma. The expression of Cyclin D1 which normally plays an important role as cell proliferation promoter in G1 phase of the cell cycle. However, when over expressed, Cyclin D1 has enhanced the cancer progression in tumorigenesis (Hunter and Pines, 1991). The importance of targeting oncogenes in molecular targeted cancer treatment was reported by Heinrich *et al.* (2002) and Casali and Messina (2004) in their study. In this study the investigators used Imatinib to affect/inhibit the KIT and PDGFR receptor kinases in chronic myelogenous leukemia.

2.6.2 Tumor Suppressor Genes

Conversely, inactivation or deletion of tumor suppressor genes (TSG) or anti-oncogenes is responsible for initiating cancer progression (Mitelman, 2005). This is because tumor suppressor genes are negative growth regulators that regulate cellular trafficking, regulation of DNA damage response and apoptosis in a recessive fashion (Weinberg, 1991). The definition of TSG is still evolving and Weinberg (1991) defined that TSG is a genetic element whose loss or inactivation allows a cell to display one or the other phenotype of neoplastic growth deregulation. The loss of function of TSG is oncogenic and has been documented to be the major event during carcinogenesis (Munger, 2002). It is common that there are few inactivated TSG in same tumors and the same suppressors can be found inactivated in different tumor types such as lung, breast and colon cancer (Sager, 1989). Loss of these genes promotes the cell cycle proliferation, signal transduction, angiogenesis and tumor growth (Sager, 1989). TSG are recessive which requires two mutated alleles for tumor formation and it was reported in a two-hit mechanism, proposed by Knudson in the early 1970 (Knudson, 1971). The two hit theory stated that both the alleles needed to be inactivated in order to promote malignant growth (Knudson, 2001).

Kinzler and Vogelstein have categorized tumor suppressor genes into 2 groups which were 'gatekeepers'/'caretakers' (Kinzler and Vogelstein, 1997) and 'landscaper' (Kinzler and Vogelstein, 1998). The gatekeeper genes like RB1 and TP53 function to inhibit the tumor growth, suppress neoplasia and promote cell death. The losses of function of these genes are rate-limiting in multi-stage tumorigenesis (Kinzler and Vogelstein, 1997). By contrast, 'caretaker' tumor suppressor genes such as BRCA1 and BRCA2 work as DNA maintenance genes to suppress growth by ensuring the fidelity of the DNA code through effective repair of DNA damage and maintain genome integrity

(Kinzler and Vogelstein, 1997; Russo *et al.*, 2006). Landscaper's actually work through a less direct mechanism by affecting tumor microenvironment in which tumor cell grow such as extracellular matrix (ECM) protein, cell surface markers, adhesion proteins and survival factors. Indeed, loss of function of landscaper TSG would lead to malignancies (Kinzler and Vogelstein, 1998). For example, loss of E-cadherin and alpha-catenin in epithelial cell junction and epithelial stromal cell interaction induce epithelial mesenchyme transition in carcinoma progression (Thiery, 2002).

Reports suggest that loss of function of FHIT, RB1, TP53 and CDKN2A has been implicated in head and neck SCC (Virgilio *et al.*, 1996; Koontongkaew *et al.*, 2000; Nakahara *et al.*, 2000). Biallelic inactivation of a specific TSG can occur through chromosomal allelic loss such as loss of heterozygosity (LOH)-allelic deletion, point mutation and deletion of both alleles (Oesterreich and Fuqua, 1999; Vogelstein and Kinzler, 1993; Knudson, 2001; Oster *et al.*, 2005). Both alleles of a TSG need to be inactivated in order to unmask recessive mutations which have effect on cell phenotype (Yokota and Sugimura, 1993). Inactivation of TSGs through epigenetic mechanism decreases the gene expression without affecting the DNA sequence. In the epigenetic silencing mechanism, CpG islands located in the promoter regions is hypermethylated and reduces the TSGs expression (Jain, 2003). For example, Yeh *et al.* (2003) have demonstrated few hypermethylated regions containing TSGs such as CDKN2B, CDKN2A and TP53 related to oral cancer. In addition to two-hit hypothesis, a new concept of haploinsufficiency has been recently proposed (Macleod, 2000). This concept is related to a gene dosage effect whereby a copy-loss of a potential TSG could have an impact on cell phenotype (Fodde and Smits, 2002).

Levine (1997) described the inactivation of TSGs which leads to the evasion of cell from the tight cell cycle regulation and predispose to the uncontrolled cell growth and cell division which contributes to the malignant phenotype of cancer. Thus, loss of TSGs may be predictive of patient outcome (Gleich and Salamone, 2002).

2.7 Carcinogenesis

In carcinogenesis, genetic instability could influence the certain enzymes that accelerate the pathogenic pathways which are involved in DNA replication and repairing, chromosomal instability, apoptosis and cell cycle regulation in response to DNA damage (Beckman and Loeb, 2005). This can be seen in oral carcinogenesis, which results from multistep process involving the accumulation of genetic alterations including chromosomal aberrations, DNA mutations and epigenetic alterations (Scully *et al.*, 2000; Choi and Myer, 2008). These events could disrupt the normal cellular processes and may lead to abnormal amplification of centrosomes, defects in DNA repair mechanism, uncontrolled cell proliferation and reduction of apoptosis (Weinstein, 2002; Chi and Jeang, 2008). This condition will contribute to different phenotypic changes that can result in elevated malignancy and more aggressive behaviour.

Cancer is a group of disease that are characterized by unregulated DNA replication, cell growth, cell division and survival which differs from their normal cell counterparts (Evan and Vousden, 2001; McSharry, 2001). Generally, cancer is the condition where there is a loss of co-operative cell behaviours that normally facilitate multicellularity which includes the formation of tissues and organs (Abott *et al.*, 2006). Cancer cells become ‘deaf’ to the usual controls on proliferation and follow their own agenda for reproduction which is beyond the constraints of normal cells (Weinberg, 1996). It has been postulated that three main steps are needed for the development of cancer:

Initiation, Promotion and Progression (Farber, 1984; Farber, 2003). The initiation process starts when there are irreversible cellular changes or mutations causing DNA damages, which are arisen either spontaneously or induced by exposure to a specific carcinogen. During the initiation phase, the initiated cells are not characterized as tumor cells due to the fact that they have not yet acquired the autonomy of growth and the DNA instability may remain undetected throughout the life (Okey *et al.*, 2005). For promotion stage, it is general characterized that when there is further unchecked proliferation of these mutated cells from the initiation phase, causes a faster increase in tumor size (Farber, 1984). This process can be enhanced by chronic exposure to carcinogenic stimuli which causes the changes of an initiated cell leading to neoplastic transformation to favor tumor growth (Okey *et al.*, 2005). The final stage of cancer development is the progression stage where successive mutations will give rise to increasingly malignant sub-populations. In this stage, the pre-neoplastic cells are transformed to a state in which they are more committed to malignant development. This process involves accumulation of further gene mutations leading to heterogeneity in cell population (Faber, 1984). As the tumor progression advances, the cells lose their adherence property, detach from the tumor mass and invade the neighboring tissues. The detached cells will also enter the circulating blood and lymph which are then transported to other tissues or organs away from the primary sites, which subsequently grows into secondary tumors and forms the distant metastases, resulting in wide spread tumors (Okey *et al.*, 2005).

2.8 Hallmarks of Cancer

Hanahan and Weinberg (2000) have proposed six hallmarks for distinguishing cancer cells from their normal counterpart: (a) self-sufficiency of growth signals; (b) insensitivity to growth-inhibitory signals; (c) evasion of programmed cell death; (d) immortality or unlimited replicative potential; (e) sustained angiogenesis, and (f) tissue invasion and metastasis.

2.8.1 Self-sufficiency of growth signals

Normal cells require growth signals which are soluble and membrane bound growth factors, to switch the cells from quiescent state into a proliferative state (Hanahan and Weinberg, 2000). These growth signals are ligands of receptors that are transduced from cell surface receptor to activate the specific intracellular signaling pathways which increase the cell proliferation activities (Aaronson, 1991; Hanahan and Weinberg, 2000). Dys-regulation of these growth signals and increase in their expression cause the tumor cell to become hyper-responsive and thus become an important driver of self-sufficiency growth in oral cancer (Todd *et al.*, 1991). For example, EGFR receptor can be amplified and over-expressed through the most common truncation mutation (Roger *et al.*, 2005; Kalyankrishna and Grandis, 2006). Grandis and Tweardy (1993) showed that over expression of the EGFR and its ligand along with transforming growth factor alpha (TGF- α) could play a critical role in oral carcinogenesis. The binding of TGF- α to EGFR receptor causes cascade of intracellular signaling events such as cell proliferation, survival, invasion, metastasis and angiogenesis (Reuter *et al.*, 2007).

2.8.2 Insensitivity to growth-inhibitory signals

Hunter and Pines (1994) indicated that loss of expression of the tumor suppressor genes which encodes cell cycle inhibitory protein will enable to increase the cell proliferations that are mainly controlled by the growth inhibitory signals such as p53 and pRB. In cancer, this particular hallmark can be seen through the dys-regulated interactions between growth inhibitory signals and cyclin-dependent kinase (CDK), which increase the progression of the cell cycle (Serrano *et al.*, 1993). Few earlier studies (Pavelic *et al.*, 1996; Pande *et al.*, 1998; Xu *et al.*, 1998) have reported that low expression of pRB genes were detected in oral cancer. This indicated that loss of pRB gene have disrupted the cell cycle progression by increased expression of transcription factor like E2F that promote cell cycling. Other growth inhibitory signals such as p21 and p16 are also been involved in cell cycle inhibition by interacting with CDK that ultimately stop the cell cycle progression (Serrano *et al.*, 1993). These reports were further supported by study done by Sartor *et al.* (1999) which revealed that the loss of p16 permits the cell cycle to progress uncontrollably due to binding of p16 with CDK4 and CDK6 which able to inhibit the cell from entering into S phase during the cell cycle.

2.8.3 Evasion of programmed cell death

Hanahan and Weinberg (2000) revealed that the ability to induce tumor growth is not only caused by uncontrolled cell proliferation mechanism but also the ability of tumor cells to eliminate them to become senescence *via* a process called apoptosis. Oren (1992) and Manning and Patierno (1996) described that cancer cell differs from their normal cell counterparts by increased survival and evades apoptosis. This phenomenon has been achieved in different ways in oral cancer and one such form is over-expression of Bcl-2 which acts as anti-apoptotic regulatory protein in cancer (Pezzella *et al.*, 1993). Study conducted by Oliver and his colleagues (Oliver *et al.*, 2004) suggested that Bcl-

xL inhibition can be an effective treatment to solve OSCC patients who were cisplatin resistance.

2.8.4 Immortality or unlimited replicative potential

In normal condition, telomere shortens after each cell cycle and this limits the life span of the cells (Hanahan and Weinberg, 2000; Snustad and Simmons, 2003). Cancer cells have an infinite lifespan (immortal) due to the ability to replicate indefinitely by increasing the length of their telomeres (Hayflick, 1997). The up regulation of telomerase which is an enzyme to protect against telomere shortening in cancer helps increase the life span simply by extending the end of telomere *via* reverse transcription (Shay and Wright, 2006). Several studies have found that human telomerase catalytic subunit gene (hTERT) were over expressed in oral cancer which was associated with poor prognosis (Kannan *et al.*, 1997; Gordon *et al.*, 2003; Lee *et al.*, 2001a ; Chen *et al.*, 2007a).

2.8.5 Sustained angiogenesis

Folkman (1990) proposed that the onset of angiogenesis was associated with tumor growth and metastasis. This is because tumors grow well in the presence of good blood supply that feeds them with nutrients through their angiogenic mechanism (Folkman, 2006). In 1996, Hanahan and Folkman (1996) suggested that angiogenic switch is dependent on the balance mechanism between pro-angiogenic signals and anti-angiogenic signals that switch from the anti-angiogenic state to pro-angiogenic state to form new blood vessels in cancer tissues. Angiogenic switching has been very well reported with tumor tissue in a study done by Udagawa *et al.* (2002) which showed there was an expansion of tumor in mice after increase in the level of VEGF through transfection method. In oral cancer, the expression of VEGF was reported higher than

normal oral mucosa tissue (Denhart *et al.*, 1997). This suggests that angiogenesis of tumor tissues of oral mucosa are correlated with tumor progression and aggressiveness.

2.8.6 Tissue Invasion and Metastasis

On top of all vital hallmarks of cancer, the most important characteristic that differ cancer cells from their normal counterpart is the ability to perform metastasis and invade into surrounding tissues. Indeed, metastasis is the leading cause of death in cancer (Liotta, 1986; Sporn, 1996; Hanahan and Weinberg, 2000). Unfortunately the factors that influence invasion and metastasis are not fully understood. However, Liotta (1986) indicated that cancer cells must interact with matrix at many stages in its metastatic cascade. This is a very complex process, which involves intravasation through angiogenesis mechanism, extravasation and growth into the surrounding tissues, survival in the bloodstream, stops in a new organ (secondary organ), cytoskeleton remodeling, initiation and maintenance of growth and neo-angiogenesis in the metastatic tumor (Chamber *et al.*, 2002).

Paget (1889) had proposed the concept of “seed and soil” where the growths of the cancer cells were dependent on the secondary organ. However, this pattern of metastasis was challenged by Ewing (1928), where he suggested that organ specific metastasis was mainly caused by circulatory pattern between the primary tumor and secondary organs. Later, Thiery (2002) proposed that metastasis was associated with epithelial-mesenchymal transition (EMT). This proposal was earlier been supported by the study of Gumbiner (1996) which showed that loss of E-cadherin expression during EMT leads to increased cell motility and invasion.

In oral cancer, in order for the cancer cell to invade and metastases, they need to degrade the basement membrane and the extracellular matrix. This degradation is achieved through the active extracellular protease such as matrix metalloproteinases (MMPs) and they may also alter several classes of proteins which are involved to couple with their surroundings such as cell adhesion molecules (CAMs), integrins, as well as loss of E-Cadherin protein in the basement membrane (Choi and Myer, 2008). For example, decrease expression of E-cadherin has been associated with lymph node metastasis in oral cancer (Diniz-Freitas *et al.*, 2006). Besides that, many other studies have showed that several of the MMP genes such as MMP-1, -3, -2,-9, -10,-11 and -13 were expressed in OSCC, which suggests that they play a role in the pathogenesis of OSCC (Polette *et al.*, 1991; Gray *et al.*, 1992; Polette *et al.*, 1993; Kusakawa *et al.*, 1993; Muller, *et al.*, 1993). This is in agreement with Jones and Walker (1997) and Chamber *et al.* (2002) studies which demonstrated that MMP expression were highly correlated with aggressive tumor such as oral cancer. This is because of the fact that for the tumor cell to disseminate, extracellular matrix (ECM) must be degraded through the proteolytic degradation mechanism mainly ruled by MMPs families (Thomas *et al.*, 1999). As metastasis is a hidden process which occurs inside the body, understanding its mechanism becomes a necessity for diagnosis and therapeutic approaches (Chamber *et al.*, 2002).

2.9 Model of Oral Squamous Cell Carcinoma (OSCC)

In 1996, Califano and Sidransky groups conducted some studies to correlate the genetic alterations and histological stages to establish the sequence of progression for specific genetic alterations in head and neck cancer (Califano *et al.*, 1996). This study employed loss of heterozygosity (LOH) assay using microsatellite markers on chromosome 3p, 4q, 6p, 8p, 8q, 9p, 11q13, 13q21, 14q and 17p13 to analyze 87 oral premalignant lesions which included 35 hyperplasias, 31 dysplasias and 21 carcinoma insitu (CIS) as well as 30 invasive tissues. The results showed that there were a high frequency of LOH occurred in benign squamous hyperplastic lesion at chromosome 9p21 (20%), 3p21(16%) and 17p13 (11%). Hence they concluded that these chromosomal aberrations were initiated at early events in head and neck cancer progression (Califano *et al.*, 1996). The chromosomal aberrations (loss of chromosome 9p21, 3p21 and 17p13) were seen frequently in benign squamous hyperplastic lesion and their frequencies of losses were showed an increasing progression from benign hyperplasia to dysplasia lesion (Califano *et al.*, 1996). Subsequent genetic alterations were observed from dysplasia to CIS stage where occurrences of LOH were detected at chromosome 11q13, 13q21 and 14q31 (Califano *et al.*, 1996). In the later event from carcinoma in situ to invasive cancer, LOH were detected at chromosome 6p, 8p, 8q and 4q26-28 (Califano *et al.*, 1996). The important implication from the genetic progression models will allow us to understand better, the biological behaviour and progression of oral carcinogenesis, based on the identification of targeted TSGs and oncogenes in the multifocality of oral carcinogenesis (Califano *et al.*, 1996).

Several studies have suggested that oral carcinogenesis are developed through different histopathological stages, beginning from premalignant, through carcinoma in situ and ultimately to invasive cancer which is highly related to dysplasia-carcinoma sequence (Sidransky, 1995; Califano *et al.*, 1996; Garnis *et al.*, 2004a). To reconcile this model, Califano and colleagues have proposed that oral carcinogenesis is resulted from multistep process which includes chromosomal aberrations, DNA mutations and epigenetic alterations (Califano *et al.*, 1996). The hypothetical model of oral carcinogenesis is illustrated in Figure I.

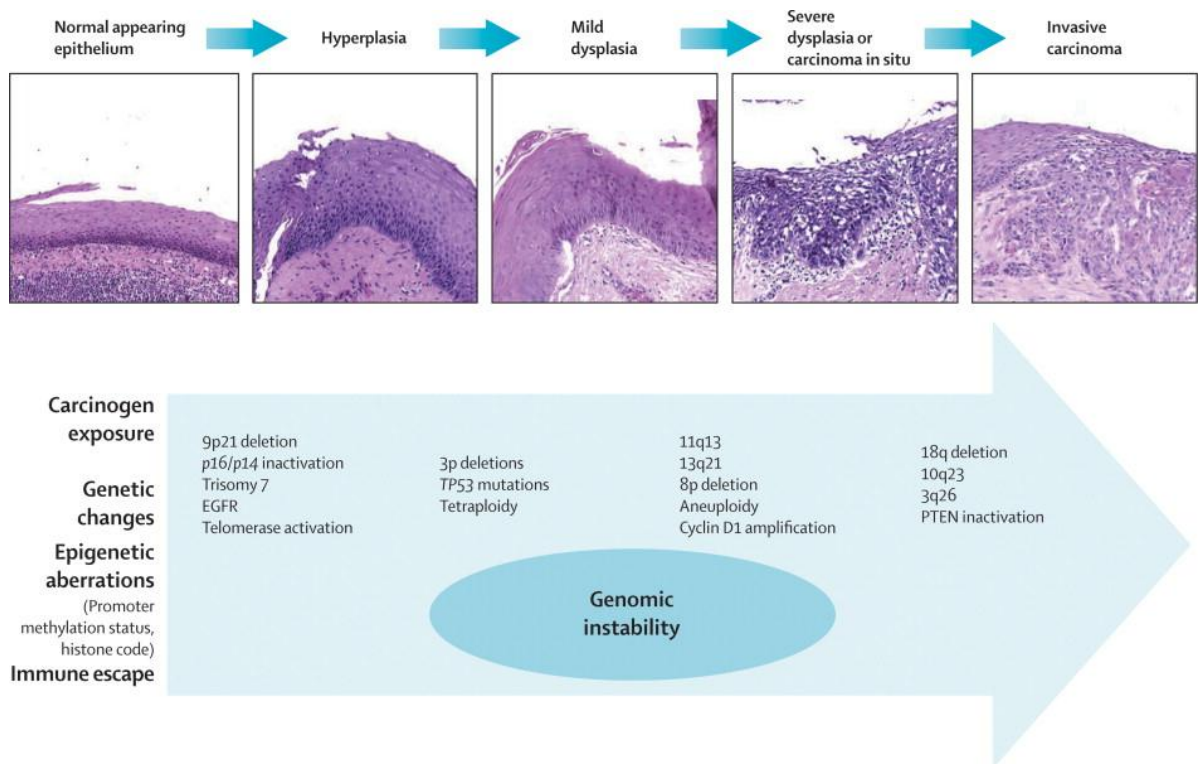


Figure 1. Hypothetical model of oral carcinogenesis. The image is reprinted with kind permission of the publisher. (Appendix 2.1)

LOH analysis was conducted by van der Riet *et al.* (1994) to define more clearly the lost function of chromosome 9p in the early event of oral carcinogenesis. They found that approximately 70% of pre-invasive lesions showed an allelic loss signal on chromosome 9p. This suggested the inactivation of chromosome 9p as the most common genetic changes that occur early in the progression of head and neck tumors. In another study Mao *et al.* (1996) have performed loss of heterozygosity (LOH) assay on

84 oral premalignant lesions using two microsatellite markers located at chromosome 9p21 and 3p14 and the results showed that almost 96% of oral premalignant lesions have yielded positive results. The genes contained within the loci of chromosome 9p21 were found to be p16 gene (Reed *et al.*, 1996) and the loss of this gene function occurs frequently in many early cases of human cancers (Rocco and Sidransky, 2001).

In addition, loss of chromosome arm 3p includes FHIT and RASSF1A which are known TSG and these genes were detected in oral premalignant lesions and suggest that this loss is an early event process in oral carcinogenesis (Mao *et al.*, 1996a). While in the progression from benign squamous hyperplasia to dysplasia, loss of chromosome 17p was observed in LOH analysis and sequence analysis of p53 exon sequence (Califano *et al.*, 1996). Besides that study of Rousseau *et al.* (2001) on 59 premalignant oral lesion tissues showed that CCND1 genes are located at chromosome 11q13. In their study this gene was over-expressed in almost 40% of mild dysplasia, 45% of moderate dysplasia and 24% of severe dysplasia oral epithelial dysplasia, when analysed through quantitative PCR and immunohistochemistry (IHC) study

2.10 Conventional Cytogenetic:

Cytogenetic studies on solid tumors have a great impact on the site of clinical genetic and basic science information of solid tumors over the past two decades (Mitelman *et al.*, 1997). Cytogenetics is known as the study of chromosomes in terms of “coloured body”, which can be divided into conventional and molecular type of analyses (Fan, 2003). Conventional cytogenetic analyses are more advanced in hematologic malignancy researches which normally involve using a variety of staining methods such as Giemsa and Leishman staining to highlight chromosome bands to study the structure and numbers of chromosome (Thompson, 1997).

2.11 Molecular Cytogenetic

The molecular cytogenetic techniques are mostly involved in the analysis of chromosomal alterations using in situ hybridization based techniques for example, Comparative genomic hybridization (CGH), Fluorescent in situ hybridization (FISH) and technology like array CGH (Fan, 2003; Varella-Garcia, 2003).

2.11.1 Comparative genomic hybridization (CGH)

The introduction of CGH was first described by Kallionoemi *et al.* (1992) as a technique that is improved from the limitation of conventional cytogenetic analysis. This technique was designed to identify the regions of amplification and deletion across the genome in a single hybridization experiment based on the comparison of hybridization signal intensities. This technique has contributed the greatest impact on molecular cytogenetic field for solid tumors. They have not only allowed to skip the karyotypic demonstration and *in vitro* tumor cell culture but also enhanced the previous knowledge about targeted chromosomal aberrations to a higher level (James, 1999). In

addition, it allows quick detection and mapping of DNA sequence copy number differences of normal and abnormal gene content (Weiss *et al.*, 1999).

Most of the applications of CGH come from cancer research due to DNA copy-number alterations which are of pathogenic importance in cancer (Tachdjian *et al.*, 2000). In addition, CGH can also be used to detect consistent recurrent chromosomal losses and gains in specific tumors, implication of specific genes in cancer development and progression, analysis of the clonal evolution of cancer *in vivo*, and the dissection of genetic changes in experimental models of carcinogenesis as well as tumor progression (Tachdjian *et al.*, 2000).

Despite all these benefits, CGH do have its own limitations. The disadvantages include failure to identify structural chromosomal aberrations such as translocation and inversion, as CGH can only detect the gains and loss of chromosomal aberrations (Lichter *et al.*, 2000). Besides that, CGH could provide only a limited mapping resolution, which will decrease the sensitivity to detect the chromosomal aberrations in the study (Lichter *et al.*, 2000). In addition, CGH technique has a lower resolution as its limited ability can only detect chromosomal aberrations in the range of 10-20 Mb (Bentz *et al.*, 1998). For example, CGH has its limitation to detect the amplicon unit which are less than 2Mb and/or deleted regions which is at least 10Mb (Piper *et al.*, 1995).

According to the standard CGH procedures as described by Kallioniemi (1992), genomic DNA isolated from test and reference samples are labelled with red and green fluorescent dyes, respectively. Each labelled DNA is subjected to competitive hybridization to normal metaphase chromosomes. The hybridization of repetitive

sequence is actually blocked by addition of Cot-1 DNA. The ratios of red and green fluorescent signals in paired samples are measured along the longitudinal axis of each chromosome. Chromosomal regions involved in deletion or amplification in test DNA appear red or green, respectively, but the chromosomal regions that are equally represented in the test and reference DNA appear yellow based on the comparison of hybridization signal intensities.

CGH has proved to be a promising tool to detect chromosomal aberrations in oral cancer. The first CGH data in head and neck cancer was published by Speicher *et al.* (1995). In their study 13 tumor samples which included 6 pharyngeal, 3 tongue, 2 larynx, 1 lip and 1 neck were analyzed. Later, another CGH study using 17 oral squamous cell carcinoma cell lines were published by Matsumura (1995). Although the samples involved were different but their findings were almost similar, where gains in chromosome 3q and losses in chromosome 3p were observed. This suggested that the copy number change on chromosome 3 is an important chromosomal alteration in oral carcinogenesis. Besides that, the CGH results from Matsumura (1995) showed that the common gain regions observed were chromosome 8q22-26, 3q25-27, 7p12, 11q13, 13q33, 14q, 15q and 20q whereas common deleted regions were detected in chromosome 3p, 18q21, 5q21-q22, 7q31 and 8p.

CGH studies of oral squamous cell carcinoma have identified non random chromosomal gains and losses affecting whole chromosomes of 2, 5p, 7p and 8q (Hermsen *et al.*, 1997; Okafuji *et al.*, 1999). However, Okafuji *et al.* (1999) study allowed a more comprehensive analysis of chromosomal alterations compared to Hermsen *et al.* (1997) due to large number of OSCC cell lines samples which were screened for chromosomal aberrations using CGH. This has benefitted Okafuji *et al.* (1999) to identify new

chromosomal aberrations, where they found that deletion of chromosome 18q and 4q and gains of chromosome 20 that were not reported in earlier studies. Similarly, Martin *et al.* (2008) have carried out CGH study on a large cohort of 31 OSCC cell lines to determine the chromosomal aberrations involved in the development of OSCC and their findings suggested that loss of chromosome 9p and gain of chromosome 11q13 are the potential prognostic markers involved. Furthermore, CGH data from Martin *et al.* (2008) (gain of chromosome 3q26-qter, 8q24 and 20q12) showed a similar pattern with the chromosomal aberrations detected from the study of Uchida *et al.* (2006).

Chromosomal aberrations in head and neck squamous cell carcinoma (HNSCC) detected using meta-analysis from a total of 13 studies revealed that the most common gains were at chromosome 3q26-27, 1q25-q44, 2p, 2q, 5p15, 8q24, 9q34, 11q13, 20q12-13.2 while losses were at chromosome 3p12-24, 4q21-31, 5q21, 6p, 6q, 7q22-qter, 8p, 9p21-24, 10q22-26, 13q, 14q, 17p, 18q 21q and 22 (See review; Gollin, 2000) Recent study by Tsantoulis *et al.*, (2007) also highlighted several important copy number aberrations such as gains of chromosome 8q22-ter, 11q13, 17q24-25, 20q and loss of chromosome 3p, 4q, 5q21, 18q, 21q and 22q which were in agreement with previous studies. Wolff *et al.* (1997) were reported to be the first to perform CGH analysis with a combination of tumor samples (20 frozen OSCC tumor tissues and four tongue OSCC cell lines). Frequent gain of DNA copy number aberrations were detected in chromosome 3q26-27, 5p15, 9q34, 1p36.3, 8q24, 10q26, 19, 20q, 11q13, 12q24, 15q22-24, 16p13.2, 17q, 6q26-qter, 7p22, 12p12.2-p13, 14q31-32.2, 1q32-41, 2q37, 16q23-24 and losses were found in chromosome 3p12, 5q21 and 6q13.

In another study, gains of chromosome 3q21-29, 11q13, and the loss of chromosome 8p12 were identified as independent prognostic (Bockmuhl *et al.*, 2000). Furthermore, Kaplan-Meier analysis revealed that gains of chromosome 2q12, 3q21-29, 6p21.1, 11q13, 14q23, 14q24, 14q31, 14q32, 15q24, 16q22 and deletions of 8p21-22 and 18q11.2 were significantly associated with both shorter disease-free interval and disease-specific survival. Similarly, another study by Ashman *et al.* (2003) revealed that gains of chromosome 3q25-27 and deletion of 22q were significantly correlated with reduced disease specific survival. This study suggested that deletion of chromosome 22q was a potential independent prognostic marker in HNSCC. The chromosomal aberrations detected from both studies of Bockmuhl *et al.*, (2000) and Ashman *et al.*, (2003) using CGH, facilitated the effort to create a tumor classification system for histological characterisation. According to this system it classifies patients into subgroups with different survival for different treatment and diagnosis of HNSCC.

Some studies have shown that there were differences of chromosomal aberrations between metastasized and non metastasized OSCC. Ashman *et al.* (2003) study reported that gain of chromosome 7p and loss of 9q were significantly more frequent in HNSCC patients with positive lymph node metastasis. However, these findings were in disagreement with Bockmuhl *et al.* (1997) and Bergamo *et al.* (2000) which showed that gain of 7p is more common in negative lymph node metastasis HNSCC. While the alteration of chromosome 7p is still controversial, it is interesting to note that gain of chromosome 7q11.2 and 7q21 was commonly detected in metastases HNSCC (Bockmuhl *et al.*, 2002; Tremmel *et al.*, 2003; Hannen *et al.*, 2004). But in another study by Noutomi *et al.* (2006), they reported that loss of 4p was associated with lymph node metastasis when analysed with 35 primary OSCC tissues. While another study showed

that gains of chromosome 1q and 2q were detected in oral cancer patients with a clinical history of recurrence and metastasis (Bergamo *et al.*, 2000).

To date, there were only 2 studies that analyzed the chromosomal alterations in tongue SCC using CGH technique (Steinhart *et al.*, 2001; Hannen *et al.*, 2004). However, the objectives of these two studies were different. Steinhart *et al.* (2001) study was aimed to determine the copy number gains and concluded that amplification of 7q could be specifically linked with tongue SCC whereas, study by Hannen *et al.* (2004) was to investigate the differences of chromosomal imbalances between metastasized and non metastasized tongue SCC using CGH. They found that gains of chromosome 3q23-qter, 5p, 12p and 13q21-22, 7q21, 14q and loss of chromosome 15q were more commonly found in metastasized tongue SCC.

More than half of HNSCC tumors showed deletion of chromosome 3p and 8p and was associated with poor differentiation of HNSCC tumors (Bockmuhl *et al.*, 1996; Stafford *et al.*, 1999). Bockmuhl *et al.* (1996) indicated that loss of chromosome 3p and 9p and gains of chromosome 3q were highly associated with well differentiated carcinomas and early tumor development. However, deletion of chromosome 4q, 8p, 18q and 21q and amplification of chromosome 11q13, 19q, 22q were highly observed in undifferentiated HNSCC tumors (Bockmuhl *et al.*, 1996).

Lin *et al.* (2002) conducted a study to determine the chromosomal aberrations of 33 OSCC associated with betel quid chewing and 14 OSCC associated with cigarette smoking. Surprisingly, most of the chromosomal alterations were found to be similar in these two groups but the loss of chromosome 4q and gain of chromosome 8q were significantly lower in betel quid chewing OSCC patients. Two years later, Lin *et al.*

(2004) subjected for chromosomal aberrations identification using CGH from OSCC cell lines. All these cell lines were derived from a long term areca (betel) quid chewer who does not smoke. It was revealed that this cell line carried unique chromosomal gain in 5q and loss in 6p as well as common loss in chromosome 4q and gain of chromosome 8q and entire chromosome 20.

Noutomi *et al.* (2006) was the first group to analyse the genetic alterations in both primary OSCC and adjacent dysplastic lesions using CGH. Results showed that gain of chromosome 8q22-23 was the most frequent alteration observed in both dysplasia and OSCC and this finding were supported by Abou-Elhamd and Habib (2008), suggesting that this genetic alteration was the earliest event in the process of oral carcinogenesis.

Redon *et al.* (2001) performed a CGH study to compare the chromosomal aberrations between early stages (I and II) and advanced stages (III and IV) of OSCC. They observed a specific pattern of chromosomal alterations in the early stages which were showing gain at chromosome 11q13, 3q, 8q and loss of chromosome 3p. Besides that, they concluded that loss of chromosome 9p and whole chromosome 18 were indicative of stage III and IV tumor which were in concordance with Brieger *et al.* (2003) study.

Huang *et al.* (2002) groups have managed to construct an oncogenetic tree for tumor progression based on the CGH data obtained from 75 head and neck cancer. This oncogenetic tree model revealed that gains of chromosome 3q21-29 were important early chromosomal alterations among all the subsites of head and neck cancer followed by the gains of chromosome at 8q. Hence they occupied the position of second most important chromosomal alterations in early event. The later event were divided into three subgroups which were loss of chromosome at 3p, 9p,13q and gains at 7p, followed

by gains of chromosome 5p, 9qter, 17p, 8q and 18p. However, the data obtained from this oncogenetic tree model was masked by others subgroup of head and neck cancer. In 2009, an oncogenetic tree model based on 97 primary OSCC samples was constructed (Pathare *et al.*, 2009). This advanced oncogenetic tree model revealed multiple pathways of oral cancer progression and showed that the gains of chromosome 8q and losses of 3p were the most common chromosomal alterations in oral cancer. This model pointed out three progression pathways which were initiated by deletion of chromosome 8p, another by loss of chromosome 3p and the third by gains of chromosome 11q and 7p.

The importance to detect chromosomal aberrations occurring due to the increasing instabilities of DNA copy number changes together with concomitant microscopic approaches were found to be highly correlated with phenotypic abnormalities (Skacel *et al.*, 2009). Thus, this suggests that this technique has wide potential application in both basic research and clinical practices.

2.11.2 Florescent in situ hybridization (FISH)

Fluorescence in situ hybridization (FISH) is a molecular cytogenetic technique in which DNA probes are labeled with different colored fluorescent tags to detect and localize the presence or absence of specific regions of the genome in chromosomes (Pinkel *et al.*, 1986). This technique has become an increasingly important diagnostic tool especially in cancer research (Thompson and Gray, 1993). The advantage of using FISH as a cytogenetic tool is its capability to analyze chromosome instability in-situ such as DNA copy number at specific loci in metaphase spread and interphase nuclei (Thompson and Gray, 1993; Hackel and Varella-Garcia, 1997). By using FISH, other informations such as genomic integrity and rearrangement such as translocations can be identified which

otherwise are unable to detect using standard karyotyping technique (Thompson and Gray, 1993). FISH also allows simultaneous analysis of different probes by using different fluorochromes which saves time (Thompson and Gray, 1993). The other advantages of using FISH are that it can be applied to fixed tissues. Hence, in-vitro cell cultures are not required and assist researchers to understand better about chromosomal aberrations including the identification of genetic endpoints, which might be useful to be developed as a chemotherapeutic agent (Thompson and Gray, 1993).

FISH technique has been applied to study chromosome aberration in many cancers including oral cancer. Paradiso *et al.* (2002) used FISH to analyze the alterations of HER-2/neu gene which is frequently found to be amplified in several tumor types including OSCC. The gene HER-2/neu is located on the chromosome 17q12-21 and these genes are responsible for encoding members of the tyrosine kinase family. The results showed that there was high amplification of this particular gene in OSCC and thus it was proved that this technique could be considered as a feasible tool to be applied on cytological brush specimens of oral mucosa. Another FISH study conducted by Uzawa *et al.* (2007) using materials from fine needle aspiration biopsy, consisting of 57 OSCC samples showed amplification of CCND1 and deletion of p16. This study thus suggested that the combination of these 2 markers can serve as better prognostic indicator for OSCC. Based on Uchida *et al.* (2006) study, FISH technique was employed to examine the regions of chromosome 5p and the FISH has detected 100% gain in this region. These results were accordance with the corresponding CGH study on 11 oral cancer cell lines. The results from this study showed there was a significant correlation between the gains of chromosome 5p indicating chromosomal aberration of 5p involved in oral cancer progression (Uchida *et al.*, 2006).

2.12 “Omic” Profiling

Human Genome Project which was completed in the year 2004 has successfully identified a total of 24,500 genetic codes (International Human Genome Consortium, 2004). This achievement has resulted in the advancement of high throughput technologies. One such technological development was “omic” profiling, which has enhanced the large scale discovery of genes that are disrupted in cancer and pathways which favoured the better understanding of the biology of cancer progression (Baak *et al.*, 2003; Nagaraj, 2009). With the introduction of high throughput technologies coupled with sophisticated bioinformatic tools for complex data analysis have enabled us to provide a comprehensive and systematic “omic” profiling to examine genes interaction within the biological pathways and network in cancer genomics study (Martin and Nelson, 2001). Through this, understanding of the mechanism behind aberrations leading to carcinogenesis can be improved (Martin and Nelson, 2001). Due to the complex nature of cancer, “omic” technologies take a central place in discovering new biomarkers that are important for developing new diagnostic and therapeutic strategies (Garnis *et al.*, 2004a). The “omic” technologies include genomics (the study of the human genome), transcriptomics (gene expression) and proteomics (the analysis of the protein complement of the genome). These technologies play an important role to reveal the mechanism of the cellular constituents in cancer cell and how they interact to produce cellular phenotypes through the information encoded from these high-throughput technologies (Nagaraj, 2009). One of the best genomic technology is the array CGH.

2.12.1 Array CGH

Array CGH is a high throughput technology where the metaphase chromosomes of the conventional CGH technique are replaced by high density array elements, which are comprised of genome sequences of DNA as the target for analysis (Lucito *et al.*, 2003). This technology is more advanced whereby it provides particular biological aspects of gene that resides within regions involved in copy-number alteration. The information of the probes/clones sequence that is spotted onto glass microscope slide (array) can be translated into genes by matching with the genome database *via* sophisticated software. Thus this technique can easily determine and pinpoint the candidate genes that are responsible for cancer and for unknown genetic diseases (Inazawa *et al.*, 2004; Lockwood *et al.*, 2006).

The limitations from conventional CGH can be overcome using array CGH (Inazawa *et al.*, 2004; Pinkel and Albertson, 2005; Kallioniemi, 2008; Theisen, 2008; Shinawi and Cheung, 2008). For example, array CGH has more significant advantages in terms of higher resolution and throughput with possibilities for automation, robustness, simplicity, high reproducibility and precise mapping of alterations compared to cytogenetic method (Inazawa *et al.*, 2004; Kallioniemi, 2008; Shinawi and Cheung, 2008). The resolution of array CGH is theoretically higher than CGH due to the smaller size of DNA probes in array CGH, where the level of resolution is determined by the size of nucleic acid targets and the density of coverage over the genome (Bejjani and Shaffer, 2006). Array consisting of 60 mer oligonucleotide offer a greater sensitivity and specificity in detection of copy number alterations (Curtis *et al.*, 2009). Besides that, the ultimate goal to use array CGH in research is to pinpoint the locations of cancer-associated genes and to understand the association of specific genetic changes or patterns of changes to know tumor characteristic, and tumor progression. This will

facilitate the discovery of new markers for tumor progression and identification of predictive and prognostic markers in cancer (Kallioniemi, 2008; Shinawi and Cheung, 2008). However, array CGH has its limitations such as it is unable to detect balanced rearrangements such as translocations and inversions, but could only detect copy number imbalances relative to other DNA regions and the cost to run an experiment using high resolution array CGH is relatively higher. Moreover, the nature of tissue heterogeneity reduces the sensitivity of copy number aberrations by shifting the reading of signal ratio due to contamination of noncancerous cell in tumor samples (Lockwood *et al.*, 2006).

The application of array CGH for DNA number "signatures" or profiling has been carried out in various cancer researches (Bejjani and Shaffer, 2006). The first oral cancer array CGH data was published by Redon *et al.* (2002) where they constructed a bacterial artificial chromosome (BAC) array CGH which was derived from 85 clones with a higher coverage for chromosome 3q25-qter region. Through the screening of head and neck cancer cell lines by house array CGH, they discovered a high level amplification of narrow chromosome 3q25.3 which comprised of clone RP5-854L10 and RP11-197N19.

One year later, Garnis *et al.*, (2003) have constructed a high resolution array CGH using a tiling set of 535 human BAC that cover almost the whole chromosome 3p which were applied to 20 OSCC tissues. Their results showed that there is a deletion of chromosome 3p14 which contains FHIT genes, a known predictor of poor outcome in oral cancer. Furthermore, they managed to detect novel chromosomal deletion regions in 3p22, 3p24, 3p26 and were the first to report the detection of chromosomal gain in chromosome 3p. Subsequently, Garnis *et al.* (2004c) carried out genome wide screening

on 20 OSCC tissues, focusing on a region at chromosome 8q22 with their newly developed array CGH derived from 165 BAC clones representing chromosome 8q21-24. Through the application of array CGH, they identified a recurring amplified region of chromosomal 8q22 which contained LRP12 gene. Gene expression analysis revealed this gene is a potential oncogene for oral cancer. In order to further identify chromosomal alterations in oral cancer, Garnis *et al.* (2004b) designed an oral cancer specific human BAC array for CGH analysis. This designed array contained 223 BAC clones where 135 clones were mapped to regions commonly altered in oral cancer which included 3p13-14, 3p24, 4q28, 7p11, 8p23, 9p22, 11q13, 13q21 and 17p13. Although coverage of selected chromosomal region was smaller, this study marked the first step in the construction of commercial oral cancer specific array CGH.

It is interesting to note that Snijders *et al.* (2005) hypothesized that rare amplicons tend to be informative for tumor growth. They performed a genome wide screening using high resolution array CGH derived from 2464 BAC clones, on 89 OSCC tissues taken from 4 different sites in oral cavity (tongue, cheek, gum and floor of mouth) and focused on narrow regions of gene amplification to discover genetic pathways involved in OSCC carcinogenesis. They managed to detect amplified genes involved in integrin signaling, survival, adhesion and migration as well as hedgehog and notch signaling pathways. The decision to select minimum common regions and match with gene expression data enabled Snijders *et al.* (2005) to identify candidate driver genes located in the amplicons. In the same year, Baldwin *et al.* (2005) carried out the first whole genome wide screening with 32433 BAC clones tiled across the genome to hybridize with chromosomal aberrations in 20 formalin fixed paraffin embedded OSCC tissues. They reported several novel chromosomal aberrations which included gain of TRIO gene located in chromosome 5p15.2 and 9 other matrix metalloproteinase (MMP) genes

(MMP7, MMP20, MMP27, MMP8, MMP10, MMP1, MMP3, MMP12, MMP13) within an amplicon located in chromosome 11q22.3.

Due to the increase in OSCC incidence among younger patients, O'Regan *et al.* (2006) conducted a study to identify chromosomal alterations differences between oral cancer patients younger than 40 years old against patients aged 40 years and above. The results showed that deletion of p16 (chromosome 9p21), MSH3 (chromosome 5q11.2-13.2) and APC (chromosome 5q21-22) were completely absent in the younger group but was present in more than 50% among the older group. A unique study conducted by Chen *et al.* (2004), used array CGH to investigate the genetic profiles of 60 microdissected OSCCs in betel quid users and non users. The results from this study showed that gain of EXT gene and loss of FHIT gene were significantly higher in betel quid users as compared to non betel quid users. Another study investigated the chromosomal aberrations between smoker and non smokers who diagnosed with OSCC (O'Regan *et al.*, 2006). However, this study only revealed that young OSCC smokers showed few genomic changes compared to older smokers.

A study by Liu *et al.* (2006) compared the genomic alterations in 8 pairs of microdissected tissues from primary OSCC and matched metastatic lymph nodes which revealed that genetic alterations for TGF β 2, CRBP1, PIK3CA, HTR1B, HRAS, ERBB3 and STK6 genes were significantly different between primary OSCC and the associated metastases. Similarly, Chen *et al.* (2004) also performed genome wide screening to detect the genomic alterations for metastasizing and non metastasizing OSCC. This study revealed that gains of EGFR genes were occurring at chromosome 7p; FDF4, CCND1 and EMS1 genes at chromosome 11q13 and AIB1 genes at chromosome 20q. Ultimately this study showed that all these genes were significantly associated with

metastasis OSCC. Another study carried out by Sparano *et al.* (2006) have attempted to narrowed down the previously published amplified regions by array CGH to smaller regions intended for yielding shorter lists of candidate genes. For example, the published region was from chromosome 5p14-15.3 (22.2 Mb), and they narrowed the documented chromosomal alterations to a minimal region which is chromosome 5p15.33 (0.4 Mb). Amplification of chromosome 11q13 in oral cancer have been well documented and to characterize the 11q13 amplicon in detail, Freier *et al.* (2007) have used matrix CGH microarray containing BAC clones derived from chromosome 11q13. They attempted this approach in 40 OSCC tissues. They showed that the genes, EMS1 and SHANK2 which were known as cytoskeleton associated proteins. These genes were amplified within the CCND1 amplicon, suggesting that these proteins coupled with CCND1 genes cooperatively will contribute to OSCC pathogenesis.

Instead of analyzing the chromosomal aberrations for OSCC in all the studies described above, Garnis *et al.* (2009) have attempted to study the genome analysis of a cohort of 86 oral lesions with longitudinal follow up. Samples taken in their study included 62 oral premalignant lesions (OPL) and 24 OSCC. They used a tiling array CGH which comprised of 26363 overlapping BAC clones covering the entire human genome. Garnis *et al.* (2009) established that the most commonly observed genomic alteration was gain of chromosome 20p among the OPL whereas gain of chromosome 8q was more common in OSCC. This study has given the first high resolution analysis of whole genome in OPLs that represented a significant step towards predicting progression risk in early pre-invasive diseases. Likewise, Cha *et al.* (2011) recently determined the genetic alteration of OSCC and the dysplastic transitional area using high resolution 4x44k oligonucleotide array CGH. They used samples derived from 7 OSCC, which included 7 adjacent oral dysplasia and 7 matched margins, normal free of oral mucosa

tissues. This study revealed that gain of RNF36 (chromosome 15q21.1) gene was highly observed in oral dysplasia compared to OSCC. They also reported that the highest frequency of deleted genes in oral dysplasia and OSCC tissues was CKAP2L at chromosome 2q13.

To date, Ambatipudi *et al.* (2011) was the first study to employ high resolution 1 x 105k oligonucleotide array CGH to perform genome wide screening for determining the chromosomal aberrations associated with poor prognosis. This study revealed that the novel chromosomal alterations such as 9p23-24.3 (PTPRD gene) and 5p15.33 (PDCD6 gene). Moreover, Ambatipudi *et al.* (2011) compiled a list of genes associated with oral cancers based on 25 previous studies.

Since 2002, a numerous number of array CGH studies have been carried out. However, some major drawbacks are profounding in these studies. In their studies they employed only the non sensitive BAC clone array CGH and low resolution oligonucleotide array CGH (4x44k and 1x105k oligonucleotide based array CGH) to screen the whole genome chromosomal aberrations in OSCC. In fact these technologies are not capable of detecting small genomic alterations in terms of amplification and deletion in the genomic regions. On the contrary, the increased density of the ultra dense higher resolution aCGH technology will result in 2.1kb median probe spacing which would have provided detailed information even at individual exon levels (Przybytkowski *et al.*, 2011).

3.0 Methodology

3.1 Study Design

This was a descriptive study to determine the copy number alterations and gene pathways that are unique to tongue and cheek SCC. This study was approved by the Medical Ethics Committee (MEC), Faculty of Dentistry with the MEC no: DFOS1001/0004(P). This was a small part of a large study ["Oral cancer and Pre-cancer in Malaysia-Risk factors, prognostic markers, gene expression and impact on quality of life"] (DFOOP0306/0018/L) which includes the banking data and tissues.

3.2 Sample selection

Samples for the study comprised of frozen tissues sections from OSCC patients. For this descriptive study, determination of number of samples are based on literature reviews and current genome wide screening study on oral cancer using array CGH (Cha *et al.*, 2011; Sparano *et al.*, 2006; Garnis *et al.*, 2004c; Garnis *et al.*, 2003). Fresh frozen tissues (n=20) consisting of 12 tongue and 8 cheek SCC were included in this study and were obtained from Malaysian Oral Cancer Database and Tissue Bank System (MOCDTBS) coordinated by Oral Cancer Research and Coordinating Center (OCRCC-UM) and were included in this study.

3.2.1 Demographic characteristics of the samples

The demographic characteristics of the sample such as age, gender and ethnic were extracted from the MOCDTBS.

3.2.2 Sample Criteria

A) Inclusion criteria

1. Oral cancers which are pathologically diagnosed as OSCC.
2. Primary tumor tissues consisting of epithelial tumor cells with evidence of more than 70% of tumor OSCC cells.

B) Exclusion criteria

1. Oral cancers which are pathologically diagnosed as other cancer (not OSCC).

3.3 Sample Preparation of DNA

3.3.1 Cryosection on frozen tissue

Frozen tissue specimens were mounted using Optimal Cutting Temperature (OCT) compound embedding medium (Leica, Surgipath) and were serially sectioned at -20 °C. The outer sections of the tissues were H&E stained for histological assessment with the help of an experienced pathologist. Tumor percentage was gauged under the microscope and macro-dissection was done to obtain $\geq 70\%$ tumor cells for this study. Then, 750 μm of the section were collected in sterile 1.5ml microcentrifuge tube and subjected for DNA extraction.

3.3.2 DNA Extraction

All tumor tissues for test gDNA samples were isolated by using DNEasy Blood & Tissue Kit (Qiagen GmbH Germany) following the manufacturer's instructions. 750 μm of the tissue sections were mixed with 1x PBS buffer and then centrifuged in 5000rpm for 5 minutes. The supernatant were discarded thereafter. Then, 180 μl of ATL buffer and 20 μl of proteinase K were added and then kept for incubation at 56 °C, overnight. Later, 200 μl AL buffer was added again and then followed by incubation at 70 °C for 10 minutes. Afterthat, absolute ethanol (200 μl) were added and thereafter all the mixtures weretransferred to QIAamp Mini spin for centrifugationat 8000rpm for 1

minute. After centrifugation the collection tubes were discarded and new columns were placed into the new collection tubes. To this 500 μ l AW1 was added and then again centrifuged at 8000 rpm for 1 minute duration. This process was repeated by changing only the buffer with 500 μ l AW2. Column was spun at 13200 rpm for 3 minutes in order to dry the column membrane. In continuation, 50 μ l AE buffer was added into the column and then incubated for 5 minutes. Final centrifugation was done at 13200 rpm for 3 minutes to collect the elution buffer which contained DNA.

3.3.3 Quantification of DNA measurement

The quality and quantity of gDNA were verified by Nanodrop ND-2000. DNA concentration (ng/ μ l) and DNA purity (A260/A280) were calculated using the spectrophotometer (NanoDrop ND-2000) according to the manufacturer's instructions.

3.4 Technique to be employed for Array CGH.

Array CGH was performed using Oxford gene technology according to the method as described in manufacturer's instructions (Agilent Technologies) and also the conditions of Srisupundit *et al.*, (2010). Oligonucleotide microarray experiments were performed using the Agilent SurePrint G3 Human CGH 1 x 1 M (Agilent Technologies, Santa Clara, CA, USA). All the methods were conducted according to manufacturer's protocol (version 5.0, June 2007) with minor modification that replaced DNA digestion step to defragmentation method. The microarray used for this study was 1 x 1 Million slide formats printed using Agilent's 60-mer SurePrint technology which was having 974,016 biological features (coding and non-coding sequences represented). The probes cross over both the non-coding and coding regions for genome wide-ranging representation.

3.4.1 Sample Preparation (Defragmentation Method) for aCGH

For defragmentation method, total of 500ng of gDNA from each frozen tumor tissue (test sample) and commercially gender matched pooled DNA sample (Promega) were performed at 98 °C for 20 minutes. In this protocol the DNA digestion step was omitted. The mixture of both test sample and reference control were transferred to thermomixer and then kept for incubation (37 °C) overnight. After this duration the mixtures were again incubated at 65 °C for 20 minutes to inactive the enzymes. Followed by this, agarose gel electrophoresis was run to determine the size of the digested gDNA in the length of 200bp-500bp. An amount of 2 µl of the digested genomic DNA (26 µl) was taken for 1% agarose gel electrophoresis. The digested products were always taken between 200bp-500bp in length. For determining the DNA size, 1 µl DNA Ladder was used. Agarose gel electrophoresis was carried out at 110 volts at room temperature. The digested DNA was visualized by ethidium bromide staining.

3.4.2 Sample Labelling for aCGH

Digested gDNA was fluorescently labelled with cyanine 5-dCTP (test sample) and cyanine 3-dCTP (reference control) using a labelling kit (Agilent Genomic DNA Labeling Kit PLUS). An amount of 5 µl random primer was added to the digested gDNA (24 µl). This was transferred to the thermomixer comfort (Eppendorf, Hamburg, Germany) at 95 °C for 3 minutes and later moved to ice and incubated for 5 minutes. Thereafter, the mixture was mixed with labelling master mix which contained 2 µl of nuclease free water, 10 µl 5X Buffer, 5 µl 10X dNTP, 3 µl cyanine 5-dCTP (Test sample) or cyanine 3-dCTP (reference control) and 1 µl Exo-Klenow fragment to make a total volume of 50 µl. Then the mixture was transferred to thermomixer at 37 °C for 2 hours followed by 10 minutes incubation at 65 °C to inactive the enzyme.

3.4.3 Probe Purification for aCGH

Probe purification was done by using Microcon YM-30 filter column (Millipore, Massachusetts, USA). The column was placed into a 1.5 ml microfuge tube and centrifuged at 8000rpm for 10 minutes. The supernatant was discarded. Then, 430 μ l of 1X TE (pH 8.0) was mixed with 50 μ l of labelled gDNA and loaded all 480 μ l mixture into the filter and centrifuged for 10 minutes at 8000rpm. Then, the filter was inverted into a fresh 1.5 ml microfuge tube and then centrifuged at 8000rpm for 1 minute to collect the purified sample. The volume of the purified probe was recommended not above 80.5 μ l, which was measured using pipetter. Then, 1.5 μ l of the purified probe was taken to determine the yield and specific activities by using the NanoDrop ND-2000 UV-VIS Spectrophotometer.

3.4.4 Microarray Hybridization

An amount of 80.5 μ l of Cy5 labelled gDNA mixture and 80.5 μ l of Cy3 labelled gDNA mixture were mixed together with 50 μ l of 1.0 mg/ml Cot-1 DNA, 52 μ l of Agilent 10X Blocking Agent and 260 μ l of Agilent 2X Hybridization Buffer in 1.5 ml microcentrifuge tube. The mixture was transferred to thermomixer comfort to incubate at 95 °C for 3 minutes followed by incubation for 30 minutes at 37 °C. Then, the sample tubes were centrifuged at 17900 x g for 1 minute to collect the sample at the bottom of the tube.

Hybridization sample mixture (490 μ l) was slowly dispensed onto the gasket slide well in a drag-dispense manner and was fully loaded before placing the Agilent SurePrint G3 Human CGH 1 x 1 M microarray slide. Then, the microarray slide was placed facing down the gasket slide to make sandwich pair. Gently, SureHyb chamber cover was placed slowly onto the sandwiched slide and clamped tighten onto the chamber. Next,

the assembled chamber was rotated vertically to wet the slides and assessed the mobility of the bubbles. The slide chamber was placed in the rotator rack in a hybridization oven set to 65 °C and rotated at 20rpm for 40 hours.

3.4.5 Washing Preparation for aCGH

After 40 hours of the hybridization process, Agilent SurePrint G3 Human CGH 1 x 1 M microarray slide in a sandwiched pair slide was taken out from the SureHyb chamber oven and disassembled in Oligo aCGH Wash buffer 1. Then, the first washing procedure was done by placing the slide rack into slide staining dish number 2 and washed for 5 minutes. This was followed by the second washing step where the slide rack was placed into the pre-warmed glass dish which was filled with Oligo aCGH buffer 2 at 37 °C for 1 minute. Later, the dish glass was transferred to Acetonitrile and was stirred for 1 minute. The slide was scanned immediately to minimize the impact of environmental oxidants on signal intensities.

3.4.6 Microarray scanning using Agilent Scanner Control and Feature Extraction (FE)

Slide was assembled into appropriate slide holder in Agilent Microarray Scanner (Agilent California) with Agilent Scanner Control software v7.0. Scan region was set to 61 x 21.6 mm and resolution was set to 2.5 µm. Dye channel was set to Red and Green where the PTM was set to 100%. Automatic file setting was selected with Prefix 1 which set to Instrument Serial Number and Prefix 2 was set to Array Barcode. Scanner ready signal was appeared after all the setting was completed. Agilent Feature Extraction software v9.5 was used to extract microarray CGH data into raw text output files. Then, FE quality control report was generated and TIFF images of the hybridized array areas was taken.

3.5 Analysis

3.5.1 Data Analysis

Extracted raw text output data were normalized and converted into cgh files which were then uploaded into OGT's CytoSure™ Interpret software and save into a database of all pre-processed arrays ready for population analysis to obtain meaningful results.

The CytoSure Interpret software (CBS) package features a circular binary segmentation algorithm, where thresholds were set at 0.3 for gains and 0.6 for losses due to a loss of chromosomal material. They were taken in the ratio of 2:1 which is the total loss of information and the gain is kept as 3:2 to reduce the false positive findings. The calling threshold was then set at four flanking probes where $Z > 0.016$.

For a region to be called as aberrant, it must be present in the both datasets within the dye-swap experiment. To complement the CBS algorithm, a second threshold based method was also applied to the individual probe \log_2 ratios. Data from the patient labelled in Cy5 and in Cy3 is combined as follows: $X^2 \times Y^2 = Z$ (where, X = individual probe normalised \log_2 ratio for Cy5-labelled patient data, Y = individual probe normalised \log_2 ratio for Cy3-labelled patient data).

3.5.2 Population analysis

The population analysis is a feature in Cytosure software that provides calculation and graphing of aberration frequencies from groups derived from the saved database. Hence, in this study, all tongue and cheek SCC array CGH files were saved into the database which was labelled as cheek and tongue, respectively and were taken for population analysis to obtain the unique copy number alterations (CNAs) among these two distinct groups. All CNAs that were found common in both tongue and cheek SCC were

excluded. As a further filter, only CNAs within genic regions were used for further interpretation.

3.5.3 Pathway Analysis

The set of genes obtained from Cytosure software included the amplified and deleted genes in tongue and cheek SCC, respectively were uploaded into MetaCore™ analytical suite version 4.5 (GeneGo, Inc., St. Joseph, MI) and the analysis was conducted in accordance with the application manual. List of amplified and deleted genes were analyzed for pathway analysis in MetaCore™ and the results were ranked by p-value. Only significant pathway with $FDR < 0.05$ was included in this study.

For determining the representation of pathways dictated by an input list, MetaCore software applies a hypergeometric distribution statistics (A standard statistic used to compare populations). In the present study, the “p value” was compared to the population as represented by the data to a selected ontology or network algorithm. Like any other statistics the “p value” or statistical significance in Metacore software was a measure of the likelihood that an event would happen purely by chance. The hypergeometric statistic in particular takes into account the number of objects in the dataset, the number of objects in the intersecting map/network (in the enrichment ontology used) and the number of objects in the entire database. This assessment would therefore return the significance level that tells us the likelihood that intersects between the dataset and a particular map/network, purely obtained by chance.

Although the hypergeometric statistic calculation could be the major drawbacks of this software, since the outcomes are mainly derived by chance but Metacore is the only tool

that has the largest collection of pathway maps to be able to do this type of analysis and that the maps are manually constructed based on high quality experiments.

4.0 Result

4.1 Demographic characteristic of the study samples

The mean age of patients whose tumor samples were used for the study was 54.85 ± 16.13 years with 7 males and 13 females. From the total samples, about 65% of the cases were in pathological stage III or IV. The details for each case are as in table 4.1.

Table 4.1: Demographic characteristics of study samples.

Case Number	Age (years)	Gender	Ethnic	Site	Differentiation	pTNM
1	43	M	Others	Cheek	WELL	III
2	70	F	Indian	Cheek	MOD	IV
3	59	M	Indian	Tongue	WELL	I
4	53	F	Others	Tongue	WELL	IV
5	29	F	Malay	Tongue	WELL	III
6	80	M	Chinese	Tongue	MOD	II
7	74	F	Indian	Cheek	WELL	III
8	42	F	Malay	Tongue	WELL	IV
9	67	M	Malay	Cheek	WELL	IV
10	78	F	Indian	Cheek	MOD	IV
11	57	M	Others	Cheek	MOD	IV
12	49	F	Malay	Tongue	MOD	IV
13	26	F	Indian	Tongue	WELL	I
14	47	M	Malay	Tongue	WELL	II
15	30	M	Malay	Tongue	MOD	IV
16	52	F	Indian	Tongue	MOD	IV
17	61	F	Indian	Cheek	MOD	II
18	75	F	Indian	Cheek	MOD	IV
19	49	F	Indian	Tongue	WELL	I
20	56	F	Indian	Tongue	MOD	I

Abberation: M=Male; F=Female; WELL=Well Differentiation; MOD=Moderate Differentiation

4.2 Chromosomal alterations (aberrations) detected using array CGH and genes involved

All 20 cases of OSCC used in this study showed chromosomal aberrations. Summaries of chromosomal gains and losses by array CGH on tongue SCC analysis are illustrated in tables 4.2 and 4.3 respectively. Tables 4.4 and 4.5 shows summary of losses and gains of chromosomal segments observed in cheek SCC, respectively. Comparison of chromosomal aberrations of tongue SCC and cheek SCC using Cytosure software showed chromosomal aberrations in 210 genes for tongue SCC. These included 187 amplified gene and 23 deleted genes. In contrast to tongue SCC, there were chromosomal aberrations of 71 genes for cheek SCC which included 23 amplified genes and 48 deleted genes. The mean number of chromosomal aberrations per tumor for tongue SCC (22.75 ± 26.58) was higher than for cheek SCC (8.63 ± 11.89).

4.2.1 Chromosomal aberrations detected genes involved in tongue SCC

Table 4.2 and 4.3 lists the 187 amplified and 23 deleted genes identified from tongue SCC, respectively. In this study, there were five most common amplified regions which were 8q24.22 (33.33%), 8q24.3 (33.33%), 11q13.2 (33.33%), 12q13.13 (33.33%), 14q32.33 (33.33%) whereas the most common deleted regions were 2q21.1 (16.67%) and 6q21 (16.67%). Genes which were amplified at chromosomal regions 8q24.22 contained WISP1 and gene discovered at 8q24.3 was DENND3. For region 11q13.2, the amplified gene was CCDC87. Amplified gene at region 12q13.13 was HOXC13 while at 14q32.33, the gene involved was CDCA4. For the deleted genes, TUBA3D and LACE1 were found at chromosomal regions at 2q21.1 and 6q21, respectively.

Table 4.2: Details of the amplified regions for each chromosome in 12 tongue SCC.

Chromosome (Cytoband)	Position Start (bp*)	Position Stop(bp*)	Size (kb*)	Frequency No	Gene Name
1p36.13	17173727	17178735	5.008	1	MFAP2
1p36.21	13786019	13816638	30.619	1	PDPN
1q32.2	207854989	207892155	37.166	1	LAMB3
1q32.2	207933242	207974704	41.462	1	HSD11B1
1q21.1	144271009	144279871	8.862	1	ANKRD35
2q31.1	176697610	176726177	28.567	1	HOXD3
2q31.1	171353649	171396655	43.006	1	GAD1
3p21.31	45777623	45784991	7.368	1	SLC6A20
3q22.3	139549846	139606863	57.017	1	MRAS
3q23	144320888	144324458	3.57	2	CHST2
3q24	150191793	150227721	35.928	2	GYG1
3q25.2	156279994	156384067	104.073	2	MME
4p16.3	2814573	2901366	86.793	2	ADD1
4p15.32	17189465	17218539	29.074	2	LAP3
4p15.2	26207471	26365538	158.067	2	TBC1D19
4p14	38723098	38804575	81.477	1	KLHL5
4p14	40953629	40965137	11.508	1	UCHL1
5q23.3	131846712	131854231	7.519	3	IRF1
5q32	149656691	149662626	5.935	1	ARSI
6p24.1	11824981	11914856	89.875	1	C6orf105
6p12.3	49626351	49637536	11.185	1	C6orf141
6q25.3	160143191	160161671	18.48	1	PNLDC1
6p21.33	31538937	31541421	2.484	1	HCP5
6p21.31	35179083	35217098	38.015	2	TCP11
6p21.31	35788157	35803837	15.68	2	FKBP5
6p21.31	36437239	36476426	39.187	2	ETV7
6p21.1	43158819	43237417	78.598	2	PTK7
6p22.1	26389604	26393724	4.12	1	HIST1H4H
6p22.1	26473707	26561548	87.841	1	BTN3A2
6p22.1	26511387	26561548	50.161	1	BTN3A2
6p22.1	26549309	26561548	12.239	1	BTN3A3
6p22.1	29071557	29080878	9.321	1	ZNF311
6p22.1	29627716	29708258	80.542	1	GABBR1
6p21.33	31204971	31215767	10.796	1	PSORS1C1
6q25.3	160141234	160161399	20.165	1	PNLDC1
7p15.3	24704967	24763847	58.88	1	DFNA5
7p15.1	29926627	29995512	68.885	1	SCRN1
7p13	45006049	45082131	76.082	1	CCM2
7p11.2	55050689	55240708	190.019	2	EGFR
7p11.2	56099397	56115745	16.348	1	SUMF2
7q11.23	72621076	72630946	9.87	3	TBL2
7q22.1	98810308	98830216	19.908	1	ARPC1B
7q36.1	150180531	150189198	8.667	1	ABP1
7p15.2	27095898	27149791	53.893	1	HOXA1, HOXA3
8p21.2	27223463	27313957	90.494	1	PTK2B
8q22.3	104104531	104153424	48.893	1	ATP6V1C1
8q22.3	104224163	104256121	31.958	1	BAALC
8p21.3	22355303	22366416	11.113	1	PPP3CC

8q24.11	118918056	119198326	280.27	1	EXT1
8q24.13	126055333	126060447	5.114	2	ZNF572
8q24.13	126196625	126445872	249.247	2	NSMCE2
8q24.22	134273951	134310617	36.666	4	WISPI
8q24.3	142207959	142274555	66.596	4	DENND3
9q34.13	134664598	134743659	79.061	1	C9orf98
9q22.33	98039129	98104004	64.875	1	HSD17B3
9q32	115278405	115397080	118.675	2	RGS3
9q33.1	116831323	116920031	88.708	1	TNC
9q34.3	136655451	136674300	18.849	1	COL5A1
9q34.3	137593925	137598618	4.693	1	PAEP
10q24.2	99407997	99510452	102.455	1	ZFYVE27
10q26.3	134906557	134940323	33.766	3	ADAM8
10p12.33	17673021	17699336	26.315	1	PTPLA
10p12.33	17900075	17992777	92.702	1	MRC1L1
10p12.1	25189493	25281202	91.709	1	PRTFDC1
10p11.22	33231119	33287171	56.052	1	ITGB1
10q11.21	44778909	44804392	25.483	2	RASSF4
10q21.2	62317787	62431144	113.357	1	RHOBTB1
10q21.3	69536826	69640844	104.018	1	MYPN
10q21.3	69730465	69762648	32.183	1	PBLD
10q24.2	99324084	99333368	9.284	1	ANKRD2
10q25.1	106103709	106106235	2.526	2	CCDC147
10q25.2	114128606	114177763	49.157	1	ACSL5
10q26.11	120779322	120780972	1.65	1	NANOS1
10q26.12	121478692	121578038	99.346	1	INPP5F
10q26.13	124215266	124264200	48.934	2	HTRA1
10q26.13	124743238	124758245	15.007	1	IKZF5
11p15.5	278928	285642	6.714	1	ATHL1
11q12.3	62973282	63015147	41.865	2	HRASLS5
11P11.2	47330263	47427282	97.019	1,2	SPI1, RAPSN
11p15.5	1898505	1916357	17.852	2	TNNT3
11p15.3	12092713	12241401	148.688	2	MICAL2
11p11.2	45637655	45643584	5.929	1	CHST1
11p11.2	47638037	47692751	54.714	1	AGBL2
11q12.1	57079389	57138518	59.129	1	UBE2L6, SERPING1
11q12.2	60438242	60447436	9.194	3	TMEM109
11q12.3	63064947	63087387	22.44	3	RARRES3, HRASLS2
11q13.2	66114705	66117058	2.353	4	CCDC87
11q13.4	73730096	73787027	56.931	1	PGM2L1
11q13.5	76980583	76998194	17.611	2	AQP11
11q14.1	77404751	77411989	7.238	2	KCTD14
11q14.2	86336369	86344063	7.694	2	FZD4
11q14.2	86426396	86709231	282.835	1	TMEM135
11q14.2	87666768	87709775	43.007	1	CTSC
11q22.2	101797089	102399026	601.937	2	MMP7, MMP10, MMP1, MMP3, MMP13
11q23.3	117570358	117600835	30.477	1	AMICA1
11q23.3	118794199	118800044	5.845	1	THY1
12q21.31	83954316	84160801	206.485	1	LRRIQ1
12p13.31	6431637	6440583	8.946	3	TAPBPL

12p13.31	6519363	6535496	16.133	3	IFFO1
12p13.31	6751897	6757816	5.919	3	LAG3
12p13.31	6894036	6902891	8.855	3	ENO2
12p13.31	6974939	7046058	71.119	1	C1S
12p13.31	7167563	7172746	5.183	1	RBP5
12p11.21	32152836	32223652	70.816	1	BICD1
12q13.12	47675392	47679321	3.929	1	DDN
12q13.12	48738561	48763522	24.961	1	ACCN2
12q13.13	51058305	51065637	7.332	1	KRT84
12q13.13	52619392	52626446	7.054	4	HOXC13
12q13.13	52972184	52975801	3.617	1	NFE2
12q13.2	54515588	54522936	7.348	1	MMP19
12q13.3	55021709	55040075	18.366	1	STAT2
12q23.3	103979457	104002318	22.861	1	ALDH1L2
12q23.3	107214336	107257185	42.849	1	CMKLR1
12q24.13	111838770	111853973	15.203	1	OAS1, OAS2
12q24.13	112320466	112359994	39.528	1	SDS, SDSL
13q21.2	59282895	59290208	7.313	2	DIAPH3
13q32.3	99413301	99436918	23.617	2	ZIC5, ZIC2
14q32.2	99870220	99912155	41.935	1	WARS
14q32.13	93713859	93810227	96.368	1	PPP4R4
14q32.13	94148362	94159975	11.613	1	SERPINA3
14q22.2	54295792	54325384	29.592	3	SAMD4A
14q31.1	80754809	80931555	176.746	1	STON2
14q31.3	85073759	85162748	88.989	1	FLRT2
14q31.3	87472049	87529642	57.593	1	GALC
14q32.13	93471727	93493383	21.656	1	ASB2
14q32.13	93617425	93652468	35.043	1	IFI27L1, IFI27
14q32.31	101298118	101462533	164.415	2	PPP2R5C
14q32.33	104547601	104558273	10.672	4	CDCA4
15q24.1	72620931	72676274	55.343	1	ARID3B
15q25.2	82949282	82967115	17.833	1	ZSCAN2
15q25.2-	82999397	83002792	3.395	1	NMB
15q25.3					
15q26.2	92612534	92822582	210.048	1	MCTP2
15q11.2	18887165	18970887	83.722	3	HERC2P2
16q24.3	88348597	88410356	61.759	1	FANCA
16p13.13	11991779	12000539	8.76	1	RUNDC2A
16q12.1	49337161	49345341	8.18	1	CYLD
16q22.1	66525882	66528243	2.361	2	PSMB10
16q22.1	66579726	66590793	11.067	2	DPEP2
16q24.1	85158357	85160018	1.661	1	FOXL1
17q21.31	40184574	40184919	0.345	1	DBF4B
17q25.1	71452224	71486882	34.658	1	ACOX1
17p11.2	18864731	18869144	4.413	1	GRAP
17p11.2	24048981	24053357	4.376	1	SUPT6H
17q21.31	39497185	39499525	2.34	1	LSM12
17q21.32	43480047	43488533	8.486	1	NFE2L1
17q21.32	43984403	44006794	22.391	1	HOXB3
17q23.1	50829759	50854316	24.557	1	MMD
17q23.2	53404297	53420590	16.293	1	VEZF1
17q23.2	54587515	54634550	47.035	1	PRR11

18q11.2	18006926	18035637	28.711	3	GATA6
18q11.2	19541601	19759913	218.312	3	LAMA3
18q12.1	27337511	27380493	42.982	1	DSG2
18q21.1	45360066	45364337	4.271	3	LIPG
19q13.42	60335823	60352240	16.417	1	TNNT1
19q13.42	60688438	60690717	2.279	1	NAT14
20q13.12	44070959	44078427	7.468	2	MMP9
20q13.12	43888615	43895771	7.156	1	TNNC2
20q11.22	33593171	33608629	15.458	1	ERGIC3
20q11.21	30820954	30860463	39.509	1	DNMT3B
20q11.22	33222904	33228740	5.836	1	PROCR
20q11.23	34954902	35013513	58.611	1	SAMHD1
20q11.23	36190424	36226992	36.568	1	TGM2
20q13.12	42538389	42556442	18.053	1	TTPAL
20q13.12	43185486	43186512	1.026	1	WFDC12
20q13.12	43185564	43238568	53.004	1	PI3
20q13.12	43875211	43878987	3.776	1	UBE2C
20q13.13	47101835	47146360	44.525	1	CSE1L
20q13.13	47289194	47328156	38.962	1	ZNFX1
20q13.2	52529429	52539281	9.852	2	DOK5
20q13.32	55612686	55719537	106.851	1	ZBP1, PMEPA1
20q13.32	56468502	56523193	54.691	1	APCDD1L
21q22.12	35082078	35098101	16.023	1	RUNX1
21q22.11	33723598	33728314	4.716	2	IFNGR2
21q11.2- 21q21.1	15263756	15355324	91.568	1	NRIP1

bp*=base pair; kb*=kilo base pair

Table 4.3: Details of the deleted regions for each chromosome in 12 tongue SCC cases.

Chromosome (Cytoband)	Position Start (bp*)	Position Stop(bp*)	Size (kb*)	Frequency No	Gene Name
2p23.2	28235311	28242770	7.459	1	BRE
2p23.2	29671435	29678404	6.969	1	ALK
2q21.1	131950134	131956930	6.796	2	TUBA3D
2q33.3	205164282	205440801	276.519	1	PARD3B
2q34	211243669	211251963	8.294	1	CPS1
2q34	214582908	214586990	4.082	1	SPAG16
4q11	52612324	52658188	45.864	1	SPATA18
4q12	54020883	54152305	131.422	1	LNX1
4q12	56731115	56733122	2.007	1	KIAA1211
6q21	108828360	108832065	3.705	2	LACE1
6q26	162979182	162994871	15.689	1	PARK2
7q32.1	126267469	126282441	14.972	1	GRM8
9p24.1	6231654	6247861	16.207	1	IL33
9p22.3	6231654	6247861	16.207	1	c9orf93
9p22.2	17568481	17782593	214.112	1	SH3GL2
10q25.3	116929774	116963934	34.16	1	ATRNL1
11p15.1	20917601	20944649	27.048	1	NELL1
18q12.1	30007775	30019752	11.977	1	NOL4
19p12	20595355	20635991	40.636	1	ZNF626
19p12	20897856	20924883	27.027	1	ZNF85
20p12.2	9467146	9767439	300.293	1	PAK7
20p12.1	13777292	13918422	141.13	1	SEL1L2
22q11.1	15822604	15868880	46.276	1	GAB4

bp*=base pair; kb*=kilo base pair

4.2.2 Chromosomal aberrations and detected genes involved in cheek SCC

Table 4.4 and 4.5 lists the amplified and deleted genes identified from cheek SCC, respectively. The genome wide information obtained from array CGH on cheek SCC showed amplification of 22q12.3 as the most frequently observed chromosomal aberration (25%) and the gene identified in this region was APOL1. On the contrary, the most deleted regions were identified in the chromosome 2q22.1 (25%), 7q35 (25%) and 19q13.33 (25%) which mapped with LRP1B, CNTNAP2 and FUT2 genes, respectively.

Table 4.4: Details of the amplified regions for each chromosome in 8 cheek SCC cases.

Chromosome (Cytoband)	Position Start (bp*)	Position Stop(bp*)	Size (kb*)	Frequency No	Gene Name
1q25.1	173300562	173379842	79.28	1	TNN
1q25.2	177335027	177335306	0.279	1	ABL2
2p23.1	30313491	30348049	34.558	1	LBH
3p25.2	12950575	12983916	33.341	1	IQSEC1
3p22.1	42701984	42717935	15.951	1	KBTBD5, HHATL
3q21.1	123741593	123776009	34.416	1	PARP9, DTX3L
4q11	52612495	52628952	16.457	1	SPATA18
6q21	112654596	112662700	8.104	1	LAMA4
7p14.3	32964292	33012655	48.363	1	FKBP9
7p14.1	41695169	41709208	14.039	1	INHBA
7q32.1	128190946	128198375	7.789	1	CALU
8p12	31759525	31766348	6.823	1	NRG1
9p13.3	35663862	35671065	7.203	1	CA9
9p13.3	35063878	35069996	6.118	1	FANCG
9p13.3	35087688	35093116	5.428	1	STOML2
9p13.3	34979730	34988341	8.611	1	DNAJB5
9p13.2	38382700	38388633	5.933	1	ALDH1B1
9q21.13	73372022	73564287	192.265	1	TMEM2
19p13.3	6615584	6621581	5.997	1	TNFSF14
19q13.33	55084766	55092186	7.42	1	IL4I1
22q12.3	34978564	34988898	10.334	2	APOL1

bp*=base pair; kb*=kilo base pair

Table 4.5: Details of the deleted regions for each chromosome in 8 cheek SCC cases.

Chromosome (Cytoband)	Position Start (bp*)	Position Stop(bp*)	Size (kb*)	Frequency No	Gene Name
1p22.3	86159237	86175251	16.014	1	COL24A1
2q22.1	131950134	131956930	6.796	2	LRP1B
3p26.3	213534	423608	210.074	1	CHL1
3p26.3	1108466	1418836	310.37	1	CNTN6
3p26.3-	2255213	3070579	815.366	1	CNTN4
3p26.2					
3p25.3	9666093	9718597	52.504	1	MTMR14
3p25.3	9720475	9746339	25.864	1	CPNE9
3p25.3	9871336	9907828	36.492	1	CIDEC
3p25.3	11009390	11055504	46.114	1	SLC6A1
3p25.2	12304193	12449328	145.135	1	PPARG
3p25.2	12813148	12850999	37.851	1	TMEM40
3p25.1	14691659	14788696	97.037	1	C3orf20
3p24.1	30623156	30867024	243.868	1	TGFBR2
3p24.1	30623156	30867024	243.868	1	GADL1
3p23	32255093	32385426	130.333	1	CMTM8
3p22.3	36397255	36562637	165.382	1	STAC
3p22.3	37468570	37835081	366.511	1	ITGA9
3p22.2	38470768	38509347	38.579	1	ACVR2B
3p22.1	41512795	41552303	39.508	1	ULK4
3p21.31	46452680	46481467	28.787	1	LTF
3p21.31	46728608	46734352	5.744	1	TSP50
3p21.31	49566923	49683032	116.109	1	BSN
3p21.31	49696399	49701098	4.699	1	MST1
3p21.2	52255225	52259610	4.385	1	PPM1M
3p14.3	54518914	54660981	142.067	1	CACNA2D3
3p14.2	62330439	62334204	3.765	1	FEZF2
3p14.1	63971362	63984450	13.088	1	PSMD6
3p14.1	69871009	70098183	227.174	1	MITF
3p13	73512066	73756515	244.449	1	PDZRN3
3p12.1	87069733	87122868	53.135	1	VGLL3
3p11.2	87391526	87408424	16.898	1	POU1F1
7q35	145776903	145923085	146.182	2	CNTNAP2
8p23.1	10501415	10606095	104.68	1	RP1L1
8p23.1	10618655	10625437	6.782	1	SOX7
8p21.3	23299350	23367704	68.354	1	ENTPD4
8p21.1	27492216	27527997	35.781	1	CLU
8p12	38153461	38189731	36.27	1	BAG4
10q11.22	46413545	46420541	6.996	1	GPRIN2
10q11.22	56841264	56851024	9.76	1	PPYR1
10q21.3	67856106	67862075	5.969	1	CTNNA3
12p12.1	25552267	25558944	6.677	1	IFLTD1
14q31.1	79176042	79183210	7.168	1	NRXN3
15q13.1	25856742	25861802	5.06	1	OCA2
16p13.2	7336826	7343082	6.256	1	A2BP1
16p12.1	48082088	48371710	289.622	1	ZNF423
18p11.23	7962551	7967038	4.487	1	PTPRM

19q13.33	53892650	53900934	8.284	2	FUT2
20q13.12	45214224	45220304	6.08	1	EYA2

bp*=base pair; kb*=kilo base pair

4.3 Significant signaling pathways analysis from data sets of chromosomal aberrations in tongue and cheek SCC

To elucidate which biological pathways are affected between tongue and cheek SCC, the amplified and deleted genes were subjected to pathway analysis using MetaCore Software. Summaries of the significant pathways ($p < 0.05$) identified from amplified and deleted genes from tongue SCC are illustrated in tables 4.6 and 4.7 respectively. The pathway analysis revealed 40 and 17 significant pathways affected by amplified and deleted genes, respectively in tongue SCC. Pathway analysis of the amplified and deleted genes that were found in cheek SCC yielded 20 and 10 significant pathways ($p < 0.05$), respectively and are illustrated in table 4.8 and 4.9.

4.3.1 Significant signaling pathways of tongue SCC

In depth biological analysis by MetaCore™ Software revealed that the most significant pathway involved in the amplified gene was cell adhesion ECM remodeling pathway. This study determined that 11/187 of amplified genes were member of cell adhesion ECM remodeling pathway. The amplified genes involved in this pathway include EGFR, ITGB1, LAMA3, LAMB3, MMP1, MMP10, MMP2, MMP13, MMP3, MMP7, MMP9. On the other hand, the most significant pathway involved in the deleted genes of tongue SCC was proteolysis (Role of Parkin) in Ubiquitin-Proteasomal Pathway. The deleted genes involved in this pathway were PARK2 and TUBA3D.

Table 4.6: Significant biological pathway associated with amplified genes from tongue SCC.

No	Pathway	P value	Gene(s)
1	Cell adhesion ECM remodeling	2.942e-10	EGFR, ITGB1, LAMA3, LAMB3, MMP1, MMP10, MME (MMP12), MMP13, MMP3, MMP7, MMP9
2	Immune response Antiviral actions of Interferons	3.888e-6	IFNGR2, IRF1, OAS1, OAS2, STAT2, WARS
3	Immune response Histamine H1 receptor signaling in immune response	4.360e-5	MMP1, MMP13, MMP3, MMP9, PPP3CC
4	Cell adhesion Cell-matrix glycoconjugates	2.568e-4	ITGB1, MMP1, MMP3, MMP9, TNC
5	Regulation of lipid metabolism Regulation of fatty acid synthase activity in hepatocytes	4.786e-4	ADD1
6	Development Role of IL-8 in angiogenesis	1.304e-3	EGFR, ADD1, ARPC1B
7	Development Angiotensin signaling via STATs	2.268e-3	PPP3CC, PTK2B, STAT2
8	Immune response Oncostatin M signaling via MAPK in mouse cells	2.941e-3	MMP1, MMP13, MMP3
9	Immune response Oncostatin M signaling via MAPK in human cells	3.450e-3	MMP1, MMP13, MMP3
10	Regulation of lipid metabolism Regulation of lipid metabolism via LXR, NF-Y and SREBP	3.724e-3	ADD1
11	Transcription Androgen Receptor nuclear signaling	6.011e-3	EGFR, FZD4, PTK2B
12	Immune response Oncostatin M signaling via JAK-Stat in mouse cells	9.490e-3	MMP1, SERPINA3
13	Immune response Oncostatin M signaling via JAK-Stat in human cells	1.166e-2	MMP1, SERPINA3
14	Cytoskeleton remodeling TGF, WNT and cytoskeletal remodeling	1.327e-2	FZD4, MMP13, MMP7, SERPING1, MMP1, ARPC1B
15	Immune response IL-17 signaling pathways	1.328e-2	MMP1, MMP3, MMP9
16	Development Regulation of epithelial-to-mesenchymal transition (EMT)	1.579e-2	EGFR, FZD4, MMP9
17	Immune response IFN alpha/beta signaling pathway	1.658e-2	IRF1, STAT2
18	Glycolysis and gluconeogenesis p.3 / Human version	1.658e-2	ENO2

19	Glycolysis and gluconeogenesis p.3	1.658e-2	ENO2
20	Transport RAB5A regulation pathway	1.793e-2	EGFR,INPP5F
21	Transcription Transcription regulation of aminoacid metabolism	1.793e-2	NFE2, NFE2L1
22	Cytoskeleton remodeling Reverse signaling by ephrin B	2.692e-2	LAP3, RGS3, ARPC1B
23	Beta-alanine metabolism/ Rodent version	2.856e-2	ABP1, GAD1
24	Development Transcription regulation of granulocyte development	2.856e-2	RUNX1, SPI1
25	Development Angiotensin activation of ERK	3.025e-2	EGFR, PTK2B
26	G-protein signaling G-Protein alpha-q signaling cascades	3.197e-2	PTK2B, RGS3
27	Immune response CXCR4 signaling via second messenger	3.197e-2	LAP3, PTK2B
28	Fatty Acid Omega Oxidation	3.197e-2	ACOX1, ACSL5
29	Chemotaxis CXCR4 signaling pathway	3.197e-2	LAP3, PTK2B
30	Keratan sulfate metabolism p.1	3.373e-2	CHST1, CHST2
31	Cell adhesion Plasmin signaling	3.373e-2	MMP1, MMP13, SERPING1
32	Beta-alanine metabolism	3.373e-2	ABP1, GAD1
33	Regulation of lipid metabolism Insulin regulation of fatty acid methabolism	3.617e-2	ADD1
34	Regulation of metabolism Bile acids regulation of glucose and lipid metabolism via FXR	3.736e-2	ADD1
35	HIV-1 signaling via CCR5 in macrophages and T lymphocytes	4.113e-2	PPP3CC, PTK2B
36	Keratan sulfate metabolism p.2	4.113e-2	CHST1, CHST2
37	Translation Non-genomic (rapid) action of Androgen Receptor	4.307e-2	EGFR, FZD4
38	Apoptosis and survival BAD phosphorylation	4.704e-2	EGFR, PPP3CC
39	Development ACM2 and ACM4 activation of ERK	4.908e-2	EGFR, PTK2B
40	Cell adhesion Chemokines and adhesion	4.973e-2	ARPC1B, ITGB1, LAMA3, MMP1, MMP13

Table 4.7: Significant biological pathway associated with deleted genes from tongue SCC.

No	Pathways	P value	Gene(s)
1	Proteolysis Role of Parkin in the Ubiquitin-Proteasomal Pathway	1.704e-4	PARK2, TUBA3D
2	UMP biosynthesis	1.693e-2	CPS1
3	Proteolysis Putative ubiquitin pathway	1.945e-2	PARK2
4	Cytoskeleton remodeling Role of PDGFs in cell migration	2.029e-2	PAK7
5	Cytoskeleton remodeling Neurofilaments	2.112e-2	TUBA3D
6	Neurophysiological process Role of CDK5 in presynaptic signaling	2.363e-2	SH3GL2
7	Cell adhesion Gap junctions	2.530e-2	TUBA3D
8	Cytoskeleton remodeling Reverse signaling by ephrin B	2.614e-2	TUBA3D
9	Cell cycle Role of Nek in cell cycle regulation	2.697e-2	TUBA3D
10	Apoptosis and survival Granzyme B signaling	2.697e-2	TUBA3D
11	Cell cycle Spindle assembly and chromosome separation	2.780e-2	TUBA3D
12	Nitrogen metabolism	2.947e-2	CPS1
13	Nitrogen metabolism/ Rodent version	3.030e-2	-
14	Cytoskeleton remodeling Keratin filaments	3.030e-2	TUBA3D
15	Aldosterone biosynthesis and metabolism	3.444e-2	CPS1
16	Apoptosis and survival TNFR1 signaling pathway	3.610e-2	BRE
17	Apoptosis and survival FAS signaling cascades	3.610e-2	BRE

4.3.2 Significant signaling pathways of cheek SCC

In cheek SCC, results obtained with MetaCore software showed cytoskeleton remodeling (Role of Activin A) in cytoskeleton remodeling and cadherin mediated cell adhesion pathway were the top biological pathway associated with amplified and deleted genes, respectively. It was found that 2 of the 48 deleted genes were members of the cadherin mediated cell adhesion pathway, including CTNNA3 and PTPRM genes. In contrast, the identified amplified gene which falls in the cytoskeleton remodeling (Role of Activin A) in cytoskeleton remodeling pathway was INHBA.

Table 4.8: Significant biological pathway associated with amplified genes from cheek SCC.

No	Pathways	P value	Gene(s)
1	Cytoskeleton remodeling Role of Activin A in cytoskeleton remodeling	1.933e-2	INHBA
2	Atherosclerosis Role of ZNF202 in regulation of expression of genes involved in Atherosclerosis	2.028e-2	APOL1
3	Cholesterol and Sphingolipids transport / Recycling to plasma membrane in lung (normal and CF)	2.220e-2	APOL1
4	Apoptosis and survival Apoptotic Activin A signaling	2.411e-2	INHBA
5	Histamine metabolism	3.076e-2	ALDH1B1
6	Signal transduction Activin A signaling regulation	3.171e-2	INHBA
7	Apoptosis and survival Role of CDK5 in neuronal death and survival	3.266e-2	NRG1
8	Ascorbate metabolism / Rodent version	3.266e-2	ALDH1B1
9	Nitrogen metabolism	3.361e-2	CA9
10	Nitrogen metabolism/ Rodent version	3.455e-2	CA9
11	Development ERBB-family signaling	3.738e-2	NRG1
12	Development Role of Activin A in cell differentiation and proliferation	3.833e-2	INHBA
13	Translation Non-genomic (rapid) action of Androgen Receptor	3.833e-2	NRG1
14	Apoptosis and survival Lymphotoxin-beta receptor signaling	3.927e-2	TNFSF14
15	Apoptosis and survival Apoptotic TNF-family pathways	4.021e-2	TNFSF14
16	Development Ligand-independent activation of ESR1 and ESR2	4.209e-2	NRG1
17	Role of alpha-6/beta-4 integrins in carcinoma progression	4.302e-2	NRG1
18	Niacin-HDL metabolism	4.490e-2	APOL1
19	Pyruvate metabolism	4.677e-2	ALDH1B1
20	Cell adhesion ECM remodeling	4.957e-2	LAMA4

Table 4.9: Significant biological pathway associated with deleted genes from cheek SCC.

No	Pathways	P value	Gene(s)
1	Cell adhesion Cadherin-mediated cell adhesion	8.569e-4	CTNNA3, PTPRM
2	Cell adhesion Role of CDK5 in cell adhesion	1.525e-2	CTNNA3
3	Regulation of lipid metabolism Regulation of fatty acid synthesis: NLTP and EHHADH	3.028e-2	PPARG
4	Development TGF-beta-induction of EMT via ROS	3.193e-2	TGFBR2
5	Cytoskeleton remodeling Role of Activin A in cytoskeleton remodeling	3.359e-2	ACVR2B
6	Cell adhesion Endothelial cell contacts by non-junctional mechanisms	4.018e-2	CTNNA3
7	Apoptosis and survival Apoptotic Activin A signaling	4.182e-2	ACVR2B
8	Cell adhesion Endothelial cell contacts by junctional mechanisms	4.346e-2	CTNNA3
9	Possible pathway of TGF-beta 1-dependent inhibition of CFTR expression	4.509e-2	TGFBR2
10	Regulation of lipid metabolism RXR-dependent regulation of lipid metabolism via PPAR, RAR and VDR	4.999e-2	PPARG

5.0 Discussion

This study was performed to determine the chromosomal aberrations (alterations) in tongue and cheek SCC using array CGH technology and to identify genes that act in similar pathway. The genes which interact together during oncogenesis that are thought to link the genes in various copy number aberrations *via* pathway analysis have been identified in order to strengthen this study.

To the best of our knowledge, this is the first study using an ultra dense array CGH for the discovery of cancer genes on the subsites of OSCC (tongue and cheek) in which ultra high (2.1kb) resolution definition of DNA copy number aberrations were generated from frozen tissue sections. Strength of our study is that the chromosomal aberrations and pathways involved in oral carcinoma genesis were analyzed separately based on the specific subsites within the oral cavity. Moreover, this study also managed to show the micro-copy number aberrations (CNAs), defined as < 1Mb in genomic length. This demonstrated the sensitivity of the ultra-dense array CGH on gDNA of frozen tissues compared with others studies which showed only macro-CNAs. (Chen *et al.*, 2004b; Garnis *et al.*, 2004b; Baldwin *et al.*, 2005; Snijders *et al.*, 2005; Liu *et al.*, 2006; Sparano *et al.*, 2006; Jarvinen *et al.*, 2008).

Previously, most of the studies performed genome wide screening using conventional CGH and PCR amplification on bacterial artificial chromosomes (BACs) clones spotted in the array CGH. But in these studies there was a limitation with probe used which started from 3-1Mb. This limitation produced high cross hybridization artifacts in the data analysis process and thus decreased the sensitivity and resolution that hindered the implementation of array CGH in research and diagnostics. Although there were few studies which began to use 60-mer oligonucleotide array CGH to analyse whole DNA

copy number profiles of OSCC, but most of these studies utilised 4x44k and 105K oligonucleotide array CGH. However, the probe resolution was not as high as the ultra dense 1 Million Mb oligonucleotide array CGH to detect small genomic regions with amplification and deletions as in this study. Besides that, the present study was the first which took an effort to study the distinct subsites of OSCC (tongue and cheek SCC) by using the ultra dense array CGH to discover the cancer genes on these 2 subsites and the biological pathways involved in these 2 subsites compared to other studies.

Although there were few studies that carried out genome wide analysis on OSCC as in the current study, none had provided signaling pathways based on the specific anatomical subsites of OSCC. A study on proteomic analysis of oral cancer cell line derived from tongue, cheek and gum SCC by Chi *et al.* (2009) showed that the most significant pathway involved in all 3 cell lines were cytoskeleton remodeling keratin filament pathway and type 1 IFN signaling pathway. However, they did not separate the pathways and proteins involved based on the specific subsite. Another study by Wang *et al.* (2009), to detect the potential pathways for the malignant transformation from dysplasia to OSCC also did not analyzed the pathways based on the different subsites.

5.1 Chromosomal aberrations (alterations) in tongue and cheek SCC using array CGH

In the present study, chromosomal aberrations were observed in all cases. Frequent copy number aberrations included gains of 3q, 7p, 8q,9q, 11q, and 14q as well as losses of 3p, 8p, 9p and 18q which is in accordance with previous studies related to OSCC (Chen *et al.*, 2004b; Garnis *et al.*, 2004b; Baldwin *et al.*, 2005; Snijders *et al.*, 2005; Liu *et al.*, 2006; Sparano *et al.*, 2006; Jarvinen *et al.*, 2008) and other solid tumors (Beroukhim *et al.*, 2010). Current study demonstrated that the number of chromosomal aberrations in tongue SCC is higher than cheek SCC and this might be attributed to the more aggressive behavior of tongue in terms of invasiveness and metastasis (Hannen *et al.*, 2004). This can be further supported by Hannen *et al.* (2004) study which demonstrated that metastatic of tongue SCC have higher copy number gains than non metastatic SCC. However, it was noticed in the current study that the deletion copy number of cheek SCC was higher than tongue SCC. This finding can be explained by the fact that those with cheek SCC in this study population were betel quid chewers. Earlier, studies of Lee *et al.* (2001b) have shown that there are high accumulations of mitochondrial DNA deletions in oral tissues collected from quid chewers. This further explained the high incidence of cheek SCC among Southeast Asian countries where betel quid chewing are commonly practiced (Silverman, 2003; Pathak *et al.*, 2009).

In tumorigenesis, gene amplification is an essential step to activate the oncogenes (Coleman and Tsongalis, 2002). Notably, the most frequently amplified genes showed in tongue SCC were WISP1 (8q24.22), DENND3 (8q24.3), CCDC87 (11q13.2), HOXC13 (12q13.13) and CDCA4 (14q32.33) whereas for cheek SCC, the most frequently amplified gene was APOL1 gene (22q12.3). WISP1 (Wnt-induced secreted protein 1) is one of the CCN (CTGF/Cyr61/Nov) family. They have the biological

properties such as regulation of cell proliferation, adhesion, apoptosis, extracellular matrix production, migration and growth arrest (Bringstock, 2003). WISP1 acts as β -catenin regulated genes that are involved in tumorigenesis. They interact with DNA binding factors which lead to activation of uncontrolled downstream genes that are normally involved in cancer (Xu *et al.*, 2000). WISP1 has previously been reported to be amplified and over expressed in cholangiocarcinoma (Tanaka *et al.*, 2003) and breast (Xie *et al.*, 2001), lung (Chen *et al.*, 2007b), colon (Pennica *et al.*, 1998) and rectal cancers (Soon *et al.*, 2003). Soon *et al.* (2003) in their study showed over expression of WISP1, which was found to inhibit lung metastasis and invasion. Contrasting to these results was reported by Davies *et al.* (2005). They revealed that WISP1 expression was inversely linked to poor prognosis. Owing to these controversy outcomes, further studies on WISP1 in OSCC is needed.

Amplification of DENND3 gene was also seen in lymphoblastoid cell lines from autism and micropapillary carcinomas (Nishimura *et al.*, 2007; Marchi *et al.*, 2009). The function of DENND3 gene remains unclear but identification of this gene in invasive ductal breast carcinoma warrants further study (Marchi *et al.*, 2009). CCDC87 genes are encoded as a member of the coiled coil domain family, which function as dimerization domains for a wide variety of proteins, but its contribution to oncogenesis remains unclear. The gene CCDC87 was reported in invasive and non invasive melanoma cells (Berthier-Vergnes *et al.*, 2011). The occurrence of this particular gene in the present study associated with oral cancers reflects further studies on this to understand its role in cancer progression.

HOXC13 is a homeobox-containing gene which encodes transcription factors that play important roles in embryogenesis and organogenesis (Yamada and Tsuchida, 2009). Dys-regulation of HOXC13 is well documented to be associated with oncogenesis (Lawrence *et al.*, 1996; Lappin *et al.*, 2006). HOXC13 gene has been shown to get amplified at chromosome 12q13.13 in tongue SCC cell lines (Tsui and Garnis, 2010) and Hodgkin lymphoma (Hartmann *et al.*, 2008). Besides that, HOXC13 gene activation could be resulted from chromosomal translocation, where they have fused with NUP98 gene when associated with acute myeloid Leukemia (La Starza *et al.*, 2003; Vogelstein and Kinzler, 2004; Yamada *et al.*, 2008). However, HOXC13 genes have been identified as methylated gene in Methylated-CpG island recovery assay-assisted microarray analysis on fresh frozen breast cancer tissues (Tommasi *et al.*, 2009). This observation was in agreement with others cancer such as colon and leukemia cell line (Yeang, 2010). To date, the regulatory mechanism of HOXC13 still remains unknown (Ansari *et al.*, 2009), thus warranting further downstream analysis of this gene to confirm their roles in tumorigenesis.

The current study also identified CDCA4 gene in the amplified region of 14q32.33. This gene encodes a protein that belongs to the E2F family of transcription factors and act as an important transcriptional regulator in DNA replication, cell cycle progression, DNA damage repair, apoptosis, cell differentiation, development and cell proliferation regulation through E2F/retinoblastoma protein pathway (Nevin, 1998; Hayashi *et al.*, 2006). CDCA4 has previously been reported to be associated with retinoblastoma (Hayashi *et al.*, 2006). However, literature on association of this gene with cancer is lacking.

The most amplified gene in cheek SCC was mapped on chromosome 22q12.3 which contains APOL1. Of note, this gene encodes secreted high density lipoprotein which is involved in activities such as trypanosome lysis, autophagic cell death, lipid metabolism, cellular senescence and others vascular activities (Tzur *et al.*, 2010). Other researchers have also identified the amplification and over expression of APOL1 gene in cervical carcinoma (Wilting *et al.*, 2008), kidney disease (Tzur *et al.*, 2010), prostate cancer (Thierry-Mieg and Thierry-Mieg, 2006), renal cell carcinoma (Kruger *et al.*, 2005; Camp *et al.*, 2006). The aggressive behaviour of this gene was showed by Huang *et al.* (2003), where they indicated that APOL1 gene was associated with metagene predictor of lymph node metastasis in breast cancer. However, a contrasting finding was reported by Kim *et al.* (2010) who showed that APOL1 was down regulated. This highlighted the need for further elucidation of the role of APOL1 in OSCC.

In cancer, losses of chromosomal materials most likely will harbor inactivated tumor suppressor genes that leads to tumorigenesis (Furukawa *et al.*, 2006; Frohling and Dohner, 2008). This study identified that the most frequently deleted genes were TUBA3D (2q21.1), LACE1 (6q21), LRP1B (2q22.1), CNTNAP2 (7q35) and FUT2 (19q13.33). TUBA3D is a gene that encodes microtubule constituents and is a part of the alpha tubulins superfamily (Khodiyar *et al.*, 2007). This gene, together with beta tubulin represents the major components of microtubules that are involved in cell division, cell polarity, structural support and intracellular transport (Dode *et al.*, 1997). The down regulation of these genes have been previously reported in Ewing Sarcoma (Mukhopadhyay *et al.*, 2010), while Ben-Chetrit *et al.* (2005) reported that the loss of TUBA3D will suppress the anti inflammatory effects in cancer by decreasing the neutrophil migration towards the inflammatory site.

LACE1 has been reported to be down regulated in B-cell lymphoma cell lines (Inomata *et al.*, 2009) and they are also reported to be deleted in early stages of lung adenocarcinoma (Mikse *et al.*, 2010). Although LACE1 has been implicated as a tumor suppressor gene, its genetic inactivation has not been observed in any cancer. Thus, further functional evaluation of this gene is worth for investigation. The deletion of LRP1B gene as found in the present study is in concordance with previous reports. In early studies this gene was frequently reported to be silenced and under expressed in oral cancer (Cengiz *et al.*, 2007), esophageal SCC (Sonada *et al.*, 2004), urothelial cancer (Langbein *et al.*, 2002) and B-cell lymphoma (Rahmatpanah *et al.*, 2006). Initially, LRP1B was termed as LRP-DIT (deleted in Tumors) and was considered as a member of the LDL receptor family (Dietrich *et al.*, 2010). As the name implies, this gene is suggested as the potential tumor suppressor gene happening due to the specific deletion through genetic and epigenetic mechanism on CpG Island (Marschang *et al.*, 2004; Dietrich *et al.*, 2010). Both studies of Liu *et al.* (2000a) and (2000b) showed that loss and inactivation of LRP1B plays an important role in enhancing tumor invasion and metastasis through regulation of extracellular proteolysis and cell motility in late stages of lung cancer progression. Although loss of LRP1B have been reported in oral cancer (Cengiz *et al.*, 2007), the mechanism of this gene remains unclear, further downstream analysis should be performed to clarify its role in oral carcinogenesis.

Traditionally, CNTNAP2 gene has been found to be associated with neurological disorders as this gene is a member of neurexins superfamily, which is a group of transmembrane protein that assist cell-cell interactions in nervous system (Baumgarther *et al.*, 1996). Deletion of this gene could lead to unwanted movement due to the disturbance of K⁺ channel distribution in the nervous system (Verkerk *et al.*, 2003). Previous studies reported the association of this gene with focal epilepsy (Friwdman *et*

al., 2008), schizophrenia (Friwdman *et al.*, 2008), autism (Alarcon *et al.*, 2008), speech and language disorders (Petrin *et al.*, 2010) and Gilles de la Tourette syndrome (Verkerk *et al.*, 2003). Interestingly, this gene was also found to be deleted in the present study. Moreover, this gene was also associated with other types of cancer such as pancreatic adenocarcinoma (Omura *et al.*, 2008), neuroblastoma (Stalling *et al.*, 2006), myeloid leukemia cell line (McAvoy *et al.*, 2007), breast (McAvoy *et al.*, 2007), ovarian (McAvoy *et al.*, 2007), endometrial (McAvoy *et al.*, 2007) and brain cancers (McAvoy *et al.*, 2007). The nature of this gene which was found at the common fragile site explains its nature as a TSG (Smith *et al.*, 2005). However, this suggestion is in disagreement with Chen *et al.* (2008) where they showed that CNTNAP2 gene was over expressed in oral squamous cell carcinoma. This contradiction might be due Chen *et al.* (2008) take an effort to compare the gene expression of OSCC with normal control which derived from normal oral tissue and dysplastic oral lesion, and this comparison data hence will be different with our study. This suggested that CNTNAP2 expression is copy number independent and could be another primary driving force behind this gene deletion in chromosome 7q35.

FUT2 gene encodes an α -(1, 2) fucosyltransferase that controls the expression of the ABH antigens in saliva and mucosal tissues (Serpa *et al.*, 2004). The current knowledge suggests that homozygous nonsense mutation of this gene is strongly associated with resistance to novovirus infections (Thorven *et al.*, 2005). The deletion of FUT2 gene found in the present study was in agreement with the study of Massion *et al.* (2008), which performed a genome wide screening on frozen non small cell lung cancer tissues. However, the findings from both these studies are in contrast to the study conducted by Madjd *et al.* (2005), where the expression of FUT2 gene was inversely proportional to

survival in lymph node negative breast carcinomas. The controversy surrounding the involvement of FUT2 gene in tumorigenesis suggests further investigation of this gene.

5.2 Significant pathways involved in tongue and cheek SCC using pathway analysis software.

There is currently a limited understanding of the biological pathways involved between tongue and cheek SCC. The discovery of potential markers has represented a substantial challenge particularly for markers that are applicable to subsite of oral cancer. The interest in the current project was to identify the genes that interact together during oral carcinogenesis. Hence, the pathway analysis was carried out to associate the genes with various copy number alterations in oral cancer. Pathway analysis was performed to determine the biological pathways of the amplified and deleted genes to further understand the mechanism of oral carcinogenesis. Thirty-five (19.7%) out of 178 amplified genes were found to be enriched in tongue SCC analysis. Pathway analysis of these 35 genes identified four significant pathways using Metacore's Genego pathway tool. Some of the important top pathways generated by Genego include; cell adhesion extracellular matrix (ECM) remodeling, proteolysis role of Parkin in the Ubiquitin-Proteasomal, development of beta-adrenergic receptors signaling *via* cAMP and cell adhesion cadherin mediated cell adhesion pathway.

Identified significant pathways in tongue SCC is as illustrated in table 4.6 where the most significant pathway was related to cell adhesion ECM remodeling. Altogether 11 key downstream regulators were found in this pathway (EFGR, ITGB1, LAMA3, LAMB3, MMP1, MMP10, MMP12, MMP13, MMP3, MMP7 and MMP9) were identified. Identification of this pathway has been previously reported by Luna *et al.* (2009) where they suggested that the alteration of ECM in cell adhesion, ECM-remodeling pathway was highly induced by chronic oxidative stress. Besides that, other

researchers has shown that this pathway was involved in chronic pancreatitis, breast cancer and ovarian cancer (Pitteri *et al.*, 2009; Reding *et al.*, 2009; Bessarabova *et al.*, 2010). Of note, ECM remodeling is important in the development of malignancies due to proteolysis of the ECM, which allow tumor cells to invade and penetrate through basement membrane barrier in the process of tumor invasion and metastasis (Stetler-Stevenson and Yu, 2001). MMPs are known as a "protease in invasion" in this pathway. They degrade the various components of the ECM and favor metastases (Stetler-Stevenson and Yu, 2001). Proteolysis of the ECM is a crucial initial step in the mechanism of tumor cell invasion. In order for cells to invade and spread they must be able to first penetrate into connective tissue barriers such as the basement membrane, provisional matrix or interstitial stroma that surround tumors and existing blood vessels. The mechanism of this pathway in tumorigenesis can be explained by the release of active HB-EGF which is cleaved by MMP7 activated its receptors. The gene, EGFR were found to induce cell proliferation, cell survival and tissue remodeling (Rehault *et al.*, 2001; Yu *et al.*, 2002). The cell adhesion movement mainly caused by the binding of ITGB1 to collagen IV during ECM remodeling pathway favors cell invasion and metastasis (Li *et al.*, 2006). The involvement of this pathway in tongue SCC further explains its aggressive behavior in oral cancer.

Dys-regulation of proteolysis role of Parkin in the ubiquitin-proteosomal pathway has been previously reported in Bovine Leukemia virus infection (Li *et al.*, 2009) and arsenic induced bladder cancer (Sen *et al.*, 2007). The role of Parkin in these pathway was regulated by E3-ubiquitin protein that ubiquitinates and degrade the tubulin alpha and tubulin beta (Ren *et al.*, 2003). Basically, the deletion of Parkin increased the level of Parkin associated endothelin receptors like PAELR, leading to the dopaminergic neuron degeneration and apoptosis inhibition (Imai *et al.*, 2001). This is one of the

hallmarks in tumorigenesis as described by Hanahan and Weinberg (2000). This is further supported by an example of numb gene which is regulated by E3-ubiquitin where the lost expression of this gene in breast cancer leads to an uncontrolled cell proliferation (Pece *et al.*, 2004). This implied that PARK2A and TUBA3D might act as a negative regulator for this pathway in oral cancer through ubiquitination and proteasomal degradation where loss of these genes can increase cell proliferation. Our findings may further contribute to the understanding of key biologic functions and pathways of certain ubiquitin and proteasomal degradation protein genes that are associated with OSCC. However, further investigations into ubiquitin genes are relevant in oral carcinogenesis.

Amplification of INHBA gene was detected using array CGH in the present study and has been recognized as one of the member in the pathway of cytoskeleton remodelling (Role of Activin A). Activin A which is also known as INHBA is a member of the transforming growth factor beta (TGF- β) (Vale *et al.*, 2006). In this pathway, INHBA plays its role by binding to either activin A or B type II receptors (ActRIIA or ActRIIB), which are membrane-spanning serine/ threonine kinase receptors to form a complex (Vale *et al.*, 2006). These complexes commence a signaling cascade which involves specific transcription factors such as SMAD2, SMAD3 and SMAD4 (Mathews, 1994). The binding of SMAD4 to SMAD2 and SMAD3 forms a complex and is translocated into the nucleus to promote cell differentiation and cell proliferation (Zhang *et al.*, 2005). Furthermore, INHBA also activates RhoA which leads to the formation of transcription of independent actin stress fibres which are accounted for actin cytoskeleton reorganization and cell migration (Zhang *et al.*, 2005). This pathway has been reported by Nambiar *et al.* (2010) in a study on colon cancer metastasis. The involvement of INHBA in tumorigenesis have been widely studied in different type of

cancers such as prostate cancer, lung adenocarcinoma, endometrial adenocarcinoma, esophageal SCC, ovarian cancer, cervical carcinomas and colon carcinomas (Petraglia *et al.*, 1998; van Schaik *et al.*, 2000; Wildi *et al.*, 2001; Simone *et al.*, 2002; Woodruff, 2002; Yoshinaga *et al.*, 2007; Seder *et al.*, 2009). These findings suggested that this gene might be a potential therapeutic target for oral cancer.

Another pathogenic pathway identified in the present study was the ERBB-family signalling pathway and the gene involved in this process was NRG1 gene. The binding of NRG1 which is an extracellular growth factor ligands to ERBB3 which causes the activation of intracellular signaling pathways that regulates diverse biologic responses including proliferation, differentiation, cell motility, and survival in cancer (Sweeney and Carraway, 2000; Marmor *et al.*, 2004; Revillion *et al.*, 2008). These finding was supported by a study in ovarian cancer cell lines where activation of NRG1/ErbB3 autocrine loops increases cell proliferation (Sheng *et al.*, 2010). Furthermore, expressions of ErbB-3 and ErbB-4 have been previously reported in OSCC, childhood medulloblastoma and breast cancer (Xia *et al.*, 1999; Gilbertson *et al.*, 1997; Yarden, 2001).

The present study now report for the first time that CTNNA3 and PTPRM genes, which are associated with cadherin mediated cell adhesion pathway are deleted in OSCC. This finding showed that besides E-cadherin, others members of the junctional complex also accounts for the impaired cell to cell adhesion. Previously, this pathway have been reported in colon cancer (Nambiar *et al.*, 2010) and the deletion of CTNNA3 and PTPRM have been identified in lung squamous cell carcinoma, endometrial carcinoma and bladder cancer (Kim *et al.*, 2002; McDoniels-Silvers *et al.*, 2002; Meehan *et al.*, 2007). Early study showed that the loss of CTNNA3 has been positively associated with

tumor invasion and metastasis (Shiozaki *et al.*, 1996). This pathway suggested that CTNNA3 which is one of the catenin members will interact with other catenins to allow the association of cadherins to form catenin/cadherin complex (Goodwin and Yap, 2004). The connections of this complex to actin cytoskeleton *via* α -catenin together with the involvement of PTPRM dephosphorylated induce a stable intercellular cell to cell adhesion (Brady-Kalnay *et al.*, 1998). The dys-regulation of this pathway explained the detachment of cancer cells for invasion *via* the loss of tight adherens junction (cell to cell adhesion), which is resulted from the defect in binding of the catenin/cadherin complex with actin microfilament network of the cellular cytoskeleton (Frixen *et al.*, 1991; Thiery, 2002). This supports the suspected behaviour of the loco-regional failure in cheek SCC.

Study Limitation:

Even though a higher resolution array CGH was employed in this study, it still suffers from several limitations despite being a powerful technology to detect genomic abnormalities in terms of copy number variation. Another limitation found in the present study was regarding the small sample size which consisted of only 20 OSCC samples. In fact the sample size considerably influenced the biological conclusion derived using array CGH. The frequency of copy number variation in 20 OSCC patients will be differed with hierarchical clustering and may affect the overall estimation of frequency of variation. Furthermore, this study did not analyze the copy number aberrations parallel to the gene expression level. This can be achieved by integrated the copy number aberrations with gene expression data that derived from a public database to identify specific target genes that altered the gene dosage caused by the chromosomal aberrations. Apart from that, one of the major difficulty using array CGH is the involvement of high cost, which could not be afforded by the public and thus making it a difficult tool for diagnostic.

6.0 Conclusion and recommendations

The primary objective of the current research was to determine the chromosomal aberrations in tongue and cheek SCC using array CGH technology. The results from the present study to determine chromosomal aberrations in tongue and cheek SCC using array CGH technology has revealed that there was difference in set of chromosomal aberrations from both subsites.

The second objective was to identify the genes involved in tongue and cheek SCC using array CGH technology. For this objective, this study has managed to compile a list of amplified and deleted genes detected from tongue and cheek SCC using array CGH.

The final objective was to determine the significant pathways involved in tongue and cheek SCC using pathway analysis software. The present study has showed that there were different biological pathways involved in tongue and cheek SCC. The results revealed that, the most significant pathway involved in tongue SCC was cell adhesion ECM remodeling pathway. This finding has further strengthened the explanation for the reason behind the more aggressive behaviors in tongue SCC than others subsites of oral cancer. Although the feature of the Metacore software is often debatable, however, it is recommended that in order to validate the said pathways, more cellular biology experiments have to be conducted.

Overall, this study has showed there are differences in chromosomal aberrations, gene and its pathways tongue and cheek SCC. The observations from the present research has opened up a stepping stone for the future researchers to explore the chromosomal aberrations, genes, their interactive pathways and role on oral carcinogenesis. Nevertheless, this study has managed to utilize high resolution array CGH to perform

genome wide copy number profiling to identify copy number alterations that are associated with oral cancer subtype in frozen tumor tissues. This has provided useful information on the molecular biology of genetic alterations in oral carcinogenesis which could be considered in the management of oral cancer. Integrative analyses of different data types and advanced molecular profiling have enabled us to relate these alterations to the activities of oncogenic pathways and subsequently this could be made useful for the identification of suitable therapeutic opportunities.

Recommendation:

In the future, the findings from the high-resolution array CGH can be validated using larger number of samples on customize lower resolution array CGH which comprising of the targeted CNAs that are specific for oral cancer based on the ultra dense array CGH platform. This is proposed due to fact that by customizing the array CGH with targeted CNAs that specific for oral cancer, this would reduce the costing for the array CGH. Through this attempt large numbers of samples are could be involved and validate using the customized array CGH to validate the specific CNAs for oral cancer. Moreover, the significant association with socio-demographic profile or clinicopathological parameters could be also attempted. It could also improve the probability of detecting recurrent segmental copy number alteration and other mechanisms such as epigenetic changes and mutations at the sequence level, which can contribute to specific genes deregulation. The present study has identified the genes that are residing in recurrent chromosomal regions of tongue and cheek SCC. Therefore, in order to obtain a more complete picture of cancer, future efforts can be made to combine the array CGH approaches with transcriptomic and proteomic technologies. This can be done by integrating the copy number data with public expression dataset to identify the potential driver genes.

Oral cancer has been documented as a heterogenous tumor group, due to their occurrence at various sites within the oral cavity and their different biological behaviours. In view of this, the genomic alterations of others subsites within the oral cavity such as gum, floor of mouth and hard palate is further recommended for future study and thus it would have been possible to determine their genetic alterations of the different subsites.

Although array CGH has been offered as the first tier diagnostic tool, the economics is not favourable to public at the moment. As this study have detected potential unique genes for tongue and cheek SCC out of 24,500 genes in the human genome sequence project, the implicated genes should be further studied. Moreover, their role in oral carcinogenesis should be also uncovered in detail. This could be carried out using the low costing quantitative PCR technique to determine their gene expression level. These results will certainly add to the present knowledge of oral carcinogenesis and might be useful for future application in the prognosis and therapy of OSCC.