DETECTION OF HOST-SPECIFIC IMMUNOGENIC PROTEINS IN THE SERA OF ORAL SQUAMOUS CELL CARCINOMA (OSCC) PATIENTS

by

SITI NURAISHAH BT AZMAN

Thesis is submitted in fulfillment of the
requirements for the Degree of
Masters of Science

May 2015

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and Most Merciful. He is worthy of all the praises with the blessings and grace that He had bestowed upon me. Without Him, I could not possibly have completed my studies and this thesis.

My deepest appreciation goes to Assoc. Prof. Dr Chen Yeng for her guidance, advices and all the knowledge she shared which had helped both in research and thesis writing. Without her kind supervision and patience, I would not have made it through.

My sincere thanks go to Prof. Rosnah Md Zain and the Oral Cancer Research and Coordinating Centre (OCRCC) for all the knowledge shared and providing me the serum samples. I would also like to thank all the INFORMM lab staffs that had helped me while doing my bench works especially Puan Noorizan, Puan Nurfatihah, Puan Maimunah, Puan Norsyazwani, Puan Norazimah, Puan Norhalida and Puan Sabariah. My gratitude also goes to all INFORMM administration staffs especially Cik Fauziah, Puan Nurul, Cik Zira, Puan Siti, En Irwan, and En Adli.

To my lab-mates; Shue, Tan, Chan and Lydia, thank you for being supportive and all the helps that I received from you, it will forever be remembered. To Mun, Dibah, Peeps, Ju, Bid, Farah, Ozah, kak Madihah, kak Anizah, Farhanah, Thanes, Kalpu, kak Syida, Darul and my many great course mates, I am glad our paths crossed and greatly indebted for all the fun, tears and memorable experience we shared together all these years.

Last and most importantly, I want to thank my parents, who have always been my pillars of strength, my source of motivation and inspiration, for their unwavering and unconditional love and support that had made me to what I am today. Not to forget, to all my family members who have been caring and supportive throughout.

Funding for this study was obtained from USM Research University grant No: 1001/CINFORMM/811099 and UM HIR grants (No.H18001-00-C00020 and No.H18001-00-C00009). The author received financial support from USM Fellowship Scheme.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS……	**ii**
TABLE OF CONTENTS……	**iv**
LIST OF TABLES……...…	**viii**
LIST OF FIGURES……	**ix**
LIST OF ABBREVIATIONS……..…	**x**
ABSTRAK……	**xi**
ABSTRACT……..	**xiii**

CHAPTER 1: INTRODUCTION

1.1. Oral cancer……..	1
1.2. Epidemiology……..	2
1.3. Etiology……	3
1.3.1. Betel quid……..…	4
1.3.2. Areca nut……..….	4
1.3.3. Slaked lime……...	5
1.3.4. Smokeless or chewing tobacco…….....	6
1.3.5. Tobacco smoking……..…	6
1.3.6. Alcohol…….	7
1.3.7. Diet and nutrition……..…	9
1.4. Microorganisms and cancer……	10
1.5. Proteomics……..	12
1.6. Cancer serology using proteomics approach……..	14

1.7. IgM antibody……..…… 16

1.8. Functional annotation and protein interaction analyses…….. 18

1.9. Statement of the problem and rationale of the study…….. 20

2.0. Objectives of the study……... 22

CHAPTER 2: MATERIALS AND METHODOLOGY

2.1. Collection of serum samples……..…… 24

2.2. Two-dimensional polyacrylamide gel electrophoresis (2-DE)…….…..... 25

2.2.1. First dimension: Isoelectric focusing (IEF)……..….. 25

2.2.1.1. Materials / standard solutions……..…….. 25

2.2.1.2. Rehydration of Immobiline™ DryStrip……. 26

2.2.1.3. Performing IEF……..…….…... 27

2.2.2. Second dimensional electrophoresis……...…… 29

2.2.2.1. Materials/ standard solutions……..……...…… 29

2.2.2.2. Casting of gradient SDS-PAGE gel……...…… 31

2.2.2.3. IPG strip equilibration……...…… 32

2.2.2.4. Running second dimension gel……..……... 32

2.2.3. Visualization of results...…….…….…….. 34

2.2.3.1. Silver staining…… 34

2.2.3.1.1. Materials/ standard solutions...……..……...…… 34

2.2.4. Evaluation of results...……..…….…… 31

2.2.4.1. Image analysis……..……...…… 31

2.2.4.2. Statistical analysis…….……... 38

2.3. Protein identification using 2-DE immunoblotting method…….….. 38

2.3.1. Materials/ standard solutions……..…….. 38

2.3.2. Transfer of proteins onto nitrocellulose membrane……..….. 40

2.3.3. Immunoprobing using primary and secondary antibodies…… 41

2.3.4. Visualization of immunoblotting results…….……..... 42

2.4. Protein identification using mass spectrometry (MS)……...… 44

2.4.1. MS compatible silver staining…….…….…… 44

2.4.1.1. Materials/ standard solutions…….…. 44

2.4.2. Protein extraction and in-gel digestion…….… 47

2.4.2.1. Pre-treatment of cored out gel pieces prior to in-gel digestion... 47

2.4.2.1.1. Destaining of spots for MS compatible silver stained gel. 47

2.4.2.1.1.1. Materials/ standard solutions……. 47

2.4.3. In-gel tryptic digestion of proteins…….……... 48

2.4.3.1. Materials/ standard solutions…….……... 49

2.4.4. Peptide purification using C18 ZipTips…….……... 50

2.4.4.1. Materials/ standard solutions…….…..….. 51

2.4.5. MALDI-TOF/TOF Tryptic peptide analysis……..…… 52

2.5. Funtional annotation and protein interaction analyses…… 53

CHAPTER 3: RESULTS AND ANALYSIS

3.1. Typical 2-DE serum protein profiles of normal controls

and OSCC patients…… 54

3.2. Image analysis of 2-DE gels…….…... 57

3.3. Identification of immunogenic protein using 2-DE Western blotting…… 57

3.4. Identification of proteins using mass spectrometry…… 61

CHAPTER 4: DISCUSSION AND CONCLUSION…… 79

REFERENCES…….……... 95

APPENDICES…….……. 119

PUBLICATION…….…….. 163

PAPER PRESENTATION……..…… 166

LIST OF TABLES

		Page
Table 1.1	Cancer serology studies using proteomics analysis approach.	15
Table 2.1	Immobiline DryStrip IEF conditions for Ettan IPGphor 3 system.	28
Table 2.2	Preparation of gradient gel (from 8% - 18% gradient SDS-PAGE) solutions.	33
Table 2.3	Electrophoresis conditions for the second dimension gel electrophoresis.	33
Table 2.4	Silver staining methods.	36
Table 2.5	MS compatible silver staining methods.	46
Table 2.6	Protein input of proteins of interest submitted to DAVID v6.7	53
Table 3.1	The relative expression of host specific proteins among the sera of patients.	62
Table 3.2	Immunogenic proteins recognized by IgM shown on the 2-DE immunoblotted nitrocellulose membrane.	63
Table 3.3	Mass spectrometric identification of host specific protein spot from serum protein profiles using Mascot search engine and NCBI database.	65
Table 3.4	Mass spectrometric identification of non-host specific protein spot from serum protein profiles using Mascot search engine and NCBI database.	67
Table 3.5	Functional annotation analysis of identified host-specific proteins using DAVID v6.7.	77

LIST OF FIGURES

		Page
Figure 1.1	Schematic depiction of IgM pentavalent structure.	17
Figure 1.2	Summary on the whole of methodology performed in this study.	23
Figure 2.1	Arrangement of transfer sandwich for electrophoretic transfer using Transblot® SD Semidry Transfer Cell (Bio-Rad, USA) semidry system.	43
Figure 3.1	Representative 2-DE serum protein profiles (pH 4-7) of unfractionated sera of normal controls.	55
Figure 3.2	Representative 2-DE serum protein profiles (pH 4-7) of unfractionated sera of OSCC patients.	56
Figure 3.3	2-DE Western blot of (a) category 1, (b) category 2, (c) category 3 and (d) category 4.	59
Figure 3.4	Host specific proteins MALDI-TOF/TOF mass spectra of tryptic peptides derived from 2-DE protein spots.	69
Figure 3.5	Non-host specific proteins MALDI-TOF/TOF mass spectra of tryptic peptides derived from 2-DE protein spots.	74
Figure 3.6	Interaction networks of identified host-specific proteins using STRING v9.1.	78

ABBREVIATIONS

OSCC	Oral Squamous Cell Carcinoma
SDS	Sodium Dodecyl Sulphate
PAGE	Polyacrylamide Gel Electrophoresis
IEF	Isoelectric Focusing
Ig	Immunoglobulin
NCP	Nitrocellulose Membrane
ECL	Enhanced Chemiluminescence
HRP	Horseradish Peroxidase
MW	Molecular Weight
APS	Ammonium Persulfate
TEMED	N, N, N', N'-tetramethylethane-1, 2-diamine
M	Molar
mM	milimolar
L	Liter
ml	milliliter
g	gram
mg	milligram
µl	microliter
rpm	Revolutions per minute
rcf	Relative centrifugal force

PENGENALPASTIAN PROTEIN PERUMAH BERIMUNOGENIK DI DALAM SERA PESAKIT SEL KARSINOMA SKUAMA ORAL (OSCC)

ABSTRAK

Kanser mulut adalah kanser ke-enam paling lazim di dunia dan sel karsinoma skuama oral (OSCC) yang berasal dari mukosa mulut membentuk lebih daripada 90% kemalignan dan menjana hampir 300,000 kes baru setiap tahun. Cabaran utama untuk meningkatkan kadar kemandirian pesakit OSCC adalah disebabkan kanser mulut sering tidak dikesan sehingga peringkat lewat dan hanya separuh daripada mereka yang disahkan menghidap penyakit tersebut masih hidup selepas lima tahun manakala pesakit peringkat awal mempunyai kadar kemandirian 5 tahun yang sangat baik iaitu lebih daripada 80%. Oleh itu, pembangunan penanda biologi untuk OSCC yang sesuai dan tepat amat diperlukan untuk membantu dalam pengesanan awal sekaligus boleh mengurangkan kematian berkaitan dengan kanser mulut. Dalam kajian ini, elektroforesis dua dimensi (2-DE) dan teknik pemedapan-immuno telah digunakan untuk mengenal pasti penanda-edaran dalam sera pesakit OSCC. Profil protein yang diperolehi, dibanding dan dianalisa menggunakan perisian PDQuest dan seterusnya dikenal pasti dengan teknik pemedapan Barat menggunakan serum terkumpul (antibodi primer) dan monoklonal anti-manusia IgM-HRP (antibodi sekunder). Protein yang dikehendaki untuk dikaji kemudiannya dikenal pasti dan dicirikan dengan menggunakan spektrometri jisim MALDI-TOF/TOF diikuti oleh maskot protein carian pangkalan data. Perbezaan dalam pengekspresan protein α1-B-Glikoprotein (ABG), haptoglobin (HAP), klusterin (CLU), leusin yang kaya dengan α2-Glikoprotein (LRG), retinol mengikat protein 4 prekursor (RBP4), PRO2044, dan

proapolipoprotein (Proapo-AI) diperhatikan dalam serum pesakit OSCC apabila dibandingkan dengan serum individu yang sihat. Daripada protein-protein tersebut, hanya empat protein perumah iaitu CLU, HAP, Proapo-AI dan RBP4 mempamerkan keimunogenikkan dalam pemedapan Barat manakala lima protein bukan perumah yang dihipotesiskan sebagai protein *Acinetobacter lwoffii, Burkholderia multivorans, Myxococcus xanthus, Laribacter hongkongensis* dan hemolysin A *Streptococcus salivarius* turut diperhatikan. Kajian ini boleh dianggap sebagai batu loncatan dalam pembangunan penanda protein yang berpotensi untuk memudahkan pengenalan OSCC di peringkat awal. Tambahan pula, pendekatan menganalisa hubungan OSCC dengan bakteria telah menunjukkan beberapa hasil yang menjanjikan dan bermaklumat.

DETECTION OF HOST-SPECIFIC IMMUNOGENIC PROTEINS IN THE SERA OF ORAL SQUAMOUS CELL CARCINOMA (OSCC) PATIENTS

ABSTRACT

Oral cancer is the sixth most prevalent cancer in the world and the predominant type, oral squamous cell carcinomas (OSCCs) deriving from the oral mucosa constitute more than 90% of malignancies and generating nearly 300,000 new cases annually. The main challenge of improving the survival rate of OSCC patients is that oral cancer is often not detected until the later stages resulting in only half of those diagnosed with the disease surviving more than five years. The early-stage patients have an excellent 5-year survival rate of more than 80% unlike the late stages patients Therefore, the development of suitable and reliable biomarkers of OSCC is greatly needed to assist in early detection which may reduce the mortality associated with oral cancer. In this study, two dimensional gel electrophoresis (2-DE) and immunoblotting techniques were employed to identify novel circulating markers in sera from OSCC patients. The protein profiles obtained were compared and analyzed using PDQuest software and were further identified with immunoblotting using pooled sera (primary antibody) and monoclonal anti-human IgM-HRP (secondary antibody). The spots of interest were identified using MALDI-TOF/TOF mass spectrometry followed by MASCOT protein database search. Differences in the expressions of alpha-1-B-glycoprotein (ABG), haptoglobin (HAP), clusterin (CLU), leucine-rich α2-glycoprotein (LRG), retinol binding protein 4 precursors (RBP4), PRO2044, and proapolipoprotein (Proapo-AI) were observed in OSCC patients' sera when compared with normal controls. Of those proteins, only four host specific

proteins namely CLU, HAP, Proapo-AI and RBP4 exhibited immunogenicity in Western blots whilst five non-host hypothetical proteins of *Acinetobacter lwoffii, Burkholderia multivorans, Myxococcus xanthus, Laribacter hongkongensis* and hemolysin A of *Streptococcus salivarius* were also observed. This study can be perceived as a stepping stone in the development of potential protein markers that facilitate the identification of OSCC at an early stage. In addition, the approach of analyzing the relationship of OSCC with bacteria has shown some very promising result which is significantly informative.

CHAPTER ONE

INTRODUCTION

1.1. Oral Cancer

Oral cancer can be subdivided into three categories namely carcinoma of the oral cavity proper, carcinomas of the lip vermilion and carcinomas arising in the oropharynx (Neville and Day, 2002; Llewellyn *et al*., 2001). Malignancies arising in the lip, teeth, gum, tongue, floor of the mouth, gingivae, palate, buccal mucosa and salivary glands are categorized as cancers of the oral cavity (Silverman, 2001). Although there are several types of malignant oral cancers, more than 90% of these malignant neoplasms are squamous cell carcinomas (SCCs) developing in the mucous membranes of the mouth and oropharynx (Chen and Myers, 2001). Oral cancer is an epithelial neoplasia and it is believed that OSCC follows a comparable pattern in its development. Epithelial neoplasia generally begins as a focal clonal overgrowth of altered stem cells near the basement membrane, expanding laterally and upward, replacing the normal epithelium and progressing through hyperplasia to dysplasia to carcinoma in situ and invasive carcinoma (Turhani *et al*., 2006). Oral cancer is atypical in that it carries a high risk of second primary tumors. Survivors of the first oral cancer have up to 20-fold increased risk of developing a second primary oral cancer. That risk lasts for five to ten years and sometimes longer (Mager *et al*., 2005).

A fundamental barrier to improving the survival rate in OSCC relates to the fact that it often remains undetected until later stages despite the accessibility of oral cavity to

direct examination. Oral cancer has been reported as having one of the highest mortality ratios amongst all malignancies by World Health Organization (Parkin *et al*., 2001). As a matter of fact, oral cancer has not seen a drop in its 50% mortality rate over the past 30 years (Kujan *et al*., 2005), and as a result only half of those diagnosed with the disease survive more than five years. In fact, it was reported that a few months delay in diagnosis can reduce a patient's chance of survival from 80% to 40% (Lin *et al*., 2006). High rates of second oral malignancies were also related to oral cancer in which up to 30% of patients suffered a recurrence of the tumor or development of a second primary tumor even with intensive follow-up (Dhooge *et al*., 1998; Poh *et al*., 2011).

1.2. Epidemiology

Albeit oral cancer being the eighth most common cause of cancer-related deaths globally, large numbers of people are still unaware of its existence (Parkin *et al*., 2005). In developing countries, mouth and pharynx is the sixth and eighth most common sites for malignant disease among men and women respectively (Johnson *et al*., 2011). High incidence rate of oral cancer as reported in South and South East Asia which accounts for more than 180,000 cases annually (Bofetta and Parkin, 1994). In Malaysia particularly, the prevalence of oral cancer was 0.04% as reported by Zain *et al*. (1997)

Predominantly, oral cancer occurs in males after the fifth decade of life. In Asian populations, the mean age of oral cancer patients is in the fifties and early sixties (Johnson *et al*., 2011). In recent years, incidence of oral cancers among younger people is on the rise. Oral cancers were increasingly reported to occur at ages

younger than forty five years especially in the high prevalence areas of the world where heavy uses of tobacco were reported (Llewellyn *et al*., 2004; Elango *et al*., 2006). Males had a higher mean global age-specific incidence of mouth and pharyngeal cancer when compared to females. However, the disparity of oral malignancies between males and females has become less prominent over the past fifty years (Neville and Day, 2002; Hooper *et al*., 2009).

Variations by ethnicity were illustrated globally. Asians are most likely to develop malignancies in the buccal mucosa, African Americans males in the USA have a high incidence of oral and pharyngeal cancer while lip cancer is increasing in Australia (Johnson et al., 2011). In Malaysia, Zain *et al*. reported in 1997 that other Bumiputera (indigenous people of Sabah and Sarawak) exhibited the highest percentage of subjects with oral mucosal lesions with 17.97%, followed by Indians, 14.81%, Chinese, 9.53% and Malays, 7.49%. Other Bumiputera (as mentioned above) also have the highest prevalence of oral carcinoma with 1.9/1000 population. In 2007, National Cancer Registry (NCR) stated that among 353 cases of oral cancer reported in Malaysia, 182 were females. Incidence of oral cancer was also reported as highest in Indian females where the ASR was 10.2/100000 female populations. Of those oral malignancies cases reported with staging, only 35% were diagnosed at stage I and II (NCR, 2007). Early diagnosis is imperative.

1.3. Etiology

Oral squamous cell carcinoma is commonly associated with tobacco in various forms, areca nut or betel quid chewing, heavy alcohol consumption and also microbial factors. The usage of tobacco and areca nut, either simultaneously or alone,

accounts for the majority of oral cancers and oral potentially malignant disorders (Amarasinghe *et al*., 2010). Areca nut which is an independent risk factor for oral cancer is a common component of different chewing habits and classified as carcinogenic to humans by the WHO (Warnakulasuriya *et al*., 2002). Poor oral hygiene due to increment of microbial load and chronic trauma from ill-fitting dentures or sharp restorations may also contribute as risk factors for oral cancer.

1.3.1. Betel Quid

Betel quid chewing habit has been mentioned as early as 600 A.D. in the Sanskrit and can be traced back locally in 1664, in which an impost duty on betel leaf imported from India into Malacca was mentioned in the Dutch archives (Muir and Kirk, 1960). In general, a betel quid consists of betel leaf, areca nut, slaked lime and also tobacco. Based on local preference, other substances such as cardamom, saffron, cloves, aniseed, turmeric, mustard or sweeteners were also included (Norton, 1998; Neville and Day, 2002). Betel quid with or without the addition of tobacco is considered as carcinogenic to humans hence provide a risk of oral cancer and potentially malignant disorders development (Warnakulasuriya *et al*., 2002; Bagan and Scully, 2008).

1.3.2. Areca Nut

The seed of the oriental palm *Areca catechu* is termed as areca nut. According to IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 85; areca nut consumption is common in India, Sri Lanka, Bangladesh, Taiwan and Melanesia. It is also popular in parts of South-East Asian countries such as Thailand, Indonesia, Cambodia, including Malaysia and in emigrant communities from these

countries. It is estimated about 20% of the world's population, approximately 600 million people worldwide used areca nut as a masticatory substance (Johnson *et al*., 2011).

Areca nut constitutes mainly of carbohydrates, fat, proteins, fiber, polyphenols (flavanols and tannins), alkaloids and minerals. In the mouth, nitrosation of the alkaloids in dried stored nuts produced nitrosamines and especially in the acid conditions of the stomach with the presence of nitric oxide generated by bacteria as reported in IARC Monographs (Volume 85). Subjects with poor oral hygiene significantly have higher endogenous nitrosation and more extensive nitrosamine formation take place if they also chew tobacco. According to IARC Monographs (Volume 82), fungi such as *Aspergillus flavus*, *Aspergillus niger* and *Rhizopus* spp. may contaminate the areca nut. Aflatoxins produced by the contaminants, which are established carcinogens were found in almost 40% of areca nut samples analyzed using thin-layer chromatography.

1.3.3. Slaked Lime

Slaked lime (calcium hydroxide) is commonly obtained by heating sea shells or harvested from corals in coastal areas of Sri Lanka and islands of the pacific whereas in inland areas, it is quarried from limestone. Erosion of the oral mucous membranes can developed when slaked lime is added into betel quids hence facilitate penetration of betel quid carcinogens through the mucosa (Johnson *et al*., 2011).

1.3.4. Smokeless or Chewing Tobacco

Tobacco, originated from cut leaves of *Nicotiana tabacum* and *Nicotiana rustica* are frequently added to the betel quid. Not less than sixteen carcinogens including tobacco-specific nitrosamines and polycyclic aromatic hydrocarbons are exposed to the oral mucosa when chewing tobacco takes place (Petti, 2009). Carcinogenic tobacco-specific nitrosamines; N-nitrosonornicotine (NNN) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) were highly expressed in saliva of tobacco chewers and oral snuff users (Idris *et al*., 1992; Nilsson, 2011). Abundance of reactive oxygen species are released when chewing betel quid and especially whilst the quid is still present. As major genotoxic agents, both tobacco-specific nitrosamines and reactive oxygen species involved in oral cancer are associated with the use of chewing tobacco as reported by IARC Monographs (Volume 85). Smokeless tobaccos mostly have high levels of nicotine, addictive and can also be initiators of smoking (Haddock *et al*., 2001). In addition, they have significant cardiovascular effects and produce oral mucosal lesions as well as local damage to the periodontium (Critchley and Unal, 2003).

1.3.5. Tobacco Smoking

As the foremost preventable cause of premature death worldwide, tobacco is a major etiological risk factor for the development of oral and pharyngeal cancer and other malignancies of upper aero-digestive tract (Johnson *et al*., 2011). It is estimated that 10 million deaths per year will occur by 2020, of which 70% will be in developing countries due to tobacco-related illness (Warnakulasuriya *et al*., 2005). Risk of developing oral cancer is five to nine times greater for smokers compared to non-

smokers. Approximately 80% of oral cancer patients smoked, that is two to three times greater compared to the general population (Neville and Day, 2002).

Tobacco contains as much as 6000 different compounds in the particulate and vapour phases. In addition to nicotine, these compounds comprise polycyclic aromatic hydrocarbons, tobacco glycoprotein and some metals; many of which are known to be antigenic, cytotoxic, mutagenic or carcinogenic (Arnson *et al*., 2010). Carcinogenic NNK, NNN and polycyclic aromatic hydrocarbons particularly have been linked to upper aero-digestive tract cancer. Smoking and exposure to nicotine influences the immune system in many ways. Immunosuppression effects involving reductions of antigen presenting activity, circulating immunoglobulins, T-cell activity and neutrophil activity were observed in smokers. Pro-inflammatory effects were also demonstrated whereby augmentation of acute phase and pro-inflammatory reactants, neutrophil count, circulatory T-cells and free radical burden take place (Arnson *et al*., 2010).

1.3.6. Alcohol

Alcohol consumption has been identified as a major public health problem worldwide as well as major risk factor for cancers of the upper aerodigestive tract. Carcinogenic effect of alcohol was first reported in 1960's and has been replicated since (Ketcham *et al*., 1963; Bofetta and Hashibe, 2006).Cook reported in 1998 that immune surveillance can be reduced with alcohol drinking thus favouring cancer development and metastatic potential (Cook, 1998). Pure ethanol does not show carcinogenicity in animal studies (Boyle *et al*., 2003), but other components of alcoholic drinks including impurities and contaminants might increase cancer risk. *N*-

nitrosodiethylamine found present in some beer and whisky is associated with an increased risk of oral cancer as well as polycyclic aromatic hydrocarbons which some are considered to be carcinogenic (Ogden and Wight, 1998).

Alcohol dehydrogenase and aldehyde dehydrogenase are the major alcohol-metabolizing enzymes where they oxidize ethanol to acetaldehyde and detoxify acetaldehyde to acetate respectively. Due to its multiple mutagenic effects on DNA, acetaldehyde is responsible for the oral carcinogenic effect of ethanol (Johnson *et al*., 2011). Alcohol can promote oncogenesis through impairment of phospholipids of cell membranes by ethanol therefore increases permeability and enhancing the penetration of carcinogens across the oral mucosa (Howie *et al*., 2001). Ethanol also acts as a solvent and impairs DNA repair mechanisms thus allowing penetration of carcinogens through the mucosa, possibly catalyzing the activation of carcinogenesis (Bofetta and Hashibe, 2006). Ethanol is also hepatotoxic and caused the effectiveness of enzyme systems central to the detoxification of carcinogens such as glutathione-S-transferases to reduce. Super-multiplicative synergistic effects of alcohol and smoking were extensively reported since 1970's wherein some subsets of patients who are extremely heavy smokers and heavy drinkers can have 100 times greater risk for developing a malignancy (Rothman and Keller, 1972; Zeka *et al*., 2003; Salaspuro and Salaspuro, 2004; Bofetta and Hashibe, 2006).

1.3.7. Diet and Nutrition

Dietary factors are estimated to account for approximately 20% of all cancers in the developing countries (Cancer: Diet and Physical Activity's Impact, WHO). There have been extensive studies in associations with diets and oral cancer and precancer in the last three decade. Insufficient intake of fresh fruits and vegetables has been consistently linked with an increase in cancer prevalence or incidence (Levi *et al*., 1998; Lucenteforte *et al*., 2009). Through a meta-analysis of existing epidemiologic studies, Pavia *et al*. (2006) observed an overall 49% and 50% reduction on oral cancer risk for each portion of fruit and vegetables consumed per day respectively (Pavia *et al*., 2006).

Vitamin A, C and E and related carotenoids (in particular beta-carotene) and selenium appear to decrease incidence of epithelial cancer. The concentrations of these nutrients are found to be relatively high in the serum although the adrenal and pituitary glands, the brain, white blood cells and platelets showed higher concentrations compared to serum (Machlin and Bendich, 1987). The protective effect of these micronutrients attributed to their antioxidant activities that act by reducing free radical reactions that can cause DNA mutations and changes in lipid peroxidation of cellular membranes and changes in enzymic activities (Machlin and Bendich, 1987; Schwartz and Shklar, 1988; Du *et al*., 2002). Micronutrients are also responsible in modulation of carcinogen metabolism, maintenance of immune function, inhibition of endogenous carcinogens formation and influence cell transformation and differentiation (Zain, 2001). Serum vitamins A, B_{12}, C, E, beta-carotene and foliate were shown by Ramaswamy and his colleagues, to be decreased in leukoplakia patients compared to controls (Ramaswamy *et al.*, 1996). In Japan,

another study also reported a significant difference in serum levels of lycopene and beta-carotene in those with leukoplakia compared to control (Nagao *et al*., 2000).

1.4. Microorganisms and Cancer

Oral cancer is considered to be a multi-factorial disease, as it can stem from exposure to several types of carcinogens, including microbial factors (Kalu U. E. Ogbureke, 2012). The existence of relationships among certain bacteria and cancers continue to be promulgated albeit bacterial mechanisms involved are at present unclear (Mager, 2006). It is estimated around 15% of malignancies (about 1.5 million cases) worldwide can be attributed to viral, bacterial and other pathogens annually (Srivastava *et al*., 2005). For most associations between infection and malignancies, it is still unknown whether the bacterial infection is a marker of disease or is causally related to tumour formation except for *Bartonella* (Dehio, 2005) and *Helicobacter pylori* (Rogers *et al*., 2005) infections which have been shown to induce tumour formation.

Several pieces of evidence have supported the association of microbial infection with oncogenesis. For instance, association of *H. pylori* with two different forms of gastric cancer; mucosa associated lymphoid tissue (MALT) lymphoma and adenocarcinoma is the utmost-studied relationship between a bacterial infection and cancer to date and are also categorized by the IARC (WHO) as a carcinogenic factor in humans (Peek and Blaser, 2002; Björkholm *et al*., 2003; Marshall and Windsor, 2005; Correa and Houghton, 2007). Similarly, *Chlamydophila pneumonia* has been associated with malignant lymphoma and lung cancer in males (Anttila *et al*., 1998; Kocazeybek, 2003), whereas *Candida albicans* and *Streptococcus anginosus* were linked to oral

carcinomas (Sasaki *et al*., 2005; Hooper *et al*., 2009). It is also known that OSCC patients tend to possess significantly elevated concentrations of certain bacteria in their saliva. Thus, changes in salivary microflora may represent a non-invasive diagnostic tool for predicting oral cancer (Hooper *et al*., 2009).

Hence, to understand the pathogenesis and prevention of certain cancers, it may perhaps be achieved by studying the bacterial infections associations and their effects on the host (Vogelmann and Amieva, 2007). Nevertheless, the mechanisms by which cancer formation attributed by bacteria are complex and recent investigations showed involvements of deleterious alterations in physiological host processes such as inflammation, antigen-driven lymphoproliferation and induction of hormones that increases epithelial cell proliferation. Bacteria may also directly affect oncogenesis through production of toxic and carcinogenic metabolites (Chang and Parsonnet, 2010). Bacteria, on their own are often insufficient to induce cancer but accompanied by chronic inflammation and independent mutations in oncogenic signaling pathways, the tumour formation process might occur (Vogelmann and Amieva, 2007). There is evidence that epidemiological and etiological links between microbial infection and oral cancer could exist. The activation of pro-carcinogenic substances by the oral microflora, specifically the conversion of ethanol to acetaldehyde, may be an important etiological factor (Hooper *et al*., 2009).

The oral cavity is home to a diverse microflora comprising many microbial species, each present in varying amounts. The composition and quantity of this microflora differ from each individual and can change throughout the lifetime in response to a variety of factors (Marsh and Percival, 2006). There have been only a few

investigations into the possible associations between bacterial species and oral carcinoma thus far. The presence of *Candida albicans*, a common oral commensal, has long been recognized as an independent risk factor for malignant transformation (Cawson, 1969). An intraoral carcinomas study done by Nagy *et al*., (1998) demonstrated an increased numbers of certain members of oral bacteria on the surface of tumours in comparison with control sites. Other studies have also demonstrated elevation of some common microflora on or in the lesions and their associated lymph nodes (Mager *et al*., 2005; Hooper *et al*., 2007; Westphal *et al*., 2008) It was presumed that such selectivity occurred because the bacteria were effectively shielded from the host immune system while being within the solid tumour (Yong *et al*., 2004). Previous research as well has shown that oral bacteria exhibit specific tropisms toward different biological surfaces such as the teeth and mucosa (Gibbons, 1996). This pattern of specificity suggests that different intra-oral surfaces and bacterial species have different receptors and adhesion molecules that influence the colonization of different oral surfaces (Mager *et al*., 2005).

1.5. Proteomics

As the final products manufactured in living cells, protein is important in the annotation of the genome. The word ‘proteome’ was first describes as a set of all proteins expressed by a given genome in which complementary to studies at the transcript level (Wasinger *et al*., 1995). In emphasizing its dynamic nature, a more specific definition of proteome was describes as a set of proteins that provides information on proteins that are expressed in a biological compartment at a particular time, under a particular set of conditions, as its composition may vary from tissue to tissue or even from cell to cell (Beranova-Giorgianni, 2003). In principal, proteome

offers richer source for the functional description of diseases. Specific properties of proteins such as post-translational modifications, various conformation states and alternative splicing illustrate the multidimensionality, high variability and dynamic nature of the proteomic information (Tambor *et al*., 2010). This may be of particular importance for diseases such as cancer, which evolve dynamically and affect many heterogeneous cell populations either as part of the cancer or as part of the host's reaction to the tumour (Kolch *et al*., 2005). Since the molecular pathways in normal and transformed cells were influenced by protein molecules, proteomic markers are more relevant to the disease state initiation and progression (Mishra and Verma, 2010). Proteomics cover of a broad scope whereby it encompasses identification and quantification of proteins in cells, tissues and biological fluids; analysis of changes in protein expression in normal versus diseases cells; characterization of post-transcriptional modifications; studies of protein-protein interactions and other applications.

Two dimensional gel electrophoresis (2-DE), mass spectrometry (MS) and bioinformatics tools are the key components of an approach that has been termed 'the classical proteomics methodology'. 2-DE was first introduced in the early 1970s by O'Farrell (1975). In 2001, serological proteome analysis (SERPA) has been proposed for a top-down approach usable for discovery-driven immunomics by Klade and his colleagues (Klade *et al*., 2001). SERPA or also termed as PROTEOMEX by Seliger and Kellner, (2002) is based on a classical proteomics workflow associating an effective separation on 2-DE gels and an identification by MS. SERPA allows the transfer and immobilization of proteins from tumour tissue or tumour cell lines to a semirigid support with the combination of Western blotting and

2-DE gels. Sera from cancer patients or healthy subjects were screened; allowing immunodetection of relevant antigens among the several thousand individual proteins separated using 2-DE. Comparative probing of blots with sera from patients and healthy subjects may allow the identification of associated antigens that elicit a humoral immune response with the sera from cancer patients specifically (Caron *et al*., 2007).

1.6. Cancer Serology Using Proteomics Approach

Cancer may be accompanied by the production and release of a substantial number of proteins or hormones into the blood that could serve as useful markers for assessing prognosis, monitoring treatment and detecting malignant disease at an early stage. Therefore, utilizing serum profiling in diagnosing cancer is certainly an appealing concept. In addition to protecting us against pathogens, the immune system is also on guard against other threats, including tumours (Tan, 2001). Generation of circulating antibodies that bind to self-protein can be deemed as the systemic amplification by the immune system of a signal that indicates presence of the tumor (Purcell and Gorman, 2004).

Serum is a commonly used matrix to screen for biological markers, which subsequently has been proven to be very useful in the diagnosis, prognosis, treatment and/or early detection of disease. Serum is derived from coagulated blood in which fibrin clots formed, blood cells and related coagulation factors are separated from serum by applying centrifugal force. During this process, proteins are released by platelets into the serum. When average metabolite concentrations were compared between subjects with different phenotypes, serum demonstrated higher sensitivity

compared to plasma thus revealed more potential in biomarkers detection (Yu *et al*., 2011).

Differences in the expression of some serum proteins may be an early sign of an altered physiology that may be indicative of disease (Poon and Johnson, 2001).The study of proteomics had contributed to significant advances in understanding cancer. Numerous studies on protein expression in different types of cancers had been done and published. These findings have provided important information on functional cellular processes. A list of cancer biomarker studies using proteomics approach is tabulated in Table 1.1.

Table 1.1: Cancer serology studies using proteomics analysis approach.

Cancer type	**References**
Breast cancer	Doustjalali *et al*., 2004; Kyselova *et al*., 2008; Gromov *et al*., 2010
Colorectal cancer	Kim *et al*., 2009
Nasopharyngeal cancer	Doustjalali *et al*., 2006
Prostate	Li *et al*., 2004; Han *et al*., 2012
Lung cancer	Yang *et al*., 2007; Hongsachart *et al*., 2009
Liver cancer	Looi *et al*., 2008; Zinkin *et al*., 2008; Liu *et al*., 2011
Pancreatic cancer	Fiedler *et al*., 2009; Wingren *et al*., 2012
Head and neck cancer	Ralhan *et al*., 2011
Gastric cancer	Liu *et al*., 2012

1.7. IgM Antibody

The term antibodies were first used by Karl Landsteiner in 1900 which is a translation from a German word "Antikörper" (Vollmers and Brändlein, 2005). Immunoglobulin M (IgM) antibodies are present in the circulation of normal humans and other mammalian species as part of the innate immunity and were conserved in all animals (Parkin and Cohen, 2001; Marchalonis *et al*., 2002). Pentameric IgM molecules made about 30% of the blood-circulating immunoglobulins in human. IgM is initially secreted by B cells upon primary stimulation with antigen (Zouali, 2001; Tchoudakova *et al*., 2009) and participates in natural defenses against foreign pathogens as well as neoplastic cells and tumors (Brändlein *et al*., 2003).

In fact, autoantibodies against specific cancer antigens have been identified for several types of tumors, including colon, breast, lung, ovary, prostate, and head and neck. These antibodies have been found to recognize over expressed (e.g., Her2), mutated (e.g., p53), or tissue-restricted (e.g., cancer-testis antigens) proteins, which are produced by cancer cells and elicit immune responses (Lin *et al*., 2007). Therefore, detection of such antibodies in patient sera can be exploited as a means of cancer diagnosis. Indeed, the specificity and sensitivity of the antibody response to low antigen levels make it an ideal screening or diagnostic tool for early identification of cancer biomarkers in serum-based assays.

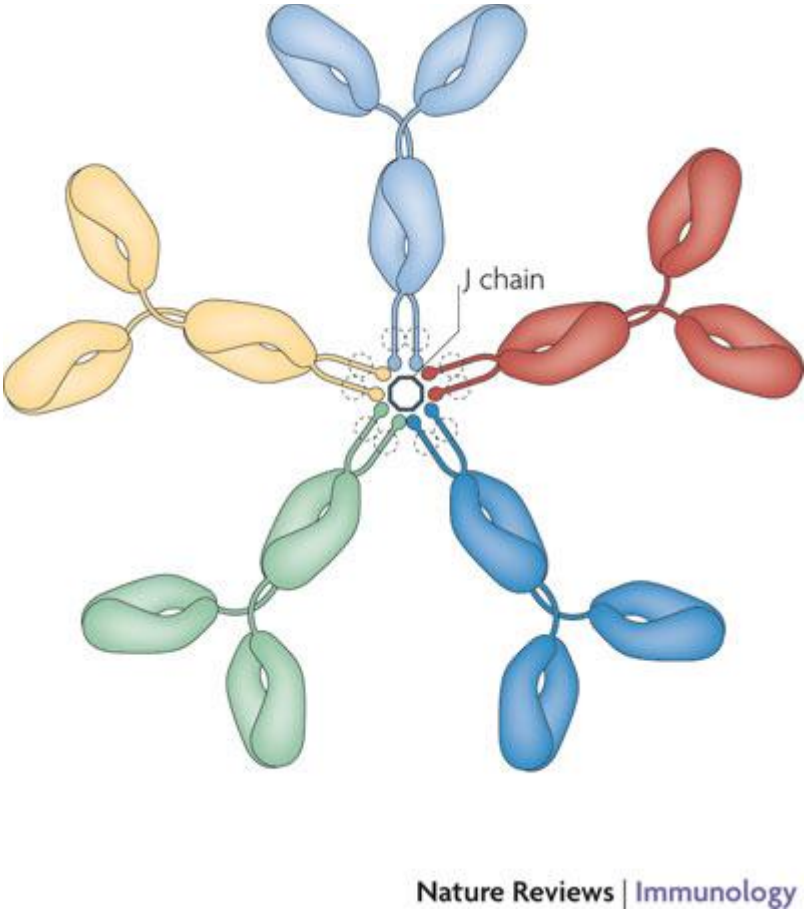

Figure 1.1: Schematic depiction of IgM pentavalent structure. The joining (J) chain promotes formation of the IgM pentamer. The central protruding region increase interactions with key ligands (Ehrenstein and Notley, 2010) (http://www.nature.com/nri/journal/v10/n11/abs/nri2849.html)

1.8. Functional Annotation and Protein Interaction Analyses

DAVID (Database for Annotation, Visualization and Integrated Discovery) bioinformatics resources are a high-throughput and integrated data-mining environment which able to systematically extracting biological features or meaning from large gene or protein lists (Huang *et al*., 2009). DAVID is uniquely characterized with integrated and expanded back-end annotation database, advanced modular enrichment algorithms and powerful exploratory ability (Sherman *et al*., 2007; Alvord *et al*., 2007)

Without ever making direct contact, proteins can catalyze subsequent reactions in a metabolic pathway, regulate each other transcriptionally or post-transcriptionally or jointly contribute to larger, structural assemblies. Furthermore, together with direct physical interactions, such indirect interactions constitute the larger superset of 'functional protein-protein associations' or 'functional protein linkages' (Eisenberg *et al*., 2000). Protein-protein associations have proven to be a useful concept by which to group and organize all protein-coding genes in a genome. The complete set of associations can be assembled into a large network, which captures the current knowledge on the functional modularity and interconnectivity in the cell. Protein network information can aid in the interpretation of functional genomics data.

The STRING (Search Tool for the Retrieval of Interacting Genes) database has been designed with the goal to assemble, evaluate and disseminate protein-protein association information comprehensively. The STRING v9.1 in particular, specializes in three ways. Firstly, it provides uniquely comprehensive coverage, with

more than one thousand organisms, five millions proteins and more than two hundred millions interactions stored. It is also one of very few sites to hold experimental, predicted and transferred interactions, together with interactions obtained through text mining. It also includes a vast of accessory information, such as protein domains and protein structures improving its day-to-day value for users.

1.9. Statement of the problem and rationale of the study

The advent of proteomic technologies have allowed the extensive fractionation of proteins in biological specimens, analysis of peptides by MS and matching of mass spectra to peptide sequences in human genome and protein databases (Omenn e*t al*., 2006). The dynamic nature of the circulatory system and its constituents reflects diverse physiological or pathological states, and the serum-based biomarkers provide noninvasive tests for screening as well as disease classification and monitoring for cancer progression and regression (Hanash *et al*., 2008). Theoretically, each disease may be uncovered and characterized by its unique biomarker. However, rather than to see this biomarker as a single molecule, a biomarker should be regarded as a panel of up- and down-regulated proteins or proteins with altered posttranslational modifications, which differ in disease and normal state (Etzioni *et al*., 2003; Rifai *et al*., 2006).

Even with advances in surgery, radiation and chemotherapy, oral cancer is one of the lowest five-year survival rate among major cancer sites and this rate has not significantly improved (Jou *et al*., 2010). In this present investigation, identification of novel OSCC biomarkers in order to promote the development of specific diagnostic tests for the early detection of oral cancer was aimed. The immunoproteomic approach employed in this study allowed the detection of immunogenic host-specific proteins in patient samples using pooled human antibodies. Thus, host- or tumor-specific proteins represented markers for the 'selected' antibody. Proteomic analysis allows the characterization and quantification of proteins and peptides in biological samples (Arnott and Emmert-Buck, 2010).

Therefore, a combined analytical platform involving two-dimensional gel electrophoresis (2-DE) and mass spectrometry (MS) has been widely used for biomarker identification (Srinivas *et al*., 2001; Kolch *et al*., 2005; Karpova *et al*., 2010).

In the present study, unfractionated sera from OSCC patients and normal controls were subjected to 2-DE and immunoblot for protein profiling, followed by characterization of potential markers using MS. Notably, since the approach used involved patient serum, which is complex and consists of various proteins, it can potentially reflect numerous events occurring *in vivo* simultaneously (Yeng *et al*., 2010). In addition to providing potential biomarkers for the early detection of oral cancers, this study contributes to the understanding of naturally occurring antibodies in cancer.

2.0. Objectives of the study

The aims of this study were as follows:

1) To develop the 2-DE serum protein profiles of patients with OSCC as well as that from normal individuals

2) To identify serum proteins that are differentially expressed in the serum of OSCC patients

3) To identify circulating immunogenic proteins and host specific proteins eliciting humoral responses in OSCC patients using immunoproteomics approach.

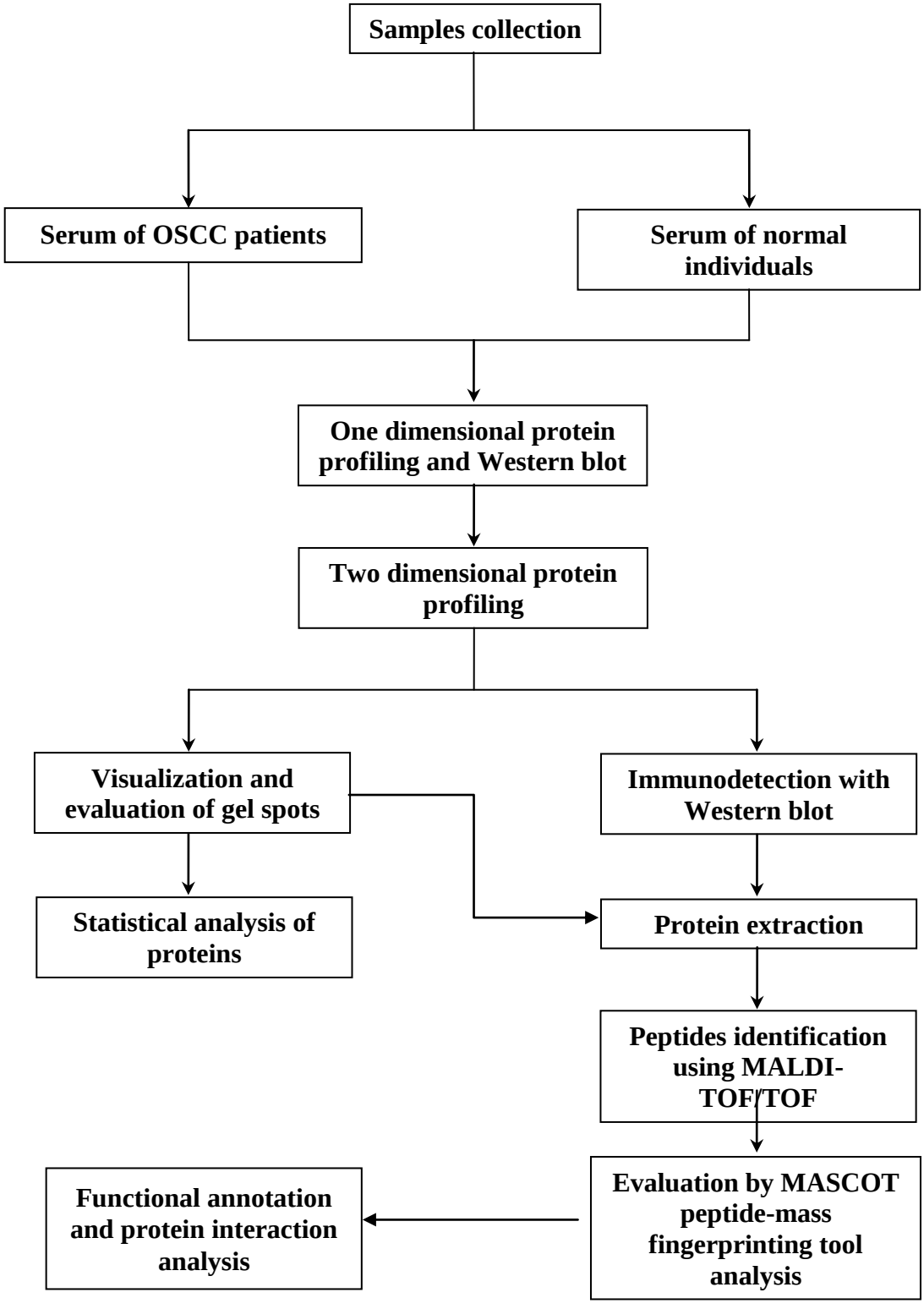

Figure 1.2: Summary on the whole of methodology performed in this study.

CHAPTER TWO

MATERIALS AND METHODOLOGY

2.1. Collection of Serum Samples

Based on clinical information and laboratory diagnosis for carcinogenesis, a total of 25 serum samples of oral squamous cell carcinoma (OSCC) patients were obtained from the Oral Cancer Research and Coordinating Centre (OCRCC), University of Malaya, Kuala Lumpur and used in this study, Additionally, 25 serum samples were also collected from healthy individuals and used as control group. The cancer group consisted of patients with early stages of oral squamous cell carcinoma and without any history with other type of malignancy. The normal healthy individuals referred to individuals that were determined not to have any history with cancer. Samples obtained were with consent and approval granted by both universities; the Medical Ethics Committee, Dental Faculty, University of Malaya (UM) (Ref: DF OP0907/0050(P)) and Universiti Sains Malaysia (USM) (Ref: USMKK/PPP/JEPeM [213.3.09]). The serum samples were stored in aliquots of 10 µl and kept at -80°C until use.