Mass Spectrometry-based Serum Peptidomic Patterns for Cervical Cancer Detection

Phetploy Rungkamoltip

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biomedical Sciences
Prince of Songkla University
2018

Mass Spectrometry-based Serum Peptidomic Patterns for Cervical Cancer Detection

Phetploy Rungkamoltip

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biomedical Sciences
Prince of Songkla University
2018

Thesis Title	Mass spectrometry-based serum peptidomic patterns for cervical cancer detection
Author	Miss Phetploy Rungkamoltip
Major Program	Biomedical Sciences

Major Advisor

.....
(Dr. Raphatphorn Navakanitworakul)

Co-advisor

.....
(Dr. Sittiruk Roytrakul)

Examining Committee:

.....Chairperson
(Dr. Chantragan Phiphobmongkol)

.....Committee
(Dr. Raphatphorn Navakanitworakul)

.....Committee
(Dr. Sittiruk Roytrakul)

.....Committee
(Asst. Prof. Dr. Kanyanatt Kanokwiroon)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfillment of the requirements for the Master of Sciences Degree in Biomedical Sciences

.....
(Prof. Dr. Damrongsak Faroongsarng)
Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

.....Signature
(Dr. Raphatphorn Navakanitworakul)
Major Advisor

.....Signature
(Dr. Sittiruk Roytrakul)
Co - Advisor

.....Signature
(Miss Phetploy Rungkamoltip)
Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

..... Signature
(Miss Phetploy Rungkamoltip)
Candidate

ชื่อวิทยานิพนธ์ การหารูปแบบของเปปไทด์ในซีรัมสำหรับการตรวจมะเร็งปากมดลูกโดยใช้แมสสเปกโทรมิทรี
ผู้เขียน นางสาวเพชรพลอย รุ่งกมลทิพย์
สาขาวิชา ชีวเวชศาสตร์
ปีการศึกษา 2561

บทคัดย่อ

มะเร็งปากมดลูกเป็นมะเร็งที่พบในผู้หญิงทั่วโลกโดยเฉพาะประเทศกำลังพัฒนา เนื่องจากมีจำนวนผู้ป่วยที่เสียชีวิตเพิ่มสูงขึ้น อีกทั้งมีข้อจำกัดทางด้านความไวและความจำเพาะของวิธีการตรวจคัดกรอง (Pap smear และ HPV DNA test) จึงจำเป็นต้องหาวิธีอื่นช่วยในการคัดกรองเบื้องต้น รูปแบบเปปไทด์จึงเป็นอีกหนึ่งวิธีที่น่าสนใจสำหรับการนำมาคัดแยกคนปกติออกจากผู้ป่วยมะเร็งปากมดลูก

การตรวจหารูปแบบเปปไทด์ในซีรัมของคนปกติและผู้ป่วยมะเร็งปากมดลูก ด้วยวิธี Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) โดยใช้ตัวอย่างซีรัม จำนวน 222 ตัวอย่าง ซึ่งประกอบด้วยผู้หญิงที่ไม่เป็นโรค จำนวน 83 ตัวอย่างและผู้ป่วยมะเร็งปากมดลูก จำนวน 139 ตัวอย่าง โดยตัวอย่างทั้งหมดจะถูกแบ่งเป็น 2 ชุดข้อมูลที่เป็นอิสระต่อกัน ได้แก่ ชุดตัวอย่างที่ทำการทดสอบ (คนปกติจำนวน 30 คน และผู้ป่วยมะเร็งปากมดลูกจำนวน 75 คน) และชุดตัวอย่างที่ใช้ในการยืนยันผลที่ได้ (ผู้หญิงที่ไม่เป็นโรคจำนวน 53 คน และผู้ป่วยมะเร็งปากมดลูกจำนวน 64 คน) โดยตัวอย่างจากผู้ป่วยแต่ละรายจะถูกนำมารวมกันในปริมาณที่เท่ากันตามกลุ่มความรุนแรงของโรค [ผู้ป่วยมะเร็งปากมดลูก (ก่อนการเป็นมะเร็ง มะเร็งระยะที่ 1 ระยะที่ 2 และระยะที่ 3) และผู้หญิงที่ไม่เป็นโรค] โดยก่อนการวิเคราะห์ด้วยเครื่อง mass spectrometry ซีรัมเปปไทด์จะถูกเตรียมด้วยวิธี ZiptipC18 และวิเคราะห์ผลที่ได้โดยใช้ซอฟต์แวร์ FlexAnalysis 3.0 ClinProTools 2.2 และ BioTyper 2.0

รูปแบบเปปไทด์ในซีรัมช่วง 1000 ถึง 10000 Da ถูกนำมาวิเคราะห์ และแสดงผลออกมาในรูปแบบโครมาโตแกรม เจลจำลอง และภาพสามมิติ ของการวิเคราะห์องค์ประกอบหลักจากผลรูปแบบของเปปไทด์ เราสามารถนำมาใช้ในการแยกและจัดกลุ่มผู้หญิงที่ไม่เป็นโรค และระยะของผู้ป่วยมะเร็งปากมดลูก (ก่อนการเป็นมะเร็ง มะเร็งระยะที่ 1 ระยะที่ 2 และระยะที่ 3) รูปแบบของเปปไทด์ทั้งหมดจากการรวมตัวอย่างซีรัมตามแต่ละกลุ่ม ถูกนำมาใช้เป็นรูปแบบเปปไทด์อ้างอิงในฐานข้อมูล biotyper และใช้ข้อมูลนี้ในการจำแนกแต่ละตัวอย่างได้ เมื่อนำมาเปรียบเทียบกับข้อมูลทางคลีนิค พบว่าผลจาก MALDI-TOF MS มีผลความไว 60.97% และ ความจำเพาะ 93.63% นอกจากนี้ตำแหน่งมวลต่อประจุ 1466.91 Da มีความแตกต่างกันระหว่างกลุ่มที่ศึกษาอย่างมี

นัยสำคัญทางสถิติ (โดยใช้สถิติ Wilcoxon/Kruskal-Wallis test (pWk), $p<0.001$) ซึ่งเป็นโมเลกุลที่น่าสนใจในการนำไปศึกษาต่อในอนาคต จากผลการศึกษาในครั้งนี้จึงสรุปได้ว่าวิธี MALDI-TOF MS น่าจะเป็นทางเลือกหนึ่งสำหรับใช้ในการวินิจฉัยมะเร็งปากมดลูก ด้วยคุณสมบัติที่สามารถทำซ้ำได้ มีความแม่นยำ และใช้เวลาน้อยในการตรวจวิเคราะห์

Thesis Title	Mass spectrometry-based serum peptidomic patterns for cervical cancer detection
Author	Miss Phetploy Rungkamoltip
Major Program	Biomedical Sciences
Academic Year	2018

ABSTRACT

Cervical cancer is the common malignancy among females worldwide, especially in developing country. Due to high mortality and lack of sensitivity and specificity of the screening tests (Pap smear and HPV DNA test), an alternative tool is crucially needed. Interestingly, the peptide pattern is an alternative diagnostic tool for discriminating healthy women from cervical cancer patients.

The serum peptide profiles from healthy subjects and cervical cancer patients were acquired using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). A total of 222 serum samples (83 healthy women and 139 cervical cancer patients) were divided into 2 independent datasets: training set (30 healthy women and 75 cervical cancer patients) and validation set (53 healthy subjects and 64 cervical cancer patients). Each specimen was pooled together according to each sample group [patients with cervical cancer (precancerous, stage I, II and III) and healthy women]. Prior to mass spectrometry analysis, serum peptides were prepared by ZiptipC18. Finally, the results were analyzed using FlexAnalysis 3.0, ClinProTools 2.2 and BioTyper 2.0 software.

The serum peptide profiles in the range 1000 and 10000 Da were analyzed and shown as the chromatogram, gel view and dimensional image of principal component analysis (PCA). Based on peptide mass fingerprint (PMF), we were able to discriminate and category healthy subjects and patients with stage of cancer: precancerous, stage I, II and III. Furthermore, all PMF from pooled stage serum were used as reference mass spectrum in the in-house database and were used to classify samples. Compare to the clinical diagnosis, MALDI-TOF MS yields a sensitivity (60.97%) and high specificity (93.69%). Furthermore, the peak at 1466.91 Da was significantly differenced among the investigated groups (Wilcoxon/Kruskal-Wallis test (pWk), $p<0.001$), it is, therefore, an interesting molecule for the further study. The results of this study conclude that MALDI-TOF MS method might be used as an

alternative approach for cervical cancer test with good repeatable, high accuracy, and less time-consuming.

ACKNOWLEDGMENTS

The successful achievement of this thesis would not have been possible without the support, counseling and assisting from many people in so many ways. First of all, I would like to express profound gratitude and deep appreciation to my major advisor, Dr. Raphatphorn Navakanitworakul. It has been an honor to be her first MSc student. I heartily thankful to my major adviser for invaluable support, encouragement, supervision, suggestions and all opportunities throughout this course.

Deepest gratitude is also to my co-adviser to Dr. Sittiruk Roytrakul for his kindness, encouragement, valuable advice and providing proteomics knowledge, and opportunity for precious experiences in GPRO221, Proteomics Research Laboratory, National Center for Genetic Engineering and Biotechnology. I would like to deepest thank Mrs. Janthima Jaresittikunchai and Miss Narumon Phaonakrop for all of very kindness help, encouragement, inspiration and continuous great advice throughout my experiment. I would like to acknowledge Miss Chadaporn Chaiden, Miss Vilailak Tiyao, Mr. Paiboon Tunsagool, Miss Wassamon Moyadee, Miss Rarinthorn Sumrit, Mr. Sekkarin Ploypetch and Miss Kisana Bhinija for all the help, caring, and every concern during my experiment.

Next, I would like to take this occasion to sincerely special thank Dr. Chantragan Phiphobmongkol at Laboratory of Biochemistry, Chulabhorn research institute for consultation, taking care and providing special equipment for Extracellular vesicles (EVs) research parts. I also give appreciation to Mr. Churat Weeraphan for teaching and supporting the EVs information. Moreover, I also would like to thank Mr. Kittirat Saharat, Miss Jirawan Panachan, Mr. Tanin-ek Sriwitool, Dr. Pukkavadee Netsirisawan, Mr. Theetat Ruangjaroon, Miss Nuttaporn Paramee and Miss Jutamaad Satayavivad for their support, both academically and emotionally.

Finally, I gratefully express to thanks the Faculty of Medicine, Prince of Songkla University for a graduate student grant. This work was also supported by PSU research fund, Prince of Songkla University. I am expressing my gratitude to Miss Piyathida Molika for help and support me all times in this laboratory. Moreover, I will thank all instructors, members and staff (especially Miss Kanokon Khannui and Miss Mayuree Boonrach) at Department of Biomedical Sciences, Faculty of Medicine, Prince of Songkla University as well as the good relationship from my friend at other Faculties, Universities and country. And lastly and foremost, my deepest gratitude

goes to my beloved father, Mr. Kimkun Rungkamoltip, and mother, Mrs. Jiraphorn Rungkamoltip for unconditioned love, encouragement, and untiring support. Furthermore, both of you made me who am I now. Thank you.

Phetploy Rungkamoltip

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

%	Percentage
µg	Microgram
µg/µl	Micrograms per microliter
ACN	Acetonitrile
ASC	Atypical squamous cells
ACTH	Adrenocorticotropic hormone
ASCCP	American society for colposcopy and cervical pathology
ATP	Adenosine triphosphate
A.u.	Arbitrary unit
CA125	Cancer antigen 125
CA15-3	Cancer antigen 15-3
CA19-9	Cancer antigen 19-9
CEA	Carcinoembryonic antigen
CHCA	Alpha-cyno-4-hydroxycinnamic acid
CIN	Cervical intraepithelial neoplasia
Conc.	Concentration
$CuSO_4$	Copper (II) sulfate
Da	Dalton
DegAla-FPA	Fibrinogen A with an alanine truncation at N-terminal
dH_2O	Distilled water
DTT	Dithiothreitol
ESI	Electrospray ionization
FIGO	The International Federation of Gynecology and Obstetrics
g	Gram
GA	Genetic Algorithm

LIST OF ABBREVIATIONS AND SYMBOLS (CONTINUED)

HPV	Human papillomavirus
HSIL	High-grade squamous intraepithelial lesion
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LSIL	Low-grade squamous intraepithelial lesion
m/z	Mass-to-charge ratio
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MCP	Microchannel plate
ml	Milliliter
MS	Mass spectrometry
MWCO	Molecular weight membrane cut off
Na_2CO_3	Sodium carbonate
Na_2HPO_4	Sodium phosphate dibasic
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ORFs	Open reading frames
Pap smear	Papanicolaou smear
PBS	Phosphate buffered saline
PCA	Principal component analysis
PEG	Polyethylene glycol
PMF	Peptide mass fingerprint
pRB	Retinoblastoma protein
PSA	Prostate-specific antigen
pWK	Wilcoxon/Kruskal-Wallis test
RPM	Revolutions per minute
SCCA	Squamous cell carcinoma antigen
SELDI-TOF	Surface-enhanced laser desorption/ionization time-of-flight
SD	Standard deviation
SDS	Sodium dodecyl sulfate

LIST OF ABBREVIATIONS AND SYMBOLS (CONTINUED)

SNN	Supervised Neural Network
sLea	Sialyl Lewis-a
TFA	Trifluoroacetic acid
TOF	Time-of-flight
v/v	Volume by volume
w/v	Weight by volume

CHAPTER 1

INTRODUCTIONS

BACKGROUND AND RATIONALE

Cervical cancer is ranked as the fourth most frequent cancer among women worldwide, with 528,000 newly diagnosed cases and 266,000 deaths annually (1). In Thailand, the most recent data indicate that cervical cancer is the second most commonly diagnosed cancer in females behind breast cancer which incidence and mortality cases are about 8,184 and 4,513 cases, respectively (2). The incidence and mortality rates of cervical cancer have considerably reduced in developed countries through the cervical cancer screening tests for human papillomavirus (HPV) and through Papanicolaou smear (Pap smear); nevertheless, the incidence rate of this disease remains high in developing countries, where 70% of the patients are at the advanced stage of the cancer and the age of the diagnosis patients is slightly decreasing (3, 4). According to American society for colposcopy and cervical pathology (ASCCP), the 30-65-year-old women are recommended to diagnose cervical cancer by using the combination of screening with Pap smear and HPV DNA test (5, 6). Pap smear is a rapid cytology-based microscopic examination to monitor precancerous changes which can detect the atypical squamous cells at transformation zone (5). Because of Pap smear screening test, cervical cancer incidences in the United States has been reduced by nearly 70% between 1955 and the mid-1980s (7). Nonetheless, specificity and sensitivity of cytology-based screening are 51% and 66.6%, respectively (8). Moreover, the conventional Pap smear test showed a false negative rate about 14-33%, two of third of false negative results from sampling and slide preparation (7). Additionally, most women are embarrassed to undergo the screening along with a pelvic examination (9). HPV DNA detection has been examined for supporting the accuracy of Pap smear which determines women at risk for cervical neoplastic without cervical cytological process. Several studies demonstrated that HPV screening test has a higher negative predictive value and sensitivity than cytology-based screening (10- 13). However, the multiple HPV type infections are contained in samples at low DNA yield. Hence, the amplification of HPV DNA with the low yield of several HPV genotypes resulted in false-negative (14). According to the limitation of screening tests, the establishment of optimal

biomarker testing from body fluids would be the alternative screening test for the early detection.

Many researchers have endeavored to identify new cancer biomarkers in serum that reflect a particular pathophysiological state (15, 16). Currently, several serum proteins are widely used as a biomarker for cancer diagnosis. For example, squamous cell carcinoma antigen (SCCA) is most commonly used as serum tumor marker in cervical cancer diagnosis. Moreover, serum SCCA has a high concentration in patients with squamous cell carcinomas of esophagus, lung, head and neck and anus (17- 20). Furthermore, other serum markers as carcinoembryonic antigen (CEA), cancer antigen 19-9 (CA19-9) and cancer antigen 125 (CA125) have demonstrated as a utility in patients with cervical adenocarcinoma (21-25). However, the detection of these proteins has been used in other cancers including lung, colorectal, ovarian, gastric, breast and pancreatic cancer (23- 34). Thus, the use of current biomarker is still less specificity and sensitivity. Moreover, the complexity of high abundant protein in serum might obstacles the low concentration of biomarker detection. Therefore, determination of serum peptide patterns has been a considerable attention.

Nowadays, the subsequent for biomarker research field is illustrated by serum peptide profiles. Mass spectrometry (MS) is an approach for protein or peptide patterns and identification, which technology distributes for high throughput, rapid and nonbiased novel biomarkers discovery (35). Matrix-associated laser desorption/ionization time-of-flight (MALDI-TOF) is one type of mass spectrometry that detects mass-to-charge (m/z) of sample compound. This measurement process composes of picomole sensitivity, high accuracy, high resolution and high-throughput (36). MALDI-TOF MS provides the tool needed to define specific proteins or peptide patterns/signatures that can be used for clinical sample identification and classification (36). Peptide patterns analysis has been widely tested on various types of cancers, including gastric, ovarian, hepatic and breast cancer (36). Nonetheless, the serum peptide patterns correlated with the developmental stages of cervical cancer have been rarely reported. In this study, we used MALDI-TOF MS to investigate the difference in serum peptide patterns of patients with cervical cancer in various developmental stages and healthy women, to discriminate the progression of disease and also distinguish healthy women from cancer patients.

LITERATURE REVIEW

1. Cervical cancer

1.1 Anatomy of cervix

The female reproductive system is the important system to enable the species reproduction (processing of pregnancy and birth). This system consists of external and internal organs. The external genital or vulva is the open of vagina to the outside which composes of Bartholin's gland, mone pubis, hymen, labia majora, labia minora, clitoris, vestibule and perineum. These external organs have 3 functions: supporting sperm to enter the body, providing the sexual pleasure and guarding the internal organs (37). The internal genital organs include ovaries, fallopian tubes, vagina and uterus. The uterus and vagina place after and above the pubic bone in the pelvis. In front of internal organs is urinary bladder and urethra. The rectum is behind the vagina and uterus as shown in Figure 1A. Ovaries are a pair of organs; each one stated on the right and left side of the uterus. The ovaries are responsible for egg generation and the estrogen and progesterone hormone production**.** Fallopian tubes are tubes extend from the uterus to ovaries which are customarily sites of fertilization. The vagina is the internal organ that connects the external organs to the uterus. Uterus or womb is the thick-walled of pear shape organ which locates near the pelvic cavity. A blastocyst or fertilized egg implants and grows in the uterus hollow. Uterus consists of 4 parts which are fundus, corpus, isthmus and cervix. The fundus is the upper part of the uterus that attaches to the fallopian tubes. The main body of the uterus is called corpus. This part accommodates a developing fetus. The isthmus is the position between the corpus and the cervix (38).

The cervix is the lowest one-third part of the uterus where connects between the uterus and the top end of the vagina. The upper two-thirds of cervix lie above vagina is called endocervix. The endocervix's opening is called the internal os. The ectocervix is the outer part of the cervix that contacts with the vagina. The endocervical canal runs through the cervix from internal os to the uterine cavity (external os) as shown in Figure 1B. In cervix and uterus, the internal iliac arteries and their branches supply, moreover the vein and lymphatic ducts run parallel to the arteries.

The cervix surface is arranged by stratified squamous epithelium (ectocervix and vagina) and columnar epithelium (cervical cannel extents outward to ectocervix). The transitional or transformation zone is position which changes columnar epithelium cell into squamous cells. This zone is most common are for cervix abnormal cell development (38).

(A)

(B)

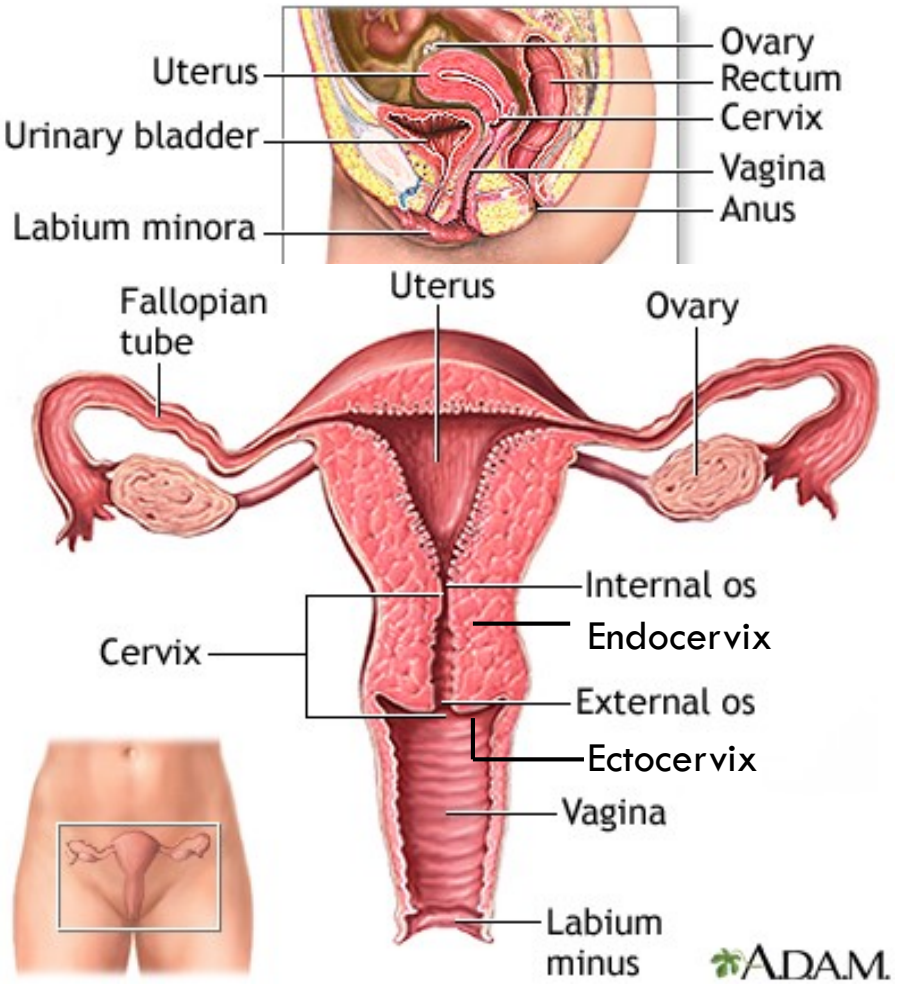


Figure 1. Female reproductive organs. The female internal organs consist of vagina, fallopian tube, ovary and uterus. (A) In the sagittal view, uterus is in front of rectum and behind the urinary bladder. (B) The lower one of third of uterus is cervix which includes 2 parts: the ectocervix and endocervix as shown in the front view of female reproductive organ. (Adapted from Baggish et al., 2010, (37))

1.2 Cervical cancer epidemiology

Worldwide, cervical cancer is a major health-related problem among woman. Cervical cancer has the fourth highest incidence worldwide, with an incidence of 527,600 estimated new cases and 265,700 estimated deaths in 2012 (1). Moreover, 87% of death cases arise in the less developed country (39). Additionally, the trend in cervical cancer incidence is still increasing every year (3). The incidence rate of cervical cancer is fewer patients under the age of 25 years; nevertheless, the incidence increases in women at age 35 to 40 years. There was the maximum cervical cancer patient in women of ages 50 and 60 years. In Thailand, cervical cancer is the second most frequent of women malignancy behind breast cancer which incidence and mortality cases are about 8,184 and 4,513 cases, respectively (2).

1.3 Etiology of cervical cancer

The risk factors of cervical cancer are immunocompromised, taking an oral contraceptive pill, smoking, *Chlamydia trachomatis* infection, using an intrauterine device, taking Diethylstilbestrol, having a family history of cervical cancer and multiple sexual partners (40). Additionally, 95% of cervical cancer cases are caused by HPV infection. HPV can be classified into 2 groups which are the low-risk and high-risk type. The low-risk HPV type is the main cause of genital wards which includes 6, 11, 42, 43 and 44. The high-risk HPV type incorporates 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 70 which are most common types for cervical cancer initiation(41). Thirty HPV types can transmit through sexual contact and infect the vagina, vulva, penis, anus and cervix (41). HPV type 16 infected patients are a half of cervical cancer cases. The HPV life cycle is closely related to the various stages of HPV-infected epithelial cells. The HPV contains double-stranded circular DNA which divided into three regions. First, the early region contains core promoter, silencer and enhancer sequence which control the transcription of the open reading frames (ORFs). Second, the late region composes of ORFs, E1, E2, E4, E5, E6 and E7 which are related in oncogenesis and viral replication. HPV E1 protein is an enzyme associated with ATP dependent viral DNA helicase which involves in DNA replication (42). HPV E2 codes for transcriptional activator proteins in viral gene transcription and replication (43). E4 protein plays a role in cytokine network destabilization (E1 and E4 fusion protein) and viral releases from epithelial layers (44).

E5 acts as a transmembrane protein, which binds with epidermal growth factor receptor resulting in activating the mitogenic pathways (45). Both E6 and E7 viral protein control cell cycle to amplify genome in epithelial layers. HPV E6 protein can act as oncogenic molecule which interacts and degrades p53 through ubiquitin degradation (46). HPV E7 oncogene binds the retinoblastoma family (pRB), which is degraded (47). Consequently, E2F releases from pRB. E2F transcription factor encodes cyclin A and E which drive the cell cycle, resulting in uncontrol cell proliferation (48). Third, genome region encodes the L1 and L2 for viral capsid (49). L1 is the major capsid protein associated self-assembly into viral-like-particles (50). HPV L2 (the minor capsid protein) involves in virion formation and the infected host cell. L1 and L2 viral capsids are undetected in infected epithelial cells, however, the differential cell at the upper layers of the squamous epithelium occur both L1 and L2 viral capsid (51).

1.4 Pathogenesis of cervical cancer

HPV particles infect the basal epithelial cell at the cervical squamocolumnar transformation zone as shown in Figure 2. The viral genome initiates amplification prior to viral genomes maintain in the infected cells. First, both E1 and E2 expression assist the genome amplification. E1 enzyme acts as ATP dependent helicase which interacts with AT enrich sequences at the replication origin. E1/E2 complex recruits the essential cellular enzyme for viral genome replication. Viral genomes are synthesized without control from cell cycle owing to E6 and E7 function (45). When the viral life cycle completes, the expression of L1 and L2 are imported into nuclear by karyopherins. After that, L1 and L2 assembly and increase the efficiency of viral-like particles (52). E4 binds with cytokeratin filaments to facilitate virion release to infect the other normal cell (44). The infected cells are not treated resulted in the precancerous and cancer development.

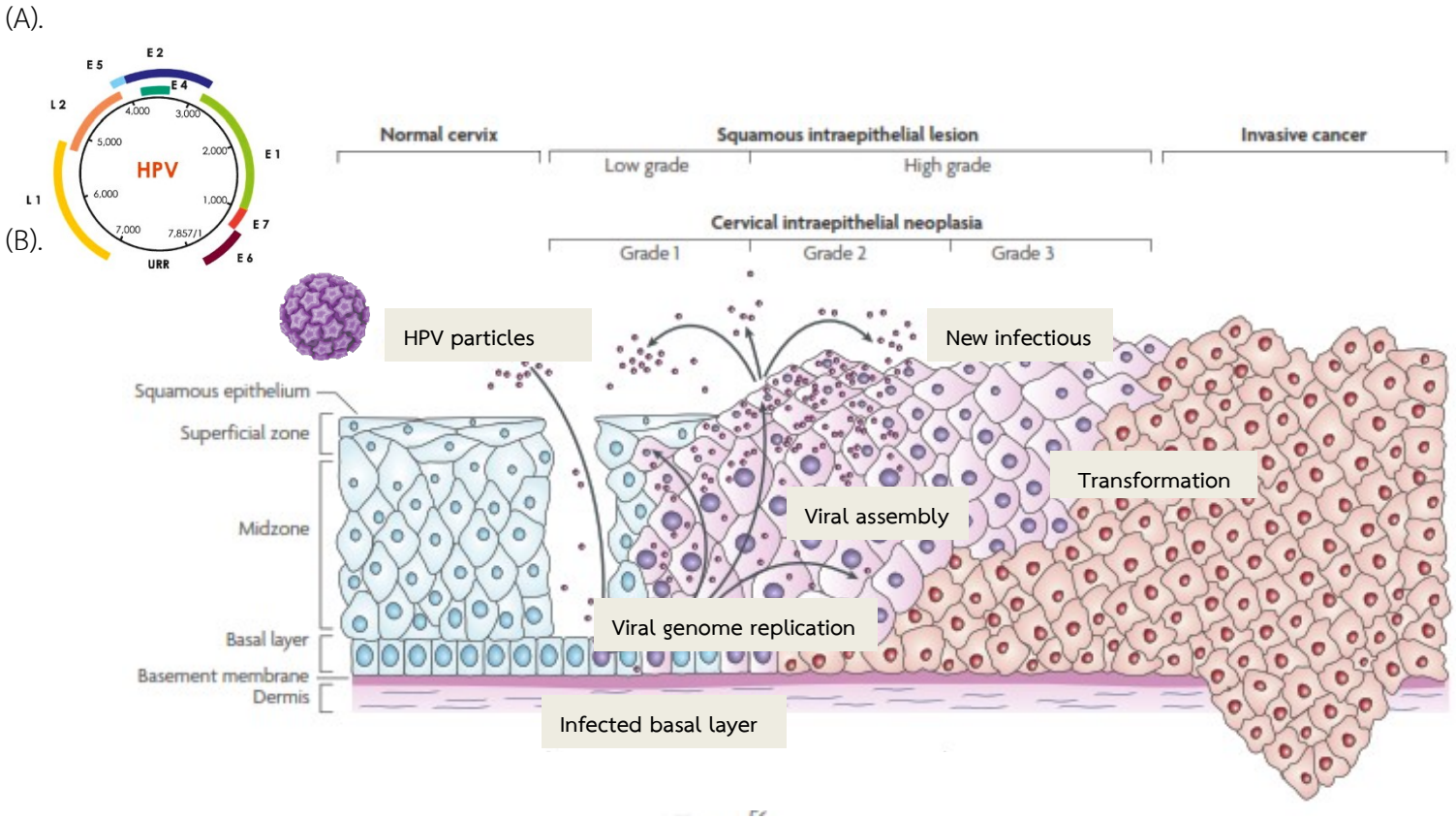


Figure 2. HPV structure and role in cervical cancer. (A) HPV encodes 6 early proteins (E1, E2, E4, E5, E6 and E7) and 2 late proteins (L1 and L2). (B) HPV infects the basal layer of host cell. The HPV genome rapidly replicates and assembles, thus the viral particles are accumulated in the host cells. The untreated lesion is developed for the dysplasia, micro-invasive and invasive cancer, subsequently. (Adapted from Kelloff et al.,2007, (53))

1.5 Type and grading of cervical cancer

Cancer is defined by abnormal and uncontrolled cell proliferation. Cervical cancer can be categorized into 4 types: squamous cell carcinoma, adenocarcinoma, adenosquamous cell carcinoma, and neuroendocrine carcinoma (54). The major type of cervical cancer is squamous cell carcinomas which initiate at the exocervix (squamocolumnar junction). Cancer arises from endocervical columnar cells called adenocarcinoma (55). Adenocarcinoma was still high the incidence cases, resulting from the limitation of preinvasive stage detection test (55). Moreover, the squamous cell carcinoma has a good prognosis than adenocarcinoma (56). Adenosquamous carcinoma is tumors including squamous and glandular carcinoma. This type is related to a pelvic lymph-node metastasis higher risk than adenocarcinomas or squamous cell carcinomas (57, 58). The cervical cancer type can identify via cytology examination.

In precancerous classification system, the precancerous condition is classified based on the histological examination (WHO descriptive classification) and cytology (Bethesda system). According to histological examination, premalignant of squamous cell at the cervix surface or cervical intraepithelial neoplastic (CIN) can be graded into 3 types that depend on the depth of changed cell on the surface of the cervix. Classifications of CIN are CIN1 as only mild dysplasia, CIN 2 as two–thirds of the surface layers and CIN3 as the full thickness of the surface layer (59). The WHO classification has been used for the diagnostic test. On the other hand, the classification based cytology examination has been applied for the screening test. For the cytological reports, cytology test can classify precancerous condition into 4 groups: normal, atypical squamous cells (ASC), low-grade squamous intraepithelial lesion (LSIL) and high grade squamous intraepithelial lesion (HSIL). ASC is the slightly abnormal finding in Pap smear. LSIL is the mildly abnormal cell. HSIL is more serious transforms in the cervix than LSIL (38).

1.6 Stage of cervical cancer

The International Federation of Gynecology and Obstetrics (FIGO) classified cervical cancer bases on size of tumor and spread of disease. The mandatory technique for staging is speculum, vagina, and rectal examination, intravenous pyelogram and abdominal ultrasound. The patients with cervical cancer can classify into 5 stages which are stage 0, I, II, III, and IV as shown in Figure 3. Stage 0 or carcinoma in situ is the cancer cell could not develop beyond the basement membrane of cervix. Stage I is micro-invasive cervical cancer. Stage II shows the carcinoma spreading beyond the cervix, but not to the pelvic wall. Stage III is carcinoma spreading onto the pelvic wall and stage IV is an invasive carcinoma which spread to the other organ (60, 61).

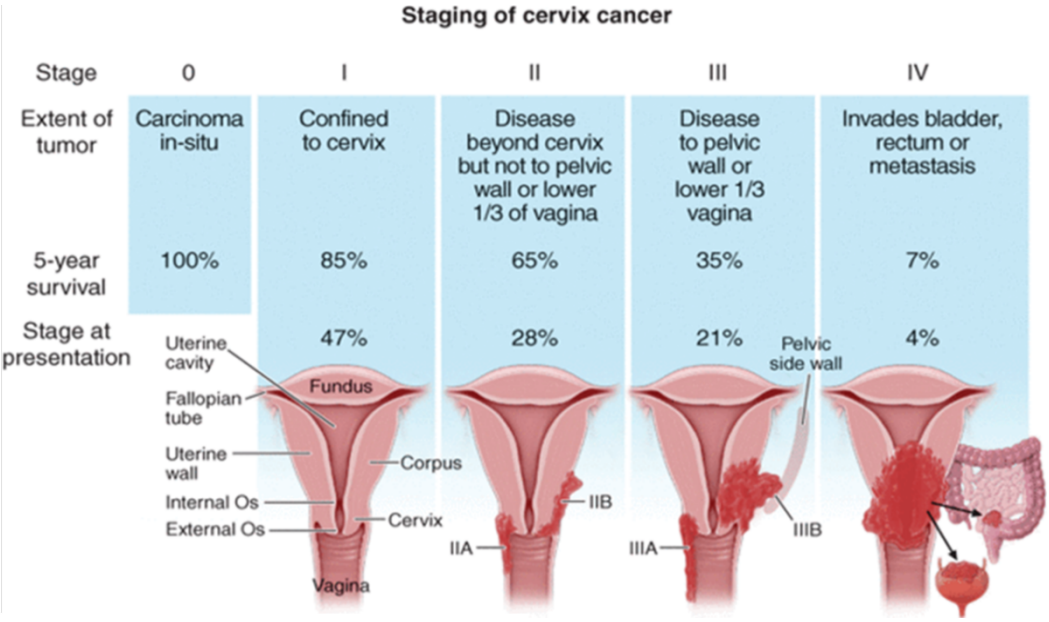


Figure 3. Cervical cancer staging according to the FIGO system. Cervical cancer staging is divided into 5 stages which is stage 0, I, II, III, and IV (60).

1.7 Cervical cancer screening test

Nowadays, the screening test in the Era of precision medicine is an approach which supports the effective and accurate treatment. Treatment in advance stage (metastasis cancer) is less effective and more difficult than treatment in early stage (localized cancer). The cancer treatment prior to metastatic spread reinforces survival rates improvement. Early detection or screening test may be the most powerful program for minimizing the unpleasant and death of cancer patients (62). Thus, the accuracy of screening test has been concerned (63).

In cervical cancer, cytology examination and HPV DNA test are the screening approach (5, 6). The cytological smear was used to detect the abnormal cervical cancer (precancerous and cancerous processes in the cervix) which collects cells from the outer opening vaginal canal (5). The cells are stained via acidic and basic dyes. Hematoxylin is the basic dye which binds to the basophilic substance (DNA and RNA). Eosin acts as acid dye and binds with basic part of the cell such as positively charged amino acid in the cytoplasm (64). In Pap smear results, atypical squamous cell (ASC cells) shows irregular and dark oval cells which increase two to three times nuclei by compares to the normal cells. LSIL contains cells with the 3-6 times larger nuclei than normal cells, nucleus to cytoplasm ratio below one-third, transparent cytoplasm, and poorly cell boundaries. Moreover, HSIL is abnormal nuclear chromatin pattern, multinuclear, two-thirds of the nuclei-cytoplasm ratio, hyperchrome with reticulate chromatin, cyanophilic, amphophilic or orangeophilic cytoplasm (65). Previous studies showed that Pap smear contributed to decreasing the incidence; nonetheless, this screening test of cervical cancer is still limited in sensitivity, specificity, high false negative rate and uncomfortable sample collection (7, 9). HPV DNA detection is another one approach which determines women at risk for cervical neoplastic without cervical cytological process. HPV screening test has a high negative predictive value and a high sensitivity when compared with cytology tests (10-13). Nevertheless, HPV DNA test has low positive predictive value and low specificity; low yield of HPV DNA in sample cases is the limitation of this technique (66, 67). Therefore, the biomarker may be an alternative approach which improves the early detection of cervical cancer.

2. Biomarker

2.1 Ideal characteristics of biomarker

According to the definition of the National Cancer Institute, a biomarker is a biological molecule occurs in body fluids or tissues which is the indicator of normal, abnormal, or disease condition (68). An ideal biomarker should describe the happening of the disease and must have several characters for clinical application(69). Firstly, the marker test must be as non-invasive collection, meaning it must be simple and safe to achieve (70). Secondly, an ideal biomarker should relate the disease progression. Thus, the biomarker detection test should classify the condition of disease and detect the treatment response (71). Thirdly, tumor marker should be high sensitivity for the early detection of disease. Fourthly, a biomarker should be specific for present in this disease an absent in the other condition. Moreover, both sensitivity and specificity should be high values, therefore false-positive and false-negative values diminished (69, 70). All characteristics of ideal biomarkers were summarized in Table 1.

Table 1. The ideal biomarker characteristic (71).

Characteristic	Description
(1) Non-invasive collection	Expression within a sample obtainable without discomfort to the patient
(2) Readily available	Presentation in an easily obtainable sample that is commonly obtained clinically such as blood or urine
(3) High sensitivity	Allows early detection of disease with little or no overlap between healthy and diseased patients
(4) High specificity	Present in the disease in question, with little or no overlap between comorbid conditions
(5) Rapid response	Changes rapidly in response to treatment
(6) Risk stratification	Provides prognostic information to the clinician, allowing classification of the disease along with diagnosis
(7) Insight to disease	Provides insight into the underlying mechanism of the disease

2.2 Source of biomarkers

A biomarker relates to a cancer characteristic or host cell-mediated response which may have clinical relevance (70, 71). Thus, biomarkers are valuable appliances for cancer screening, diagnosis, monitoring and treatment (33). The molecules as biomarkers are elevated during cancer progression which can be detected in a variety of cell lines (in vitro), animal base and clinical model (tissues and biological fluid). The specimens from human model are high complexity and clinically relevant than cell culture and animal base model as shown in Figure 4. Nevertheless, cell culture and animal model are commonly used as a preliminary phase of biomarker development (70).

Firstly, the use of in vitro experiment or cell culture model is simple model for cancer biomarker development research. Studies have appeared cell signaling pathways leading to test and therapy development (71). The two-dimensional cell cultures characterize the molecules separating disease cells from their innate environment. Cell lines are high homogeneity (high similarity with initiate cancer cells) resulted in reproducibility of results. Moreover, this model provides the easy substitution, immediate accessibility and easy to operate. Nevertheless, characteristic in long-term of cancer cell lines, loss of heterogeneity, distinguishes environment of tumors are the limitation of two-dimensional cell cultures (70, 72). As a result, these limitations in the differentiation of cancer environment, 3-dimensional (3D) cell culture methods have been studied (73). The markers relate to clinical phenotypes. 3D culture experiments are more likely to strengthen the knowledge of tumor biology (73, 74). The 3D culture simulates cancer cell behavior as factors similar to animal base models (75, 76).

Secondly, the animal model is a realistic model for clinical biomarker development. An ideal characteristic of model has been similar with both physiological and pathological features as human cancer and high-quality control samples (microenvironmental and genetic variability). The experimental cases in this model have less variable and high homogeneous samples than human specimen resulted in successful biomarker discovery (77). Nevertheless, the limitation of animal model is the imitation of complicated mechanisms in human such as carcinogenesis and cancer progression. Therefore, the average success rate of the biomarker from animal to clinical phase is less than 8%. Animal-based models are preliminary models prior to the models of human (62, 77- 79).

Thirdly, the biomarker investigation of human specimen is more specific with cancer progression than others models. Thus the candidate molecules from tissues and biological fluid are applied to the clinical test. The most common source for biomarker development is tissues, cerebrospinal fluid, saliva, urine, blood (80). The biomarkers originate from the tissue at disease organ. As a result, these markers are the most specific with cancer (71). Nevertheless, these tissues are complicated integrating many different cell types. Specimen collection may not simply incorporate tumor region but also contains blood as well as normal tissue (81). Urine is non-invasive biospecimen which assists diagnostic information. Advantages of urinary biomarker are accurate, timely and cost-effective. However, the biomolecules as biomarker should concentrate prior to biomarker discovery step (82). The blood source is that the cell secretes many molecules into the bloodstream resulting from the response to the disease. Thus, the molecules reflect both physiological and pathological states. Biomarkers from blood are also high clinically relevant, easily collection, minimally invasive source that are most commonly for clinical biomarker investigation (71, 80).

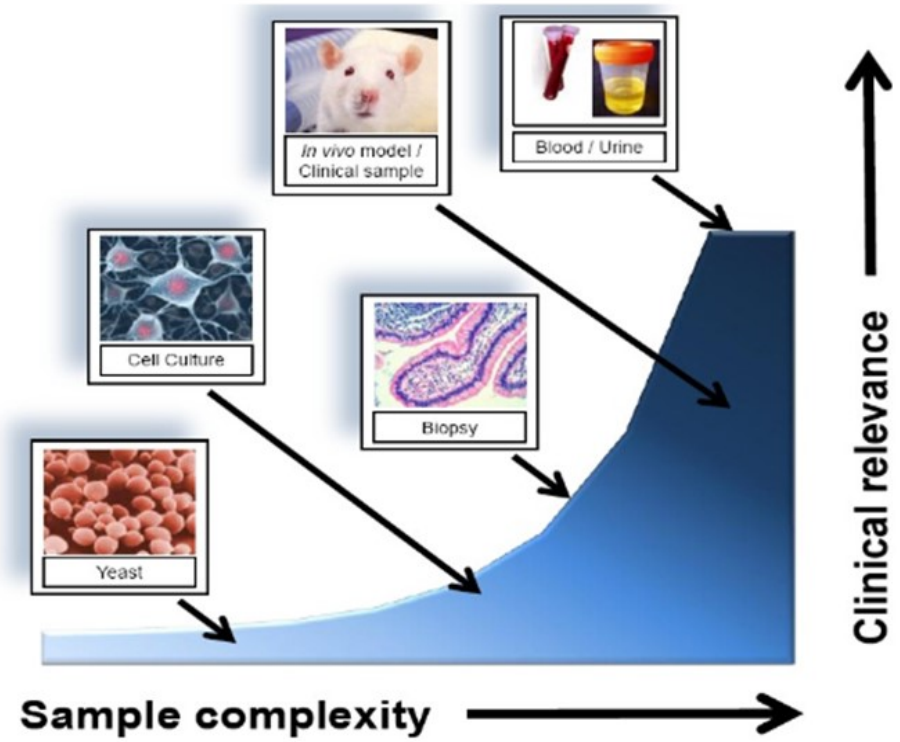


Figure 4. The potential source of biomarker. Biomarkers have been established from several sources such as tissue biopsy, urine, whole blood, plasma and serum. The blood-based biomarkers are also high clinically relevant and a minimally invasive source (71).

2.3 Current biomarker for cancer diagnosis

Serum biomarkers are used for diagnosis of cancer as shown in Table 2 (15, 16). The prostate-specific antigen (PSA) is a serine protease of Kallikrein genes family that is secreted by prostate cells both normal and neoplastic epithelium. PSA represents the invasion and metastases states. Nowadays, PSA has used as an early diagnostic biomarker for prostate cancer. This molecule is available in small quantities in serum from healthy and increases in the prostate cancer (83, 84). The CEA is an oncofetal antigen produced by human colonic carcinoma (85). The first in developing CEA blood test is detected in colon cancer patients. Elevated CEA level is also detected in many other cancers including bladder, thyroid, pancreas, lung, liver, stomach, prostate, ovarian, breast and cervical cancer (86, 87). This antigen is useful in the prognostic marker of lung cancer. Moreover, CEA levels have been used as a prognostic or surveillance in colorectal cancer (26, 27). CA125 is a glycoprotein in

mucin family presents on coelomic epithelium which is highly expressed in ovarian cancers (88). CA125 antigen is commonly used as a biomarker for prognosis and monitoring of ovarian cancer patients (28, 29). Moreover, the CA-125 arises in other types of malignant, such as pancreatic, colon, endometrium, lung, breast and cervical cancer (86, 89, 90). Cancer antigen 15-3 (CA15-3) is mucin-like glycoprotein that is produced by secretory mammary epithelial cells. Its function is lubrication and cell protection. Additionally, CA15-3 reduces cell adhesion which is involved in metastasis. Elevated CA15-3 levels are found in breast cancer and others, such as lung, ovarian, colorectal and pancreatic cancer. The breast cancer patients with low CA15-3 are good prognosis than patients with high level. Therefore, the CA15-3 has been used as the biomarker for monitoring treatment and recurrence of cancer (30-33). CA19-9 is sialyl Lewis-a (sLea) that occurs in the cancer cell surfaces (91). The interaction between CA19-9 ligand and E-selectin receptor on the cytokine-activated endothelial cell induces cancer invasion and metastasis. CA 19-9 is expressed in tissues and serum of cancer patients, such as pancreatic, colorectal, esophageal, biliary, gastric and cervical cancer. Due to the high specificity, the CA19-9 is used to universal application biomarker in diagnosis, monitoring progression and follow up treatment of hepatocarcinoma, biliary carcinoma, colorectal, gastric, pancreatic and cervical cancer. Moreover, an elevated level of CA19-9 can also occur in nonmalignancy diseases as cholangitis, cholecystolithiasis, toxic hepatitis (34, 92-94). The SCCA is a squamous cell marker of cervical cancer. SCCA level is correlated tumor size, stage, invasion and metastasis (86). According to the European group of tumors cancer, SCCA is useful for following squamous cell carcinoma progression (95). In addition to cervical cancer, elevated SCCA levels are found in other squamous cell carcinomas such as esophagus, lung, head and neck and anal cancer (17, 18, 20). Therefore, the single biomarker displayed low values of specificity, sensitivity and accuracy.

Table 2. Biomarker for diagnosis of cancers. Several serum proteins are widely utilized as a tumor marker for cancer diagnoses.

Biomarker	Cancer type	Reference
PSA	prostate cancer	(83)
CEA	bladder cancer, colon cancer, thyroid cancer, lung cancer, stomach cancer, liver cancer, prostate cancer, ovary cancer, pancreatic cancer, breast cancer, cervical cancer	(85- 87)
CA125	ovarian cancer, pancreatic cancer, colon cancer, fallopian tubes cancer, breast cancer, endometrium cancer, cervical cancer	(28, 86, 89, 90)
CA15-3	breast cancer, lung cancer, ovarian cancer, colorectal cancer, pancreatic cancer	(30- 33)
CA19-9	pancreatic cancer, colorectal cancer, esophageal cancer, gastric cancer, hepatocarcinoma, biliary carcinoma, cervical cancer	(34, 86, 92- 94)
SCCA	esophagus cancer, cervical cancer, head and neck cancer, lung cancer, anal cancer	(17-20, 86)

2.4 Combination of markers

The biomarker is the utility for clinical test; nevertheless, a single biomarker is incompetent in describing complicate pathological processes. The combination of biomarkers measurement achieves more understands in the complex transformation and an improved the biomarker detection. In 2000, the first report of protein biomarker panel was published (96). The 4 protein biomarkers (s-100 protein, thrombomodulin, myelin basic protein, and neuron-specific enolase) play important roles in acute stroke. 93% of patients with acute ischemic stroke were diagnosed by using the least one of biomarkers combination (96). In the same manner, CXCL10 is the biomarker for sleeping sickness which classified stage 1 and stage 2 patients, with a specificity of 100% and sensitivity of 84%. Moreover, the panel biomarker CXCL8, CXCL10 and heart-fatty acid binding protein supported the detection test for sleeping sickness stage 2 patients. The combination displayed 97% sensitivity and 100% specificity (97). A combination of biomarkers including CEA, New York esophageal cancer-1 antibody cytokeratin-19 fragment 21-1, hepatocyte growth factor, and CA-125 showed the differential expression among lung cancer and healthy subjects. The biomarker panels as clinical feature had 49% sensitivity and 96% specificity (98). For instance, 6-biomarker panel (cyclin D1, glutathione s-transferase pi 1, lemur tyrosine kinase 2, hepsin, myosin VI and fibronectin 1) were applied to distinguish prostate cancer from benign prostate. As a result, the panel of 6 biomarkers has higher specificity than PSA at the 4.0 ng/ml cut off value (99). Nowadays, the combination of biomarkers was studied by using the mass spectrometry. For example, in 9 protein peaks (3972, 5336, 11185, 4062, 4071, 4609, 6950, 8115 and 8133) were detected by SELDI-TOF MS. This SELDI pattern as screening application can discriminate the patients with breast cancer from the non-cancer subjects. Therefore, mass spectrometry is the one approach for generating the combination of biomarkers (100).

3. Mass spectrometry

3.1 Basic of mass spectrometry

Mass spectrometry is an approach to protein profiling and identification (35). The typical MS has incorporated 3 parts that are an ion source, analytical part and detector as shown in Figure 5. A mass spectrometer detects the molecule by m/z measurement.

The molecules are induced to become the ionized molecules by MALDI and electrospray ionization or ESI (101). Generally, MALDI is the ionization approach which uses the laser energy absorbed matrix to generate ionized molecules. MALDI bestows the 1+ charge to the molecules (102). In contrast, the ESI generated a range of charged species (+1, +2, +3, +4 and so on) for each molecule. ESI generates the ionize analyst in capillary tip via liquid dispersion prior to travel in mass analyzer (101, 103).

The different mass analyzer types detect the ions in different principle. Most often these mass analyzers are time of flight (TOF), quadrupole and ion trap. First, TOF mass analyzer separates ions by time in the flight tube under the vacuum condition. The kinetic energy of ion is a principle of TOF. Whereas this principle is not in quadrupole instruments (104). Quadrupole mass analyzer separates molecules based on m/z by detection of ion movement in the dynamic electrostatic field in area between each opposing rod pair. The ion trap separates the molecule in the electric field same as the quadrupole. The mass analyzer composes of a specific voltage ring electrode and end cap electrodes which trap ions of certain m/z values until ions more stabilized. Then the ions travel to the detector (104). Samples are subjected to MS prior to vaporization and ionization by the ion source. The ions of molecules are accelerated under vacuum. The molecule ions encounter the mass analyzer (electric or magnetic field); consequently, the molecules deflect their own paths based on their m/z then hit the detector (electron multipliers or microchannel plates). Nowadays, the liquid chromatography-tandem mass spectrometry (LC-MS/MS), surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) and MALDI-TOF MS were the most common applications in the proteomics/peptidomics field. SELDI-TOF MS and MALDI-TOF MS are applied for protein fingerprint detection. SELDI-TOF MS analyzes the differences fingerprints among group via capture peptide on a solid-phase protein chip surface.

Some proteins in the samples bind with the surface modified plate, thus the only targeted peptide/protein group is detected. Although the sample preparation step of SELDI-TOF MS is unique, the co-crystallization and ionization step are similar to MALDI-TOF MS (105).

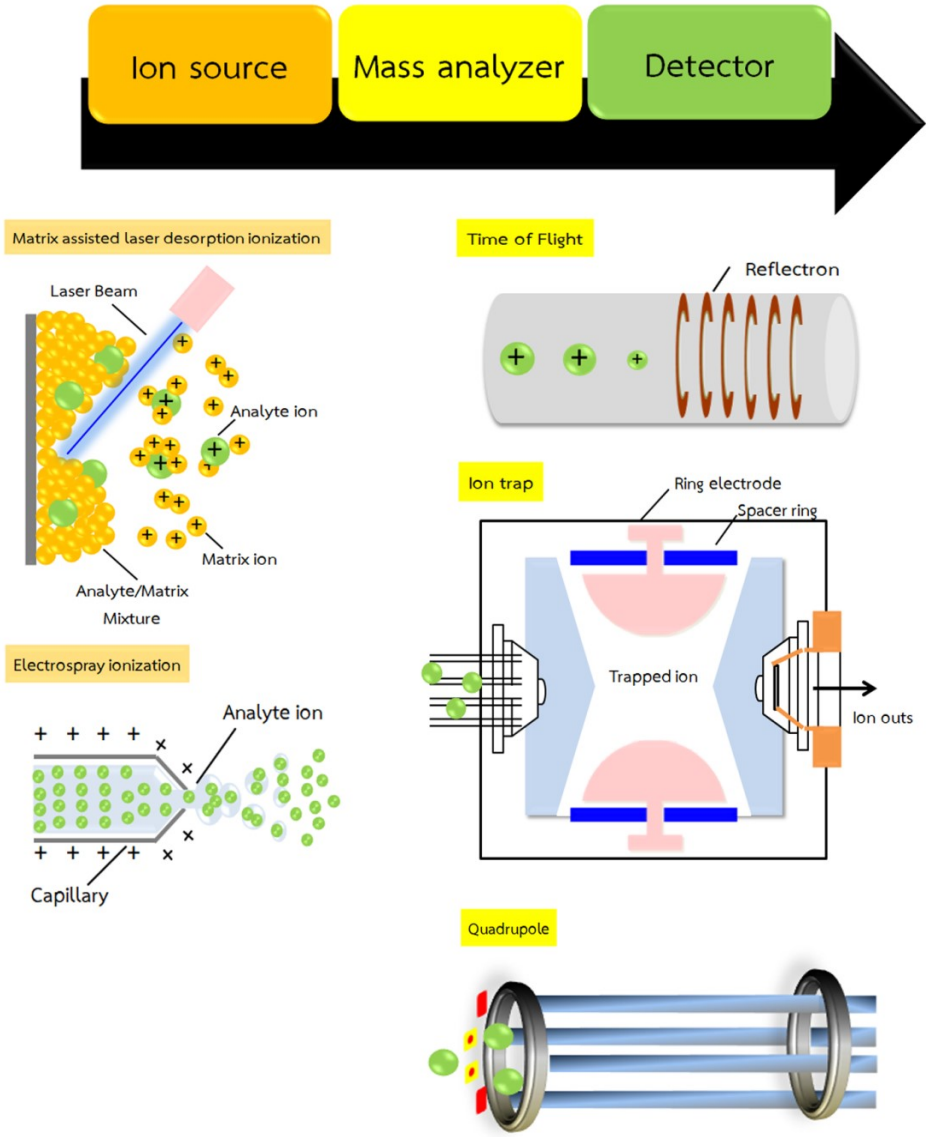


Figure 5. The typical MS has incorporated 3 parts that are an ion source, analytical part and detector, consequence there are platforms based on the sample properties, the mass spectrometer type. The ionization can divide into 2 platforms (MALDI and ESI). In mass analyzer, there are three platforms: time of flight, ion trap and quadrupole.

3.2 Principle of MALDI-TOF MS

The peptide patterns research field was performed by MALDI-TOF MS. MALDI-TOF MS is used as the cancer diagnosis approach, resulting from picomole sensitivity, high accuracy, high resolution and high-throughput of this technique (63). MALDI and TOF were used as the ionization and mass analyzer step in this platform. First, the sample is co-crystallized with matrix prior to laser energy absorbing matrix. Then The matrix transfers the energy to molecules resulted in molecules in solid are changed into the gas phase (106). The ions pass sample electrode and extraction electrode to the Einzl lens. The suppressed undesirable ions in ion path were performed by Einzl lens (107, 108). Then, the ions under the vacuum travel in field-free drift region in order to detect by microchannel plate I (MCP I) in linear mode. As a result, the peptide mass spectrums are shown in this mode. For identification step, the ions pass through the reflectron under high voltage. The peptide ions breakdown to sequence and hit with the MCPII detection as shown in Figure 6. The consequence, the MALDI-TOF MS base peptide/protein fingerprints are displayed after MCP II detection (36, 107).

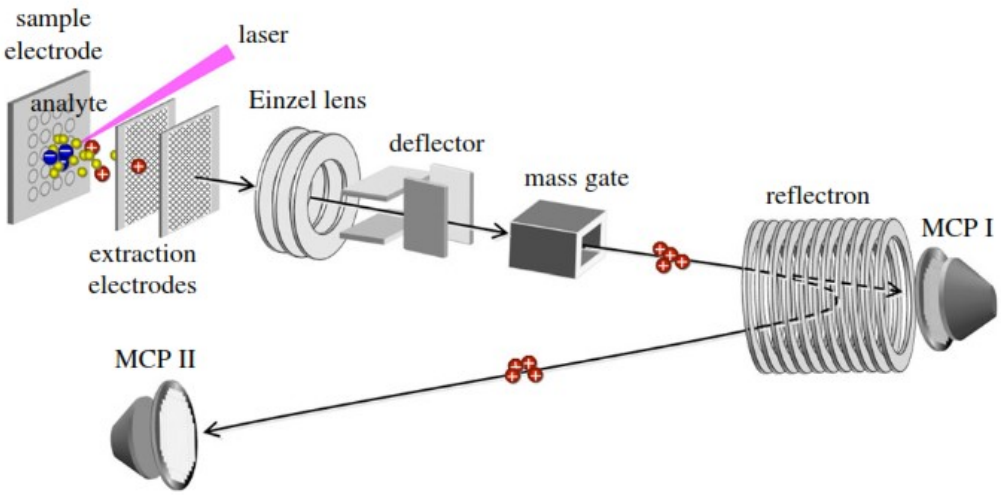


Figure 6. Pictorial diagrams of MALDI-TOF MS/MS (74). MALDI-TOF MS is a platform of mass spectrometry which detects the mass per charge measurement. MALDI-TOF MS composes of 3 steps. The first step is ionization; matrix absorbs the energy which transfers to sample. The sample becomes ions. The ion travels in the field free drift region in order to the detector (MCP I & II). The smaller molecules travel shorter time than the larger one (108).

4. Peptide fingerprints in cancer

Mass spectrometry is an effective diagnostic tool for proteins or peptides detection in the biological fluid (109). The previous study shows that several low molecular weight peptide spectrums were associated with the pathology condition. Peptidomic pattern analysis has been widely observed on the microorganism and several diseases especially, cancers e.g., gastric, ovarian, hepatic and breast cancer as shown in Table 3 (110- 112). Four peaks (at m/z 2752, 6277, 5866 and 10093) were differentially expressed between primary breast and lung adenocarcinoma. A peak at m/z 10093 was the higher expression in only primary lung cancer than breast cancer (113). The peak at 1465.63 (m/z) was observed as a peak to discriminate between gastric cancer patients with and without lymph node (110). The specific low molecular weight protein profiles in saliva were used to discriminate healthy subject from oral lichen planus, chronic periodontitis, and oral cancer. For example, the peaks at 5592.26 and 8301.46 Da were significant increase in oral cancer. The 12964.55 and 13279.08 (m/z) peaks were overexpression in oral lichen planus (114). The low molecular weight peptides (m/z 1310, 2135, 2411, 2585, 3591, 3973, and 4299) were discriminated healthy from pancreatic cancer and chronic pancreatitis. On the other hand, 1209, 1258, 1276, 1448, 1544, 1614 and 1685 (m/z) peaks were expressed only in the healthy subject (109, 115). The previous study shows the several low molecular weight peptide spectrums were associated with the pathology condition. Valerio et al. demonstrated that the low molecular weight peptide patterns were higher expressed in pancreatic cancer in healthy subjects (115). The serum profiles using the SELDI-TOF MS were used to improve the diagnosis of breast cancer patients. The peaks at m/z 3808, 6850, 7926, 8115, 8143, 8916 and 13870 were significantly correlated with breast cancer patients. The peptides at m/z 1452, 2670, 3972, 5354, 5523, 6624 and 28268 were significantly higher in healthy serum than breast cancer serum (100, 116). In hepatocellular carcinoma research, 11 m/z peaks were constructed to the biomarkers panel. Serum peptides at m/z 1073, 1450, 1866, 3883, 4054, 4644, 5064, 5247, 5805, 6579 and 7637 were clearly different among the investigated groups: hepatocellular carcinoma, chronic hepatitis, liver cirrhosis and healthy control (117). In ovarian cancer serum, 27 peptide peaks were used as biomarker panel to discriminate between the ovarian cancer patients and healthy controls. Peaks m/z 5486, 5643, 5253, 8149, 16450, 8054, 11650, 3933, 7773, and 4560 were upregulated in ovarian cancer samples when compared with healthy, although peaks at m/z 8943, 13720, 6440, 8142, 6890, 8700,

8575, 13980, 6369, 4348 and 3260 were down-regulated in ovarian cancer (118). The m/z peaks at 7350, 8446, 10850 and 14693 were higher expression in malignant Barrett's epithelium than benign epithelium cells. This panel was developed to search malignant Barrett's epithelium biomarkers in tissues (119). Furthermore, up to 56 different peaks have been found in patients with and without multiple myeloma, and three significant peaks at m/z 8131, 11660 and 22752 were used to identify multiple myeloma and non-multiple myeloma (109). Similarly, the MALDI peptide patterns were used for discriminating among esophageal adenocarcinoma, Barrett's esophagus, high-grade dysplasia and healthy person. They found that six mass spectrums significantly differed, and the peaks at m/z 1908, 2112, 3158, 3404, 3766, and 4562 were overexpressed in patients with esophageal adenocarcinoma (120). In the range of 700-10000 m/z, the mass spectrums were different among biliary tract cancer, benign biliary disease and healthy controls. The 887, 2903 and 5803 Da were higher expressed in biliary tract cancer than healthy controls (121). The candidate marker for colorectal cancer was generated via MALDI-TOF MS. The results show 5 different peptide peaks at m/z 1895, 2020, 2080, 2656 and 3238 were significant differences between patients with colorectal cancer and healthy control. The sensitivity and specificity of this panel displayed 95.6% and 91.8%, respectively (122). For cervical cancer stratification development, the MALDI imaging MS and Pap smear were combined. Pap smears were analyzed via MALDI imaging MS. 5 peptide peaks (2012, 2172, 2339, 3441 and 4740) were significantly different between the positive and negative cytology (123). In cervical cancer, the peaks at 3974, 4175 and 5906 Da have identified the specific serum tumor marker to discriminate between healthy subjects and patients with cervical cancer (124). More researches focus on peptide fingerprints which may improve the cancer screening/ diagnosis test. However, there are minimal researches concerning serum peptide fingerprints in various stage of cervical cancer.

Table 3. The peptide signature peaks in different cancer types. The serum/plasma signature peptide patterns in the range of 1000 – 10000 m/z in each cancer type were reported using MALDI-TOF MS approach. The proposed m/z peaks may be the candidate biomarkers. The biomarker candidate was identified via MALDI-TOF MS/MS. (Adapted from Hajduk et al., 2018, (125))

Cancer type	**m/z**	**Biomarker candidates**
Stomach	1089, 1466, 1470, 1626, 5904, 5917	Fibrinopeptide A
	3217, 3316, 3952, 6431, 6630	-
	6431, 6629, 9443	Apolipoproteins C-I, Apolipoproteins C-III
Colorectal	1532, 1781, 1867, 2132, 2234, 2490, 2880	Alpha-fetoprotein, prothrombin, Isoform 2 of inter-alpha-trypsin, Inhibitor heavy chain H4, Isoform 1 of autophagy-related protein, transthyretin, Fibrinogen beta chain
	1779, 1866, 1935, 2022, 4589	Complement C3f
	7772	-
Pancreas	3884, 5959	Platelet factor 4
	2084, 2178, 2770, 2899, 3096, 8760, 8939	-
	1489, 2913, 4112	Apolipoprotein A-IV
Lung	1760, 2881, 2940 ,3884, 5248, 5773, 5851, 7172 , 7764, 8140	-
Breast	1041, 1061, 2660	Fibrinogen alpha chain, Inter-alpha-trypsin inhibitor heavy chain H4, apolipoprotein A-II
	1046	Angiotensin II
	4209, 4264	-

Table 3. (Continued)

Cancer type	**m/z**	**Biomarker candidates**
Myeloma	2900, 3316, 7763, 2661	-
Renal	1083, 1445, 6879	-
Bladder	3525, 4282, 4963, 5804, 5903	-
Ovary	1066,1082,1088,1277,1293, 4467,4963, 8602	-
Uterus	1026, 1889, 1899, 3210, 3265, 3282	Isoform 2 of fibrinogen alpha chain, Albumin, Isoform CRA, Complement C4-B-like isoform 1

OBJECTIVE

The aim of study is to characterize the serum peptide patterns at various stages of cervical cancer patients (precancerous, stage I, stage II and stage III) and healthy women by MALDI-TOF MS analysis.

CHAPTER 2

RESEARCH METHODOLOGY

Material

1. Chemicals and reagents

Alpha-cyno-4-hydroxycinnamic acid (CHCA), ProteoMass™ peptide& protein MALDI-MS calibration kit and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich, Germany. Folin-Ciocalteu phenol reagent and sodium hydroxide (NaOH) were bought from MERCK, GERMANY. Acetonitrile (ACN), copper(II) sulfate ($CuSO_4$), sodium carbonate (Na_2CO_3), sodium dodecyl sulfate (SDS) and tataric acid were obtained from RCI Labscan Ltd (Thailand), Fisher scientific (United Kingdom), Panreac Applichem(United States of America), USB corporation(United States of America), and Cario Erba (Italy).

METHODS

1. Sample collection and preparation

Blood specimens from 139 preoperative cervical cancer patients and 83 age-matched healthy volunteers without cancer and HIV infection history were obtained from the Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Songklanagarind hospital. The diagnosis of cervical cancer specimen was histopathologically confirmed by the clinician. The clinical staging is according to the criteria of the FIGO. The subjects ranged in age from 30 to 65 years old. Informed consents were obtained from all participants and the study was approved by the ethics committee of the faculty of Medicine of Prince of Songkla University. The detail characteristics of all participants and experimental sets were shown in Table 4-6. Prior to clinical operation, 5 ml of whole blood was collected from each subject who had not been fasting prior to any invasive process and allowed to clot at room temperature for 1 hour. Within 2 hours after collection, the blood samples were centrifuged at 2,500 xg for 10 minutes at 4 °C. To avoid multiple freeze-thaw cycles, all serum samples were separated into multiple aliquots and immediately stored at -80°C until use.

2. Experimental design

The methodology processes are divided into 2 phases (optimal condition and experiment phase). First, the optimal condition phase was divided for finding the suitable condition for the experimental phase. This phase consists of sample preparation using cut off and ZiptipC18 (vary ACN concentration for wash and elute fraction). The samples were obtained from 40 healthy subjects, 39 patients with precancerous, 40 patients with early stage and 29 patients with advanced stage. The serum from patients with same stage was pooled and used as shown in Table 4.

The optimal condition obtained from in the first phase was used in the second phase (experimental phase). The 222 specimens including 83 healthy controls, 51 patients with precancerous, 24 patients with stage I, 37 patients with stage II and 27 patients with stage III were performed for experimental phase. The individual sample was pooled together in the equal protein amount for reducing the bias based on clinical diagnosis. For the experiment, the serum specimens were divided into 2 sets as training (30 healthy women and 75 cervical cancer patients)

and validation set (53 healthy subjects and 64 cervical cancer patients) as shown in Table 5. Each set was then assigned to perform in a distinct day, as well as a duplicate of each set was carried out in the following week in order to validate a mass spectrum and test the reproducibility. For MS analysis, the samples at each stage were pooled and quantified using Lowry assay. The overview of experimental design is presented in Figure 7.

All individual samples as same cases with experimental phase were used to observe the peptides mass spectrums. Then, the results from dendrogram showed the peptide mass spectrums were separated into 5 groups including healthy subject, precancerous group 1 (precancerous mass spectrum which close stage I), precancerous group 2 (precancerous PMF which resembles stage III), stage I & IIA (stage II peptide spectrum which looks like stage I) and stage III& IIB (stage II peptides profile similar with stage III). The samples were pooled in the same protein amount based on dendrogram to creating database. The number of cases and their characteristics in this group was shown in Table 6. Individual cases for each pooled serum group were shown in Appendix B (Table 11-14).

Table 4. The characteristics of participants in the optimal condition phase. Specimens from healthy control (HC), precancerous, early and advanced stage were used for selection of the suitable sample preparation procedure

Sample preparation	Condition	Samples group (no. of cases)
Cut off		HC (40), Precancerous (39), Early (40) and Advance (29)
ZiptipC18	Vary ACN	HC (40)
	Fraction selection	HC (40), Precancerous (39), Early (40) and Advance (29)
	Sample per matrix ratio	HC (40), Precancerous (39), Early (40) and Advance (29)

Table 5. Characteristics of participants in the experimental phase. Specimens from both training (90 cases) and validation set (105 cases) were used.

Sample group	Pathological stage	Overall no. of cases	Training set		Validation set	
			no. of cases	Age (median/range)	no. of cases	Age (median/range)
Healthy controls	-	83	30	42/ (31-60)	53	42/ (30-59)
Cervical cancer patients	Precancerous	51	30	44/ (30-61)	21	45/ (30-61)
	Stage I	24	15	45/ (32-54)	9	48/ (32-65)
	Stage II	37	15	45/ (31-65)	22	51/ (39-61)
	Stage III	27	15	44/ (32-61)	12	52/ (43-63)

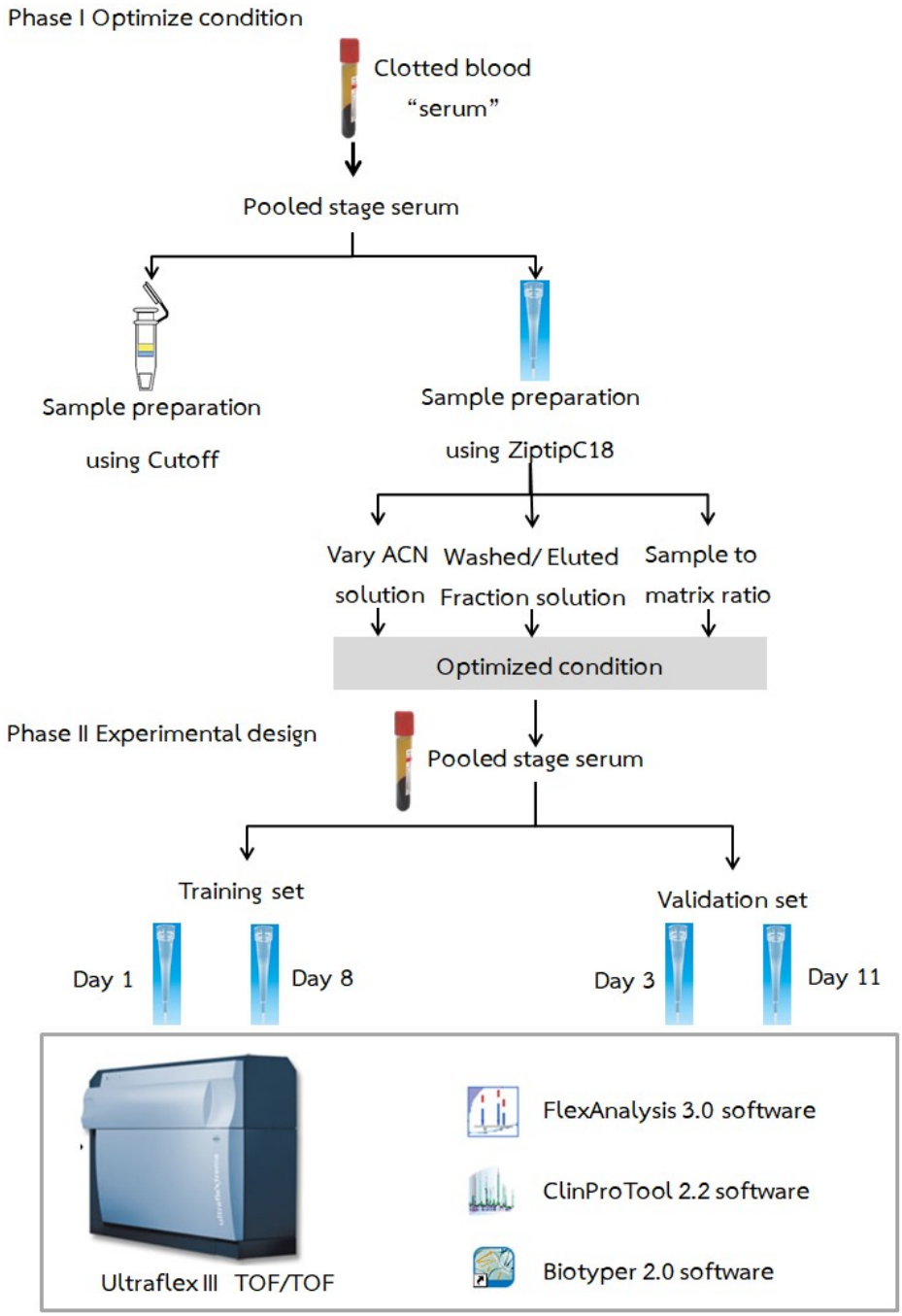


Figure 7. Experimental design of serum peptide profiles. The optimal condition and experimental design phase were performed in this study. In an optimal condition phase, pooled stage serum was used for selecting the optimal protocol (sample preparation condition). The experimental phase was generated based on the optimal protocol. The pooled serum specimens were divided into 2 sets as training and validation set. Peptides in all samples were prepared by using ZipTip C18 prior to MALDI-TOF MS analysis followed by bioinformatics tool analysis.

Table 6. Patient characteristic for creating the database.

Sample group	Charecteristic of investigated group			
	Biotyper group	Age (median/range)	Pathological stage	Histology
Healthy controls	Healthy (n=25)	39/(30-57)	-	-
Cervical cancer patients	Precancerous set 1 (n=4)	37/(35-49)	LSIL (n=3) HSIL (n=1)	Adenosquamos (n=1) SCCA (n=3)
	Precancerous set 2 (n=6)	48.5/(30-53)	LSIL (n=2) HSIL (n=4)	Adenosquamos (n=1) SCCA (n=5)
	Stage I and II set 1 (n=19)	50/(32-65)	IA (n=6) IB (n=3) IIA (n=1) IIB (n=9)	Adenocarcinoma (n=1) Adenosquamos (n=3) SCCA (n=15)
	Stage II set 2 and III (n=22)	49/(31-61)	IIA (n=1) IIB (n=12) IIA (n=0) IIIB (n=10)	Adenocarcinoma (n=1) Adenosquamos (n=2) SCCA (n=16) No data (n=4)

3. Preparation of serum peptides by cut off device

The principle of cut off device is membrane ultrafiltration which separates molecules based on size (molecular weight). The centrifugation forces solution to flow through the molecular weight membrane cut off (MWCO). As a result, the low molecular weight molecules flow through the permeable membrane, whereas the larger molecules than the pore size of membrane are still retained. In proteomic research filed, the MWCO membrane is used to concentrate the peptide/protein (125).

In this study, the 10 kDa cut off membrane filtration (Nanosep, PALL, USA) was equilibrated by sterile water prior to centrifuge at 5,000 rpm for 5 min. Then, the sterile water in cut off tube was removed. For specimen preparation, 150 µl of pooled serum was diluted with 300 µl of sterile water. The diluted serum was added on the top of 10 kDa cut off membrane filtration tubes and centrifuged at 10,000 rpm for 20 min. Then the solution was collected prior to concentrating under vacuum. The protein concentrations of samples were measured by using Lowry assay. The serum peptides were diluted with 0.1% TFA to the final concentration of 1 µg prior to MS processing and bioinformatics analysis. The cut off process was shown in Figure 8.

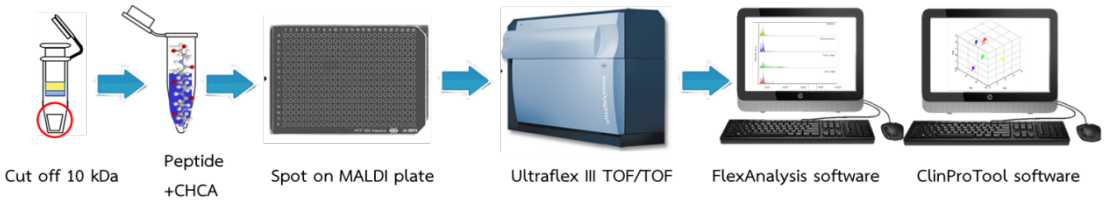


Figure 8. The peptides were prepared by using the cut off. The sample was mixed with MALDI matrix and directly applied to the MALDI steel target plate. Mass spectrums were performed using Ultraflex III TOF/TOF. All results were evaluated by using FlexAnalysis 3.0 and ClinProTool 2.2 software.

4. Preparation of serum peptides by ZipTipC18

In order to remove the polymer peaks, peptide samples were purified by ZiptipC18 reverse phase chromatography (MERCK, Germany). The C18 binds with resin acts as the stationary phase. When sample flows through the ZiptipC18, the hydrophobic peptides bind with C18 at the end of tip. Then ACN as mobile phase was used to elute the hydrophobic peptides to mass spectrometry (125). The peptide preparation was processed by ZipTipC18 was shown in Figure 9

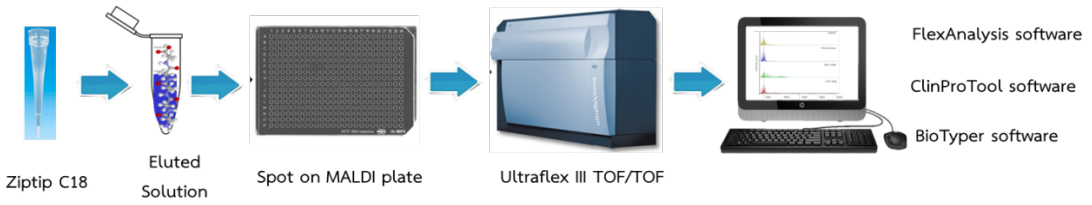


Figure 9. Serum peptides preparation using ZiptipC18 reverse phase chromatography. Then, the peptides were mixed with MALDI matrix and directly applied to the MALDI steel target plate prior to mass spectrum analysis by the combination of Ultraflex III TOF/TOF and bioinformatics tools (FlexAnalysis 3.0, ClinProTool 2.2 and Biotyper 2.0 software).

4.1 Selection of appropriated solution

In this step, the peptides from pooled samples were applied to observe the condition. Four concentrations of ACN (25%, 50%, 75% and 100% ACN) were used for eluting the peptides in ZiptipC18. First, ZipTipC18 were wetted and equilibrated by 100% ACN and 0.1% TFA, respectively. The 100 µg serum peptides were subjected onto ZipTipC18 prior to wash the unbound peptides with 20 µl of 0.1% TFA. The peptide fractions were eluted from ZiptipC18 with 20 µl of 25%, 50%, 75% and 100% ACN. Each eluted fractions were collected and dried under vacuum before resuspending in 20 µl of 0.1% TFA.

4.2 Optimal conditions of sample to matrix ratio

All peptide from eluted fractions were tested for selecting appropriated sample to matrix ratio. The ratios yielding maximum number of MS signal were used for the experimental phase.

5 sets of 100 µl serum peptides were eluted with 100%ACN prior to mix with CHCA matrix solution at the ratio (v/v) of 1:1, 1:2, 1:3, 1:4 and 1:5, respectively. Then, the mixtures were directly spotted onto MALDI plate and dried at room temperature.

4.3 Experiment phase in both training and validation datasets

Both pooled serum based on staging (healthy subjects and patients with precancerous, stage I, II and III) and individual serum sample were analyzed according to the optimal procedure. According to the optimized protocol shown in 4.1, the tip was pre-wetted and equilibrated with 100% ACN and 0.1% TFA, respectively. The samples were diluted to 100 µg peptides with 0.1% TFA and loaded onto the tip. The tip was washed to remove the unbound peptide with 0.1% TFA. Then, sample fractions were eluted with 20 µl of 100% ACN. The eluted solutions from all samples were mixed with CHCA matrix at the ratio of 1:5 prior to spotting on the MALDI plate.

5. MALDI TOF preparation

The peptides were mixed with MALDI matrix consisting of 10 mg/ml of CHCA in 50% ACN containing 0.1%TFA (sample to matrix ratio; 1:5), and directly applied onto the MALDI steel target plate (MTP384 ground steel plate, Bruker, Germany). The replicates were used in turn to prevent sample preparation bias in pooled sample. In order to eliminate the difference in the matrix crystallization, all preparation steps were performed in a precise air-conditioned room. After air drying, the mass spectral analysis was performed using Ultraflex III TOF/TOF (Bruker, Germany) in a linear positive mode with a mass range of 1000- 10000 Da. Laser shots 1500 shots/spot. Before the MS analysis, external calibration was performed using standard peptide and protein mixtures (ProteoMass™ peptide& protein MALDI-MS calibration kit; Sigma Alderich, United States), which consist of human Angiotensin II (m/z 1046), $P_{14}R$ (m/z 1533), human ACTH fragment 18-39 (m/z 2465), bovine insulin

oxidized B chain (m/z 3465) and bovine insulin (m/z 5731). Both the training and validation sets used the same processing steps of MALDI peptide preparation.

6. MS Data analysis

All raw MALDI spectrums were operated to spectra processing, incorporating MALDI spectra detection, smoothing, baseline peak subtraction and internal spectral alignment by using FlexAnalysis 3.0 (Bruker Deltonik, GmbH, Germany). ClinProtTools 2.2 software (Bruker Deltonik, GmbH, Germany) was used to determine total average spectrum, principal component analysis (PCA), virtual gel view. The significant peptides were distinguished by Genetic algorithm (GA), different average, ANOVA, supervised neural network (SNN), and wilcoxon/kruskal-wallis test. Two-Tailed ($P<0.001$) was analyzed significantly different among investigate groups. The 8 replicated individual patterns in gel view were examined for observing the pattern. Then the dendrogram from individual samples was used as the criteria for pooled serum prior to generating the database. The reference mass spectrum of pooled stage sample was subjected into BioTyper 2.0 software (Bruker Deltonik, GmbH, Germany). The mass spectrum from individual sample was compared to the peptide patterns in the reference database. Bruker outliner criteria were used in this analysis (Bruker Deltonik, GmbH, Germany). Each peptide pattern which matched with the database was exhibited in the log score ranging from 0- 3. A score between 1.9-3.0 indicate PMF from individual case quite similar to the top database. The numbers of matched cases in each database have calculated for the sensitivity and specificity by using 2 by 2 tables as shown in Table 7.

Table 7. Sensitivity and specificity calculation.

		Individual sample		Number
		Disease	Non disease	
Database	Positive	A True positive	B False positive	Match database
	Negative	C False negative	D True negative	Unmatch database
		Disease cases	Non-disease cases	Total number

$$\% \text{ Sensitivity} = \frac{A}{(A+C)} * 100$$

$$\% \text{ Specificity} = \frac{D}{(B+D)} * 100$$

CHAPTER 3

RESULTS

1 Results from optimizing condition phase

1.1 Mass spectrums from ultrafiltration preparation (cut off)

For low molecular weight peptide collection, ultrafiltration or cut off was used. Individual specimens were pooled together based on clinical diagnosis. Four pooled serum groups (healthy, precancerous, early and advanced stage) were examined to find the optimal condition.

Pooled serum samples from ultrafiltration preparation generated the results in 3 views including total average spectrum, gel view and PCA. The total average spectra in 1000-10000 m/z were shown in Figure 10. The x and y-axis in chromatogram represented the mass per charge value and the intensity, respectively. Each group had the own signature peptide pattern. Nonetheless, the repetitive peaks were found in the range of 1000- 5000 Da of all samples.

To observe the consistency result with total average spectrum, the profiles in the pseudo-gel view were displayed in Figure 11. The linear grayscale for intensity display was used for the spectra of four model generation group's classification. The peak intensity in arbitrary units (a.u.) shows in the right y-axis. The high intensity represents in a darker line when compared with the low intensity of peak. The left y-axis exhibits the replication number of the spot. The x-axis represents the m/z value. The 24 replicated mass spectrums were arranged into 1 box, consequence same m/z in all spectrums were exhibited the line parallel with the y-axis. The continuous lines from no. 1 to 24 represent the repeatable mass spectra. In this figure, the continuous line was occurred in all groups of patients. Moreover, the peptide patterns in gel view were consistency with the chromatogram which showed the repetitive peak in the range 1000-10000 Da.

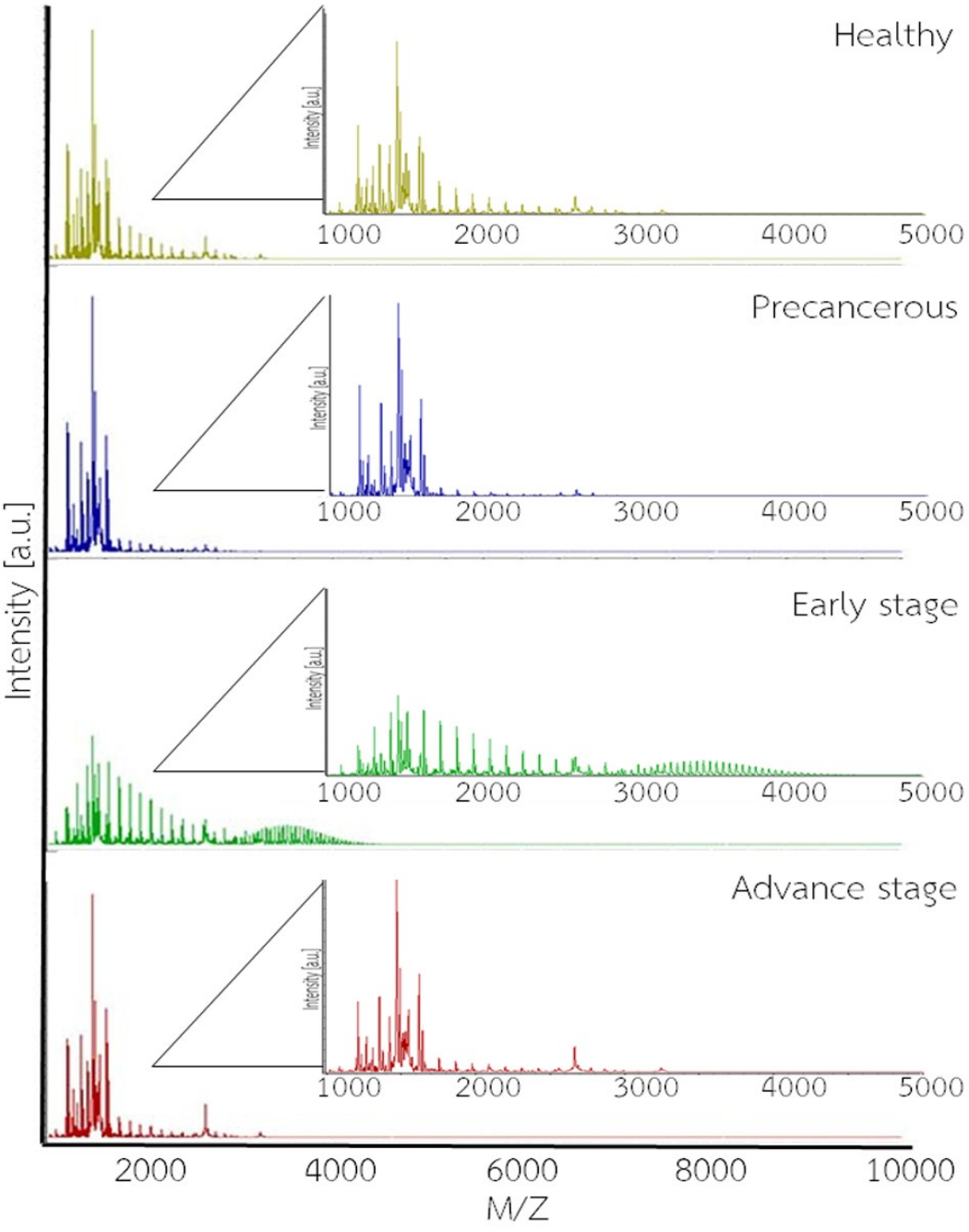


Figure 10. Mass spectrums of pooled sera prepared by cut off. The total average spectrums in range 1000-10000 Da were performed. The x and Y -axis represents the m/z value and the intensity, respectively. The repetitive mass spectrums in range of 1000-5000 were showed in all sample groups.

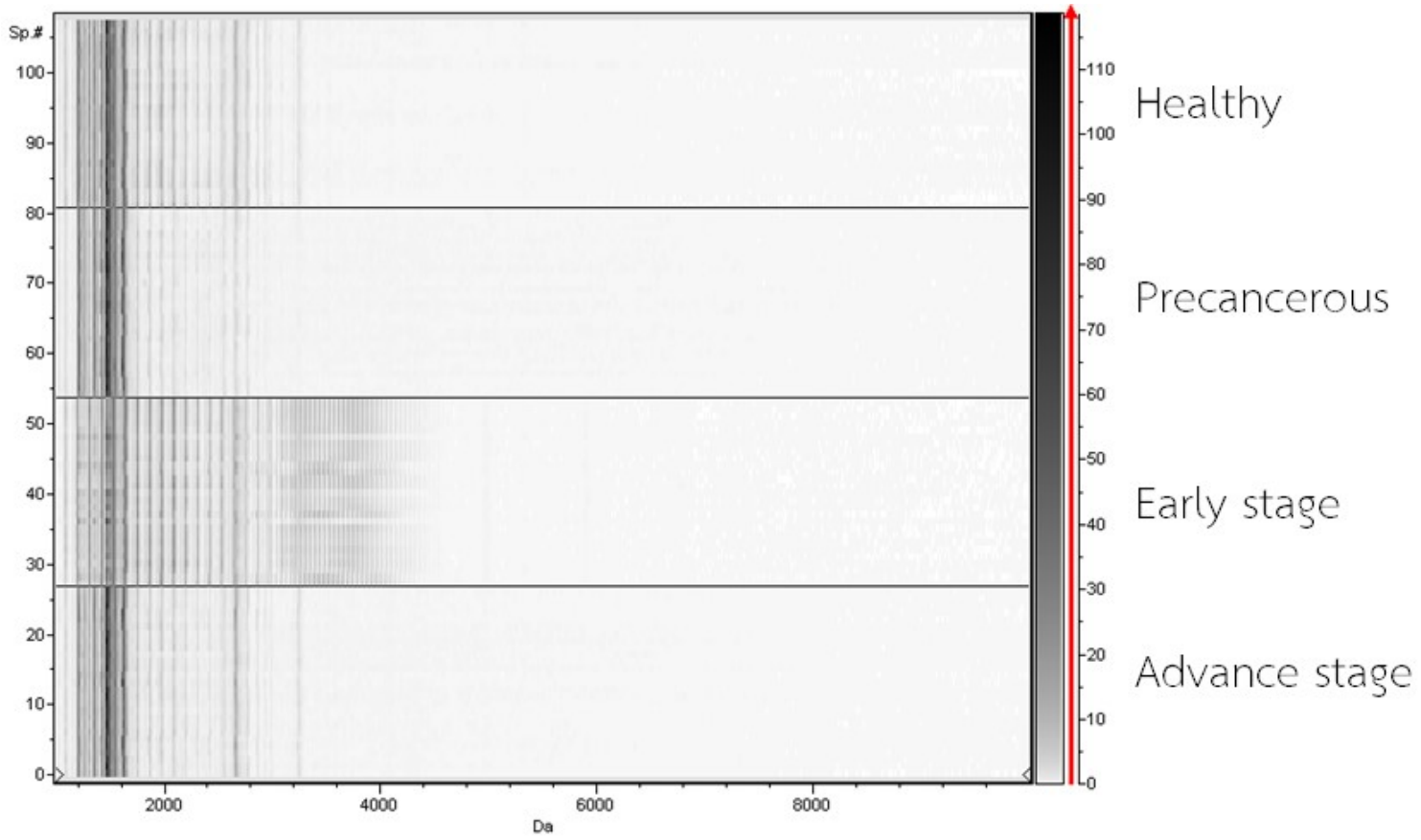


Figure 11. Pseudo-gel of pooled serum prepared by cut off. The four model generation classes (healthy, precancerous, early and advanced stage) were classified using a linear grayscale for intensity display. The x-axis, right and left axis exhibits the m/z value, peak intensity level and number of spots.

The distinguish peptide profiles in terms of total peaks and intensity were generated in PCA. This view is an extensively utilized mathematical technique created to demonstrate the variance within a data. In the event of mass profile, the variables are illustrated by the intensity at defined masses. Each mass spectrum from 1 spot generated 1 sphere. The distant of spot shows the resemblance when sphere close together which explained this peptide profile was similar in both intensity and mass. In this result, yellow, blue, green and red sphere exhibited the mass spectrums from healthy, patients with precancerous, stage I, II and III, respectively. Regrettably, spheres from all groups were mixed up; therefore the pooled stage represent sample after ultrafiltration could not completely separate distribution among groups as shown in Figure 12. This problem may be caused by repetitive peaks.

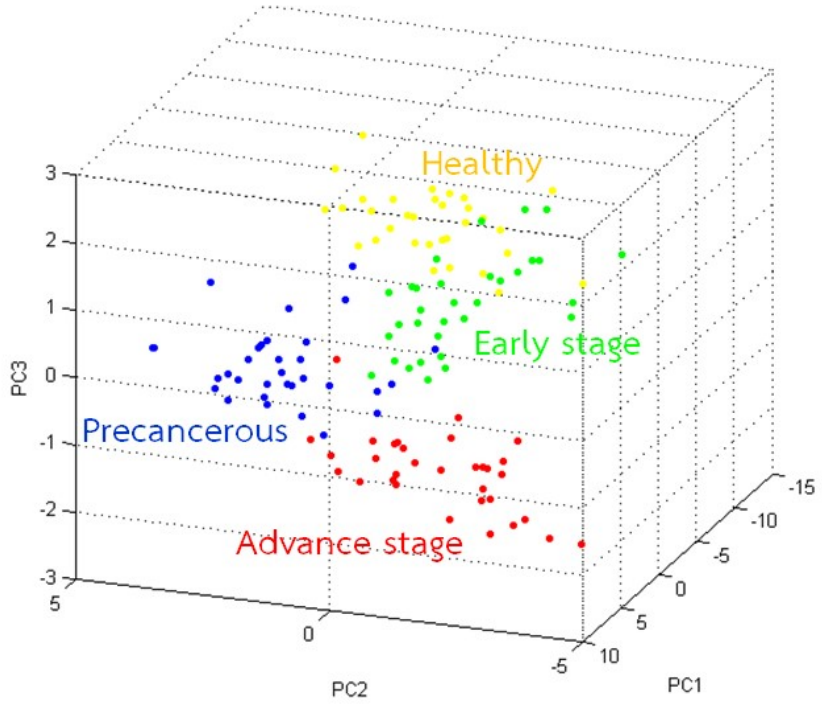


Figure 12. PCA plot for the pooled samples prepared by cut off. Healthy group was shown in yellow sphere. Cervical cancer patients with precancerous, early stage and advanced stage were exhibited in blue, green and red sphere, respectively.

The 111 Da repeated mass spectrums occurred in all investigated group. These repeat units were presented in MALDI spectrums expanded region between m/z 1000 and 3000. The only mass spectrum from early stage cervical cancer occurred 2 repeat unit sets which were mass spectrums with 111 and 43

repeat units as shown in the Figure 13. The mass spectrum with 43 repeat units was shown in the range 3000-5000 Da. The previous study showed the repeated mass distributions were commonly found in the polymer research filed. The repeated units in mass spectrum are the monomer of polymer. Normally, polymer was applied for medical application such as the containers, syringe. The degradation mode of polymer resulted in the monomers which were detected via MALDI-TOF MS (126, 127).

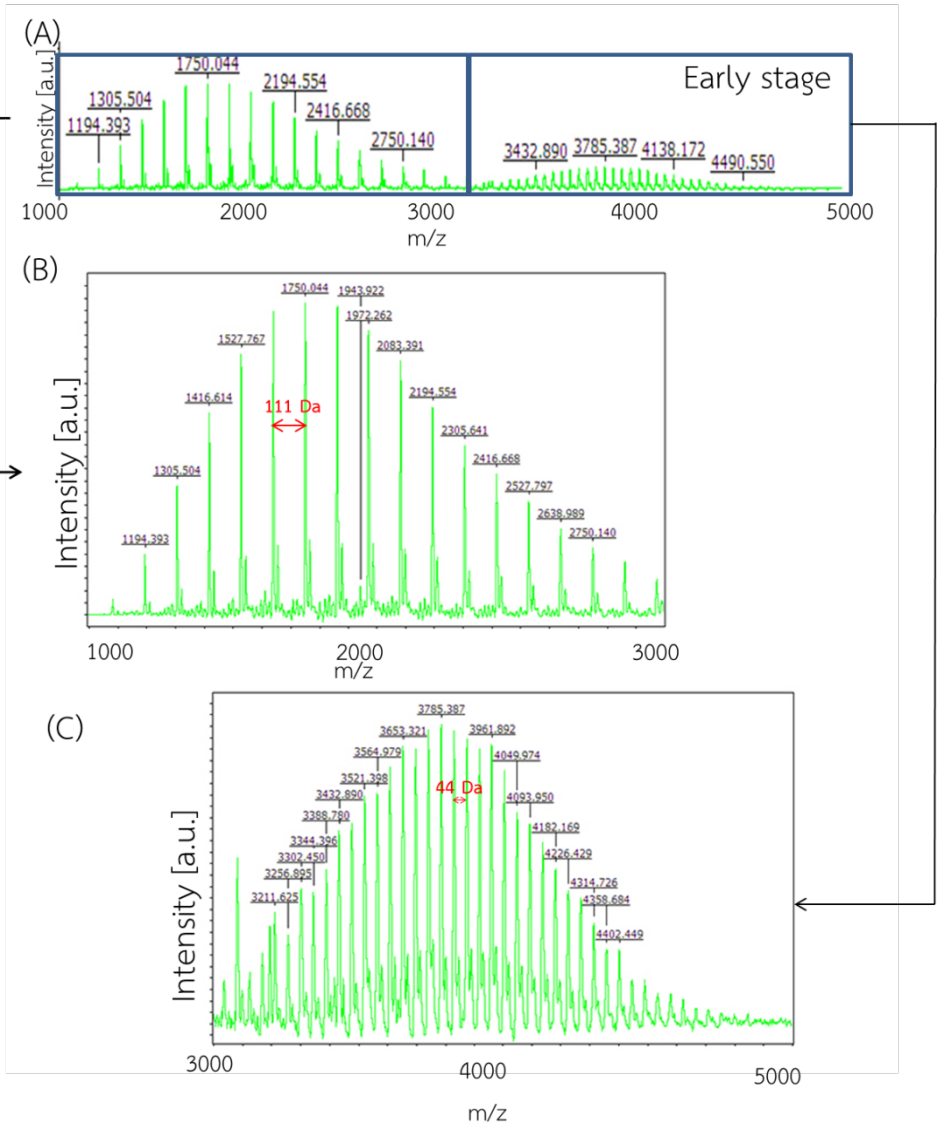


Figure 13. MALDI-TOF peptide fingerprints in the m/z range 1000 – 5000 Da. (A) The mass spectrums in early stage represent both peptides and polymer peaks. (B, C) The mass spectrum with 111 and 44 Da repeat units occurred in the range of 1000- 3000 and 3001- 5000 Da, respectively.

1.2 Mass spectrums from ZiptipC18 reverse phase chromatography

The polymer mass spectrums were removed by using reverse phase chromatography. Pooled serums of healthy, precancerous, early and advanced stage were used as the investigated group for optimizing the condition of peptide patterns.

1.2.1 Mass spectrums comparison among the different concentration of ACN

The individual samples from healthy subject group were pooled prior to cut off and ZiptipC18 comparison. The peptide profiles from ZiptipC18 showed the decrease of repetitive pattern when compared with mass spectrums from cut off; therefore, ZiptipC18 was selected for peptide preparation.

The various concentrations of ACN were shown in the range of 1000-5000 m/z (Figure 14). The mass spectrums of the eluted solution in 25%, 50% and 75% ACN exhibited both peptide signal and the repetitive polymer peaks. Whereas, the peptide pattern from 100% ACN displayed only the peptide pattern; therefore the 100% ACN is the best concentration for sample preparation.

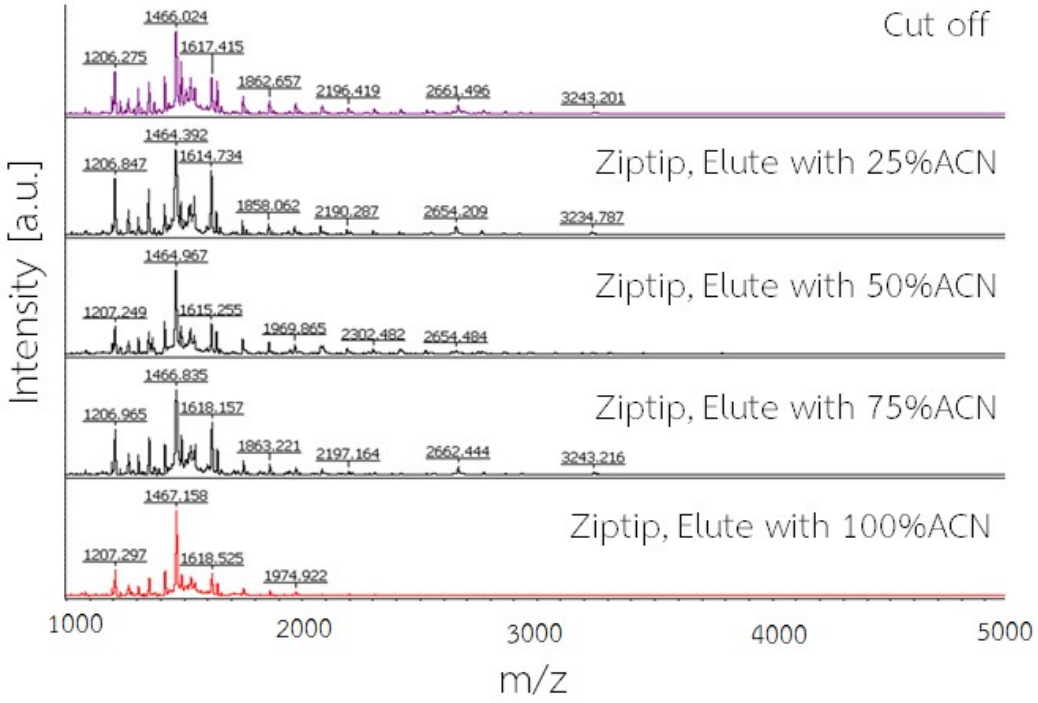


Figure 14. Comparison of mass spectrums of pooled healthy subjects by cut off and Ziptip C18 preparation method. The eluted solution at 25%, 50%, 75% and 100% ACN were performed for the trial condition of ZiptipC18. The mass spectrum from 100% ACN eluted solution showed an absence of the repetitive peaks.

1.2.2 Comparison of mass spectrums between washed and eluted solution

After ZiptipC18 processing, the hydrophilic and hydrophobic peptides were collected in washed and eluted solution, respectively. For selecting the solution, the peptide patterns in the range of 1000- 10000 m/z from washed and eluted solution were compared via intensity and number of peaks (signal to noise more than 5). In washed fraction, peptide patterns had a low both intensity (lower 3500 a.u.) and total peaks (Figure 15A). On the other hand, the peptide patterns in the eluted fraction generated many peaks with high intensity (more than 1.5×10^4 a.u.) as shown in Figure 15B. Moreover, repetitive peaks were decreased. As a result, the peptides mostly presented in the eluted solution resulted in the eluted fraction were used in the experimental phase.

(A)

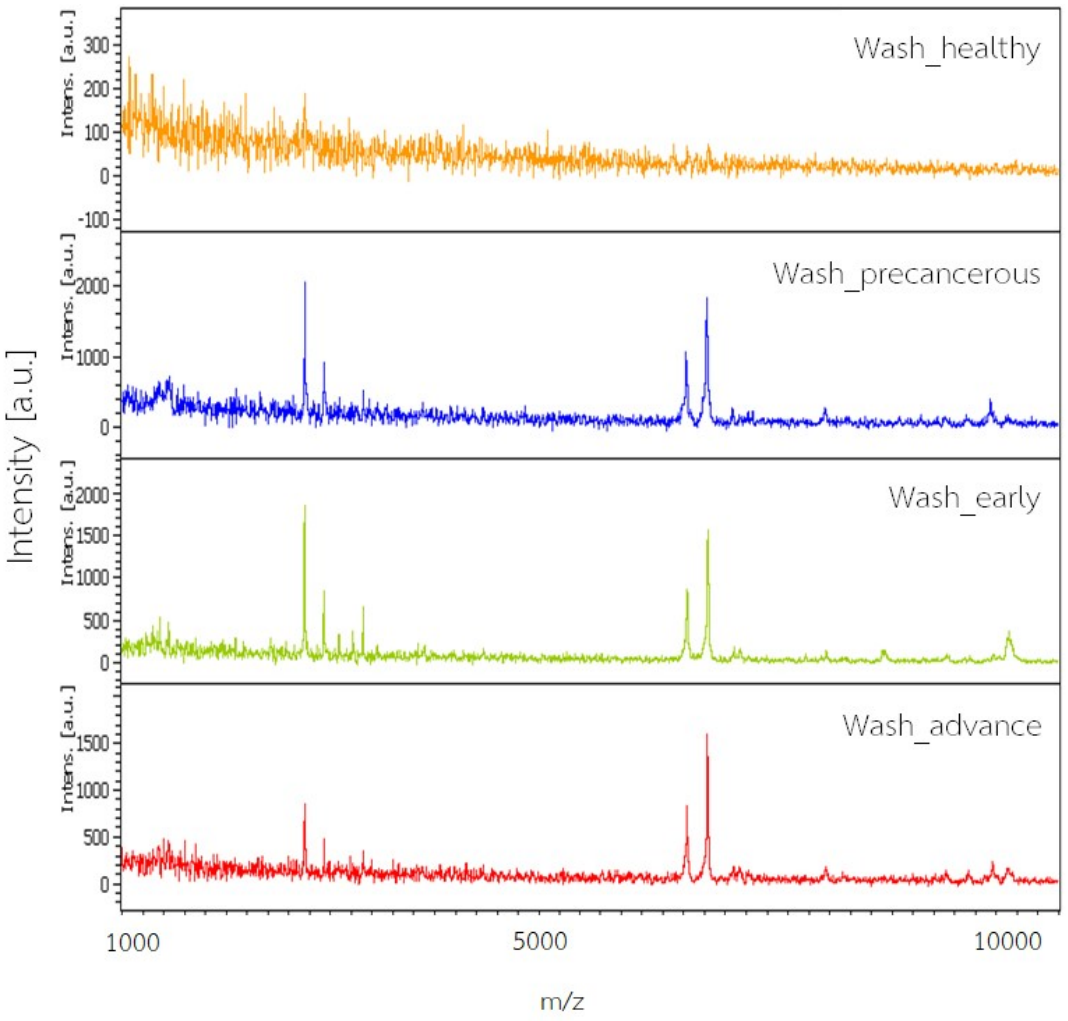


Figure 15. The MALDI-TOF profiles comparison of the optimal condition in sample preparation. Both of washed and eluted fractions were performed by using pooled sample from healthy women, patients with precancerous, early and advanced stage of cervical cancer. The mass spectrums from the washed fraction showed lower intensity and number of peaks (A). Nevertheless, the spectrums from the eluted fraction were high intensity and number of peaks (B).

(B)

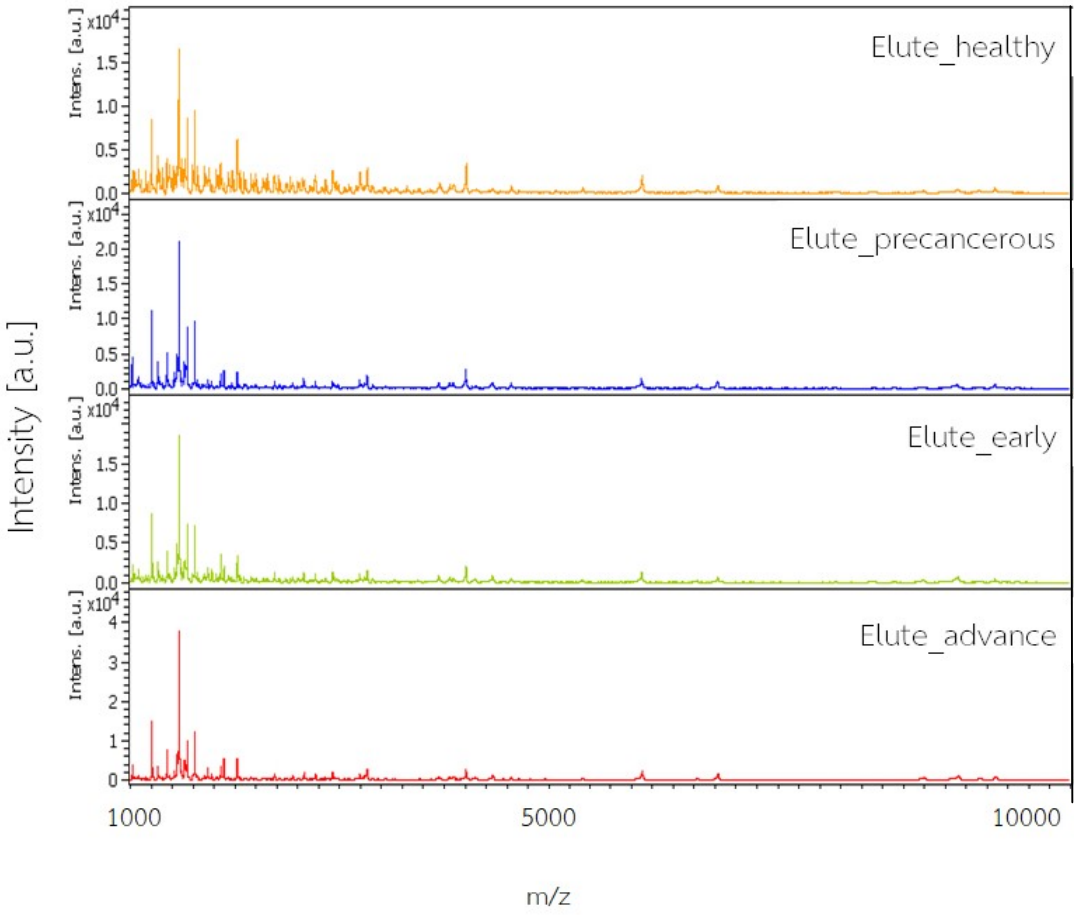


Figure 15. (Continued)

1.2.3 Comparison of spectrums derived from different ratios of sample to matrix ratio

The sample to matrix ratio is a factor which influences the ionization step. The appropriate ratio results in mass spectrum with high number of peaks.

The eluted fractions from all of pooled serum (healthy, precancerous, early and advanced stage) were used to optimize the sample to matrix ratio. The mass spectrums obtained from matrix ratio of 1:1 to 1:5 were shown in Figure 16. The total number peaks in each fraction were listed in Table 8. The high number of peaks in healthy subjects was ratio 1:1. The maximum number of peaks in all group except healthy subject was ratio 1:5. Therefore, the appropriate sample to matrix ratio for the experimental phase is 1:5 (v/v).

Table 8. The optimization of sample to matrix ratio. The sample to matrix ratio was varied from 1:1 to 1:5 which compared in term of total peaks.

	Total number of peaks			
Matrix ratio	Healthy	Precancerous	Early	Advance
1:1	29	19	23	26
1:2	27	19	31	35
1:3	20	19	21	26
1:4	18	18	18	32
1:5	22	35	32	36
Maximum	29	35	32	36

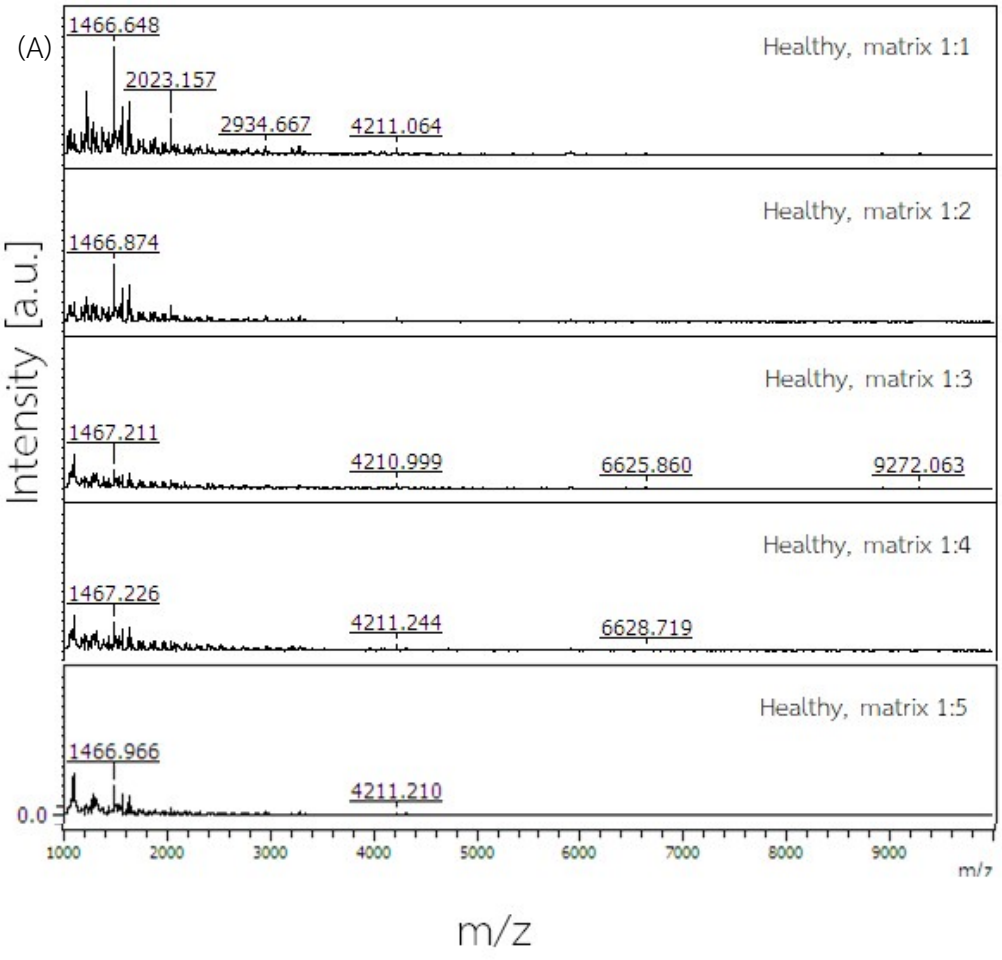


Figure 16. The mass spectrums from optimization of sample to matrix ratio. All of pooled sample groups; healthy (A), precancerous (B), early (c) and advance (D) stage group showed the mass range 1000-5000 Da. Mass spectrums of samples to matrix ratio from 1: 1 to 1:5 were shown.

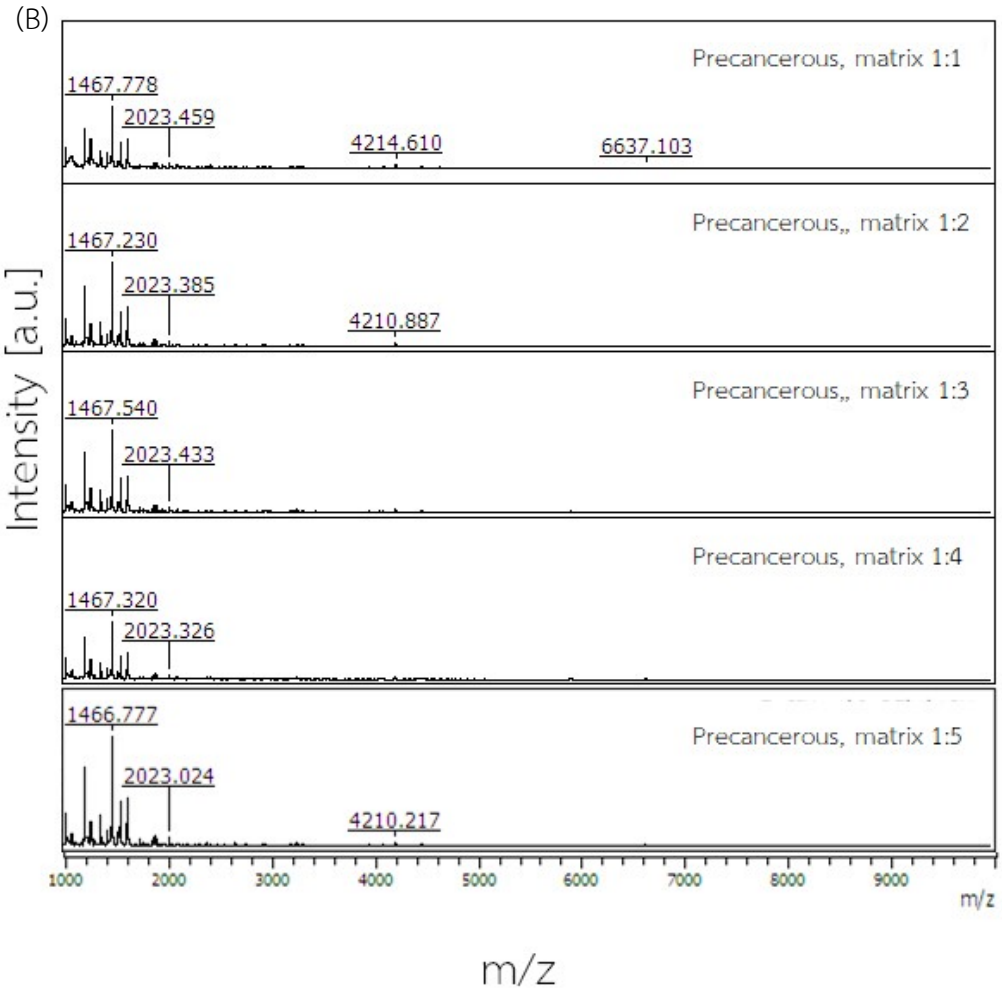


Figure 16. (Continued)

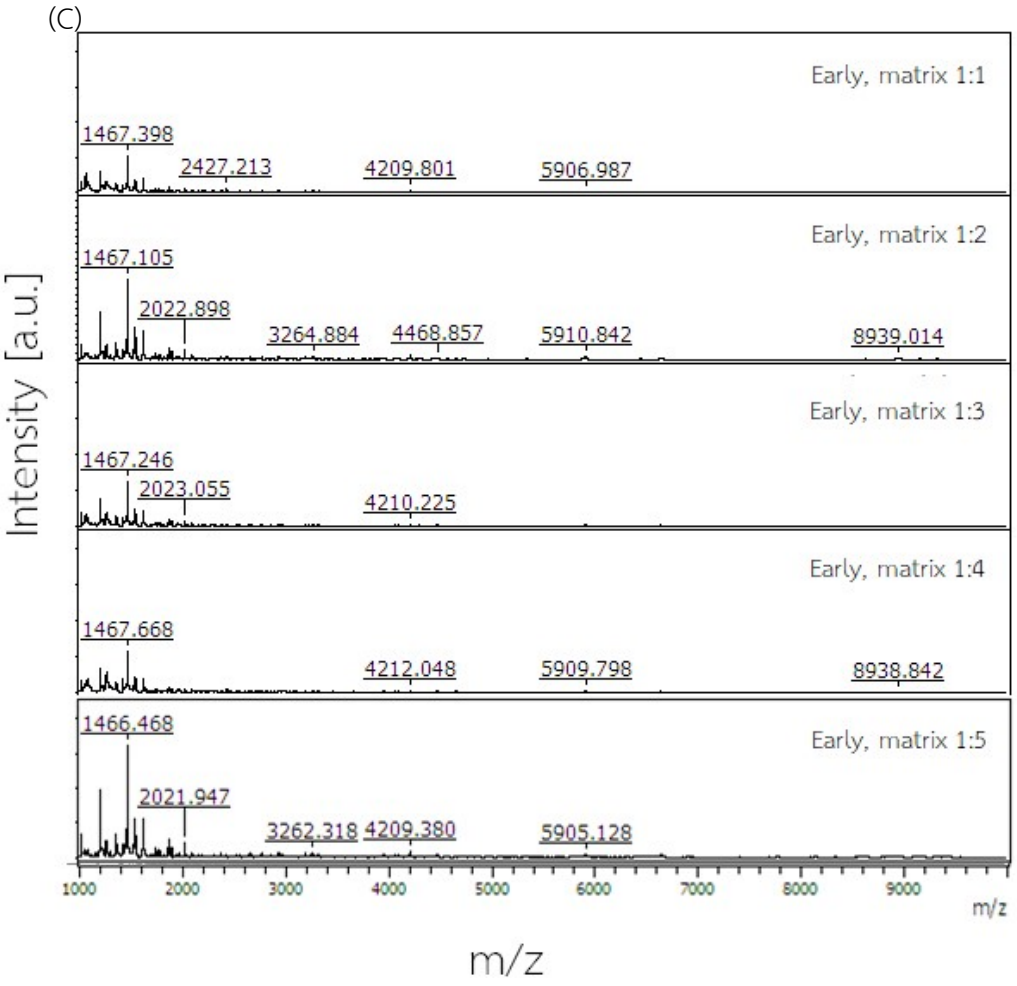


Figure 16. (Continued)

(D)

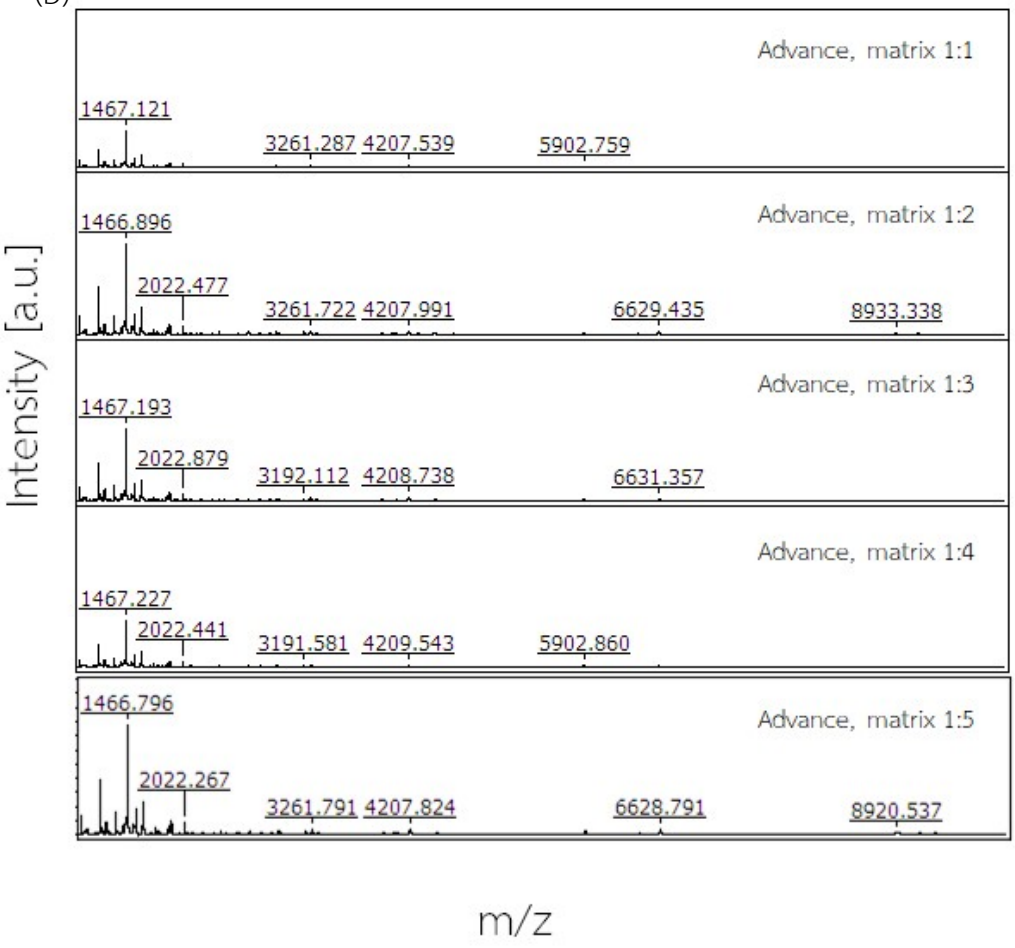


Figure 16. (Continued)

2. Investigation of peptide patterns in both training and validation set

2.1 Peptide profiles of pooled serum

The 2 independence data sets (training and validation sets) were performed in the experimental phase. The validation dataset was used to confirm the results from training set. As a result, the peptide mass spectrum of 24 spots of each sample groups was shown results in form of chromatogram, gel view, PCA, signature and significant peaks. The total average spectrums in the range of 1000-10000 m/z were generated from pooled sample among training set, validation set, and their duplication. As a result, each sample group had its own signature peptide patterns. According to the independence dataset, the peptide profiles between training and validation set had some different m/z peaks (Figure 17). The average total numbers of peaks in each group of training set were 52 (healthy subjects), 49 (precancerous), 65 (stage I), 66 (stage II) and 73 (stage III) as shown in Figure 17A. In the validation set, the average total number of peak from healthy subject, patients with precancerous, stage I, II and III were 52, 59, 63, 53 and 71, respectively. The result in gel view form was shown in Figure 18. Gel view in each dataset and its duplicate showed continuous y-axis line which means the result is the high accuracy. The representative pooled sample group was presented as PCA. The color in PCA exhibited the sample group; healthy (purple), yellow (patients with precancerous stage), blue (patients with stage I), green (patients with stage II) and red sphere (patients with stage III). The similar color was closely and the different group was long distant. Consequently, the PCA model showed the discrimination of the peptide patterns that were found not only in cervical cancer and healthy group but also among different cancer stages as shown in Figure 19.

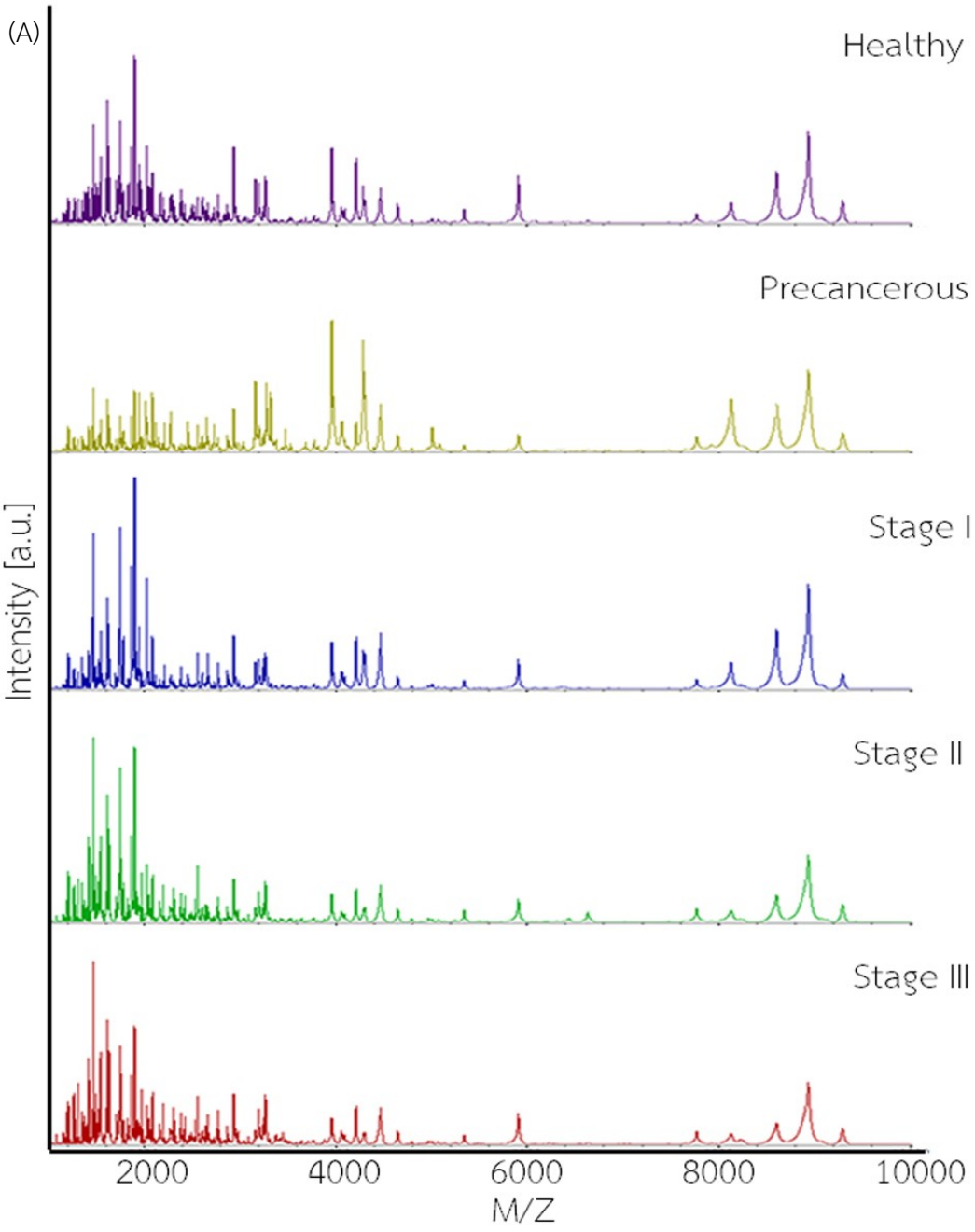


Figure 17. MALDI spectrum of pooled serum sample obtained from healthy women and various stages of cervical cancer in the range of 1000- 10000 m/z. The total average spectrums were the representative patterns in experimental phase which acquired from the training (A) and validation datasets (B).

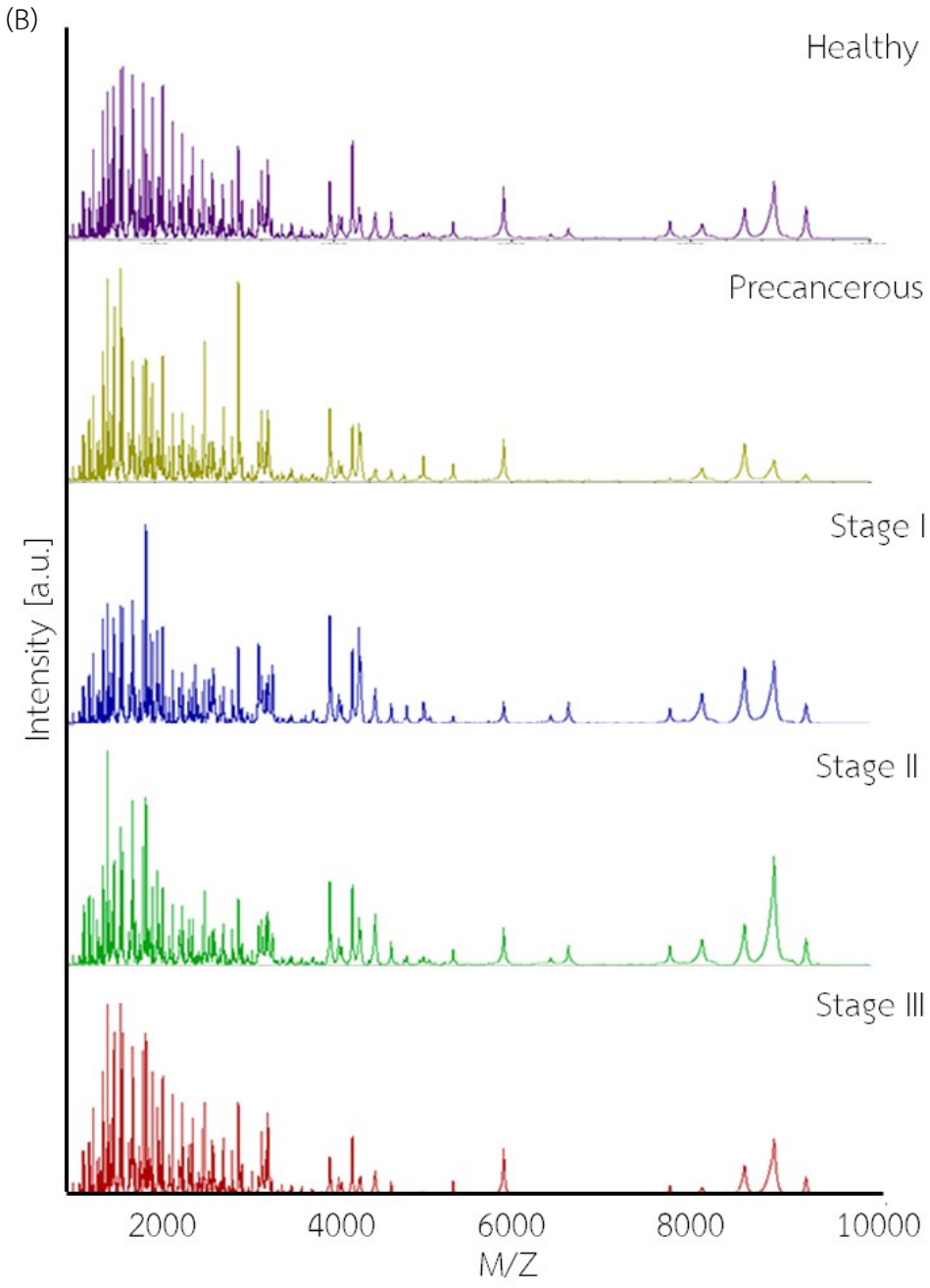


Figure 17. (Continued)

(A)

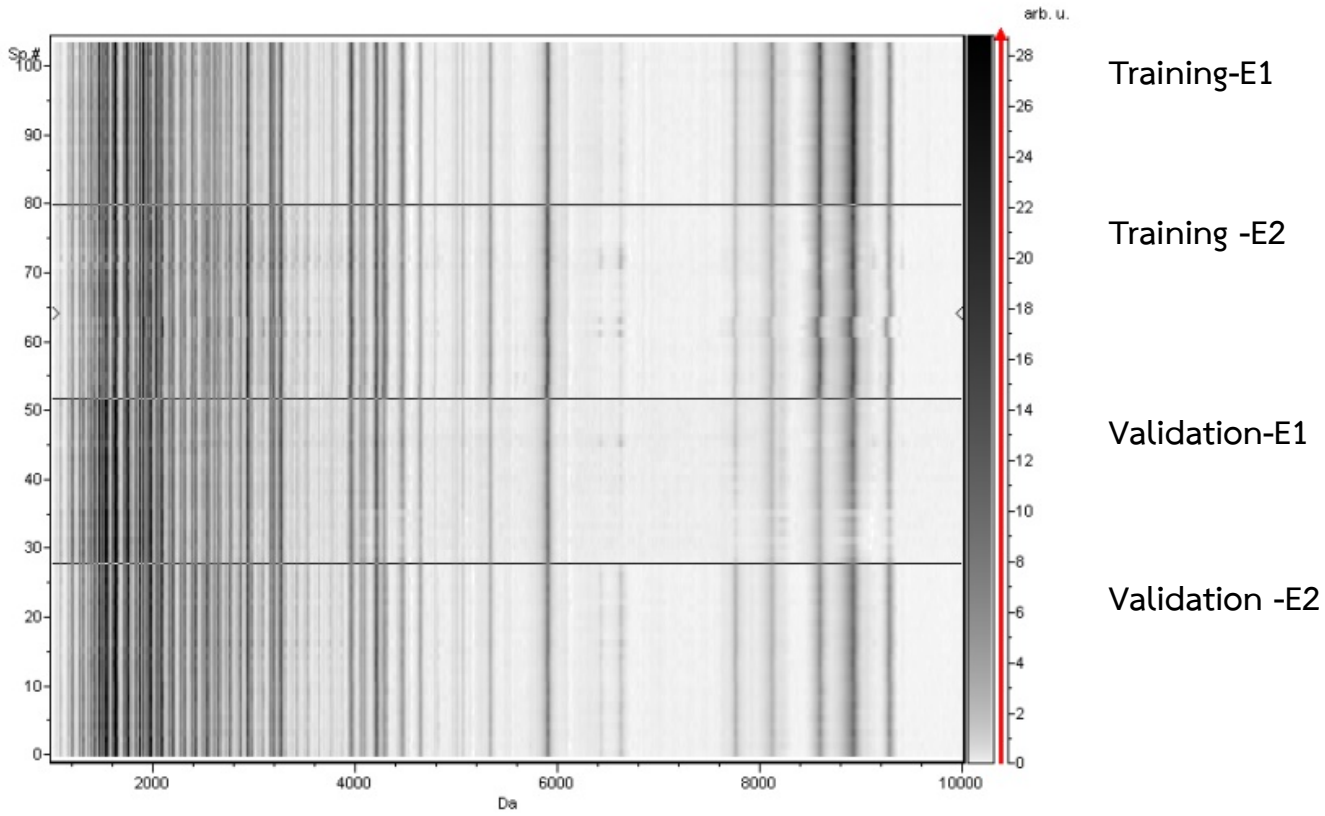


Figure 18. The gel view of pooled sample groups from training, validation set and its duplicated. Pseudo-gel of pooled serum sample from healthy women (A) and patients with various stages of cervical cancer (precancerous (B), stage I (C), stage II (D), and stage III (E)) in the range of 1000- 10000 m/z generated by ClinProTool2.2 software.

(B)

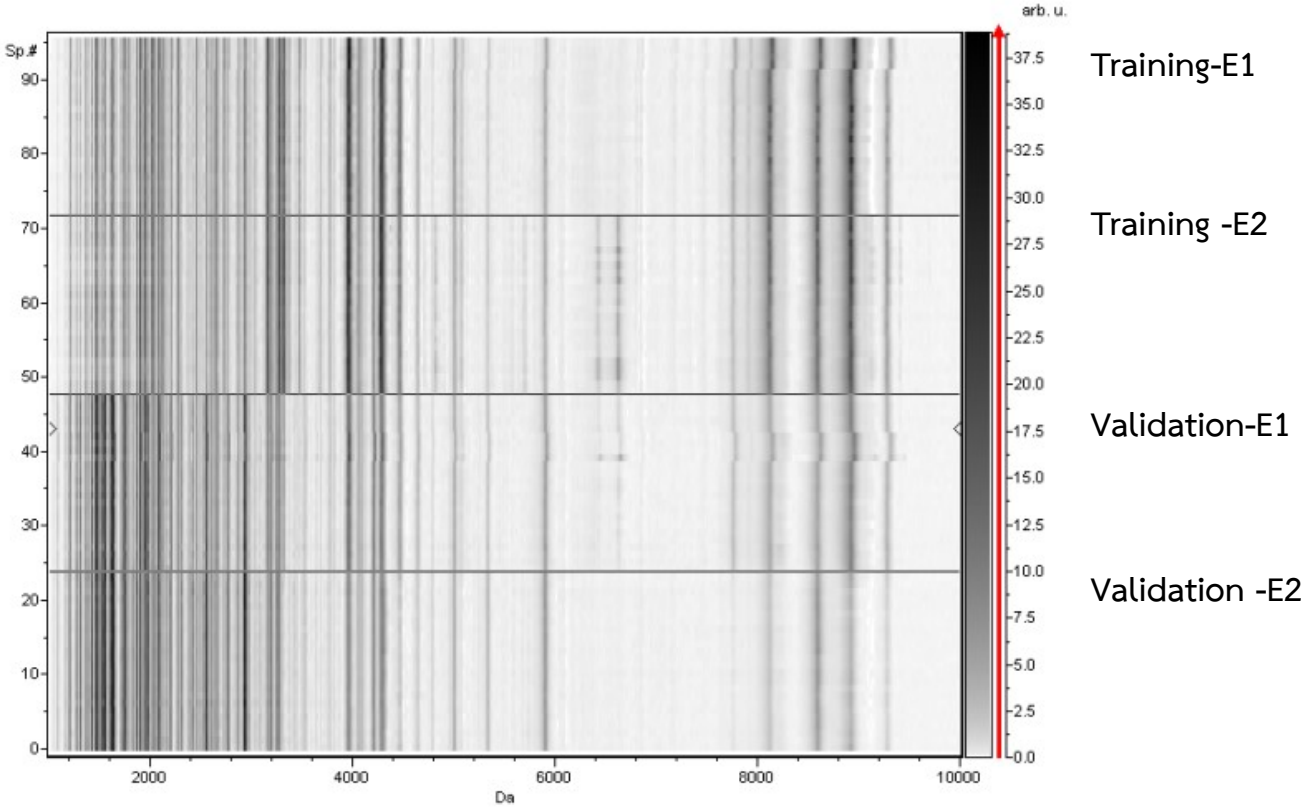


Figure 18. Continued

(C)

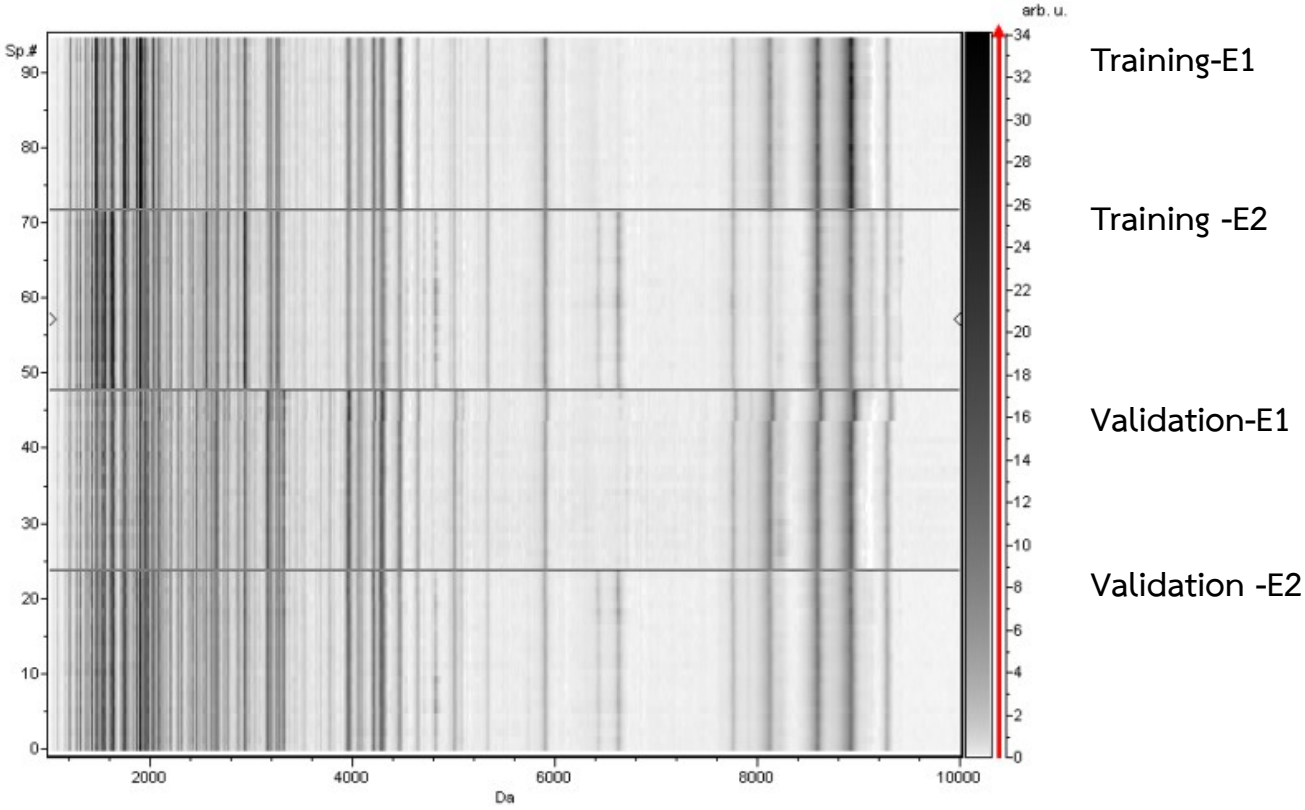


Figure 18. Continued

(D)

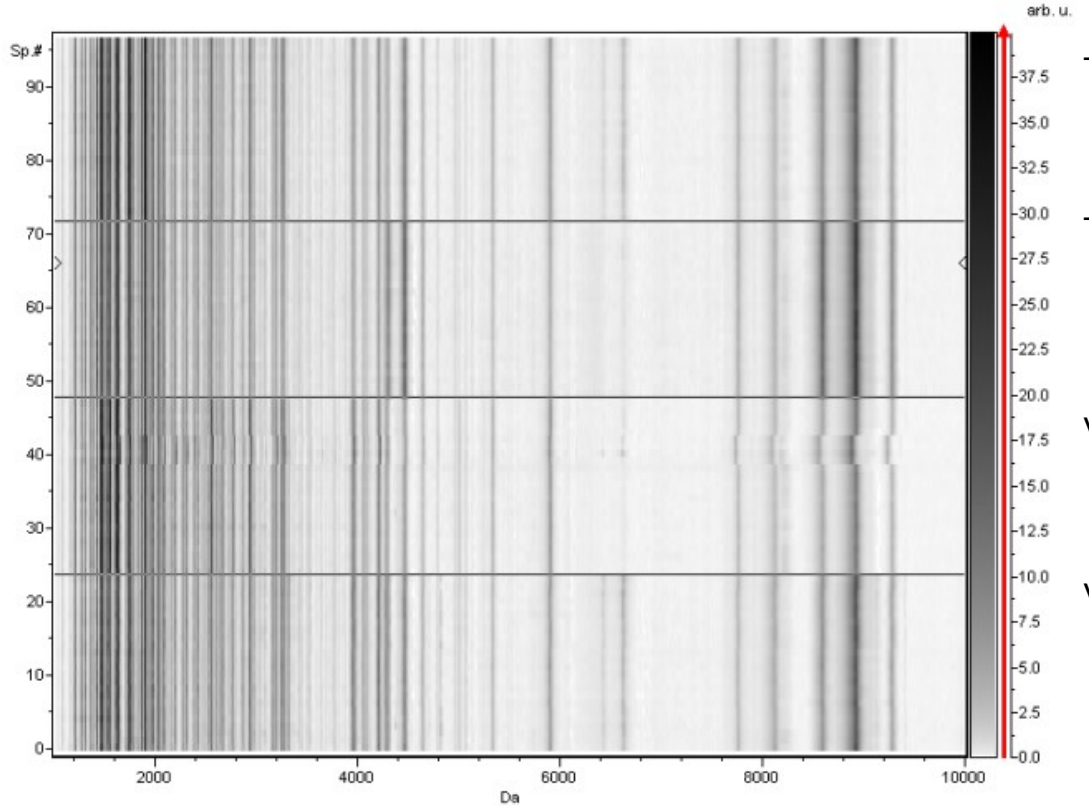


Figure 18. Continued

(E)

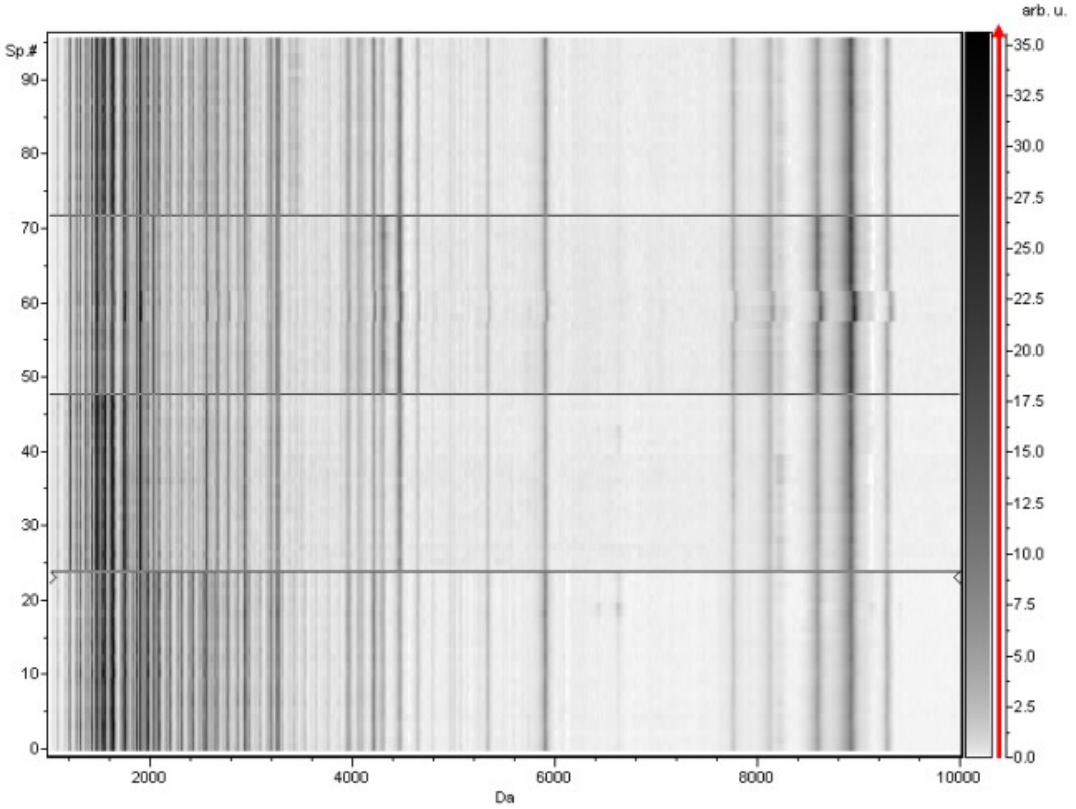


Figure 18. Continued

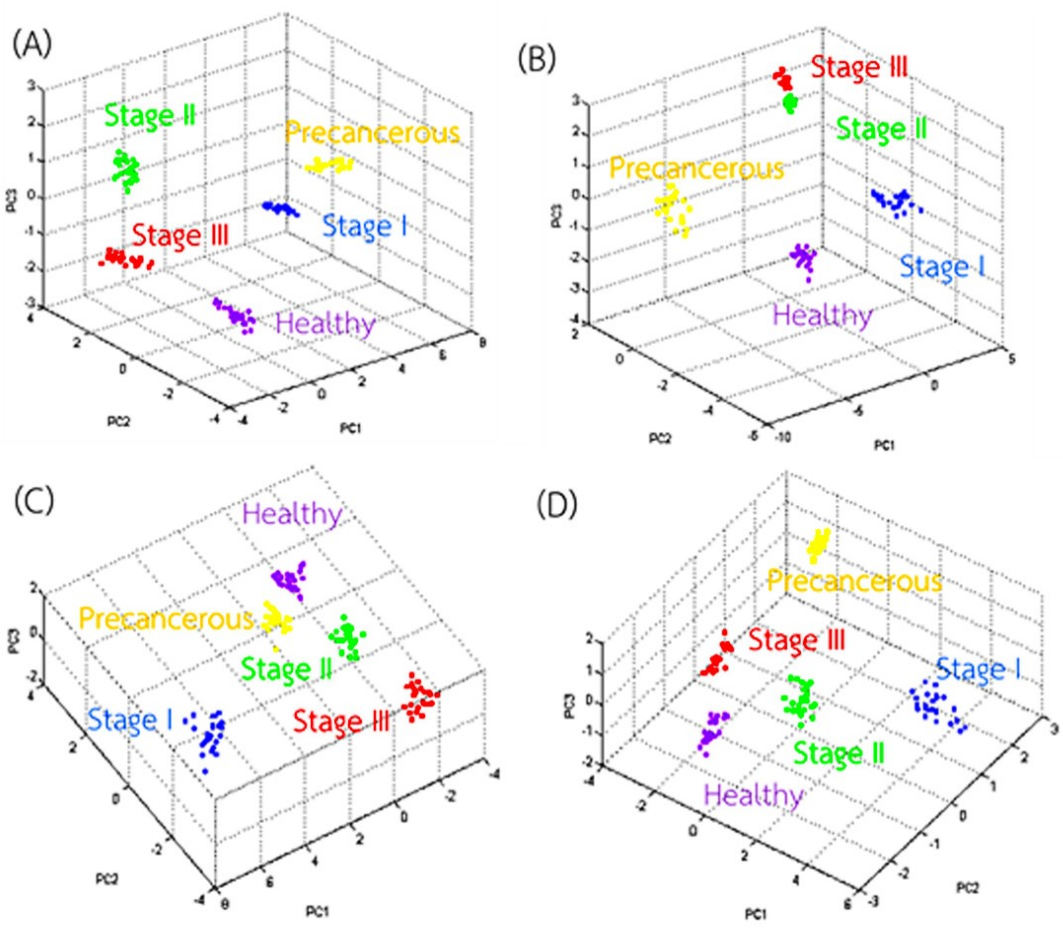


Figure 19. PCA of pooled sample prepared by using ZiptipC18 reverse phase chromatography. The data were generated from healthy women and cervical cancer patients (precancerous, stage I, II and III). The 24 replications (24 spheres) per group were performed to observe the precision. The purple, yellow, blue, green and red spheres represent the mass spectrums from healthy, precancerous, stage I, II and III, subsequently. The same colors of spot were closely and the different groups were far away. Cluster analysis exhibited separately distribution with healthy, precancerous and cervical cancer (stage I, II and III). (A) and (B) showed results from pooled sample in training data set and it's duplicated. (C) and (D) represented results from pooled sample in validation data set and it's duplicated.

Differences in the peptide mass profiles among examined groups were analyzed by multivariate and univariate algorithm (different average, ANOVA t-test and Wilcoxon/ Kruskal-Wallis test) as shown in Table 9. The different peaks were in the range between start and end mass. The averages of different mass peak were shown in column mass in the table. The weight shows the value from the algorithm calculation. In the Table 9, the group with high intensity of peaks represent in the dominant group. P<0.001 was considered statistically significant. The multivariate algorithms in this study were genetic algorithm (GA) and supervised neural network (SNN). First, the genetic algorithm is model which classified all peaks via evolutionary characteristics. The distances between all spectra were calculated with the k-nearest neighbor. The class membership of the neighboring point was used to categorize each mass spectrum (128). As this result, the average molecular weight of 1488, 1741, 2307 and 3242 Da were high intensity in stage III, I, II and II, respectively. SNN algorithm is another multivariate classification model which categorized base on proteotype. The optimal distribution of spectra was generated prior to use matrix of data space which is relevant for the classification (129). The mass signals of m/z 1898 and 2044 were increased in stage I and healthy subject (SNN, p-value <0.001). Quick classifier algorithm is the univariate classification model. The peak areas among group were compared with the statistic (p-value). Difference Average compares the difference between the maximal and minimal average area of mass spectrum (130). In case of this study, the peak at 1466.91 Da was high intensity in cervical cancer patients with stage III. ANOVA is a statistic in univariate model for the data expecting normal distribution (130,131). The result show that the average molecular weight of 1898, 3159 and 4299 were significantly high intensity in stage I, precancerous and stage III (ANOVA, p<0.001). The Wilcoxon and Kruskal-Wallis Test are the non-parameter method which does not suppose a normal population (132). In this study, the difference in the mass spectrums among the five investigated groups (healthy, precancerous, stage I, II and III) were examined by the Wilcoxon/Kruskal-Wallis test. A mass signal at 1466 Da was significantly differenced among cervical cancer and healthy group. This peak was differentially high expressed in cervical cancer patients with stage III. On the other hand, m/z 1466 in a group of healthy and precancerous represents the low intensity as shown in Figure 20.

Table 9. The different mass signal among pooled stage group (healthy, precancerous, stage I, II and III). GA, SNN, different average, ANOVA and Wilcoxon/Kruskal-Wallis test (pWk) were used as algorithm for observing the significant peak (p-value <0.001).

Mass	Start Mass	End Mass	Weight	P-value	Dominant group
GA					
1488.72	1481.63	1498.28	39.7622503		Stage III
1741.72	1736.11	1747.68	48.0993892		Stage I
2307.55	2300.88	2316.39	20.3088624		Stage II
3242.1	3231.43	3252.41	30.2284207		Stage II
SNN					
1898.01	1892.44	1912.26	0.07462128		Stage I
2044.73	2038.47	2052.65	0.09665862		Healthy subjects
Different average					
1466.91	1460.72	1481.63	64.7121248		Stage III
ANOVA					
1898.01	1892.44	1912.26	9.36E-66	0.000008	Stage I
3159.09	3148.93	3176.56	3.39E-56	0.000008	precancerous
4299.4	4293.28	4336.07	4.49E-76	0.000002	Stage III
pWK					
1466.91	1460.72	1481.63	1.49E-21	0.000139	Stage III

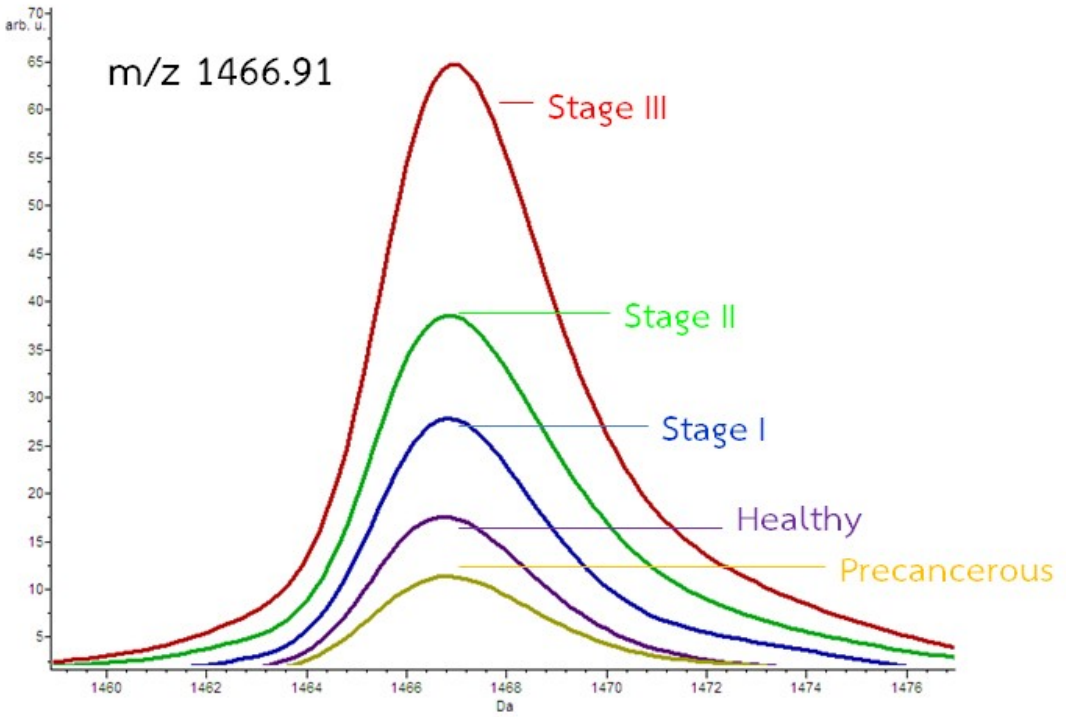


Figure 20. The significant different peaks among the investigated groups. The total average spectra in the range 1000- 10000 m/z were analyzed using the Wilcoxon/Kruskal-Wallis test. In cervical cancer patients with stage III, the peptide signals at 1466.91 Da was significantly differenced as compared with other groups ($p<0.001$).

2.2 Mass spectrum of the individual sample

All mass spectra from the individual sample were overlaid based on the staging and dataset. At m/z 1000- 10000, the peptide profiles in each sample had their own pattern; however, the individual sample in the same group still showed some similar peaks. In pseudo-gel view result, the 20 representative individual patterns of each sample group were aligned as shown in Figure 21. In this case, the results from individual sample exhibited each sample had the distinguished profile which consistency with chromatogram view.

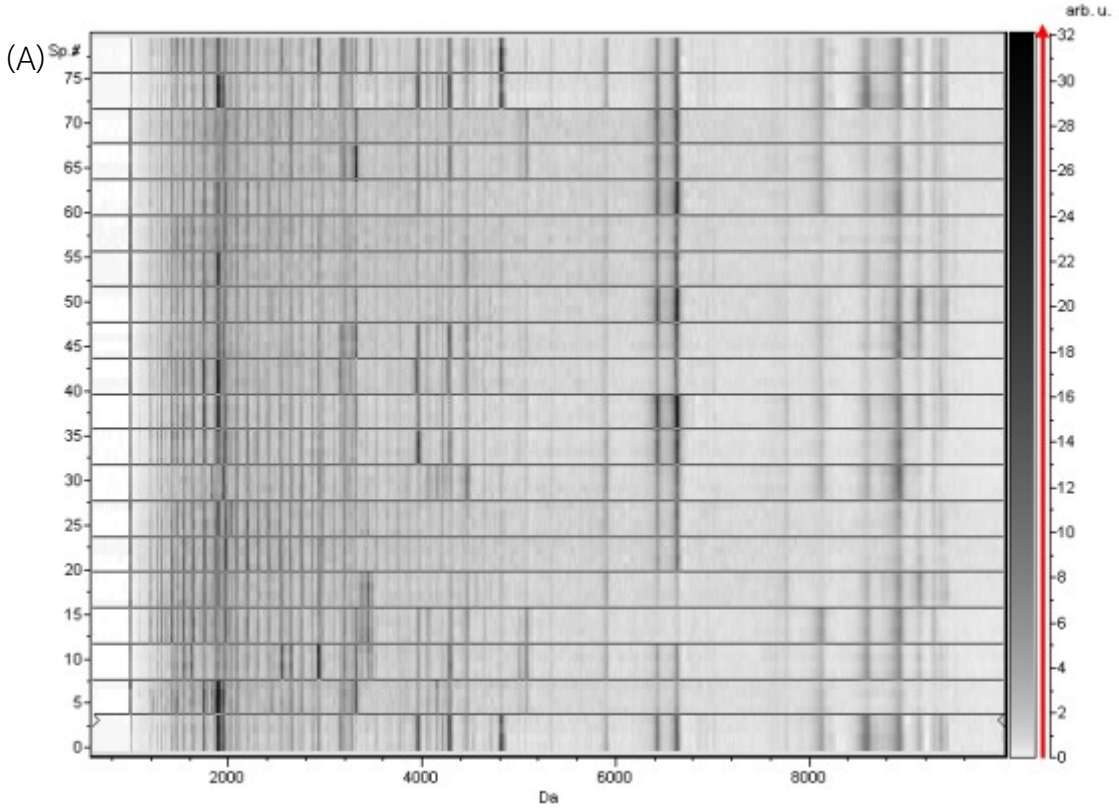


Figure 21. The gel view with the range of 1000- 10000 m/z of 20 individual samples per groups. Individual sample from healthy subjects and cervical cancer patient with precancerous, stage I, stage II, and stage III exhibited the patterns in A, B, C, D and E, respectively.

(B)

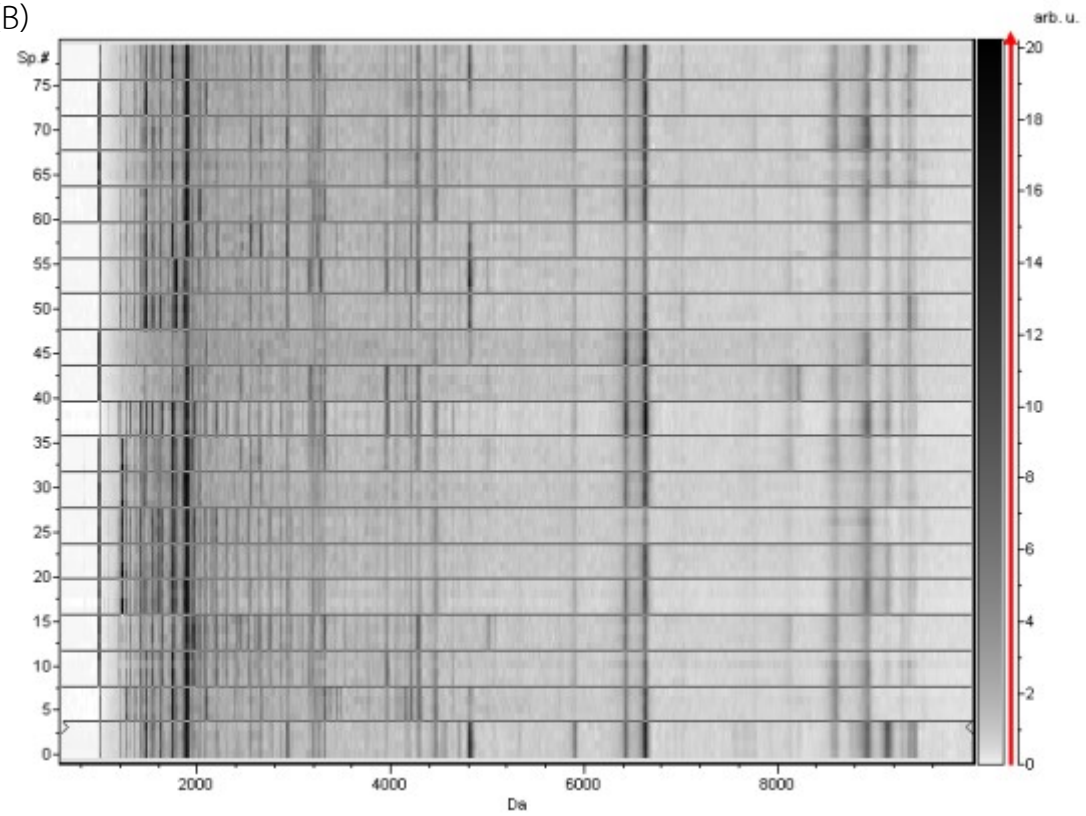


Figure 21. (Continued)

(C)

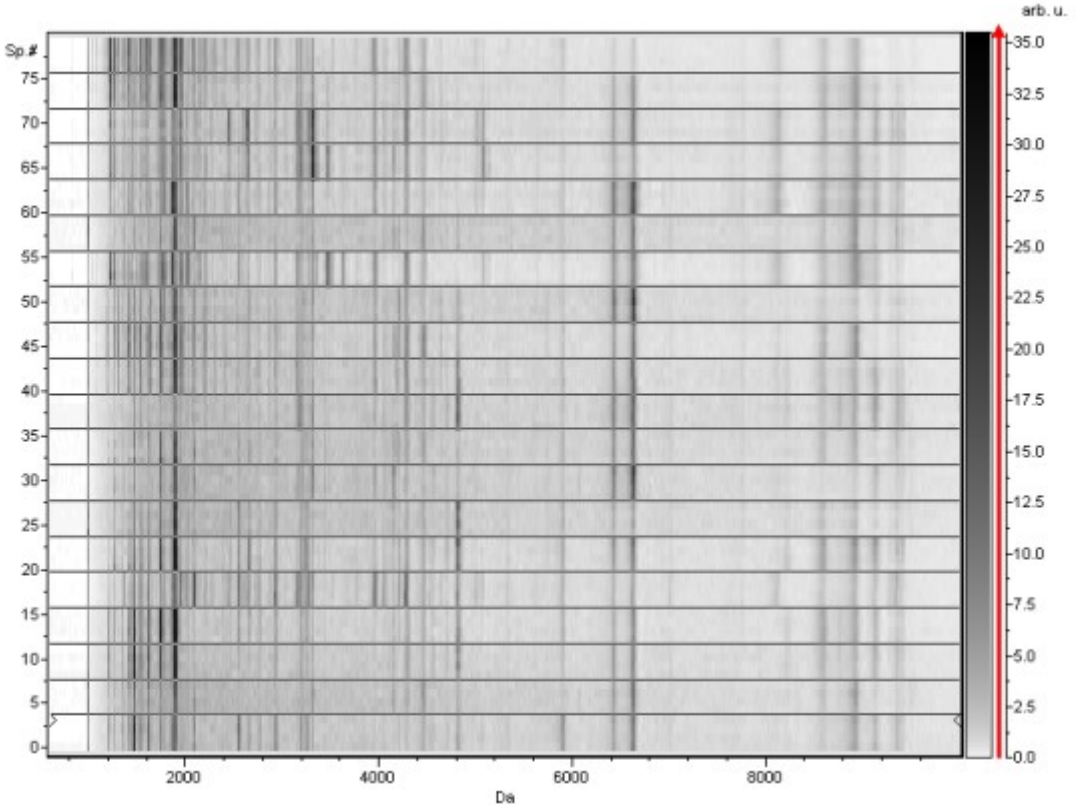


Figure 21. (Continued)

(D)

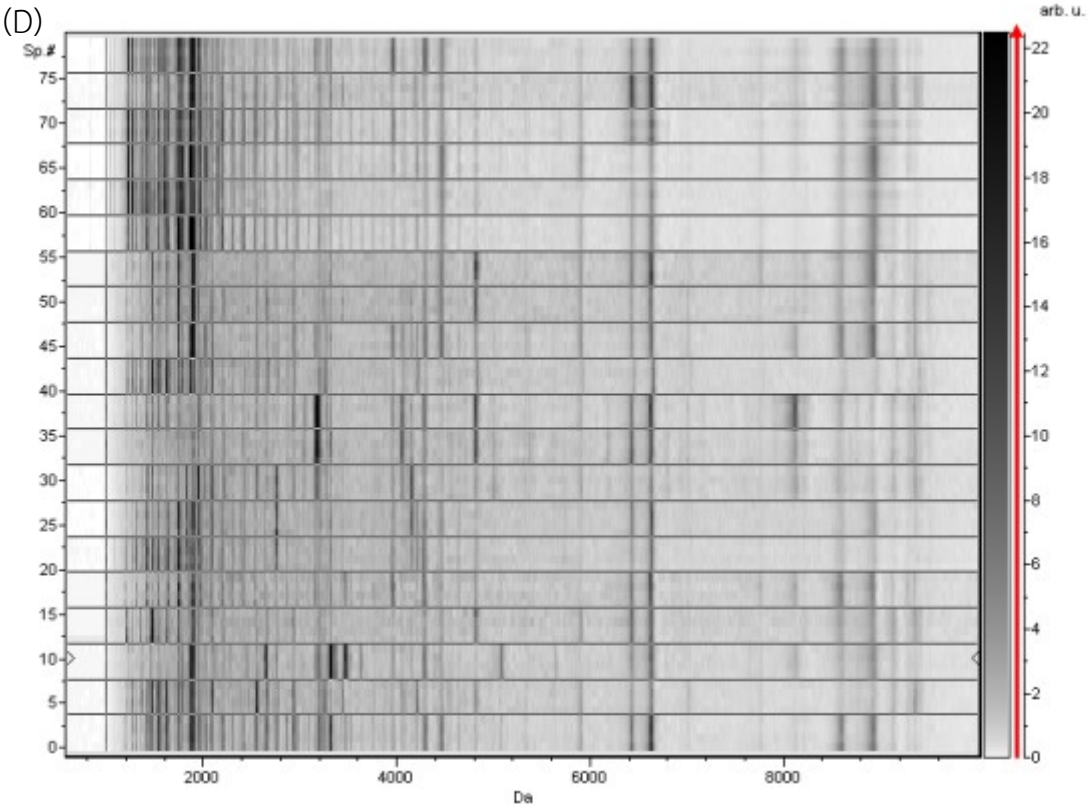


Figure 21. (Continued)

(E)

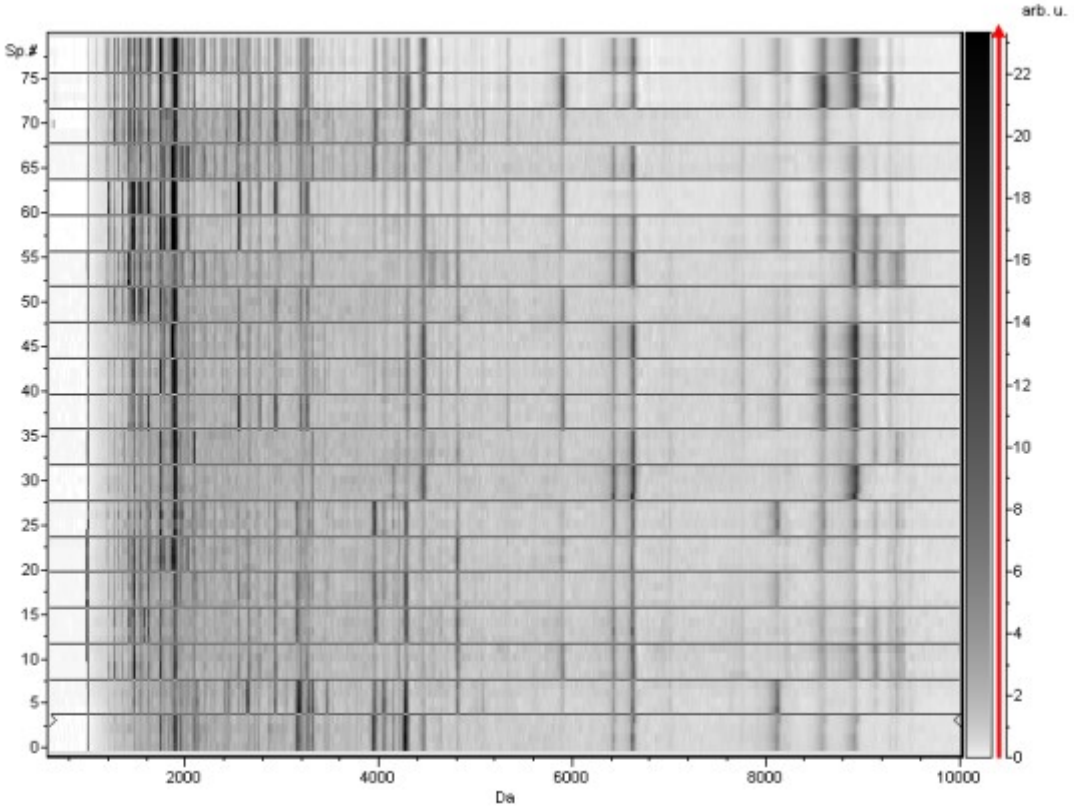


Figure 21. (Continued)

2.3 Comparison peptide profiles between pooled and individual samples

For generating the database, the appropriated data for creating reference mass spectrums were concerned. The dendrogram is the cluster analysis method in group similar mass spectrums of all peptide spectrums and intensity of peaks. From ClinProTool 2.2 software analysis, dendrogram revealed 5 groups of mass spectrums. Group 1 was mass spectrum from healthy subjects. Group 2 and 3 were precancerous group 1 (PMF from precancerous similar with stage I) and precancerous group 2 (peptide patterns from precancerous similar with stage III), respectively. Stage I&IIA (mass spectrums from stage II similar with stage I) was the sample group 4. Then, the 5 group of dendrogram was generated from stage IIB&III (IIB: peptide profiles from stage II similar with stage III). The mass spectrum of individual sample from group 1, 2, 3 and 5 were clustered and showed the shortest distance than mass spectrum from group 4. Mass spectrum from group 3 and 5 were grouped closer. The clade of stage I&IIA showed the difference clade from the others (Figure 22). The dendrogram in each group was shown in appendix B (Figure 25).

The specimens were pooled based on dendrogram prior to creating the database. PCA from 5 groups of serum (healthy, precancerous group 1, precancerous group 2, stage I&IIA and stage IIB&III) were analyzed. PCA score plot of spectra of 32 replicates of each group was performed. In this view, the spheres from precancerous group 1, precancerous group 2 and stage IIB&III were grouped closer. However, all studied groups showed the distinctly separate distribution in this dimension, as elucidated by PCA score plot (Figure23).

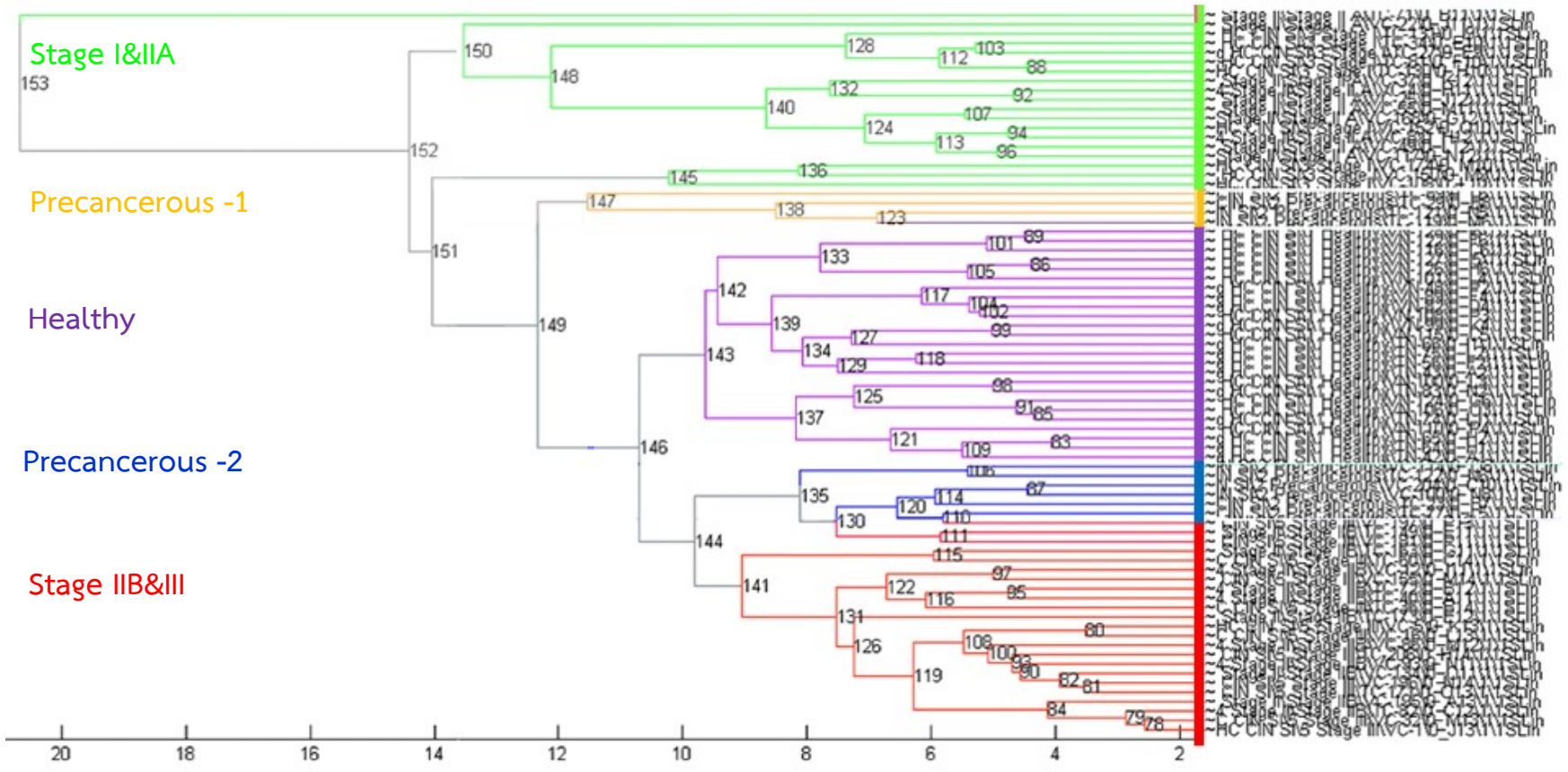


Figure 22. Dendrogram from the individual mass spectrum. Dendrogram generated from peptide profiles by using the Biotyper 2.0 software. The profiles were separated into 5 groups: healthy subject, precancerous group 1, precancerous group 2, stage I&IIA and stage IIB&III.

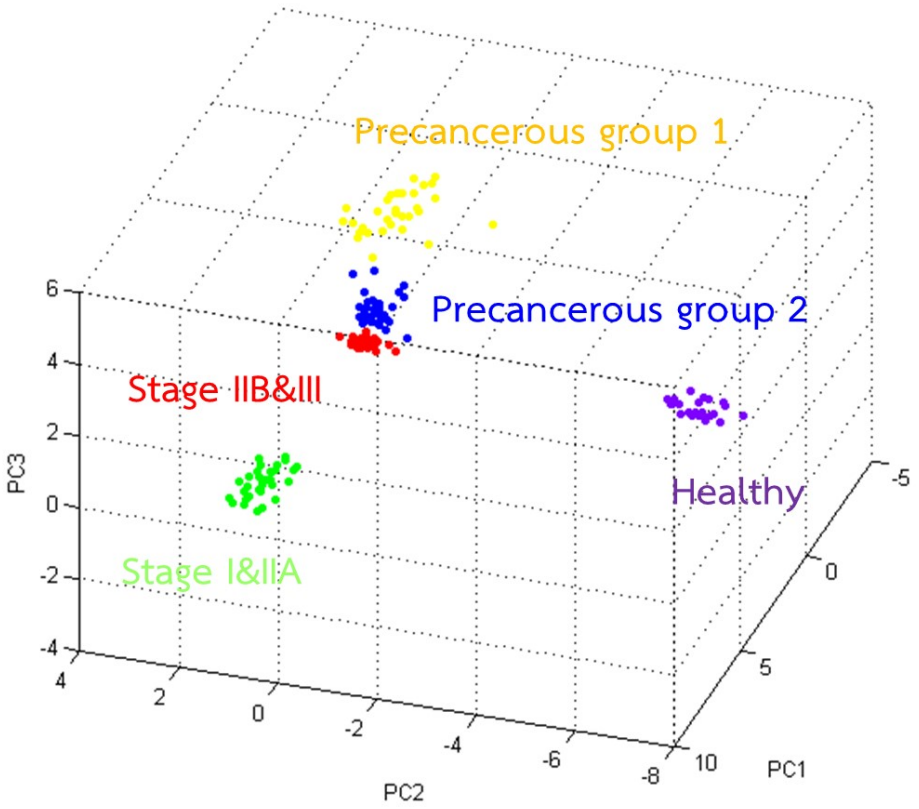


Figure 23. PCA for all spectrums of 5 pooled serum group based on dendrogram. The 5 separated clusters show purple (healthy), yellow (precancerous group 1), blue (precancerous group 2), green (stage I&IIA) and red sphere (stage IIB&III). Cluster analysis exhibited separately distribution among sample groups.

As a result, shown in Figure 24, peptide profiles of individual serum in each sample group were matched with the in-house reference mass (pooled serum). The inverted blue mass spectrum is the reference peptide profiles. The upper mass spectrum was obtained from the representative individual sample. The colors of pattern matching result were correlated with the similarity score ranging between 0 and 3. The score range from 2 to 3 was regarded as the same m/z peaks between subjected spectrum and database which displayed in the green color (score value 2.000-3.000) and yellow color (score value range of 1.700-1.999). The score value lower 1.699 and red color mass spectrum were presented the unmatched mass spectrum with the database spectrum (133). In the matching model, the individual groups from healthy, precancerous, stage I, stage II and stage III compared with in

house database resulted in the score value of all groups were more than 1.700. The perfect matching of database is shown in the Figure 24.

The mass spectrums were used for database generation. The in house database was contained mass spectrum of the patients with various stages of cervical cancer and healthy subjects. The peptide profiles of patients were compared with this in house database. Then, the patients was diagnosed the cervical cancer. Therefore, the in house database is a tool for supporting the early detection in cervical cancer. As results from Biotyper 2.0 software, PMF of each individual sample were calculated for comparing with the referenced mass spectrum. The numbers of matching between individual and reference mass spectrum were counted for sensitivity and specificity calculation. Sensitivity is the probability of matching number between individual and database group. Specificity is the percentage which the number mass spectrums from other group were showed the unmatched results. The sensitivity and specificity in each group of sample were shown in Table 10. In this experiment, the sensitivity and specificity with the database were 16.69-84.21% and 77.19-98.68%, respectively. The sensitivity and specificity of databases were shown in the Table 10.

(A)

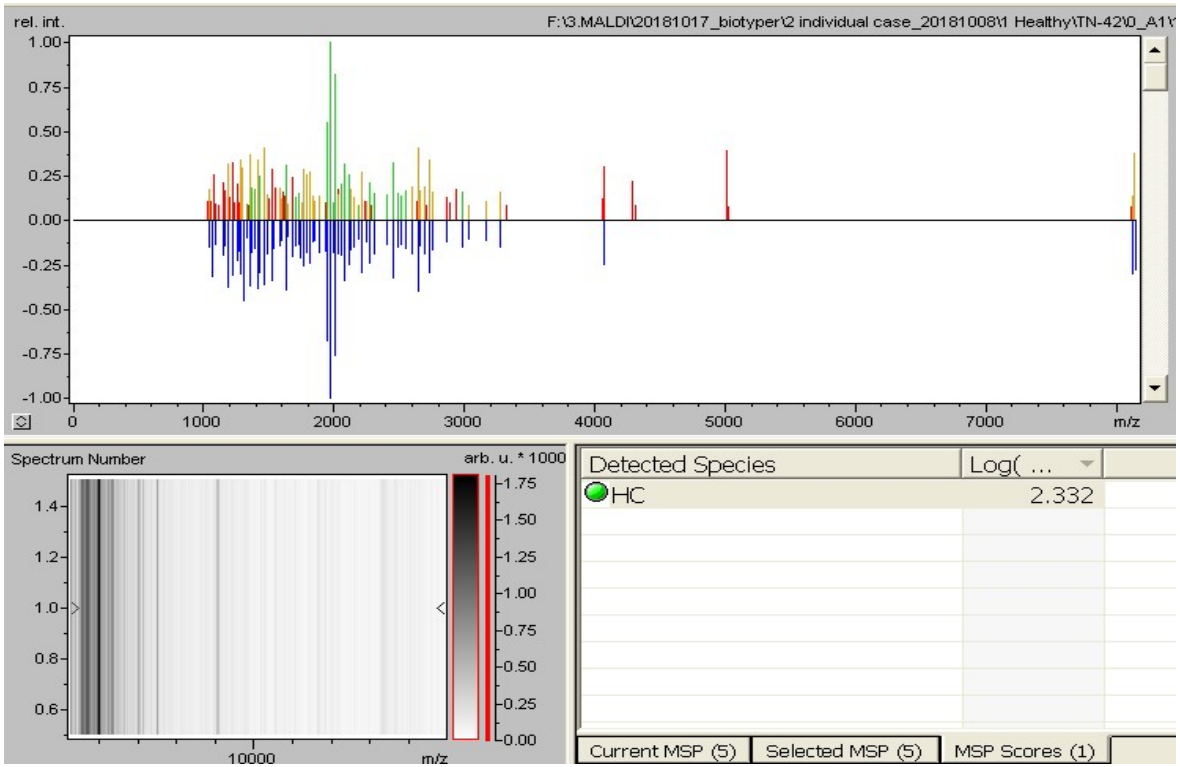


Figure 24. Comparison of the peptide profiles obtained from individual sample with the in-house database by using BioTyper 2.0 software. The pooled sample in each group was used as the reference spectrums in the database. The 5 selected individual profiles were subjected to match with the reference mass spectrum. As result of healthy women (A), patients with precancerous (B), stage I (C), II (D) and III (E), the mass spectrum in all group showed high matching score (the green and yellow color).

(B)

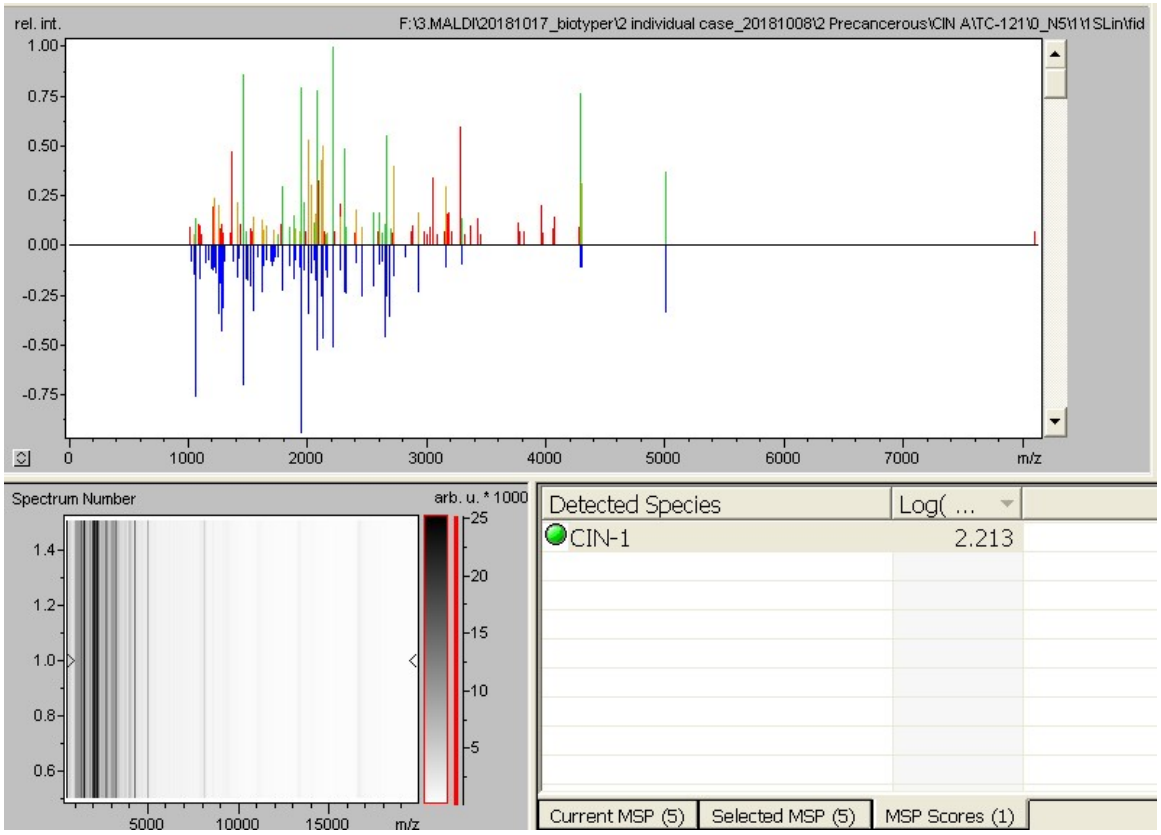


Figure 24. (Continued)

(C)

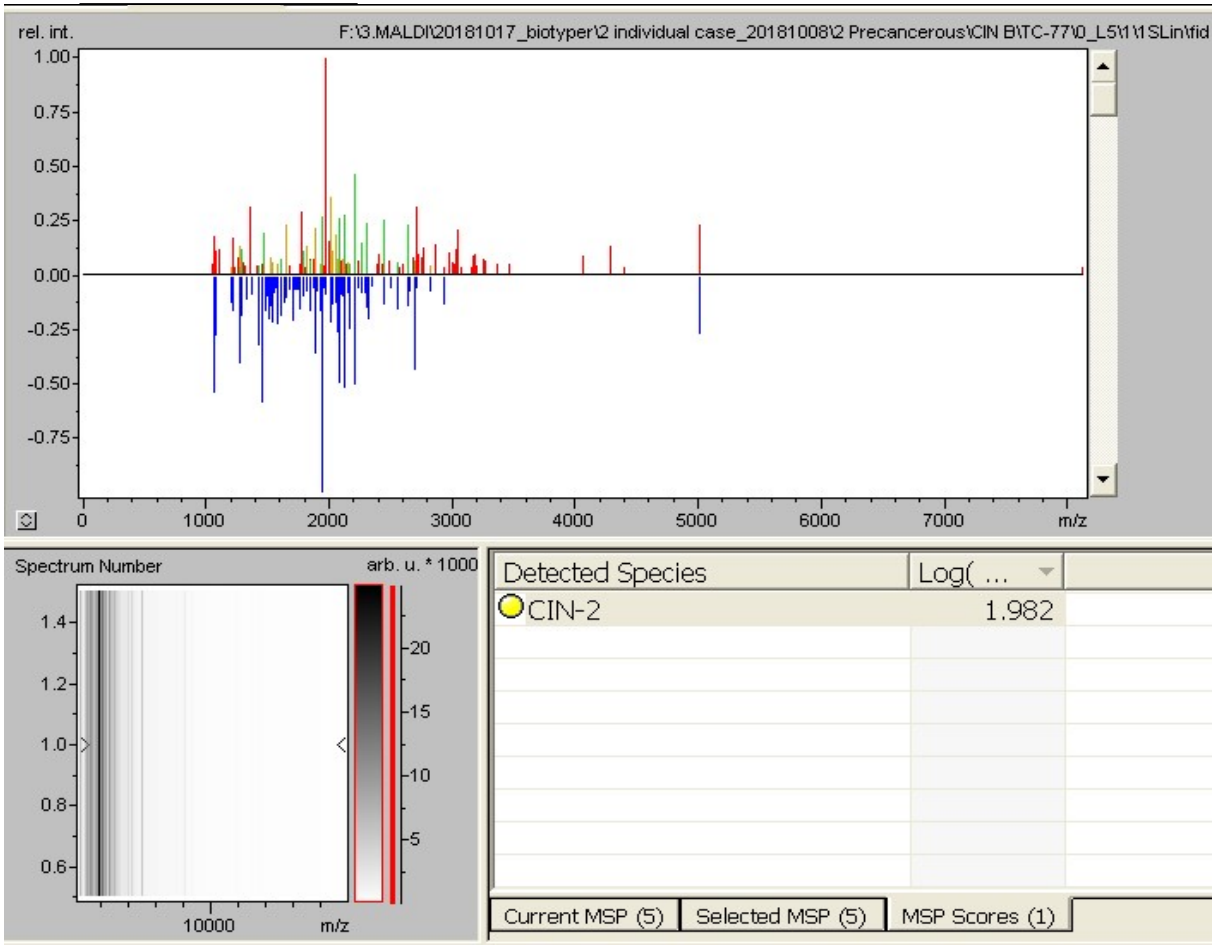


Figure 24. (Continued)

(D)

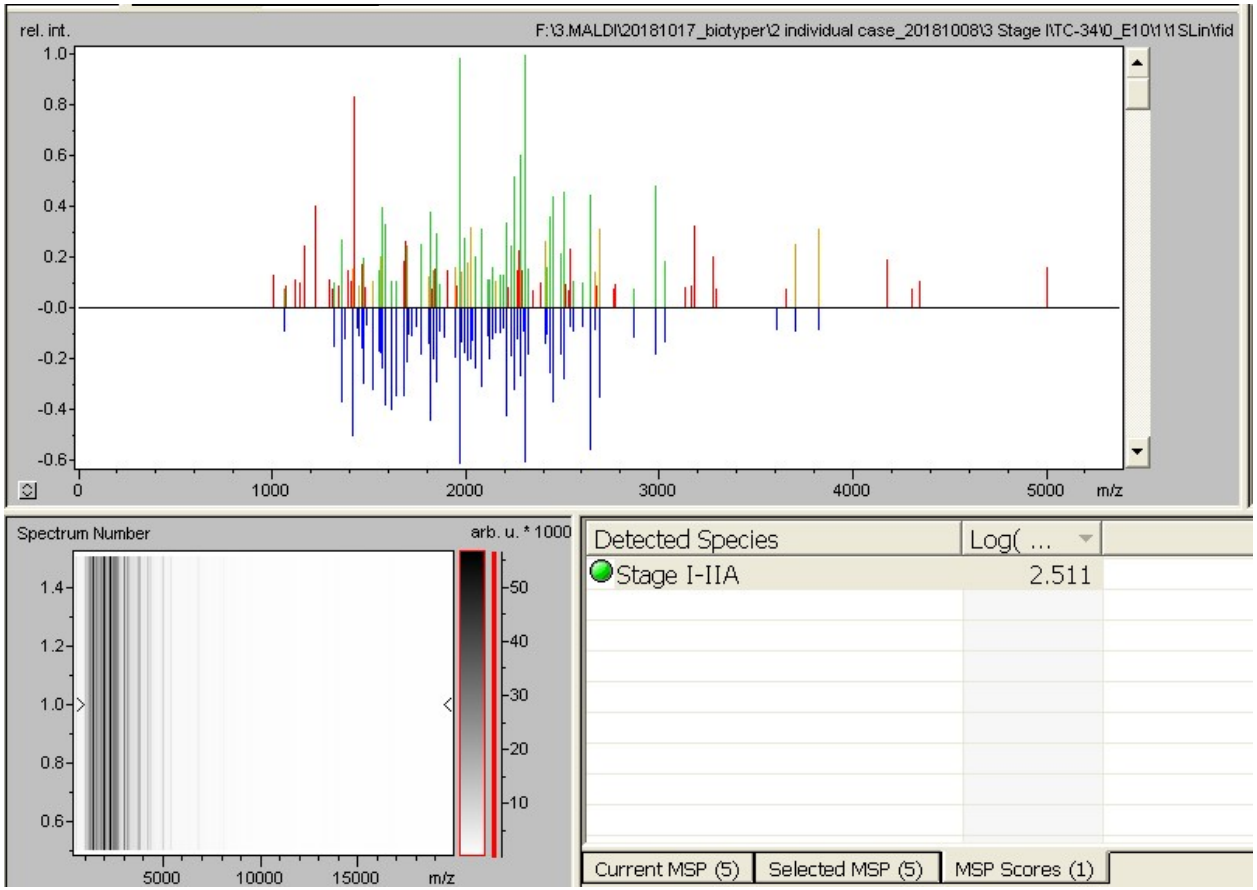


Figure 24. (Continued)

(E)

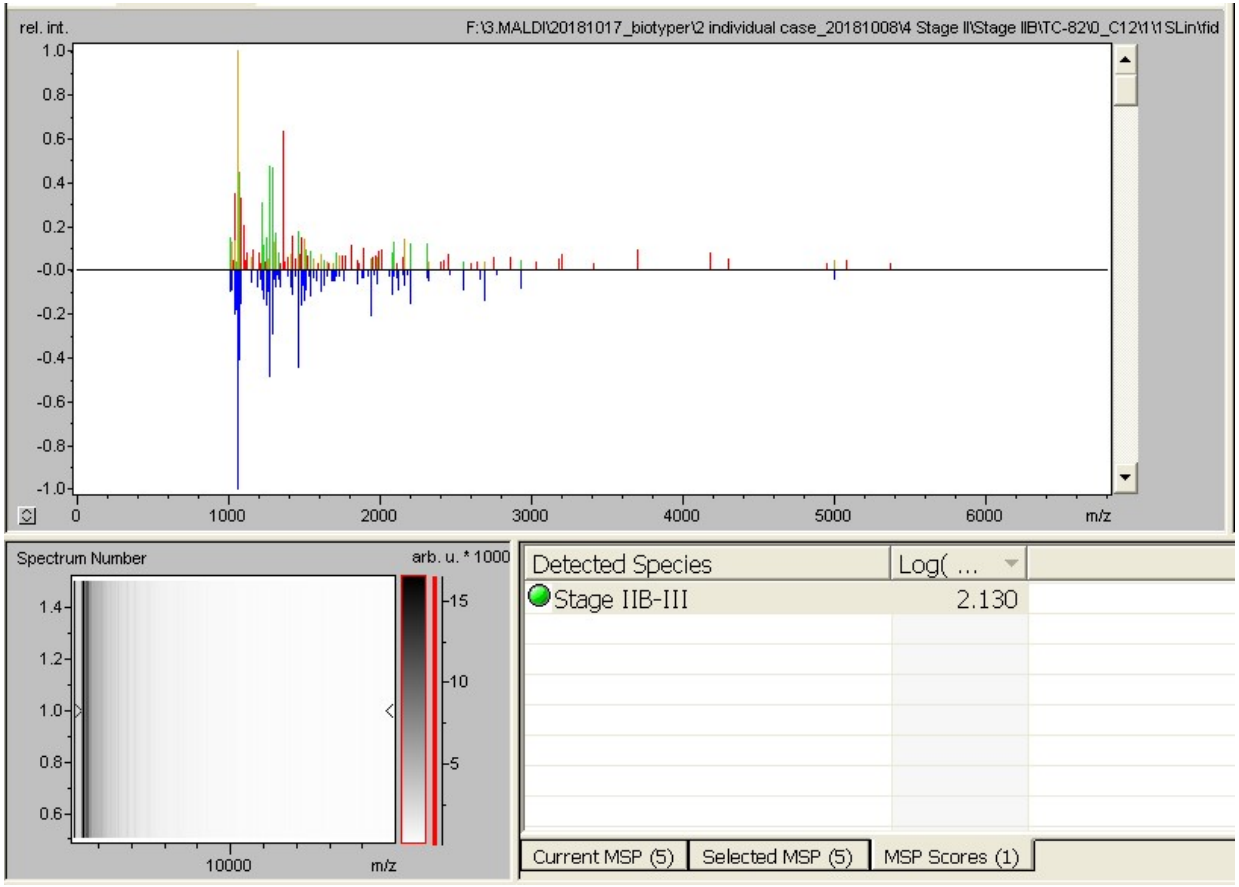


Figure 24. (Continued)

Table 10. Sensitivity and specificity values of the in-house database. Five groups of samples were used to generate the in-house database by Biotyper 2.0 software. The matched number between individual and this database were calculated.

		Sensitivity	Specificity
Database	Healthy	84.00	77.19
	Precancerous-1	75.00	93.59
	Precancerous-2	16.67	98.68
	Stage I & II set 1	84.21	98.41
	Stage II set 2 & III	32.14	98.15

CHAPTER 4

DISCUSSION

Nowadays, the most popular tests for cancer diagnosis are a protein-based assay to discover cancer-associated peptides. The current use of single biomarker such as SCCA for cervical cancer diagnosis is unsuccessful in early phase of cancer due to its low expression. Although other molecules including CA19-9, CEA and CA125 have been acted as biomarkers in cervical cancer detection, all of these proteins have also been utilized in other cancers (23- 34). It seems likely that no early diagnostic biomarkers with high sensitivity and specificity can be applied as a routine screening assay. Nowadays, the combined biomarkers could increase sensitivity and specificity for the screening test (100). Lately, mass spectrometry-based peptide profiles as the combination have developed a strategy to replace non-specific of single biomarker tests for the early diagnosis of cancer. Peptidomic patterns analysis focuses on the patterns of peptide rather than the identification of a single relevant biomarker.

In this study, the ultrafiltration was the first sample preparation technique before mass spectrum analysis. In this method, the signal profiles of peptide and polymer were detected resulted in false clustering (Figure 10-12). The 110 Da repeated spectrums of polymer were found in the range of m/z 1000-3000. In the range of 3000-5000 Da, which the 43 da repeated polymer were found in sample of polymer showed of 3000- 5000 da (Figure 13). In the previous research, the repeat mass distribution of the interval 43 Da represented the mass spectrums of polyethylene glycol monododecyl ether or PEG (134). Moreover, dihydroxy PEG usually shows the polymer spectrum with 110 Da repeat units (135). Commonly, the blood collection tube wall composes of many polymers (polyester, polyethylene, polypropylene, polyacrylic, polysiloxane, polytetrafluoroethylene, polyvinyl chloride, polyacrylonitrile, and polystyrene). Furthermore, tube surfactant structure is polyethylene oxide and polypropylene oxide attached with the polydimethylsiloxane (136). Further studies have examined whether polyethylene in plastic wall and surfactant that may interfere with MALDI-TOF MS peaks. They found that surfactant from test tubes may influence the ionization process which generates the signal peak in the range of 1000-5000 Da (127). Therefore, the specimen for the

MALDI-TOF MS approach should be amount of the blood containing and plastic tube without dihydroxy PEG and poly (ethylene glycol) monododecyl ether.

The reverse phase chromatography or ZipTipC18 was used for removing the polymer. The polyethylene glycol is the hydrophobic long chain molecules which strongly attach with the stationary phase in ZiptipC18 (137, 138). The peptides elutes while some polymer still binds to C18 on the resin (stationary phase). The polymer with the high polarity is eluted with low percentage of ACN or high proportion of distilled water. On the other hand, the high concentration of ACN in the eluted solution is suitable to elute the low polarity of polymer. The sample preparation step is an important factor MALDI-TOF MS analysis; therefore, the optimal condition is necessary. The fraction chromatographic separation sample to matrix ratio was performed to optimize sample preparation by using ZipTipC18 (Figure 14-16). The 100% ACN was suitable for eluting the peptide from ZiptipC18 with less amount of polymer. For the crystallization step, the CHCA matrix 1:5 was suitable to ionize the biomolecules. The biomolecules or sample were generated the ion when the biomolecules were collision with matrix ions. The matrix ions to analysis are small, resulting in a small number of biomolecules were ionized. The small number of ion generated low number of peaks and low intensity in the mass spectrum. On the other hand, the large number of matrixes induced to generate many kinds of biomolecules ions, especially in complexity sample (139).

The previous studies reported that the use of MALDI-TOF MS and SELDI-TOF MS could differentiate the status of healthy subjects from disease. A little attention has been paid to signature peptide patterns and the presence of disease. For example, 55 signature peptides in primary glomerulonephritis patients specimens which were different from m/z 1769, 1898, 1913, 1945, 2491, 2756, 2977, 3389 and 4752 were significantly correlated with poor kidney function. Plasma peptide at m/z 8602 and serum peptides at m/z 3242 was increased in endocapillary hypercellularity. This peptide profiles can improve the diagnosis and prognosis of this glomerular disease (140). In cancer research, 5 MALDI peaks at 1021, 1467, 8944, 3149 and 4137 as the peptide patterns used to distinguish the small cell lung cancer from healthy subjects (141). For gastric cancer classification, peptides at m/z 2046, 3179, 1817, 1725 and 1929 were constructed to differentiate patients with gastric cancer from healthy controls (142). Moreover, the peptide peaks at 1741 and 4210 m/z were correlated with gastric cancer when compared with chronic atrophic gastritis and healthy subjects (143). The m/z peak at 2049 and 2305 was higher

expressed with pancreatic disease than gastric cancer, chronic gastritis, diabetes mellitus, and healthy (144). The differently-expressed peptides peaks in both of familial adenomatous polyposis and sporadic colorectal adenomas were detected using SELDI- TOF MS. Peak m/z 5640, 3160, 4180 and 4290 were most frequently found in patients with FAP adenomas, peaks of 3940 and 3400 m/z were positively associated with sporadic colorectal adenomas (145). In agreement with the previous studies, we found different peptide among healthy subjected and patients with cancer. Nine differently expression of mass peaks were detected in cervical cancer (precancerous, stage I, II and III) and healthy subjected. These peptide peaks include 1741.72, 3242.10, 1488.72, 2307.55, 2044.73, 1898.01, 3159.09, 4299.40 and 1466.91Da. Moreover, we found that peak at m/z 1466.91 was significantly differentially expressed in cervical cancer. This peak was detected in serum from patients with stage III, whereas its intensity was low in healthy and precancerous group. Moreover, previous reports showed that 1466 Da peptide was highly expressed in cirrhotic with hepatocellular carcinoma compared with cirrhosis patients (146). The peak at m/z 1466 was detected in gastric cancer patients with lymph node metastasis but absence in gastric cancer without lymph node metastasis and normal control sample. The exopeptidases carboxypeptidase Y and P is the enzyme which digested myoglobin at the c terminal sequence. Consequently, the mass peaks at 1465 with sequence DIAA was detected via MALDI-TOF MS (147). Moreover, the 1466 Da was identified as fibrinogen A with an alanine truncation at N-terminal (degAla-FPA) with sequence DSGEGDFLAEGGGVR (110). FPA have many fragments (the peaks at 758, 905, 1020, 1077, 1206, 1263, 1350, 1466, 1470, 1626, 5904, and 5917 Da) which were generated from the tumor-specific exoprotease activity (125). This data supports that the peptidome in serum is amplified by exoprotease activity. Tumor cell may provide the unique exoprotease, therefore the peptides in cancer patients may differ from the healthy (148, 149).

Interestingly, peak at 1466.91 may be the digested peptides form tumor specific peptides which would be a candidate biomarker in advance stage of cancer. Therefore, this peak would be further identified. Use of other MS technique such as LC-MS/MS is needed to further identify this peptide sequence peak at m/z 1466.91. However, the disadvantage of this technique is time-consuming, more complicated and expensive (114). Furthermore, electrospray ionization (ESI) incorporated with LC-MS/MS produced multiply charged ions resulted in the loss of singly charged ions, whereas MALDI generated peptides containing single charged ions (101, 103). Thus, use of high-resolution tandem mass spectrometry (MALDI-TOF/TOF)

could enable the highly confident identification of peptide sequence analysis. Nevertheless, the peptide patterns may have the utility for cervical cancer screening/ diagnosis test via generating in house database.

The in-house database was created using pooled samples. The sensitivity and specificity from the matching score with this database are ranged 16.67- 84.21% and range 77.19- 98.68%, respectively. The combinations of MALDI-TOF MS and Biotyper 2.0 software in this study have higher specificity than cytological examination from the previous study (8). However, this sensitivity is lower than the traditional screening test for cervical cancer as the number of samples per group is too small which affects the representation of the analytical group.

However, our current data is informative and needs to validate the signature peptide patterns for screening and diagnostic test of cervical cancer. Further studies, the generated databases are required to develop the diagnostic model and subsequently proven this model by using a large number of individual patients and healthy subjects. Moreover, the use of externally validated samples such as patients with other cancer types is required to confirm the specific peptide patterns to cervical cancer.

CHAPTER 5

CONCLUSION

The specific serum peptide profiles of healthy and patients with various stages of cervical cancer were exhibited and used to construct the in-house database. However, a larger number of individual cervical cancer patients, healthy controls and other types of cancer should be recruited to fully explore the reliability of the database and use as a diagnostic tool. Taken together, utilization of MALDI technology combination with bioinformatics tools could promote the discovery of cancer-associated peptides and contribute to better early diagnosis with reproducibility, less time consuming, and high throughput.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
2. ICO. Human Papillomavirus and Related Cancers. Fact Sheet. 2016.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66:7-30.
4. Tewari KS, Sill MW, Long HJ, Penson RT, Huang H, Ramondetta LM, et al. Improved survival with bevacizumab in advanced cervical cancer. N Engl J Med. 2014;370:734-43.
5. Fateme CPJ, Masoumi K, Forouzan A and Jafar RM. Comparison results of Pap smear and colposcopy and histopathology in women with post-coital bleeding. J Obstet Gynaecol Res. 2015;8:10-5.
6. Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam S, Cain J, et al. American Cancer Society, American Society for colposcopy and cervical pathology, and American Society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin. 2012;62:147-72.
7. Gibb RK, Martens MG. The impact of liquid-based cytology in decreasing the incidence of cervical cancer. Rev Obstet Gynecol. 2011;4:S2-11.
8. Karimi-Zarchi M, Peighmbari F, Karimi N, Rohi M, Chiti Z. A comparison of 3 ways of conventional Pap smear, liquid-based cytology and colposcopy vs cervical biopsy for early diagnosis of premalignant lesions or cervical cancer in women with abnormal conventional Pap test. Int J Biomed Sci. 2013;9:205-10.
9. Al-Naggar RA, Isa ZM. Perception and opinion of medical students about Pap smear test: a qualitative study. Asian Pac J Cancer Prev. 2010; 11: 435-40.
10. Wright TC, Jr. Cervical cancer screening in the 21st century: is it time to retire the PAP smear?. Clin Gynecol Obstet. 2007;50:313-23.
11. Cuzick J, Mayrand MH, Ronco G, Snijders P, Wardle J. Chapter 10: New dimensions in cervical cancer screening. Vaccine. 2006;24:S3/90-7.
12. Schiffman M, Herrero R, Hildesheim A, Sherman ME, Bratti M, Wacholder S, et al. HPV DNA testing in cervical cancer screening: results from women in a high-risk province of Costa Rica. Jama. 2000;283:87-93.

13. Clavel C, Masure M, Bory J, Putaud I, Mangeonjean C, Lorenzato M, et al. Human papillomavirus testing in primary screening for the detection of high-grade cervical lesions: a study of 7932 women. Br J Cancer. 2001;84:1616-23.
14. Brandstetter T, Böhmer S, Prucker O, Bissé E, zur Hausen A, Alt-Mörbe J, et al. A polymer-based DNA biochip platform for human papilloma virus genotyping.I Virol Methods. 2010;163(1):40-8.
15. Malati T. Tumour markers: An overview. Indian J Clin Biochem. 2007; 22: 17-31.
16. Kirwan A, Utratna M, Dwyer ME, Joshi L, Kilcoyne M. Glycosylation-based serum biomarkers for cancer diagnostics and prognostics. Biomed Res Int. 2015;2015: 490531.
17. Kosugi S, Nishimaki T, Kanda T, Nakagawa S, Ohashi M, Hatakeyama K. Clinical significance of serum carcinoembryonic antigen, carbohydrate antigen 19-9, and squamous cell carcinoma antigen levels in esophageal cancer patients. World J Surg. 2004;28:680-5.
18. Molina R, Filella X, Auge JM, Fuentes R, Bover I, Rifa J, et al. Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell lung cancer as an aid in histological diagnosis and prognosis. comparison with the main clinical and pathological prognostic factors. Tumour Biol. 2003;24:209-18.
19. Ayude D, Gacio G, Cadena PDL, Pallas M, E., Martínez-Zorzano V, De Carlos S, et al. Combined use of established and novel tumour markers in the diagnosis of head and neck squamous cell carcinoma. Oncol Rep. 2003;10(5):1345-1350
20. Petrelli NJ, Palmer M, Herrera L, Bhargava A. The utility of squamous cell carcinoma antigen for the follow-up of patients with squamous cell carcinoma of the anal canal. Cancer. 1992;70:35-9.
21. Ngan HYS, Cheung ANY, Lauder IJ, Cheng DKL, Wong LC, Ma HK. Tumour markers and their prognostic value in adenocarcinoma of the cervix. Tumour Biol 1998; 19:439-44.
22. Tsai CC, Liu YS, Huang EY, Huang SC, Chang HW, Tseng CW, et al. Value of preoperative serum CA125 in early-stage adenocarcinoma of the uterine cervix without pelvic lymph node metastasis. Gynecol Oncol. 2006; 100:591-5.
23. Crombach G, Scharl A, Wurz H. CA 125 in normal tissues and carcinomas of the uterine cervix, endometrium and Fallopian tube. II. Immunoradiometric determination in secretions, tissue extracts and serum. Arch Gynecol Obstet. 1989;244:113-22.

24. Borras G, Molina R, Xercavins J, Ballesta A, Iglesias J. Tumor antigens CA 19.9, CA 125, and CEA in carcinoma of the uterine cervix. Gynecol Oncol. 1995;57: 205-11.
25. Gadducci A, Cosio S, Carpi A, Nicolini A, Genazzani AR. Serum tumor markers in the management of ovarian, endometrial and cervical cancer. Biomed Pharmacother. 2004;58:24-38.
26. Wang JY, Lu CY, Chu KS, Ma CJ, Wu DC, Tsai HL, et al. Prognostic significance of pre- and postoperative serum carcinoembryonic antigen levels in patients with colorectal cancer. Eur Surg Res. 2007;39:245-50.
27. Mujagic Z, Mujagic H, Prnjavorac B. The relationship between circulating carcinoembryonic antigen (CEA) levels and parameters of primary tumor and metastases in breast cancer patients. Med Arh. 2004;58:23-6.
28. D'Aloia A, Faggiano P, Aurigemma G, Bontempi L, Ruggeri G, Metra M, et al. Serum levels of carbohydrate antigen 125 in patients with chronic heart failure: relation to clinical severity, hemodynamic and doppler echocardiographic abnormalities, and short-term prognosis. J Am Coll Cardiol. 2003;41:1805-11.
29. Nguyen L, Cardenas-Goicoechea SJ, Gordon P, Curtin C, Momeni M, Chuang L, et al. Biomarkers for early detection of ovarian cancer. Womens Health (Lond). 2013;9(2):171-85.
30. Duffy chem. MJ. CA 15-3 and related mucins as circulating markers in breast cancer. Ann Clin Bio. 1999;36:579-86.
31. Park BW, Oh JW, Kim JH, Park SH, Kim KS, Kim JH, et al. Preoperative CA 15-3 and CEA serum levels as predictor for breast cancer outcomes. Ann Oncol. 2008; 19:675-81.
32. Shering SG, Sherry F, McDermott EW, O'Higgins NJ, Duffy MJ. Preoperative CA 15-3 concentrations predict outcome of patients with breast carcinoma. Cancer. 1998;83:2521-7.
33. Bhatt AN, Mathur R, Farooque A, Verma A, Dwarakanath BS. Cancer biomarkers - current perspectives. Indian J Med Res. 2010;132:129-49.
34. Kannagi R. Carbohydrate antigen sialyl Lewis a--its pathophysiological significance and induction mechanism in cancer progression. Chang Gung Med J. 2007;30: 189-209.
35. Jones MB, Krutzsch H, Shu H, Zhao Y, Liotta LA, Kohn EC, et al. Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. Proteomics. 2002;2:76-84.

36. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics. 2004;3:367-78.
37. Karram MB. Atlas of pelvic anatomy and gynecologic surgery. 2010.
38. Organization GWH. Comprehensive cervical cancer control: a guide to essential practice. 2nd edition. 2014.
39. Torre LA, Siegel RL, Ward EM, Jemal A. global cancer incidence and mortality rates and trends--an update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27.
40. Franco EL, Duarte-Franco E, Ferenczy A. Cervical cancer: epidemiology, and the role of human papillomavirus infection. Cmaj. 2001;164(7):1017-25.
41. Waggoner SE. Cervical cancer. Lancet. 2003; 361: 2217-25.
42. Bergvall M, Melendy T, Archambault J. The E1 proteins. Virology. 2013;445:35-56.
43. McBride AA. The papillomavirus E2 proteins. Virology. 2013;445:57-79.
44. Doorbar J. The E4 protein; structure, function and patterns of expression. Virology. 2013;445:80-98.
45. Harden ME, Munger K. Human papillomavirus molecular biology. Mutat Res Rev Mutat Res. 2017;772:3-12.
46. Vande Pol SB, Klingelhutz AJ. Papillomavirus E6 oncoproteins. Virology. 2013;445 :115-37.
47. Yim EK, Park JS. The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. Cancer Res Treat. 2005;37:319-24.
48. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. Lancet . 2013;382:889-99.
49. Burd EM. Human papillomavirus and cervical cancer. Clin Microbiol Rev. 2003;16 :1-17
50. Buck CB, Day PM, Trus BL. The papillomavirus major capsid protein L1. Virology. 2013;445:169-74.
51. Wang JW, Roden RBS. L2, the minor capsid protein of papillomavirus. Virology. 2013;445:175-86.
52. Zhou J, Sun XY, Stenzel DJ, Frazer IH. Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. Virology. 1991;185:251-7.
53. Kelloff GJ, Sigman CC. Assessing intraepithelial neoplasia and drug safety in cancer-preventive drug development. Nat Rev Cancer. 2007;7(7):508-18.

54. Wang T-y, Chen B-f, Yang Y-c, Chen H, Wang Y, Cviko A, et al. Histologic and immunophenotypic classification of cervical carcinomas by expression of the p53 homologue p63: a study of 250 cases. Hum Pathol. 2001;32:479-86.
55. Greer BE, Figge DC, Tamimi HK, Cain JM. Stage IB adenocarcinoma of the cervix treated by radical hysterectomy and pelvic lymph node dissection. Am J Obstet Gynecol. 1989;160:1509-14.
56. Eifel PJ, Morris M, Oswald MJ, Wharton JT, Delclos L. Adenocarcinoma of the uterine cervix. prognosis and patterns of failure in 367 cases. Cancer. 1990; 65: 2507-14.
57. Hale RJ, Wilcox FL, Buckley CH, Tindall VR, Ryder WDJ, Logueh JP. Prognostic factors in uterine cervical carcinoma: a clinicopathological analysis. Int J Gynecol Cancer. 1991;1:19-23.
58. Yazigi R, Sandstad J, Munoz AK, Choi DJ, Nguyen PD, Risser R. Adenosquamous carcinoma of the cervix: prognosis in stage IB. Obstet Gynecol. 1990;75:1012-5.
59. Park J, Sun D, Genest DR, Trivijitsilp P, Suh I, Crum CP. Coexistence of low and high grade squamous intraepithelial lesions of the cervix: morphologic progression or multiple Papillomaviruses. Gynecol Oncol. 1998;70:386-91.
60. Cervical and vulva cancer: chages in FIGO definitions of staging. Br J Obstet Gynaecol. 1996;103:405-6.
61. Loscalzo D. Harrison's Principles of Internal Medicine. 18th ed: McGraw-Hill, 2012.
62. Kelly-Spratt KS, Kasarda AE, Igra M, Kemp CJ. A mouse model repository for cancer biomarker discovery. J Proteome Res. 2008;7(8):3613-8.
63. Shin SH, Bode AM, Dong Z. Precision medicine: the foundation of future cancer therapeutics. NPJ Precis Oncol. 2017;1(1):12-4.
64. Blatt AJ, Kennedy R, Luff RD, Austin RM, Rabin DS. Comparison of cervical cancer screening results among 256,648 women in n\multiple clinical practices. Cancer Cytopathol. 2015;123:282-8.
65. Muntean M, Simionescu C, Taslîcă R, Gruia C, Comanescu A, Pătrană N, et al. cytological and histopathological aspects concerning preinvasive squamous cervical Lesions. Curr Health Sci J. 2010;36:26-32.
66. Franco EL, Ferenczy A. Cervical cancer screening following the implementation of prophylactic human papillomavirus vaccination. Future Oncol. 2007;3:319-27.
67. Arbyn M, Sasieni P, Meijer CJ, Clavel C, Koliopoulos G, Dillner J. Chapter 9: Clinical applications of HPV testing: a summary of meta-analyses. Vaccine. 2006;24: 78-89.

68. Kulasingam V, Pavlou MP, Diamandis EP. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. Nat Rev Cancer. 2010; 10:371-8.
69. Fathi E, Sam N, Farahzadi R. Biomarkers in medicine: an overview.Br J Med Med Res. 2014;4(8):1707-18.
70. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies.Nat Clin Pract Oncol. 2008;5(10):588-99.
71. Orton DJ, Doucette AA. Proteomic workflows for biomarker identification using mass spectrometry technical and statistical considerations during initial discovery. Proteomes. 2013;1:109-27.
72. Raju KL, Augustine D, Rao RS, S VS, Haragannavar VC, Nambiar S, et al. Biomarkers in tumorigenesis using cancer cell lines: a systematic review. Asian Pac J Cancer Prev. 2017;18(9):2329-37.
73. Lenas P, Moos M, Luyten FP. Developmental engineering: a new paradigm for the design and manufacturing of cell-based products. part I: from three-dimensional cell growth to biomimetics of in vivo development. Tissue Eng Part B Rev. 2009;15(4):381-94.
74. Martin KJ, Patrick DR, Bissell MJ, Fournier MV. Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets. PLoS One. 2008;3(8):e2994.
75. Weigelt B, Bissell MJ. Unraveling the microenvironmental influences on the normal mammary gland and breast cancer. Semin Cancer Biol. 2008;18(5):311-21.
76. Schmeichel KL, Bissell MJ. Modeling tissue-specific signaling and organ function in three dimensions. J Cell Sci. 2003;116(12):2377-88.
77. Carboni L. Peripheral biomarkers in animal models of major depressive disorder. Dis Markers. 2013;35(1):33-41.
78. Denayer T, Stöhr T, Van Roy M. Animal models in translational medicine: validation and prediction. New Horiz Transl Med. 2014;2(1):5-11.
79. Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. Am J Transl Res. 2014;6(2):114-8.
80. Tambor V, Fucikova A, Lenco J, Kacerovsky M, Rehacek V, Stulik J, et al. Application of proteomics in biomarker discovery: a primer for the clinician. Physiol Res. 2010;59(4):471-97.

81. Jurisicova A, Jurisica I, Kislinger T. Advances in ovarian cancer proteomics: the quest for biomarkers and improved therapeutic interventions. Expert Rev Proteomics. 2008;5(4):551-60.
82. Harpole M, Davis J, Espina V. Current state of the art for enhancing urine biomarker discovery. Expert Rev Proteomics. 2016;13(6):609-26.
83. Stephan C, Jung K, Diamandis EP, Rittenhouse HG, Lein M, Loening SA. Prostate-specific antigen, its molecular forms, and other kallikrein markers for detection of prostate cancer. Urol . 2002;59:2-8.
84. Barry MJ. Prostate-specific–antigen testing for early diagnosis of prostate cancer. N Engl J Med. 2001;344:1373-7.
85. Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. J Exp Med. 1965;121:439-62.
86. Gadducci A, Cosio S, Carpi A, Nicolini A, Genazzani AR. Serum tumor markers in the management of ovarian, endometrial and cervical cancer. Biomed Pharmacother 2004;58:24-38.
87 Alaoui-Jamali MA, Xu Y-j. Proteomic technology for biomarker profiling in cancer: an update. J Zhejiang Univ Sci B. 2006;7:411-20.
88. Bast RC, Jr., Urban N, Shridhar V, Smith D, Zhang Z, Skates S, et al. Early detection of ovarian cancer: promise and reality. Cancer Treat Res. 2002;107:61-97.
89. Meden H, Fattahi-Meibodi A. CA 125 in benign gynecological conditions. Int J Biol Markers. 1998;13:231-7.
90. Bast RC, Jr, Xu FJ, Yu YH, Barnhill S, Zhang Z, Mills GB. CA 125: the past and the future. Int J Biol Markers. 1998;13:179-87.
91. Koprowski H, Steplewski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P. Colorectal carcinoma antigens detected by hybridoma antibodies. Somatic Cell Genet. 1979;5:957-71.
92. Safi F, Roscher R, Bittner R, Schenkluhn B, Dopfer HP, Beger HG. High sensitivity and specificity of CA 19-9 for pancreatic carcinoma in comparison to chronic pancreatitis. Serological and immunohistochemical findings. Pancreas. 1987;2: 398-403.
93. Duraker N, Hot S, Polat Y, Hobek A, Gencler N, Urhan N. CEA, CA 19-9, and CA 125 in the differential diagnosis of benign and malignant pancreatic diseases with or without jaundice. J Surg Oncol. 2007;95:142-7.
94. Duffy MJ. CA 19-9 as a marker for gastrointestinal cancers: a review. Ann Clin Biochem. 1998;35:364-70.

95. Tumour markers in gynaecological cancers--EGTM recommendations. European group on tumor markers. Anticancer Res. 1999;19:2807-10.
96. Hill MD, Jackowski G, Bayer N, Lawrence M, Jaeschke R. Biochemical markers in acute ischemic stroke. CMAJ. 2000;162(8):1139-40.
97. Hainard A, Tiberti N, Robin X, Lejon V, Ngoyi DM, Matovu E, et al. A Combined cxcl 10, cxcl 8 and h-fabp panel for the staging of human African Trypanosomiasis patients. PLoS Negl Trop Dis. 2009;3(6):e459.
98. Mazzone PJ, Wang XF, Han X, Choi H, Seeley M, Scherer R, et al. Evaluation of a serum lung cancer biomarker panel. Biomark Insights. 2018;13:1-5.
99. Guo J, Yang J, Zhang X, Feng X, Zhang H, Chen L, et al. A panel of biomarkers for diagnosis of prostate cancer using urine samples. Anticancer Res. 2018;38(3):1471-7.
100. Yigitbasi T, Calibasi-Kocal G, Buyukuslu N, Atahan MK, Kupeli H, Yigit S, et al. An efficient biomarker panel for diagnosis of breast cancer using surface-enhanced laser desorption ionization time-of-flight mass spectrometry. Biomed Rep. 2018;8(3):269-74.
101. Siuzdak G. An Introduction to mass spectrometry ionization: an excerpt from the expanding role of mass spectrometry in biotechnology, 2nd ed.; mcc press: San Diego, 2005. JALA Charlottesv Va. 2004;9(2):50-63.
102. Bhardwaj C, Hanley L. Ion sources for mass spectrometric identification and imaging of molecular species. Nat Prod Rep. 2014;31(6):756-67.
103. Aebersold R, Mann M. Mass spectrometry-based proteomics.ature. 2003;422: 198-207.
104. Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. Nat Rev Drug Discov. 2003;2:140-50.
105. Liu C. The Application of SELDI-TOF-MS in clinical diagnosis of cancers. J Biomed Biotechnol. 2011;2011:1-6.
106. Karas M, Krüger R. Ion formation in MALDI: the cluster ionization mechanism. Chem Rev. 2003;103:427-40
107. Jurinke C, Oeth P, Boom D. MALDI-TOF mass spectrometry. Mol Biotechnol. 2004;26:147-63.
108. Wang C-C, Lai Y-H, Ou Y-M, Chang H-T, Wang Y-S. Critical factors determining the quantification capability of matrix-assisted laser desorption/ionization– time-of-flight mass spectrometry. Philos Trans A Math Phys Eng Sci. 2016;374:1-14.
109. Wang QT, Li YZ, Liang YF, Hu CJ, Zhai YH, Zhao GF, et al. Construction of a multiple myeloma diagnostic model by magnetic bead-based MALDI-TOF

mass spectrometry of serum and pattern recognition software. Anat Rec (Hoboken). 2009;292:604-10.

110. Zhang MH, Xu XH, Wang Y, Linq QX, Bi YT, Miao XJ, et al. A prognostic biomarker for gastric cancer with lymph node metastases. Anat Rec (Hoboken). 2013; 296:590-4.
111. Ahmed N, Oliva KT, Barker G, Hoffmann P, Reeve S, Smith IA, et al. Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer. Proteomics. 2005;5:4625-36.
112. Veenstra TD, Prieto DA, Conrads TP. Proteomic patterns for early cancer detection. Drug Discov Today. 2004;9:889-97.
113. Reyzer ML PJ, Allen JL, Chertov O, Kim DY, et al. MALDI MS profiles distinguish ER-negative breast bancers from lung adenocarcinoma. J Proteomics Bioinform. 2013:S6-9.
114. Chaiyarit P, Taweechaisupapong S, Jaresitthikunchai J, Phaonakrop N, Roytrakul S. Comparative evaluation of 5-15-kDa salivary proteins from patients with different oral diseases by MALDI-TOF/TOF mass spectrometry. Clin Oral Investig. 2015;19:729-37.
115. Valerio A, Basso D, Mazza S, Baldo G, Tiengo A, Pedrazzoli S, et al. Serum protein profiles of patients with pancreatic cancer and chronic pancreatitis: searching for a diagnostic protein pattern. Rapid Commun Mass Spectrom. 2001;15: 2420-5.
116. Chung L, Moore K, Phillips L, Boyle FM, Marsh DJ, Baxter RC. Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. Breast Cancer Res. 2014;16(3):R63.
117. Ying X, Han SX, Wang JL, Zhou X, Jin GH, Jin L, et al. Serum peptidome patterns of hepatocellular carcinoma based on magnetic bead separation and mass spectrometry analysis. Diagn Pathol. 2013;8:130.
118. Wu S, Xu K, Chen G, Zhang J, Liu Z, Xie X. Identification of serum biomarkers for ovarian cancer using MALDI-TOF-MS combined with magnetic beads. Int J Clin Oncol. 2012;17(2):89-95.
119. Streitz JM, Jr., Madden MT, Marimanikkuppam SS, Krick TP, Salo WL, Aufderheide AC. Analysis of protein expression patterns in Barrett's esophagus using MALDI mass spectrometry, in search of malignancy biomarkers. Dis Esophagus. 2005;18(3):170-6.
120. Lavine BK, White CG, DeNoyer L, Mechref Y. Multivariate classification of disease phenotypes of esophageal adenocarcinoma by pattern recognition analysis of

MALDI-TOF mass spectra of serum N-linked glycans. Microchem J. 2017;132: 83-8.

121. Sandanayake NS, Camuzeaux S, Sinclair J, Blyuss O, Andreola F, Chapman MH, et al. Identification of potential serum peptide biomarkers of biliary tract cancer using MALDI MS profiling. BMC Clin Patho. 2014;14(1):1-11.
122. Wang H, Luo C, Zhu S, Fang H, Gao Q, Ge S, et al. Serum peptidome profiling for the diagnosis of colorectal cancer: discovery and validation in two independent cohorts. Oncotarget. 2017;8(35):59376-86.
123. Schwamborn K, Krieg RC, Uhlig S, Ikenberg H, Wellmann A. MALDI imaging as a specific diagnostic tool for routine cervical cytology specimens. Int J Mol Med. 2011;27(3):417-21.
124. Liu C, Pan C, Shen J, Wang H, Yong L, Zhang R. Discrimination analysis of mass spectrometry proteomics for cervical cancer detection. Med Oncol. 2011;28 Suppl 1:S553-9.
125. Hajduk J, Matysiak J, Kokot ZJ. Challenges in biomarker discovery with MALDI-TOF MS. Clin Chim Acta. 2016;458:84-98.
126. Maitz MF. Applications of synthetic polymers in clinical medicine. Biosurf Biotribol. 2015;1(3):161-76.
127. Drake SK, Bowen RA, Remaley AT, Hortin GL. Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. Clin Chem. 2004;50:2398-401.
128. Pusch W, Kostrzewa M. Application of MALDI-TOF mass spectrometry in screening and diagnostic research. Curr Pharm Des. 2005;11(20):2577-91.
129. Hammer B, Strickert M, Villmann T. Supervised neural gas with general similarity measure. Neural Process Lett. 2003;21(1):21–44.
130. Walpole RE, Mayers RH, Mayers SL, Ye KE. Probability and statistics for engineers and scientists, 5th edition: Macmillan. 1993.
131. Chambers J, Hastie T, Pregibon D. Statistical models in s. Compstat.1990:317-321
132. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583-621.
133. Brioude G, Brégeon F, Trousse D, Flaudrops C, Secq V, De Dominicis F, et al. Rapid diagnosis of lung tumors, a feasability study using maldi-tof mass spectrometry. PLoS One. 2016;11(5):e0155449.
134. Walterová Z, Horský J. Quantification in MALDI-TOF mass spectrometry of modified polymers. Anal Chim Acta. 2011;693:82-8.

135. Myers BK, Zhang B, Lapucha JE, Grayson SM. The characterization of dendronized poly(ethylene glycol)s and poly(ethylene glycol) multi-arm stars using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chim Acta. 2014;808:175-89.
136. Bowen RAR, Remaley AT. Interferences from blood collection tube components on clinical chemistry assays. Biochem Med. 2014;24:31-44.
137. D. SW, E. MD. Effect of poly(ethylene glycol) on the hydrophobic interactions and rheology of proanthocyanidin biopolymers from Pinus radiata. J Appl Polym Sci. 2005;97(3):1254-60.
138. Boone C, Adamec J. 10-top-down proteomics. proteomic profiling and analytical chemistry (Second Edition). Boston: Elsevier; 2016.175-91.
139. Smolira A, Wessely-Szponder J. Importance of the matrix and the matrix/sample ratio in MALDI-TOF-MS analysis of cathelicidins obtained from porcine neutrophils. Appl Biochem Biotechnol. 2015;175(4):2050-65.
140. Graterol F, Navarro-Muñoz M, Ibernon M, López D, Troya MI, Pérez V, et al. Poor histological lesions in IgA nephropathy may be reflected in blood and urine peptide profiling. BMC Nephrol. 2013;14:82.
141. Li Z, Tang C, Li X, Li J, Wang W, Qin H, et al. Detection and significance of small-cell lung cancer serum protein markers using MALDI-TOF-MS. Int J Clin Exp Med. 2017.929-36.
142. Lu HB, Zhou JH, Ma YY, Lu HL, Tang YL, Zhang QY, et al. Five serum proteins identified using SELDI-TOF-MS as potential biomarkers of gastric cancer. Jpn J Clin Oncol. 2010;40(4):336-42.
143. Li P, Ma D, Zhu ST, Tang XD, Zhang ST. Serum peptide mapping in gastric precancerous lesion and cancer. J Dig Dis. 2014;15(5):239-45.
144. Padoan A, Seraglia R, Basso D, Fogar P, Sperti C, Moz S, et al. Usefulness of MALDI-TOF/MS identification of low-MW fragments in sera for the differential diagnosis of pancreatic cancer. Pancreas. 2013;42(4):622-32.
145. Cai SR, Yu JK, Jiang WZ, Zhang SZ, Zheng S. [Application of serum protein markers to distinguish familial adenomatous polyposis (FAP) and sporadic colorectal adenomas. Zhonghua Zhong Liu Za Zhi. 2009;31(3):192-5.
146. Xianyin Lai FAW, Suthat Liangpunsakul. Charecterization of peptides and low molecular weight proteins in plasma from subjects with hepatocellular carcinoma. Proteomics. 2014;1:1-6.

147. Thiede B, Wittmann-Liebold B, Bienert M, Krause E. MALDI-MS for c-terminal sequence determination of peptides and proteins degraded by carboxypeptidase Y and P. FEBS Letters. 1995;357(1):65-9.
148. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271-84.
149. Villanueva J, Martorella AJ, Lawlor K, Philip J, Fleisher M, Robbins RJ, et al. Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. Mol Cell Proteomics. 2006;5:1840-52.

APPENDIX A

REAGENTS FOR EXPERIMENTS

CTC

$CuSO_4 \cdot 5 H_2O$	0.1 g
Tatalic acid	0.2 g
dH_2O	50 ml

20% Na_2CO_3

Na_2CO_3	10 g
dH_2O	50 ml

0.8 N NaOH

NaOH	1.6 g
dH_2O	50 ml

5% SDS

SDS	2.5 g
dH_2O	50 ml

Lowry reagent A

CTC	5 ml
20% Na_2CO_3	5 ml
0.8 N NaOH	10 ml
5% SDS	20 ml

Lowry reagent B

Folin-Ciocalteu phenol reagent	1 ml
dH_2O	5 ml

CHCA matrix solution

CHCA	10 mg
Organic solvent (0.1%TFA/100%ACN (ratio 1:1))	1 ml

APPENDIX B

RESULTS

The individual cases were obtained from the Songklanakarin hospital. The samples are divided into 5 groups (healthy subjects, cervical cancer patients with precancerous, stage I, II and III). The individual cases were pooled for optimizing experimental phase and creating the database as shown in Table 11-14.

Table 11. Individual cases for pooled stage serum (optimal condition phase).

No.	Healthy		CIN		Early		late	
	Case	Age	Case	Age	Case	Age	Case	Age
1	NC42	36	CCA2	43	CCA7	50	CCA13	39
2	NC43	42	CCA43	40	CCA14	45	CCA26	38
3	NC44	35	CCA59	47	CCA27	46	CCA36	46
4	NC45	49	CCA77	53	CCA34	50	CCA56	50
5	NC47	33	CCA89	35	CCA78	52	CCA60	39
6	NC49	42	CCA97	48	CCA81	32	CCA87	44
7	NC50	50	CCA119	38	CCA98	51	CCA113	34
8	NC51	37	CCA121	36	CCA103	35	CCA127	38
9	NC53	35	CCA122	51	CCA129	34	CCA135	49
10	NC54	53	CCA167	36	CCA130	33	CCA137	56
11	NC55	51	CCA191	30	CCA131	54	CCA142	61
12	NC56	43	CCA194	54	CCA139	44	CCA146	58
13	NC61	31	CCA201	32	CCA169	42	CCA170	40
14	NC63	34	CCA9	36	CCA172	51	CCA205	52
15	NC64	33	CCA10	56	CCA177	38	CCA206	32
16	NC65	35	CCA23	32	CCA19	58	CCA120	52
17	NC66	31	CCA29	49	CCA21	32		
18	NC67	42	CCA33	59	CCA40	53		
19	NC68	44	CCA38	49	CCA53	37		
20	NC71	43	CCA61	35	CCA71	65		
21	NC72	43	CCA63	48	CCA72	47		
22	NC73	46	CCA65	44	CCA74	33		
23	NC74	49	CCA91	35	CCA82	57		
24	NC75	40	CCA95	36	CCA101	34		
25	NC76	34	CCA140	54	CCA141	45		
26	NC78	60	CCA145	57	CCA149	31		
27	NC79	43	CCA90	61	CCA173	42		
28	NC83	42	CCA158	51	CCA186	43		
29	NC84	35	CCA160	41	CCA189	56		
30	NC86	56	CCA164	45	CCA163	61		
	average	42	average	44	average	45	average	46
	max	60	max	61	max	65	max	61
	min	31	min	30	min	31	min	32
	mode	42	mode	36	mode	50	mode	39
	median	42	median	44	median	45	median	45
	SD	7.65	SD	9.14	SD	9.67	SD	8.82

Table 12. Individual cases for pooled stage serum (experimental phase: training dataset).

No.	Healthy		CIN		Stage I		Stage II		Stage III	
	Case	Age	Case	Age	Case	Age	Case	Age	Case	Age
1	NC42	36	CCA2	43	CCA7	50	CCA19	58	CCA13	39
2	NC43	42	CCA43	40	CCA14	45	CCA21	32	CCA26	38
3	NC44	35	CCA59	47	CCA27	46	CCA40	53	CCA36	46
4	NC45	49	CCA77	53	CCA34	50	CCA53	37	CCA56	50
5	NC47	33	CCA89	35	CCA78	52	CCA71	65	CCA60	39
6	NC49	42	CCA97	48	CCA81	32	CCA72	47	CCA87	44
7	NC50	50	CCA119	38	CCA98	51	CCA74	33	CCA113	34
8	NC51	37	CCA121	36	CCA103	35	CCA82	57	CCA127	38
9	NC53	35	CCA122	51	CCA129	34	CCA101	34	CCA135	49
10	NC54	53	CCA167	36	CCA130	33	CCA141	45	CCA137	56
11	NC55	51	CCA191	30	CCA131	54	CCA149	31	CCA142	61
12	NC56	43	CCA194	54	CCA139	44	CCA173	42	CCA146	58
13	NC61	31	CCA201	32	CCA169	42	CCA186	43	CCA170	40
14	NC63	34	CCA9	36	CCA172	51	CCA189	56	CCA205	52
15	NC64	33	CCA10	56	CCA177	38	CCA163	61	CCA206	32
16	NC65	35	CCA23	32						
17	NC66	31	CCA29	49						
18	NC67	42	CCA33	59						
19	NC68	44	CCA38	49						
20	NC71	43	CCA61	35						
21	NC72	43	CCA63	48						
22	NC73	46	CCA65	44						
23	NC74	49	CCA91	35						
24	NC75	40	CCA95	36						
25	NC76	34	CCA140	54						
26	NC78	60	CCA145	57						
27	NC79	43	CCA90	61						
28	NC83	42	CCA158	51						
29	NC84	35	CCA160	41						
30	NC86	56	CCA164	45						
	average	42	average	44	average	44	average	46	average	45
	max	60	max	61	max	54	max	65	max	61
	min	31	min	30	min	32	min	31	min	32
	mode	42	mode	36	mode	50	mode	-	mode	39
	median	42	median	44	median	45	median	45	median	44
	SD	7.65	SD	9.14	SD	7.67	SD	11.47	SD	8.96

Table 13. Individual cases for pooled stage serum (experimental phase: validation dataset).

No.	Healthy		CIN		Stage I		Stage II		Stage III	
	Case	Age	Code	Age	Code	Age	Code	Age	Code	Age
1	NC46	50	CCA66	36	CCA57	32	CCA168	61	CCA1	56
2	NC48	32	CCA111	40	CCA70	34	CCA4	50	CCA3	43
3	NC52	55	CCA138	37	CCA108	43	CCA8	49	CCA5	45
4	NC57	46	CCA202	54	CCA150	59	CCA12	45	CCA11	48
5	NC58	43	CCA6	45	CCA174	48	CCA15	58	CCA16	55
6	NC59	58	CCA30	30	CCA175	44	CCA22	49	CCA18	55
7	NC69	50	CCA31	55	CCA42	59	CCA25	52	CCA32	57
8	NC70	50	CCA45	33	CCA69	65	CCA35	58	CCA41	63
9	NC77	47	CCA76	43	CCA152	57	CCA37	48	CCA110	50
10	NC82	55	CCA92	48			CCA48	44	CCA155	57
11	NC85	45	CCA100	30			CCA49	50	CCA171	47
12	NC87	59	CCA102	56			CCA55	54	CCA196	49
13	NC88	48	CCA114	48			CCA86	54		
14	NC89	35	CCA115	35			CCA93	60		
15	NC90	41	CCA125	61			CCA117	50		
16	NC91	30	CCA136	56			CCA134	49		
17	NC92	33	CCA162	51			CCA147	58		
18	NC93	40	CCA183	49			CCA181	57		
19	NC94	40	CCA188	33			CCA182	44		
20	NC95	32	CCA193	41			CCA195	39		
21	NC96	44	CCA204	49			CCA197	57		
22	NC97	41					CCA203	57		
23	NC98	33								
24	NC99	30								
25	NC100	35								
26	NC101	43								
27	NC102	34								
28	NC103	38								
29	NC104	39								
30	NC105	49								
31	NC106	53								
32	NC107	35								
33	NC108	44								
34	NC110	38								
35	NC111	39								
36	NC112	42								
37	NC113	36								
38	NC114	52								
39	NC115	39								
40	NC116	57								
41	NC117	30								
42	NC118	34								
43	NC119	36								
44	NC120	31								
45	NC121	32								
46	NC122	41								
47	NC123	48								
48	NC124	38								
49	NC125	44								
50	NC126	33								
51	NC127	38								
52	NC128	41								
53	NC129	46								
	average	42	average	44	average	49	average	52	average	52
	max	59	max	61	max	65	max	61	max	63
	min	30	min	30	min	32	min	39	min	43
	mode	41	mode	30	mode	59	mode	50	mode	55
	median	41	median	45	median	48	median	51	median	52.5
	SD	7.91	SD	9.44	SD	11.70	SD	5.92	SD	5.95

Table 14. Individual cases for pooled stage serum (database generation)

No.	Group 1 (Healthy subjects)		Group 2 (Precancerous set 1)				Group 3 (Precancerous set 2)				Group 4 (Stage I & II set 1)				Group 4 (Stage II set 2 & III)			
	Case number	Age	Case number	Age	Stage	Histology	Case number	Age	stage	histology	Case number	Age	stage	histology	Case number	Age	stage	histology
1	NC 42	36	CCA 29	49	HSIL	Adenosquamous carcinoma	CCA 23	32	HSIL	Squamous cell carcinoma	CCA 4	50	IIB	Squamous cell carcinoma	CCA 1	56	IIIB	Squamous cell carcinoma
2	NC 43	42	CCA 89	35	LSIL	Squamous cell carcinoma	CCA 77	53	LSIL	Squamous cell carcinoma	CCA 8	49	IIB	Squamous cell carcinoma	CCA 5	45	IIIB	Squamous cell carcinoma
3	NC 48	32	CCA 119	38	LSIL	Squamous cell carcinoma	CCA 100	30	HSIL	Squamous cell carcinoma	CCA 22	49	IIB	Squamous cell carcinoma	CCA 12	45	IIB	-
4	NC 56	43	CCA 121	36	LSIL	Squamous cell carcinoma	CCA 114	48	HSIL	Squamous cell carcinoma	CCA 25	52	IIB	Adenosquamous carcinoma	CCA 16	55	IIIB	Squamous cell carcinoma
5	NC 64	33					CCA 122	51	LSIL	Squamous cell carcinoma	CCA 27	46	IB	Squamous cell carcinoma	CCA 32	57	IIIB	Squamous cell carcinoma
6	NC 65	35					CCA 204	49	HSIL	Adenosquamous carcinoma	CCA 34	50	IA	Squamous cell carcinoma	CCA 36	46	IIIB	Squamous cell carcinoma
7	NC 66	31									CCA 37	48	IIB	Squamous cell carcinoma	CCA 40	53	IIB	Squamous cell carcinoma
8	NC 74	49									CCA 49	50	IIB	Squamous cell carcinoma	CCA 60	39	IIIB	Adenosquamous carcinoma
9	NC 75	40									CCA 55	54	IIB	Squamous cell carcinoma	CCA 72	47	IIB	Squamous cell carcinoma
10	NC 82	55									CCA 71	65	IIB	Squamous cell carcinoma	CCA 82	57	IIB	Squamous cell carcinoma
11	NC 83	42									CCA 81	32	IA	Adenocarcinoma	CCA 86	54	IIB	Squamous cell carcinoma
12	NC 89	35									CCA 117	50	IIB	Squamous cell carcinoma	CCA 93	60	IIB	Adenosquamous carcinoma
13	NC 99	30									CCA 130	33	IB	Squamous cell carcinoma	CCA 134	49	IIB	Squamous cell carcinoma
14	NC 100	35									CCA 131	54	IA	Adenosquamous carcinoma	CCA 149	31	IIB	-
15	NC 101	43									CCA 108	43	IA	Squamous cell carcinoma	CCA 155	57	IIIB	Squamous cell carcinoma
16	NC 106	53									CCA 150	59	IA	Adenosquamous carcinoma	CCA 163	61	IIA	Squamous cell carcinoma
17	NC 108	44									CCA 152	57	IB	Squamous cell carcinoma	CCA 171	47	IIIB	Adenocarcinoma
18	NC 110	38									CCA 168	61	IIA	Squamous cell carcinoma	CCA 173	42	IIB	-
19	NC 115	39									CCA 174	48	IA	Squamous cell carcinoma	CCA 181	57	IIB	-
20	NC 116	57													CCA 195	39	IIB	Squamous cell carcinoma
21	NC 122	41													CCA 196	49	IIIB	Squamous cell carcinoma
22	NC 124	38													CCA 197	57	IIB	Squamous cell carcinoma
23	NC 126	33													CCA 206	32	IIIB	Squamous cell carcinoma
24	NC 127	38																
25	NC 128	41																
	average	40	average	40	LSIL	3	average	44	LSIL	2	average	50	IA	6	average	49	IIA	1
	max	57	max	49	HSIL	1	max	53	HSIL	4	max	65	IB	3	max	61	IIB	12
	min	30	min	35	Adeno	0	min	30	Adeno	0	min	32	IIA	1	min	31	IIIA	0
	mode	35	mode	-	AdenoSCCA	1	mode	-	AdenoSCCA	1	mode	50	IIB	9	mode	57	IIIB	10
	median	39	median	37	SCCA	3	median	48.5	SCCA	5	median	50	Adeno	1	median	49	Adeno	1
	SD	7.23	SD	6.45			SD	10.11			SD	8.16	AdenoSCCA	3	SD	8.58	AdenoSCCA	2
	count	25	count	4			count	6			count	19	SCCA	15	count	23	SCCA	16

Figure 25. Dendrogram of each investigation group (healthy, precancerous group1, precancerous group 2, stage I&II and stage II& III) were shown in Figure 25 A-E, respectively.

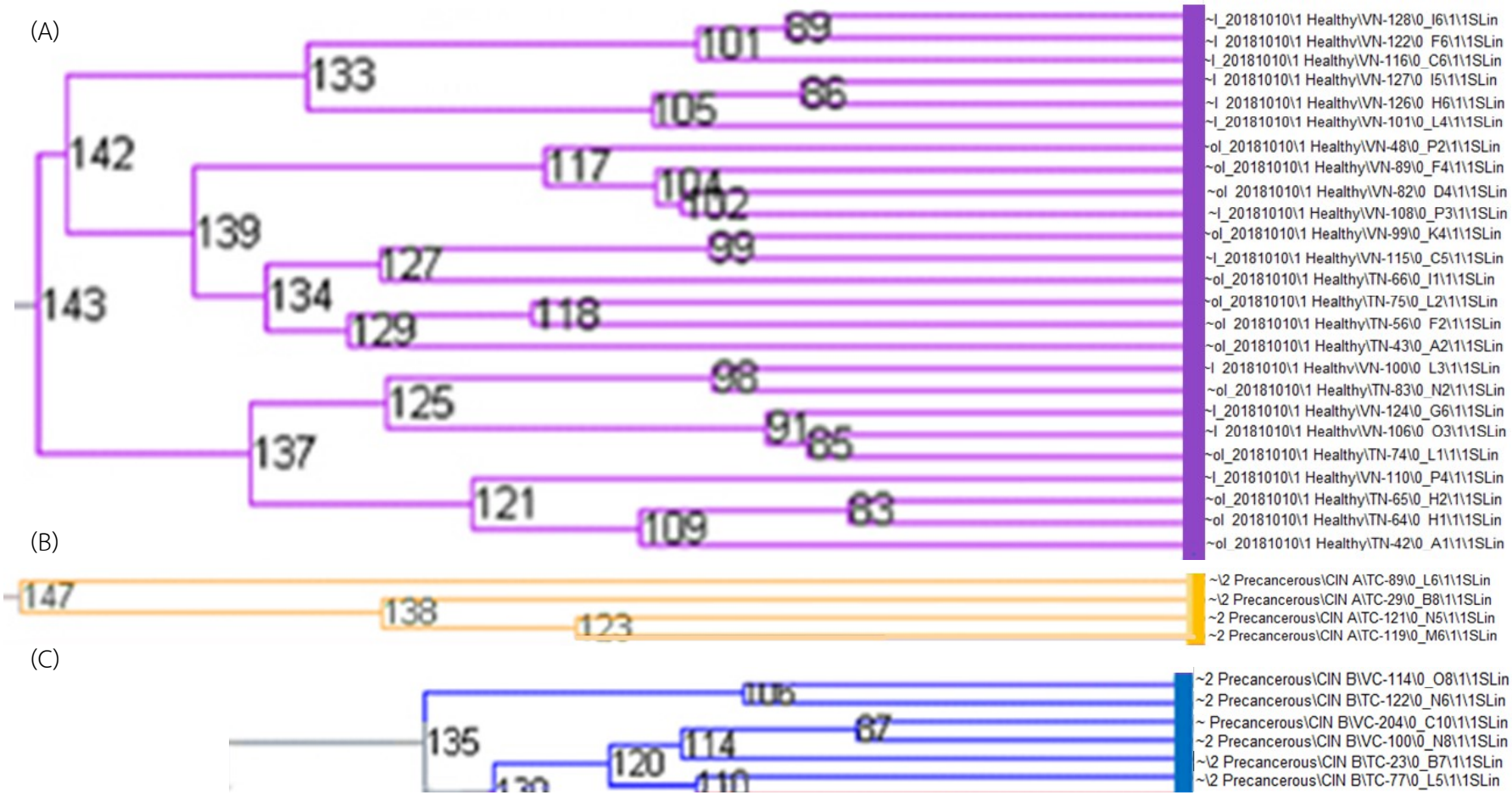
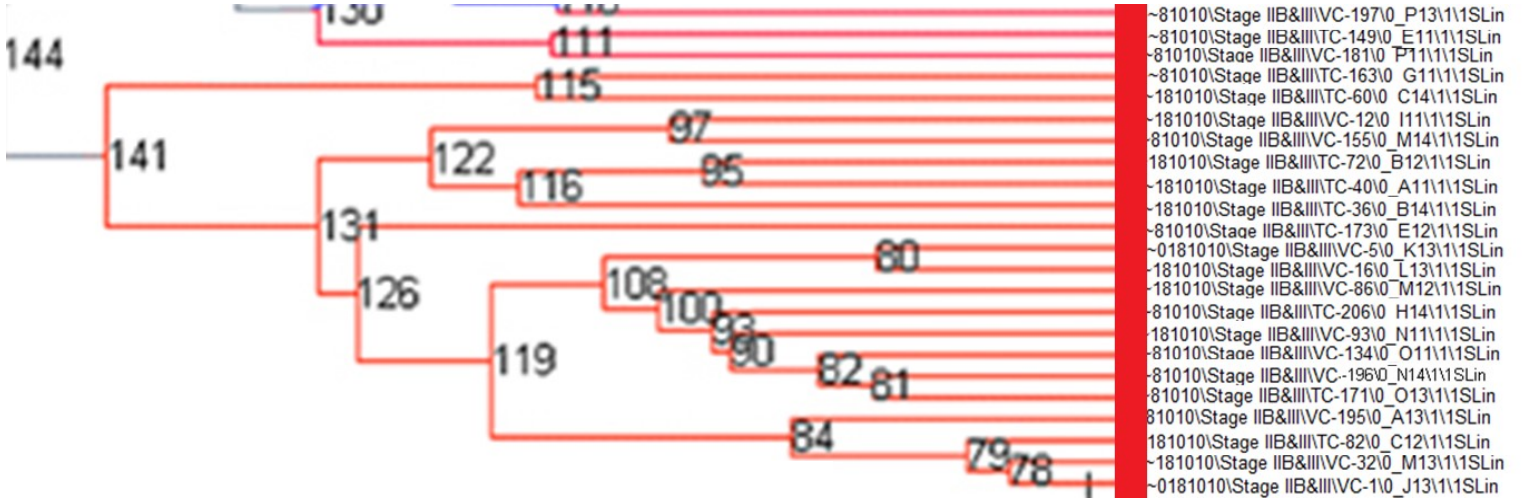


Figure 25. (Continued)

(D)

(E)



VITAE

Name Miss Phetploy Rungkamoltip

Student ID 5810320017

Educational Attainment

Degree	**Name of Institution**	**Year of Graduation**
Bachelor of science (Medical Technology)	Prince of Songkla University, Thailand	2015

Scholarship Awards During Enrolment

Research Grant, Faculty of Medicine, Prince of Songkla University

Graduate School Prince of Songkla University scholarship

List of Publication and Proceedings

1. **Rungkamoltip P**, Hanprasertpong J, Jaresitthikunchai J, Roytrakul S, and Navakanitworakul R. MALDI-TOF peptide signature in patient serum with cervical cancer. Poster presentation at 5th Asia Pacific Protein Association Conference/ 12th International Symposium of the Protein Society of Thailand 2017. July 11- 14, 2017. Bangsean, Chonburi, Thailand (Poster presentation)
2. **Rungkamoltip P**, Hanprasertpong J, Jaresitthikunchai J, Roytrakul S, and Navakanitworakul R. Utilization of MALDI-MS based serum peptide patterns to distinguish cervical cancer patients from healthy women. Oral presentation at the 33rd conference of the faculty of Medicine 2017. August 2-4, 2017. Faculty of Medicine, Prince of Songkla University, Thailand (Oral presentation)