SECRETOME MAPPING OF THE HUMAN GUT AND MICROBIOME FROM COLORECTAL CANCER PATIENTS

PUTRI INTAN HAFIZAH BT MEGAT MOHD AZLAN

UNIVERSITI KEBANGSAAN MALAYSIA

SECRETOME MAPPING OF THE HUMAN GUT AND MICROBIOME FROM COLORECTAL CANCER PATIENTS

PUTRI INTAN HAFIZAH BT MEGAT MOHD AZLAN

THESIS SUBMITTED IN FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE

UKM MEDICAL MOLECULAR BIOLOGY INSTITUTE
UNIVERSITI KEBANGSAAN MALAYSIA
KUALA LUMPUR

2019

PEMETAAN SEKRETOM USUS MANUSIA DAN MIKROBIOME DARIPADA PESAKIT KANSER KOLOREKTAL

PUTRI INTAN HAFIZAH BT MEGAT MOHD AZLAN

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH
IJAZAH SARJANA SAINS

INSTITUT PERUBATAN BIOLOGI MOLEKUL UKM
UNIVERSITI KEBANGSAAN MALAYSIA
KUALA LUMPUR

2019

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

24 June 2019

PUTRI INTAN HAFIZAH
BT MEGAT MOHD AZLAN
P83382

PERAKUAN TESIS SARJANA / DOKTOR FALSAFAH
(CERTIFICATION OF MASTERS / DOCTORAL THESIS)

Nama Penuh Pengarang
(Author's Full Name) : PUTRI INTAN HAFIZAH MEGAT MOHD AZLAN

No. Pendaftaran Pelajar
(Student's Registration No.) : P83382

Sesi Akademik
(Academic Session) : 2018/2019

Tajuk Tesis
(Thesis Title) : SECRETOME MAPPING OF THE HUMAN GUT AND MICROBIOME FROM COLORECTAL CANCER PATIENTS

Merujuk kepada Klausa 4.2 Dasar Harta Intelek Pelajar UKM (Tambahan), tesis adalah hak milik pelajar. Saya mengaku tesis ini sebagai:
(With regard to Clause 4.2 of the UKM Student Intellectual Property Policy (Supplementary), the thesis is the student's property. I hereby declare this thesis as:)

☐	**RAHSIA (*CONFIDENTIAL*)**	Mengandungi maklumat rahsia di bawah AKTA RAHSIA RASMI 1972 *(Consisting of classified information under the OFFICIAL SECRETS ACT 1972)*
☐	**TERHAD (*RESTRICTED*)**	Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan *(Consisting of RESTRICTED information which has been determined by the organisation/body where the research was conducted)*
☑	**AKSES TERBUKA /TIDAK TERHAD (*OPEN ACCESS/ NON-RESTRICTED*)**	Saya membenarkan tesis ini diterbitkan secara akses terbuka, teks penuh atau dibuat salinan untuk tujuan pengajian, pembelajaran, penyelidikan sahaja. *(I allow this thesis to be published through open access, full text or copied for study, learning and research purposes only.)*

Bagi kategori Akses Terbuka/Tidak Terhad, saya membenarkan tesis (Sarjana/Doktor Falsafah) ini di simpan di Perpustakaan Universiti Kebangsaan Malaysia (UKM)* dengan syarat-syarat kegunaan seperti berikut:
(For the Open Access/Non-Restricted category, I allow this (Master's/Doctoral) Thesis to be kept in the Universiti Kebangsaan Malaysia (UKM) Library with the following usage conditions:)

1. Perpustakaan UKM mempunyai hak untuk membuat salinan untuk tujuan pengajian, pembelajaran, penyelidikan sahaja.
   (UKM Library has the right to reproduce the thesis for study, learning and research purposes only.)
2. Perpustakaan Universiti Kebangsaan Malaysia dibenarkan membuat satu (1) salinan tesis ini untuk tujuan pertukaran antara institusi pengajian tinggi dan mana-mana badan/ agensi kerajaan, tertakluk kepada terma dan syarat.
   (UKM Library is allowed to make one (1) copy of this thesis for exchange purpose among higher education institutions and any government body/agency, subject to terms and conditions.)

DISAHKAN OLEH:
(VERIFIED BY:)

TANDATANGAN PELAJAR
(STUDENT'S SIGNATURE)

TANDATANGAN PENYELIA /
PENGERUSI JK SISWAZAH
(SUPERVISOR'S / CHAIRPERSON SUPERVISION COMMITTEE SIGNATURE)

910820-14-6180
KAD PENGENALAN /
NO. PASPORT
(IDENTITY CARD/PASSPORT NO.)

DR. CHIN SIOK FONG
NAMA PENYELIA/
PENGERUSI JK SISWAZAH
(SUPERVISOR'S /CHAIRPERSON SUPERVISION COMMITTEE NAME)

Tarikh/ *Date:* 24-06-2019

Tarikh/ *Date:* 24-06-2019

ACKNOWLEDGEMENT

All praise to Almighty Allah for all his blessings and for giving me the patience and good health to complete this Master research. Alhamdullilah. All works and analysis presented in this thesis were carried out at UKM Medical Molecular Biology Institute (UMBI). First of all, I would like to gratefully acknowledge financial support from the Trans-Disciplinary Research Grant Scheme (TRGS) TRGS/2/2014/UKM/02/3/2 from Ministry of Higher Education (MOHE).

I would like to extend my deepest gratitude to my supervisor, Dr Chin Siok Fong, who has accepted me as one of her student and exposed me to the world of proteomics and secretomics. For the past three years, she has been an incredible mentor and supervised my work closely. I appreciate her generous guidance, support and patience throughout my research works, manuscript and thesis writing. Without her guidance, it would be difficult for me to complete this thesis.

My heartfelt appreciation to my co-supervisor, Prof Datuk Dr A. Rahman A. Jamal for his intellectual contribution, motivation and enthusiasm in scientific research. I am thankful for all his critical thoughts and advices from beginning to end of this study. I also would like to thank Assoc. Prof Dr Neoh Hui Min for her advices and willingness to share her bright thoughts during our research team meetings, which were very fruitful for shaping my research outcomes.

In addition, I wish to thank Mr Luqman Mazlan from Department of Surgery, UKM Medical Centre (UKMMC), Dr Ngiu Chai Soon from Department of Medicine, UKMMC, Mr Andee Dzulkarnaen Zakaria from Department of Surgery, University of Science Malaysia Hospital (HUSM) and all the staffs at Endoscopy Unit, UKMMC who had given me indispensable help during my stool sample collection. It was a pleasure to work with them.

I also would like to dedicate a special thanks to my beloved parents, Megat Mohd Azlan and Aabasiyah Tengku Ali, my dear sister Putri Intan Mawadah and my love one, Muhamad Hafiz for their encouragement, support, patience and understanding. Their endless love is the great force that has kept me going throughout these difficult years. May Allah bless them all.

Last but not least, I gratefully acknowledge UMBI staffs, Biobank team, lab mates and friends for their help and support in creating a pleasant working environment throughout my years in UMBI.

ABSTRAK

Di Malaysia, kanser kolorektal (CRC) merupakan kanser kedua paling biasa dihidapi dalam kalangan lelaki dan wanita. Usus manusia dihuni oleh trilion flora usus dalam komposisi yang seimbang bagi membantu hos memelihara homeostasis usus dan saling memberi manfaat antara satu sama lain. Namun, perubahan drastik ke atas komposisi mikrob boleh mengganggu homeostasis usus lalu menyebabkan kewujudan persekitaran mikro tumor yang mendoroang kepada pembentukan kanser kolorektal (CRC). Sehingga kini, kajian terhadap sekretom usus manusia masih belum dilaksanakan. Objektif bagi kajian ini adalah, memprofil protein yang dirembeskan oleh usus manusia dan mikrob daripada pesakit CRC dan individu kawalan yang mempunyai morfologi kolon yang sihat menggunakan teknologi spektrometri jisim. Sampel najis daripada 26 pesakit CRC dan 20 individu kawalan telah dikumpul, dihomogen dan ditapis sebelum proses pengekstrakan dan analisis protein. Sampel seterusnya dicerna dalam bentuk larutan diikuti dengan pengenalpastian dan pengkuantitian protein. Perisian bioinformatik seperti SPSS, MaxQuant, DAVID dan String telah digunakan bagi analisis statistik, pemvisualan data, anotasi fungsian dan ramalan interaksi serta laluan protein. Hasil kajian mendapati bahawa lebih banyak protein manusia telah dikenalpasti bagi sampel CRC berbanding kawalan, dan sebaliknya bagi protein mikrob. Kebanyakan protein manusia yang telah dikenalpasti sebagai protein eksklusif terhadap CRC berperanan dalam sel lekatan, transduksi isyarat intrasel dan proses angiogenesis. Manakala, ekspresi protein yang tertinggi telah dipetakan pada sampel CRC Tahap 1 dan 2. Model ramalan terbaik bagi CRC dihasilkan daripada kombinasi protein manusia Huntingtin dan RNA eksonuklease 5. Model ini adalah sensitif tetapi tidak spesifik dalam membezakan kawalan berbanding CRC. Manakala, laluan KEGG beranotasi yang teratas bagi protein manusia-eksklusif CRC ialah laluan Faktor Teraruh Hipoksia-1 (HIF-1). Bagi protein mikrob yang eksklusif untuk CRC pula, protein daripada yis menunjukkan ekspresi tertinggi dengan ramalan interaksi protein dipetakan pada peranan yang melibatkan pembaikpulihan DNA, regulasi transkripsi dan pengikatan ATP. Kesimpulannya, flora usus dan kolon manusia merembeskan banyak protein ke persekitaran luar sel dan berkemungkinan terlibat dalam pelbagai proses reaksi dan respons hos-mikrob dalam CRC.

ABSTRACT

In Malaysia, colorectal cancer (CRC) was the second most common cancer among men and women. The human gut is home to trillions of gut flora that thrive in delicate balance, which has helped maintain host's gut homeostasis and mutually benefited both parties tremendously. However, a drastic perturbation of microbial composition has hampered gut homeostasis initiating tumour microenvironment for development of CRC. Hitherto, the study of human gut secretome is still scarce with gap of knowledge to be filled. The objective of this study was to profile secreted proteins released from the human gut and microbial of CRC patients and control with healthy colon morphology by assessing the secretome in stool samples, using mass spectrometry technology. Stool samples from 26 CRC and 20 controls were collected, homogenized and filtered prior to protein extraction and analysis. Samples were subjected for in-solution digestion, followed by protein identification and quantification. Bioinformatics tools such as SPSS, MaxQuant, DAVID and String were used for statistical analysis, data visualization, functional annotations and prediction of protein interactions and pathways. We identified more human origin proteins in CRC as compared to control and inversely for proteins from microbial origin. The identified human proteins that are exclusive for CRC were mostly related to cell adhesion, intracellular signal transduction and angiogenesis functions and the top identified proteins were mapped to Stage I and II of CRC. The best prediction model for CRC was built upon the combination of human Huntingtin and RNA exonuclease 5 proteins. The model was sensitive but not specific in discriminating the control from CRC. Meanwhile, the top annotated KEGG pathway for human CRC-exclusive proteins was Hypoxia-inducible factor-1 (HIF-1). In addition, yeast proteins were topping the microbial CRC-exclusive proteins list, with the predicted protein interactions mapped to DNA repair, transcription regulation and ATP binding. In conclusion, gut flora and human colon released abundance of microbial proteins to the external environment possibly mediating various host-microbe reactions and responses in CRC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ASR	Age-Specific Rates
AUC	Area Under The Curve
BSS	Bristol Stool Scale
CEA	Carcinoembryonic Antigen
CRC	Colorectal Cancer
CagA	Cytotoxin-associated gene A protein
DAVID	David Functional Annotation Tool
dH_2O	Distilled water
FDR	False Discovery Rate
FIT	Fecal Immunochemical Testing
GI	Gastrointestinal Tract
GO	Gene Ontology
gFOBT	guaiac-based Fecal Occult Blood Test
IBD	Inflammatory Bowel Diseases
MS	Mass Spectrometry
NPV	Negative Predictive Value
PLS-DA	Partial Least Square Discriminant Analysis
PSM	Peptide Spectral Match
PPV	Positive Predictive Value
PCA	Principal Component Analysis
ROS	Reactive Oxygen Species
RT	Room Temperature
SES	Socioeconomic Status
UKM	Universiti Kebangsaan Malaysia

CHAPTER I

INTRODUCTION

1.1 RESEARCH BACKGROUND

Colorectal cancer (CRC) is a tumor that develops in the colon or rectum slowly over the years, transition from normal to adenomatous polyps and further progress to benign adenomas before becoming invasive and metastasized (Vargas and Thompson 2012). In Malaysia, the incidence rate of CRC was ranked second after breast cancer and in the world, CRC placed third among men and women worldwide with mortality rates reached up to ~ 690,000 people per year (Ervik et al. 2016; Ferlay et al. 2014). In 2017, about ~ 136,00 new CRC cases and about 50,000 deaths are estimated in United States (American Cancer Society 2017a).

Over the years, the death rate of CRC has been declining in both men and women resulted from early screening program and improved treatments (American Cancer Society 2017a). However, more deaths from CRC were reported in less-developed areas due to socioeconomic difference thus, slowed down success rate of the program (Ferlay et al. 2014). Several factors that may increase one's chances of developing CRC include imbalanced diets, lack of physical activity, obesity, smoking, excessive alcohol intake, aging, genetics alteration, history of CRC in first degree relatives, gastrointestinal tract (GI) inflammation and dysbiosis of gut microbiome (Ghee 2014; Johnson et al. 2013; Ma et al. 2013; Zackular et al. 2014).

Human gut especially the large intestine, hosted trillions of microorganisms including virus, fungi and bacteria from thousands of different species that make up the microbiota (Dulal and Keku 2014; Gagnière et al. 2016; Sonnenburg et al. 2016). Gut microbiota and host have co-existed in a symbiotic interaction by maintaining gut

homeostasis through regulating metabolic functions and intestinal epithelium structure as well as influencing host immune system and protecting against pathogens (Chow et al. 2010; Gagnière et al. 2016; Kamada et al. 2013; Marchesi et al. 2016; Sun and Chang 2014; Zhang et al. 2015).

Changes in bacterial composition may hamper gut ecosystem and exposed host to risk of developing inflammatory, metabolic and gastrointestinal tract disorders (Seyfried et al. 2014; Zhang et al. 2015), suggesting the role of dysbiosis in pathogenesis of the disease (Carding et al. 2015; Marchesi et al. 2011, 2016; Sheflin et al. 2014; Sobhani et al. 2011; Tjalsma et al. 2012; Zhang et al. 2015). Dysbiosis in general is a situation when the gut ecosystem is enriched with pathogenic bacteria instead of the beneficial ones (Schwabe and Jobin 2013). Colonic mucosa that are often exposed to dysbiosis are believed to have higher chances of developing inflammation and dysplasia over period of time (Carding et al. 2015; Gagnière et al. 2016; Gao et al. 2015). According to the Hallmark of Cancer, inflammation was mentioned as the seventh hallmark and in most tumors including CRC, chronic inflammation often displayed even at the early onset of the disease (Hanahan and Weinberg 2011; Lasry et al. 2016).

Some studies have confirmed certain bacterial species that were highly abundant in CRC to induce inflammatory and/or pro-carcinogenic pathway (Boleij et al. 2011; Boleij and Tjalsma 2012, 2013; Guo and Li 2014; Kostic et al. 2012; Sobhani et al. 2013). Protract exposure of the colon and gastrointestinal tract to dysbiosis and mucosal immune system dysregulation have consequently promoted chronic inflammation that gives rise to inflammatory bowel disease (IBD) (Kostic et al. 2014; Maloy and Powrie 2011; Sartor 2008). Ulcerative colitis and Crohn's disease are two clinical forms of IBD that differ in pathology and location in the GI (Maloy and Powrie 2011). A few studies reported that, patients who suffered IBD were more susceptible and risky to contract CRC (Triantafillidis et al. 2009; Zackular et al. 2013). Therefore, in order to study the relationship between gut microbiome and CRC manifestation, a proper sample that is able to represent the colonic microbiota has to be carefully chosen.

Stool can be a suitable sample since it is physically close and in contact with the epithelial lining of the colon space and able to reflect the host-microbial relationship (Lichtman et al. 2015; Verberkmoes et al. 2008; Xiong et al. 2015). Stool consists of a significant amount of fibrous material, food matters, microbes, host cells, and also molecules from both host and microbes that can be used as a representative of the gut system (Lichtman et al. 2015; Verberkmoes et al. 2008; Xiong et al. 2015). Moreover, the non-invasive feature of stool has made it suitable for further biomarker analysis and more specific to gut intestinal contribution (Eckburg et al. 2005; Lichtman et al. 2013; Verberkmoes et al. 2008). A study conducted by Zackular et al. (2014) has reported significant difference of microbiome population in stool of healthy individuals compared with ones that have adenomas and colorectal carcinomas. Individuals with both adenomas and carcinomas showed a dramatic loss of common gut commensal such as *Clostridium*, *Bacteroides*, and *Lachnospiraceae* (Ivanov and Honda 2012; Zackular et al. 2014). The potential use of stool as a CRC diagnosis bacterial marker were also evaluated recently by using a large sample cohort (Liang et al. 2017). The study demonstrated that *Fusobacterium nucleatum* was significantly abundant in CRC AUROC of 0.868 ($P < 0.0001$) and discriminated CRC from the control samples with a sensitivity of 77.7%, and specificity of 79.5%. However, when this bacterium was combined together with three other bacterial marker (*Clostridium hathewayi*, *Bacteroides clarus* and undefined species named as *m7),* the diagnostic ability had improved to more than 80% when used together with fecal immunochemical testing (FIT) (Liang et al. 2017).

Over the years, next-generation sequencing has helped us to understand the overall role of gut microbiota in terms of genetic, composition, ecology, roles and diversity in dybiosis-related diseases like CRC (Ji and Nielsen 2015; Kostic et al. 2014). However, relying on metagenomic findings alone is not sufficient to define the interaction between host-gut microbiota and its metabolic activity (Ji and Nielsen 2015; Lichtman et al. 2013). Therefore, integrations of functional metagenomic such as proteomics with the available knowledge are critical to fill in the gaps (Ji and Nielsen 2015; Lichtman et al. 2013). Proteomics is able to provide evidence for gene functionality in terms of its expression ability, conditions and to what extend it has been transcribed and translated (Verberkmoes et al. 2008; Wang et al. 2015). Whilst, by

utilizing the technology and knowledge in metaproteomics, we will be able to characterize human gut microbiota roles, metabolic activities and host-microbial interactions and development (Xiong et al. 2015).

Incorporating metaproteomics in studying gut microbiome seems promising since, the microbes itself (including the growing tumors) secreted abundance of proteins to the extracellular environment that may mediate host-bacterial interactions (Makridakis and Vlahou 2010). The study of these secreted proteins or better known as secretome may help us to understand and unravel cross communication between the cells and molecular mechanism of a disease such as CRC (Makridakis and Vlahou 2010). In general, secretome is a subdiscipline of oncoproteomics and a powerful tool to identify biomarkers responsible for growth factors, cytokines, proteases and others that are vital for cancer progression (Barderas et al. 2013).

At the moment, knowledge on secretome has been broadly applied in cell secretome studies (Barderas et al. 2013; de Wit et al. 2014; Shi et al. 2009; Wu et al. 2008; Wu, et al. 2008). In CRC, using metastasized CRC cell lines and serum samples, three prominent proteins (S100A8/A9, GDF15 and SERPINI1) were discovered to differentiate between healthy and CRC individuals (Barderas et al. 2013). In a different study, 76 proteins were found to be differentially identified in CRC tissue presumably of neoplastic cell origin as compared to their normal counterparts (de Wit et al. 2014). These proteins were associated with cancer and responsible in DNA replication including cell division (de Wit et al. 2014). Wu et al. (2008) in two separate studies found collapsin response mediator protein-2 (CRMP-2) and Mac-2 binding protein (Mac-2BP) in CRC cell lines conditioned media. CRMP-2 has biological function related to differentiation, while Mac-2BP is important in cell adhesion (Consortium 2018). Downstream analysis demonstrated overexpression of both proteins in tissues and plasma of CRC as compared to the controls (Wu et al. 2008; Wu et al. 2008). Shi et al. studied the secretome of CRC liver metastasis colon mucosa in comparison to normal (Shi et al. 2009). Her findings revealed 32 protein spots being differentially expressed in cancer but only 2 proteins were identified by the MS, which were Desmocolin-2 and fibrinogen γ chain proteins (Shi et al. 2009).

Despite all the findings mentioned, analyses on gut microbiome and host secretome in CRC individuals and healthy control with normal colon morphology by using stool samples are yet to be explored. Therefore, this study was carried out to profile main protein components of host and microbial secretome extracted from stool samples of CRC and healthy control individuals.

1.2 RESEARCH OBJECTIVES

1.2.1 General Objective

This study is carried out to identify the secreted proteins released from the human gut and microbial of CRC patients and control individuals by assessing the secretome in stool samples.

1.2.2 Specific Objectives

i. To profile and compare the secreted proteins of human gut and microbial in CRC and control individuals using mass spectrometry analysis and evaluate potential candidate proteins as risk predictor for CRC.

ii. To identify the molecular and biological functions of the identified secreted proteins in CRC and control individuals.

iii. To determine the correlation of identified proteins in CRC and control individuals, and predicted the interactions and pathways of secreted proteins using STRING database analysis.

1.3 HYPOTHESIS

The identified secreted proteins from human gut and microbial in CRC individuals can be mapped to tumorigenesis-associated molecular and biological functions that may be responsible for pathogenesis of CRC.

CHAPTER II

LITERATURE REVIEW

2.1 COLORECTAL CANCER

2.1.1 The Epidemiology of Colorectal Cancer

Colorectal cancer (CRC) is described as the abnormal growth tissue in large intestine either at colon or rectum (colorectal) (American Cancer Society 2017b). The tumorigenic events of CRC may be caused by genetic and epigenetic alterations, thus triggering the transformation of normal epithelium to adenoma and later proceeded to carcinoma and metastatic tumour (Armaghany et al. 2012). The sequence of these events are also known as adenoma-carcinoma sequence (Armaghany et al. 2012; Bujanda et al. 2010) in which was proposed in the 80's and represented in almost 95% of CRC cases (Bujanda et al. 2010). Adenomas or adenomatous polyps are benign lesions that have malignancy potentials (Aarons et al. 2014; Bujanda et al. 2010).

According to Global Cancer Incidence, Mortality and Prevalence (GLOBOCAN) 2012, CRC was ranked as the third most common cancer in both male and female, worldwide (Figure 2.1) (Ferlay et al. 2017). The incidence of CRC was prevalent among the western populations, until recently that Asian countries started to portray similar trends too (Wong and Ng 2013). It is believed that the rising trends were probably due to adoption of westernized diets and lifestyles among the Asians (Ferlay et al. 2014; Pericleous et al. 2013; Pourhoseingholi 2014). Globally, Asia has the highest number of CRC incidents for both genders with more than 600,000 incidents as compared to the world continents with age-specific rates (ASR) incidence of 13.7 and ASR mortality of 7.2 per 100,000 people (Figure 2.2) (Ferlay et al. 2017).

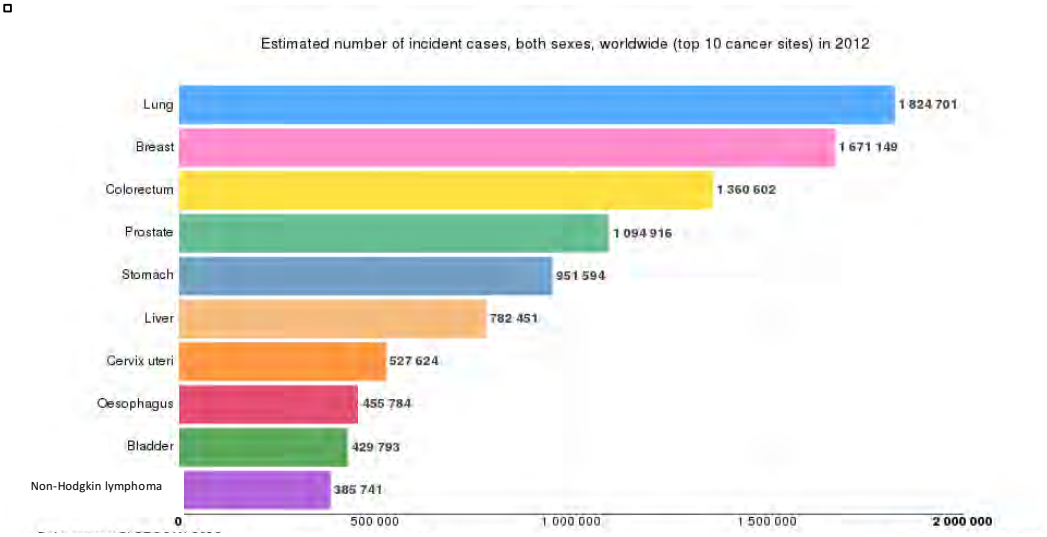


Figure 2.1 The estimated number of cancer incidence for both genders, worldwide in year 2012. From the bar chart, CRC was ranked third among the top ten most common cancer worldwide.

Source: Ferlay et al. 2017

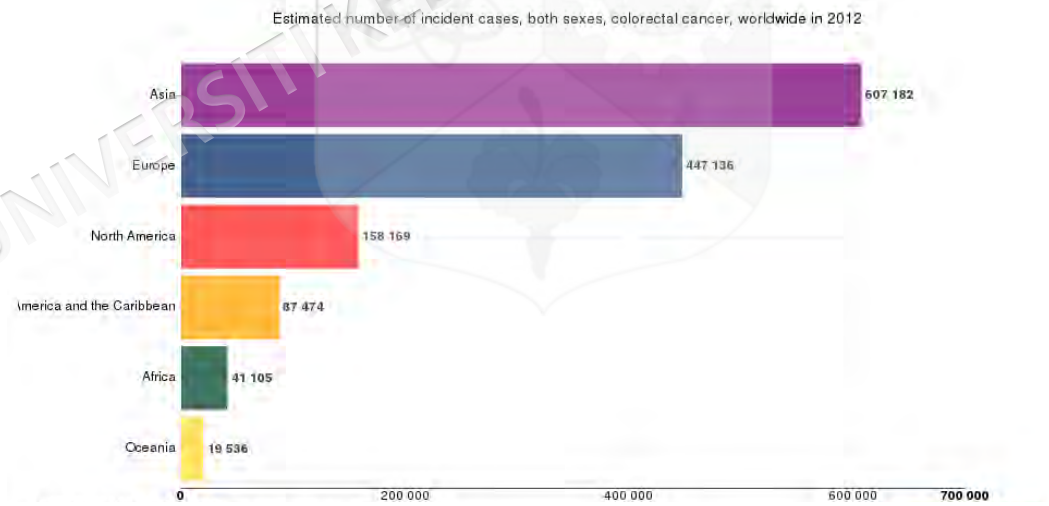


Figure 2.2 The estimated number of colorectal cancer (CRC) incidence for both genders among the world continents in 2012. Asia had the highest number of CRC incidents for both genders with more than 600,000 incidents as compared to the other continents.

Source: Ferlay et al. 2017

The CRC burden in developed countries in Asia such as Japan, South Korea, and Singapore were on the rise as compared to their developing neighbours (including Malaysia), however the incidence and mortality rates were stable and even declining (Ferlay et al. 2014; Sung et al. 2015; Veettil et al. 2016). This healthy pattern was associated with effective CRC screening programs, improved treatments and increased public awareness (Torre et al. 2015; Veettil et al. 2016).

In Malaysia, CRC was the second most common cancer after breast cancer with more than 4000 incident cases (Figure 2.3) and over 2000 death for both genders in year 2012 (Figure 2.4) (Ferlay et al. 2017). Between the genders, men were more susceptible to suffer from CRC as compared to women, and reported to have higher mortality rates (Abu Hassan et al. 2016; Ferlay et al. 2014). Among three main ethnics in Malaysia, Chinese had the highest number of CRC incidence and mortality rates as compared to Malays and Indians (Abu Hassan et al. 2016). Similar trend was also observed among the Singaporean Chinese and genetic factor was suggested to be one of the cause (Lee et al. 2017).

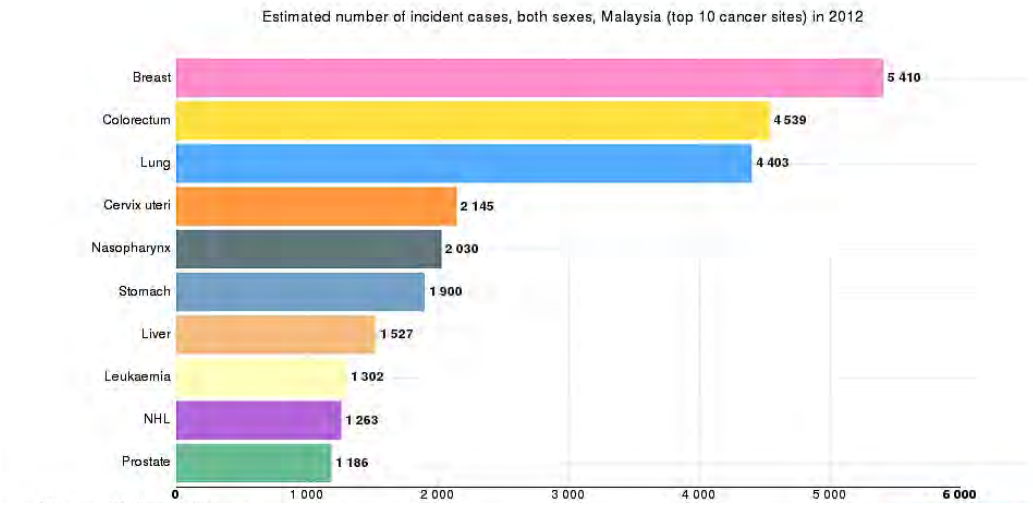


Figure 2.3 Estimated number of colorectal cancer incidence for both genders in Malaysia, year 2012. Colorectal cancer was ranked second among the top ten most common cancers with more than 4000 incident cases.

Source: Ferlay et al. 2017

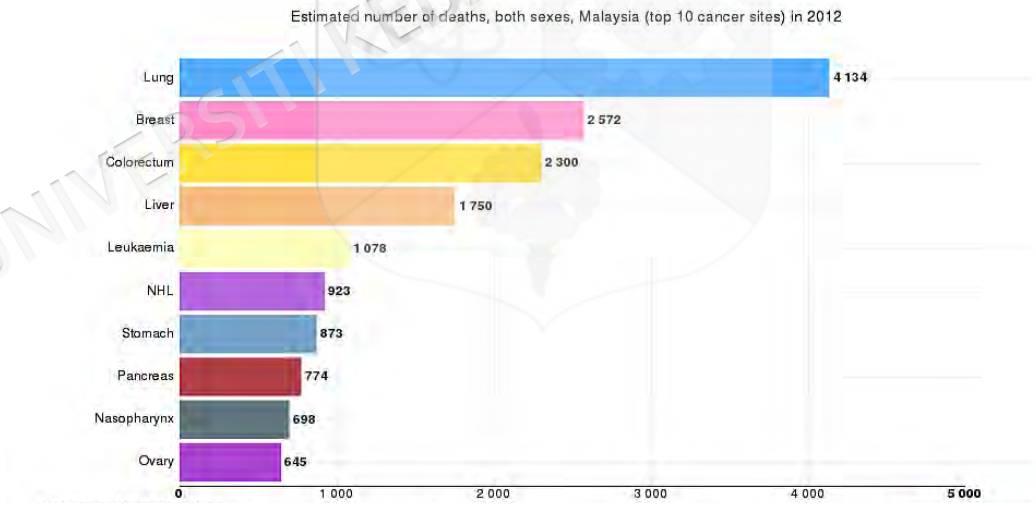


Figure 2.4 Estimated mortality cases of colorectal cancer for both genders in Malaysia, year 2012. Colorectal cancer was placed third with over 2000 death cases.

Source: Ferlay et al. 2017

2.1.2 Colorectal Cancer Screening

Early detection of tumours resulted in 90% of survival rates among CRC patients (Siegel et al. 2015). However, for advanced CRC, the chance of survival is only 8% (Scully and Cheung 2016; Siegel et al. 2015) thus forging CRC as an ideal disease for screening (Wong and Ng 2013). Significant number of death from CRC can be reduced with the utilization of screening tests and effective early intervention programs (Bibbins-Domingo et al. 2016; Rawl et al. 2012). All of these tests and screening programs primarily focused on both the detection of the cancer and also the removal of adenomatous polyps (Rawl et al. 2012). Screening tests of CRC can be divided into three; stool-based tests, direct visualization tests and serological tests (Bibbins-Domingo et al. 2016).

Colonoscopy and flexible sigmoidoscopy are the gold standards for accurate diagnostic of colorectal cancer. These visualization tests allow clinicians to examine lining of the rectum and colon, detect lesions and remove polyps (Macrae et al. 2016). Prior to the procedure, patients are required to consume laxatives, specially prepared diet or enemas to cleanse their colons (Macrae et al. 2016). Colonoscopy enables clinician to examine the entire colon (Nicholson and Korman 2005). This procedure requires mild sedation and there are cases of bleeding and complication in 0.08% individuals (Holme et al. 2013). This procedure is advocated for individuals that are at high risk towards CRC (Nicholson and Korman 2005). Meanwhile, flexible sigmoidoscopy is a procedure that examine one-third of the colon until the splenic flexure and no sedation is required for it (Macrae et al. 2016; Nicholson and Korman 2005).

Combination of stool-based test and flexible sigmoidoscopy have been reported to greatly reduce mortality rates of CRC and considered as an effective screening strategy (Holme et al. 2013). Other randomized control trials have also reported improved survival rates of CRC on individuals that carried out stool-based screening annually or biennially in addition to flexible sigmoidoscopy once in every three to five years (Lin et al. 2016).

Stool-based tests are used to detecting traces of blood in stool samples that may come from vascularized polyps, adenomas and tumours in the colon (Carroll et al. 2014; Scully and Cheung 2016). These tests are considered as preliminary test and if the tests are positive (traces of blood are detected), visualization test will be performed to take a closer look on the cause of bleeding (Holme et al. 2013). Stool based test comprised of fecal immunochemical test (FIT), guaiac-based fecal occult blood test (gFOBT) and stool DNA test (Bibbins-Domingo et al. 2016; El Zoghbi and Cummings 2016; Scully and Cheung 2016).

gFOBT detects blood traces in stool based on heme pseudoperoxidase activity (Allison et al. 2014). If heme are present, hydrogen peroxide added during the analysis will oxidize guaiac acid to yield blue coloured dye (Allison et al. 2014; El Zoghbi and Cummings 2016). gFOBT is recommended to be performed annually or biennially to reduce CRC mortality (Hardcastle et al. 1996). However, this test is prone to false positive and false negative results, depending on patient's diet intake prior to test and also the types of medication they are taking (Mitchell et al. 2004).

FIT, is a fecal immunochemical test that use specific antibodies which binds to intact globin components of human hemoglobin (Allison et al. 2014; El Zoghbi and Cummings 2016). It measures quantity of antibodies bound to hemoglobin thus, making it clinically sensitive and specific (Allison et al. 2014). However, the sensitivity of FIT will reduce with any delay in analysis reading and it also costs more compared to gFOBT (Allison et al. 2014; El Zoghbi and Cummings 2016). In terms of performance, a randomized control trials have reported the superiority of FIT over gFOBT in sensitivity of detecting adenomas and tumour (B. Levin et al. 2008; van Rossum et al. 2008). Stool DNA is a test to detect DNA commonly present in tumour (Kanthan et al. 2012). According to El Zoghbi and Cummings (2016) and Imperiale et al. (2014), stool DNA is more sensitive but less specific than FIT.

There are many tumour biomarkers that have been discovered but only a few are applied for diagnostic screening of CRC (Chan et al. 2010). The commonly used clinical tumour biomarkers are carcinoembryonic antigen (CEA), carbohydrate antigen CA 24, CA 19-9, CA 50, tissue plasminogen activator, tissue-polypeptide-specific

antigen and tissue inhibitor of metalloproteinase 1 (Chan et al. 2010). CEA is commonly used as a follow-up test for detection of recurrent CRC following primary curative treatment (Chan et al. 2010; Nicholson et al. 2015). However, it is recommended to combine CEA monitoring with other diagnostic modalities in order to augment result's sensitivity and reduce missed cases (Nicholson et al. 2015). Evidently, an improved detection and sensitivity of CRC diagnosis was reported when CEA and CA 19-9 were used together (Zhang et al. 2015).

2.2 GUT MICROBIOME

2.2.1 The Relationship Between Gut Microbiome and Human

Human colon is colonized by trillions of colonic microbiota mainly composed of bacteria (Gagnière et al. 2016). Our gastrointestinal tract hosts microbial communities that maintain stable ecological niches and established a dynamic relationship to host's health and diseases (Gagnière et al. 2016). The complex microbial community of human gut is composed of at least 1013 microorganisms with approximately 800 species of anaerobic bacteria (which have not yet been successfully cultured in the laboratory) (Martin et al. 2014). Human gut is colonized by 4 different phyla of microbes (*Firmicutes, Bacteroidetes, Proteobacteria* and *Actinobacteria*) that tend to vary in population depending on each individual diet (Sun and Chang 2014).

Majority of colonic bacteria are gram-positive bacteria, virtually all of which are obligate anaerobes (Sun and Chang 2014). Guts of ruminants and termites are the good examples of anaerobic bacteria that take charge in breaking down ingested polysaccharides and fermenting the resultant products into short-chain fatty acids (Brune and Friedrich 2000). Through these chain series, the hosts are able to gain carbon and energy while, the microbes provide a rich source of glycans (Brune and Friedrich 2000). On the other hand, our gut also contains microbes that play roles in synthesizing vitamins (such as folate and riboflavin) and metabolites, detoxifying xenobiotics, maintaining gut epithelial structure and participating in immune system protection (Cipriani et al. 2010; Delzenne et al. 2011). Deficiency of these beneficial bacteria due to broad spectrum antibiotics intake may cause improper development of gut immune

system and antibiotic-related colitis, consequently promoting dysbiosis, one of the risk factor of CRC (DeWeerdt 2015; Garrett 2015; Kostic et al. 2012). Therefore, a better understanding on gut microbiota function is critical to understand their role in human homeostasis and disease pathogenesis.

2.2.2 Dysbiosis of Gut Microbiome Leads to Colorectal Cancer

Microbes residing in the human gut exist in delicate balance among the microbial communities (Sun and Chang 2014). In situations that may lead to dysbiosis, the beneficial microbes were reduced greatly due to several causes such as diet alteration, medications and environmental factors (Sun and Chang 2014). This situation leads up to saturation of pathogenic microbes in the human gut. To counteract this matter, host's immune and inflammatory responses will be activated and consequently initiate inflammation (Sun and Chang 2014). Prolonged colon inflammation may cause dysplasia and if left untreated will later develop into tumours (Hanahan and Weinberg 2011).

There is a significant relationship between gut microbes, inflammation and host's immune system. When an inflammation occurs, gut mucosal barrier will be compromised allowing microbes to breach and influence host's immune responses (Garrett 2015; Pylayeva-Gupta et al. 2012). Once the barrier is breached, innate and adaptive immune responses will be activated via cytokine-mediated signalling pathways leading to tumour progression and growth (Garrett 2015; Pylayeva-Gupta et al. 2012).

The population of microbes in the human intestine is more saturated at the large intestine and for that reason, CRC is prone to develop at that location (Sobhani et al. 2013). Host's small intestine is created in such a way that promote fast flow of contents (Salyers 2005). Thus, discouraging microbial growth by washing away microbes before they have any chance to establish (Salyers 2005). However, the flow slows down as it passed through the colon, which eventually allowed microbes to establish themselves and could reach up to 30 percent of human colonic content (Salyers 2005).

On the other hand, it is important to note that, microbes are not the only 'culprit' for the formation of tumour on colon surfaces. Instead, microbes may find a tumor's oxygen tension or carbon sources permissive and take advantage of an underused nutritional niche (Garrett 2015). This situation attracts the accumulation of microbes and subsequently influences genomic stability, anti-apoptosis and proliferative signals (Garrett 2015). High saturation of these microbes in a tumour microenvironment could aggravate the situation further.

The gut microbes also have evolved mechanisms to damage DNA (Garrett 2015). They are equipped with this feature as a means of survival against other competitors but in some way may contribute to the aetiology of cancer (Garrett 2015). These microbes possessed defensive factors such as colibactin encoded by the *pks* locus, expressed by *Escherichia coli* as well as by other Enterobacteriaceae and cytolethal distending toxin (CDT) produced by several ε- and γ-proteobacteria (Nougayrède et al. 2006; Putze et al. 2009). Both colibactin and CDT can cause double-stranded DNA damage in mammalian cells (Garrett 2015; Guerra et al. 2011). *Bacteroides fragilis* also possess a similar effect with its toxin that facilitate increment of reactive oxygen species (ROS) level thus promoting mutation (Goodwin et al. 2011). Some bacteria are normal flora in the human gut but their relationships with β-catenin signaling may cause appropriate boundaries between the host and microbes to be lost (Garrett 2015). β-catenin is a signaling pathway frequently altered in many malignancies and can be activated by bacterial toxin. For example the toxin produced by *B. fragilis* stimulates cleavage of E-cadherin which later activates β-catenin (Nelson and Nusse 2004). E-cadherin is a glycosylated Type I transmembrane protein which is critical to the formation of intercellular adhesion junctions (in the intestinal epithelium) and to cellular signaling, proliferation and differentiation (Nelson and Nusse 2004).

Fusobacterium nucleatum is a stool microbe that have been cultured previously from biopsies of inflamed gut mucosa (Rubinstein et al. 2013; Sears and Garrett 2014). This bacteria was found abundant on colonic mucosa and adenomas, supporting its possible role in mucosal inflammation (Rubinstein et al. 2013; Sears and Garrett 2014). *F. nucleatum* was previously shown to participate in tumorigenesis via FadA expression

(a cell surface adhesion component) that binds to host E-cadherin and activates β-catenin (Rubinstein et al. 2013; Sears and Garrett 2014).

Helicobacter pylori of oncogenic type 1 strain also targeted β-catenin pathway via injection of Cytotoxin-associated gene A protein (CagA) into the host's cytoplasm (Abreu and Peek 2014) thus facilitates up-regulation of genes related to cellular proliferation, survival, migration and angiogenesis in gastric cancer (Abreu and Peek 2014). A large population-based case-control study in Germany had tested against the serum antibodies to *H. pylori* and the CagA protein, and found that *H. pylori* could result in a small yet relevant increase of CRC risk (Zhang et al. 2012).

2.3 SECRETOME

2.3.1 Secretome proteomics

CRC can be studied by using metagenomics approach such as sequencing and has been widely used to discover novel bacteria species (Kolmeder et al. 2012; Redon et al. 2005). Transcriptomics analysis in prokaryotic expression study can be considered as a powerful tool, but it requires a stable mRNA which is extremely low in prokaryotes. The stability of a particular mRNA is controlled by specific interactions between its structural elements and RNA-binding proteins that can be general or mRNA-specific (Guhaniyogi and Brewer 2001). Besides that, mRNA transcripts do not necessarily represent the microbial function which is ultimately mediated by proteins (Kolmeder et al. 2012; Redon et al. 2005). Therefore, settling on metagenomic analysis alone is not enough. Additional information such as functional metagenomics (transcriptomics and proteomics) is required in order for us to identify not only the expressed genes but also its mechanisms in CRC.

The term secretome is first used in a study of signal peptide-dependent protein transport in *Bacillus subtilis* (Tjalsma et al. 2000). Secretome includes all of the proteins that are being secreted by the cells, tissues and organism via various secretion pathways (Karagiannis et al. 2010; Loei et al. 2012; Volmer et al. 2005). These proteins are comprised of proteases, growth factors and cytokines (Karagiannis et al. 2010; Loei et

al. 2012; Makridakis and Vlahou 2010). Growing tumours secrete similar proteins as the normal cells with additional angiogenic factors crucial for cancer progression and metastasis (Liotta and Kohn 2001).

The first study on secretome was performed in an in-vitro setting using interstitial fluid extracted from breast tumour (Celis et al. 2004). Besides interstitial fluid, conditioned media (CM) of cancerous cell lines had also been used for secretome studies since the cells possess a rich source of secreted and cellular proteins (Kulasingam and Diamandis 2008; Minamida et al. 2011; Xu et al. 2010). Secretome analysis of cardiac stem cells has successfully identified soluble proteins secreted by the cardiac myocytes, fibroblasts, and/or (cardiac) stem cells which are able to regenerate myocardium (Burchfield and Dimmeler 2008; Doyle et al. 2008; Jackson et al. 2001; Timmers et al. 2008) hence, introduced the probability of protein therapy. A study on nine metastatic bone cell lines had discovered several functional classes of protein that were associated with bone metastasis (Blanco et al. 2012). Interestingly, there were experiments designed for biomarker discovery purpose whereby the identified secreted proteins from cell lines were subsequently validated in the patients' samples. Such studies had identified Collapsin response mediator protein-2 (CRMP-2) (Wu et al. 2008) and Melanotransferrin (TRFM) (Shin et al. 2014) from the CRC cell line secretome analyses, followed by confirmation on their elevated levels in the tumour and plasma samples of CRC patients.

Previous studies have shown, a co-culture experiment of HT-29 colon cancer cell line with NCM460, a colon mucosal epithelial cell line stimulating the tumour microenvironment. The outcome showed deregulation of Acrogranin, Insulin-like growth factor binding protein 6 (IGFBP6) and Vimentin, thus these proteins were suggested as inducer proteins for tumour progression (Zeng et al. 2013). In metastatic CRC, proteins that are critical in metastatic events such as Trefoil factor 3 (TFF3) and Growth/differentiation factor 15 (GDF15) were significantly expressed in secretome of SW620 (a lymph node metastatic cell line) and these proteins were also found elevated in the serum of the metastatic CRC patients (Xue et al. 2009).

Studies on the secretome of gut microbiota are scarce and have important knowledge gap waiting to be filled. To investigate the secretome of both gut bacteria and the host, we suggest that stool samples would be suitable for the meta-proteomic study of such kind and the data should be acquired and quantitated using high resolution mass spectrometry (MS) technologies (Cameron and Takáts 2018; Jin et al. 2017). The study outcomes may enable us to discover proteins secreted by the host's tumourous tissues as well as from the microbiome that are related to CRC-stricken guts. Providing information on host-microbe interactions, as well as the crosstalk and signaling pathways related to the pathogenesis of the disease.

The MS technology is a powerful tool to characterise atomic or molecular species as gas-phase ions and performs the analysis based on their different mass-to-charge (m/z) ratios (Cameron and Takáts 2018). It capable of providing qualitative and quantitative analytical data on nanomolar to attomolar amounts of analyte (Mellon 2003). The advantages of MS include, high sensitivity and mass accuracy as well as, structural information (Vandell and Limbach 2010). MS often coupled with a chromatographic technique to separate the chemical constituents of a sample based on their physical and/or chemical properties prior to MS analysis (Cameron and Takáts 2018). This technique helps to reduce the molecules which are present at the ion source, decreasing competition between molecules and thus the potential for ion suppression; improving the precision, sensitivity, and limit of detection of MS analysis (Cameron and Takáts 2018). The MS comprise four main components, an ion source where sample molecules may be ionized, a mass analyzer that separates ion according to their mass-to-charge ratio, a detector that measures the abundances of the separated ions as an electrical signal and a recording device that convert the detector signal into a form suitable for further study and processing (Mellon 2003).

CHAPTER III

METHODOLOGY

3.1 MATERIALS AND EQUIPMENTS

3.1.1 Chemical Reagents

Chemical reagents used in this study were as follows; acetonitrile (Merck Millipore, Germany), Trizma base (Sigma-Aldrich, USA), hydrochloric acid fuming 37% (Merck, Germany), DL-Dithiothreitol (Sigma-Aldrich, USA), iodoacetamide (Sigma-Aldrich, USA), sequencing grade modified trypsin, porcine (Promega, USA), ammonium bicarbonate (Sigma-Aldrich, USA), formic acid (Sigma-Aldrich, USA), quick start Bradford 1x dye reagent (Bio-Rad, USA), quick start bovine serum albumin standard (Bio-Rad, USA), 2x Laemmli sample buffer (Bio-Rad, USA), 2-mercaptoethanol (Sigma-Aldrich, USA), 10x Tris/Glycine/SDS buffer (Bio-Rad, USA), Precision Plus Protein Standards (Bio-Rad, USA), and Bio-Safe Coomasie G-250 stain (Bio-Rad, USA).

3.1.2 Equipments

Equipments used to carry out this study were Eppendorf Research® plus pipettes (Eppendorf, Germany), Mettler Toledo Jewellery JP weighing scale (Mettler Toledo, USA), SaStec Lab Equipment fume hood (Segar Alatan Sains Sdn, Bhd., Malaysia), Eppendorf 5810R centrifuge (Eppendorf, Germany), LaboGene Scan Speed Mini centrifuge (LaboGene, Denmark), Fine Vortex (Fine PCR, Korea), Eppendorf Concentrator Plus (Eppendorf, Germany), Varioskan Flash spectrophotometer machine (Thermo Scientific, USA), PowerPac HC high-current power supply (Bio-Rad, USA), Criterion vertical electrophoresis cell (Bio-Rad, USA), Biometra TS1 Thermoshaker

thermal block and shaker (Analytik Jena AG, Germany), KS 260 basic flat orbital shaker, 230 V (Sigma-Aldrich, USA), Bio-Rad ChemiDoc XRS+ System (Bio-Rad, USA), S220 Seven Compact pH/Ion benchtop meter (Mettler Toledo, USA), Eksigent ekspert nanoLC 415 HPLC system (Eksigent, USA), and TripleTOF 5600+ System mass spectrometer (AB Sciex, USA).

3.1.3 Consumables

Consumables used in this study were as follows; Millex-GV filter unit Durapore PVDF membrane size 0.22 µm (Merck Millipore, Germany), 20 ml/cc syringe without needle (Muzamal Industries Sdn. Bhd., Malaysia), disposable Pasteur pipette (Kumpulan Saintifik F.E. Sdn. Bhd., Malaysia), Falcon 50 ml conical centrifuge tubes (Fisher Scientific, USA), Eppendorf safe–lock tubes 1.5 ml (Eppendorf, German), 650 ml Fisherbrand commode specimen collection system (Fisher Scientific, USA), 52 ml clear stool container with PE-spoon stopper (Greiner Bio-One, Japan), Reynolds zipper bags large size (Reynolds, USA), chillers soft ice substitute size small (Coleman, USA), foam cooler box mini with string handle (325 mm x 200 mm x 250 mm) (Venue Star Sdn. Bhd., Malaysia), Pierce C18 Tips 100 µl (Thermo Scientific, USA), Eppendorf epT.I.P.S. (10 µl, 100 µl and 1000 µl) (Sigma-Aldrich, USA), Nunc micro well 96-well microplate (Thermo Fisher Scientific, USA), Mini-Protean TGX Precast Gels (Bio-Rad, USA), and polypropylene vial (Agilent, USA), nanoAcquity UPLCTM Trap Column, Symmetry® C18, 5 µm, 180 µm × 20 mm (Waters, USA).

3.2 STUDY DESIGN

This study is an exploratory research that investigates protein profiles of gut microbiome and host secretome in CRC individuals as compared to control (with healthy colon morphology). We have explored areas related to stool samples collection and optimized protein extraction from gut microbial secretome which will be useful for future studies.

Therefore, we have included all stool samples that we had managed to collect over one (1) year period (March 2016 to March 2017) into this study. The methodology

of this study can be divided into two parts, 1) secretome protein profiling of CRC and control samples using proteomic analysis, 2) visualisation of protein expression, statistical analysis, functional annotation and determination of pathways correlation using bioinformatic tools. The overview of research methodology was summarized in Figure 3.1.

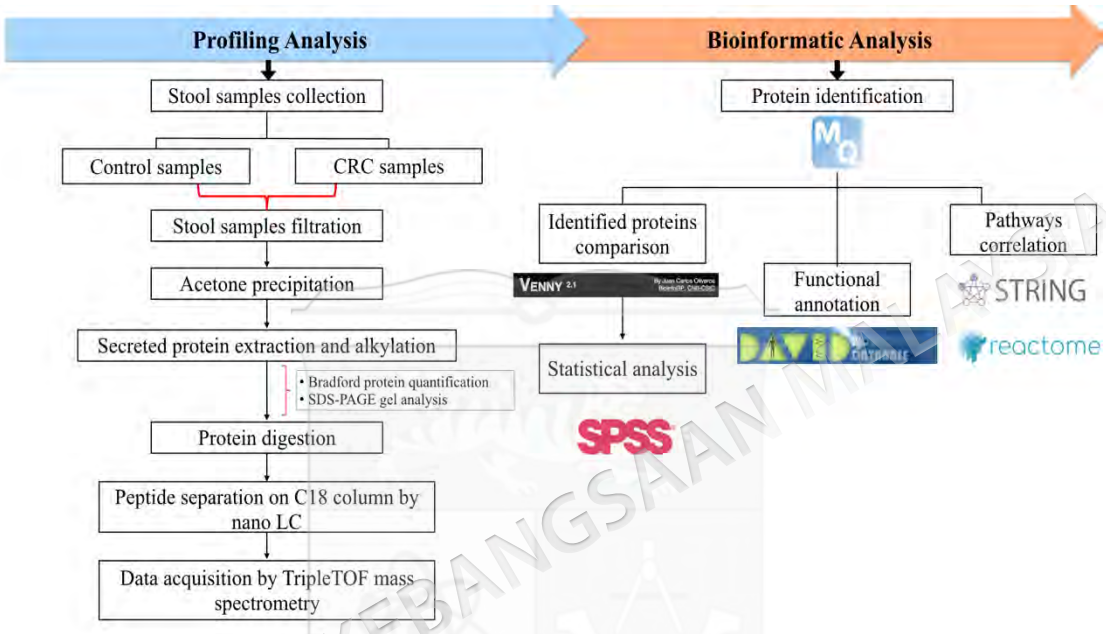


Figure 3.1 Overview of research methodology

3.3 ETHICAL APPROVAL

This study was approved by the UKMMC Research Ethics Committee (Appendix A). All participants were informed and have consented to participate in this study. The template of information sheets and consent forms for research participants were attached in Appendix B, C, D and E.

3.4 STOOL SAMPLES COLLECTION

A total of 46 stool samples were collected from CRC patients (26 samples) and controls with healthy colon morphology (20 samples) at the UKM Medical Centre, Cheras (UKMMC). The inclusion criteria for research participants were as follows; (1) individuals within the age range of 30 to 70 years old regardless of gender and ethnicity,

(2) diagnosed with CRC or free from CRC (for controls) with confirmation made by colonoscopy procedure at the Endoscopy Unit, UKMMC and histopathological diagnosis from Pathology Department, UKMMC, and (3) CRC was their primary cancer and did not undergo preoperative chemotherapy or radiotherapy treatments. Meanwhile, the exclusion criteria were; (1) recurrent CRC patients or cancer surveillance patients, (2) patients who had undergone adjuvant or neoadjuvant treatment for CRC or other cancers, and (3) patients who had medical history of antibiotic usage for one year or six months prior to stool sample collection. Our sample collection workflow at UKMMC was summarized in a flow chart (Appendix M).

Every participant recruited was provided with a foam cooler box contained, one (1) Fisher commode specimen collection system, six (6) pieces of chillers soft ice substitute, size small and one (1) piece of Reynolds zipper bags, size large (Appendix N). This box was given to all 46 participants after their respective colonoscopy procedure at the Endoscopy Unit, UKMMC. Participants were requested to self-collect their stool samples a day before their return for their respective next clinic appointment (scheduled 2 weeks after colonoscopy procedure). Guideline for stool samples collection was provided in the given foam cooler box (Appendix O). Participants were also required to note the time and date of stool collection on a sticker pasted on the box (Appendix P). The boxes containing the stool samples were retrieved by our research team during the clinic appointments.

3.5 STOOL SAMPLES FILTRATION

The physical characteristics of collected stool was observed and noted. Characteristics such as colour, consistency, shape, constituents and Bristol stool scale (BSS) classification (Figure 3.2) were recorded for reference (Lewis and Heaton 1997). According to BSS, stool characteristics can be divided into seven types, from hard lumps to watery stools. Then, the stool was spooned at approximate equal amount into five (5) sterile 50 ml tubes. The weigh was recorded and later stored in UMBI’s Biobank at -80 °C within 24 hours after their collections. Patient's information such as age, sex, ethnicity, stool collection time stamp and histopathological report were recorded.

Prior to sample's filtration, a tube of frozen stool was retrieved from UMBI's Biobank -80 °C freezer and left to thaw at RT. About 28 ml of distilled water was added to 2 ml of thawed stool sample and centrifuged at 3000 x g at 4 °C for 10 minutes using Eppendorf 5810R centrifuge. Obtained supernatant was transferred into a 50 ml conical centrifuge tube by using Pasteur pipette and remaining pellet was discarded. Supernatant was filtered using Millex-GV filter unit Durapore PVDF with membrane size of 0.22 µm through 20 ml/cc syringe.

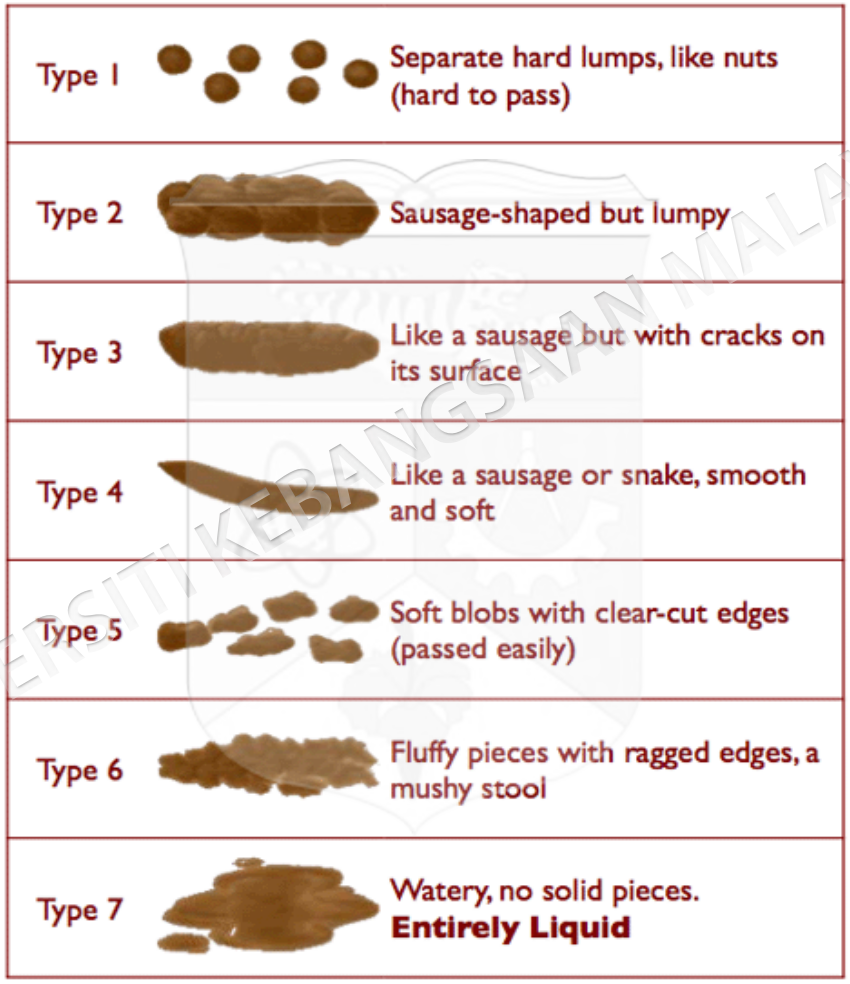


Figure 3.2 Bristol stool scale

Source: Lewis and Heaton 1997

3.6 PROTEIN EXTRACTION

Collected filtrate was precipitated using acetone precipitation method. Briefly, four times volume of cold acetone was added to the filtrate, vortexed and stored overnight at -20 °C. Filtrate was centrifuged at 15,000 x g, 4 °C for 10 minutes and the supernatant was discarded. Obtained pellet was rinsed twice using 20 ml mixture of cold acetone and distilled water at the ratio of 4:1, followed by centrifugation at 15,000 x g, 4 °C for 10 minutes during each rinse. At the final rinse, the supernatant was discarded while the remaining pellet was left to air-dry for approximately 20 minutes. The pellet was then dissolved in 40 mM Tris-HCL, pH 7.4 with 3% of DL-Dithiothreitol at 2:1 ratio of buffer solution to pellet. Preparation of buffers used were tabulated in Appendix Q.

Approximately, 100 µl of dissolved pellet was aliquoted into two 1.5 ml Eppendorf safe–lock tubes and stored at -20 °C for standard Bradford protein quantification. Whereas, the remaining dissolved pellet were aliquoted into several 1.5 ml Eppendorf safe–lock tubes, 100 µl each and stored at -80 °C until further analysis.

3.7 PROTEIN ESTIMATION

Seven (7) dilutions of bovine serum albumin (BSA) standard range between 1.25 to 25 µg/ml protein were prepared to establish the protein estimation standard curve. TrisHCL-DTT buffer was used for the blank tubes. Aliquoted pellet samples were retrieved from -20 °C, thawed on ice and diluted to a dilution factor of 1:200. Approximately, 100 µl of each standard solution or diluted pellet sample was pipetted into triplicates of 1.5 ml Eppendorf tubes and transferred to Nunc micro well 96-well microplate (Thermo Fisher Scientific, USA). Following that, 100 µl of Bradford reagent was added to each tube, mixed well and allowed to incubate for 5 minutes at RT. Absorbance of samples were measured at 595 nm using Varioskan Flash spectrophotometer machine (Thermo Scientific, USA) and the protein amount for each sample was determined from the standard curve.

3.8 SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE) ANALYSIS

SDS-PAGE analysis was used to observe the protein content and complexity, from each extracted sample. This step was performed to ensure that each sample to be proceeded for protein digestion and mass spectrometry acquisition contained substantial amount of proteins. Prior to SDS-PAGE analysis, one (1) litre of running buffer consisting 100 ml of 10x Tris/Glycine/SDS buffer with 900 ml distilled water was prepared. The Mini-Protean TGX stain-free precast gels were assembled into a Criterion vertical electrophoresis cell and the buffer chambers were filled with running buffer until it reached the indicator mark. For sample preparation, the sample was retrieved from the -80 °C storage and thawed. Five (5) µl of the sample was aliquoted into 1.5 ml tube followed by addition of 4.75 µl Laemmli sample buffer and 0.25 µl 2-Mercaptoethanol. The mixture was mixed well and denatured using Biometra TS1 Thermoshaker thermal block and shaker at 100 °C for 10 minutes prior to its loading into the respective well on the gel. Precision Plus Protein Standards was used as a protein ladder to estimate the size of protein bands. Next, the electrophoresis cell was connected to PowerPac HC high-current power supply at 200 V for 30 minutes until the dye front reached the reference line. Then, the gels were rinsed with distilled water twice and stained with Bio-Safe coomasie G-250 stain overnight with continuous shaking by using KS 260 basic flat orbital shaker. Later, the gels were rinsed again with distilled water prior to gel viewing using ChemiDoc XRS+ System.

3.9 IN-SOLUTION DIGESTION

Approximately, 20 µg of pellet for each sample was retrieved from -80 °C, added with 20 µl of acetonitrile and incubated for 20 minutes, reduced with 40 µl 10 mM DL-Dithiothreitol at 60 °C for 10 minutes and alkylated with 20 µl of 55 mM iodoacetamide for 35 minutes at RT. Then, 20 µg of sequencing grade modified trypsin was added to digest the sample, followed by addition of 10 µl of 25 mM ammonium bicarbonate, pH 7.5. The mixture was then left to incubate at 37 °C for 18 hours on Biometra TS1 Thermoshaker thermal block and shaker. Later, 20 µl of 0.3% formic acid was added to the sample. The obtained peptides were then zip-tipped, to purify and concentrate the

protein samples. The 100 µl Pierce C18 Tips were used for this process according to manufacturer's protocol. The obtained samples were dried up using Eppendorf Concentrator Plus followed by addition of 0.3% formic acid. Samples were pipetted into polypropylene vial and subjected to Eksigent ekspert nanoLC 415 system injection. The preparation of buffers used were tabulated in Appendix R. The in-solution digestion protocol used in present study was established in our lab.

3.10 LC-MS/MS ACQUISITION AND ANALYSIS

Analytical separation and analysis were performed using an Eksigent NanoLC system connected to a quadrupole time-of-flight 5600+ TripleTOF™ mass spectrometer (AB Sciex, USA). Peptides were directly injected onto a nanoAcquity UPLCTM Trap Column, Symmetry® C18, 5 µm, 180 µm × 20 mm (Waters, USA) and then separated using a 120 min gradient of buffer A: 0.1% (v/v) formic acid, 2% (v/v) acetonitrile and buffer B: 0.1% (v/v) formic acid, 90% (v/v) acetonitrile with a flow rate of 0.3 µL/min. The data was acquired using TripleTOF 5600 system fitted with a Nanospray source. Mass spectra were acquired in 250 ms accumulation time per spectrum and mass range of 300–1250 m/z to obtain MS/MS spectra for the most abundant parent ions following each survey MS1 scan. A maximum of 50 precursors with a charge state between 2 and 4 were selected for fragmentation for each mass spectrum. The tandem mass spectrum was acquired in the high-sensitivity mode, in which the signals were accumulated for a minimum of 50 ms per spectrum, and the dynamic exclusion time was set at 24 s.

3.11 PROTEIN IDENTIFICATION

Each sample was analysed individually in a data dependent acquisition (DDA) mode and the spectral data files were submitted to MaxQuant software (version 1.6.0.1) (Max Planck Institute, Germany) for simultaneous database matching for CRC and control, respectively. Uniprot fasta file from human and microbial (with archaea, bacteria, eukaryota, fungi and virus specification), downloaded on September 2017 and June 2017 were used for human and microbial database searching. Trypsin digestion enzyme was selected and two maximum missed cleavages were accepted. The variable modifications for the searches were acetyl (N-terminal) and oxidation (methionine),

with maximum five modifications per peptide and carbamidomethyl (C) was selected as a fixed modification. Initial search peptide tolerance was set at 0.07 Da and fragment mass tolerance was at 0.006 Da. The false discovery rates (FDR) for peptide spectral matches (PSM) and proteins were set to 0.01. AB Sciex Q-TOF was chosen as the instrument. Label-free quantification (LFQ) was set to minimum ratio count of two and normalization was selected. The minimal peptide length was set to seven amino acids and the match between runs was selected for each experiment with a match time window of 0.7 min. Under identification tab, minimum razor peptide was set to one and a minimum score of 40 was accepted for modified peptides.

3.12 DATA AND STATISTICAL ANALYSIS

Expressed proteins from MaxQuant analysis were used to generate comparative Venn's diagram using Venny software (version 2.1) (Oliveros 2007). The protein interactions and pathways correlations were determined using STRING (version 10.5) database, using the multiple proteins search feature against the *Homo sapiens* and various microbial organism databases. Functional annotation and prediction of protein-protein interaction of the identified proteins were mined using Database for Annotation, Visualization and Integrated Discovery (DAVID) software (version 6.8) (Szklarczyk et al. 2017). The data were mined using functional annotation feature whereby Uniprot Accession was set as the identifier and the search was specified to *Homo sapiens* and various microbial species databases. IBM SPSS (version 22) software package (SPSS, Inc., Chicago, IL) was used for Mann-Whitney, Kruskal-Wallis and Spearman's correlation statistical analysis tests, to determine the significant proteins according to demography and clinical factors. MetaboAnalyst (version 4.0) was used to generate the heat map analysis, principal component analysis (PCA), orthogonal partial least square discriminant analysis (OPLS-DA) and protein-protein correlation analysis using Spearman's correlation (Chong et al. 2018).

CHAPTER IV

RESULTS

4.1 STOOL SAMPLES COLLECTION AND DEMOGRAPHY

A total of 46 stool samples were obtained from 26 CRC patients of various stages and 20 individuals with healthy colon morphology at UKM Medical Centre, Cheras, Kuala Lumpur. These samples were collected from March 2016 to March 2017. Among the 26 individuals, 15 are Malay, 10 Chinese and one Indian with majority of whom are males (16) and only 10 females. According to Dukes classification (a staging system for CRC), three samples were classified as Stage I, eight (8) samples Stage II, 13 samples Stage III and only two (2) samples classified as Stage IV. The demography and clinical data for 26 CRC individuals and 20 control with healthy colon morphology used in this study were tabulated in Table 4.1.

Table 4.1 The demography and clinical data for CRC and control samples used in the study

Characteristics	**CRC samples (n= 26)**	**Control with healthy colon morphology samples (n = 20)**
Age (years old)		
<40	1	-
40-49	-	2
50-59	11	3
60-69	5	11
>70	9	4
Gender		
Male	16	10
Female	10	10
Ethnicity		
Malay	15	7
Chinese	10	10
Indian	1	3
Tumour Stage		
I	3	-

to be continued…

…continuation

II	8	-
III	13	-
IV	2	-

4.2 STOOL SAMPLES CHARACTERISTICS AND BRISTOL STOOL SCALE

All stool samples collected were aliquoted into five sterile containers and stored at -80°C freezer, UMBI's Biobank within 24 hours time frame after defecation. Based on our observation following the Bristol stool scale (BSS) category guidelines, most samples from CRC individuals were brown in colour with blood in the stool. The stool consistency was mushy and categorized as Type 6 in BSS. Meanwhile, stools from the control group appeared to have light to dark brown in colour with soft-formed and semi-solid consistency and can be categorized as Type 3. A table on collected stool samples categorization based on the stool scale was tabulated in Appendix F.

4.3 SDS-PAGE GEL IMAGE OF CRC AND CONTROL SAMPLES

SDS-page gel image analysis on extracted secretome was carried out to observe protein content and bands distribution in each sample. A representative SDS-PAGE gel image of nine secretome samples from CRC and control is shown in Figure 4.1. The other gel images of the secretome samples were shown in Appendix G.

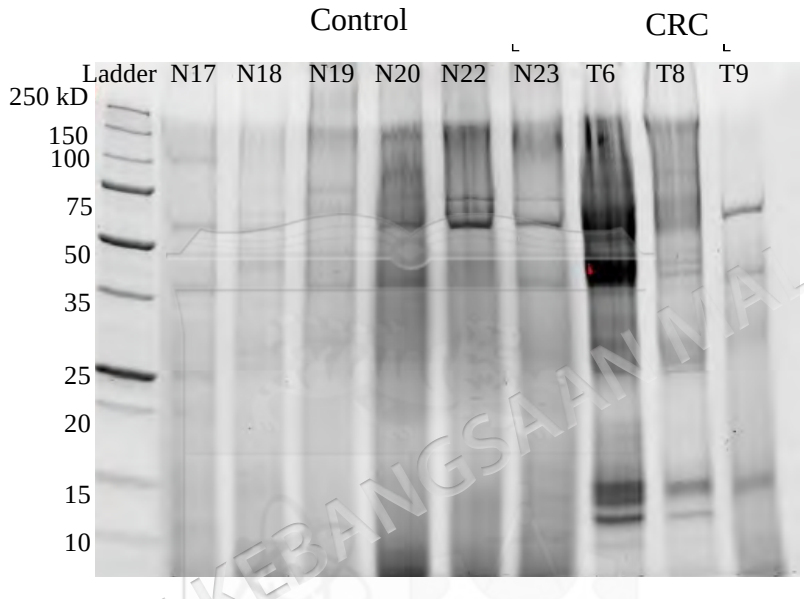


Figure 4.1 Representative SDS-PAGE gel image comprised of 9 secretome samples, 6 control (labelled N17, N18, N19, N20, N22 and N23) and 3 CRC samples (labelled T6, T8 and T9). Visible thicker band at size 10 to 15 kD were observed in CRC samples but the bands were consistently faint in the control groups.

4.4 THE IDENTIFIED HUMAN ORIGIN SECRETOME

4.4.1 Overview of Identified Human Proteins

We had analysed all 46 secretome samples (26 CRC and 20 controls) using MaxQuant software search against human Uniprot database. A total of 2031 proteins were discovered in both CRC and control samples. As shown in Venn's diagram, over half of the proteins were exclusive to CRC, 56.9% as compared to proteins in exclusive in control, 39.8% and only 3.3% proteins were found in both groups (Figure 4.2).

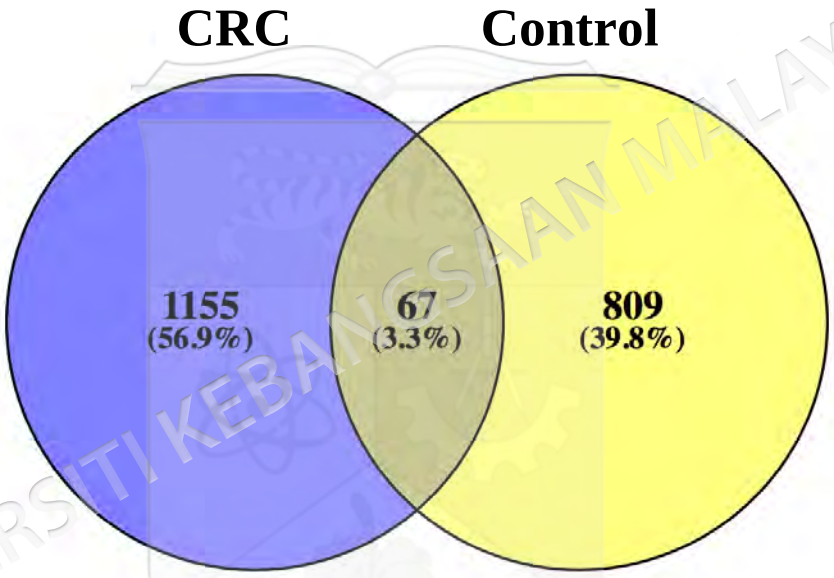


Figure 4.2 Venn's diagram of human proteins identified in 26 CRC and 20 control secretomes. A total of 2031 proteins were discovered, 1155 proteins (56.9%) were unique to CRC, 809 proteins (39.7%) were unique to control and 67 proteins (3.3%) were shared by both groups.

4.4.2 Exclusively Identified Human Proteins in Colorectal Cancer (CRC)

a. Protein pathways analysis of exclusive human proteins in CRC

STRING database was used to identify protein interactions and pathways for CRC exclusive proteins (Figure 4.3). Ten (10) KEGG annotated pathways were correlated with these proteins, including Hypoxia-inducible factors 1 (HIF-1) signaling pathway, Cholinergic synapse, Oocyte meiosis, Leukocyte transendothelial migration, signaling pathways regulating pluripotency of stem cells, Jak-STAT signaling pathway, cGMP-PKG signaling pathway, Transcriptional misregulation in cancer, Viral carcinogenesis and Huntington's disease.

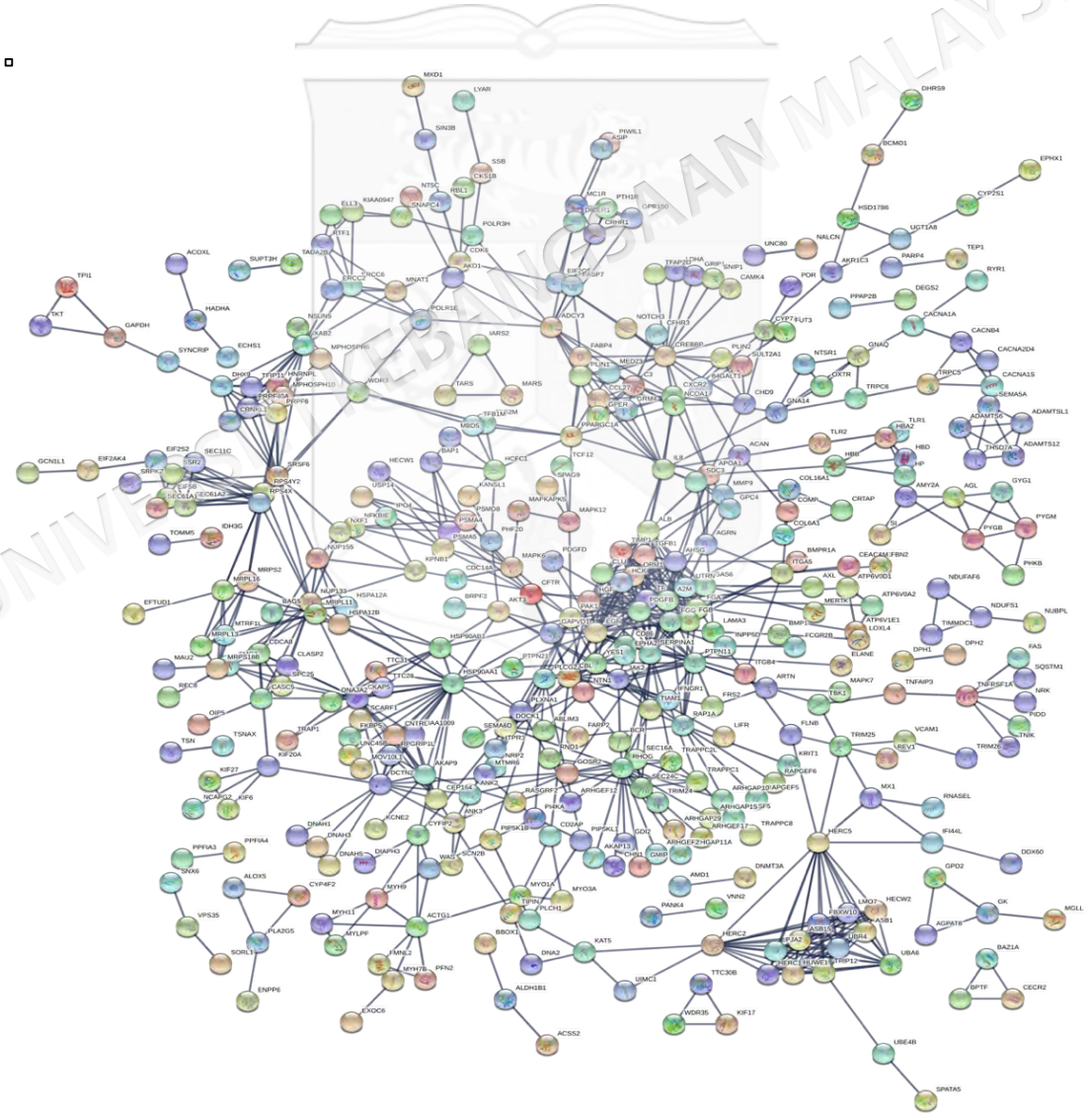


Figure 4.3 The overview of protein interactions for exclusively identified human proteins in CRC.

b. Exclusively Identified human protein based on CRC staging

We carried out Venn's diagram analysis to determine proteins that were unique to CRC stages. We discovered a total of 503 proteins were unique to Stage III, the highest number of proteins among all st66ages and only one protein in Stage IV (Figure 4.4). Overall, there were great numbers of protein shared between two to three different CRC stages. However, none were shared by all four stages together. List of unique and shared proteins between these stages were compiled in Appendix S.

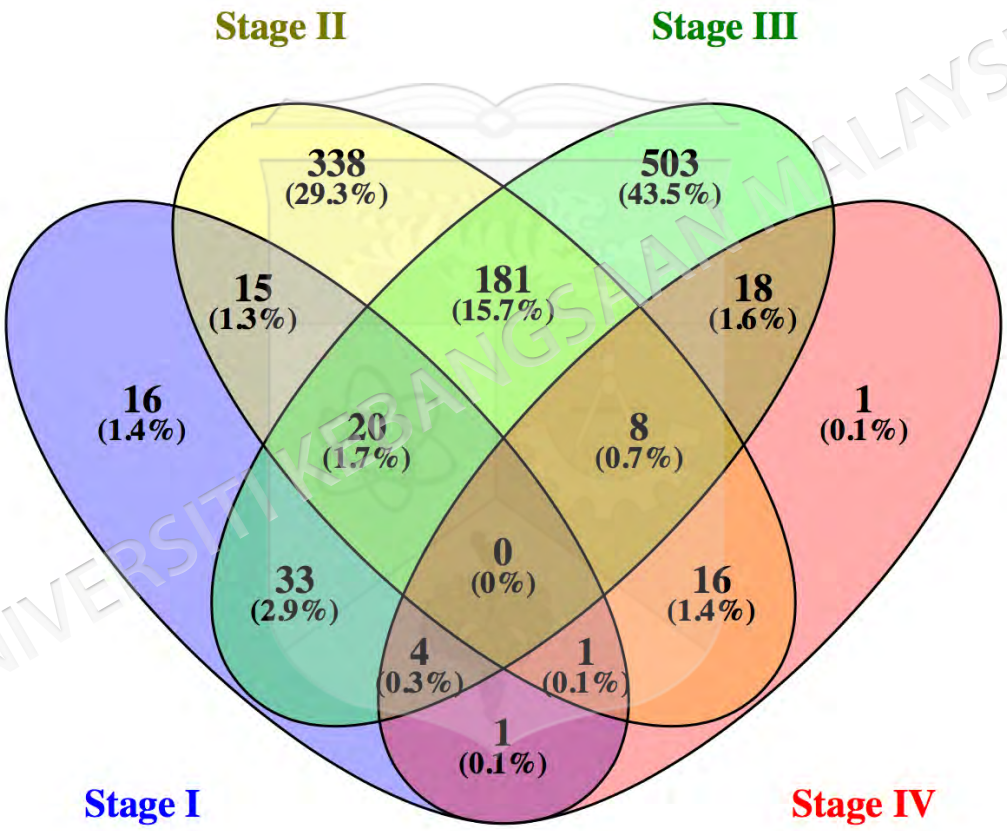


Figure 4.4 Venn's diagram of human proteins expression based on CRC stages. Stage I had 16 unique proteins, Stage II 338 proteins, Stage III 503 proteins and Stage IV had only one unique protein. Stage I shared 15 proteins in common with Stage II, 33 proteins with Stage III and 1 protein shared with Stage IV. Stage II shared 181 proteins with Stage III and 16 proteins with Stage IV. Meanwhile, Stage III and IV shared 18 proteins together. Stage I, II and III shared 20 proteins in common, four proteins were commonly shared between Stage I, III and IV and only one protein was shared between Stage I, II and IV. Lastly, eight proteins were shared between Stage II, III and IV.

c. Heat map analysis of exclusive human proteins in CRC

We then divided the CRC samples into two main CRC staging groups; early CRC (Stage I and II) and late CRC (Stage III and IV). The protein intensity patterns in both groups were observed via heat map analysis, to provide intuitive visualization of the data. Each coloured cell on the map corresponds to protein intensity value, with protein names in rows and biological samples in column. The top 100 proteins out of 1155 proteins exclusively identified in CRC were shown in Figure 4.5. Two clusters of different protein expressions were observed which subsequently discriminate early CRC from late CRC.

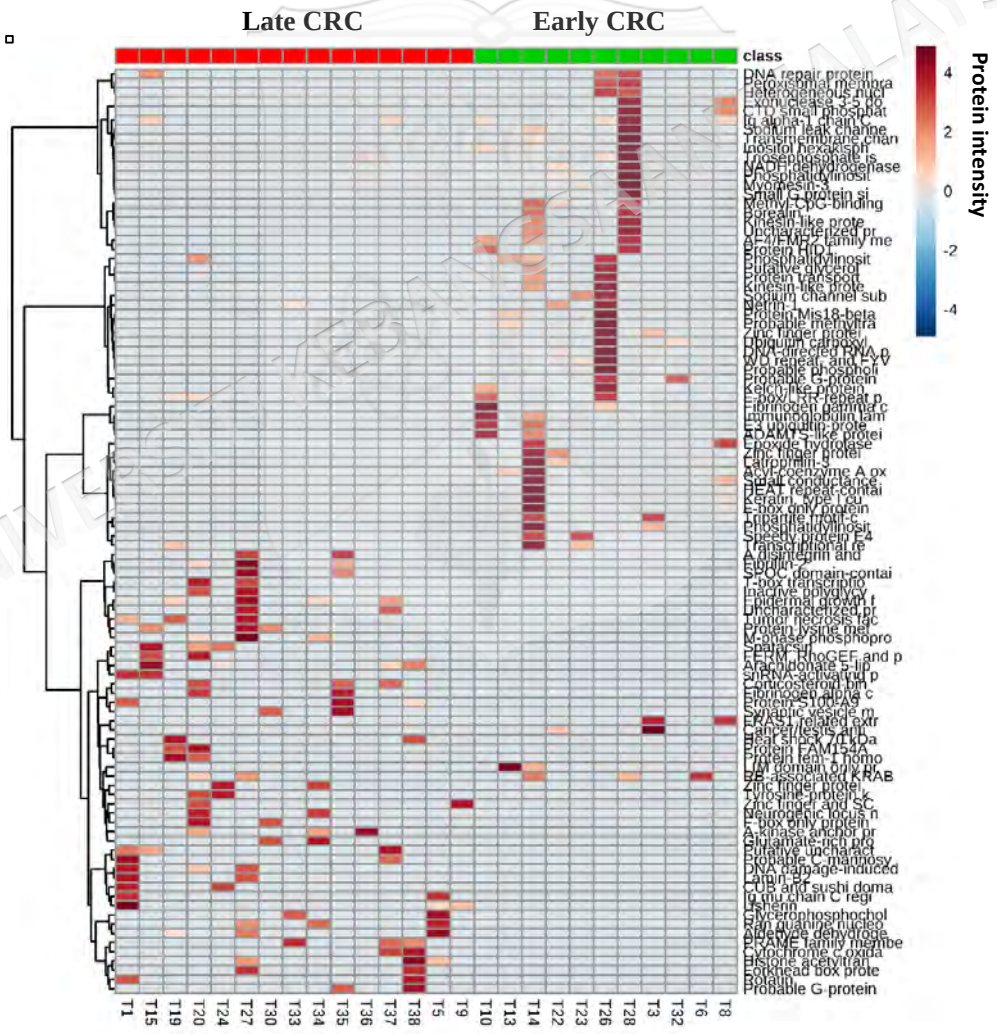


Figure 4.5 A heat map analysis of top 100 proteins identified exclusively in CRC samples. Based on the map, discovered proteins were arranged in rows and samples in column. Column are coloured according to group (green = early CRC; Stage I and II, red = late CRC; Stage III and IV). In the heat map image, grey indicates missing value.

d. Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA) of exclusive human proteins in CRC

An orthogonal partial least squares - discriminant analysis (orthoPLS-DA) was performed on the same data in addition to principle analysis component (PCA) (Appendix H), in order to observe the clustering of samples according to tumour stages. As shown in Figure 4.6, a distinguished separation between early and late CRC groups was observed.

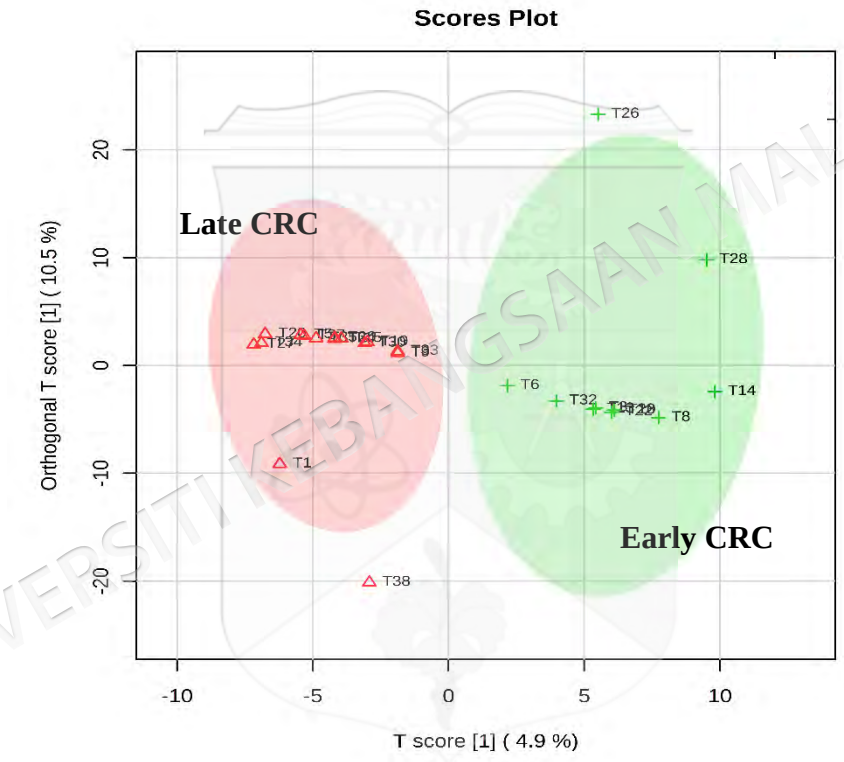


Figure 4.6 An orthogonal partial least squares - discriminant analysis (orthoPLS-DA) of human proteins exclusively identified in colorectal cancer (CRC) grouped according to tumour stages, early and late CRC. A distinct separation was observed between early stage of CRC and late stage CRC.

e. Functional annotation of exclusive human proteins in CRC

By using DAVID software, we had mined functional annotation for all 1155 CRC exclusive proteins based on gene ontology (GO) for biological processes. The top functional annotations were related to positive regulation of GTPase activity, cell adhesion, protein phosphorylation, intracellular signal transduction and positive regulation of cell proliferation (Figure 4.7).

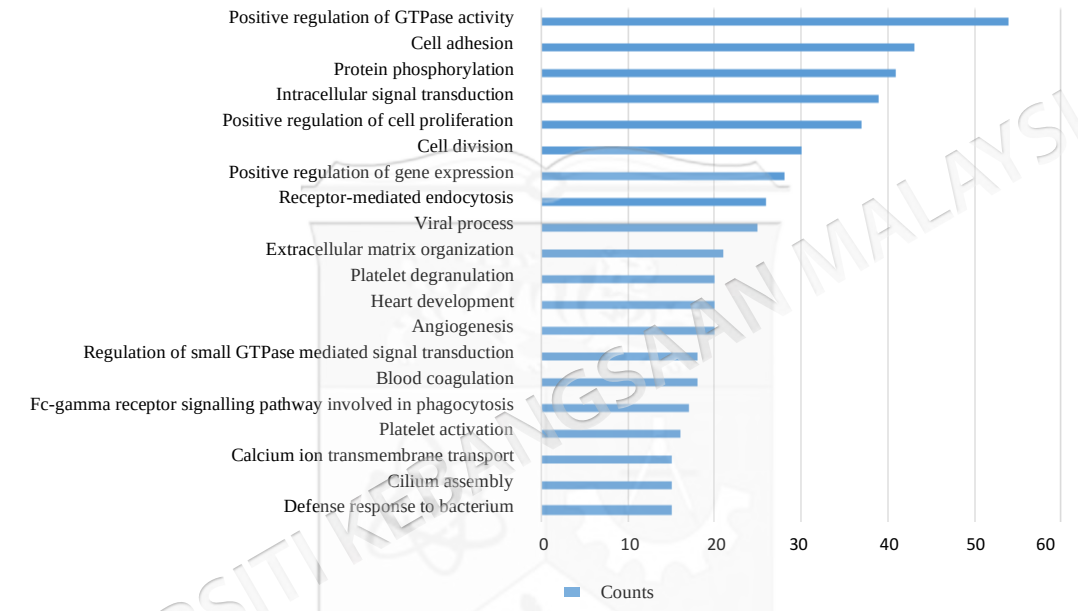


Figure 4.7 The top 20 functional annotation of CRC exclusive human proteins according to gene ontology (GO) of biological processes.

f. Protein - protein correlation of CRC exclusive human proteins

Besides that, we also analyzed the correlation between proteins using Spearman's rank correlation and the top 12 significant correlations with *p*-values less than 0.05 were shown in Table 4.2. The most significant correlations were between protein Olfactory receptor 4E2 with C-Jun-amino-terminal kinase-interacting protein 4 and Putative elongation factor 1-alpha-like 3 and correlation between protein C-Jun-amino-terminal kinase-interacting protein 4 with Putative elongation factor 1-alpha-like 3.

Table 4.2 Top 12 significant (p<0.05) protein - protein correlation of human proteins exclusive in CRC

No.	Protein IDs	Protein names	Correlation Coefficient	*p*-value
1	P07900;Q58FG0 - Q8NGC2	Heat shock protein HSP 90-alpha - Olfactory receptor 4E2	0.83287	1.30E-07
2	P07900;Q58FG0 - O60271	Heat shock protein HSP 90-alpha - C-Jun-amino-terminal kinase-interacting protein 4	0.83287	1.30E-07
3	P07900;Q58FG0 - Q5VTE0;P68104; Q05639	Heat shock protein HSP 90-alpha - Putative elongation factor 1-alpha-like 3;Elongation factor 1-alpha 1;Elongation factor 1-alpha 2	0.83287	1.30E-07
4	P07900;Q58FG0 - Q5VUA4	Heat shock protein HSP 90-alpha - Zinc finger protein 318	0.48152	0.012751
5	P07900;Q58FG0 - Q9NYW8	Heat shock protein HSP 90-alpha - RB-associated KRAB zinc finger protein	0.4091	0.037966
6	P07900;Q58FG0 - P68871;CON__P02070;CON__Q3SX09;P69892;P69891; P02100	Heat shock protein HSP 90-alpha - Hemoglobin subunit beta;LVV-hemorphin-7;Spinorphin	0.53086	0.0052675
7	Q8NGC2 - O60271	Olfactory receptor 4E2 - C-Jun-amino-terminal kinase-interacting protein 4	1.00	<0.001
8	Q8NGC2 - Q5VTE0;P68104; Q05639	Olfactory receptor 4E2 - Putative elongation factor 1-alpha-like 3;Elongation factor 1-alpha 1;Elongation factor 1-alpha 2	1.00	<0.001
9	O60271 - Q5VTE0;P68104; Q05639	C-Jun-amino-terminal kinase-interacting protein 4 - Putative elongation factor 1-alpha-like 3;Elongation factor 1-alpha 1;Elongation factor 1-alpha 2	1.00	<0.001
10	Q5VUA4 - A2VDJ0	Zinc finger protein 318 - Transmembrane protein 131-like	0.44662	0.022182
11	Q5VUA4 - Q9NYW8	Zinc finger protein 318 - RB-associated KRAB zinc finger protein	0.54183	0.0042491

to be continued…

…continuation

12	Q5VUA4 - P68871; CON__P02070; CON__Q3SX09; P69892;P69891; P02100	Zinc finger protein 318- Hemoglobin subunit beta;LVV-hemorphin-7;Spinorphin	0.44018	0.024424

g. Exclusive proteins in CRC based on gender and tumour staging

i. Gender

Mann Whitney analysis, was used to identify the most significant proteins based on gender. Sixteen (16) significant proteins were found with *p*-values less than or equal to 0.05. Fifth teen (15) proteins were significant in female and one (1) in male (Table 4.3). Some of the significant proteins found, were Epidermal growth factor-like protein 7, Serum albumin and Transmembrane protein 131-like. Some of these proteins were related to angiogenesis mechanism and promotes tumour growth (Hansen et al. 2014; Pinte et al. 2016).

Table 4.3 Sixteen human exclusive proteins in colorectal cancer (CRC) found significant in gender

No.	Protein IDs	Peptides identified	Protein names	Gender	*p*-value
1	A2VDJ0	308	Transmembrane protein 131-like	Female	0.05
2	P0CF74: P0DOY3; P0DOY2; A0M8Q6	2;61	Immunoglobulin lambda constant 6; Immunoglobulin lambda constant 3; Immunoglobulin lambda constant 2; Immunoglobulin lambda constant 7	Female	0.023
3	P30837	371	Aldehyde dehydrogenase X, mitochondrial	Female	0.023
4	Q06520	1221	Bile salt sulfotransferase	Female	0.023
5	Q13227	590	G protein pathway suppressor 2	Female	0.023
6	Q49A26	30	Putative oxidoreductase GLYR1	Female	0.023
7	Q5T5N4	1125	Uncharacterized protein C6orf118	Female	0.023
8	Q5T848	1305	Probable G-protein coupled receptor 158	Female	0.023
9	Q8N2C7	645	Protein unc-80 homolog	Female	0.023
10	Q8N6M0	755	OTU domain-containing protein 6B	Female	0.023
11	Q96Q35	829	Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 12 protein	Female	0.026
12	Q99547	796	M-phase phosphoprotein 6	Female	0.023
13	Q9HCM1	264	Uncharacterized protein KIAA1551	Female	0.023
14	Q9HD47	671	Ran guanine nucleotide release factor	Female	0.023
15	Q9UHF1	98	Epidermal growth factor-like protein 7	Female	0.007
16	P02768	44	Serum albumin	Male	0.017

ii. Tumour Staging

Kruskal-Wallis analysis was applied to identify proteins that were significant to CRC stages. A total of four (4) proteins found to be significant; two (2) in early CRC (Stage 1); Cancer/testis antigen 47B and Unconventional myosin-IA and another two (2) proteins in late stages CRC (Stage IV); Haptoglobin and Extracellular serine/threonine protein kinase FAM20C (Table 4.4).

Table 4.4 Colorectal cancer (CRC) exclusive proteins found to be significant based on tumour staging

No.	Protein IDs	Peptides identified	Protein names	CRC Stage	*p*-value
1	P0C2W7	1008	Cancer/testis antigen 47B	Stage I	0.001
2	Q9UBC5	95	Unconventional myosin-IA	Stage I	0.013
3	P00738;P00739	63;1369	Haptoglobin; Haptoglobin-related protein	Stage IV	0.007
4	Q8IXL6	682	Extracellular serine/threonine protein kinase FAM20C	Stage IV	0.029

4.4.3 Exclusively Identified Human Proteins in Control

a. Functional annotation of exclusive human proteins in control

Based on our Venn's diagram (Figure 4.2), 809 of the proteins were uniquely found in the control samples. The functional annotation of these proteins were analysed based on gene ontology (GO) of biological processes (Figure 4.8). Some of the top biological functions were related to positive regulation of GTPase activity, intracellular signal transduction, protein ubiquitination, viral process and protein autophosphorylation.

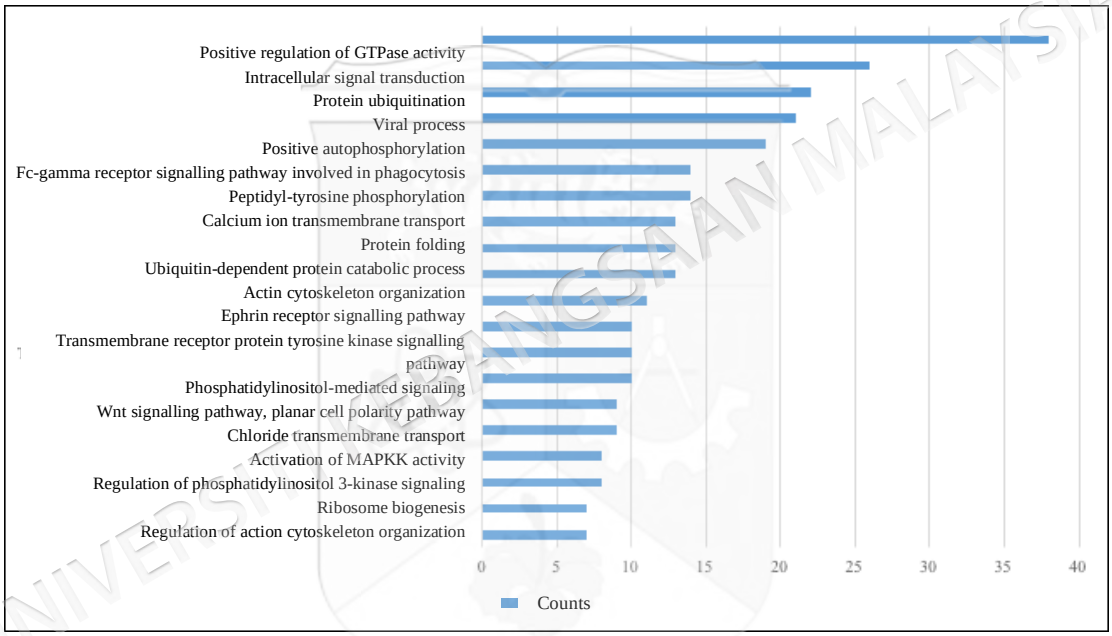


Figure 4.8 The top 20 functional annotation of control exclusive human proteins according to gene ontology (GO) biological processes.

b. Exclusive human proteins in control based on gender

Mann Whitney analysis, had identified four (4) significant proteins in gender. Three (3) proteins were significant in female and one (1) in male with *p*-values less than or equal to 0.05 (Table 4.5). The top proteins found were, Rho guanine nucleotide exchange factor 5, Transient receptor potential cation channel subfamily M member 1, DNA-directed RNA polymerase III subunit RPC7-like and Shugoshin 1.

Table 4.5 Four human exclusive proteins in control found significant in gender

No.	Protein IDs	Peptides identified	Protein names	Gender	*p*-value
1	Q12774	1178	Rho guanine nucleotide exchange factor 5	Female	0.033
2	Q7Z4N2	1908	Transient receptor potential cation channel subfamily M member 1	Female	0.031
3	Q9BT43	930	DNA-directed RNA polymerase III subunit RPC7-like	Female	0.031
4	Q5FBB7	117	Shugoshin 1	Male	0.031

4.4.4 Human Proteins Identified Both in Colorectal Cancer (CRC) and Control

a. Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA) of human proteins identified both in CRC and control

Sixty-eight proteins were found in both CRC and control samples. The clustering of CRC and control were indistinct with some CRC and controls were observed close to each other (Figure 4.9). The PCA plot of the data was included in Appendix I.

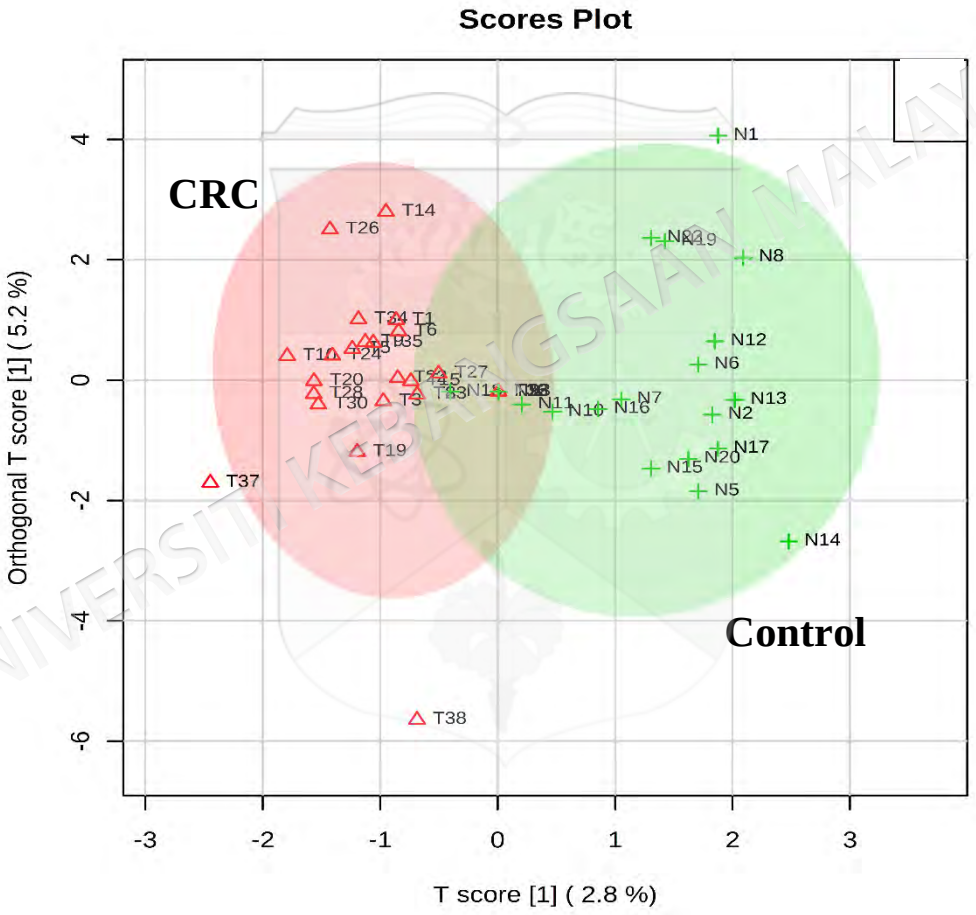


Figure 4.9 OrthoPLS-DA of human proteins discovered in colorectal cancer (CRC) and control. A less distinct samples clustering were observed between both groups. The CRC samples was shaded red and green for controls.

b. Fold change differences of human proteins identified in CRC and control

Out of the 67 proteins, we found 18 proteins that had fold change values more than or equal to 2.00 (Table 4.6). However, none of these 18 proteins achieved significant values.

Table 4.6 Shared proteins with fold change values more than or equal to 2.00

No.	Protein IDs	Peptides identified in CRC	Peptides identified in Control	Protein names	Fold change	*p*-value
1	P0C0L5; P0C0L4	664;793	113	Complement C4-B;Complement C4 beta chain;Complement C4-B alpha chain;C4a anaphylatoxin;C4b-B;C4d-B;Complement C4 gamma chain;Complement C4-A;Complement C4 beta chain;Complement C4-A alpha chain;C4a anaphylatoxin;C4b-A;C4d-A;Complement C4 gamma chain	442.1	0.082
2	Q4AC94	607	1042	C2 domain-containing protein 3	200.6	0.428
3	Q15262	1348	2212	Receptor-type tyrosine-protein phosphatase kappa	35.32	0.234
4	Q9Y6R7	28	547	IgGFc-binding protein	24.70	0.856
5	Q8N9W4	790	310	Golgin subfamily A member 6-like protein 2	17.84	0.875
6	Q6ZNG9	178	475	KRAB-A domain-containing protein 2	9.10	0.679
7	O00461	208	1120	Golgi integral membrane protein 4	8.65	0.875
8	Q8NCX0	149	502	Coiled-coil domain-containing protein 150	8.6	0.875
9	O60447	678	1486	Ecotropic viral integration site 5 protein homolog	7.72	0.717
10	Q8TE54	1061	620	Anion exchange transporter	6.36	0.321
11	Q96K78	813	2056	Probable G-protein coupled receptor 128	5.54	0.679
12	Q8NHM5	1260	1164	Lysine-specific demethylase 2B	4.88	0.437
13	Q13029	205	2592	PR domain zinc finger protein 2	3.37	0.251
14	Q6R2W3	1174	697	SCAN domain-containing protein 3	3.31	0.875
15	O94988	1175	1285	Protein FAM13A	3.04	0.875

to be continued…

…continuation

16	Q9BRR9	324	1887	Rho GTPase-activating protein 9	2.31	0.286
17	P09668	1169	1069	Pro-cathepsin H;Cathepsin H mini chain;Cathepsin H;Cathepsin H heavy chain;Cathepsin H light chain	2.19	0.875
18	O95251	1330	1409	Histone acetyltransferase KAT7	2.17	0.875

c. The significant human proteins in CRC and control samples

When we compared the 67 proteins, we discovered two (2) significant proteins and both were found in control samples with *p*-value less than and equal to 0.05 (Table 4.7). These proteins were Huntingtin and RNA exonuclease 5.

Table 4.7 Significant proteins found in common between colorectal cancer (CRC) and control

No.	Protein IDs	Peptides identified in CRC	Peptides identified in Control	Protein names	Fold change	*p*-value
1	P42858	601	2399	Huntingtin	-0.78	0.008
2	Q96IC2	268	1175	RNA exonuclease 5	-0.96	0.002

d. Receiver operating characteristics curve (ROC) analysis of significant proteins

i. ROC analysis for Huntingtin and RNA exonuclease 5 proteins

ROC analysis was carried out on significant proteins; Huntingtin and RNA exonuclease 5 proteins to predict its diagnostic characteristics as a biomarker. Based on the analysis, area under the curve (AUC) for both, Huntingtin and RNA exonuclease 5 proteins were less than 0.700. The specificity for both proteins were similar, 0.962. However, the positive predictive value (PPV) and negative predictive value (NPV) were less than 50.00 (Table 4.8 and Figure 4.10). According to the analysis, both proteins were not effective to be used for diagnostic testing.

Table 4.8 The predictive diagnostic characteristics of Huntingtin and RNA exonuclease 5 proteins

No.	Protein IDs	Protein names	Area under the curve (AUC)	Sensitivity	Specificity	Positive predictive value (PPV)	Negative predictive value (NPV)
1	P42858	Huntingtin	0.651	0.350	0.962	12.50	34.21
2	Q96IC2	RNA exonuclease 5	0.685	0.400	0.962	11.11	32.43

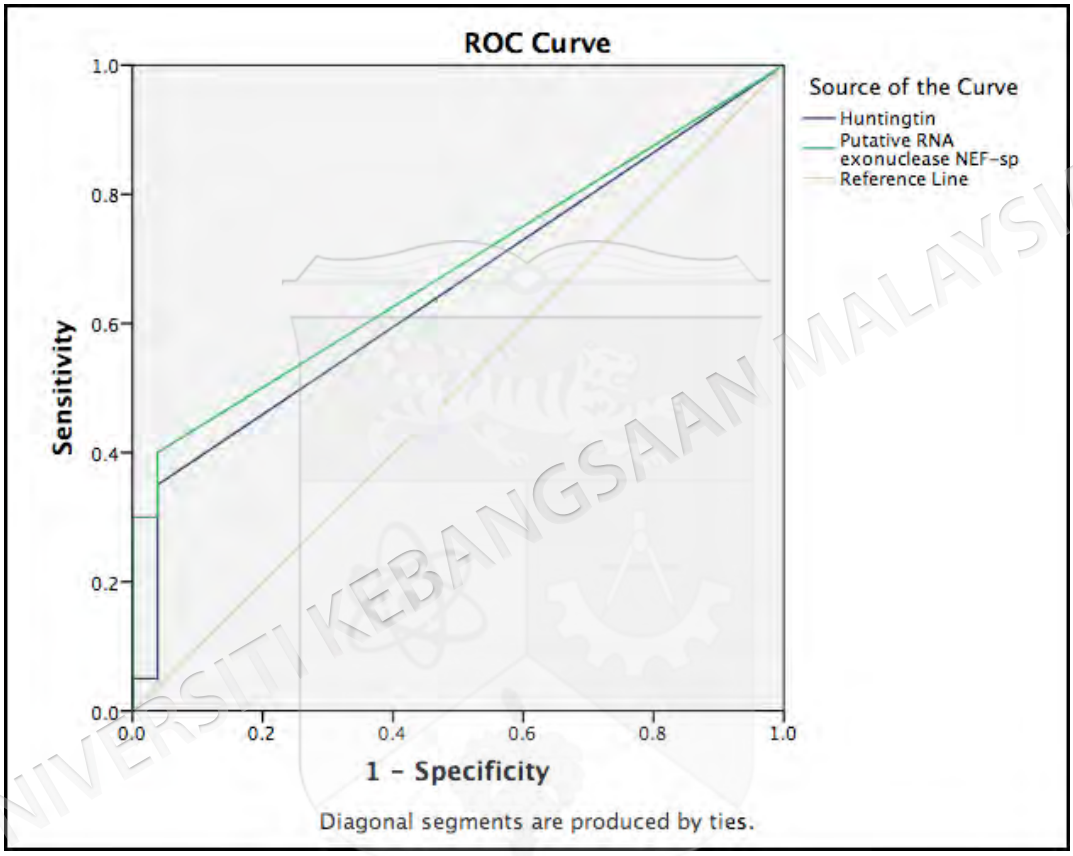


Figure 4.10 Receiver operating characteristics (ROC) curves for Huntingtin and RNA exonuclease 5 proteins.

ii. ROC analysis for combination of Huntingtin and RNA exonuclease 5 proteins

The combination of Huntingtin and RNA exonuclease 5 proteins yielded better AUC with values more than 0.700, improved sensitivity (0.923), specificity (0.50), PPV (70.59) and NPV (83.33) (Table 4.9 and Figure 4.11). The combination of proteins was considered to be sensitive but not specific, despite its ability to prove a true positive and a true negative between control from CRC.

Table 4.9 The predictive diagnostic characteristics for combination of Huntingtin and RNA exonuclease 5 proteins

No.	Protein combination	Area under the curve (AUC)	Sensitivity	Specificity	Positive predictive value (PPV)	Negative predictive value (NPV)	Negative predictive value (NPV)
1	Huntingtin and RNA exonuclease 5	0.715	0.923	0.50	70.59	83.33	34.21

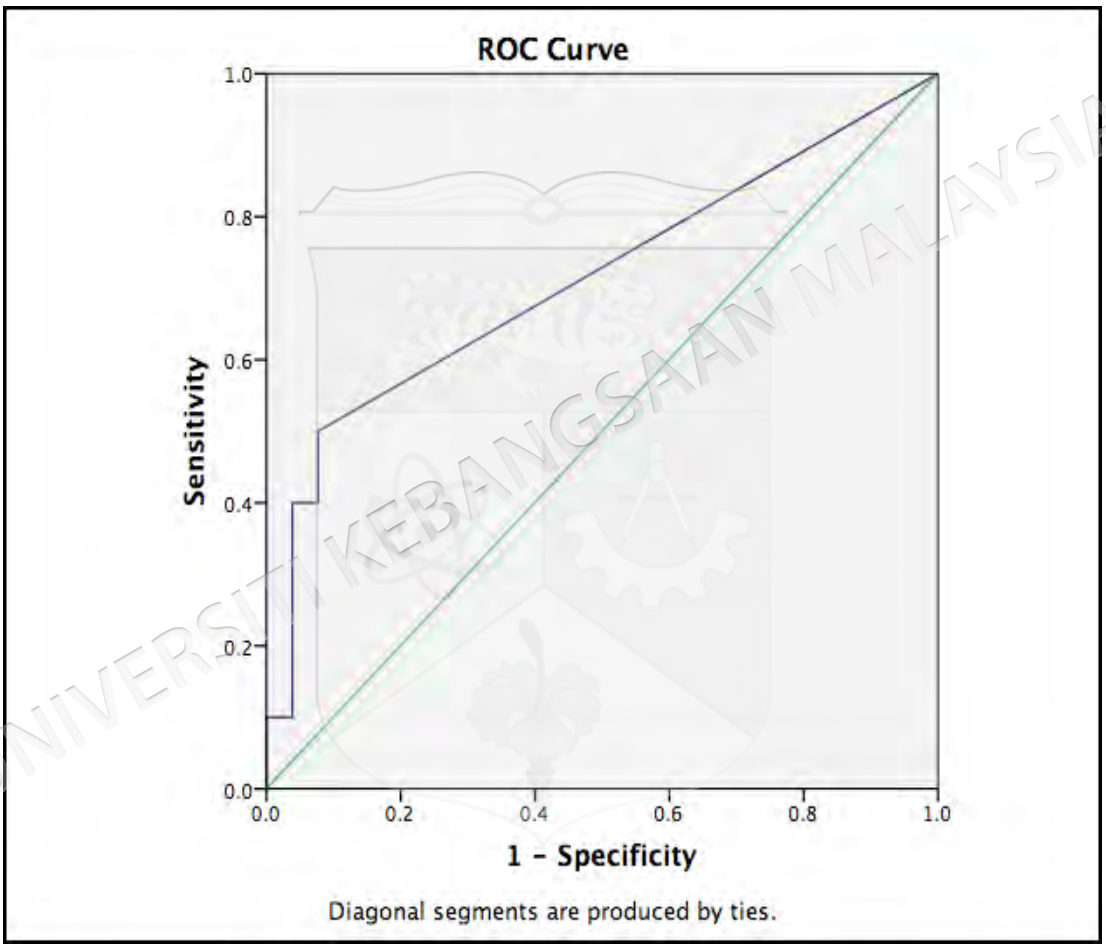


Figure 4.11 Receiver operating characteristics (ROC) curve of Huntingtin and RNA exonuclease 5 proteins combination.

e. Human proteins identified both in CRC and control based on gender

A statistical analysis was carried out to observe proteins that were significant in gender. We discovered only one (1) protein, Centromere protein F was significant in female (*p*-value less than or equal to 0.05) (Table 4.10).

Table 4.10 Significant protein identified in colorectal cancer (CRC) and control based on gender

No.	Protein IDs	Peptides identified	Protein names	Gender	Fold change	*p*-value
1	P49454	2538	Centromere protein F	Female	-0.26	0.043

4.5 THE IDENTIFIED MICROBIAL ORIGIN SECRETOME

4.5.1 Overview of Identified Microbial Proteins

Forty-six secretome samples were analysed and 2126 of microbial proteins were identified from the CRC and control samples. Based on Venn's diagram, most of the proteins were unique to control individuals as compared to CRC, and only one (1) bacterial protein was found in both samples (Figure 4.12).

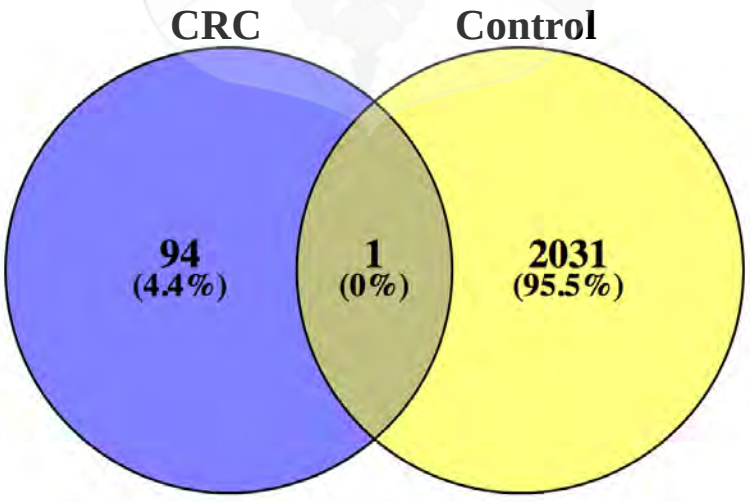


Figure 4.12 Venn's diagram of overall proteins from microbial origin identified in secretome of 46 samples, 26 CRC and 20 control with healthy colon morphology. Most of the proteins, were uniquely discovered in control samples (2031 proteins), meanwhile only 94 proteins were found in CRC and one (1) protein was found in both sample groups.

a. Heat map analysis of microbial protein expressions in CRC and control

Heat map analysis of the top 100 microbial protein intensity for CRC and control samples was shown in Figure 4.13. Most of the top 100 proteins were identified in control samples only. Among the top five proteins observed were E3 ubiquitin-protein ligase listerin protein (*Schizosaccharomyces pombe*), Type-2 restriction enzyme Sau3AI protein (*Staphylococcus aureus*), protein rep (*Streptomyces lividans*), Zinc finger protein ZPR1 (*Saccharomyces cerevisiae*) and 4-hydroxythreonine-4-phosphate dehydrogenase (*Campylobacter jejuni subsp. jejuni serotype O:2*).

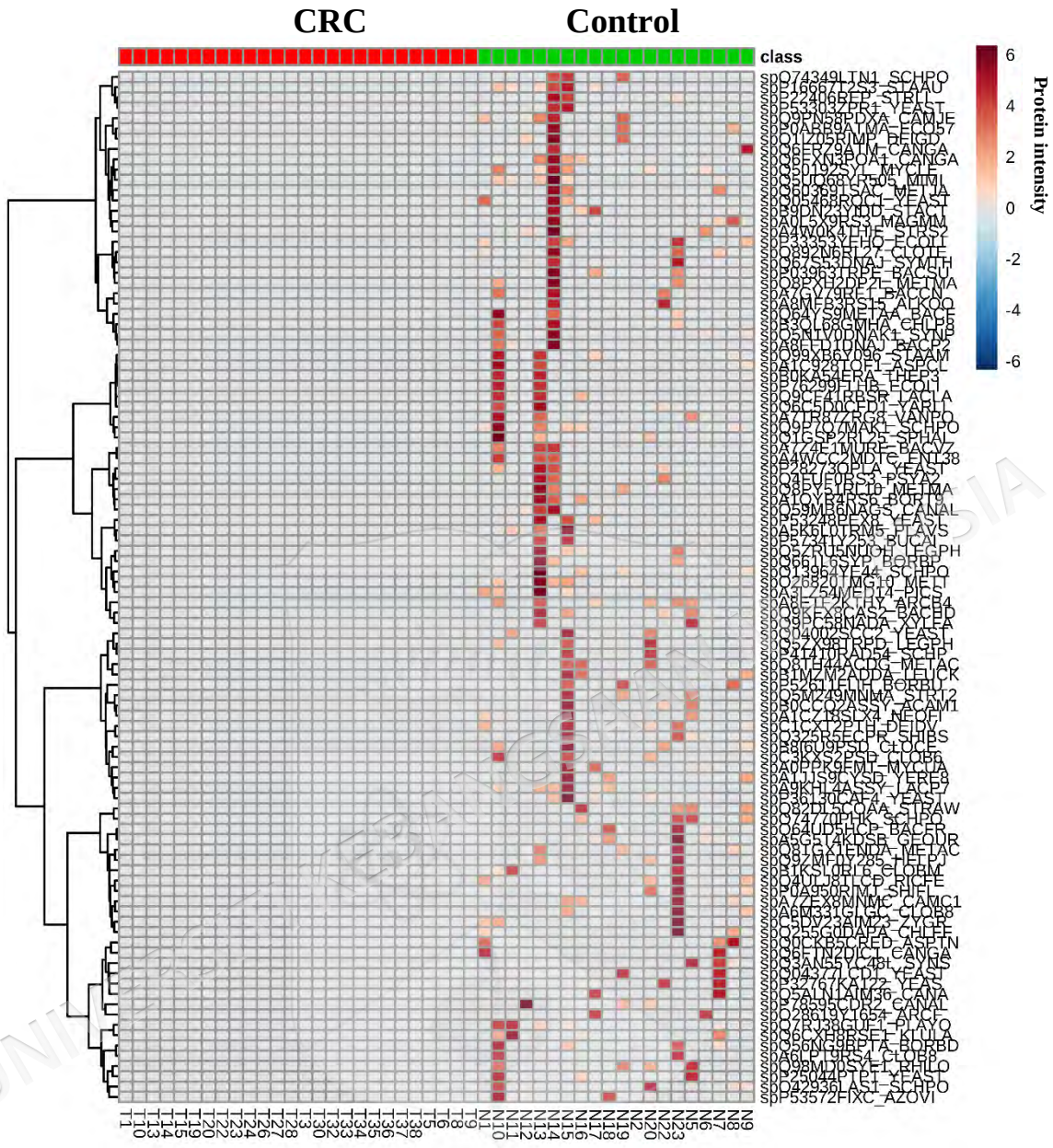


Figure 4.13 Heat map analysis of top 100 microbial protein intensity in CRC and control samples. Most of the top proteins were observed in control samples (green) as compared to CRC (red).

b. Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA) of microbial proteins in CRC and control

An orthoPLS-DA was carried out on microbial proteins from CRC and control. As shown in Figure 4.14, two different clustering were observed distinguishing both samples from each other. The PCA plot of the data was included in Appendix J.

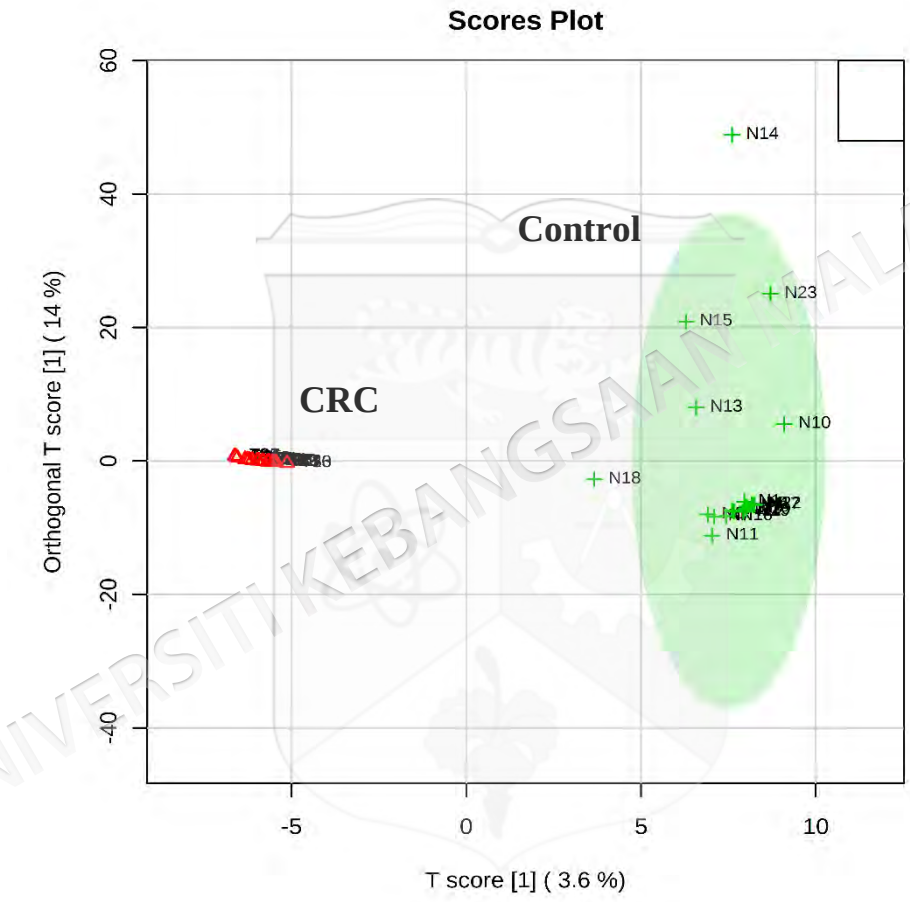


Figure 4.14 OrthoPLS-DA of microbial proteins found in colorectal cancer (CRC) and control. A distinguished clustering was observed between both groups, CRC (red) and control (green).

4.5.2 Classification of Identified Proteins Based on Microbial Species

A total of 2126 microbial proteins were discovered in our CRC and control samples. About 1354 proteins were mapped to 109 bacterial genera, 571 proteins were secreted by 23 genera of fungi, 128 proteins were recognised from eight (8) archaea genera, 54 proteins were secreted by virus from 11 genera and 19 proteins from three (3) eukaryota genera. Details of the top 10 genera for overall microbial organism were shown Table 4.11.

Table 4.11 Genus of microbial organism mapped to proteins recognised in CRC and control samples

Microbial Domain	Colorectal cancer	Control
Bacteria	Escherichia	Escherichia
	Xanthomonas	Bacillus
	Clostridium	Salmonella
	Campylobacter	Streptococcus
	Rickettsia	Staphylococcus
	Bacillus	Rickettsia
	Yersinia	Clostridium
	Bacteroides	Mycobacterium
	Helicobacter	Helicobacter
	Mycobacterium	Shewanella
Fungi	Saccharomyces	Saccharomyces
	Aspergillus	Schizosaccharomyces
	Candida	Aspergillus
	Schizosaccharomyces	Candida
	Meyerozyma	Nakaseomyces
	Nakaseomyces	Eremothecium
	Eremothecium	Kluyveromyces
	Kluyveromyces	Ustilago
	Yarrowia	Yarrowia
	Histoplasma	Vanderwaltozyma
Archaea	Archaeoglobus	Methanocaldococcus
	Methanocaldococcus	Methanosarcina
	Methanosarcina	Pyrobaculum
Eukaryota	Plasmodium	Plasmodium
		Manduca
		Phaeodactylum
Virus	Not present	Mimivirus
		Rotavirus
		Lambdavirus
		Muvirus

4.5.3 Exclusively Identified Proteins in Colorectal Cancer

Ninety-four microbial proteins were exclusively discovered in our CRC samples. These proteins were secreted by bacteria (59 proteins), fungi (32 proteins), archaea (2 proteins) and eukaryota (one protein).

a. Protein pathway analysis of exclusive microbial proteins in CRC

Based on STRING database, we analysed the 94 microbial proteins and found four different protein interactions. Two interactions involved bacteria (*Ureaplasma parvum* and *Mycoplasma synoviae*) and the remaining two interactions involved fungi (*Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*) (Figure 4.15, Figure 4.16, Figure 4.17 and Figure 4.18). These proteins; Energy-coupling factor transporter ATP-binding protein (ecfA1), DNA primase (dnaG), DNA-directed RNA polymerase subunit beta protein (rpoC), ATP synthase subunit beta (atpD), DNA repair protein rhp57 (rhp57), Structural maintenance of chromosomes protein 5 (smc5), SWR1 complex bromodomain subunit (bdf1), Actin (Act1), Cytochrome B2 (CYB2), Actin (Act1) and Chromatin-remodeling complex subunit IES6 (IES6), were related to transmembrane transport, DNA replication and repair, transcription regulation and chromatin modelling (Consortium 2018). However, no specific pathways were mapped to the interactions.

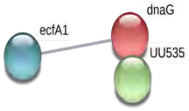


Figure 4.15 The overview of protein interactions for bacteria *Ureaplasma parvum*. A correlation was observed between Energy-coupling factor transporter ATP-binding protein (ecfA1) and DNA primase (dnaG).

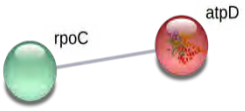


Figure 4.16 The overview of protein interactions for bacteria *Mycoplasma synoviae*. A single correlation between DNA-directed RNA polymerase subunit beta protein (rpoC) and ATP synthase subunit beta (atpD) was discovered.

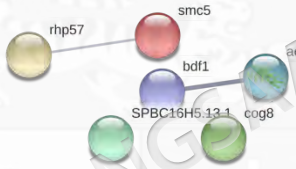


Figure 4.17 The overview of protein interactions for fungi *Schizosaccharomyces pombe*. Two protein interactions were discovered between DNA repair protein rhp57 (rhp57) and Structural maintenance of chromosomes protein 5 (smc5) and SWR1 complex bromodomain subunit (bdf1) and Actin (Act1).

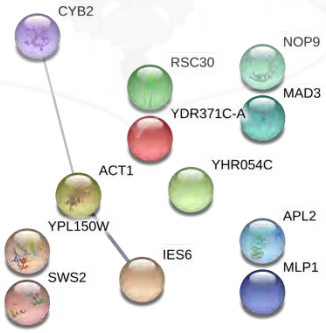


Figure 4.18 The overview of protein interactions for fungi *Saccharomyces cerevisiae*. Three proteins were observed to interact and connected with each other; Cytochrome B2 (CYB2), Actin (Act1) and Chromatin-remodeling complex subunit IES6 (IES6).

b. Exclusively identified microbial proteins based on CRC staging

We carried out a Venn's diagram analysis on 94 CRC exclusive proteins to determine the specificity of these proteins to any CRC stages. Based on Figure 4.19, we observed almost all stages had unique protein except for Stage I. All of Stage II, III and IV CRC had proteins that were secreted by bacteria, fungi and parasite species, including *U. parvum*, *S. pombe* and *S. cerevisiae* species. The full list of unique and shared proteins between these stages were compiled in Appendix T.

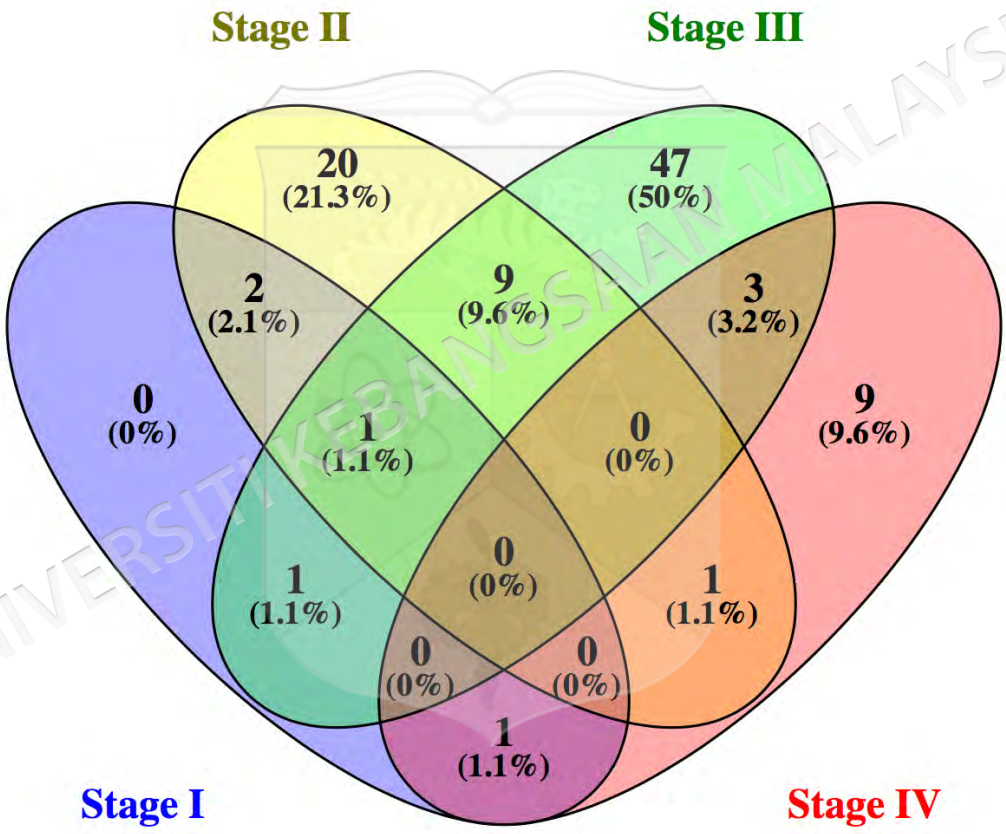


Figure 4.19 Venn's diagram of microbial proteins exclusively found in CRC classified based on CRC stages. There was no unique microbial protein identified in Stage I, but 20 microbial proteins were found in Stage II, 47 proteins in Stage III and 9 proteins in Stage IV. Stage I and II shared two proteins in common. Meanwhile, only one protein was shared between Stage I and III and Stage I and IV. Nine proteins were shared in common between Stage II and III and only one protein was shared between Stage II and IV. Stage III and IV shared three proteins and lastly only one protein was shared between three CRC stages (I, II and III).

c. Heat map analysis of exclusive microbial proteins in CRC

The protein intensity patterns in late and early CRC were observed via heat map analysis, to provide intuitive visualization of the data. Figure 4.20 shown the top 100 microbial proteins intensity identified in CRC secretome. A distinguished pattern of was observed between both groups, although the protein intensity was directed by individual samples.

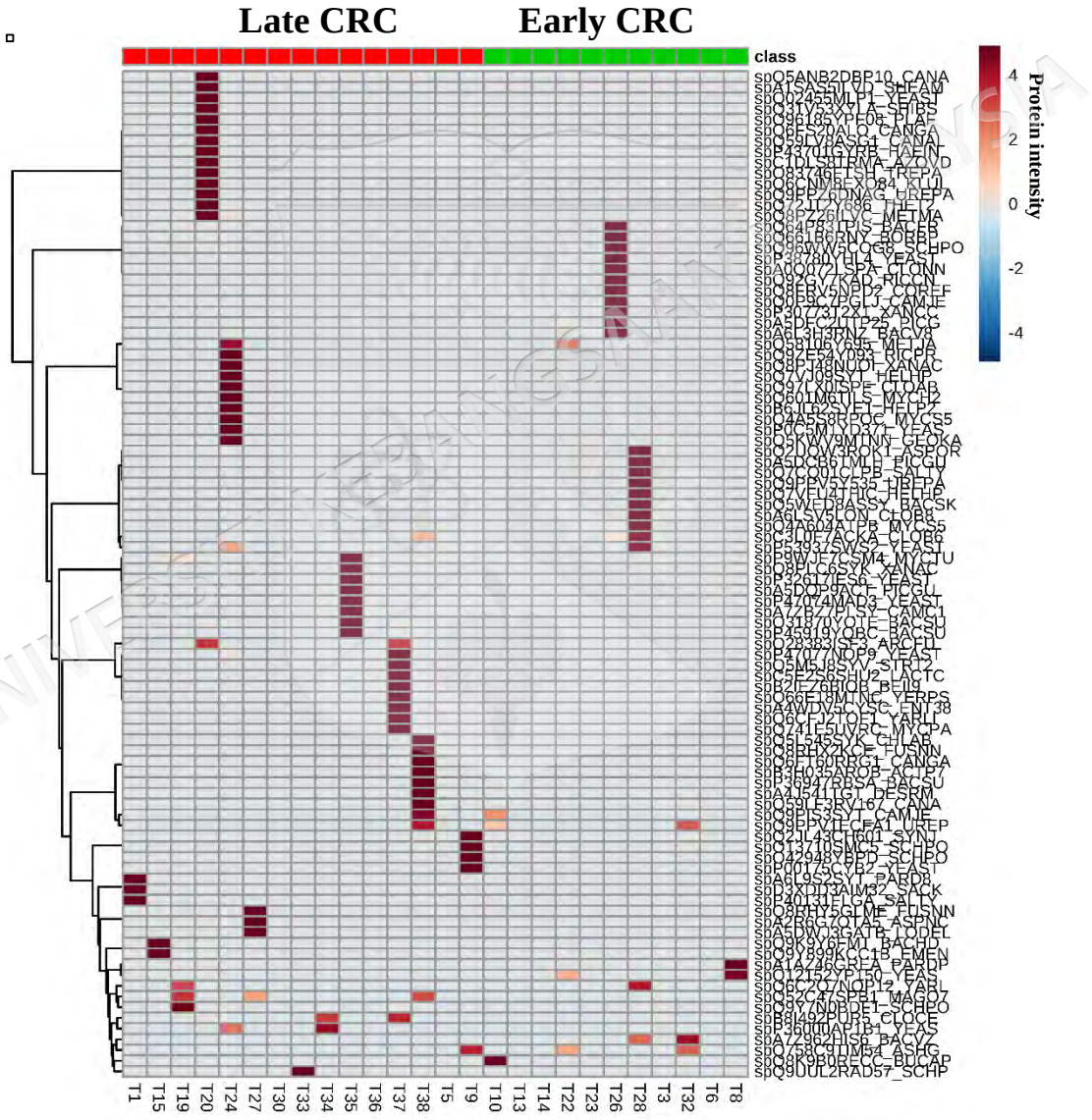


Figure 4.20 Heat map analysis of top 100 microbial proteins exclusively discovered in colorectal cancer (CRC). A distinct protein expression pattern was observed separating the early stages CRC (green) and late stages CRC (red).

d. Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA) of exclusive microbial proteins in CRC

OrthoPLS-DA was performed using microbial proteins in exclusive CRC based on early and late CRC stages. Figure 4.21 showed a distinctive clustering distinguishing both samples from each other. PCA plot on this data was compiled in Appendix K.

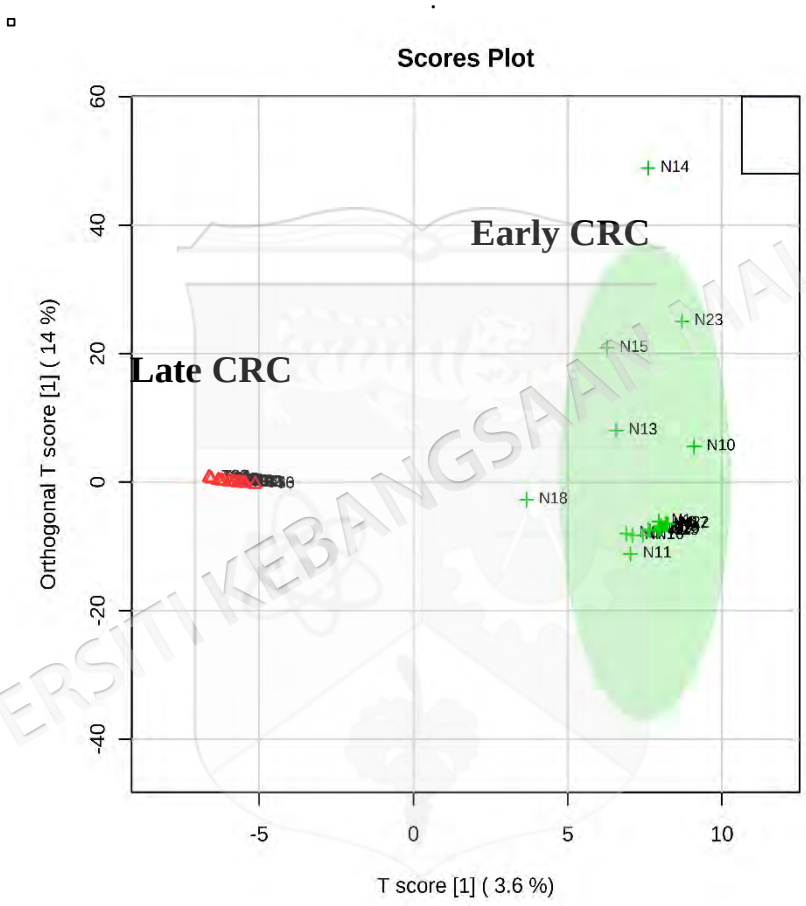


Figure 4.21 The orthoPLS-DA of microbial proteins exclusively identified in colorectal cancer (CRC). A clear separation was observed between early (green circle) and late (red circle) CRC samples.

e. Protein - protein correlation of CRC exclusive microbial proteins

The correlation between proteins is studied using Spearman's rank correlation. Top 28 significant correlations with *p*-values less than 0.05 were tabulated in Table 4.12.

Table 4.12 Top 28 protein - protein correlation of microbial proteins exclusive in CRC.

No.	Proteins correlated	Protein names and Microbe species	Correlation Coefficient	*p*-value
1	Q8PZ26 - Q72JT2;Q5SJG0	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - UPF0365 protein TT_C0686 *Thermus thermophilus*	0.5	0.0092937
2	Q8PZ26 - Q9PPZ6	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - DNA primase *Ureaplasma parvum serovar 3*	0.72111	3.24E-05
3	Q8PZ26 - Q6FS20	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - D-arabinono-1,4-lactone oxidase *Candida glabrata*	0.72111	3.24E-05
4	Q8PZ26 - Q6CNM8	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - Exocyst complex component EXO84 *Kluyveromyces lactis*	0.72111	3.24E-05
5	Q8PZ26 - Q5ANB2	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - ATP-dependent RNA helicase DBP10 *Candida albicans*	0.72111	3.24E-05
6	Q8PZ26 - Q59LV8	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - Activator of stress genes protein 1 *Candida albicans*	0.72111	3.24E-05
7	Q8PZ26 - Q02455	Ketol-acid reductoisomerase (NADP(+)) *Methanosarcina mazei* - Protein MLP1 *Saccharomyces cerevisiae*	0.72111	3.24E-05
8	Q72JT2;Q5SJG0 - Q9PPZ6	UPF0365 protein TT_C0686 *Thermus thermophilus* - DNA primase *Ureaplasma parvum serovar 3*	0.72111	3.24E-05
9	Q72JT2;Q5SJG0 - Q6FS20	UPF0365 protein TT_C0686 *Thermus thermophilus* - D-arabinono-1,4-lactone oxidase *Candida glabrata*	0.72111	3.24E-05
10	Q72JT2;Q5SJG0 - Q6CNM8	UPF0365 protein TT_C0686 *Thermus thermophilus* - Exocyst complex component EXO84 *Kluyveromyces lactis*	0.72111	3.24E-05
11	Q72JT2;Q5SJG0 - Q5ANB2	UPF0365 protein TT_C0686 *Thermus thermophilus* - ATP-dependent RNA helicase DBP10 *Candida albicans*	0.72111	3.24E-05
12	Q72JT2;Q5SJG0 - Q59LV8	UPF0365 protein TT_C0686 *Thermus thermophilus* - Activator of stress genes protein 1 *Candida albicans*	0.72111	3.24E-05
13	Q72JT2;Q5SJG0 - Q02455	UPF0365 protein TT_C0686 *Thermus thermophilus* - Protein MLP1 *Saccharomyces cerevisiae*	0.72111	3.24E-05
14	Q9PPZ6 - Q6FS20	DNA primase *Ureaplasma parvum* serovar 3 - D-arabinono-1,4-lactone oxidase *Candida glabrata*	1.00	<0.001

to be continued…

…continuation

15	Q9PPZ6 - Q6CNM8	DNA primase *Ureaplasma parvum* serovar - Exocyst complex component EXO84 *Kluyveromyces lactis*	1.00	<0.001
16	Q9PPZ6 - Q5ANB2	DNA primase *Ureaplasma parvum* serovar 3 - ATP-dependent RNA helicase DBP10 *Candida albicans*	1.00	<0.001
17	Q9PPZ6 - Q59LV8	DNA primase *Ureaplasma parvum* serovar 3 - Activator of stress genes protein 1 *Candida albicans*	1.00	<0.001
18	Q9PPZ6 - Q02455	DNA primase *Ureaplasma parvum* serovar 3 - Protein MLP1 *Saccharomyces cerevisiae*	1.00	<0.001
19	Q6FS20 - Q6CNM8	D-arabinono-1,4-lactone oxidase *Candida glabrata* - Exocyst complex component EXO84 *Kluyveromyces lactis*	1.00	<0.001
20	Q6FS20 - Q5ANB2	D-arabinono-1,4-lactone oxidase *Candida glabrata* - ATP-dependent RNA helicase DBP10 *Candida albicans*	1.00	<0.001
21	Q6FS20 - Q59LV8	D-arabinono-1,4-lactone oxidase *Candida glabrata* - Activator of stress genes protein 1 *Candida albicans*	1.00	<0.001
22	Q6FS20 - Q02455	D-arabinono-1,4-lactone oxidase *Candida glabrata* - Protein MLP1 *Saccharomyces cerevisiae*	1.00	<0.001
23	Q6CNM8 - Q5ANB2	Exocyst complex component EXO84 *Kluyveromyces lactis* - ATP-dependent RNA helicase DBP10 *Candida albicans*	1.00	<0.001
24	Q6CNM8 - Q59LV8	Exocyst complex component EXO84 *Kluyveromyces lactis* - Activator of stress genes protein 1 *Candida albicans*	1.00	<0.001
25	Q6CNM8 - Q02455	Exocyst complex component EXO84 *Kluyveromyces lactis* - Protein MLP1 *Saccharomyces cerevisiae*	1.00	<0.001
26	Q5ANB2 - Q59LV8	ATP-dependent RNA helicase DBP10 *Candida albicans* - Activator of stress genes protein 1 *Candida albicans*	1.00	<0.001
27	Q5ANB2 - Q02455	ATP-dependent RNA helicase DBP10 *Candida albicans* - Protein MLP1 *Saccharomyces cerevisiae*	1.00	<0.001
28	Q59LV8 - Q02455	Activator of stress genes protein 1 *Candida albicans* - Protein MLP1 *Saccharomyces cerevisiae*	1.00	<0.001

f. Exclusive microbial proteins in CRC based on gender and tumour staging

i. Gender

A Mann-Whitney analysis between the gender groups had found phosphoribosylformylglycinamidine cyclo-ligase protein secreted by *Clostridium cellulolyticum*. This protein was significantly found in female only (Table 4.13). Data mining on GO biological processes of this protein had shown its function in 'de novo' inosine monophosphate (IMP) biosynthetic process.

Table 4.13 The top 10 significant microbial proteins discovered in colorectal cancer (CRC) based on gender

No.	Protein IDs	Peptides identified	Protein names	Microbe species	Gender	*p*-value
1	B8I492	66	Phosphoribosylformylglycinamidine cyclo-ligase	*Clostridium cellulolyticum*	Female	0.007*
2	P47077; C8ZBK1; C7GV42; B3LQ92; A6ZPV9	84	Nucleolar protein 9	*Saccharomyces cerevisiae*	Female	0.068
3	O28383	19	Iron-sulfur flavoprotein AF_1896	*Archaeoglobus fulgidus*	Female	0.068
4	P36000	24	AP-1 complex subunit beta-1	*Saccharomyces cerevisiae*	Female	0.068
5	Q8PZ26	158	Ketol-acid reductoisomerase (NADP(+))	*Methanosarcina* mazei	Female	0.068
6	A2R6G7; A0A1R3RGJ2	168	Ochratoxin halogenase ota5	*Aspergillus niger carbonarius*	Female	0.206
7	A0Q072	29	Lipoprotein signal peptidase	*Clostridium novyi*	Female	0.206
8	A1SAS5	4	Dihydroxy-acid dehydratase	*Shewanella amazonensis*	Female	0.206
9	A4WDV5	39	Adenylyl-sulfate kinase	*Enterobacter sp.*	Female	0.206
10	Q66E18; B2K631; B1JIK5; A9R2Z7; A4TPN1; A1JP11	59	Enolase-phosphatase E1	*Yersinia pseudotuberculosis*	Female	0.206

The significant protein with *p*-value less than 0.05 was marked with asterisk (*)

ii. Tumour staging

Ten microbial proteins were found significant in Stage IV CRC (Table 4.19), nine (9) proteins were identified from bacterial species and one (1) protein belongs to fungi. The top GO biological processes of these proteins were attributed to glutamyl-tRNA aminoacylation, transcription, DNA-templated and L-methionine salvage from methylthioadenosine.

Table 4.14 Significant microbial proteins found in colorectal cancer (CRC) based on tumour staging

No.	Protein IDs	Peptides identified	Protein names	Gene ontology (GO) biological processes	Microbe species	Tumour stage	*p*-value
1	B6JL62	161	Glutamate-tRNA ligase 1	Glutamyl-tRNA aminoacylation	*Helicobacter pylori*	Stage IV	0.007
2	P0C5M1	165	Uncharacterized protein YDR371C-A	Not available	*Saccharomyces cerevisiae*	Stage IV	0.007
3	Q8PJ48; Q3BRN7; Q2P0W3; Q8P7U0; Q5GXU0; Q4UWB0	136	NADH-quinone oxidoreductase subunit I	Not available	*Xanthomonas axonopodis pv. citri, Xanthomonas campestris pv. Vesicatoria, Xanthomonas oryzae pv. oryzae, Xanthomonas campestris pv. campestris, Xanthomonas oryzae pv. oryzae, Xanthomonas campestris pv. campestris*	Stage IV	0.007
4	Q4A5S8	49	DNA-directed RNA polymerase subunit beta'	Transcription, DNA-templated	*Mycoplasma synoviae*	Stage IV	0.007
5	Q58106	40	Uncharacterized protein MJ0695	Not available	*Methanocaldococcus jannaschii*	Stage IV	0.025
6	Q5KWV9	93	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	L-methionine salvage from methylthioadenosine and nucleoside catabolic process	*Geobacillus kaustophilus*	Stage IV	0.007
7	Q601M6	138	tRNA(Ile)-lysidine synthase	tRNA processing	*Mycoplasma hyopneumoniae*	Stage IV	0.007
8	Q7VJ09	73	Threonine--tRNA ligase	Threonyl-tRNA aminoacylation	*Helicobacter hepaticus*	Stage IV	0.007
9	Q9ZE54	67	Uncharacterized protein RP093	Not available	*Rickettsia prowazekii*	Stage IV	0.007
10	Q97LX0	44	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway and terpenoid biosynthetic process	*Clostridium acetobutylicum*	Stage IV	0.007

4.5.4 Microbial Proteins Exclusively Identified in Control Based on Gender

Only two proteins were significant in female and these proteins were secreted by *Schizosaccharomyces pombe* and *Sphingopyxis alaskensis* (Table 4.15). However, only GO biological processes of *S. alaskensis* was found and this protein was related to translation.

Table 4.15 Significant microbial proteins found in control based on gender

No.	Protein IDs	Peptides identified	Protein names	Gene ontology (GO) biological processes	Microbe species	Gender	*p*-value
1	O13964	2226	Uncharacteri-zed protein C24C9.04	Not available	*S. pombe*	Female	0.013
2	Q1GSP2	261	50S ribosomal protein L25	Translation	*Sphingopyxis alaskensis* (*S. alaskensis*)	Female	0.031

4.5.5 Microbial Proteins Identified in Colorectal Cancer (CRC) and Control Based on Gender

There was only one (1) bacterial protein found in both sample groups. Protein Holliday junction ATP-dependent DNA helicase RuvA was secreted by *Borrelia recurrentis*. However, when we carried out statistical analysis with respect to gender, ethnicity and age, the *p*-value was not significant, less than or equal to 0.05 in all three categories (Table 4.16).

Table 4.16 Statistical outcome of bacterial protein found in both colorectal cancer (CRC) and control based on gender

No.	Protein IDs	Peptides identified in CRC	Peptides identified in Control	Protein names	Microbe species	Gender (*p-value)*	Fold change
1	B5RQK6; B5RLN1	56	1560	Holliday junction ATP-dependent DNA helicase RuvA	*Borrelia recurrentis, Borrelia duttonii*	0.826	0.232

CHAPTER V

DISCUSSION

5.1 STOOL CHARACTERISTICS OF CRC-STRICKEN INDIVIDUALS

Bristol stool scale (BSS) has been used as a guideline to describe stool characteristics and estimate transit time assisting medical practitioners in ruling out bowel related issues (Degen and Phillips 1996; Heaton et al. 1992; Lewis and Heaton 1997; Vandeputte et al. 2015). BSS is comprised of seven (7) points scale separated by three main category, the hard stool form (Type 1 or 2), normal stool form (Type 3, 4 or 5) and loose stool form (Type 6 or 7) (Lee et al. 2017). The hard stool forms were attributed to longer transit rate, meanwhile the loose forms were equivalent to fast transit rate (Vandeputte et al. 2015). Protracted transit time could reduce moisture content in stools causing it to be hard and eventually affects stool microbial affluence (Vandeputte et al. 2015).

Change in bowel habit (such as constipation or diarrhea) is one of many symptoms for CRC (Hamilton and Sharp 2004). However, according to a prospective study, having a loose stool is likely predictive for CRC (Selvachandran et al. 2002). In our sample cohort, most stools from the CRC stricken gut were categorised as loose form with visible blood traces meanwhile, most stools from the control group were categorised as normal form. Loose stools differs to diarrhea based on bowel motions frequency per day and stools consistency (Kune 2012). However, both situations risked inflammation to colon mucosa and facilitate progressive growth of cells known as neoplasia (Kune 2012; Park et al. 2009). This condition could probably triggered by increased production of prostaglandin E_2 (PGE_2), known to induce aberrant colon motility and foster diarrheal state in rabbit model (Burakoff and Percy 1992). PGE_2 is highly expressed in CRC tissue as well as in vitro CRC stem cells (Arai et al. 2016;

Wang et al. 2015). Quite a number of studies had supported loose stools to correlate with increase risk of CRC (Inoue et al. 1995; Park et al. 2009; Simons et al. 2010). However, a few studies had omitted association between stool consistency with CRC and stressed on rectal bleeding as one of the major symptoms, accompanied by other indications such as loss of weight and alteration of bowel pattern (Adelstein et al. 2011; Astin et al. 2011; Olde Bekkink et al. 2009; Otani et al. 2006). In our opinion, all of the mentioned CRC symptoms should be considered when ruling out the disease supported by information on patient's genetic and epigenetic factors.

5.2 UTILIZATION OF STOOL FOR SECRETOMIC STUDY

Stools are composed of more than 70% of water and the remaining percentage are made up of solid fractions which are divided into two, organic and inorganic compositions (Rose et al. 2015). The organic composition is mainly comprised of microbial biomass followed by protein, carbohydrate, non-nitrogenous undigested plant fiber and undigested fat (Rose et al. 2015; Stephen and Cummings 1980; Volk and Rummel 1987). While, the inorganic composition is made up from calcium phosphate, iron phosphate, secreted intestinal fluid, small amount of shed colonocytes and mucus (Iyengar et al. 1991; Loktionov 2007; Rose et al. 2015).

Besides, stools are also high in microbial constitution and regularly bathe with host's intestinal protein in the gut lumen. Therefore, we reckon it as a suitable specimen for the present study. Although we do aware that harvesting secretome from stool specimen is relatively new, but the potential of the approach promises an outcome that may enlighten us on the host-microbe interactions and possible crosstalks in a tumour microenvironment.

Proteins are able to enter the gut's lumen via three mechanisms; leakage, secretion and exfoliation (Osborn and Ahlquist 2005; Wang et al. 2014). Leakage of proteins into gut lumen and stool may occur when there is a bleeding or blood and lymph vessels damage on tumour surface (Osborn and Ahlquist 2005; Wang et al. 2014). Hemoglobin, leukocyte protein (calprotectin, lactoferrin and lysozyme) and serum proteins (α1-antitrypsin and albumin) are example of proteins that could be found

in gut lumen (Dubrow et al. 1992; Dubrow and Yannielli 1992; Gilbert et al. 1996; Nakayama et al. 1987; Osborn and Ahlquist 2005; Roseth et al. 1995). Besides that, the host colon and cancer cells themselves are secreting abundance of molecules such as mucin, carcinoembryonic antigen (CEA) and mini chromosome maintenance proteins (MCM2) (Wang et al. 2014). Furthermore, exfoliated colonocytes resulted from increased cell turnover rate of tumour as compared to normal epithelium may also be found (Osborn and Ahlquist 2005; Wang et al. 2014) in which subsequently, leads to abundance of exfoliated colonocytes trapped in the mucus layer (Loktionov 2007; Osborn and Ahlquist 2005). Apart from that, collected secretome can be contaminated with intracellular proteins, as a result from cell or microbial autolysis and non-classical transportation of proteins to the outer environment (Loei et al. 2012; Makridakis and Vlahou 2010; Rabouille 2017). In the present study, we indeed identified proteins possibly related to the leaked mechanism in our secretome samples, which are coherent to recently mentioned information.

A proper filtration step is necessary to filter out cells and stool's undigested materials from the collected secretome. Prior to the analysis, we have carried out a proof of concept experiment using PVDF filter with 0.22 µm membrane size. This membrane size was chosen since, most bacterial and archaea cells size are approximately between 0.2 to 750 µm (Ginzberg et al. 2015; Levin and Angert 2015; Lyth et al. 2018; Milo and Phillips 2015; Portillo et al. 2013; Wang and Lin 2012). The size of fungi cell are roughly around 2 to 16 µm and parasites such as *Plasmodium falciparum* is close to 1 to 1.5 µm (Ginzberg et al. 2015; Levin and Angert 2015; Lyth et al. 2018; Milo and Phillips 2015; Portillo et al. 2013; Wang and Lin 2012). Meanwhile, the size of human cells span a wide range between 30 to 105 µm, except for viruses which appeared to be smaller (17 - 1000 nm) (Milo and Phillips 2015). The obtained secretome filtrate was cultured in agar and broth media and no bacterial growth was observed.

During recruitment, an instruction sheet on stool collection steps was given to each patient (Appendix O). They were thoroughly briefed on the self-collection procedures and were handed a medium size foam box (Appendix N) containing one (1) unit of stool container and several packets of gel packs. The stool container helps to keep specimen from environmental contamination, while the gel pack aid to maintain

cool temperature during transportation and eventually prevent significant microbial growth alteration (Choo et al. 2015; Franzosa et al. 2014; Kim et al. 2017; Tedjo et al. 2015; Wu et al. 2018). Stool microbiota are mostly anaerobic thus, it is recommended to be process and store at ultra low temperature within six (6) to 48 hours after defecation (Terveer et al. 2017; Wu et al. 2018).

5.3 CRC STOOL SAMPLES EPIDEMIOLOGY AND CLINICAL DATA

Most of our CRC specimens were from Malays (58%) followed by Chinese (38%) and Indian (3.8%). Based on staging, 50% of our recruits were diagnosed with Stage III CRC and the later stages (Stage III and IV) were mostly suffered by Malay group. Our findings were coherent with a local review on five-year CRC histopathology that manifested a similar trend (Azmi et al. 2007; Ghee 2014). Almost 62% of our CRC samples were from male and 38% were from female, with majority aged 50-year and above. The large number of CRC male patients have been highlighted in several studies as well (Abu Hassan et al. 2016; Ferlay et al. 2014). The age factor has been well established as one of the main contributor towards increasing risk of CRC (Frampton and Houlston 2016). Previously, the American Cancer Society has suggested for routine CRC screening to be done at the age of 50 until a recent guideline recommended a lower screening age (starting at 45 years old), in concomitant to rising CRC incidence among the young adults (Dyer 2018).

5.4 THE RELATIONSHIP BETWEEN TOP HUMAN PROTEINS IDENTIFIED IN HEATMAP ANALYSIS WITH CRC STAGES

Based on Figure 4.5, heatmap analysis, the top 50 proteins out of 100 had high protein intensity in the early CRC group with three (3) topmost proteins include, DNA repair protein REV1, Peroxisomal membrane protein 4 and Heterogeneous nuclear ribonucleoprotein Q. In reference to the Venn's diagram (Figure 4.4), both Peroxisomal membrane protein 4 and Heterogeneous nuclear ribonucleoprotein Q were mapped to Stage II samples, whereas DNA repair protein REV1 was expressed in Stage II and III.

DNA repair protein REV1, is one of Y-family DNA polymerase that participates in DNA translesion tolerance responsible in blocking unrepaired damage area of DNA from further replication (Guo et al. 2009; Lange et al. 2011; Lin et al. 1999). REV1 is not able to synthesize DNA polymer but it possesses a deoxycytidyl transferase activity that enable it to insert deoxycytidine monophosphate (dCMP) opposite guanine, uridine and at apurinic or apyrimidinic (AP) sites (Guo et al. 2009; Lange et al. 2011; Lin et al. 1999). Suppression of REV1 in mice caused chromosome instability, which had led to spontaneous cancer development and single nucleotide polymorphism in the REV1 gene (Guo et al. 2009; Lange et al. 2011; Xie et al. 2010), hence increased cancer risk (Guo et al. 2009; Lange et al. 2011; Xie et al. 2010). We observed lack of REV1 expression in our late stage CRC samples, thus we hypothesize the absent of this protein could probably facilitate progression of disease.

Peroxisomal membrane protein 4 is encoded by PXMP4 gene (Consortium 2018). A proteome analysis using iTRAQ labeling method and verification by SRM/MRM has identified 20 proteins including PXPM4 protein to be significantly overexpressed in early CRC tissue and predicted to be one of the biomarkers for clinical diagnosis of CRC (Kume et al. 2014), in agreement to our findings.

Heterogeneous nuclear ribonucleoprotein Q protein or hnRNP Q, plays an important role in RNA processing and its suppression caused proliferation rate to decline in CRC cell line (Yoo et al. 2009). The hnRNP Q protein reported to interact with Galectin-3 (protein responsible in regulating cancer cell proliferation), and its stability is dependent on the interaction (Yoo et al. 2009). Suppression of Galectin-3 caused hnRNP Q down regulation and leads to reduce CRC cells proliferation (Yoo et al. 2009). Outcome of the study also suggested, hnRNP Q protein as a target for 5-Fluorouracil chemotherapeutic agent to halt cell proliferation (Yoo et al. 2009).

5.5 PREDICTIVE DIAGNOSTIC OF HUNTINGTIN AND RNA EXONUCLEASE 5 PROTEINS AS CRC BIOMARKERS

The receiver operating characteristic (ROC) curve is frequently being used as a prediction model to study the reliability of a model to distinguish positive from negative

cases (Lasko et al. 2005). Our findings have shown that the combination of Huntingtin and RNA exconuclease 5 proteins as a predictive diagnostic model (Table 4.9 and Figure 4.11) was sensitive but not specific. The protein combination was shown to be able to distinguish true positive CRC cases yet, not fit to discriminate true negative healthy control from the CRC cases (Ekelund 2012). The justification for such outcome was probably due to small sample size. According to Hanczar et. al., an area under the curve (AUC) can be inadequately evaluated if the samples size is less than or equal to 100 ($N \leq 100$) and this may lead to inaccurate estimation of validated parameters (Hanczar et al. 2010).

5.6 FUNCTIONAL ANNOTATION OF CRC EXCLUSIVE HUMAN PROTEINS

Among the top functional annotations for our identified CRC exclusive proteins were cell adhesion, intracellular signal transduction and angiogenesis. Cell adhesion is an important attribute for tissue construction (Lodish et al. 2000). The cell adhesion molecules (CAMs), refers to group of transmembrane proteins on cell surface responsible in helping cells to bind with other cells or extracellular matrix (Lodish et al. 2000; Nasser and Spencer 2017). During metastasis event, most cells will loss its intercellular adhesion capability that enable it to enter blood or lymphatic stream and due to that, CAMs were linked to the initiation and progression of cancer (Okegawa et al. 2004). The intracellular signal transduction is a chain of signals received from cell surfaces then transmitted to various intracellular targets initiated by ligand binding that occur in a cell (Cooper and Hausman 2000). These signaling cascades are responsible in regulating cell proliferation, differentiation, angiogenesis, apoptosis and senescence (Adjei and Hidalgo 2005). In cancer, most of the cellular signaling network have been altered genetically or epigenetically resulting in ‘different’ features commonly acquired by most cancer cells (Hanrahan et al. 2014). Meanwhile, the angiogenesis describe formation of new blood vessels, but in cancer setting the new blood vessels will sprout from the existing vessels (Mousa et al. 2015; Rmali et al. 2006). This step is important to ensure supplies of nutrients and oxygen for growing tumour tissue (Sun 2012). The functional annotations that we observed in our findings enable us to conclude that most of the proteins identified from the gut secretome of CRC individuals showed the

characteristics that support the development of CRC in the individuals and most importantly, were able to be extracted non-invasively from the stool samples.

5.7 THE ROLE OF CRC EXCLUSIVE HUMAN SECRETOME IN KEGG ANNOTATED PATHWAYS

Overall, we have identified over thousands of human secretome proteins exclusive in CRC and related to ten (10) different KEGG annotated pathways (Figure 4.3) in particular the Hypoxia-inducible factor-1 (HIF-1) signaling pathway, Oocyte meiosis and Leukocyte transendothelial migration. Hypoxia-inducible factors 1 (HIF-1) signaling pathway is responsible to help cancer cells discern and acclimatize in hypoxic environment (Nagy 2011). Hypoxia describe condition with insufficient oxygen supplies to the surrounding cells (Eales et al. 2016; Muz et al. 2015; Nagy 2011). In cancer, rapid proliferation of the cancer cells exhaust oxygen supplies and increase the distance of cells from existing vessels (Eales et al. 2016; Muz et al. 2015; Nagy 2011). In order to keep up with the demand, HIF-$\propto$ one of HIF-1 subunit will be overexpressed and subsequently activates genes that were responsible in angiogenesis, the formation of blood vessels from the existing ones (Shi and Fang 2004). HIF-1 expression was reported to be markedly upregulated in immunohistochemical staining of tumour biopsies for most cancers (Shi and Fang 2004; Talks et al. 2000).

Meiosis is a cell division cycle following DNA replication that equally segregate chromosome pair to gametes (Kim et al. 2015). Oocytes enter meiosis cycle during fetal development, halt at prophase I until being induced by luteinizing hormone during puberty to complete the first meiotic division (Cohen and Holloway 2010; Kim et al. 2015). At second meiotic division, oocytes are arrested again at metaphase II before released during estrous cycle and the cell division is complete upon fertilization (Cohen and Holloway 2010; Kim et al. 2015). The association of oocyte meiosis with CRC development is still scarce in literature. However, there are studies that relate the similarity in behavior of cancer cells with oocyte (Majerus 2002; Terret et al. 2013).

Leukocytes or white blood cells are important in defending body from infectious agents (Ashton 2010). The human vascular system is comprised of large number of

connected blood vessels surrounding the tissues which enable immune cells to travel across the stream (Schimmel et al. 2016). Inflammation is a common occurrence among CRC patients that resulted from impairment of mucosal barrier (Heidemann et al. 2006; Janakiram and Rao 2014). During this event, local leukocytes will migrate to site of inflammation in amoeboid manner on apical surface of endothelial cells that lined the luminal site of blood vessels (Muller 2011; Schimmel et al. 2016). The adhesion molecules expressed on the leukocytes' surface and the counter-receptors on endothelial cells guide the leukocytes drifting towards endothelial border, crossing the endothelium and subendothelial basement membrane, then enter the interstitial tissue (Ley et al. 2007; Muller 2011; Rijcken et al. 2007; Strell and Entschladen 2008). The infiltration of immune cells through the blood vessels are recognized as transendothelial migration (Schimmel et al. 2016) and this phenomena is commonly reported in cancer (Barry et al. 2013; J. M. Drake et al. 2009; J. Drake et al. 2008; Enarsson et al. 2007; Laferrière et al. 2001; Peyri et al. 2009; Reymond et al. 2012)

5.8 PROTEIN - PROTEIN CORRELATION OF CRC EXCLUSIVE HUMAN PROTEINS

According to our result (Table 4.2), Heat shock protein HSP 90-alpha (Hsp90α) was significantly correlated with several proteins. However, a literature search has shown an interesting relationship between HSP90α with Putative elongation factor 1-alpha-like 3.

Hsp90α is a cytoplasmic chaperone and part of Hsp90 superfamily (Milicevic et al. 2008). All of Hsp90 family performed similar roles, protect against apoptosis, promote cell proliferation and transformation (Hostein et al. 2001; Milicevic et al. 2008; Münster et al. 2001; Pandey et al. 2000). Hsp90 is essential in various signaling pathway, folding, conformation and stability of cancerous proteins (Milicevic et al. 2008; Pratt and Toft 2003; Workman 2004). Besides that, Hsp90 acted as a final activator for numerous regulatory cells that responsible in invasion, angiogenesis and metastasis (Maloney and Workman 2002; Milicevic et al. 2008; Sőti et al. 2005). The overexpression of Hsp90 family in CRC were associated with metastasis and poor outcome including CRC initiation (Milicevic et al. 2008). Putative elongation factor 1-

alpha-like 3 or eEF1A1 is involved in assisting GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis (Consortium 2018). In addition, differential expression of eEF1A1 was previously reported in laryngeal carcinoma tissue (Li et al. 2014).

Our findings have identified a correlation between Hsp90 and eEF1A proteins. A direct correlation between these proteins has yet to be reported in CRC. However, based on our literature search, a group of researchers have reported on the relationship between activation-induced deaminase (AID) (an enzyme that caused mutation on DNA through somatic hypermutation (SHM) and class switch recombination (CSR)) with Hsp90 and eEF1A (Methot et al. 2015). Deficiency of AID was reported to cause a person to suffer from immune deficiency (Revy et al. 2000), meanwhile immoderate of AID action can trigger chromosomal translocation (Häsler et al. 2011), autoimmunity (Diaz 2013) or B cell lymphomas (Methot et al. 2015; Robbiani et al. 2009). Most of AID are localized in cytoplasm (Methot et al. 2015). However, its activity is more active in nucleus which, caused it to be transported in and out of nucleus (Häsler et al. 2011; Methot et al. 2015). Disruption of AID interaction with eEF1A resulted from mutation in AID residues, could lead to AID accumulation in nucleus and increased CSR which subsequently promotes chromosomal translocation (Häsler et al. 2011; Methot et al. 2015). Based on this information, we hypothesize that the relationship between Hsp90 and eEF1A may induce mutagenic processes in cells and could predisposed to genesis of cancer.

5.9 SIGNIFICANT HUMAN PROTEINS BASED ON EPIDEMIOLOGICAL AND CLINICAL FACTORS

Our results indicate, most CRC exclusive proteins were significant in female although our CRC samples were mostly from male patients. The top most significant proteins in female were Epidermal growth factor-like protein 7 and Serum albumin. Epidermal growth factor-like protein 7 (EGFL7) is a secreted proteins expressed by endothelial cells (Delfortrie et al. 2011). It was commonly expressed in late tumours and associated with poor prognosis in CRC (Delfortrie et al. 2011; Hansen et al. 2014). EGFL7 protein is responsible to support angiogenesis mechanism and promotes tumour growth by

evading immunity control, via leukocyte adhesion expression suppression produced by endothelial cell (Hansen et al. 2014; Pinte et al. 2016). Serum albumin is the most abundant protein in blood plasma (Fujii et al. 2012; Knekt et al. 2000). High level of serum albumin was detected among patients with distal colon cancer and it was claimed as being able to predict CRC recurrence (Fujii et al. 2012; Knekt et al. 2000). The level of serum albumin was associated with poor CRC surgical outcome as well (Fujii et al. 2012; Knekt et al. 2000). Individuals suffered from preoperative hypoalbuminemia (serum albumin level less than 3.5 g/L) were predicted to have poor prognosis following curative operation (Gupta and Lis 2010; Lai et al. 2011). Hypoalbuminemia was suspected to be caused by malnutrition (which was a common problem among cancer patients) and comorbidities (Lai et al. 2011).

We have also identified four CRC exclusive proteins significantly expressed in early and late CRC stages. Two of the proteins which are the Unconventional myosin-IA (MYO1A) and Haptoglobin (HP) were frequently reported in the literature. MYO1A protein is part of structural protein in cytoskeleton underpinning the brush border of intestinal epithelial tissue (Mazzolini et al. 2012). The absent or reduced level of the protein will lead to cell polarity and differentiation attritions (Mazzolini et al. 2012). Thus, down regulation of MYO1A was associated with increased tumour growth (Mazzolini et al. 2012). In our findings, the protein was significantly expressed only in Stage I CRC. This result was contradicting with the report from Mazzolini et al. However, we perceived the breakdown of tissue architecture due to tumour progression is a gradual process. Hence, there were still remnants of MYOIA proteins in the collected secretome, originated from the early stage of CRC. HP protein was significantly found in our Stage IV samples and associated with poor prognosis of CRC (Sun et al. 2016).

Most of the proteins identified from healthy colon had biological functions mapped to protein ubiquitination and protein autophosphorylation (refer Figure 4.8). Protein ubiquitination is a post-translational modification that involved transferring of ubiquitin to lysine residues (Hershko and Ciechanover 1998; G. Xu and Jaffrey 2011). Malfunction of protein ubiquitination will cause malignancies and other diseases (Ahmed et al. 2011; Cole and Timiras 1987; Giasson and Lee 2003; Herrmann et al.

2004; Hoeller and Dikic 2009). Protein phosphorylation is another vital mechanism in cell regulation and processes (Ardito et al. 2017). The guardian of the genome, p53 proteins were among the proteins activated by phosphorylation (Ardito et al. 2017). Here, we summarised that most secretome proteins identified in healthy controls were aiding in maintenance of established colon.

5.10 THE IDENTIFIED MICROBIAL ORIGIN SECRETOME

The human gut secretome does not constitute only the human proteins alone. Over 10^{11} microbial cells thriving in every millimeter of the large intestine (Davenport et al. 2017; Walter and Ley 2011), thus it is evident to see fractions of microbial proteins in our samples as well. Overall, we have discovered over 2000 microbial proteins with more than half of the proteins were secreted by the controls. The small number of microbial proteins identified in our CRC samples is justified by a study that reported significant reduction of microbial communities in CRC stool as compared to healthy controls (Ahn et al. 2013). Hence, contribute to low number of microbial proteins identified in our findings (refer subtopic 4.5.2).

Apart from that, we studied the overall microbial structure for both sample groups and observed four (4) main microbial communities in our CRC samples, such as bacteria, fungi, archaea and eukaryote. The control group also shared similar communities but with an addition of virus domain. According to literature, virus or bacteriophage were abundant and diverse in healthy gut (Breitbart et al. 2003; Łusiak-Szelachowska et al. 2017). Their presence is important in the regulation and maintenance of bacterial community through lysis of microbial host, that enable bacteriophage to exert its antibiotic resistance feature and exotoxins production (Breitbart et al. 2003; Davis and Waldor 2002; Łusiak-Szelachowska et al. 2017). In reference to the percentage of composition for each community, we noticed an almost 13% increase of fungi and a slight reduction of bacteria and archaea communities in CRC as opposed to the control samples.

Fungal colonization tend to increase in an immuno-compromised hosts affecting various organs including colon (Praneenararat 2014). Suffering from malignancy is one

of many risk factors for fungi infections (Praneenararat 2014). Archaea are prokaryotes that are known as extreme bacteria that thrive in extreme environments with unique metabolisms feature (Bond Jr et al. 1971; Borrel et al. 2011; Gaci et al. 2014; Hackstein and van Alen 1996; López et al. 2014). At the moment, a direct link between archaea and CRC is still vague, but in other studies archaea were associated with increase calorie intake in animal model and pose an impact on other microbiota metabolism (Praneenararat 2014; Samuel and Gordon 2006).

5.11 TOP MICROBIAL PROTEINS IDENTIFIED FROM STOOL MATERIALS

We also observed and compared expressions between identified CRC exclusive microbial proteins with controls. Our findings showed, all top 50 expressions were found in control samples only and these top proteins were secreted by the yeast and bacterial species such as *Saccharomyces cerevisiae*, *Candida glabrata*, *Escherichia coli* and *Escherichia coli*.

The *Saccharomyces cerevisiae*, also known as baker's yeast often used in food and beverages as a fermentation agent (Argueso et al. 2009). In control samples, Ribosome quality control complex subunit 1 (RQC1) protein and Zinc finger protein ZPR1 were among the top proteins secreted by *S. cerevisiae.* RQC1 is responsible in maintaining cellular protein homeostasis (Consortium 2018; Defenouillère et al. 2016). Meanwhile, ZPR1 protein is important in regulation of normal nucleolar function for proliferating cells (Galcheva-Gargova et al. 1998). Absent of this protein will affect cell's nucleolar activity (Galcheva-Gargova et al. 1998).

Candida glabrata is a non-pathogenic fungi colonizing the colon mucosal tissues in healthy individuals (Fidel Jr et al. 1999). Based on the heatmap, Serine/threonine-protein kinase TEL1 and ADP-ribose 1"-phosphate phosphatase proteins were among the top proteins released by *C. glabrata.* The Serine/threonine protein kinase is involved in regulation of DNA damage response mechanism and maintaining chromatin organization (Consortium 2018). Meanwhile, *Escherichia coli* is a pathogen that have caused multiple gut illnesses, such as colitis and diarrhea (Ratnam et al. 1988). It releases Magnesium-transporting ATPase, P-type 1 protein that

has molecular function related to ATP binding, magnesium-importing ATPase activity and metal ion binding (Consortium 2018). We also identified Protein YehQ released by *Escherichia coli* that plays important role in zinc ion binding. E. *coli* is a non-pathogenic bacteria, widely used in bench study for genetic cloning and as a probiotic treatment for gut inflammation such as ulcerative colitis (Kruis et al. 2004), Crohn's disease (Fujimori et al. 2007) and acute diarrhea (Bereswill et al. 2013; Henker et al. 2007).

5.12 MICROBIAL PROTEINS RELATED TO CRC STAGING

Next, we performed another heatmap analysis (Figure 4.20) to compare expression of CRC exclusive microbial proteins based on CRC staging. Fourteen proteins topping the heatmap were identified in the late stages of CRC. Some of microbial species that released these proteins were *Candida albicans, Shewanella amazonensis, Saccharomyces cerevisiae* and *Candida glabrata*. Screening through the microbial species, we noticed a few similar fungi species identified previously in controls were also present in CRC, but different proteins were identified. Additionally, we also noticed most of the top expressed proteins in early and late groups were unique to samples.

Protein MLP1 released by *S. cerevisiae* was previously reported in late stage CRC and this protein was related to biological function of mRNA export and DNA repair (Consortium 2018; Fasken et al. 2008; Galy et al. 2004). The D-arabinono-1,4-lactone oxidase protein secreted by *C. glabrata* was involved in dehydro-D-arabinono-1,4-lactone biosynthetic process (Consortium 2018) whereas the ATP-dependent RNA helicase DBP10 protein released by *C. albicans* was responsible in rRNA processing (Consortium 2018).

S. cerevisiae has been widely linked to cancer study. This yeast has a conserved amino acid sequence and protein functions similar to human, hence was commonly used as a model to study mutations in hereditary nonpolyposis colorectal carcinoma and CRC (Cazzanelli et al. 2018; Jeyaprakash et al. 1996; Pang et al. 1997). An animal model study had shown that selenium-enriched *S. cerevisae* is an effective food supplement to reduce cancerous and pre-cancerous lesion in CRC (Abedi et al. 2018). Meanwhile,

combination of *S. cerevisiae* with paclitaxel (a chemotherapy medication) has helped to promote anticancer effect in breast cancer by inducing cancerous cell apoptotic activity in vitro (Badr El-Din et al. 2017). Apart from that, *S. cerevisiae* is also known as yeast probiotics (Saber et al. 2017). The ß-glucans in *S. cerevisiae*, was indicated to improve immunity system (Javmen et al. 2015), inhibit spreading of tumour (Yoon et al. 2008), enhance wound repair (Majtan and Jesenak 2018) and lowered blood cholesterol (Kusmiati and Dhewantara 2016; Saber et al. 2017). To date, publications that linked the potential of *S. cerevisiae* in promoting CRC carcinogenesis are still scarce. Yet, we found plenty of proteins from this species in our case samples.

On the other hand, *Candida sp.* is a commensal pathogen that can trigger serious infection in an immunocompromised individual (Mohd Bakri et al. 2010). It is known as opportunistic pathogen (Chung et al. 2017) and able to adjust to host's ecology (Mohd Bakri et al. 2010). Factors promoting *Candida* infection include dysbiosis of gut microbiota, weaken immunity system, protracted use of antibiotics, drug abuse, lack of digestive secretions, imbalanced diets and liver disease (Chung et al. 2017; Martins et al. 2014). *Candida* infection promoted inflammation and increased overall risk of acquiring CRC (Chung et al. 2017; Kumamoto 2011). The colonization of *Candida sp.* was reported to initiate formation of biofilm on the mucosal surface of the colon in which comprised of multiple cell types (Ganguly and Mitchell 2011) that are resistance to human immune system (Nobile and Johnson 2015).

5.13 CRC EXCLUSIVE-MICROBIAL PROTEIN PATHWAY ANALYSIS

Analysis on the predicted pathways related to our CRC exclusive microbial proteins were carried out as well but we do not obtain any predicted pathways related to these proteins. However, we discovered three (3) protein interactions involving proteins from the yeast species (*Schizosaccharomyces pombe* and *S. cerevisiae).* The interaction between DNA repair protein rhp57 (rhp57) and structural maintenance of chromosomes protein 5 (smc5) in *S. pombe* has not been reported yet, but both proteins are related to DNA repair. The rhp57 protein is involved in mending recombinant DNA and fixing the gamma-induced damage (Consortium 2018) whereas the smc5 is one of the six classes in structural maintenance of chromosome (SMC) proteins, crucial in DNA repair

as well (Pebernard et al. 2006; Sergeant et al. 2005). The second protein interaction found in *S. pombe,* was between SWR1 complex bromodomain subunit (bdf1) protein with actin (Act1) protein. Bdf forms a complex with SWR1 to aid in integration of H2A.Z (a histone variant), mediating transcriptional regulation of genes through chromatin modeling (Consortium 2018; García-Oliver et al. 2017; Wu et al. 2009) whereas Act1 protein is an integral part of cell motile process (Consortium 2018; Takaine and Mabuchi 2007). Although there is no direct functional linkage between these two proteins, we postulate that both proteins do play an interlinking role respectively and this remains to be elucidated. *S. pombe* was commonly used as a model to study actin cytoskeleton action (Takaine and Mabuchi 2007).

The next interaction observed was between three proteins from *S. cerevisiae,* namely cytochrome B2 (CYB2), actin (Act1) and chromatin-remodeling complex subunit IES6 (IES6). CYB2 has biological process related to lactate metabolism (Consortium 2018) and IES6 is associated with transcriptional regulation with INO80 (a chromatin modelling complex crucial in DNA processes). Apart from that, this protein also participated in mediating telomere length (Consortium 2018; Morrison et al. 2004; Smith et al. 2004; Yu et al. 2007). Again, we do not see any direct functional linkages but all three proteins are crucial for the organism's survival.

Based on the acquired information, most proteins that participate in the interactions had local functions associated to microbial species, but its effects towards the host and cancer genesis are still unknown. This gap of knowledge opens up the opportunity for future research on to the host-microbe interactions.

5.14 MICROBIAL PROTEIN-PROTEIN CORRELATION

Twenty-eight significant protein-protein correlations were identified among the CRC exclusive microbial proteins. Based on our literature search, we could not establish any direct correlations between these proteins with CRC with regards to their molecular and biological functions within the microbial species. To date, we have accumulated abundance of information on the microbial species that were implicated in CRC, mostly identified via their genomic make-up. However, data on their proteins are still lacking,

let alone studies that report specifically on various microbial secretome proteins, as well as its roles in CRC microenvironment. In the present study, no postulation can be made yet but interestingly, we do observe some interspecies protein-protein correlations that may require further analysis to be done.

5.15 SIGNIFICANT MICROBIAL PROTEINS BASED ON EPIDEMIOLOGICAL AND CLINICAL FACTORS

Based on gender, the most significant CRC exclusive microbial protein was found in female samples. The protein is identified as Phosphoribosylformylglycinamidine cyclo-ligase protein, secreted by *Clostridium cellulolyticum* and associated with 'de novo' IMP biosynthetic process.

Aside from that, we have identified ten significant CRC exclusive microbial proteins based on tumour staging and the proteins were significantly expressed in Stage IV of the disease. Proteins from *Helicobacter pylori* and *Helicobacter hepaticus* were among the microbial molecules identified. Both bacteria species were reported to cause carcinogenic effect to the host, particularly H. *pylori* which was classified as class I human carcinogen (Strofilas et al. 2012; WHO et al. 1994). In our current work, the significant presence of *H. pylori* proteins in the late stage of CRC supports this finding.

CHAPTER VI

CONCLUSION

6.1 RESEARCH SUMMARY

We have identified more human origin proteins in CRC as compared to control and inversely for proteins from microbial origin. The best prediction model was built with combination of human Huntingtin and RNA exonuclease 5 proteins. The model was sensitive, but not specific in discriminating control from CRC. Our findings have shown that human exclusive proteins for CRC were mostly mapped to cell adhesion, intracellular signal transduction and angiogenesis functions. Meanwhile, the top annotated KEGG pathway for human CRC-exclusive proteins was Hypoxia-inducible factor-1 (HIF-1). Yeast proteins were topping the microbial CRC-exclusive protein list, with predicted protein interactions mapped to DNA repair, transcription regulation and ATP binding.

6.2 STUDY LIMITATION

The limitation of this study includes the small sample size. Increasing the size of sample cohort may enable us to have an overview of secretome proteomic landscape of CRC individuals in Malaysia and to verify the present findings. Also, only 3% of our CRC sample was from the Indian ethnic. In our opinion, an ethnic-balanced sampling is a better study design if we were to investigate the ethnicity factor.

Apart from that, the utilization of stool samples requires filtration step during sample processing. This step may result in sample loss thus underestimation. Current findings require in-vivo and in-vitro downstream analysis to understand further the carcinogenic or protective potentials of the candidate human and microbial proteins.

Moreover, there are chances of intracellular protein contamination in the secretome samples. This contamination could occur in two ways; autolysis of cell or bacteria and the transportation of proteins to the outer environment via non-classical secretion system. Both situations are beyond our control, since most of the processes occur in the host's guts and all we are 'collecting' are the products and sieving through these 'contaminations' is a challenge. Information on proteins localization can be obtained from online protein database search.

6.3 FUTURE WORKS

The relationship between human gut microbiota and pathogenesis of CRC involved vast numbers of role players. Focusing on a single area of study may not accurately portray the actual landscape of the disease hence, incorporation of findings from various area of studies are necessary. For example, the integration of high-resolution meta-proteomics data with meta-genomics will provide a comprehensive view on biological network in the gut tumour microenvironment. The development of Integrative Personal Omics Profile (iPOP), will merge findings from candidate microbiome with multiple omics technologies; genomic, transcriptomic, proteomic, metabolomic and autoantibody profiles, as an initiative towards precision medicine in the coming years. Correlating patient's dietary information with meta-genomic, meta-proteomic and metabolite analysis may provide better understanding on diet influence, the abundance of microbial species discovered, the identified host and microbial secretome and the level of short chain fatty acids. The sequence of information may unveil the complexity of human gut flora and CRC pathogenesis. Apart from that, the identified protein signatures from microbes and host can be utilized as novel biomarkers and lead to establishment of a new stool-based non-invasive CRC screening panel.

REFERENCE

Aarons, C. B., Shanmugan, S. & Bleier, J. I. S. 2014. Management of malignant colon polyps: Current status and controversies. *World Journal of Gastroenterology : WJG* 20(43): 16178–16183.

Abedi, J., Saatloo, M. V., Nejati, V., Hobbenaghi, R., Tukmechi, A., Nami, Y. & Khosroushahi, A. Y. 2018. Selenium-enriched Saccharomyces cerevisiae reduces the progression of colorectal cancer. *Biological Trace Element Research* 1–9.

Abreu, M. T. & Peek, R. M. 2014. Gastrointestinal malignancy and the microbiome. *Gastroenterology* 146(6): 1534–1546.e3.

Abu Hassan, M. R., Ismail, I., Mohd Suan, M. A., Ahmad, F., Wan Khazim, W. K., Othman, Z., Mat Said, R., Tan, W. L., Mohammed, S. R., Soelar, S. A. & Nik Mustapha, N. R. 2016. Incidence and mortality rates of colorectal cancer in Malaysia. *Epidemiology Health* 38(0): e2016007-0.

Adelstein, B.-A., Macaskill, P., Chan, S. F., Katelaris, P. H. & Irwig, L. 2011. Most bowel cancer symptoms do not indicate colorectal cancer and polyps: a systematic review. *BMC Gastroenterology* 11(1): 65.

Adjei, A. A. & Hidalgo, M. 2005. Intracellular Signal Transduction Pathway Proteins As Targets for Cancer Therapy. *Journal of Clinical Oncology* 23(23): 5386–5403.

Ahmed, N., Zeng, M., Sinha, I., Polin, L., Wei, W.-Z., Rathinam, C., Flavell, R., Massoumi, R. & Venuprasad, K. 2011. The E3 ligase Itch and deubiquitinase Cyld act together to regulate Tak1 and inflammation. *Nature Immunology* 12(12): 1176–1183.

Ahn, J., Sinha, R., Pei, Z., Dominianni, C., Wu, J., Shi, J., Goedert, J. J., Hayes, R. B. & Yang, L. 2013. Human Gut Microbiome and Risk of Colorectal Cancer. *Journal of The National Cancer Institute* .

Allison, J. E., Fraser, C. G., Halloran, S. P. & Young, G. P. 2014. Population Screening for Colorectal Cancer Means Getting FIT: The Past, Present, and Future of Colorectal Cancer Screening Using the Fecal Immunochemical Test for Hemoglobin (FIT). *Gut and Liver* 8(2): 117–130.

American Cancer Society. 2017a. *Cancer Facts & Figures 2017*. Atlanta. https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html [2 January 2018].

American Cancer Society. 2017b. *Colorectal Cancer Facts & Figures 2017-2019*. Atlanta. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf [2 January 2018].

Arai, Y., Matsuura, T., Matsuura, M., Fujiwara, M., Okayasu, I., Ito, S. & Arihiro, S. 2016. Prostaglandin E-major urinary metabolite as a biomarker for inflammation in ulcerative colitis: prostaglandins revisited. *Digestion* 93(1): 32–39.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. 2017. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine* 40(2): 271–280.

Argueso, J. L., Carazzolle, M. F., Mieczkowski, P. A., Duarte, F. M., Netto, O. V. C., Missawa, S. K., Galzerani, F., Costa, G. G. L., Vidal, R. O., Noronha, M. F., Dominska, M., Andrietta, M. G. S., Andrietta, S. R., Cunha, A. F., Gomes, L. H., Tavares, F. C. A., Alcarde, A. R., Dietrich, F. S., McCusker, J. H., Petes, T. D. & Pereira, G. A. G. 2009. Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research* 19(12): 2258–2270.

Armaghany, T., Wilson, J. D., Chu, Q. & Mills, G. 2012. Genetic Alterations in Colorectal Cancer. *Gastrointestinal Cancer Research : GCR* 5(1): 19–27.

Ashton, N. 2010. Physiology of red and white blood cells. *Anaesthesia & Intensive Care Medicine* 11(6): 236–241.

Astin, M., Griffin, T., Neal, R. D., Rose, P. & Hamilton, W. 2011. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. *The British Journal of General Practice* 61(586): e231–e243.

Azmi, M. N., Zailani, M. A. & Norashikin, M. N. 2007. Five-year review of histopathological findings of colorectal cancer patients operated in Hospital Tengku Ampuan Afzan Kuantan Pahang Malaysia. *International Medical Journal Malaysia* 6: 15–20.

Badr El-Din, N. K., Mahmoud, A. Z., Hassan, T. A. & Ghoneum, M. 2017. Baker's yeast sensitizes metastatic breast cancer cells to paclitaxel in vitro. *Integrative Cancer Therapies* 17(2): 542–550.

Barderas, R., Mendes, M., Torres, S., Bartolomé, R. A., López-Lucendo, M., Villar-Vázquez, R., Peláez-García, A., Fuente, E., Bonilla, F. & Casal, J. I. 2013. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration and invasion. *Molecular & Cellular Proteomics : MCP* 12(6): 1602–1620.

Barry, S., Chelala, C., Lines, K., Sunamura, M., Wang, A., Marelli-Berg, F. M., Brennan, C., Lemoine, N. R. & Crnogorac-Jurcevic, T. 2013. S100P is a metastasis-associated gene that facilitates transendothelial migration of pancreatic cancer cells. *Clinical and Experimental Metastasis* 30(3): 251–264.

Bereswill, S., Fischer, A., Dunay, I. R., Kühl, A. A., Göbel, U. B., Liesenfeld, O. & Heimesaat, M. M. 2013. Pro-inflammatory potential of *Escherichia coli* strains K12 and Nissle 1917 in a murine model of acute ileitis. *European Journal of Microbiology & Immunology* 3(2): 126–134.

Bibbins-Domingo, K., Grossman, D. C., Curry, S. J., Davidson, K. W., Epling, J. W., García, F. A. R., Gillman, M. W., Harper, D. M., Kemper, A. R. & Krist, A. H. 2016. Screening for colorectal cancer: us preventive services task force recommendation statement. *Jama* 315(23): 2564–2575.

Blanco, M. A., LeRoy, G., Khan, Z., Aleckovic, M., Zee, B. M., Garcia, B. A. & Kang, Y. 2012. Global secretome analysis identifies novel mediators of bone metastasis. *Cell. Res.* 22(9): 1339–1355.

Boleij, A. & Tjalsma, H. 2012. Gut bacteria in health and disease: A survey on the interface between intestinal microbiology and colorectal cancer. *Biological Reviews* 87(3): 701–730.

Boleij, A. & Tjalsma, H. 2013. The itinerary of Streptococcus gallolyticus infection in patients with colonic malignant disease. *The Lancet Infectious Diseases* 13(8): 719–724.

Boleij, A., van Gelder, M. M. H. J., Swinkels, D. W. & Tjalsma, H. 2011. Clinical Importance of Streptococcus gallolyticus infection among colorectal cancer patients: systematic review and meta-analysis. *Clinical Infectious Diseases* 53(9): 870–878.

Bond Jr, J. H., Engel, R. R. & Levitt, M. D. 1971. Factors influencing pulmonary methane excretion in man. An indirect method of studying the in situ metabolism of the methane-producing colonic bacteria. *The Journal of Experimental Medicine* 133(3): 572–588.

Borrel, G., Jézéquel, D., Biderre-Petit, C., Morel-Desrosiers, N., Morel, J.-P., Peyret, P., Fonty, G. & Lehours, A.-C. 2011. Production and consumption of methane in freshwater lake ecosystems. *Research in Microbiology* 162(9): 832–847.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P. & Rohwer, F. 2003. Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology* 185(20): 6220–6223.

Brune, A. & Friedrich, M. 2000. Microecology of the termite gut: structure and function on a microscale. *Current Opinion in Microbiology* 3(3): 263–269.

Bujanda, L., Cosme, A., Gil, I. & Arenas-Mirave, J. I. 2010. Malignant colorectal polyps. *World Journal of Gastroenterology : WJG* 16(25): 3103–3111.

Burakoff, R. & Percy, W. H. 1992. Studies in vivo and in vitro on effects of PGE2 on colonic motility in rabbits. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 262(1): G23–G29.

Burchfield, J. S. & Dimmeler, S. 2008. Role of paracrine factors in stem and progenitor cell mediated cardiac repair and tissue fibrosis. *Fibrogenesis & Tissue Repair* 1(1): 4.

Cameron, S. J. S. & Takáts, Z. 2018. Mass spectrometry approaches to metabolic profiling of microbial communities within the human gastrointestinal tract. *Methods*.

Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. 2015. Dysbiosis of the gut microbiota in disease. *Microbial Ecology in Health and Disease;* (26).

Carroll, M. R. R., Seaman, H. E. & Halloran, S. P. 2014. Tests and investigations for colorectal cancer screening. *Clinical Biochemistry* 47(10): 921–939.

Cazzanelli, G., Pereira, F., Alves, S., Francisco, R., Azevedo, L., Dias Carvalho, P., Almeida, A., Côrte-Real, M., Oliveira, M. & Lucas, C. 2018. The yeast *Saccharomyces cerevisiae* as a model for understanding RAS proteins and their role in human tumorigenesis. *Cells* 7(2): 14.

Celis, J. E., Gromov, P., Cabezón, T., Moreira, J. M. A., Ambartsumian, N., Sandelin, K., Rank, F. & Gromova, I. 2004. Proteomic Characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery . *Molecular & Cellular Proteomics* 3(4): 327–344.

Chan, C., Fan, C., Kuo, Y., Chen, Y., Chang, P., Chen, K., Hung, R. & Chan, E. 2010. Multiple serological biomarkers for colorectal cancer detection. *International Journal of Cancer* 126(7): 1683–1690.

Chin, S.-F., Megat Mohd Azlan, P. I. H., Mazlan, L. & Neoh, H. 2018. Identification of *Schizosaccharomyces pombe* in the guts of healthy individuals and patients with colorectal cancer: preliminary evidence from a gut microbiome secretome study. *Gut Pathogens* 10(1): 29.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S. & Xia, J. 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*. https://www.metaboanalyst.ca/ [1 August 2018]

Choo, J. M., Leong, L. E. & Rogers, G. B. 2015. Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports* 5: 16350.

Chow, J., Lee, S. M., Shen, Y., Khosravi, A. & Mazmanian, S. K. 2010. Host–bacterial symbiosis in health and disease. *Advances in Immunology* 107: 243–274.

Chung, L.-M., Liang, J.-A., Lin, C.-L., Sun, L.-M. & Kao, C.-H. 2017. Cancer risk in patients with candidiasis: a nationwide population-based cohort study. *Oncotarget* 8(38): 63562–63573.

Cipriani, S., Mencarelli, A., Palladino, G. & Fiorucci, S. 2010. FXR activation reverses insulin resistance and lipid abnormalities and protects against liver steatosis in Zucker (fa/fa) obese rats. *Journal of Lipid Research* 51(4): 771–784.

Cohen, P. E. & Holloway, J. K. 2010. Predicting gene networks in human oocyte meiosis. *Biology of Reproduction* 82(3): 469–472.

Cole, G. M. & Timiras, P. S. 1987. Ubiquitin-protein conjugates in Alzheimer's lesions. *Neuroscience Letters* 79(1): 207–212.

Consortium, U. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 46(5): 2699. https://www.uniprot.org/ [1 August 2018].

Cooper, G. M. & Hausman, R. E. 2000. Pathways of Intracellular Signal Transduction. *The Cell: A Molecular Approach*, 2nd Edition. Sunderland (MA): Sinauer Associates.

Davenport, E. R., Sanders, J. G., Song, S. J., Amato, K. R., Clark, A. G. & Knight, R. 2017. The human microbiome in evolution. *BMC Biology* 15(1): 127.

Davis, B. M. & Waldor, M. K. 2002. Mobile genetic elements and bacterial pathogenesis. *Mobile DNA II*: 1040–1059. American Society of Microbiology.

de Wit, M., Kant, H., Piersma, S. R., Pham, T. V, Mongera, S., van Berkel, M. P. A., Boven, E., Pontén, F., Meijer, G. A., Jimenez, C. R. & Fijneman, R. J. A. 2014. Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. *Journal of Proteomics* 99: 26–39.

Defenouillère, Q., Zhang, E., Namane, A., Mouaikel, J., Jacquier, A. & Fromont-Racine, M. 2016. Rqc1 and Ltn1 prevent C-terminal alanine-threonine tail (CAT-tail)-induced protein aggregation by efficient recruitment of Cdc48 on stalled 60S subunits. *The Journal of Biological Chemistry* 291(23): 12245–12253.

Degen, L. P. & Phillips, S. F. 1996. How well does stool form reflect colonic transit? *Gut* 39(1): 109–113.

Delfortrie, S., Pinte, S., Mattot, V., Samson, C., Villain, G., Caetano, B., Lauridant-Philippin, G., Baranzelli, M.-C., Bonneterre, J., Trottein, F., Faveeuw, C. & Soncin, F. 2011. Egfl7 promotes tumor escape from immunity by repressing endothelial cell activation. *Cancer Research* 71(23): 7176 LP-7186.

Delzenne, N. M., Neyrinck, A. M., Backhed, F. & Cani, P. D. 2011. Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nat. Rev. Endocrinol.* 7(11): 639–646.

DeWeerdt, S. 2015. Microbiome: Microbial mystery. *Nature* 521(7551): S10–S11.

Diaz, M. 2013. The role of activation-induced deaminase in Lupus Nephritis. *Autoimmunity* 46(2): 115–120.

Doyle, B., Sorajja, P., Hynes, B., Kumar, A. H. S., Araoz, P. A., Stalboerger, P. G., Miller, D., Reed, C., Schmeckpeper, J. & Wang, S. 2008. Progenitor cell therapy in a porcine acute myocardial infarction model induces cardiac hypertrophy,

mediated by paracrine secretion of cardiotrophic factors including TGFβ1. *Stem Cells and Development* 17(5): 941–952.

Drake, J. M., Strohbehn, G., Bair, T. B., Moreland, J. G., Henry, M. D. & Hynes, R. O. 2009. ZEB1 Enhances Transendothelial Migration and Represses the Epithelial Phenotype of Prostate Cancer Cells. *Molecular Biology of The Cell* 20(8): 2207–2217.

Drake, J., Strohbehn, G., Varzavand, A., Winterwood, N., Moreland, J., Stipp, C. & Henry, M. 2008. Transendothelial migration selects for prostate cancer cells with a mesenchymal phenotype and loss of laminin 332 expression. *Cancer Research* 68(9 Supplement): 1952 LP-1952.

Dubrow, R., Kim, C. S. & Eldred, A. K. 1992. Fecal lysozyme: An unreliable marker for colorectal cancer. *American Journal of Gastroenterology* 87(5).

Dubrow, R. & Yannielli, L. 1992. Fecal protein markers of colorectal cancer. *American Journal of Gastroenterology* 87(7).

Dulal, S. & Keku, T. O. 2014. Gut Microbiome and Colorectal Adenomas. *The Cancer Journal* 20(3): 225–231.

Dyer, O. 2018. Colorectal cancer: US guidelines urge screening from age 45 as incidence soars in younger adults. *BMJ* 361.

Eales, K. L., Hollinshead, K. E. R. & Tennant, D. A. 2016. Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis* 5(1): e190.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. 2005. Diversity of the Human Intestinal Microbial Flora. *Science* 308(5728): 1635–1638.

Ekelund, S. 2012. ROC curves—what are they and how are they used? *point of care* 11(1): 16–21.

El Zoghbi, M. & Cummings, L. C. 2016. New era of colorectal cancer screening. *World Journal of Gastrointestinal Endoscopy* 8(5): 252–258.

Enarsson, K., Lundin, B. S., Johnsson, E., Brezicka, T. & Quiding-Järbrink, M. 2007. CD4+CD25high regulatory T cells reduce T cell transendothelial migration in cancer patients. *European Journal of Immunology* 37(1): 282–291.

Ervik, M., Lam, F., Ferlay, J., Mery, L., Soerjomataram, I. & Bray, F. 2016. Cancer Today. *Cancer Today*. Lyon, France. http://gco.iarc.fr/today [11 December 2017].

Fasken, M. B., Stewart, M. & Corbett, A. H. 2008. Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. *The Journal of Biological Chemistry* 283(40): 27130–27143.

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I. & F, B. 2017. Global Cancer Observatory: Cancer Today. *International Agency for Research on Cancer*. https://gco.iarc.fr/today [10 January 2017].

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. 2014. *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Lyon, France.

Fidel Jr, P. L., Vazquez, J. A. & Sobel, J. D. 1999. *Candida glabrata*: review of epidemiology, pathogenesis, and clinical disease with comparison to C. *albicans*. *Clinical Microbiology Reviews* 12(1): 80–96.

Frampton, M. & Houlston, R. S. 2016. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. *Genetics In Medicine* 19: 314.

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D. & Gevers, D. 2014. Relating the metatranscriptome and metagenome of the human gut. *Proceedings of The National Academy of Sciences*.

Fujii, T., Sutoh, T., Morita, H., Katoh, T., Yajima, R., Tsutsumi, S., Asao, T. & Kuwano, H. 2012. Serum albumin is superior to prealbumin for predicting short-term recurrence in patients with operable colorectal cancer. *Nutrition and Cancer* 64(8): 1169–1173.

Fujimori, S., Tatsuguchi, A., Gudis, K., Kishida, T., Mitsui, K., Ehara, A., Kobayashi, T., Sekita, Y., Seo, T. & Sakamoto, C. 2007. High dose probiotic and prebiotic cotherapy for remission induction of active Crohn's disease. *Journal of Gastroenterology and Hepatology* 22(8): 1199–1204.

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W. & Brugère, J.-F. 2014. Archaea and the human gut: new beginning of an old story. *World Journal of Gastroenterology* 20(43): 16062–16078.

Gagnière, J., Raisch, J., Veziant, J., Barnich, N., Bonnet, R., Buc, E., Bringer, M.-A., Pezet, D. & Bonnet, M. 2016. Gut microbiota imbalance and colorectal cancer. *World Journal of Gastroenterology* 22(2): 501–518.

Galcheva-Gargova, Z., Gangwani, L., Konstantinov, K. N., Mikrut, M., Theroux, S. J., Enoch, T. & Davis, R. J. 1998. The cytoplasmic zinc finger protein ZPR1 accumulates in the nucleolus of proliferating cells. *Molecular Biology of The Cell* 9(10): 2963–2971.

Galy, V., Gadal, O., Fromont-Racine, M., Romano, A., Jacquier, A. & Nehrbass, U. 2004. Nuclear retention of unspliced mrnas in yeast is mediated by perinuclear

Mlp1. *Cell* 116(1): 63–73.

Ganguly, S. & Mitchell, A. P. 2011. Mucosal biofilms of *Candida albicans*. *Current Opinion in Microbiology* 14(4): 380–385.

Gao, Z., Guo, B., Gao, R., Zhu, Q. & Qin, H. 2015. Microbiota dysbiosis is associated with colorectal cancer. *Frontiers in Microbiology* 6: 20.

García-Oliver, E., Ramus, C., Perot, J., Arlotto, M., Champleboux, M., Mietton, F., Battail, C., Boland, A., Deleuze, J.-F., Ferro, M., Couté, Y. & Govin, J. 2017. Bdf1 bromodomains are essential for meiosis and the expression of meiotic-specific genes. *PLOS Genetics* 13(1): e1006541.

Garrett, W. S. 2015. Cancer and The Microbiota. *Science* 348(6230): 80–86.

Ghee, L. K. 2014. A review of colorectal cancer research in malaysia. *Medical Journal Malaysia* 69.

Giasson, B. I. & Lee, V. M.-Y. 2003. Are ubiquitination pathways central to parkinson's disease? *Cell* 114(1): 1–8.

Gilbert, J. A., Ahlquist, D. A., Mahoney, D. W., Zinsmeister, A. R., Rubin, J. & Ellefson, R. D. 1996. Fecal marker variability in colorectal cancer: calprotectin versus hemoglobin. *Scandinavian Journal of Gastroenterology* 31(10): 1001–1005.

Ginzberg, M. B., Kafri, R. & Kirschner, M. 2015. On being the right (cell) size. *Science* 348(6236): 1245075.

Goodwin, A. C., Shields, C. E. D., Wu, S., Huso, D. L., Wu, X., Murray-Stewart, T. R., Hacker-Prietz, A., Rabizadeh, S., Woster, P. M., Sears, C. L. & Casero, R. A. 2011. Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. *Proceedings of The National Academy of Sciences of The United States of America* 108(37): 15354–15359.

Guerra, L., Guidi, R. & Frisan, T. 2011. Do bacterial genotoxins contribute to chronic inflammation, genomic instability and tumor progression? *The FEBS Journal* 278(23): 4577–4588.

Guhaniyogi, J. & Brewer, G. 2001. Regulation of mRNA stability in mammalian cells. *Gene* 265(1): 11–23.

Guo, C., Kosarek-Stancel, J. N., Tang, T.-S. & Friedberg, E. C. 2009. Y-family DNA polymerases in mammalian cells. *Cellular and Molecular Life Sciences* 66(14): 2363.

Guo, Y. & Li, H.-Y. 2014. Association between Helicobacter pylori infection and colorectal neoplasm risk: a meta-analysis based on east asian population. *Journal*

of Cancer Research and Therapeutics 10(8): 263.

Gupta, D. & Lis, C. G. 2010. Pretreatment serum albumin as a predictor of cancer survival: a systematic review of the epidemiological literature. *Nutrition Journal* 9(1): 69.

Hackstein, J. H. P. & van Alen, T. A. 1996. Fecal methanogens and vertebrate evolution. *Evolution* 50(2): 559–572.

Hamilton, W. & Sharp, D. 2004. Diagnosis of colorectal cancer in primary care: the evidence base for guidelines. *Family Practice* 21(1): 99–106.

Hanahan, D. & Weinberg, R. A. 2011. Hallmarks of cancer: the next generation. *Cell* 144(5): 646–74.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M. & Dougherty, E. R. 2010. Small-sample precision of ROC-related estimates. *Bioinformatics* 26(6): 822–830.

Hanrahan, A. J., Iyer, G. & Solit, D. B. 2014. Intracellular Signaling. Niederhuber., Armitage., Doroshow., Kastan. & Tepper. *Abeloff's Clinical Oncology*, 5th Edition ., 22–39.e8. Philadelphia.

Hansen, T. F., Nielsen, B. S., Sørensen, F. B., Johnsson, A. & Jakobsen, A. 2014. Epidermal growth factor–like domain 7 predicts response to first-line chemotherapy and bevacizumab in patients with metastatic colorectal cancer. *Molecular Cancer Therapeutics* 13(9): 2238 LP-2245.

Hardcastle, J. D., Chamberlain, J. O., Robinson, M. H. E., Moss, S. M., Amar, S. S., Balfour, T. W., James, P. D. & Mangham, C. M. 1996. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *The Lancet* 348(9040): 1472–1477.

Häsler, J., Rada, C. & Neuberger, M. S. 2011. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1α (eEF1A). *Proceedings of The National Academy of Sciences of The United States of America* 108(45): 18366–18371.

Heaton, K. W., Radvan, J., Cripps, H., Mountford, R. A., Braddon, F. E. & Hughes, A. O. 1992. Defecation frequency and timing, and stool form in the general population: a prospective study. *Gut* 33(6): 818–824.

Heidemann, J., Domschke, W., Kucharzik, T. & Maaser, C. 2006. Intestinal microvascular endothelium and innate immunity in inflammatory bowel disease: a second line of defense? *Infection and Immunity* 74(10): 5425 LP-5432.

Henker, J., Laass, M., Blokhin, B. M., Bolbot, Y. K., Maydannik, V. G., Elze, M., Wolff, C. & Schulze, J. 2007. The probiotic *Escherichia coli* strain Nissle 1917 (EcN) stops acute diarrhoea in infants and toddlers. *European Journal of Pediatrics* 166(4): 311–318.

Herrmann, J., Ciechanover, A., Lerman, L. O. & Lerman, A. 2004. The ubiquitin–proteasome system in cardiovascular diseases—a hypothesis extended. *Cardiovascular Research* 61(1): 11–21.

Hershko, A. & Ciechanover, A. 1998. The Ubiquitin System. *Annual Review of Biochemistry* 67(1): 425–479.

Hoeller, D. & Dikic, I. 2009. Targeting the ubiquitin system in cancer therapy. *Nature* 458: 438.

Holme, Ø., Bretthauer, M., Fretheim, A., Odgaard-Jensen, J. & Hoff, G. 2013. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. *Cochrane Database of Systematic Reviews* (9).

Hostein, I., Robertson, D., DiStefano, F., Workman, P. & Clarke, P. A. 2001. Inhibition of signal transduction by the Hsp90 inhibitor 17-allylamino-17-demethoxygeldanamycin results in cytostasis and apoptosis. *Cancer Research* 61(10): 4003–4009.

Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Levin, T. R., Lavin, P., Lidgard, G. P., Ahlquist, D. A. & Berger, B. M. 2014. Multitarget stool dna testing for colorectal-cancer screening. *New England Journal of Medicine* 371(2): 184–188.

Inoue, M., Tajima, K., Hirose, K., Hamajima, N., Takezaki, T., Hirai, T., Kato, T. & Ohno, Y. 1995. Subsite-specific risk factors for colorectal cancer: a hospital-based case-control study in Japan. *Cancer Causes & Control* 6(1): 14–23.

Ivanov, I. I. & Honda, K. 2012. Intestinal commensal microbes as immune modulators. *Cell Host & Microbe* 12(4): 496–508.

Iyengar, V., Albaugh, G. P., Lohani, A. & Nair, P. P. 1991. Human stools as a source of viable colonic epithelial cells. *The FASEB Journal* 5(13): 2856–2859.

Jackson, K. A., Majka, S. M., Wang, H., Pocius, J., Hartley, C. J., Majesky, M. W., Entman, M. L., Michael, L. H., Hirschi, K. K. & Goodell, M. A. 2001. Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *Journal of Clinical Investigation* 107(11): 1395.

Janakiram, N. B. & Rao, C. V. 2014. The role of inflammation in colon cancer. *Advances in Experimental Medicine and Biology*. Springer.

Javmen, A., Nemeikaitė-Čėnienė, A., Bratchikov, M., Grigiškis, S., Grigas, F., Jonauskienė, I., Zabulytė, D. & Mauricas, M. 2015. β-Glucan from *Saccharomyces cerevisiae* induces ifn-γ production in vivo in BALB/c mice. *In Vivo* 29(3): 359–363.

Jeyaprakash, A., Das Gupta, R. & Kolodner, R. 1996. *Saccharomyces cerevisiae* pms2

mutations are alleles of MLH1, and pms2-2 corresponds to a hereditary nonpolyposis colorectal carcinoma-causing missense mutation. *Molecular and Cellular Biology* 16(6): 3008–3011.

Ji, B. & Nielsen, J. 2015. From next-generation sequencing to systematic modeling of the gut microbiome. *Frontiers in Genetics* 6: 219.

Jin, P., Wang, K., Huang, C. & Nice, E. C. 2017. Mining the fecal proteome: from biomarkers to personalised medicine. *Expert Review of Proteomics* 14(5): 445–459.

Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B. & Berry, D. A. 2013. Meta-analyses of colorectal cancer risk factors. *Cancer Causes & Control* 24(6): 1207–1222.

Kamada, N., Seo, S.-U., Chen, G. Y. & Núñez, G. 2013. Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews Immunology* 13(5): 321–335.

Kanthan, R., Senger, J.-L. & Kanthan, S. C. 2012. Fecal molecular markers for colorectal cancer screening. *Gastroenterology Research and Practice* 2012: 184343.

Karagiannis, G. S., Pavlou, M. P. & Diamandis, E. P. 2010. Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. *Molecular Oncology* 4(6): 496–510.

Kim, D., Hofstaedter, C. E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C. & Kelsen, J. 2017. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* 5(1): 52.

Kim, J., Singh, A. K., Takata, Y., Lin, K., Shen, J., Lu, Y., Kerenyi, M. A., Orkin, S. H. & Chen, T. 2015. LSD1 is essential for oocyte meiotic progression by regulating CDC25B expression in mice. *Nature Communications* 6: 10116.

Knekt, P., Hakulinen, T., Leino, A., Heliövaara, M., Reunanen, A. & Stevens, R. 2000. Serum albumin and colorectal cancer risk. *European Journal of Clinical Nutrition* 54: 460.

Kolmeder, C. A., de Been, M., Nikkilä, J., Ritamo, I., Mättö, J., Valmu, L., Salojärvi, J., Palva, A., Salonen, A., de Vos, W. M., Nikkil??, J., Ritamo, I., M??tt??, J., Valmu, L., Saloj??rvi, J., Palva, A., Salonen, A. & de Vos, W. M. 2012. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLOS One* 7(1): e29913.

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W., Huttenhower, C., Garrett, W. S. & Meyerson, M. 2012. Genomic analysis identifies association of *Fusobacterium* with colorectal

carcinoma. *Genome Research* 22(2): 292–298.

Kostic, A. D., Xavier, R. J. & Gevers, D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146(6): 1489–1499.

Kruis, W., Fric, P., Pokrotnieks, J., Lukás, M., Fixa, B., Kascák, M., Kamm, M. A., Weismueller, J., Beglinger, C., Stolte, M., Wolff, C. & Schulze, J. 2004. Maintaining remission of ulcerative colitis with the probiotic *Escherichia coli* Nissle 1917 is as effective as with standard mesalazine. *Gut* 53(11): 1617–1623.

Kulasingam, V. & Diamandis, E. P. 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice. Oncology* 5(10): 588–99.

Kumamoto, C. A. 2011. Inflammation and gastrointestinal *Candida* colonization. *Current Opinion in Microbiology* 14(4): 386–391.

Kume, H., Muraoka, S., Kuga, T., Adachi, J., Narumi, R., Watanabe, S., Kuwano, M., Kodera, Y., Matsushita, K., Fukuoka, J., Masuda, T., Ishihama, Y., Matsubara, H., Nomura, F. & Tomonaga, T. 2014. Discovery of colorectal cancer biomarker candidates by membrane proteomic analysis and subsequent verification using selected reaction monitoring (SRM) and tissue microarray (TMA) analysis. *Molecular & Cellular Proteomics : MCP* 13(6): 1471–1484.

Kune, G. A. 2012. Causes and control of colorectal cancer: a model for cancer prevention, 78. Springer Science & Business Media.

Kusmiati & Dhewantara, F. X. R. 2016. Cholesterol-lowering effect of beta glucan extracted from *Saccharomyces cerevisiae* in Rats. *Scientia Pharmaceutica* 84(1): 153–165.

Laferrière, J., Houle, F., Taher, M. M., Valerie, K. & Huot, J. 2001. Transendothelial migration of colon carcinoma cells requires expression of E-selectin by endothelial cells and activation of stress-activated protein kinase-2 (SAPK2/p38) in the tumor cells. *Journal of Biological Chemistry* 276(36): 33762–33772.

Lai, C.-C., You, J.-F., Yeh, C.-Y., Chen, J.-S., Tang, R., Wang, J.-Y. & Chin, C.-C. 2011. Low preoperative serum albumin in colon cancer: a risk factor for poor outcome. *International Journal of Colorectal Disease* 26(4): 473–481.

Lange, S. S., Takata, K. & Wood, R. D. 2011. DNA polymerases and cancer. *Nature Reviews Cancer* 11(2): 96–110.

Lasko, T. A., Bhagwat, J. G., Zou, K. H. & Ohno-Machado, L. 2005. The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics* 38(5): 404–415.

Lasry, A., Zinger, A. & Ben-Neriah, Y. 2016. Inflammatory networks underlying colorectal cancer. *Nature Immunology* 17(3): 230–240.

Lee, D., Koo, J. S., Kang, S., Kim, S. Y., Hyun, J. J., Jung, S. W., Yim, H. J. & Lee, S. W. 2017. Association between bowel habits and quality of bowel preparation for colonoscopy. *Medicine* 96(29): e7319.

Lee, H. P., Ling, A. & Foo, L. L. 2017. *Singapore Cancer Registry Annual Registry Report 2015*. Singapore. https://www.nrdo.gov.sg/docs/librariesprovider3/Publications -Cancer/cancer-registry-annual-report-2015_web.pdf?sfvrsn=10 [1 August 2018].

Levin, B., Lieberman, D. A., McFarland, B., Smith, R. A., Brooks, D., Andrews, K. S., Dash, C., Giardiello, F. M., Glick, S., Levin, T. R., Pickhardt, P., Rex, D. K., Thorson, A. & Winawer, S. J. 2008. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US multi-society task force on colorectal cancer, and the American College of Radiology*†. *CA: A Cancer Journal for Clinicians* 58(3): 130–160.

Levin, P. A. & Angert, E. R. 2015. Small but mighty: Cell size and bacteria. *Cold Spring Harbor Perspectives in Biology* 7(7): a019216.

Lewis, S. J. & Heaton, K. W. 1997. Stool form scale as a useful guide to intestinal transit time. *Scandinavian Journal of Gastroenterology* 32(9): 920–924.

Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. 2007. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nature Reviews Immunology* 7: 678.

Li, L., Zhang, Z., Wang, C., Miao, L., Zhang, J., Wang, J., Jiao, B. & Zhao, S. 2014. Quantitative proteomics approach to screening of potential diagnostic and therapeutic targets for laryngeal carcinoma. *PLOS One* 9(2): e90181–e90181.

Liang, Q., Chiu, J., Chen, Y., Huang, Y., Higashimori, A., Fang, J., Brim, H., Ashktorab, H., Ng, S. C., Ng, S. S. M., Zheng, S., Chan, F. K. L., Sung, J. J. Y. & Yu, J. 2017. Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. *Clinical Cancer Research* 23(8): 2061 LP-2070.

Lichtman, J. S., Marcobal, A., Sonnenburg, J. L. & Elias, J. E. 2013. Host-centric proteomics of stool: a novel strategy focused on intestinal responses to the gut microbiota. *Molecular & Cellular Proteomics : MCP* 12(11): 3310–3318.

Lichtman, J. S., Sonnenburg, J. L. & Elias, J. E. 2015. Monitoring host responses to the gut microbiota. *The ISME Journal* 9(9): 1–8.

Lin, J. S., Piper, M. A., Perdue, L. A., Rutter, C. M., Webber, E. M., O'Connor, E., Smith, N. & Whitlock, E. P. 2016. Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. *Jama* 315(23): 2576–2594.

Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F. & Wang, Z. 1999. The human REV1

gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Research* 27(22): 4468–4475.

Liotta, L. A. & Kohn, E. C. 2001. The microenvironment of the tumour-host interface. *Nature* 411(6835): 375.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. & Darnell, J. 2000. Cell-cell adhesion and communication. *Molecular Cell Biology*, 4th Edition . New York: WH Freeman.

Loei, H., Tan, H. T., Lim, T. K., Lim, K. H., So, J. B.-Y., Yeoh, K. G. & Chung, M. C. M. 2012. Mining the gastric cancer secretome: Identification of GRN as a potential diagnostic marker for early gastric cancer. *Journal of Proteome Research* 11(3): 1759–1772.

Loktionov, A. 2007. Cell exfoliation in the human colon: myth, reality and implications for colorectal cancer screening. *International Journal of Cancer* 120(11): 2281–2289.

López, P. J. T., Albero, J. S. & Rodríguez-Montes, J. A. 2014. Primary and Secondary Prevention of Colorectal Cancer. *Clinical Medicine Insights. Gastroenterology* 7: 33–46.

Łusiak-Szelachowska, M., Weber-Dąbrowska, B., Jończyk-Matysiak, E., Wojciechowska, R. & Górski, A. 2017. Bacteriophages in the gastrointestinal tract and their implications. *Gut pathogens* 9: 44.

Lyth, O., Vizcay-Barrena, G., Wright, K. E., Haase, S., Mohring, F., Najer, A., Henshall, I. G., Ashdown, G. W., Bannister, L. H., Drew, D. R., Beeson, J. G., Fleck, R. A., Moon, R. W., Wilson, D. W. & Baum, J. 2018. Cellular dissection of malaria parasite invasion of human erythrocytes using viable Plasmodium knowlesi merozoites. *Scientific Reports* 8(1): 10165.

Ma, Y., Yang, Y., Wang, F., Zhang, P., Shi, C., Zou, Y. & Qin, H. 2013. Obesity and risk of colorectal cancer: a systematic review of prospective studies. *PLOS One* 8(1): e53916.

Macrae, F. A., Bendell, J., Tanabe, K. K. & Savarese, D. M. F. 2016. Clinical presentation, diagnosis, and staging of colorectal cancer. Post TW, ed. UpToDate. Waltham, MA: UpToDate Inc. https://www.uptodate.com [5 August 2018]

Majerus, M. A. 2002. The relationship between the cancer cell and the oocyte. *Medical Hypotheses* 58(6): 544–551.

Majtan, J. & Jesenak, M. 2018. β-Glucans: Multi-functional modulator of wound healing. *Molecules (Basel, Switzerland)* 23(4): 806.

Makridakis, M. & Vlahou, A. 2010. Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics* 73(12): 2291–2305.

Maloney, A. & Workman, P. 2002. HSP90 as a new therapeutic target for cancer therapy: the story unfolds. *Expert Opinion on Biological Therapy* 2(1): 3–24.

Maloy, K. J. & Powrie, F. 2011. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* 474(7351): 298–306.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quraishi, M. N., Kinross, J., Smidt, H. & Tuohy, K. M. 2016. The gut microbiota and host health: a new clinical frontier. *Gut* 65(2): 330–339.

Marchesi, J. R., Dutilh, B. E., Hall, N., Peters, W. H. M., Roelofs, R., Boleij, A. & Tjalsma, H. 2011. Towards the human colorectal cancer microbiome. *PLOS One* 6(5): e20447.

Martin, R., Miquel, S., Ulmer, J., Langella, P. & Bermudez-Humaran, L. G. 2014. Gut ecosystem: how microbes help us. *Beneficial Microbes* 5(3): 219–233.

Martins, N., Ferreira, I. C. F. R., Barros, L., Silva, S. & Henriques, M. 2014. Candidiasis: predisposing factors, prevention, diagnosis and alternative treatment. *Mycopathologia* 177(5–6): 223–240.

Mazzolini, R., Dopeso, H., Mateo-Lozano, S., Chang, W., Rodrigues, P., Bazzocco, S., Alazzouzi, H. et al. 2012. Brush border Myosin Ia has tumor suppressor activity in the intestine. *Proceedings of The National Academy of Sciences of The United States of America* 109(5): 1530–1535.

Megat Mohd Azlan, P.-I.-H., Chin, S.-F., Low, T. Y., Hui-min, N. & Jamal, R. 2018. Analysing the secretome of gut microbiota as the next strategy for early detection of colorectal cancer. *Proteomics* 0(ja): 1800176.

Mellon, F. A. 2003. Mass Spectrometry: Principles and Instrumentation. Caballero. Oxford: Academic Press.

Methot, S. P., Litzler, L. C., Trajtenberg, F., Zahn, A., Robert, F., Pelletier, J., Buschiazzo, A., Magor, B. G. & Di Noia, J. M. 2015. Consecutive interactions with HSP90 and eEF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. *The Journal of Experimental Medicine* 212(4): 581–596.

Milicevic, Z., Bogojevic, D., Mihailovic, M., Petrovic, M. & Krivokapic, Z. 2008. Molecular characterization of hsp90 isoforms in colorectal cancer cells and its association with tumour progression. *International Journal of Oncology* 32(6): 1169–1178.

Milo, R. & Phillips, R. 2015. Cell biology by the numbers. Garland Science.

Minamida, S., Iwamura, M., Kodera, Y., Kawashima, Y., Ikeda, M., Okusa, H., Fujita, T., Maeda, T. & Baba, S. 2011. Profilin 1 overexpression in renal cell carcinoma. *International Journal of Urology* 18(1): 63–71.

Mitchell, S. H., Schaefer, D. C. & Dubagunta, S. 2004. A new view of occult and obscure gastrointestinal bleeding. *American Family Physician* 69(4): 875–884.

Mohd Bakri, M., Mohd Hussaini, H., Rachel Holmes, A., David Cannon, R. & Mary Rich, A. 2010. Revisiting the association between candidal infection and carcinoma, particularly oral squamous cell carcinoma. *Journal of Oral Microbiology* 2: 10.3402/jom.v2i0.5780.

Morrison, A. J., Highland, J., Krogan, N. J., Arbel-Eden, A., Greenblatt, J. F., Haber, J. E. & Shen, X. 2004. INO80 and γ-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* 119(6): 767–775.

Mousa, L., Salem, M. E. & Mikhail, S. 2015. Biomarkers of Angiogenesis in Colorectal Cancer. *Biomarkers in Cancer* 7(Suppl 1): 13–19.

Muller, W. A. 2011. Mechanisms of leukocyte transendothelial migration. *Annual Review of Pathology* 6: 323–344.

Münster, P. N., Basso, A., Solit, D., Norton, L. & Rosen, N. 2001. Modulation of Hsp90 function by ansamycins sensitizes breast cancer cells to chemotherapy-induced apoptosis in an rb-and schedule-dependent manner: see the biology behind: ea sausville, combining cytotoxics and 17-allylamino, 17-demethoxygeldanamycin: seq. *Clinical Cancer Research* 7(8): 2228–2236.

Muz, B., de la Puente, P., Azab, F. & Azab, A. K. 2015. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia* 3: 83–92.

Nagy, M. A. 2011. HIF-1 is the Commander of Gateways to Cancer. *Journal of Cancer Science and Therapy* 3(2): 35–40.

Nakayama, T., Yasuoka, H., Kishino, T., Ohguchi, H. & Takada, M. 1987. ELISA for occult faecal albumin. *The Lancet* 329(8546): 1368–1369.

Nasser, T. I. N. & Spencer, G. E. B. T.-R. M. in B. S. 2017. Neurite Outgrowth. Elsevier.

Nelson, W. J. & Nusse, R. 2004. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303(5663): 1483–1487.

Nicholson, B. D., Shinkins, B., Pathiraja, I., Roberts, N. W., James, T. J., Mallett, S., Perera, R., Primrose, J. N. & Mant, D. 2015. Blood CEA levels for detecting recurrent colorectal cancer. *Cochrane Database of Systematic Reviews* (12).

Nicholson, F. B. & Korman, M. G. 2005. Acceptance of flexible sigmoidoscopy and colonoscopy for screening and surveillance in colorectal cancer prevention.

Journal of Medical Screening 12(2): 89–95.

Nobile, C. J. & Johnson, A. D. 2015. Candida albicans Biofilms and Human Disease. *Annual Review of Microbiology* 69: 71–92.

Nougayrède, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U. & Oswald, E. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313(5788): 848–851.

Okegawa, T., Pong, R.-C., Li, Y. & Hsieh, J.-T. 2004. The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochimica Polonica-English Edition*- 51: 445–458.

Olde Bekkink, M., McCowan, C., Falk, G. A., Teljeur, C., Van de Laar, F. A. & Fahey, T. 2009. Diagnostic accuracy systematic review of rectal bleeding in combination with other symptoms, signs and tests in relation to colorectal cancer. *British Journal of Cancer* 102: 48.

Oliveros, J. C. 2007. Venny. An interactive tool for comparing lists with Venn's diagrams. BioinfoGP Service.

Osborn, N. K. & Ahlquist, D. A. 2005. Stool screening for colorectal cancer: molecular approaches. *Gastroenterology* 128(1): 192–206.

Otani, T., Iwasaki, M., Inoue, M., Sasazuki, S. & Tsugane, S. 2006. Bowel movement, state of stool, and subsequent risk for colorectal cancer: the japan public health center–based prospective study. *Annals of Epidemiology* 16(12): 888–894.

Pandey, P., Saleh, A., Nakazawa, A., Kumar, S., Srinivasula, S. M., Kumar, V., Weichselbaum, R., Nalin, C., Alnemri, E. S. & Kufe, D. 2000. Negative regulation of cytochrome c-mediated oligomerization of Apaf-1 and activation of procaspase-9 by heat shock protein 90. *The EMBO Journal* 19(16): 4310–4322.

Pang, Q., Prolla, T. A. & Liskay, R. M. 1997. Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Molecular and Cellular Biology* 17(8): 4465–4473.

Park, J. Y., Mitrou, P. N., Luben, R., Khaw, K.-T. & Bingham, S. A. 2009. Is bowel habit linked to colorectal cancer? results from the EPIC-Norfolk study. *European Journal of Cancer* 45(1): 139–145.

Pebernard, S., Wohlschlegel, J., McDonald, W. H., Yates 3rd, J. R. & Boddy, M. N. 2006. The Nse5-Nse6 dimer mediates DNA repair roles of the Smc5-Smc6 complex. *Molecular and Cellular Biology* 26(5): 1617–1630.

Pericleous, M., Mandair, D. & Caplin, M. E. 2013. Diet and supplements and their impact on colorectal cancer. *Journal of Gastrointestinal Oncology* 4(4): 409–423.

Peyri, N., Berard, M., Fauvel-Lafeve, F., Trochon, V., Arbeille, B., Lu, H., Legrand, C. & Crepin, M. 2009. Breast tumor cells transendothelial migration induces endothelial cell anoikis through extracellular matrix degradation. *Anticancer Research* 29(6): 2347–2355.

Pinte, S., Caetano, B., Le Bras, A., Havet, C., Villain, G., Dernayka, R., Duez, C., Mattot, V. & Soncin, F. 2016. Endothelial cell activation is regulated by epidermal growth factor-like domain 7 (Egfl7) during inflammation. *The Journal of Biological Chemistry* 291(46): 24017–24028.

Portillo, M. C., Leff, J. W., Lauber, C. L. & Fierer, N. 2013. Cell size distributions of soil bacterial and archaeal taxa. *Applied and Environmental Microbiology* 79(24): 7610–7617.

Pourhoseingholi, M. A. 2014. Epidemiology and burden of colorectal cancer in Asia-Pacific region: what shall we do now? *Translational Gastrointestinal Cancer Biostatistics and Epidemiology Models in Gastrointestinal Cancers* 3(4).

Praneenararat, S. 2014. Fungal infection of the colon. *Clinical and Experimental Gastroenterology* 7: 415–426.

Pratt, W. B. & Toft, D. O. 2003. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Experimental Biology and Medicine* 228(2): 111–133.

Putze, J., Hennequin, C., Nougayrède, J.-P., Zhang, W., Homburg, S., Karch, H., Bringer, M.-A., Fayolle, C., Carniel, E. & Rabsch, W. 2009. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infection and Immunity* 77(11): 4696–4703.

Pylayeva-Gupta, Y., Lee, K. E., Hajdu, C. H., Miller, G. & Bar-Sagi, D. 2012. Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. *Cancer Cell* 21(6): 836–847.

Rabouille, C. 2017. Pathways of unconventional protein secretion. *Trends in Cell Biology* 27(3): 230–240.

Ratnam, S., March, S. B., Ahmed, R., Bezanson, G. S. & Kasatiya, S. 1988. Characterization of *Escherichia coli* serotype O157:H7. *Journal of Clinical Microbiology* 26(10): 2006–2012.

Rawl, S. M., Menon, U., Burness, A. & Breslau, E. S. 2012. Interventions to promote colorectal cancer screening: an integrative review. *Nursing Outlook* 60(4): 172–181.e13.

Redon, E., Loubière, P. & Cocaign-Bousquet, M. 2005. Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *Journal of Biological Chemistry* 280(43): 36380–36385.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N. et al.

2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-igm syndrome (HIGM2). *Cell* 102(5): 565–575.

Reymond, N., Im, J. H., Garg, R., Vega, F. M., Borda d'Agua, B., Riou, P., Cox, S., Valderrama, F., Muschel, R. J. & Ridley, A. J. 2012. Cdc42 promotes transendothelial migration of cancer cells through β1 integrin. *The Journal of Cell Biology* 199(4): 653–668.

Rijcken, E., Mennigen, R. B., Schaefer, S. D., Laukoetter, M. G., Anthoni, C., Spiegel, H.-U., Bruewer, M., Senninger, N. & Krieglstein, C. F. 2007. PECAM-1 (CD 31) mediates transendothelial leukocyte migration in experimental colitis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 293(2): G446–G452.

Rmali, K. ., Puntis, M. . & Jiang, W. . 2006. Tumour-associated angiogenesis in human colorectal cancer. *Colorectal Disease* 9(1): 3–14.

Robbiani, D. F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K. M., Klein, I. A., Stone, G., Eisenreich, T. R., Ried, T., Nussenzweig, A. & Nussenzweig, M. C. 2009. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Molecular Cell* 36(4): 631–641.

Rose, C., Parker, A., Jefferson, B. & Cartmell, E. 2015. The characterization of feces and urine: a review of the literature to inform advanced treatment technology. *Critical Reviews in Environmental Science and Technology* 45(17): 1827–1879.

Roseth, A. G., Kristinsson, J., Nygaard, K., Aadland, E., Sundset, A. & Fagerhol, M. K. 1995. Fecal lactoferrin, a novel marker of colorectal-cancer in rats, a longitudinal in-vivo study. *Gastroenterology*, 108: A530–A530.

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G. & Han, Y. W. 2013. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host and Microbe* 14(2): 195–206.

Saber, A., Alipour, B., Faghfoori, Z. & Yari Khosroushahi, A. 2017. Cellular and molecular effects of yeast probiotics on cancer. *Critical Reviews in Microbiology* 43(1): 96–115.

Sakiyama, T., Kohno, T., Mimaki, S., Ohta, T., Yanagitani, N., Sobue, T., Kunitoh, H., Saito, R., Shimizu, K. & Hirama, C. 2005. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *International Journal of Cancer* 114(5): 730–737.

Salyers, A. A. 2005. Session I: Host-Pathogen Interactions: Defining the concepts of pathogenicity, virulence, colonization, commensalism, and symbiosis. *Presentation at The Forum on Microbial Threats Workshop Ending The War Metaphor: The Changing Agenda for Unraveling The Host-Microbe Relationship*.

Washington, DC: Institute of Medicine, Forum on Microbial Threats.

Samuel, B. S. & Gordon, J. I. 2006. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proceedings of The National Academy of Sciences of The United States of America* 103(26): 10011–10016.

Sartor, R. B. 2008. Microbial influences in inflammatory bowel diseases. *Gastroenterology* 134(2): 577–594.

Schimmel, L., Heemskerk, N. & van Buul, J. D. 2016. Leukocyte transendothelial migration: A local affair. *Small GTPases* 8(1): 1–15.

Schwabe, R. F. & Jobin, C. 2013. The microbiome and cancer. *Nature Reviews Cancer* 13(11): 800–812.

Scully, A. & Cheung, I. 2016. Colorectal Cancer Screening: Fecal occult blood test literature review for occupational health nurses. *Workplace Health and Safety* 64(3): 114–122.

Sears, C. L. & Garrett, W. S. 2014. Microbes, microbiota and colon cancer. *Cell Host and Microbe* 15(3): 317–328.

Selvachandran, S. N., Hodder, R. J., Ballal, M. S., Jones, P. & Cade, D. 2002. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *The Lancet* 360(9329): 278–283.

Sergeant, J., Taylor, E., Palecek, J., Fousteri, M., Andrews, E. A., Sweeney, S., Shinagawa, H., Watts, F. Z. & Lehmann, A. R. 2005. Composition and architecture of the *Schizosaccharomyces pombe* Rad18 (Smc5-6) complex. *Molecular and Cellular Biology* 25(1): 172–184.

Seyfried, T. N., Flores, R. E., Poff, A. M. & D'Agostino, D. P. 2014. Cancer as a metabolic disease: implications for novel therapeutics. *Carcinogenesis* 35(3): 515–527.

Sheflin, A. M., Whitney, A. K. & Weir, T. L. 2014. Cancer-promoting effects of microbial dysbiosis. *Current Oncology Reports* 16(10): 406.

Shi, H. J., Stubbs, R. & Hood, K. 2009. Characterization of de novo synthesized proteins released from human colorectal tumour explants. *Electrophoresis* 30(14): 2442–2453.

Shi, Y.-H. & Fang, W.-G. 2004. Hypoxia-inducible factor-1 in tumour angiogenesis. *World Journal of Gastroenterology* 10(8): 1082–1087.

Shin, J., Kim, H.-J. H. H. J., Kim, G., Song, M., Woo, S. J., Lee, S.-T. T., Kim, H.-J. H. H. J. & Lee, C. 2014. Discovery of melanotransferrin as a serological marker of colorectal cancer by secretome analysis and quantitative proteomics. *Journal of Proteome Research* 13(11): 4919–4931.

Siegel, R. L., Miller, K. D. & Jemal, A. 2015. Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians* 65(1): 5–29.

Simons, C. C. J. M., Schouten, L. J., Weijenberg, M. P., Goldbohm, R. A. & van den Brandt, P. A. 2010. Bowel movement and constipation frequencies and the risk of colorectal cancer among men in the Netherlands cohort study on diet and cancer. *American Journal of Epidemiology* 172(12): 1404–1414.

Smith, S., Hwang, J.-Y., Banerjee, S., Majeed, A., Gupta, A. & Myung, K. 2004. Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in *Saccharomyces cerevisiae*. *Proceedings of The National Academy of Sciences* 101(24): 9039–9044.

Sobhani, I., Amiot, A., Le Baleur, Y., Levy, M., Auriault, M.-L., Van Nhieu, J. T. & Delchier, J. C. 2013. Microbial dysbiosis and colon carcinogenesis: Could colon cancer be considered a bacteria-related disease?. *Therapeutic Advances in Gastroenterology* 6(3): 215–229.

Sobhani, I., Tap, J., Roudot-Thoraval, F., Roperch, J. P., Letulle, S., Langella, P., Corthier, G., Van Nhieu, J. T. & Furet, J. P. 2011. Microbial dysbiosis in colorectal cancer (CRC) patients. *PLOS One* 6(1): e16393.

Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S. & Sonnenburg, J. L. 2016. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529(7585): 212–215.

Sőti, C., Pál, C., Papp, B. & Csermely, P. 2005. Molecular chaperones as regulatory elements of cellular networks. *Current Opinion in Cell Biology* 17(2): 210–215.

Stephen, A. M. & Cummings, J. H. 1980. The microbial contribution to human faecal mass. *Journal of Medical Microbiology* 13(1): 45–56.

Strell, C. & Entschladen, F. 2008. Extravasation of leukocytes in comparison to tumor cells. *Cell Communication and Signaling : CCS* 6: 10.

Strofilas, A., Lagoudianakis, E. E., Seretis, C., Pappas, A., Koronakis, N., Keramidaris, D., Koukoutsis, I., Chrysikos, I., Manouras, I. & Manouras, A. 2012. Association of *Helicobacter Pylori* Infection and Colon Cancer. *Journal of Clinical Medicine Research* 4(3): 172–176.

Sun, J. & Chang, E. B. 2014. Exploring gut microbes in human health and disease: Pushing the envelope. *Genes & Diseases* 1(2): 132–139.

Sun, L., Hu, S., Yu, L., Guo, C., Sun, L., Yang, Z., Qi, J. & Ran, Y. 2016. Serum haptoglobin as a novel molecular biomarker predicting colorectal cancer hepatic metastasis. *International Journal of Cancer* 138(11): 2724–2731.

Sun, W. 2012. Angiogenesis in metastatic colorectal cancer and the benefits of targeted therapy. *Journal of Hematology & Oncology* 5: 63.

Sung, J. J. Y., Ng, S. C., Chan, F. K. L., Chiu, H. M., Kim, H. S., Matsuda, T., Ng, S. S. M. et al. 2015. An updated Asia Pacific consensus recommendations on colorectal cancer screening. *Gut* 64(1): 121 LP-132.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J. & von Mering, C. 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research* 45(Database issue): D362–D368.

Takaine, M. & Mabuchi, I. 2007. Properties of actin from the fission yeast *Schizosaccharomyces pombe* and interaction with fission yeast profilin. *Journal of Biological Chemistry* 282(30): 21683–21694.

Talks, K. L., Turley, H., Gatter, K. C., Maxwell, P. H., Pugh, C. W., Ratcliffe, P. J. & Harris, A. L. 2000. The expression and distribution of the hypoxia-inducible factors HIF-1α and HIF-2α in normal human tissues, cancers, and tumor-associated macrophages. *The American Journal of Pathology* 157(2): 411–421.

Tedjo, D. I., Jonkers, D. M. A. E., Savelkoul, P. H., Masclee, A. A., van Best, N., Pierik, M. J. & Penders, J. 2015. The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. *PLOS One* 10(5): e0126685.

Terret, M.-E., Chaigne, A. & Verlhac, M.-H. 2013. Mouse oocyte, a paradigm of cancer cell. *Cell Cycle* 12(21): 3370–3376.

Terveer, E. M., van Beurden, Y. H., Goorhuis, A., Seegers, J. F. M. L., Bauer, M. P., van Nood, E., Dijkgraaf, M. G. W., Mulder, C. J. J., Vandenbroucke-Grauls, C. & Verspaget, H. W. 2017. How to: establish and run a stool bank. *Clinical Microbiology and Infection*.

Timmers, L., Lim, S. K., Arslan, F., Armstrong, J. S., Hoefer, I. E., Doevendans, P. A., Piek, J. J., El Oakley, R. M., Choo, A. & Lee, C. N. 2008. Reduction of myocardial infarct size by human mesenchymal stem cell conditioned medium. *Stem Cell Research* 1(2): 129–137.

Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. 2012. A bacterial driver-passenger model for Colorectal cancer: Beyond the usual suspects. *Nature Reviews. Microbiology* 10(8): 575–582.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D. H., Bron, S. & van Dijl, J. M. 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews : MMBR* 64(3): 515–47.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J. & Jemal, A. 2015. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 65(2): 87–108.

Triantafillidis, J. K., Nasioulas, G. & Kosmidis, P. A. 2009. Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. *Anticancer Research* 29(7): 2727–2737.

van Rossum, L. G., van Rijn, A. F., Laheij, R. J., van Oijen, M. G., Fockens, P., van Krieken, H. H., Verbeek, A. L., Jansen, J. B. & Dekker, E. 2008. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology* 135(1): 82–90.

Vandell, V. E. & Limbach, P. A. 2010. Overview of biochemical applications of mass spectrometry. Lindon. Oxford: Academic Press.

Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R. Y., Joossens, M. & Raes, J. 2015. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*.

Vargas, A. J. & Thompson, P. A. 2012. Diet and nutrient factors in colorectal cancer risk. *Nutrition in Clinical Practice* 27(5): 613–623.

Veettil, S. K., Lim, K. G., Chaiyakunapruk, N., Ching, S. M. & Abu Hassan, M. R. 2016. Colorectal cancer in Malaysia: Its burden and implications for a multiethnic country. *Asian Journal of Surgery*.

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L. & Jansson, J. K. 2008. Shotgun metaproteomics of the human distal gut microbiota. *ISME J* 3(2): 179–189.

Volk, T. & Rummel, J. D. 1987. Mass balances for a biological life support system simulation model.

Volmer, M. W., Stühler, K., Zapatka, M., Schöneck, A., Klein-Scory, S., Schmiegel, W., Meyer, H. E. & Schwarte-Waldhoff, I. 2005. Differential proteome analysis of conditioned media to detect Smad4 regulated secreted biomarkers in colon cancer. *Proteomics* 5(10): 2587–2601.

Walter, J. & Ley, R. 2011. The human gut microbiome: ecology and recent evolutionary changes. *Annual Review of Microbiology* 65(1): 411–429.

Wang, D., Fu, L., Sun, H., Guo, L. & DuBois, R. N. 2015. Prostaglandin E(2) promotes colorectal cancer stem cell expansion and metastasis in mice. *Gastroenterology* 149(7): 1884–1895.e4.

Wang, H.-P., Wang, Y.-Y., Pan, J., Cen, R. & Cai, Y.-K. 2014. Evaluation of specific fecal protein biochips for the diagnosis of colorectal cancer. *World Journal of Gastroenterology : WJG* 20(5): 1332–1339.

Wang, L. & Lin, X. 2012. Morphogenesis in fungal pathogenicity: shape, size, and surface. *PLOS Pathogens* 8(12): e1003027.

Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W. & Zheng, S.-S. 2015. Application of metagenomics in the human gut microbiome. *World Journal of Gastroenterology : WJG* 21(3): 803–814.

WHO, I., Organization, W. H. & Cancer, I. A. for R. on. 1994. IARC monographs on the evaluation of carcinogenic risks to humans: Schistosomes, liver flukes and *Helicobacter pylori*. *IARC, Lyon*.

Wong, S. H. & Ng, S. C. 2013. Colorectal Cancer Screening in Asia. *British Medical Bulletin* 1–14.

Workman, P. 2004. Altered states: selectively drugging the Hsp90 cancer chaperone. *Trends in Molecular Medicine* 10(2): 47–51.

Wu, C.-C., Chen, H.-C., Chen, S.-J., Liu, H.-P., Hsieh, Y.-Y., Yu, C.-J., Tang, R., Hsieh, L.-L., Yu, J.-S. & Chang, Y.-S. 2008. Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics* 8(2): 316–332.

Wu, C.-C., Huang, Y.-S., Lee, L.-Y., Liang, Y., Tang, R.-P., Chang, Y.-S., Hsieh, L.-L. & Yu, J.-S. 2008. Overexpression and elevated plasma level of tumor-associated antigen 90K/Mac-2 binding protein in colorectal carcinoma. *Proteomics. Clinical Applications* 2(12): 1586–1595.

Wu, W.-H., Wu, C.-H., Ladurner, A., Mizuguchi, G., Wei, D., Xiao, H., Luk, E., Ranjan, A. & Wu, C. 2009. N terminus of Swr1 binds to histone H2AZ and provides a platform for subunit assembly in the chromatin remodeling complex. *Journal of Biological Chemistry* 284(10): 6200–6207.

Wu, W.-K., Chen, C.-C., Panyod, S., Chen, R.-A., Wu, M.-S., Sheen, L.-Y. & Chang, S.-C. 2018. Optimization of fecal sample processing for microbiome study — the journey from bathroom to bench. *Journal of The Formosan Medical Association*.

Xie, K., Doles, J., Hemann, M. T. & Walker, G. C. 2010. Error-prone translesion synthesis mediates acquired chemoresistance. *Proceedings of The National Academy of Sciences of The United States of America* 107(48): 20792–20797.

Xiong, W., Abraham, P. E., Li, Z., Pan, C. & Hettich, R. L. 2015. Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics* 15(20): 3424–3438.

Xu, G. & Jaffrey, S. R. 2011. The new landscape of protein ubiquitination. *Nature Biotechnology* 29(12): 1098–1100.

Xu, X., Qiao, M., Zhang, Y., Jiang, Y., Wei, P., Yao, J., Gu, B., Wang, Y., Lu, J. & Wang, Z. 2010. Quantitative proteomics study of breast cancer cell lines isolated

from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* 10(7): 1374–1390.

Xue, H., Lü, B., Zhang, J., Wu, M., Huang, Q., Wu, Q., Sheng, H., Wu, D., Hu, J. & Lai, M. 2009. Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *Journal of proteome research* 9(1): 545–555.

Yoo, B. C., Hong, S. H., Ku, J. L., Kim, Y. H., Shin, Y. K., Jang, S. G., Kim, I. J., Jeong, S. Y. & Park, J. G. 2009. Galectin-3 stabilizes heterogeneous nuclear ribonucleoprotein Q to maintain proliferation of human colon cancer cells. *Cellular and Molecular Life Sciences* 66(2): 350–364.

Yoon, T. J., Kim, T. J., Lee, H., Shin, K. S., Yun, Y. P., Moon, W. K., Kim, D. W. & Lee, K. H. 2008. Anti-tumor metastatic activity of β-glucan purified from mutated *Saccharomyces cerevisiae*. *International Immunopharmacology* 8(1): 36–42.

Yu, E. Y., Steinberg-Neifach, O., Dandjinou, A. T., Kang, F., Morrison, A. J., Shen, X. & Lue, N. F. 2007. Regulation of telomere structure and functions by subunits of the INO80 chromatin remodeling complex. *Molecular and Cellular Biology* 27(16): 5639–5649.

Zackular, J. P., Baxter, N. T., Iverson, K. D., Sadler, W. D., Petrosino, J. F., Chen, G. Y. & Schloss, P. D. 2013. The gut microbiome modulates colon tumorigenesis. *mBio* 4(6): e00692-13.

Zackular, J. P., Rogers, M. A. M., Ruffin, M. T. & Schloss, P. D. 2014. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prevention Research* 7(11): 1112–1121.

Zeng, X., Yang, P., Chen, B., Jin, X., Liu, Y., Zhao, X. & Liang, S. 2013. Quantitative secretome analysis reveals the interactions between epithelia and tumor cells by in vitro modulating colon cancer microenvironment. *Journal of Proteomics* 89: 51–70.

Zhang, S.-Y., Lin, M. & Zhang, H.-B. 2015. Diagnostic value of carcinoembryonic antigen and carcinoma antigen 19-9 for colorectal carcinoma. *International Journal of Clinical and Experimental Pathology* 8(8): 9404–9409.

Zhang, Y.-J., Li, S., Gan, R.-Y., Zhou, T., Xu, D.-P. & Li, H.-B. 2015. Impacts of gut bacteria on human health and diseases. *International Journal of Molecular Sciences* 16(4): 7493–7519.

Zhang, Y., Hoffmeister, M., Weck, M. N., Chang-Claude, J. & Brenner, H. 2012. Helicobacter pylori infection and colorectal cancer risk: evidence from a large population-based case-control study in Germany. *American Journal of Epidemiology* 175(5): 441–450.

APPENDIX A

RESEARCH ETHICS APPROVAL LETTER

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

Universiti Kebangsaan Malaysia | *The National University of Malaysia*

Rujukan : UKM1.5.3.5/244/JEP-2016-072
Tarikh :29 Februari 2016

Profesor Madya Dr. Raja Affendi Ali
Institut Perubatan Molekul UKM (UMBI)
Cheras

Y. Bhg. Profesor/Datuk/Dato'/Datin/Tuan/Puan,

Kelulusan Etika Menjalankan Penyelidikan di UKM

Tajuk : *Gut Microbiome: The Exciting New Frontier In Understanding Colorectal Cancer*

Perkara yang tersebut di atas adalah dirujuk.

Sukacita dimaklumkan, Jawatankuasa Etika Penyelidikan UKM meluluskan permohonan penyelidikan Y. Bhg. Profesor/Datuk/Dato'/Datin/Tuan/Puan bagi tajuk diatas. Tempoh kelulusan penyelidikan adalah daripada **25 Februari 2016 – 24 Februari 2019.** Sila kemukakan sebarang **Laporan Kesan Sampingan, Laporan Kemajuan Setiap 6 Bulan** dan **Laporan Akhir** sebaik sahaja penyelidikan tamat kepada Jawatankuasa Etika Penyelidikan UKM.

Sekian, terima kasih.

Yang benar,

Profesor Madya (K) Dato' Dr. Fuad Ismail
Pengerusi
Jawatankuasa Etika Penyelidikan
Universiti Kebangsaan Malaysia

s.k. - **Timbalan Dekan (Penyelidikan & Inovasi)**
Sekretariat Penyelidikan Perubatan & Inovasi
Hospital Canselor Tuanku Muhriz
Pusat Perubatan UKM

- **Pengarah**
Institut Perubatan Molekul UKM (UMBI)
Cheras

- **Profesor Datuk Dr. A Rahman A Jamal**
- **Profesor Dato' Dr. Wan Zurinah Wan Ngah**
- **Dr. Neoh Hui-Min,**
- **Dr. Chin Siok Fong**
- **Dr. Asiah Kassim**
- **Muhammad Afiq Osman (P83380-Calon PhD)**
- **Putri Intan Hafizah Megat Mohd Azlan (P83382-Calon PhD)**
- Institut Perubatan Molekul UKM (UMBI)
Cheras

Sekretariat Etika Penyelidikan Universiti Kebangsaan Malaysia, Tingkat 1, Blok Klinikal, Hospital Canselor Tuanku Muhriz, Pusat Perubatan UKM
Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras Kuala Lumpur. MALAYSIA.
Telefon: **+603-9145 5046 / 5048**

Mengilham Harapan, Mencipta Masa Depan • *Inspiring Futures, Nurturing Possibilities* | www.ukm.my

RESEARCH ETHICS APPROVAL LETTER (page 2)

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

Universiti Kebangsaan Malaysia | ***The National University of Malaysia***

NAME OF ETHICS COMMITTEE/IRB: Research Ethics Committee, The National University of Malaysia	**ETHICS COMMITTEE/IRB REF NO :** **UKM 1.5.3.5/244/JEP-2016-072**
PROTOCOL TITLE : Gut Microbiome: The Exciting New Frontier In Understanding Colorectal Cancer	
PRINCIPAL INVESTIGATOR : Associate Professor Dr. Raja Affendi Ali UKM Medical Molecular Biology Institute (UMBI) Cheras	

The following items (/) have been received and reviewed in connection with the above study to be conducted by the above investigator.

Documents

(/) Research Application Form
(/) Research Proposal
(/) Non-Disclosure Agreement
( ) Project Agreement
(/) Publication Policy
(/) Information Sheet (Malay & English) & Consent Form (Malay & English)
( ) Questionnaire (Malay & English)
(/) Curriculum Vitae of Researcher

The Research Ethics Committee, The National University of Malaysia operates in accordance to the International Conference of Harmonization Good Clinical Practice Guidelines.

Comments (if any): Associate Professor Dr. Norfilza Mokhtar is the co-investigator for this study and also member of Research Ethics Committee. She Strictly was not involved in the decision of Research Ethics Committee to approve this study.

Date of Approval: 25 February 2016

Associate Professor (Clinical) Dato' Dr. Fuad Ismail
Chairman
Research Ethics Committee
The National University of Malaysia

APPENDIX B

RESEARCH PARTICIPANT'S INFORMATION SHEET IN ENGLISH LANGUAGE

INFORMATION SHEET FOR PARTICIPANT

Research Title
Gut Microbiome: The Exciting New Frontier in Understanding Colorectal Cancer.

Introduction
You are invited to participate in a research study. Before participating in this study, it is important that you take time to read and understand the information in this Information Sheet.

Purpose of Study
Colorectal cancer (CRC) is the most common cancer for men and the second most common for women. Several studies have shown that gut microbiome may play a role in triggering intestinal inflammation that leads to the development of CRC. Gut microbiome is the collection of microorganisms that inhabit the gut. Therefore, manipulation of the gut microbiome via administration of probiotics may potentially improve the health and nutritional status in patients with CRC. The aims of this study are to determine the gut microbiome of Malaysian CRC patients, to profile the corresponding secreted proteins of these microbial communities, and to investigate the role of probiotic functional foods in reducing CRC-related inflammatory markers and symptom alleviation.

What will the study involve?
You will be required to complete a questionnaire which includes information on age, medical history, background details and diet. Swabs of colonic tissue biopsies, blood and fecal samples will be collected from you during your visit to the colonoscopy clinic (routine colonoscopic procedure). These samples will be processed and analysed. Any excess specimens will be stored. Future use of these samples for other studies would be anonymised.

The benefits
There is unlikely to be a direct benefit for you in participating in this study. The probiotics given may or may not help to improve your CRC condition. However, your participation could help us improve on the measures to alleviate CRC condition in future patients.

The risks
There are no additional risks involved. Only routine procedures will be performed during sample collection.

Data and confidentiality
The data from this study will be made into a report which may be published. Access to the data is only by the research team and the REC UKM. The data will be reported in a collective manner with no reference to an individual. Hence your identity will be kept confidential.

Do I have to take part?
Participation in this study is voluntary. If you agree to take part, then you will be asked to sign the "Informed Consent Form". You will be given a copy of the form and this Information Sheet.

Your treatment is not affected if you decide not to participate in this study. You will undergo usual endoscopy examination.

RESEARCH PARTICIPANT'S INFORMATION SHEET IN ENGLISH LANGUAGE (PAGE 2)

Should you decide to participate, you can still withdraw from the study without penalty. Your data will not be used and will be discarded. The researcher may also remove you from the study for a variety of reason. In this event, you will not be penalised or lose your rights as a patient.

Payment and compensation
You do not have to pay nor will you be paid to participate in this study. You do have to pay for the usual hospital charges.

Should you develop complications related to the study, then you will be compensated accordingly.

Who can I ask about the study?
If you have any questions, you can direct them to the research team. You can also contact the researcher for clarifications.

Assoc. Professor Dr. Raja Affendi Raja Ali
UKM Medical Molecular Biology Institute (UMBI)
UKM Medical Centre (UKMMC)
Jalan Ya'acob Latiff, Bandar Tun Razak
56000 Cheras, Kuala Lumpur
Tel : 03-91718459
Fax : 03-91717185

APPENDIX C

RESEARCH PARTICIPANT'S INFORMATION SHEET IN MALAY LANGUAGE

MAKLUMAT UNTUK PESERTA

Tajuk Penyelidikan
Gut Microbiome: The Exciting New Frontier in Understanding Colorectal Cancer.

Pengenalan
Anda dijemput untuk mengambil bahagian dalam kajian penyelidikan ini. Sebelum mengambil bahagian, adalah penting bagi anda mengambil masa untuk membaca dan memahami maklumat di dalam helaian maklumat ini.

Tujuan Penyelidikan
Kanser kolorektal (CRC) adalah kanser paling umum bagi lelaki dan kedua umum bagi wanita. Kajian telah menunjukkan bahawa *microbiome* usus berkemungkinan besar memainkan peranan sebagai pencetus inflamasi usus yang membawa kepada CRC. *Microbiome* usus ialah kumpulan mikroorganisme yang menghuni usus. Oleh yang demikian, manipulasi ke atas *microbiome* usus melalui pengambilan probiotik mungkin dapat meningkatkan status kesihatan dan nutrisi pesakit CRC. Tujuan penyelidikan ini adalah untuk menentukan *microbiome* usus bagi pesakit CRC di Malaysia, memprofil protein yang dirembeskan oleh komuniti mikrob tersebut dan menentukan peranan makanan fungsian probiotik dalam menurunkan penanda inflamasi berkaitan CRC serta pengurangan simptom penyakit.

Apa yang akan dilakukan?
Anda perlu melengkapkan borang soal-selidik yang merangkumi perincian tentang umur, sejarah perubatan, latar belakang dan diet. Sampel swab daripada biopsi tisu kolon, darah dan najis akan dikumpulkan daripada anda semasa lawatan anda ke klinik kolonoskopi (prosedur rutin bagi kolonoskopi). Sampel anda akan diproses dan dianalisis. Lebihan spesimen akan disimpan. Penggunaan sampel untuk kajian yang akan datang akan dikodkan.

Faedah Penyelidikan
Pengambilan bahagian dalam penyelidikan ini tidak memberi faedah terus kepada anda. Probiotik yang diberi mungkin membantu atau tidak membantu memperbaiki keadaan CRC anda. Namun begitu, penglibatan anda dapat membantu kami membaiki cara untuk mengurangkan masalah keadaan CRC bagi pesakit di masa hadapan.

Risiko
Tiada risiko tambahan yang terlibat. Hanya prosedur rutin akan dilakukan semasa pengambilan sampel.

Data dan kerahsiaan
Data yang diperolehi daripada penyelidikan ini akan dihasilkan dalam bentuk laporan yang akan diterbit. Akses kepada data hanya oleh ahli kumpulan penyelidikan ini dan Jawatankuasa Etika Penyelidikan UKM. Data akan dilaporkan secara keseluruhan tanpa merujuk kepada seseorang individu. Oleh yang demikian, identiti anda adalah sulit.

RESEARCH PARTICIPANT'S INFORMATION SHEET IN MALAY LANGUAGE (PAGE 2)

Perlukah saya mengambil bahagian?
Penglibatan anda dalam penyelidikan ini adalah secara sukarela. Sekiranya anda bersetuju untuk mengambil bahagian, anda akan diminta untuk menandatangani "Borang Keizinan". Anda akan diberi satu salinan borang tersebut dan helaian maklumat ini.

Sekiranya anda tidak bersetuju untuk mengambil bahagian dalam penyelidikan ini, rawatan anda tidak akan terjejas. Anda akan menjalani pemeriksaan endoskopi seperti biasa.

Sekiranya anda mengambil keputusan untuk mengambil bahagian, anda masih boleh menarik diri daripada penyelidikan ini tanpa dikenakan denda. Data anda tidak akan digunakan dan akan dibuang. Penyelidik juga boleh mengeluarkan anda daripada kajian di atas pelbagai sebab. Dalam hal ini, anda tidak akan dikenakan denda atau kehilangan hak anda sebagai pesakit.

Bayaran dan pampasan
Anda tidak akan dikenakan bayaran dan anda juga tidak akan dibayar bagi penglibatan dalam penyelidikan ini. Anda hanya perlu membayar caj hospital yang biasa. Sekiranya anda mengalami komplikasi yang berkaitan dengan penyelidikan ini, anda akan dibayar pampasan yang sewajarnya.

Siapakah yang boleh saya hubungi mengenai penyelidikan ini?
Sekiranya anda mempunyai sebarang pertanyaan, anda boleh menghubungi ahli kumpulan penyelidikan ini. Anda juga boleh menghubungi penyelidik untuk maklumat lanjut.

Profesor Madya Dr. Raja Affendi Raja Ali
Institut Penyelidikan Molekul UKM (UMBI)
Pusat Perubatan UKM (PPUKM)
Jalan Ya'acob Latiff, Bandar Tun Razak
56000 Cheras, Kuala Lumpur
Tel : 03-91718459
Fax : 03-91717185

APPENDIX D

RESEARCH PARTICIPANT'S INFORMED CONSENT FORM IN ENGLISH LANGUAGE

INFORMED CONSENT FORM

Research Title:

Gut Microbiome: The Exciting New Frontier in Understanding Colorectal Cancer.

Researcher's Name: Assoc. Prof. Dr. Raja Affendi Raja Ali

Consent:

I, ..., IC No : .....

- have read the information in the Patient Information Sheet **including information regarding the risk in this study**
- have been given time to think about it and all of my questions have been answered to my satisfaction.
- understand that I may freely choose to withdraw from this study at anytime without reason and without repercussion
- understand that my anonymity will be ensured in the write-up.

I hereby, voluntarily agree to be part of this research study, to follow the study procedures, and to provide necessary information to the doctor, nurses, or other staff members, as requested.

Signature : Date :

Witness	**Medical Officer**
..... (Name)	..... (Name)
..... (IC Number)	..... (IC Number)
..... (Signature)	..... (Signature)
..... (Date)	..... (Date)

APPENDIX E

RESEARCH PARTICIPANT'S INFORMED CONSENT FORM IN MALAY LANGUAGE

BORANG KEIZINAN

Tajuk Penyelidikan:

Gut Microbiome: The Exciting New Frontier in Understanding Colorectal Cancer.

Nama Penyelidik: Prof. Madya Dr. Raja Affendi Raja Ali

Persetujuan:

Saya, ..., No Kad Pengenalan: .....

- telah membaca maklumat dalam Helaian Maklumat untuk Peserta **termasuk maklumat tentang risiko dalam penyelidikan ini**
- telah diberi masa untuk memikirkan tentang penyelidikan ini dan semua pertanyaan daripada saya telah dijawab serta memenuhi kepuasan saya
- faham bahawa saya bebas untuk menarik diri daripada penyelidikan ini pada bila-bila masa tanpa memberi sebab
- faham bahawa identiti saya tidak akan didedahkan dalam laporan.

Saya, dengan ini bersetuju secara sukarela untuk mengambil bahagian dalam penyelidikan ini, mengikut prosedur kajian and memberi maklumat yang diperlukan kepada doktor, jururawat atau kakitangan.

Tandatangan : Tarikh :

Saksi	**Pegawai Perubatan**
..... (Nama)	..... (Nama)
..... (No. Kad Pengenalan)	..... (No. Kad Pengenalan)
..... (Tandatangan)	..... (Tandatangan)
..... (Tarikh)	..... (Tarikh)

APPENDIX F

THE CATEGORIZATION OF STOOL SAMPLES ACCORDING TO BRISTOL STOOL SCALE

The categorization of CRC and control stool samples collected in present study according to Bristol Stool Scale (BSS)

No.	Samples	CRC Staging	BSS Category	Appearance and Consistency
1	T1	III	7	Watery consistency with traces of blood
2	T3	I	6	Fluffy pieces with ragged edges consistency with traces of blood
3	T5	III	5	Normal consistency with no traces of blood
4	T6	II	7	Watery consistency with traces of blood and mucus
5	T8	II	7	Watery consistency with traces of blood and mucus
6	T9	III	7	Watery consistency with traces of blood
7	T10	II	6	Fluffy pieces with ragged edges consistency with traces of blood
8	T13	II	7	Watery consistency with traces of blood and mucus
9	T14	II	6	Fluffy pieces with ragged edges consistency with normal constituents
10	T15	III	7	Watery consistency with traces of blood and mucus
11	T19	III	6	Fluffy pieces with ragged edges consistency with normal constituents
12	T20	III	6	Fluffy pieces with ragged edges consistency with normal constituents
13	T22	I	7	Watery consistency no blood traces observed
14	T23	I	6	Fluffy pieces with ragged edges consistency with traces of blood and mucus
15	T24	IV	7	Watery consistency with traces of blood and mucus
16	T26	II	6	Fluffy pieces with ragged edges consistency with traces of blood
17	T27	III	3	Hard form consistency with normal constituents
18	T28	II	6	Fluffy pieces with ragged edges consistency with traces of blood
19	T30	IV	4	Smooth formed stool with no blood traces
20	T32	II	6	Fluffy pieces with ragged edges consistency with traces of blood

to be continued…

…continuation

21	T33	III	4	Soft form semisolid with traces of blood and mucus
22	T34	III	7	Watery consistency with blood traces
23	T35	III	7	Watery consistency with blood traces and mucus
24	T36	III	7	Watery consistency with blood traces and mucus
25	T37	III	6	Fluffy pieces with ragged edges consistency with traces of blood and mucus
26	T38	III	6	Fluffy pieces with ragged edges consistency with traces of blood and mucus
27	N1	-	7	Watery consistency with normal constituents and small amounts of undigested roughage
28	N2	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
29	N5	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
30	N6	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
31	N7	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
32	N8	-	3	Soft form semisolid consistency with normal constituents
33	N9	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
34	N10	-	3	Soft form semisolid consistency with normal constituents
35	N11	-	6	Fluffy pieces with ragged edges consistency with normal constituents
36	N12	-	3	Soft form semisolid consistency with normal constituents
37	N13	-	5	Fluffy pieces with ragged edges consistency with normal constituents and small amounts of undigested roughage
38	N14	-	3	Soft form semisolid consistency with normal constituents

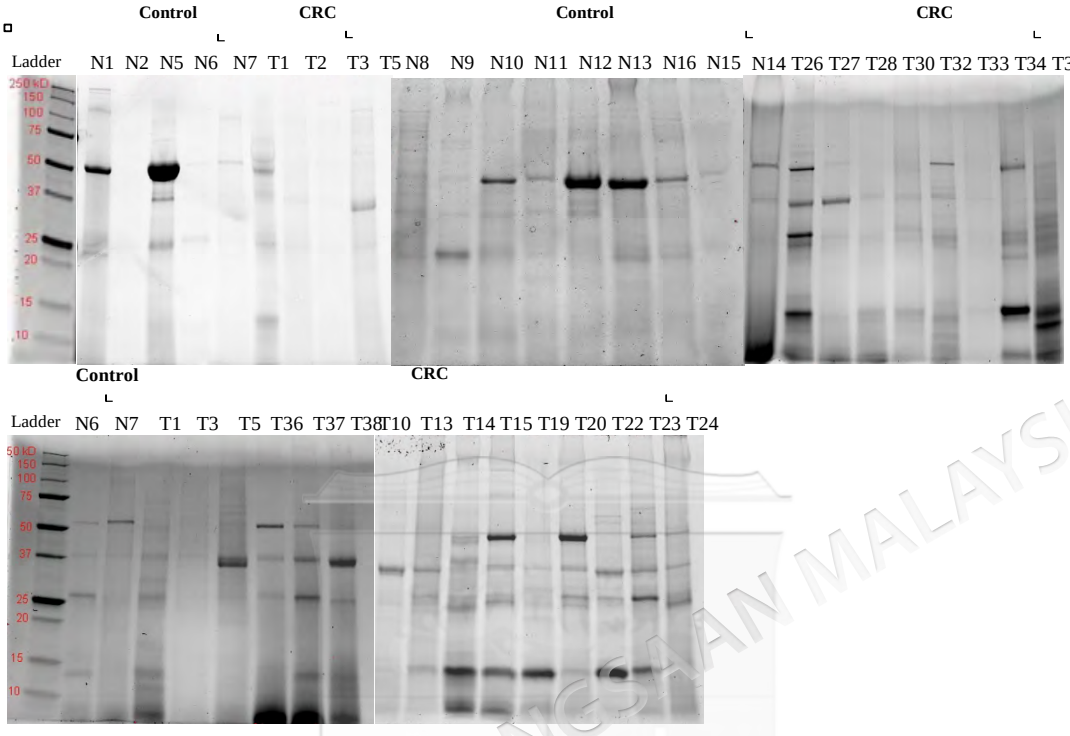
to be continued…

…continuation

39	N15	-	3	Soft form semisolid consistency with normal constituents and small amounts of undigested roughage
40	N16	-	3	Soft form semisolid consistency with normal constituents
41	N17	-	3	Soft form semisolid consistency with normal constituents
42	N18	-	5	Fluffy pieces with ragged edges consistency with normal constituents
43	N19	-	2	Hard form consistency with normal constituents
44	N20	-	5	Fluffy pieces with ragged edges consistency with normal constituents
45	N22	-	3	Soft form semisolid consistency with normal constituents
46	N23	-	3	Fluffy pieces with ragged edges consistency with normal constituents

APPENDIX G

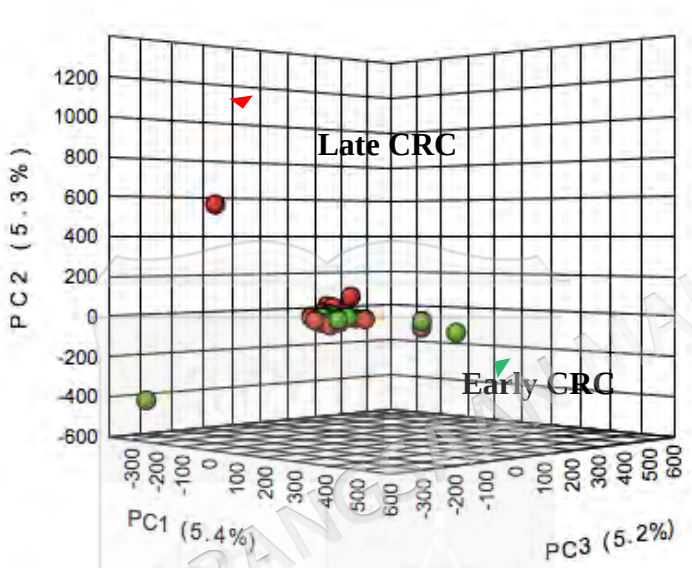
SDS-PAGE GEL IMAGES OF CRC AND CONTROL SAMPLES



SDS-PAGE gel images of remaining CRC and control secretome samples aside shown in Figure 4.1.

APPENDIX H

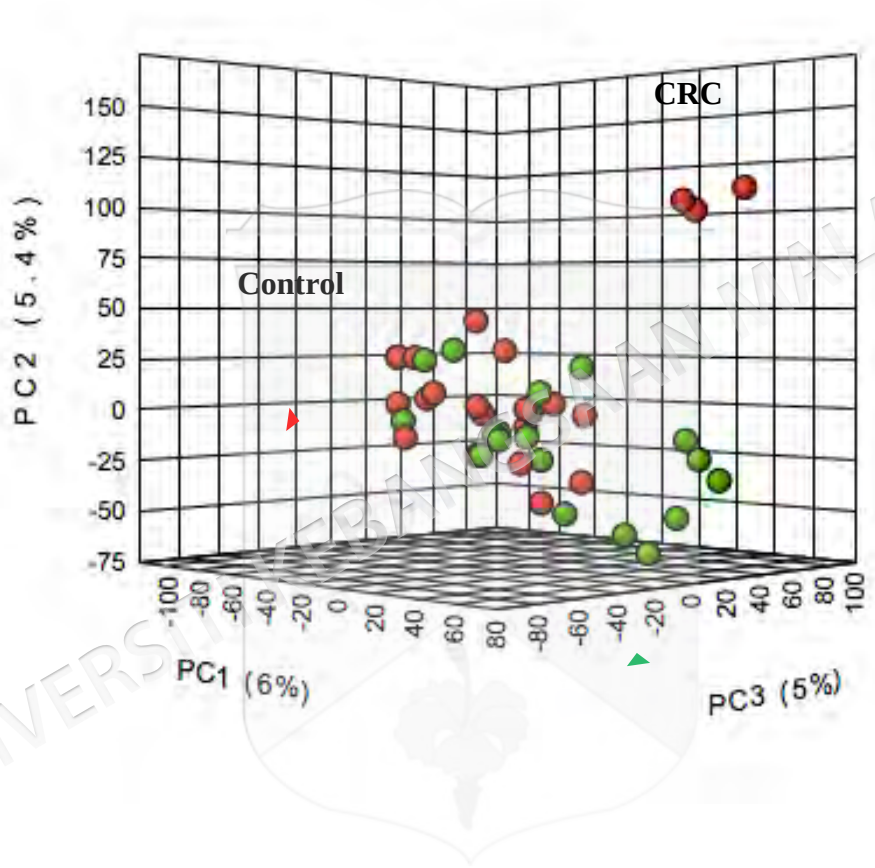
PRINCIPAL COMPONENT ANALYSIS (PCA) OF EXCLUSIVE HUMAN PROTEINS IN CRC



A principal component analysis (PCA) of human proteins exclusively found in colorectal cancer (CRC) grouped according to tumour stages, early and late CRC.

APPENDIX I

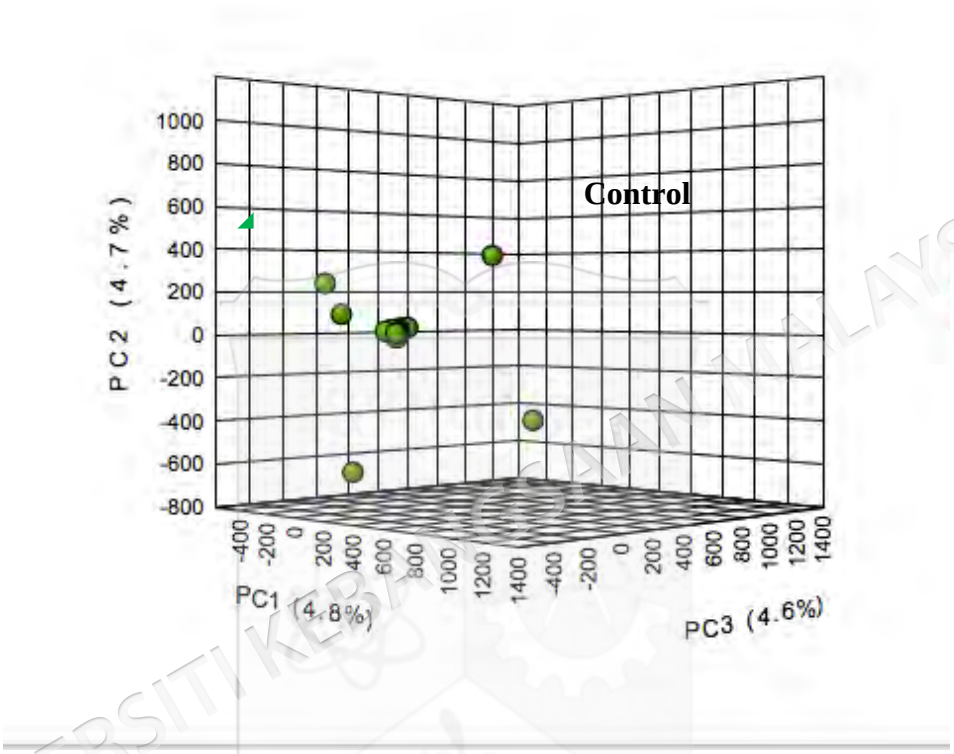
PRINCIPAL COMPONENT ANALYSIS (PCA) OF HUMAN PROTEINS IDENTIFIED IN BOTH CRC AND CONTROL



The PCA plot of human proteins commonly found in colorectal cancer (CRC) and control. The clustering of both samples were less distinguished.

APPENDIX J

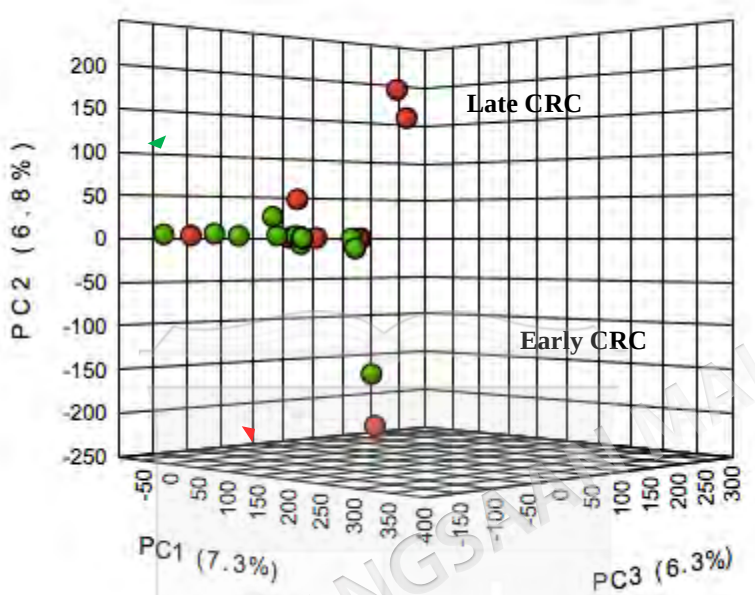
PRINCIPAL COMPONENT ANALYSIS (PCA) OF MICROBIAL PROTEINS IN CRC AND CONTROL



The PCA plot of microbial proteins found in colorectal cancer (CRC) and control. However, in above plot only the controls were clearly shown.

APPENDIX K

PRINCIPAL COMPONENT ANALYSIS (PCA) OF CRC EXCLUSIVE MICROBIAL PROTEINS



The PCA plot of microbial proteins found in colorectal cancer (CRC). However, in above plot both group were not distinctively separated.

APPENDIX L

LIST OF PUBLICATION

Wiley Online Library

Search

Login / Register

Proteomics

Review

Analysing the Secretome of Gut Microbiota as the Next Strategy For Early Detection of Colorectal Cancer

Putri-Intan-Hafizah Megat Mohd Azlan, Siok-Fong Chin, Teck Yew Low, Neoh Hui-min, Rahman Jamal

First published: 17 December 2018 | https://doi.org/10.1002/pmic.201800176

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as https://doi.org/10.1002/pmic.201800176

PDF TOOLS SHARE

Abstract

Dysbiosis of gut microbiome can contribute to inflammation, and subsequently initiation and progression of colorectal cancer (CRC). Throughout these stages, various proteins and metabolites are secreted to the external environment by microorganisms or the hosts themselves. Studying these proteins may help enhance our understanding of the host-microorganism relationship or they may even serve as useful biomarkers for CRC. However, secretomic studies of gut microbiome of CRC patients, until now, are scarcely performed. In this viewpoint article, we focus on the roles of gut microbiome in CRC, highlight the current findings on CRC secretome and address the emerging challenges and strategies to drive forward this area of research.

This article is protected by copyright. All rights reserved

Accepted Articles

Accepted, unedited articles published online and citable. The final edited and typeset version of record will appear in the future.

1800176

Related Information

Metrics

Details

This article is protected by copyright. All rights reserved

Keywords

colorectal cancer gut microbiome mass spectrometry metaproteomics secretome

Publication History

Accepted manuscript online: 17 December 2018

Manuscript accepted: 14 December 2018

Source: Megat Mohd Azlan et al. 2018

Chin *et al. Gut Pathog (2018) 10:29*
https://doi.org/10.1186/s13099-018-0258-5

Gut Pathogens

LETTER TO THE EDITOR

Open Access

Identification of *Schizosaccharomyces pombe* in the guts of healthy individuals and patients with colorectal cancer: preliminary evidence from a gut microbiome secretome study

Siok-Fong Chin[1*], Putri Intan Hafizah Megat Mohd Azlan[1], Luqman Mazlan[2] and Hui-min Neoh[1]

Abstract

Over the years, genetic profiling of the gut microbiome of patients with colorectal cancer (CRC) using genome sequencing has suggested over-representation of several bacterial taxa. However, little is known about the protein or metabolite secretions from the microbiota that could lead to CRC pathology. Proteomic studies on the role of microbial secretome in CRC are relatively rare. Here, we report the identification of proteins from *Schizosaccharomyces pombe* found in the stool samples of both healthy individuals and patients with CRC. We found that distinctive sets of *S. pombe* proteins were present exclusively and in high intensities in each group. Our finding may trigger a new interest in the role of gut mycobiota in carcinogenesis.

Keywords: *Schizosaccharomyces pombe*, Human gut, Microbiome, Secretome, Colorectal cancer

Gut microbiome and colorectal cancer

Dysbiosis of the gut microbiome has been postulated to be the causative event in colorectal cancer (CRC). For many years, most microbiome studies focused primarily on genetic sequencing and the subsequent identification of gut bacteria; however, the role of bacteria in CRC carcinogenesis remains unclear. We hypothesised that protein secretions from these microorganisms play a role in the pathophysiology of CRC. We aimed to identify and compare the secreted proteins released from the human gut microflora by assessing the stool samples of both patients with CRC and healthy control individuals.

*Correspondence: chinsiokfong@ppukm.ukm.edu.my
[1] UKM Medical Molecular Biology Institute, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia
Full list of author information is available at the end of the article

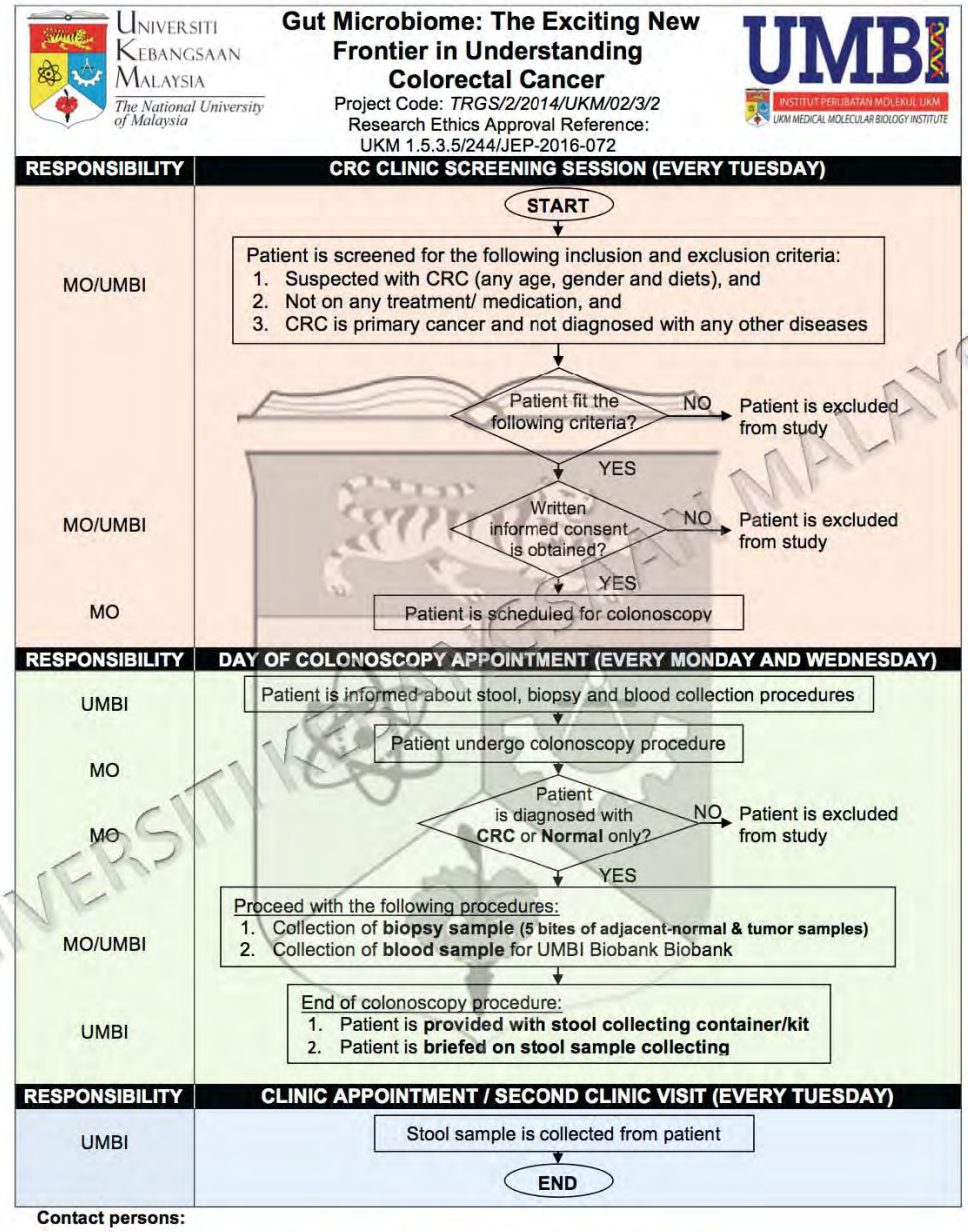
Faecal sample extraction and protein identification

A pilot study was performed at the endoscopy unit and surgery and medical clinics at the UKM Medical Centre, Kuala Lumpur, from 2016 to 2017. Stool samples from 26 patients with clinically diagnosed CRC and 20 non-CRC control individuals were collected from the UKM Medical Centre. The samples were homogenised and filtered, followed by precipitation with acetone, reduction with DL-Dithiothreitol, alkylation with iodoacetamide and digestion with trypsin. The resulting peptides were analysed by reverse-phase LC–ESI–MS/MS using Eksigent NanoLC system connected to a quadrupole time-of-flight 5600+ TripleTOF™ mass spectrometer (AB Sciex, USA). Briefly, mass spectra were acquired in the data-dependent mode (250 ms accumulation time per spectrum and mass range of 300–1250 *m/z*) to obtain MS/MS spectra for the most abundant parent ions following each survey MS1 scan. For each mass spectrum, a maximum of 50 precursors with a charge state between 2 and 4 were selected for fragmentation. The tandem mass spectrum was acquired in the high-sensitivity mode, in which the

Source: Chin et al. 2018

APPENDIX M

SAMPLE COLLECTION WORKFLOW AT ENDOCOPY UNIT, UKM MEDICAL CENTRE CHERAS (UKMMC)



Contact persons:

Muhammad Afiq Osman
UMBI Master Student
H/P: 012-4955915

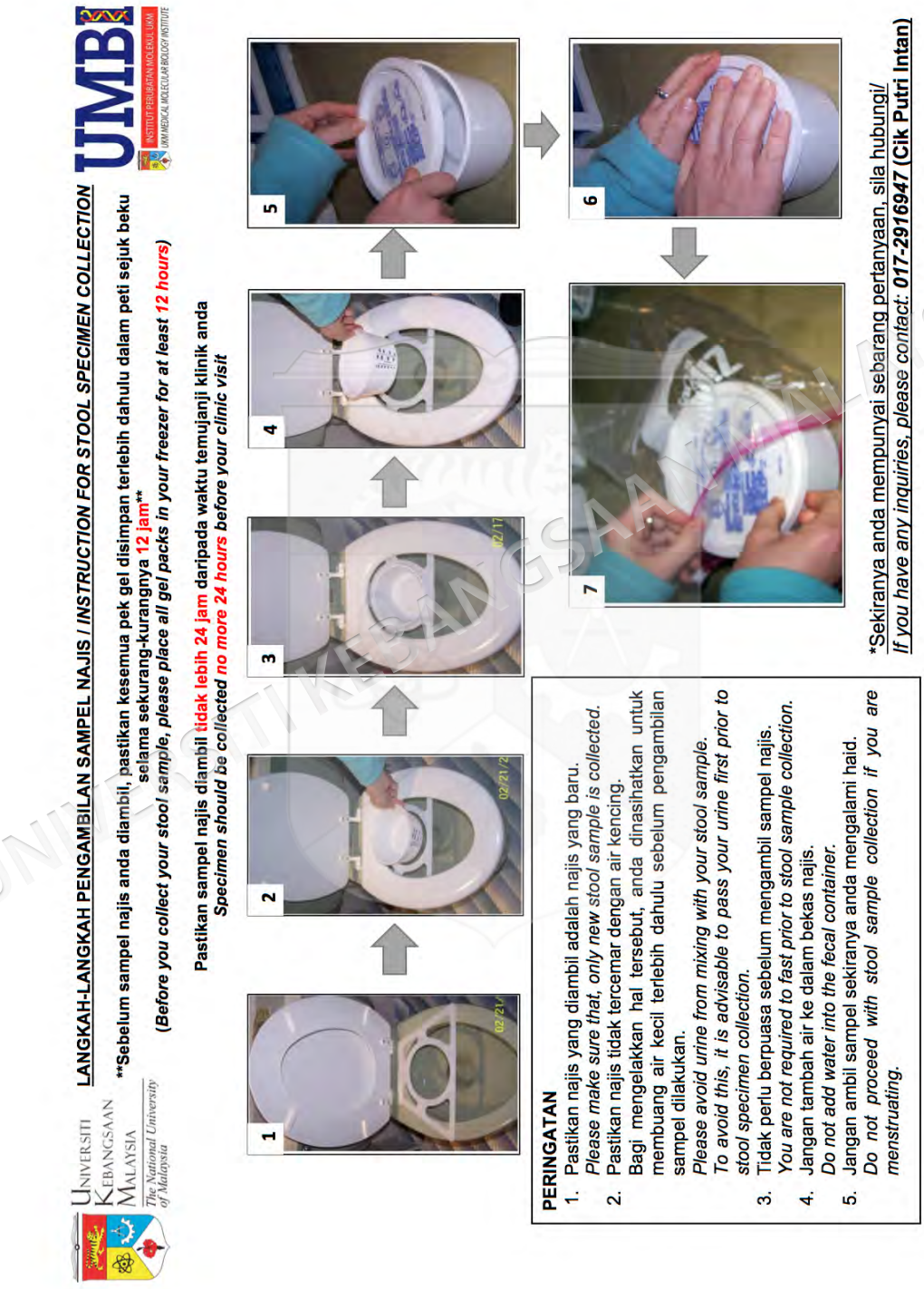
Putri Intan Hafizah
UMBI Master Student
H/P: 017-2916947

APPENDIX N

THE TOP VIEW OF FOAM COOLER BOX GIVEN TO PARTICIPANTS FOR STOOL SAMPLE COLLECTION

APPENDIX O

STOOL COLLECTION AND HANDLING GUIDELINES



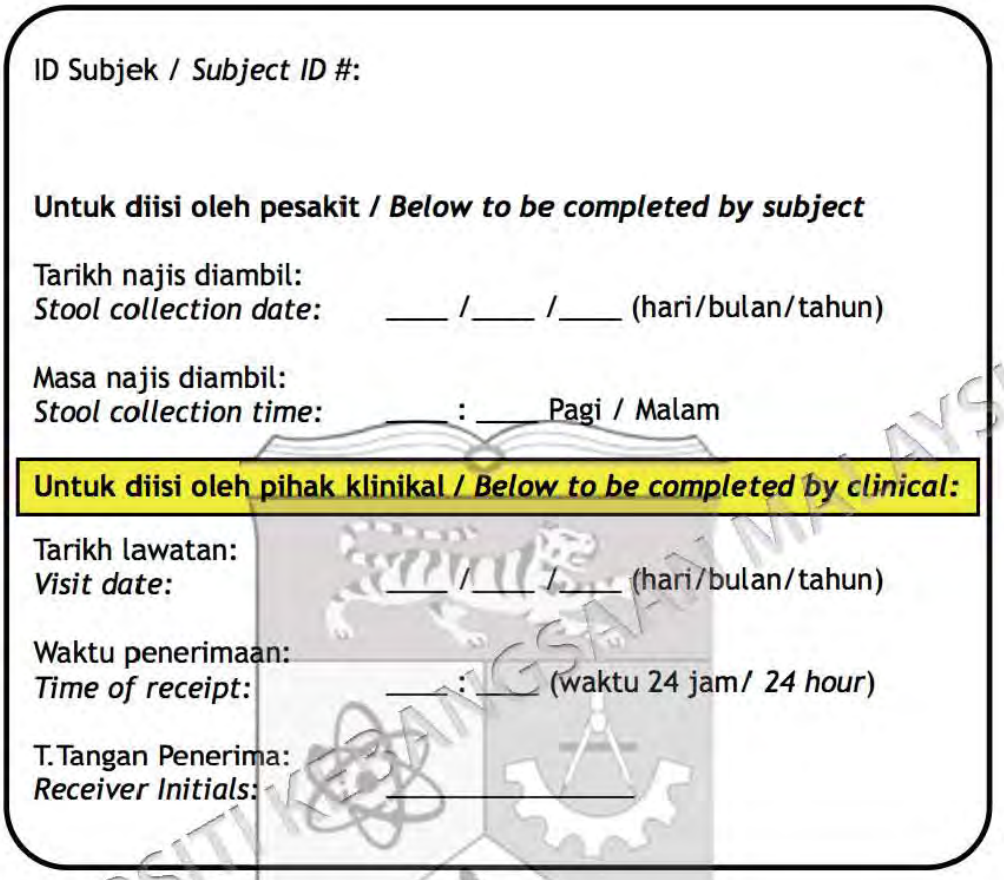
STOOL COLLECTION AND HANDLING GUIDELINES (PAGE 2)

LANGKAH MEMBUNGKUS SAMPEL NAJIS / *STOOL SAMPLE PACKAGING INSTRUCTIONS*

1. Simpan kesemua pek gel dalam peti sejuk beku (seperti Rajah 1). Jangan mula mengambil sampel najis selagi kesemua pek gel tidak disejukkan terlebih dahulu, sekurang-kurangnya selama 12 jam. *Place all gel packs in your freezer (see Figure 1). Do not collect your specimen until gel packs have been in your freezer for at least 12 hours.*	**Rajah 1/ *Figure 1*** 
2. Keluarkan pek gel daripada peti sejuk beku dan letakkan satu (1) atau dua (2) pek gel di bahagian dasar kotak *Styrofoam* (seperti Rajah 2). *Remove the gel packs from your freezer and place one (1) or two (2) of the gel packs in the bottom of the Styrofoam box (see Figure 2).*	**Rajah 2/ *Figure 2*** 
3. Tutup dengan kemas beg *Ziplock* yang mengandungi bekas sampel najis. Kemudian, letakkan beg tersebut di atas dua (2) pek gel tadi (seperti Rajah 3). *Place the sealed Ziplock bag containing your stool specimen on top of the two (2) frozen gel packs in the Styrofoam container (see Figure 3).*	**Rajah 3/ *Figure 3*** 
4. Letakkan empat (4) pek gel yang lain di sekeliling bekas sampel najis (seperti Rajah 4). *Place four (4) of the frozen gel packs around the specimen container (see Figure 4).*	**Rajah 4/ *Figure 4*** 
5. Letakkan satu (1) pek gel di atas bekas sampel najis (seperti Rajah 5). *Place one (1) gel pack on top of the container (see Figure 5).*	**Rajah 5/ *Figure 5*** 
6. Tutup bekas *Styrofoam* dengan penutup yang disediakan (seperti Rajah 2) dan bawa bersama bekas *Styrofoam* tersebut semasa temujanji klinik anda. *Place the Styrofoam lid on the Styrofoam container (see Figure 6) and bring it along during your clinic appointment.*	 **Rajah 6/ *Figure 6***
7. Sila tuliskan maklumat seperti; tarikh dan masa sampel najis diambil pada label yang telah disediakan. *Please note details such as; date and time of stool sample collection on label provided.*	ID Subjek / *Subject ID #:* **Untuk diisi oleh pesakit / *Below to be completed by subject*** Tarikh najis diambil: *Stool collection date:* ___/___/___ (hari/bulan/tahun) Masa najis diambil: *Stool collection time:* ___:___ Pagi / Malam

APPENDIX P

STICKER TEMPLATE PROVIDED ON FOAM COOLER BOX


ID Subjek / *Subject ID #*:

Untuk diisi oleh pesakit / *Below to be completed by subject*

Tarikh najis diambil:
Stool collection date: ____ /____ /____ (hari/bulan/tahun)

Masa najis diambil:
Stool collection time: ____ : ____ Pagi / Malam

Untuk diisi oleh pihak klinikal / *Below to be completed by clinical:*

Tarikh lawatan:
Visit date: ____ /____ /____ (hari/bulan/tahun)

Waktu penerimaan:
Time of receipt: ____ : ____ (waktu 24 jam/ *24 hour*)

T.Tangan Penerima:
Receiver Initials: _____

APPENDIX Q

PREPARATION OF BUFFERS FOR PROTEIN EXTRACTION

Buffer preparation of 40 mM Tris-HCL, pH 7.4 in 100 ml

Components	Weight or Volume
Tris base	0.488 g
Deionized H_2O (dH_2O)	80 ml
Adjust pH to 7.4 with HCL	-
dH_2O	to 100 ml

Preparation of 30% Dithiothreitol in 5 ml

Components	Weight or Volume
DL-Dithiothreitol	1.5 g
40 mM Tris-HCL, pH 7.4	5 ml

Preparation of 40mM Tris-HCL, pH 7.4 with 3% Dithiothreitol in 10 ml

Components	Volume
40 mM Tris-HCL, pH 7.4	9 ml
30% DL-Dithiothreitol	1 ml

APPENDIX R

PREPARATION OF BUFFERS FOR IN-SOLUTION DIGESTION

Preparation of 10 mM DL-Dithiothreitol in 0.4 mL

Components	Weight or Volume
DL-Dithiothreitol	0.000616 g
25mM Ammonium bicarbonate, pH 7.5	0.4 mL

Preparation of 55 mM Iodoacetamide in 0.2 mL

Components	Weight or Volume
Iodoacetamide	0.00203 g
25mM Ammonium bicarbonate, pH 7.5	0.2 mL

Preparation of 25 mM Ammonium bicarbonate, pH 7.5 in 100 mL

Components	Weight or Volume
Ammonium bicarbonate	0.1977 g
Deionized H_2O (dH_2O)	95 mL
Adjust pH to 7.5	-
dH_2O	to 100 mL

Preparation of 0.3% Formic acid in 100 mL

Components	Weight or Volume
Formic acid	0.2 mL
Deionized H_2O (dH_2O)	99.8 mL

APPENDIX S

EXCLUSIVE HUMAN PROTEIN EXPRESSIONS BASED ON CRC STAGING

Human proteins exclusively found in Stage I colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A8K979	ERI1 exoribonuclease 2	ERI2
2	O95837	Guanine nucleotide-binding protein subunit alpha-14	GNA14
3	P0C2W7	Cancer/testis antigen 47B	CT47B1
4	P28066	Proteasome subunit alpha type-5	PSMA5
5	P29317	Ephrin type-A receptor 2	EPHA2
6	P35749	Myosin-11	MYH11
7	P51449	Nuclear receptor ROR-gamma	RORC
8	Q13449	Limbic system-associated membrane protein	LSAMP
9	Q14207	Protein NPAT	NPAT
10	Q502W6	von Willebrand factor A domain-containing protein 3B	VWA3B
11	Q5SW79	Centrosomal protein of 170 kDa	CEP170
12	Q6UWR7	Ectonucleotide pyrophosphatase/phosphodiesterase family member 6	ENPP6
13	Q92824	Proprotein convertase subtilisin/kexin type 5	PCSK5
14	Q96JB6	Lysyl oxidase homolog 4	LOXL4
15	Q9UHD2	Serine/threonine-protein kinase TBK1	TBK1
16	Q9UPZ6	Thrombospondin type-1 domain-containing protein 7A	THSD7A

Human proteins exclusively discovered in Stage II colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A0A075B6K0	Immunoglobulin lambda variable 3-16	IGLV3-16
2	A0A075B6K6	Immunoglobulin lambda variable 4-3	IGLV4-3
3	A1A4G5	Leukemia NUP98 fusion partner 1	LNP1
4	A6NC57	Ankyrin repeat domain-containing protein 62	ANKRD62
5	A6NKF7	Transmembrane protein 88B	TMEM88B
6	A7E2Y1	Myosin-7B	MYH7B
7	A8K2U0	Alpha-2-macroglobulin-like protein 1	A2ML1
8	B2RXH4	BTB/POZ domain-containing protein 18	BTBD18
9	P20930	Filaggrin	FLG
10	O00522	Krev interaction trapped protein 1	KRIT1
11	O00566	U3 small nucleolar ribonucleoprotein protein MPP10	MPHOSPH10
12	O14607	Histone demethylase UTY	UTY
13	O14717	tRNA (cytosine(38)-C(5))-methyltransferase	TRDMT1
14	O14978	Zinc finger protein 263	ZNF263
15	O15016	Tripartite motif-containing protein 66	TRIM66
16	O15119	T-box transcription factor TBX3	TBX3
17	O15540	Fatty acid-binding protein, brain	FABP7
18	O42043	Endogenous retrovirus group K member 18 Env polyprotein;Surface protein;Transmembrane protein	ERVK-18
19	O43314	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase 2	PPIP5K2
20	O43482	Protein Mis18-beta	OIP5
21	O43866	CD5 antigen-like	CD5L
22	P37275;O60315	Zinc finger E-box-binding homeobox 1;Zinc finger E-box-binding homeobox 2	ZEB1;ZEB2
23	O60462	Neuropilin-2	NRP2
24	O60486	Plexin-C1	PLXNC1
25	O60506	Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP
26	O60518	Ran-binding protein 6	RANBP6
27	O60551	Glycylpeptide N-tetradecanoyltransferase 2	NMT2
28	O60884	DnaJ homolog subfamily A member 2	DNAJA2
29	O75027	ATP-binding cassette sub-family B member 7, mitochondrial	ABCB7
30	O75056	Syndecan-3	SDC3
31	O75145	Liprin-alpha-3	PPFIA3
32	O75335	Liprin-alpha-4	PPFIA4
33	Q9UN37;O75351	Vacuolar protein sorting-associated protein 4A;Vacuolar protein sorting-associated protein 4B	VPS4A;VPS4B
34	O75369	Filamin-B	FLNB
35	O75487	Glypican-4;Secreted glypican-4	GPC4
36	O75592	E3 ubiquitin-protein ligase MYCBP2	MYCBP2
37	O75638	Cancer/testis antigen 2	CTAG2
38	O75694	Nuclear pore complex protein Nup155	NUP155
39	O75718	Cartilage-associated protein	CRTAP
40	O76015	Keratin, type I cuticular Ha8	KRT38
41	O94768	Serine/threonine-protein kinase 17B	STK17B
42	O94906	Pre-mRNA-processing factor 6	PRPF6

to be continued…

…continuation

43	O95072	Meiotic recombination protein REC8 homolog	REC8
44	O95714	E3 ubiquitin-protein ligase HERC2	HERC2
45	P01127	Platelet-derived growth factor subunit B	PDGFB
46	P01137	Transforming growth factor beta-1;Latency-associated peptide	TGFB1
47	P02679	Fibrinogen gamma chain	FGG
48	P02765	Alpha-2-HS-glycoprotein;Alpha-2-HS-glycoprotein chain A;Alpha-2-HS-glycoprotein chain B	AHSG
49	P02788	Lactotransferrin;Lactoferricin-H;Kaliocin-1;Lactoferroxin-A;Lactoferroxin-B;Lactoferroxin-C	LTF
50	P04746;P19961; P04745	Pancreatic alpha-amylase;Alpha-amylase 2B;Alpha-amylase 1	AMY2A;AMY2B; AMY1A
51	P06401	Progesterone receptor	PGR
52	P07099	Epoxide hydrolase 1	EPHX1
53	P0CG22	Putative dehydrogenase/reductase SDR family member 4-like 1	DHRS4L1
54	P0CG32	Putative zinc finger CCHC domain-containing protein 18	ZCCHC18
55	P0CG42	Putative protein FAM157B	FAM157B
56	P0DM35	Metallothionein 1H-like protein 1	MT1HL1
57	P10909	Clusterin;Clusterin beta chain;Clusterin alpha chain	CLU
58	P11274	Breakpoint cluster region protein	BCR
59	P13727	Bone marrow proteoglycan;Eosinophil granule major basic protein	PRG2
60	P14410	Sucrase-isomaltase, intestinal;Sucrase;Isomaltase	SI
61	P17023	Zinc finger protein 19	ZNF19
62	P17707	S-adenosylmethionine decarboxylase proenzyme;S-adenosylmethionine decarboxylase alpha chain;S-adenosylmethionine decarboxylase beta chain	AMD1
63	P18074	TFIIH basal transcription factor complex helicase XPD subunit	ERCC2
64	P19320	Vascular cell adhesion protein 1	VCAM1
65	P21673	Diamine acetyltransferase 1	SAT1
66	P26639	Threonine--tRNA ligase, cytoplasmic	TARS
67	P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	NDUFS1
68	P28749	Retinoblastoma-like protein 1	RBL1
69	P30530	Tyrosine-protein kinase receptor UFO	AXL
70	P30989	Neurotensin receptor type 1	NTSR1
71	P35080	Profilin-2	PFN2
72	P36543	V-type proton ATPase subunit E 1	ATP6V1E1
73	Q13948;P39880	Protein CASP;Homeobox protein cut-like 1	CUX1
74	P42081	T-lymphocyte activation antigen CD86	CD86
75	P46939	Utrophin	UTRN
76	P48775	Tryptophan 2,3-dioxygenase	TDO2
77	P51674	Neuronal membrane glycoprotein M6-a	GPM6A
78	P51816	AF4/FMR2 family member 2	AFF2
79	P51911	Calponin-1	CNN1
80	P52823	Stanniocalcin-1	STC1
81	P53677	AP-3 complex subunit mu-2	AP3M2

to be continued…

…continuation

82	P51688	N-sulphoglucosamine sulphohydrolase	SGSH
83	P51530	DNA replication ATP-dependent helicase/nuclease DNA2;DNA replication nuclease DNA2;DNA replication ATP-dependent helicase DNA2	DNA2
84	P49327	Fatty acid synthase;[Acyl-carrier-protein] S-acetyltransferase;[Acyl-carrier-protein] S-malonyltransferase;3-oxoacyl-[acyl-carrier-protein] synthase;3-oxoacyl-[acyl-carrier-protein] reductase;3-hydroxyacyl-[acyl-carrier-protein] dehydratase;Enoyl-[acyl-carrier-protein] reductase;Oleoyl-[acyl-carrier-protein] hydrolase	FASN
85	P51610	Host cell factor 1;HCF N-terminal chain 1;HCF N-terminal chain 2;HCF N-terminal chain 3;HCF N-terminal chain 4;HCF N-terminal chain 5;HCF N-terminal chain 6;HCF C-terminal chain 1;HCF C-terminal chain 2;HCF C-terminal chain 3;HCF C-terminal chain 4;HCF C-terminal chain 5;HCF C-terminal chain 6	HCFC1
86	P53992	Protein transport protein Sec24C	SEC24C
87	P55011	Solute carrier family 12 member 2	SLC12A2
88	P61024	Cyclin-dependent kinases regulatory subunit 1	CKS1B
89	P61371	Insulin gene enhancer protein ISL-1	ISL1
90	P61619	Protein transport protein Sec61 subunit alpha isoform 1	SEC61A1
91	P78347	General transcription factor II-I	GTF2I
92	P98198	Phospholipid-transporting ATPase ID	ATP8B2
93	Q00534	Cyclin-dependent kinase 6	CDK6
94	Q01484	Ankyrin-2	ANK2
95	Q02040	A-kinase anchor protein 17A	AKAP17A
96	Q03468	DNA excision repair protein ERCC-6	ERCC6
97	Q05195	Max dimerization protein 1	MXD1
98	Q05D32	CTD small phosphatase-like protein 2	CTDSPL2
99	Q08188	Protein-glutamine gamma-glutamyltransferase E;Protein-glutamine gamma-glutamyltransferase E 50 kDa catalytic chain;Protein-glutamine gamma-glutamyltransferase E 27 kDa non-catalytic chain	TGM3
100	Q08AM6	Protein VAC14 homolog	VAC14
101	Q12830	Nucleosome-remodeling factor subunit BPTF	BPTF
102	Q12866	Tyrosine-protein kinase Mer	MERTK
103	Q12887	Protoheme IX farnesyltransferase, mitochondrial	COX10
104	Q12988	Heat shock protein beta-3	HSPB3
105	Q13153	Serine/threonine-protein kinase PAK 1	PAK1
106	Q13395	Probable methyltransferase TARBP1	TARBP1
107	Q13501	Sequestosome-1	SQSTM1
108	Q13591	Semaphorin-5A	SEMA5A
109	Q13642	Four and a half LIM domains protein 1	FHL1
110	Q13796	Protein Shroom2	SHROOM2
111	Q14004	Cyclin-dependent kinase 13	CDK13
112	Q14135	Transcription cofactor vestigial-like protein 4	VGLL4
113	Q14185	Dedicator of cytokinesis protein 1	DOCK1
114	Q14393	Growth arrest-specific protein 6	GAS6

to be continued…

…continuation

115	Q14409	Putative glycerol kinase 3	GK3P
116	Q14520	Hyaluronan-binding protein 2;Hyaluronan-binding protein 2 50 kDa heavy chain;Hyaluronan-binding protein 2 50 kDa heavy chain alternate form;Hyaluronan-binding protein 2 27 kDa light chain;Hyaluronan-binding protein 2 27 kDa light chain alternate form	HABP2
117	Q14573	Inositol 1,4,5-trisphosphate receptor type 3	ITPR3
118	Q14677	Clathrin interactor 1	CLINT1
119	Q14773	Intercellular adhesion molecule 4	ICAM4
120	Q14974	Importin subunit beta-1	KPNB1
121	Q14D04	Ventricular zone-expressed PH domain-containing protein homolog 1	VEPH1
122	Q15165	Serum paraoxonase/arylesterase 2	PON2
123	Q15697	Zinc finger protein 174	ZNF174
124	Q16627	C-C motif chemokine 14;HCC-1(3-74);HCC-1(4-74);HCC-1(9-74)	CCL14
125	Q16659	Mitogen-activated protein kinase 6	MAPK6
126	Q2TBE0	CWF19-like protein 2	CWF19L2
127	Q3KR16	Pleckstrin homology domain-containing family G member 6	PLEKHG6
128	Q3SXM5	Inactive hydroxysteroid dehydrogenase-like protein 1	HSDL1
129	Q49AM3	Tetratricopeptide repeat protein 31	TTC31
130	Q52LW3	Rho GTPase-activating protein 29	ARHGAP29
131	Q53G44	Interferon-induced protein 44-like	IFI44L
132	Q53HI1	Protein unc-50 homolog	UNC50
133	Q53HL2	Borealin	CDCA8
134	Q58EX7	Puratrophin-1	PLEKHG4
135	Q5CZC0	Fibrous sheath-interacting protein 2	FSIP2
136	Q5FWF7	F-box only protein 48	FBXO48
137	Q5HYM0	Probable ribonuclease ZC3H12B	ZC3H12B
138	Q5JRX3	Presequence protease, mitochondrial	PITRM1
139	Q5JUK3	Potassium channel subfamily T member 1	KCNT1
140	Q5JVG8	Zinc finger protein 506	ZNF506
141	Q5M9Q1	NKAP-like protein	NKAPL
142	Q5MJ10	Sperm protein associated with the nucleus on the X chromosome N2	SPANXN2
143	Q5R3I4	Tetratricopeptide repeat protein 38	TTC38
144	Q5S007	Leucine-rich repeat serine/threonine-protein kinase 2	LRRK2
145	Q5T124	UBX domain-containing protein 11	UBXN11
146	Q5T3U5	Multidrug resistance-associated protein 7	ABCC10
147	Q5T4S7	E3 ubiquitin-protein ligase UBR4	UBR4
148	Q5TCS8	Adenylate kinase 9	AK9
149	Q5TF21	Protein SOGA3	SOGA3
150	Q5THJ4	Vacuolar protein sorting-associated protein 13D	VPS13D
151	Q5VTH2	Protein Flattop	CFAP126
152	Q5VVS0	Putative uncharacterized protein C1orf140	C1orf140
153	Q5VWK0	Neuroblastoma breakpoint family member 6	NBPF6

to be continued…

…continuation

154	Q5VYM1	Uncharacterized protein C9orf131	C9orf131
155	Q68BL7	Olfactomedin-like protein 2A	OLFML2A
156	Q68DK7	Male-specific lethal 1 homolog	MSL1
157	Q6IED9	Putative diacylglycerol O-acyltransferase 2-like protein DGAT2L7P	DGAT2L7P
158	Q6IPM2	IQ domain-containing protein E	IQCE
159	Q6P047	Uncharacterized protein C8orf74	C8orf74
160	Q6P1M9	Armadillo repeat-containing X-linked protein 5	ARMCX5
161	Q6P1S2	Protein C3orf33	C3orf33
162	Q6P4F7	Rho GTPase-activating protein 11A	ARHGAP11A
163	Q6PIS1	Solute carrier family 23 member 3	SLC23A3
164	Q6T311	ADP-ribosylation factor-like protein 9	ARL9
165	Q6TDP4	Kelch-like protein 17	KLHL17
166	Q6TFL4	Kelch-like protein 24	KLHL24
167	Q6UX71	Plexin domain-containing protein 2	PLXDC2
168	Q6XZF7	Dynamin-binding protein	DNMBP
169	Q6ZMK1	Cysteine and histidine-rich protein 1	CYHR1
170	Q6ZMR5	Transmembrane protease serine 11A	TMPRSS11A
171	Q6ZMV9	Kinesin-like protein KIF6	KIF6
172	Q6ZN08	Putative zinc finger protein 66	ZNF66
173	Q6ZSZ6	Teashirt homolog 1	TSHZ1
174	Q6ZTR5	Uncharacterized protein CXorf22	CXorf22
175	Q702N8	Xin actin-binding repeat-containing protein 1	XIRP1
176	Q70Z44	5-hydroxytryptamine receptor 3D	HTR3D
177	Q76N89	E3 ubiquitin-protein ligase HECW1	HECW1
178	Q76NI1	Protein very KIND	KNDC1
179	Q7KZ85	Transcription elongation factor SPT6	SUPT6H
180	Q7RTR0	NACHT, LRR and PYD domains-containing protein 9	NLRP9
181	Q7Z3B3	KAT8 regulatory NSL complex subunit 1	KANSL1
182	Q7Z410	Transmembrane protease serine 9;Serase-1;Serase-2;Serase-3	TMPRSS9
183	Q7Z4Q2	HEAT repeat-containing protein 3	HEATR3
184	Q7Z5H5	Vomeronasal type-1 receptor 4	VN1R4
185	Q7Z5V6	Protein phosphatase 1 regulatory subunit 32	PPP1R32
186	Q86TE4	Leucine zipper protein 2	LUZP2
187	Q86UE8	Serine/threonine-protein kinase tousled-like 2	TLK2
188	Q86UK7	Zinc finger protein 598	ZNF598
189	Q86VH2	Kinesin-like protein KIF27	KIF27
190	Q86WA9	Sodium-independent sulfate anion transporter	SLC26A11
191	Q86WS4	Uncharacterized protein C12orf40	C12orf40
192	Q8IV36	Protein HID1	HID1
193	Q8IW41	MAP kinase-activated protein kinase 5	MAPKAPK5
194	Q8IWD5	Major facilitator superfamily domain-containing protein 6-like	MFSD6L
195	Q8IWX7	Protein unc-45 homolog B	UNC45B
196	Q8IY82	Dynein regulatory complex subunit 7	DRC7
197	Q8IZF0	Sodium leak channel non-selective protein	NALCN
198	Q8N4F7	RING finger protein 175	RNF175

to be continued…

…continuation

199	Q8N394	Transmembrane and TPR repeat-containing protein 2	TMTC2
200	Q8N4M1	Choline transporter-like protein 3	SLC44A3
201	Q8N4S0	Coiled-coil domain-containing protein 82	CCDC82
202	Q8N5U6	RING finger protein 10	RNF10
203	Q8N6G6	ADAMTS-like protein 1	ADAMTSL1
204	Q8N8X9	Protein mab-21-like 3	MAB21L3
205	Q8N9H8	Exonuclease mut-7 homolog	EXD3
206	Q8N9N8	Probable RNA-binding protein EIF1AD	EIF1AD
207	Q8NCN2	Zinc finger and BTB domain-containing protein 34	ZBTB34
208	Q8NEN9	PDZ domain-containing protein 8	PDZD8
209	Q8NEV4	Myosin-IIIa	MYO3A
210	Q8NGU9	Probable G-protein coupled receptor 150	GPR150
211	Q8NH94	Olfactory receptor 1L1	OR1L1
212	Q8NHP7	Exonuclease 3-5 domain-containing protein 1	EXD1
213	Q8NHY3	GAS2-like protein 2	GAS2L2
214	Q8TAD8	Smad nuclear-interacting protein 1	SNIP1
215	Q8TAG9	Exocyst complex component 6	EXOC6
216	Q8TBB5	Kelch domain-containing protein 4	KLHDC4
217	Q8TBG4	Ethanolamine-phosphate phospho-lyase	ETNPPL
218	Q8TD57	Dynein heavy chain 3, axonemal	DNAH3
219	Q8TDI8	Transmembrane channel-like protein 1	TMC1
220	Q8TF50	Zinc finger protein 526	ZNF526
221	Q8WTQ1	Beta-defensin 104	DEFB104A
222	Q8WU20	Fibroblast growth factor receptor substrate 2	FRS2
223	Q8WVM0	Dimethyladenosine transferase 1, mitochondrial	TFB1M
224	Q8WWI1	LIM domain only protein 7	LMO7
225	Q8WWW0	Ras association domain-containing protein 5	RASSF5
226	Q8WXK1	Ankyrin repeat and SOCS box protein 15	ASB15
227	Q92541	RNA polymerase-associated protein RTF1 homolog	RTF1
228	Q92616	Translational activator GCN1	GCN1L1
229	Q92730	Rho-related GTP-binding protein Rho6	RND1
230	Q969X0	RILP-like protein 2	RILPL2
231	Q96A23	Copine-4	CPNE4
232	Q96A28	SLAM family member 9	SLAMF9
233	Q96AY4	Tetratricopeptide repeat protein 28	TTC28
234	Q96BU1	S100P-binding protein	S100PBP
235	Q96C03	Mitochondrial dynamics protein MID49	MIEF2
236	Q96G75	Protein RMD5 homolog B	RMND5B
237	Q96GM8	Target of EGR1 protein 1	TOE1
238	Q96HP0	Dedicator of cytokinesis protein 6	DOCK6
239	Q96JK2	DDB1- and CUL4-associated factor 5	DCAF5
240	Q96K80	Zinc finger CCCH domain-containing protein 10	ZC3H10
241	Q96KD3	Protein FAM71F1	FAM71F1
242	Q96KX2	F-actin-capping protein subunit alpha-3	CAPZA3
243	Q96LP6	Uncharacterized protein C12orf42	C12orf42
244	Q96LW2	Uncharacterized serine/threonine-protein kinase SgK494	SGK494
245	Q96M53	Protein TBATA	TBATA

to be continued…

…continuation

246	Q96M86	Dynein heavy chain domain-containing protein 1	DNHD1
247	Q96MC4	Protein DDC8 homolog	KIAA1731NL
248	Q96MR6	Cilia- and flagella-associated protein 57	CFAP57
249	Q96NG8	Zinc finger protein 582	ZNF582
250	Q96S66	Chloride channel CLIC-like protein 1	CLCC1
251	Q96SC8	Doublesex- and mab-3-related transcription factor A2	DMRTA2
252	Q99598	Translin-associated protein X	TSNAX
253	Q99685	Monoglyceride lipase	MGLL
254	Q99973	Telomerase protein component 1	TEP1
255	Q9BQC3	Diphthamide biosynthesis protein 2	DPH2
256	Q9BQI7	PH and SEC7 domain-containing protein 2	PSD2
257	Q9BQT9	Calsyntenin-3	CLSTN3
258	Q9BRZ2	E3 ubiquitin-protein ligase TRIM56	TRIM56
259	Q9BXF3	Cat eye syndrome critical region protein 2	CECR2
260	Q9BXT6	Putative helicase Mov10l1	MOV10L1
261	Q9BXU1	Serine/threonine-protein kinase 31	STK31
262	Q9BY64	UDP-glucuronosyltransferase 2B28	UGT2B28
263	Q9BZZ5	Apoptosis inhibitor 5	API5
264	Q9C0K1	Zinc transporter ZIP8	SLC39A8
265	Q9H0F7	ADP-ribosylation factor-like protein 6	ARL6
266	Q9H2B2	Synaptotagmin-4	SYT4
267	Q9H2U9	Disintegrin and metalloproteinase domain-containing protein 7	ADAM7
268	Q9H330	Transmembrane protein 245	TMEM245
269	Q9H3P7	Golgi resident protein GCP60	ACBD3
270	Q9H4B0	Probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial	OSGEPL1
271	Q9H4B7	Tubulin beta-1 chain	TUBB1
272	Q9H4F8	SPARC-related modular calcium-binding protein 1	SMOC1
273	Q9H4W6	Transcription factor COE3	EBF3
274	Q9H665	IGF-like family receptor 1	IGFLR1
275	Q9H6T3	RNA polymerase II-associated protein 3	RPAP3
276	Q9H987	Synaptopodin 2-like protein	SYNPO2L
277	Q9H8Y5	Ankyrin repeat and zinc finger domain-containing protein 1	ANKZF1
278	Q9H9G7	Protein argonaute-3	AGO3
279	Q9H9S4	Calcium-binding protein 39-like	CAB39L
280	Q9HAT8	E3 ubiquitin-protein ligase pellino homolog 2	PELI2
281	Q9HAY6	Beta,beta-carotene 15,15-monooxygenase	BCO1
282	Q9HB65	RNA polymerase II elongation factor ELL3	ELL3
283	Q9HC10	Otoferlin	OTOF
284	Q9NR99	Matrix-remodeling-associated protein 5	MXRA5
285	Q9NSV4	Protein diaphanous homolog 3	DIAPH3
286	Q9NTN9	Semaphorin-4G	SEMA4G
287	Q9NU02	Ankyrin repeat and EF-hand domain-containing protein 1	ANKEF1
288	Q9NUN5	Probable lysosomal cobalamin transporter	LMBRD1
289	Q9NUZ1	Acyl-coenzyme A oxidase-like protein	ACOXL

to be continued…

…continuation

290	Q9NVC3	Putative sodium-coupled neutral amino acid transporter 7	SLC38A7
291	Q9NX20	39S ribosomal protein L16, mitochondrial	MRPL16
292	Q9NX58	Cell growth-regulating nucleolar protein	LYAR
293	Q9NYQ8	Protocadherin Fat 2	FAT2
294	Q9NZ63	Uncharacterized protein C9orf78	C9orf78
295	Q9NZH0	G-protein coupled receptor family C group 5 member B	GPRC5B
296	Q9NZR2	Low-density lipoprotein receptor-related protein 1B	LRP1B
297	Q9NZS9	Bifunctional apoptosis regulator	BFAR
298	Q9P107	GEM-interacting protein	GMIP
299	Q9P2D7	Dynein heavy chain 1, axonemal	DNAH1
300	Q9P2E2	Kinesin-like protein KIF17	KIF17
301	Q9P2H5	Ubiquitin carboxyl-terminal hydrolase 35	USP35
302	Q9P2K8	Eukaryotic translation initiation factor 2-alpha kinase 4	EIF2AK4
303	Q9P2P5	E3 ubiquitin-protein ligase HECW2	HECW2
304	Q9UBK2	Peroxisome proliferator-activated receptor gamma coactivator 1-alpha	PPARGC1A
305	Q9UBT7	Alpha-catulin	CTNNAL1
306	Q9UBU2	Dickkopf-related protein 2	DKK2
307	Q9UGH3	Solute carrier family 23 member 2	SLC23A2
308	Q9UGI6	Small conductance calcium-activated potassium channel protein 3	KCNN3
309	Q9UGT4	Sushi domain-containing protein 2	SUSD2
310	Q9UJA5	tRNA (adenine(58)-N(1))-methyltransferase non-catalytic subunit TRM6	TRMT6
311	Q9UKZ4	Teneurin-1;Ten-1 intracellular domain;Teneurin C-terminal-associated peptide	TENM1
312	Q9UL16	Cilia- and flagella-associated protein 45	CFAP45
313	Q9UL33	Trafficking protein particle complex subunit 2-like protein	TRAPPC2L
314	Q9UL62	Short transient receptor potential channel 5	TRPC5
315	Q9ULE0	Protein WWC3	WWC3
316	Q9ULF5	Zinc transporter ZIP10	SLC39A10
317	Q9ULI1	NACHT and WD repeat domain-containing protein 2	NWD2
318	Q9ULT0	Tetratricopeptide repeat protein 7A	TTC7A
319	Q9UNH5	Dual specificity protein phosphatase CDC14A	CDC14A
320	Q9UPE1	SRSF protein kinase 3	SRPK3
321	Q9UPS8	Ankyrin repeat domain-containing protein 26	ANKRD26
322	Q9UPY3	Endoribonuclease Dicer	DICER1
323	Q9UQE7	Structural maintenance of chromosomes protein 3	SMC3
324	Q9Y2D0	Carbonic anhydrase 5B, mitochondrial	CA5B
325	Q9Y2G9	Protein strawberry notch homolog 2	SBNO2
326	Q9Y2H6	Fibronectin type-III domain-containing protein 3A	FNDC3A
327	Q9Y2K2	Serine/threonine-protein kinase SIK3	SIK3
328	Q9Y2L8	Zinc finger protein with KRAB and SCAN domains 5	ZKSCAN5
329	Q9Y3R5	Protein dopey-2	DOPEY2

to be continued…

…continuation

330	Q9Y487	V-type proton ATPase 116 kDa subunit a isoform 2	ATP6V0A2
331	Q9Y5A6	Zinc finger and SCAN domain-containing protein 21	ZSCAN21
332	Q9Y5G1	Protocadherin gamma-B3	PCDHGB3
333	Q9Y5G4	Protocadherin gamma-A9	PCDHGA9
334	Q9Y5I3	Protocadherin alpha-1	PCDHA1
335	Q9Y5Z7	Host cell factor 2	HCFC2
336	Q9Y6I8	Peroxisomal membrane protein 4	PXMP4
337	Q9Y6X3	MAU2 chromatid cohesion factor homolog	MAU2
338	Q9Y6X4	Soluble lamin-associated protein of 75 kDa	FAM169A

Human proteins exclusively identified in Stage III colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A0A1B0GTD5	Testis-expressed protein 49	TEX49
2	A0AVT1	Ubiquitin-like modifier-activating enzyme 6	UBA6
3	A2A288	Probable ribonuclease ZC3H12D	ZC3H12D
4	A4D0S4	Laminin subunit beta-4	LAMB4
5	A4D2P6	Delphilin	GRID2IP
6	A4UGR9	Xin actin-binding repeat-containing protein 2	XIRP2
7	A6NCI4	von Willebrand factor A domain-containing protein 3A	VWA3A
8	A6NCM1	IQ and AAA domain-containing protein 1-like	IQCA1L
9	A6NE01	Protein FAM186A	FAM186A
10	A6NEN9	Uncharacterized protein CXorf65	CXorf65
11	A6NFA0	Putative family with sequence similarity 205, member C protein	FAM205CP
12	P0C7N4;A6NGB0	Transmembrane protein 191B;Transmembrane protein 191C	TMEM191B; TMEM191C
13	A6NGG8	Uncharacterized protein C2orf71	C2orf71
14	A6NIM6	Solute carrier family 15 member 5	SLC15A5
15	A6NK53	Zinc finger protein 233	ZNF233
16	A6NLI5	Putative tripartite motif-containing protein 64C	TRIM64C
17	O60309;A6NMS7;A6NM11	Leucine-rich repeat-containing protein 37A3;Leucine-rich repeat-containing protein 37A;Leucine-rich repeat-containing protein 37A2	LRRC37A3; LRRC37A;LRRC37A2
18	Q5VT03;Q8IVF1; B1AL46;A6NNL0	NUT family member 2D;NUT family member 2A;NUT family member 2E;NUT family member 2B	NUTM2D; NUTM2A; NUTM2E; NUTM2B
19	A8MT82	Putative centaurin-gamma-like family member 11P	CTGLF11P
20	A8MTA8	Protein FAM166B	FAM166B
21	A8MV72	Putative UPF0607 protein ENSP00000382826	
22	A8MW92	PHD finger protein 20-like protein 1	PHF20L1
23	A8MWE9	EF-hand calcium-binding domain-containing protein 8	EFCAB8
24	A8MWY0	UPF0577 protein KIAA1324-like	KIAA1324L
25	B1ANS9	WD repeat-containing protein 64	WDR64
26	Q12799;B9ZVM9	T-complex protein 10A homolog;Putative t-complex protein 10A homolog 2	TCP10;TCP10L2
27	O76011	Keratin, type I cuticular Ha4	KRT34
28	P60413	Keratin-associated protein 10-12	KRTAP10-12
29	E7EW31	Proline-rich basic protein 1	PROB1
30	O00189	AP-4 complex subunit mu-1	AP4M1
31	O00221	NF-kappa-B inhibitor epsilon	NFKBIE
32	O00305	Voltage-dependent L-type calcium channel subunit beta-4	CACNB4
33	O00534	von Willebrand factor A domain-containing protein 5A	VWA5A

to be continued…

…continuation

34	O14495	Lipid phosphate phosphohydrolase 3	PPAP2B
35	O00555	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	CACNA1A
36	O00468	Agrin;Agrin N-terminal 110 kDa subunit;Agrin C-terminal 110 kDa subunit;Agrin C-terminal 90 kDa fragment;Agrin C-terminal 22 kDa fragment	AGRN
37	O14776	Transcription elongation regulator 1	TCERG1
38	O14827	Ras-specific guanine nucleotide-releasing factor 2	RASGRF2
39	O14979	Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL
40	O15027	Protein transport protein Sec16A	SEC16A
41	O15050	TPR and ankyrin repeat-containing protein 1	TRANK1
42	O15164	Transcription intermediary factor 1-alpha	TRIM24
43	O43149	Zinc finger ZZ-type and EF-hand domain-containing protein 1	ZZEF1
44	O43173	Sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase	ST8SIA3
45	O43301	Heat shock 70 kDa protein 12A	HSPA12A
46	O43435	T-box transcription factor TBX1	TBX1
47	O43900	Prickle-like protein 3	PRICKLE3
48	O60240	Perilipin-1	PLIN1
49	O60502	Protein O-GlcNAcase	MGEA5
50	O60507	Protein-tyrosine sulfotransferase 1	TPST1
51	O60543	Cell death activator CIDE-A	CIDEA
52	O60841	Eukaryotic translation initiation factor 5B	EIF5B
53	O75064	DENN domain-containing protein 4B	DENND4B
54	O75095	Multiple epidermal growth factor-like domains protein 6	MEGF6
55	O75122	CLIP-associating protein 2	CLASP2
56	O75182	Paired amphipathic helix protein Sin3b	SIN3B
57	O75319	RNA/RNP complex-1-interacting phosphatase	DUSP11
58	O75427	Leucine-rich repeat and calponin homology domain-containing protein 4	LRCH4
59	O75443	Alpha-tectorin	TECTA
60	O75445	Usherin	USH2A
61	O75486	Transcription initiation protein SPT3 homolog	SUPT3H
62	O75521	Enoyl-CoA delta isomerase 2, mitochondrial	ECI2
63	O75752	UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1	B3GALNT1
64	O75844	CAAX prenyl protease 1 homolog	ZMPSTE24
65	O75882	Attractin	ATRN
66	O76003	Glutaredoxin-3	GLRX3
67	O94808	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 2	GFPT2
68	O94851	Protein-methionine sulfoxide oxidase MICAL2	MICAL2

to be continued…

…continuation

69	O94875	Sorbin and SH3 domain-containing protein 2	SORBS2
70	O94887	FERM, RhoGEF and pleckstrin domain-containing protein 2	FARP2
71	O94929	Actin-binding LIM protein 3	ABLIM3
72	O94966	Ubiquitin carboxyl-terminal hydrolase 19	USP19
73	O95235	Kinesin-like protein KIF20A	KIF20A
74	O95248	Myotubularin-related protein 5	SBF1
75	O95377	Gap junction beta-5 protein	GJB5
76	O95428	Papilin	PAPLN
77	O95521	PRAME family member 1	PRAMEF1
78	O95672	Endothelin-converting enzyme-like 1	ECEL1
79	O95755	Ras-related protein Rab-36	RAB36
80	P00439	Phenylalanine-4-hydroxylase	PAH
81	P00915	Carbonic anhydrase 1	CA1
82	P01023	Alpha-2-macroglobulin	A2M
83	P01834;P0DOX7	Ig kappa chain C region	IGKC
84	P02647	Apolipoprotein A-I;Proapolipoprotein A-I;Truncated apolipoprotein A-I	APOA1
85	P02671	Fibrinogen alpha chain;Fibrinopeptide A;Fibrinogen alpha chain	FGA
86	P02763	Alpha-1-acid glycoprotein 1	ORM1
87	P02787	Serotransferrin	TF
88	P06702	Protein S100-A9	S100A9
89	P40198;P06731	Carcinoembryonic antigen-related cell adhesion molecule 3;Carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM3; CEACAM5
90	P07947	Tyrosine-protein kinase Yes	YES1
91	P08185	Corticosteroid-binding globulin	SERPINA6
92	P08238	Heat shock protein HSP 90-beta	HSP90AB1
93	P08246	Neutrophil elastase	ELANE
94	P0C7I6	Coiled-coil domain-containing protein 159	CCDC159
95	P10145	Interleukin-8;MDNCF-a;Interleukin-8;IL-8(5-77);IL-8(6-77);IL-8(7-77);IL-8(8-77);IL-8(9-77)	CXCL8
96	P30512;P30459; P30447;P16190; P16189;P16188; P10314	HLA class I histocompatibility antigen, A-29 alpha chain;HLA class I histocompatibility antigen, A-74 alpha chain;HLA class I histocompatibility antigen, A-23 alpha chain;HLA class I histocompatibility antigen, A-33 alpha chain;HLA class I histocompatibility antigen, A-31 alpha chain;HLA class I histocompatibility antigen, A-30 alpha chain;HLA class I histocompatibility antigen, A-32 alpha chain	HLA-A
97	P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF
98	P11217	Glycogen phosphorylase, muscle form	PYGM
99	P13073	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1

to be continued…

…continuation

100	P12109	Collagen alpha-1(VI) chain	COL6A1
101	P13497	Bone morphogenetic protein 1	BMP1
102	P13640	Metallothionein-1G	MT1G
103	P13688	Carcinoembryonic antigen-related cell adhesion molecule 1	CEACAM1
104	P14210	Hepatocyte growth factor;Hepatocyte growth factor alpha chain;Hepatocyte growth factor beta chain	HGF
105	P14373	Zinc finger protein RFP	TRIM27
106	P14780	Matrix metalloproteinase-9;67 kDa matrix metalloproteinase-9;82 kDa matrix metalloproteinase-9	MMP9
107	P14866	Heterogeneous nuclear ribonucleoprotein L	HNRNPL
108	P15090	Fatty acid-binding protein, adipocyte	FABP4
109	P15260	Interferon gamma receptor 1	IFNGR1
110	P15882;P52757	N-chimaerin;Beta-chimaerin	CHN1;CHN2
111	P16144	Integrin beta-4	ITGB4
112	P16284	Platelet endothelial cell adhesion molecule	PECAM1
113	P16435	NADPH--cytochrome P450 reductase	POR
114	P16885	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2	PLCG2
115	P17025	Zinc finger protein 182	ZNF182
116	P18433	Receptor-type tyrosine-protein phosphatase alpha	PTPRA
117	P20042	Eukaryotic translation initiation factor 2 subunit 2	EIF2S2
118	P21128	Poly(U)-specific endoribonuclease	ENDOU
119	P21217	Galactoside 3(4)-L-fucosyltransferase	FUT3
120	P22680	Cholesterol 7-alpha-monooxygenase	CYP7A1
121	P22681	E3 ubiquitin-protein ligase CBL	CBL
122	P23515	Oligodendrocyte-myelin glycoprotein	OMG
123	P25025	C-X-C chemokine receptor type 2	CXCR2
124	P25445	Tumor necrosis factor receptor superfamily member 6	FAS
125	P26196	Probable ATP-dependent RNA helicase DDX6	DDX6
126	P26374	Rab proteins geranylgeranyltransferase component A 2	CHML
127	P28288	ATP-binding cassette sub-family D member 3	ABCD3
128	P29401	Transketolase	TKT
129	P29973	cGMP-gated cation channel alpha-1	CNGA1
130	P30084	Enoyl-CoA hydratase, mitochondrial	ECHS1
131	P30837	Aldehyde dehydrogenase X, mitochondrial	ALDH1B1
132	P31994	Low affinity immunoglobulin gamma Fc region receptor II-b	FCGR2B
133	P32297	Neuronal acetylcholine receptor subunit alpha-3	CHRNA3
134	P34998	Corticotropin-releasing factor receptor 1	CRHR1
135	P35237	Serpin B6	SERPINB6
136	P35573	Glycogen debranching enzyme;4-alpha-glucanotransferase;Amylo-alpha-1,6-glucosidase	AGL

to be continued…

…continuation

137	P35590	Tyrosine-protein kinase receptor Tie-1	TIE1
138	P35713	Transcription factor SOX-18	SOX18
139	P35858	Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS
140	P36894	Bone morphogenetic protein receptor type-1A	BMPR1A
141	P39877	Calcium-dependent phospholipase A2	PLA2G5
142	P40939	Trifunctional enzyme subunit alpha, mitochondrial;Long-chain enoyl-CoA hydratase;Long chain 3-hydroxyacyl-CoA dehydrogenase	HADHA
143	P42127	Agouti-signaling protein	ASIP
144	P42330	Aldo-keto reductase family 1 member C3	AKR1C3
145	P42356	Phosphatidylinositol 4-kinase alpha	PI4KA
146	P42702	Leukemia inhibitory factor receptor	LIFR
147	P43243	Matrin-3	MATR3
148	P43304	Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2
149	P43490	Nicotinamide phosphoribosyltransferase	NAMPT
150	P48378	DNA-binding protein RFX2	RFX2
151	P48556	26S proteasome non-ATPase regulatory subunit 8	PSMD8
152	P49747	Cartilage oligomeric matrix protein	COMP
153	P49917	DNA ligase 4	LIG4
154	P50395	Rab GDP dissociation inhibitor beta	GDI2
155	Q9NQ69;P50458	LIM/homeobox protein Lhx9;LIM/homeobox protein Lhx2	LHX9;LHX2
156	P50579	Methionine aminopeptidase 2	METAP2
157	P51553	Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial	IDH3G
158	P51788	Chloride channel protein 2	CLCN2
159	P53778	Mitogen-activated protein kinase 12	MAPK12
160	P54578	Ubiquitin carboxyl-terminal hydrolase 14	USP14
161	P56192	Methionine--tRNA ligase, cytoplasmic	MARS
162	P58397	A disintegrin and metalloproteinase with thrombospondin motifs 12	ADAMTS12
163	P58546	Myotrophin	MTPN
164	P63261;P60709;Q9BYX7;Q562R1;P63267;P68133;P68032;P62736;Q6S8J3;A5A3E0	Actin, cytoplasmic 2;Actin, cytoplasmic 2, N-terminally processed;Actin, cytoplasmic 1;Actin, cytoplasmic 1, N-terminally processed;Putative beta-actin-like protein 3;Putative beta-actin-like protein 3, N-terminally processed;Beta-actin-like protein 2;Actin, gamma-enteric smooth muscle;Actin, alpha skeletal muscle;Actin, alpha cardiac muscle 1;Actin, aortic smooth muscle;POTE ankyrin domain family member E;POTE ankyrin domain family member F	ACTG1;ACTB;POTEKP;ACTBL2;ACTG2;ACTA1;ACTC1;ACTA2;POTEE;POTEF
165	P61421	V-type proton ATPase subunit d 1	ATP6V0D1
166	P62834	Ras-related protein Rap-1A	RAP1A

to be continued…

…continuation

167	P78357	Contactin-associated protein 1	CNTNAP1
168	P78362	SRSF protein kinase 2;SRSF protein kinase 2 N-terminal;SRSF protein kinase 2 C-terminal	SRPK2
169	P80188	Neutrophil gelatinase-associated lipocalin	LCN2
170	P98073	Enteropeptidase;Enteropeptidase non-catalytic heavy chain;Enteropeptidase catalytic light chain	TMPRSS15
171	P98194	Calcium-transporting ATPase type 2C member 1	ATP2C1
172	Q01726	Melanocyte-stimulating hormone receptor	MC1R
173	Q01814	Plasma membrane calcium-transporting ATPase 2	ATP2B2
174	Q02985	Complement factor H-related protein 3	CFHR3
175	Q03052	POU domain, class 3, transcription factor 1	POU3F1
176	Q03252	Lamin-B2	LMNB2
177	Q03431	Parathyroid hormone/parathyroid hormone-related peptide receptor	PTH1R
178	Q06730	Zinc finger protein 33A	ZNF33A
179	Q06889	Early growth response protein 3	EGR3
180	Q07283	Trichohyalin	TCHH
181	Q10713	Mitochondrial-processing peptidase subunit alpha	PMPCA
182	Q12767	Uncharacterized protein KIAA0195	KIAA0195
183	Q12931	Heat shock protein 75 kDa, mitochondrial	TRAP1
184	Q12955	Ankyrin-3	ANK3
185	Q13018	Secretory phospholipase A2 receptor;Soluble secretory phospholipase A2 receptor	PLA2R1
186	Q13045	Protein flightless-1 homolog	FLII
187	Q13164	Mitogen-activated protein kinase 7	MAPK7
188	Q13247	Serine/arginine-rich splicing factor 6	SRSF6
189	Q13316	Dentin matrix acidic phosphoprotein 1	DMP1
190	Q13451	Peptidyl-prolyl cis-trans isomerase FKBP5;Peptidyl-prolyl cis-trans isomerase FKBP5, N-terminally processed	FKBP5
191	Q13523	Serine/threonine-protein kinase PRP4 homolog	PRPF4B
192	Q13561	Dynactin subunit 2	DCTN2
193	Q14008	Cytoskeleton-associated protein 5	CKAP5
194	Q14435	Polypeptide N-acetylgalactosaminyltransferase 3	GALNT3
195	Q14527	Helicase-like transcription factor	HLTF
196	Q14669	E3 ubiquitin-protein ligase TRIP12	TRIP12
197	Q14678	KN motif and ankyrin repeat domain-containing protein 1	KANK1
198	Q14721	Potassium voltage-gated channel subfamily B member 1	KCNB1
199	Q15399	Toll-like receptor 1	TLR1
200	Q15532	Protein SSXT	SS18
201	Q15751	Probable E3 ubiquitin-protein ligase HERC1	HERC1
202	Q16288	NT-3 growth factor receptor	NTRK3

to be continued…

…continuation

203	Q16445	Gamma-aminobutyric acid receptor subunit alpha-6	GABRA6
204	Q16513	Serine/threonine-protein kinase N2	PKN2
205	Q16566	Calcium/calmodulin-dependent protein kinase type IV	CAMK4
206	Q16600	Zinc finger protein 239	ZNF239
207	Q16670	Zinc finger and SCAN domain-containing protein 26	ZSCAN26
208	Q16787	Laminin subunit alpha-3	LAMA3
209	Q16821	Protein phosphatase 1 regulatory subunit 3A	PPP1R3A
210	Q17RC7	Exocyst complex component 3-like protein 4	EXOC3L4
211	Q2M389	WASH complex subunit 7	KIAA1033
212	Q2PZI1	Probable C-mannosyltransferase DPY19L1	DPY19L1
213	Q38SD2	Leucine-rich repeat serine/threonine-protein kinase 1	LRRK1
214	Q3KQV3	Zinc finger protein 792	ZNF792
215	Q3KQV9	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	UAP1L1
216	Q3L8U1	Chromodomain-helicase-DNA-binding protein 9	CHD9
217	Q3ZCV2	Uncharacterized protein C1orf177	C1orf177
218	Q460N5	Poly [ADP-ribose] polymerase 14	PARP14
219	Q4G0P3	Hydrocephalus-inducing protein homolog	HYDIN
220	Q4G0T1	Scavenger receptor cysteine-rich domain-containing protein	SCART1
221	Q4KWH8	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-1	PLCH1
222	Q4VC12	Putative protein MSS51 homolog, mitochondrial	MSS51
223	Q502W7	Coiled-coil domain-containing protein 38	CCDC38
224	Q53EP0	Fibronectin type III domain-containing protein 3B	FNDC3B
225	Q53QZ3	Rho GTPase-activating protein 15	ARHGAP15
226	Q562E7	WD repeat-containing protein 81	WDR81
227	Q58EX2	Protein sidekick-2	SDK2
228	Q5D1E8	Ribonuclease ZC3H12A	ZC3H12A
229	Q5EBM4	Putative zinc finger protein 542	ZNF542P
230	Q5H9M0	PWWP domain-containing protein MUM1L1	MUM1L1
231	Q5JRC9	Protein FAM47A	FAM47A
232	Q5JY77	G-protein coupled receptor-associated sorting protein 1	GPRASP1
233	Q5MJ68	Speedy protein C	SPDYC
234	Q5SVZ6	Zinc finger MYM-type protein 1	ZMYM1
235	Q5SWX8	Protein odr-4 homolog	ODR4
236	Q5SXM2	snRNA-activating protein complex subunit 4	SNAPC4
237	Q5SZB4	Uncharacterized protein C9orf50	C9orf50
238	Q5T0F9	Coiled-coil and C2 domain-containing protein 1B	CC2D1B
239	Q5T1B1	Uncharacterized protein C10orf91	C10orf91
240	Q5T1H1	Protein eyes shut homolog	EYS
241	Q5T4I8	Putative uncharacterized protein C6orf52	C6orf52

to be continued…

…continuation

242	Q5T4W7	Artemin	ARTN
243	Q5TB30	DEP domain-containing protein 1A	DEPDC1
244	Q5TB80	Centrosomal protein of 162 kDa	CEP162
245	Q5UE93	Phosphoinositide 3-kinase regulatory subunit 6	PIK3R6
246	Q5VTA0	PRAME family member 17	PRAMEF17
247	Q5VU43	Myomegalin	PDE4DIP
248	Q5VU57	Cytosolic carboxypeptidase 6	AGBL4
249	Q5VU65	Nuclear pore membrane glycoprotein 210-like	NUP210L
250	Q5VUJ6	Leucine-rich repeat and calponin homology domain-containing protein 2	LRCH2
251	Q5VVY1	Alpha N-terminal protein methyltransferase 1B	METTL11B
252	Q5W0N0	Uncharacterized protein C9orf57	C9orf57
253	Q5XG92	Carboxylesterase 4A	CES4A
254	Q5XX13	F-box/WD repeat-containing protein 10	FBXW10
255	Q63HK3	Zinc finger protein with KRAB and SCAN domains 2	ZKSCAN2
256	Q63ZY3	KN motif and ankyrin repeat domain-containing protein 2	KANK2
257	Q674R7	Autophagy-related protein 9B	ATG9B
258	Q68CR7	Leucine-rich repeat-containing protein 66	LRRC66
259	Q68CZ1	Protein fantom	RPGRIP1L
260	Q68DX3	FERM and PDZ domain-containing protein 2	FRMPD2
261	Q6IAA8	Ragulator complex protein LAMTOR1	LAMTOR1
262	Q6IV72	Zinc finger protein 425	ZNF425
263	Q6N043	Zinc finger protein 280D	ZNF280D
264	Q6NXP6	NADP-dependent oxidoreductase domain-containing protein 1	NOXRED1
265	Q6NZY4	Zinc finger CCHC domain-containing protein 8	ZCCHC8
266	Q6PXP3	Solute carrier family 2, facilitated glucose transporter member 7	SLC2A7
267	Q6QHC5	Sphingolipid delta(4)-desaturase/C4-hydroxylase DES2	DEGS2
268	Q6U841	Sodium-driven chloride bicarbonate exchanger	SLC4A10
269	Q6UWM7	Lactase-like protein	LCTL
270	Q6UWQ7	Insulin growth factor-like family member 2	IGFL2
271	Q6UWY2	Serine protease 57	PRSS57
272	Q6UXQ4	Uncharacterized protein C2orf66	C2orf66
273	Q6Y7W6	PERQ amino acid-rich with GYF domain-containing protein 2	GIGYF2
274	Q6ZMY3	SPOC domain-containing protein 1	SPOCD1
275	Q6ZN19	Zinc finger protein 841	ZNF841
276	Q6ZN68	Putative C-mannosyltransferase DPY19L2P2	DPY19L2P2
277	Q6ZQT0	Putative uncharacterized protein FLJ45035	Not available

to be continued…

…continuation

278	Q6ZR37	Pleckstrin homology domain-containing family G member 7	PLEKHG7
279	Q6ZRH7	Cation channel sperm-associated protein subunit gamma	CATSPERG
280	Q6ZRI6	Uncharacterized protein C15orf39	C15orf39
281	Q6ZRV2	Protein FAM83H	FAM83H
282	Q6ZT12	E3 ubiquitin-protein ligase UBR3	UBR3
283	Q6ZUX3	Protein FAM179A	FAM179A
284	Q6ZVT0	Inactive polyglycylase TTLL10	TTLL10
285	Q6ZXV5	Transmembrane and TPR repeat-containing protein 3	TMTC3
286	Q7L7V1	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX32	DHX32
287	Q7Z2Y5	Nik-related protein kinase	NRK
288	Q7Z3B1	Neuronal growth regulator 1	NEGR1
289	Q7Z3F1	Integral membrane protein GPR155	GPR155
290	Q7Z3S7	Voltage-dependent calcium channel subunit alpha-2/delta-4;Voltage-dependent calcium channel subunit alpha-2-4;Voltage-dependent calcium channel subunit delta-4	CACNA2D4
291	Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	HUWE1
292	Q7Z745	Maestro heat-like repeat-containing protein family member 2B	MROH2B
293	Q86SE5	RNA-binding Raly-like protein	RALYL
294	Q86TJ2	Transcriptional adapter 2-beta	TADA2B
295	Q86UL3	Glycerol-3-phosphate acyltransferase 4	AGPAT6
296	Q86VM9	Zinc finger CCCH domain-containing protein 18	ZC3H18
297	Q86VV8	Rotatin	RTTN
298	Q86WB0	Nuclear-interacting partner of ALK	ZC3HC1
299	Q86WV1	Src kinase-associated phosphoprotein 1	SKAP1
300	Q86XD5	Protein FAM131B	FAM131B
301	Q86XI2	Condensin-2 complex subunit G2	NCAPG2
302	Q86YA3	Protein ZGRF1	ZGRF1
303	Q86YQ8;Q9HCH3	Copine-8;Copine-5	CPNE8;CPNE5
304	Q86YW5	Trem-like transcript 1 protein	TREML1
305	Q8IUA7	ATP-binding cassette sub-family A member 9	ABCA9
306	Q8IUX8	Epidermal growth factor-like protein 6	EGFL6
307	Q8IVG5	Sterile alpha motif domain-containing protein 9-like	SAMD9L
308	Q8IWD4	Coiled-coil domain-containing protein 117	CCDC117
309	Q8IX07	Zinc finger protein ZFPM1	ZFPM1
310	Q8IXT1	DNA damage-induced apoptosis suppressor protein	DDIAS
311	Q8IY21	Probable ATP-dependent RNA helicase DDX60	DDX60
312	Q8IYB7	DIS3-like exonuclease 2	DIS3L2
313	Q8IYX4	Dead end protein homolog 1	DND1

to be continued…

…continuation

314	Q8IZJ0	Interferon lambda-2	IFNL2
315	Q8IZJ6	Inactive L-threonine 3-dehydrogenase, mitochondrial	TDH
316	Q8N2C3	DEP domain-containing protein 4	DEPDC4
317	Q8N2C7	Protein unc-80 homolog	UNC80
318	Q8N2H3	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 2	PYROXD2
319	Q8N2R0	Protein odd-skipped-related 2	OSR2
320	Q8N4B1	Sesquipedalian-1	FAM109A
321	Q8N4T0	Carboxypeptidase A6	CPA6
322	Q8N5C7	DTW domain-containing protein 1	DTWD1
323	Q8N6Y2	Leucine-rich repeat-containing protein 17	LRRC17
324	Q8N7X4	Melanoma-associated antigen B6	MAGEB6
325	Q8N7X8	SIGLEC family-like protein 1	SIGLECL1
326	Q8N815	Cyclin N-terminal domain-containing protein 1	CNTD1
327	Q8N9K5	Zinc finger protein 565	ZNF565
328	Q8N9M1	Uncharacterized protein C19orf47	C19orf47
329	Q8NCR0	UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 2	B3GALNT2
330	Q8ND24	RING finger protein 214	RNF214
331	Q8NDA2	Hemicentin-2	HMCN2
332	Q8NDV2	G-protein coupled receptor 26	GPR26
333	Q8NDZ2	SUMO-interacting motif-containing protein 1	SIMC1
334	Q8NEG4	Protein FAM83F	FAM83F
335	Q8NEP9	Zinc finger protein 555	ZNF555
336	Q8NEY8	Periphilin-1	PPHLN1
337	Q8NF86	Serine protease 33	PRSS33
338	Q8NFM7	Interleukin-17 receptor D	IL17RD
339	Q8NFW9	Rab effector MyRIP	MYRIP
340	Q8NG31	Protein CASC5	CASC5
341	Q8NHU2	Cilia- and flagella-associated protein 61	CFAP61
342	Q8NHY6	Zinc finger protein 28 homolog	ZFP28
343	Q8TAA1	Probable ribonuclease 11	RNASE11
344	Q8TAD2	Interleukin-17D	IL17D
345	Q8TAQ5	Zinc finger protein 420	ZNF420
346	Q8TC84	Fibronectin type 3 and ankyrin repeat domains protein 1	FANK1
347	Q8TCB6	Olfactory receptor 51E1	OR51E1
348	Q8TD47	40S ribosomal protein S4, Y isoform 2	RPS4Y2
349	Q8TDI7	Transmembrane channel-like protein 2	TMC2
350	Q8TE73	Dynein heavy chain 5, axonemal	DNAH5
351	Q8TF40	Folliculin-interacting protein 1	FNIP1
352	Q8WUA2	Peptidyl-prolyl cis-trans isomerase-like 4	PPIL4

to be continued…

…continuation

353	Q8WWQ8	Stabilin-2;190 kDa form stabilin-2	STAB2
354	Q92560	Ubiquitin carboxyl-terminal hydrolase BAP1	BAP1
355	Q92624	Amyloid protein-binding protein 2	APPBP2
356	Q92673	Sortilin-related receptor	SORL1
357	Q92729	Receptor-type tyrosine-protein phosphatase U	PTPRU
358	Q92793	CREB-binding protein	CREBBP
359	Q92835	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	INPP5D
360	Q92974	Rho guanine nucleotide exchange factor 2	ARHGEF2
361	Q92993	Histone acetyltransferase KAT5	KAT5
362	Q93015	N-acetyltransferase 6	NAT6
363	Q969P0	Immunoglobulin superfamily member 8	IGSF8
364	Q96A32	Myosin regulatory light chain 2, skeletal muscle isoform	MYLPF
365	Q96AV8	Transcription factor E2F7	E2F7
366	Q96BQ5	Coiled-coil domain-containing protein 127	CCDC127
367	Q96BZ8	Leukocyte receptor cluster member 1	LENG1
368	Q96C10	Probable ATP-dependent RNA helicase DHX58	DHX58
369	Q96C57	Uncharacterized protein C12orf43	C12orf43
370	Q96CS7	Pleckstrin homology domain-containing family B member 2	PLEKHB2
371	Q96D09	G-protein coupled receptor-associated sorting protein 2	GPRASP2
372	Q96F07	Cytoplasmic FMR1-interacting protein 2	CYFIP2
373	Q96GE4	Centrosomal protein of 95 kDa	CEP95
374	Q96H20	Vacuolar-sorting protein SNF8	SNF8
375	Q96HA7	Tonsoku-like protein	TONSL
376	Q96HJ3	Coiled-coil domain-containing protein 34	CCDC34
377	Q96I24	Far upstream element-binding protein 3	FUBP3
378	Q96IR2	Zinc finger protein 845	ZNF845
379	Q96IV0	Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase	NGLY1
380	Q96JL9	Zinc finger protein 333	ZNF333
381	Q96LU5	Mitochondrial inner membrane protease subunit 1	IMMP1L
382	Q96M83	Coiled-coil domain-containing protein 7	CCDC7
383	Q96MU6	Zinc finger protein 778	ZNF778
384	Q96N16	Janus kinase and microtubule-interacting protein 1	JAKMIP1
385	Q96NY7	Chloride intracellular channel protein 6	CLIC6
386	Q96NZ1	Forkhead box protein N4	FOXN4
387	Q96P71	N-terminal EF-hand calcium-binding protein 3	NECAB3
388	Q96PY5	Formin-like protein 2	FMNL2
389	Q96QH2	PML-RARA-regulated adapter molecule 1	PRAM1
390	Q96QK1	Vacuolar protein sorting-associated protein 35	VPS35

to be continued…

...continuation

391	Q96RD2	Olfactory receptor 52B2	OR52B2
392	Q96RL7	Vacuolar protein sorting-associated protein 13A	VPS13A
393	Q96RS6	NudC domain-containing protein 1	NUDCD1
394	Q96SQ9	Cytochrome P450 2S1	CYP2S1
395	Q99081	Transcription factor 12	TCF12
396	Q99527	G-protein coupled estrogen receptor 1	GPER1
397	Q99541	Perilipin-2	PLIN2
398	Q99547	M-phase phosphoprotein 6	MPHOSPH6
399	Q99676	Zinc finger protein 184	ZNF184
400	Q99679	Probable G-protein coupled receptor 21	GPR21
401	Q99700	Ataxin-2	ATXN2
402	Q99757	Thioredoxin, mitochondrial	TXN2
403	Q99959	Plakophilin-2	PKP2
404	Q9BPW9	Dehydrogenase/reductase SDR family member 9	DHRS9
405	Q9BQ61	Uncharacterized protein C19orf43	C19orf43
406	Q9BRL6	Serine/arginine-rich splicing factor 8	SRSF8
407	Q9BRS8	La-related protein 6	LARP6
408	Q9BVI0	PHD finger protein 20	PHF20
409	Q9BVW5	TIMELESS-interacting protein	TIPIN
410	Q9BXR3	Endogenous retrovirus group K member 6 Pol protein;Reverse transcriptase;Ribonuclease H;Integrase	ERVK-6
411	Q9BY15	EGF-like module-containing mucin-like hormone receptor-like 3;EGF-like module-containing mucin-like hormone receptor-like 3 subunit alpha;EGF-like module-containing mucin-like hormone receptor-like 3 subunit beta	EMR3
412	Q9BY50	Signal peptidase complex catalytic subunit SEC11C	SEC11C
413	Q9BZL6	Serine/threonine-protein kinase D2	PRKD2
414	Q9C0B9	Zinc finger CCHC domain-containing protein 2	ZCCHC2
415	Q9C0F3	Zinc finger protein 436	ZNF436
416	Q9GZR1	Sentrin-specific protease 6	SENP6
417	Q9H093	NUAK family SNF1-like kinase 2	NUAK2
418	Q9H1K0	Rabenosyn-5	RBSN
419	Q9H267	Vacuolar protein sorting-associated protein 33B	VPS33B
420	Q9H295	Dendritic cell-specific transmembrane protein	DCSTAMP
421	Q9H2S6	Tenomodulin	TNMD
422	Q9H2X9	Solute carrier family 12 member 5	SLC12A5
423	Q9H3S7	Tyrosine-protein phosphatase non-receptor type 23	PTPN23
424	Q9H5L6	DNA transposase THAP9	THAP9
425	Q9H5V8	CUB domain-containing protein 1	CDCP1
426	Q9H777	Zinc phosphodiesterase ELAC protein 1	ELAC1
427	Q9H8M5	Metal transporter CNNM2	CNNM2

to be continued…

…continuation

428	Q9H9A6	Leucine-rich repeat-containing protein 40	LRRC40
429	Q9HB75	p53-induced death domain-containing protein 1	PIDD1
430	Q9HBF4	Zinc finger FYVE domain-containing protein 1	ZFYVE1
431	Q9HBV1	Popeye domain-containing protein 3	POPDC3
432	Q9HCS7	Pre-mRNA-splicing factor SYF1	XAB2
433	Q9HCU4	Cadherin EGF LAG seven-pass G-type receptor 2	CELSR2
434	Q9HD47	Ran guanine nucleotide release factor	RANGRF
435	Q9NPB8	Glycerophosphocholine phosphodiesterase GPCPD1	GPCPD1
436	Q9NPH9	Interleukin-26	IL26
437	Q9NR09	Baculoviral IAP repeat-containing protein 6	BIRC6
438	Q9NR22	Protein arginine N-methyltransferase 8	PRMT8
439	Q9NRK6	ATP-binding cassette sub-family B member 10, mitochondrial	ABCB10
440	Q9NRL2	Bromodomain adjacent to zinc finger domain protein 1A	BAZ1A
441	Q9NRX6	Protein kish-B	TMEM167B
442	Q9NSE4	Isoleucine--tRNA ligase, mitochondrial	IARS2
443	Q9NSY1	BMP-2-inducible protein kinase	BMP2K
444	Q9NT22	EMILIN-3	EMILIN3
445	Q9NVE7	Pantothenate kinase 4	PANK4
446	Q9NVP4	Double zinc ribbon and ankyrin repeat-containing protein 1	DZANK1
447	Q9NX65	Zinc finger and SCAN domain-containing protein 32	ZSCAN32
448	Q9NXA8	NAD-dependent protein deacylase sirtuin-5, mitochondrial	SIRT5
449	Q9NXC5	WD repeat-containing protein mio	MIOS
450	Q9NYF8	Bcl-2-associated transcription factor 1	BCLAF1
451	Q9NYK6	Protein EURL homolog	EURL
452	Q9NYT0	Pleckstrin-2	PLEK2
453	Q9NYU1	UDP-glucose:glycoprotein glucosyltransferase 2	UGGT2
454	Q9NZI6	Transcription factor CP2-like protein 1	TFCP2L1
455	Q9NZJ4	Sacsin	SACS
456	Q9NZJ7	Mitochondrial carrier homolog 1	MTCH1
457	Q9NZN5	Rho guanine nucleotide exchange factor 12	ARHGEF12
458	Q9NZQ8	Transient receptor potential cation channel subfamily M member 5	TRPM5
459	Q9P0K9	DOMON domain-containing protein FRRS1L	FRRS1L
460	Q9P2D3	HEAT repeat-containing protein 5B	HEATR5B
461	Q9P2D6	Protein FAM135A	FAM135A
462	Q9UBC1	NF-kappa-B inhibitor-like protein 1	NFKBIL1
463	Q9UBU9	Nuclear RNA export factor 1	NXF1
464	Q9UFF9	CCR4-NOT transcription complex subunit 8	CNOT8
465	Q9UGL9	Cysteine-rich C-terminal protein 1	CRCT1

to be continued…

…continuation

466	Q9UHF1	Epidermal growth factor-like protein 7	EGFL7
467	Q9UHF3	Putative N-acetyltransferase 8B	NAT8B
468	Q9UII4	E3 ISG15 protein ligase HERC5	HERC5
469	Q9UIW2	Plexin-A1	PLXNA1
470	Q9UK73	Protein fem-1 homolog B	FEM1B
471	Q9UKE5	TRAF2 and NCK-interacting protein kinase	TNIK
472	Q9UKK3	Poly [ADP-ribose] polymerase 4	PARP4
473	Q9UKP5	A disintegrin and metalloproteinase with thrombospondin motifs 6	ADAMTS6
474	Q9UKQ2	Disintegrin and metalloproteinase domain-containing protein 28	ADAM28
475	Q9ULD4	Bromodomain and PHD finger-containing protein 3	BRPF3
476	Q9ULK4	Mediator of RNA polymerase II transcription subunit 23	MED23
477	Q9ULP9	TBC1 domain family member 24	TBC1D24
478	Q9ULZ2	Signal-transducing adaptor protein 1	STAP1
479	Q9UM47	Neurogenic locus notch homolog protein 3;Notch 3 extracellular truncation;Notch 3 intracellular domain	NOTCH3
480	Q9UN72	Protocadherin alpha-7	PCDHA7
481	Q9UNI6	Dual specificity protein phosphatase 12	DUSP12
482	Q9UNX4	WD repeat-containing protein 3	WDR3
483	Q9UPV0	Centrosomal protein of 164 kDa	CEP164
484	Q9UPV9	Trafficking kinesin-binding protein 1	TRAK1
485	Q9Y210	Short transient receptor potential channel 6	TRPC6
486	Q9Y217	Myotubularin-related protein 6	MTMR6
487	Q9Y2F5	Little elongation complex subunit 1	ICE1
488	Q9Y2H2	Phosphatidylinositide phosphatase SAC2	INPP5F
489	Q9Y2I9	TBC1 domain family member 30	TBC1D30
490	Q9Y2L5	Trafficking protein particle complex subunit 8	TRAPPC8
491	Q9Y2P4	Long-chain fatty acid transport protein 6	SLC27A6
492	Q9Y399	28S ribosomal protein S2, mitochondrial	MRPS2
493	Q9Y3B7	39S ribosomal protein L11, mitochondrial	MRPL11
494	Q9Y3R0	Glutamate receptor-interacting protein 1	GRIP1
495	Q9Y3V2	RWD domain-containing protein 3	RWDD3
496	Q9Y520	Protein PRRC2C	PRRC2C
497	Q9Y576	Ankyrin repeat and SOCS box protein 1	ASB1
498	Q9Y5R8	Trafficking protein particle complex subunit 1	TRAPPC1
499	Q9Y5W3	Krueppel-like factor 2	KLF2
500	Q9Y676	28S ribosomal protein S18b, mitochondrial	MRPS18B
501	Q9Y6J6	Potassium voltage-gated channel subfamily E member 2	KCNE2
502	Q9Y6K1	DNA (cytosine-5)-methyltransferase 3A	DNMT3A
503	Q9Y6V0	Protein piccolo	PCLO

Human proteins exclusively identified in Stage IV colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	P00738;P00739	Haptoglobin;Haptoglobin alpha chain;Haptoglobin beta chain;Haptoglobin-related protein	HP;HPR

Human proteins shared between Stage I and Stage II in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A6NLX3	Speedy protein E4	SPDYE4
2	O00459	Phosphatidylinositol 3-kinase regulatory subunit beta	PIK3R2
3	O60939	Sodium channel subunit beta-2	SCN2B
4	Q12899	Tripartite motif-containing protein 26	TRIM26
5	Q2NKQ1	Small G protein signaling modulator 1	SGSM1
6	Q330K2	NADH dehydrogenase (ubiquinone) complex I, assembly factor 6	NDUFAF6
7	Q5SZK8	FRAS1-related extracellular matrix protein 2	FREM2
8	Q5T9C9	Phosphatidylinositol 4-phosphate 5-kinase-like protein 1	PIP5KL1
9	Q5VTT5	Myomesin-3	MYOM3
10	Q6ZS81	WD repeat- and FYVE domain-containing protein 4	WDFY4
11	Q8WV37	Zinc finger protein 480	ZNF480
12	Q9HAR2	Latrophilin-3	LPHN3
13	Q9NPA5	Zinc finger protein 64 homolog, isoforms 1 and 2	ZFP64
14	Q9P267	Methyl-CpG-binding domain protein 5	MBD5
15	Q9Y2G3	Probable phospholipid-transporting ATPase IF	ATP11B

Human proteins shared between Stage I and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A6NCQ9	RING finger protein 222	RNF222
2	O14756	17-beta-hydroxysteroid dehydrogenase type 6	HSD17B6
3	O75936	Gamma-butyrobetaine dioxygenase	BBOX1
4	O95498	Vascular non-inflammatory molecule 2	VNN2
5	P30559	Oxytocin receptor	OXTR
6	P35556	Fibrillin-2	FBN2
7	P42768	Wiskott-Aldrich syndrome protein	WAS
8	P43308	Translocon-associated protein subunit beta	SSR2
9	P78329	Phylloquinone omega-hydroxylase CYP4F2	CYP4F2
10	Q03519	Antigen peptide transporter 2	TAP2
11	Q05D60	Deuterosome protein 1	CCDC67
12	Q12802	A-kinase anchor protein 13	AKAP13
13	Q16654	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial	PDK4
14	Q69YN2	CWF19-like protein 1	CWF19L1
15	Q6ECI4	Zinc finger protein 470	ZNF470
16	Q6IEE8	Schlafen family member 12-like	SLFN12L
17	Q6ZMT4	Lysine-specific demethylase 7A	KDM7A
18	Q7Z6R9	Transcription factor AP-2-delta	TFAP2D
19	Q86W28	NACHT, LRR and PYD domains-containing protein 8	NLRP8
20	Q8IXN7	N-acetylaspartylglutamate synthase A	RIMKLA
21	Q8N8U9	BMP-binding endothelial regulator protein	BMPER
22	Q8NDZ6	Transmembrane protein 161B	TMEM161B
23	Q8NHP8	Putative phospholipase B-like 2;Putative phospholipase B-like 2 32 kDa form;Putative phospholipase B-like 2 45 kDa form	PLBD2
24	Q9BZG8	Diphthamide biosynthesis protein 1	DPH1
25	Q9GZP0	Platelet-derived growth factor D;Platelet-derived growth factor D, latent form;Platelet-derived growth factor D, receptor-binding form	PDGFD
26	Q9H2C1	LIM/homeobox protein Lhx5	LHX5
27	Q9H9S3	Protein transport protein Sec61 subunit alpha isoform 2	SEC61A2
28	Q9HBL0	Tensin-1	TNS1
29	Q9NPG1	Frizzled-3	FZD3
30	Q9NR19	Acetyl-coenzyme A synthetase, cytoplasmic	ACSS2
31	Q9UBC5	Unconventional myosin-Ia	MYO1A
32	Q9UKJ3	G patch domain-containing protein 8	GPATCH8
33	Q9Y228	TRAF3-interacting JNK-activating modulator	TRAF3IP3

Human proteins shared between Stage I and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	Q8IXL6	Extracellular serine/threonine protein kinase FAM20C	FAM20C

Human proteins shared between Stage II and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A0AUZ9	KAT8 regulatory NSL complex subunit 1-like protein	KANSL1L
2	A1A4S6	Rho GTPase-activating protein 10	ARHGAP10
3	H7BZ55	Putative ciliary rootlet coiled-coil protein-like 3 protein	CROCC2
4	O14545	TRAF-type zinc finger domain-containing protein 1	TRAFD1
5	O14986	Phosphatidylinositol 4-phosphate 5-kinase type-1 beta	PIP5K1B
6	Q08ET2;O15389	Sialic acid-binding Ig-like lectin 14;Sialic acid-binding Ig-like lectin 5	SIGLEC14;SIGLEC5
7	O43164	E3 ubiquitin-protein ligase Praja-2	PJA2
8	O43815	Striatin	STRN
9	O60266	Adenylate cyclase type 3	ADCY3
10	O60271	C-Jun-amino-terminal kinase-interacting protein 4	SPAG9
11	O60477	BMP/retinoic acid-inducible neural-specific protein 1	BRINP1
12	O60603	Toll-like receptor 2	TLR2
13	O60784	Target of Myb protein 1	TOM1
14	O60825	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2;6-phosphofructo-2-kinase;Fructose-2,6-bisphosphatase	PFKFB2
15	O75165	DnaJ homolog subfamily C member 13	DNAJC13
16	O75400	Pre-mRNA-processing factor 40 homolog A	PRPF40A
17	O94818	Nucleolar protein 4	NOL4
18	O95155	Ubiquitin conjugation factor E4 B	UBE4B
19	P00338	L-lactate dehydrogenase A chain	LDHA
20	P01009	Alpha-1-antitrypsin;Short peptide from AAT	SERPINA1
21	P01024	Complement C3;Complement C3 beta chain;C3-beta-c;Complement C3 alpha chain;C3a anaphylatoxin;Acylation stimulating protein;Complement C3b alpha chain;Complement C3c alpha chain fragment 1;Complement C3dg fragment;Complement C3g fragment;Complement C3d fragment;Complement C3f fragment;Complement C3c alpha chain fragment 2	C3
22	P01033	Metalloproteinase inhibitor 1	TIMP1
23	P01133	Pro-epidermal growth factor;Epidermal growth factor	EGF
24	P01833	Polymeric immunoglobulin receptor;Secretory component	PIGR
25	P02042	Hemoglobin subunit delta	HBD
26	P02675	Fibrinogen beta chain;Fibrinopeptide B;Fibrinogen beta chain	FGB

to be continued…

…continuation

27	P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH
28	P07900;Q58FG0	Heat shock protein HSP 90-alpha	HSP90AA1
29	P08631	Tyrosine-protein kinase HCK	HCK
30	P08648	Integrin alpha-5;Integrin alpha-5 heavy chain;Integrin alpha-5 light chain	ITGA5
31	P0CF74;P0DOY3;P0DOY2;A0M8Q6	Ig lambda-6 chain C region;Ig lambda-7 chain C region	IGLC6;IGLC7
32	P0CF97	Putative protein FAM200B	FAM200B
33	P0DP72	V-set and immunoglobulin domain-containing protein 10-like 2	VSIG10L2
34	P11216	Glycogen phosphorylase, brain form	PYGB
35	P13569	Cystic fibrosis transmembrane conductance regulator	CFTR
36	P16112	Aggrecan core protein;Aggrecan core protein 2	ACAN
37	P19438	Tumor necrosis factor receptor superfamily member 1A;Tumor necrosis factor receptor superfamily member 1A, membrane form;Tumor necrosis factor-binding protein 1	TNFRSF1A
38	P20591	Interferon-induced GTP-binding protein Mx1;Interferon-induced GTP-binding protein Mx1, N-terminally processed	MX1
39	P21580	Tumor necrosis factor alpha-induced protein 3;A20p50;A20p37	TNFAIP3
40	P21817	Ryanodine receptor 1	RYR1
41	P28069	Pituitary-specific positive transcription factor 1	POU1F1
42	P28906	Hematopoietic progenitor cell antigen CD34	CD34
43	P50148;P29992	Guanine nucleotide-binding protein G(q) subunit alpha;Guanine nucleotide-binding protein subunit alpha-11	GNAQ;GNA11
44	P33552	Cyclin-dependent kinases regulatory subunit 2	CKS2
45	P35579	Myosin-9	MYH9
46	P48052	Carboxypeptidase A2	CPA2
47	P48634	Protein PRRC2A	PRRC2A
48	P50549	ETS translocation variant 1	ETV1
49	P51864	Putative teratocarcinoma-derived growth factor 3	TDGF1P3
50	P51948	CDK-activating kinase assembly factor MAT1	MNAT1
51	P60174	Triosephosphate isomerase	TPI1
52	P62701	40S ribosomal protein S4, X isoform	RPS4X
53	Q5VTE0;P68104;Q05639	Putative elongation factor 1-alpha-like 3;Elongation factor 1-alpha 1;Elongation factor 1-alpha 2	EEF1A1P5;EEF1A1;EEF1A2
54	P69905	Hemoglobin subunit alpha	HBA1
55	P84095	Rho-related GTP-binding protein RhoG	RHOG

to be continued…

…continuation

56	Q05823	2-5A-dependent ribonuclease	RNASEL
57	Q06124	Tyrosine-protein phosphatase non-receptor type 11	PTPN11
58	Q06520	Bile salt sulfotransferase	SULT2A1
59	Q07092	Collagen alpha-1(XVI) chain	COL16A1
60	Q08211	ATP-dependent RNA helicase A	DHX9
61	Q13009	T-lymphoma invasion and metastasis-inducing protein 1	TIAM1
62	Q13227	G protein pathway suppressor 2	GPS2
63	Q13231	Chitotriosidase-1	CHIT1
64	Q13698	Voltage-dependent L-type calcium channel subunit alpha-1S	CACNA1S
65	Q14258	E3 ubiquitin/ISG15 ligase TRIM25	TRIM25
66	Q14833	Metabotropic glutamate receptor 4	GRM4
67	Q14896	Myosin-binding protein C, cardiac-type	MYBPC3
68	Q14C86	GTPase-activating protein and VPS9 domain-containing protein 1	GAPVD1
69	Q15011	Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 protein	HERPUD1
70	Q15118	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial	PDK1
71	Q15631	Translin	TSN
72	Q15788	Nuclear receptor coactivator 1	NCOA1
73	Q3BBW0; Q3BBV1	Neuroblastoma breakpoint family member 9;Neuroblastoma breakpoint family member 20	NBPF9;NBPF20
74	Q3SXY8	ADP-ribosylation factor-like protein 13B	ARL13B
75	Q49A26	Putative oxidoreductase GLYR1	GLYR1
76	Q58A45	PAB-dependent poly(A)-specific ribonuclease subunit PAN3	PAN3
77	Q5EBL4	RILP-like protein 1	RILPL1
78	Q5HYW2	NHS-like protein 2	NHSL2
79	Q5T5N4	Uncharacterized protein C6orf118	C6orf118
80	Q5T848	Probable G-protein coupled receptor 158	GPR158
81	Q5VUA4	Zinc finger protein 318	ZNF318
82	Q6P3X3	Tetratricopeptide repeat protein 27	TTC27
83	Q6P4F1	Alpha-(1,3)-fucosyltransferase 10	FUT10
84	Q6UXG2	UPF0577 protein KIAA1324	KIAA1324
85	Q6ZMW3	Echinoderm microtubule-associated protein-like 6	EML6
86	Q6ZPD9	Probable C-mannosyltransferase DPY19L3	DPY19L3
87	Q6ZU35	Uncharacterized protein KIAA1211	KIAA1211
88	Q6ZUT4	Putative uncharacterized protein	FLJ43343
89	Q70CQ2	Ubiquitin carboxyl-terminal hydrolase 34	USP34

to be continued…

…continuation

90	Q7L0Y3	Mitochondrial ribonuclease P protein 1	TRMT10C
91	Q7Z2Z2	Elongation factor Tu GTP-binding domain-containing protein 1	EFTUD1
92	Q7Z443	Polycystic kidney disease protein 1-like 3	PKD1L3
93	Q7Z4H3	HD domain-containing protein 2	HDDC2
94	Q7Z7A1	Centriolin	CNTRL
95	Q7Z7M0	Multiple epidermal growth factor-like domains protein 8	MEGF8
96	Q86W56	Poly(ADP-ribose) glycohydrolase	PARG
97	Q8N4P2; Q86WT1	Tetratricopeptide repeat protein 30B;Tetratricopeptide repeat protein 30A	TTC30B;TTC30A
98	Q86XZ4	Spermatogenesis-associated serine-rich protein 2	SPATS2
99	Q8IVF2	Protein AHNAK2	AHNAK2
100	Q8IVW6	AT-rich interactive domain-containing protein 3B	ARID3B
101	Q8IXQ8	PDZ domain-containing protein 9	PDZD9
102	Q8IYJ0	PILR alpha-associated neural protein	PIANP
103	Q8IYS2	Uncharacterized protein KIAA2013	KIAA2013
104	Q8N0Z8	tRNA pseudouridine synthase-like 1	PUSL1
105	Q8N3K9	Cardiomyopathy-associated protein 5	CMYA5
106	Q8N4H5	Mitochondrial import receptor subunit TOM5 homolog	TOMM5
107	Q8N6C5	Immunoglobulin superfamily member 1	IGSF1
108	Q8N6M0	OTU domain-containing protein 6B	OTUD6B
109	Q8N6S4	Ankyrin repeat domain-containing protein 13C	ANKRD13C
110	Q8NB12	Histone-lysine N-methyltransferase SMYD1	SMYD1
111	Q8NBF1	Zinc finger protein GLIS1	GLIS1
112	Q8NDZ0	BEN domain-containing protein 2	BEND2
113	Q8NE01	Metal transporter CNNM3	CNNM3
114	Q8NE09	Regulator of G-protein signaling 22	RGS22
115	Q8NFR9	Interleukin-17 receptor E	IL17RE
116	Q8NFY4	Semaphorin-6D	SEMA6D
117	Q8NGC2	Olfactory receptor 4E2	OR4E2
118	Q8NH85	Olfactory receptor 5R1	OR5R1
119	Q8TB37	Iron-sulfur protein NUBPL	NUBPL
120	Q8TC76	Protein FAM110B	FAM110B
121	Q8TEU7	Rap guanine nucleotide exchange factor 6	RAPGEF6
122	Q8WUM0	Nuclear pore complex protein Nup133	NUP133
123	Q8WW33	Gametocyte-specific factor 1	GTSF1
124	Q8WWY7	WAP four-disulfide core domain protein 12	WFDC12
125	Q92911	Sodium/iodide cotransporter	SLC5A5

to be continued…

…continuation

126	Q93100	Phosphorylase b kinase regulatory subunit beta	PHKB
127	Q96EG1	Arylsulfatase G	ARSG
128	Q96IU4	Alpha/beta hydrolase domain-containing protein 14B	ABHD14B
129	Q96J94	Piwi-like protein 1	PIWIL1
130	Q96K21	Abscission/NoCut checkpoint regulator	ZFYVE19
131	Q96K31	Uncharacterized protein C8orf76	C8orf76
132	Q96MM6	Heat shock 70 kDa protein 12B	HSPA12B
133	Q96NU1	Sterile alpha motif domain-containing protein 11	SAMD11
134	Q96P11	Probable 28S rRNA (cytosine-C(5))-methyltransferase	NSUN5
135	Q96Q42	Alsin	ALS2
136	Q96RL1	BRCA1-A complex subunit RAP80	UIMC1
137	Q9BQP9	BPI fold-containing family A member 3	BPIFA3
138	Q9BV23	Monoacylglycerol lipase ABHD6	ABHD6
139	Q9BXB5	Oxysterol-binding protein-related protein 10	OSBPL10
140	Q9BYW2	Histone-lysine N-methyltransferase SETD2	SETD2
141	Q9BYX2	TBC1 domain family member 2A	TBC1D2
142	Q9BZ72	Membrane-associated phosphatidylinositol transfer protein 2	PITPNM2
143	Q9BZJ0	Crooked neck-like protein 1	CRNKL1
144	Q9BZL4	Protein phosphatase 1 regulatory subunit 12C	PPP1R12C
145	Q9C091	GREB1-like protein	GREB1L
146	Q9GZN7	Protein rogdi homolog	ROGDI
147	Q9H019	Mitochondrial fission regulator 1-like	MTFR1L
148	Q9H0E7	Ubiquitin carboxyl-terminal hydrolase 44	USP44
149	Q9H2J7	Sodium-dependent neutral amino acid transporter B(0)AT2	SLC6A15
150	Q9H2W1	Membrane-spanning 4-domains subfamily A member 6A	MS4A6A
151	Q9H313	Protein tweety homolog 1	TTYH1
152	Q9H3N1	Thioredoxin-related transmembrane protein 1	TMX1
153	Q9H6K4	Optic atrophy 3 protein	OPA3
154	Q9HAW9; Q9HAW8; Q9HAW7	UDP-glucuronosyltransferase 1-8;UDP-glucuronosyltransferase 1-10;UDP-glucuronosyltransferase 1-7	UGT1A8;UGT1A10; UGT1A7
155	Q9HB20	Pleckstrin homology domain-containing family A member 3	PLEKHA3
156	Q9HB21	Pleckstrin homology domain-containing family A member 1	PLEKHA1
157	Q9HBM1	Kinetochore protein Spc25	SPC25
158	Q9NPC7	Myoneurin	MYNN
159	Q9NSY0	Nuclear receptor-binding protein 2	NRBP2

to be continued…

…continuation

160	Q9NPL8	Complex I assembly factor TIMMDC1, mitochondrial	TIMMDC1
161	Q9NR16	Scavenger receptor cysteine-rich type 1 protein M160	CD163L1
162	Q9NTJ4	Alpha-mannosidase 2C1	MAN2C1
163	Q9NV79	Protein-L-isoaspartate O-methyltransferase domain-containing protein 2	PCMTD2
164	Q9NYL5	24-hydroxycholesterol 7-alpha-hydroxylase	CYP39A1
165	Q9NYW8	RB-associated KRAB zinc finger protein	RBAK
166	Q9P000	COMM domain-containing protein 9	COMMD9
167	Q9P2E5	Chondroitin sulfate glucuronyltransferase	CHPF2
168	Q9UBS9	SUN domain-containing ossification factor	SUCO
169	Q9UBZ9	DNA repair protein REV1	REV1
170	Q9UGC7	Peptide chain release factor 1-like, mitochondrial	MTRF1L
171	Q9UIA9	Exportin-7	XPO7
172	Q9UKC9	F-box/LRR-repeat protein 2	FBXL2
173	Q9UKL2	Olfactory receptor 52A1	OR52A1
174	Q9UKY0	Prion-like protein doppel	PRND
175	Q9UL36	Zinc finger protein 236	ZNF236
176	Q9ULU8	Calcium-dependent secretion activator 1	CADPS
177	Q9UNH7	Sorting nexin-6;Sorting nexin-6, N-terminally processed	SNX6
178	Q9Y243	RAC-gamma serine/threonine-protein kinase	AKT3
179	Q9Y4X3	C-C motif chemokine 27	CCL27
180	Q9Y535	DNA-directed RNA polymerase III subunit RPC8	POLR3H
181	Q9Y5K6	CD2-associated protein	CD2AP

Human proteins shared between Stage II and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	P15291	Beta-1,4-galactosyltransferase 1;Lactose synthase A protein;N-acetyllactosamine synthase;Beta-N-acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase;Beta-N-acetylglucosaminyl-glycolipid beta-1,4-galactosyltransferase;Processed beta-1,4-galactosyltransferase 1	B4GALT1
2	Q9UKY3;P23141	Putative inactive carboxylesterase 4;Liver carboxylesterase 1	CES1P1;CES1
3	P31629	Transcription factor HIVEP2	HIVEP2
4	P46976	Glycogenin-1	GYG1
5	Q14246	EGF-like module-containing mucin-like hormone receptor-like 1	EMR1
6	Q29980	MHC class I polypeptide-related sequence B	MICB
7	Q8N8A8	Protein FAM169B	FAM169B
8	Q8WVF1	Protein OSCP1	OSCP1
9	Q96HA4	Uncharacterized protein C1orf159	C1orf159
10	Q99519	Sialidase-1	NEU1
11	Q99996	A-kinase anchor protein 9	AKAP9
12	Q9BYD1	39S ribosomal protein L13, mitochondrial	MRPL13
13	Q9BZW5	Transmembrane 6 superfamily member 1	TM6SF1
14	Q9GZS1	DNA-directed RNA polymerase I subunit RPA49	POLR1E
15	Q9UFC0	Leucine-rich repeat and WD repeat-containing protein 1	LRWD1
16	Q9UL15	BAG family molecular chaperone regulator 5	BAG5

Human proteins shared between Stage III and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	O60674	Tyrosine-protein kinase JAK2	JAK2
2	O95490	Latrophilin-2	LPHN2
3	P09917	Arachidonate 5-lipoxygenase	ALOX5
4	Q14162	Scavenger receptor class F member 1	SCARF1
5	Q15021	Condensin complex subunit 1	NCAPD2
6	Q4G163	F-box only protein 43	FBXO43
7	Q5RHP9	Glutamate-rich protein 3	ERICH3
8	Q68DK2	Zinc finger FYVE domain-containing protein 26	ZFYVE26
9	Q7Z408	CUB and sushi domain-containing protein 2	CSMD2
10	Q7Z7G8	Vacuolar protein sorting-associated protein 13B	VPS13B
11	Q86WZ6	Zinc finger protein 227	ZNF227
12	Q8IWS0	PHD finger protein 6	PHF6
13	Q8N8R3	Mitochondrial basic amino acids transporter	SLC25A29
14	Q8NB90	Spermatogenesis-associated protein 5	SPATA5
15	Q99536	Synaptic vesicle membrane protein VAT-1 homolog	VAT1
16	Q9BXB7	Spermatogenesis-associated protein 16	SPATA16
17	Q9H867	Protein-lysine methyltransferase METTL21D	VCPKMT
18	Q9P2L0	WD repeat-containing protein 35	WDR35

Human proteins shared between Stage I, Stage III and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	Q2V2M9	FH1/FH2 domain-containing protein 3	FHOD3
2	Q9H5Q4	Dimethyladenosine transferase 2, mitochondrial	TFB2M
3	Q9HCK4	Roundabout homolog 2	ROBO2
4	Q9NVE5	Ubiquitin carboxyl-terminal hydrolase 40	USP40

Human proteins shared between Stage I, Stage II and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	P02768	Serum albumin	ALB
2	O14653	Golgi SNAP receptor complex member 2	GOSR2
3	O15229	Kynurenine 3-monooxygenase	KMO
4	O75030	Microphthalmia-associated transcription factor	MITF
5	O75891	Cytosolic 10-formyltetrahydrofolate dehydrogenase	ALDH1L1
6	O95631	Netrin-1	NTN1
7	P01876;P0DOX2; P01877	Ig alpha-1 chain C region;Ig alpha-2 chain C region	IGHA1;IGHA2
8	P05455	Lupus La protein	SSB
9	P20073	Annexin A7	ANXA7
10	Q16825	Tyrosine-protein phosphatase non-receptor type 21	PTPN21
11	Q4G0A6	Protein FAM188B	FAM188B
12	Q6PCB5	Round spermatid basic protein 1-like protein	RSBN1L
13	Q8NI51	Transcriptional repressor CTCFL	CTCFL
14	Q8TCD5	5(3)-deoxyribonucleotidase, cytosolic type	NT5C
15	Q8TEX9	Importin-4	IPO4
16	Q96PE2	Rho guanine nucleotide exchange factor 17	ARHGEF17
17	Q99972	Myocilin;Myocilin, N-terminal fragment;Myocilin, C-terminal fragment	MYOC
18	Q9BWT3	Poly(A) polymerase gamma	PAPOLG
19	Q9UBG0	C-type mannose receptor 2	MRC2
20	Q9Y2P7	Zinc finger protein 256	ZNF256

Human proteins shared between Stage I, Stage II and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	Q9HCM1	Uncharacterized protein KIAA1551	KIAA1551

Human proteins shared between Stage II, Stage III and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Gene names
1	A2VDJ0	Transmembrane protein 131-like	KIAA0922
2	A6QL64	Ankyrin repeat domain-containing protein 36A	ANKRD36
3	O94850	Dendrin	DDN
4	P35670	Copper-transporting ATPase 2;WND/140 kDa	ATP7B
5	P68871;P69892; P69891;P02100	Hemoglobin subunit beta;LVV-hemorphin-7;Spinorphin	HBB
6	Q8N323	NXPE family member 1	NXPE1
7	Q8TE76	MORC family CW-type zinc finger protein 4	MORC4
8	Q96Q35	Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 12 protein	ALS2CR12

APPENDIX T

EXCLUSIVE MICROBIAL PROTEIN EXPRESSIONS BASED ON CRC STAGING

Microbial proteins exclusively identified in Stage II colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	A0Q072	Lipoprotein signal peptidase	Clostridium novyi	lspA
2	A1AZ46	Transcription elongation factor GreA	Paracoccus denitrificans	greA
3	A5DCB6	Trimethyllysine dioxygenase	Meyerozyma guilliermondii	PGUG_00921
4	A5DFC2	U3 small nucleolar RNA-associated protein 25	Meyerozyma guilliermondii	UTP25
5	A6L3H3	Ribonuclease Z	Bacteroides vulgatus	rnz
6	A6LSV5	Lon protease	Clostridium beijerinckii	lon
7	A7Z962	Imidazole glycerol phosphate synthase subunit HisF	Bacillus velezensis	hisF
8	Q7CQ01;Q7AMH5	Chaperone protein ClpB	Salmonella typhimurium, Salmonella typhi	clpB
9	P30773	Type-2 restriction enzyme XcyI	Xanthomonas campestris pv. cyanopsidis	xcyIR
10	P38780;P38781	Uncharacterized protein YHR054C; Chromatin structure-remodeling complex protein RSC30	Saccharomyces cerevisiae	YHR054C
11	Q4A604	ATP synthase subunit beta	Mycoplasma synoviae	atpD

to be continued…

…continuation

12	Q5WED8	Argininosuccinate synthase	Bacillus clausii	argG
13	Q661B6	Ribonuclease Y	Borreliella bavariensis	rny
14	Q96WW5	Conserved oligomeric Golgi complex subunit 8	Schizosaccharomyces pombe	cog8
15	Q7VFU4	Phosphomethylpyrimidine synthase	Helicobacter hepaticus	thiC
16	Q92GY7;Q4UMQ8;C4K2F8; C3PP86;B0BUN9;A8GT48; A8GPC9	Adenylate kinase	Rickettsia conorii, Rickettsia felis, Rickettsia peacockii, Rickettsia africae, Rickettsia rickettsia, Rickettsia rickettsia, Rickettsia akari	adk
17	Q0P9C7	N-acetylgalactosamine-N,N'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	Campylobacter jejuni subsp. jejuni serotype O:2	pglJ
18	Q8FRV5	NAD-dependent protein deacylase 2	Corynebacterium efficiens	cobB2
19	Q8K9B0	RecBCD enzyme subunit RecC	Buchnera aphidicola subsp. Schizaphis graminum	recC
20	Q9PPV5	DegV domain-containing protein UU535	Ureaplasma parvum serovar 3	UU535

Microbial proteins exclusive in Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	A2R6G7; A0A1R3RGJ2	Ochratoxin halogenase ota5; Ochratoxin halogenase	Aspergillus niger; Aspergillus carbonarius	ota5; OTAhal
2	Q66E18;B2K631;B1JIK5; A9R2Z7;A4TPN1;A1JP11	Enolase-phosphatase E1	Yersinia pseudotuberculosis serotype I, Yersinia pseudotuberculosis serotype IB, Yersinia pseudotuberculosis serotype O:3, Yersinia pestis bv. Antiqua, Yersinia pestis, Yersinia enterocolitica serotype O:8 / biotype 1B	mtnC
3	A1SAS5	Dihydroxy-acid dehydratase	Shewanella amazonensis	ilvD
4	B3H035;A3MYR5	3-dehydroquinate synthase	Actinobacillus pleuropneumoniae serotype 7 Actinobacillus pleuropneumoniae serotype 5b	aroB
5	A4J541	Queuine tRNA-ribosyltransferase	Desulfotomaculum reducens	tgt
6	A4WDV5	Adenylyl-sulfate kinase	Enterobacter sp.	cysC
7	Q31V53;Q3YVV0;B2U560; Q8FCE3;Q7A9X4;Q0TBN7; P00944;C4ZXF6;B7ULC6; B7NP65;B7NEL7;B7N1L9; B7MES1;B7M3I8;B7L6Y0; B6I3D6;B5YVL8;B1X8I1; B1LJC7;B1IZM7;A8A623; A7ZTB2	Xylose isomerase	Shigella boydii serotype 4; Shigella sonnei; Shigella boydii serotype	xylA
8	A5DWJ3	Glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial	Lodderomyces elongisporus	PET112
9	A6L9S2	Threonine--tRNA ligase	Parabacteroides distasonis	thrS
10	B2IEZ6	Biotin synthase	Beijerinckia indica subsp. indica	bioB
11	A7ZBZ7	Glycerol-3-phosphate acyltransferase	Campylobacter concisus	plsY
12	C1DLS8	tRNA/tmRNA (uracil-C(5))-methyltransferase	Azotobacter vinelandii	trmA
13	C5E2S6	Suppressor of hydroxyurea sensitivity protein 2	Lachancea thermotolerans	SHU2

to be continued…

…continuation

14	A5DQP9;Q9UVX4;Q9UVW9;P60011;P60009;P53455;P48465;P20359;P17128;P02578;Q9UVZ8;Q75D00;P14235;P60010;P10989	Actin	Meyerozyma guilliermondii, Coprinopsis cinerea, Acremonium chrysogenum (Cephalosporium acremonium), Saccharomyces bayanus; Candida glabrata, Ajellomyces capsulatus, Cryptococcus neoformans var. grubii serotype A, Emericella nidulans, Kluyveromyces lactis, Acanthamoeba castellanii, Candida dubliniensis, Ashbya gossypii, Candida albicans, Saccharomyces cerevisiae, Schizosaccharomyces pombe	ACT1;act1
15	D3XDD3	Altered inheritance of mitochondria protein 32	Saccharomyces kudriavzevii	AIM32
16	O13710	Structural maintenance of chromosomes protein 5	Schizosaccharomyces pombe	smc5
17	O28383	Iron-sulfur flavoprotein AF_1896	Archaeoglobus fulgidus	AF_1896
18	O96185	Uncharacterized protein PFB0460c	Plasmodium falciparum	PFB0460c
19	P00175	Cytochrome b2, mitochondrial	Saccharomyces cerevisiae	CYB2
20	O31870	SPBc2 prophage-derived uncharacterized protein YotE	Bacillus subtilis	yotE
21	O42948	Uncharacterized WD repeat-containing protein C16H5.13	Schizosaccharomyces pombe	SPBC16H5.13
22	O83746	ATP-dependent zinc metalloprotease FtsH	Treponema pallidum	ftsH
23	P43701;P13364	DNA gyrase subunit B	Haemophilus influenza; Pseudomonas putida	gyrB
24	P36947	Ribose import ATP-binding protein RbsA	Bacillus subtilis	rbsA
25	P40131	Flagella basal body P-ring formation protein FlgA	Salmonella typhimurium	flgA
26	P45919	Uncharacterized protein YqbC	Bacillus subtilis	yqbC
27	Q02455	Protein MLP1	Saccharomyces cerevisiae	MLP1
28	P47074	Spindle assembly checkpoint component MAD3	Saccharomyces cerevisiae	MAD3
29	P9WJF7;P9WJF6	CRISPR type III-associated RAMP protein Csm4	Mycobacterium tuberculosis, Mycobacterium tuberculosis	csm4

to be continued…

...continuation

30	Q2JL43	60 kDa chaperonin 1	Synechococcus sp.	groL1
31	Q52C47	AdoMet-dependent rRNA methyltransferase SPB1	Magnaporthe oryzae	SPB1
32	Q59LF3	Regulator of cytoskeleton and endocytosis RVS167	Candida albicans	RVS167
33	Q59LV8	Activator of stress genes protein 1	Candida albicans	ASG1
34	Q5ANB2	ATP-dependent RNA helicase DBP10	Candida albicans	DBP10
35	Q5L545	Lysine--tRNA ligase	Chlamydia abortus	lysS
36	Q9PPZ6	DNA primase	Ureaplasma parvum serovar 3	dnaG
37	Q5M5J8;Q5M111	Valine--tRNA ligase	Streptococcus thermophilus, Streptococcus thermophilus	valS
38	Q6CFJ2	Topoisomerase 1-associated factor 1	Yarrowia lipolytica	TOF1
39	Q6CNM8	KLULA Exocyst complex component EXO84	Kluyveromyces lactis	EXO84
40	Q6FS20	D-arabinono-1,4-lactone oxidase	Candida glabrata	ALO1
41	Q6FT60	Required for respiratory growth protein 1, mitochondrial	Candida glabrata	RRG1
42	Q741E5	UvrABC system protein C	Mycobacterium paratuberculosis	uvrC
43	Q8RHX2	3-keto-5-aminohexanoate cleavage enzyme	Fusobacterium nucleatum subsp. nucleatum	kce
44	Q8RHY5	Glutamate mutase epsilon subunit	Fusobacterium nucleatum subsp. nucleatum	glmE
45	Q9UUL2	DNA repair protein rhp57	DNA repair protein rhp57 OS=Schizosaccharomyces pombe	rhp57
46	Q9Y7N0	SWR1 complex bromodomain subunit bdf1	Schizosaccharomyces pombe	bdf1
47	Q9Y899	Calcium/calmodulin-dependent protein kinase cmkB	Emericella nidulans	cmkB

Microbial proteins exclusive in Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	B6JL62	Glutamate--tRNA ligase 1	Helicobacter pylori	gltX1
2	P0C5M1	Uncharacterized protein YDR371C-A	Saccharomyces cerevisiae	YDR371C-A
3	Q8PJ48;Q3BRN7;Q2P0W3; Q8P7U0;Q5GXU0;Q4UWB0	NADH-quinone oxidoreductase subunit I	Xanthomonas axonopodis pv. Citri, Xanthomonas campestris pv. vesicatoria	nuoI
4	Q4A5S8	DNA-directed RNA polymerase subunit beta	Mycoplasma synoviae	rpoC
5	Q5KWV9	5-methylthioadenosine/S-adenosylhomocysteine nucleosidase	Geobacillus kaustophilus	mtnN
6	Q601M6	tRNA(Ile)-lysidine synthase	Mycoplasma hyopneumoniae	tilS
7	Q7VJ09	Threonine--tRNA ligase	Helicobacter hepaticus	thrS
8	Q97LX0	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Clostridium acetobutylicum	ispF
9	Q9ZE54	Uncharacterized protein RP093	Rickettsia prowazekii	RP093

Microbial proteins shared between Stage I and Stage II in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	Q12152	Putative serine/threonine-protein kinase YPL150W	Saccharomyces cerevisiae	YPL150W
2	Q2UQW3	ATP-dependent RNA helicase rok1	Aspergillus oryzae	rok1

Microbial proteins shared between Stage I and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	B8I492	Phosphoribosylformylglycinamidine cyclo-ligase	Clostridium cellulolyticum	purM

Microbial proteins shared between Stage I and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	Q58106	Uncharacterized protein MJ0695	Methanocaldococcus jannaschii	MJ0695

Microbial proteins shared between Stage II and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	Q9PIS3;Q5HWW3; A8FK07;A7H1S0; A1VXT5;B9KEB1	Threonine--tRNA ligase	Campylobacter jejuni subsp. jejuni serotype O:2, Campylobacter jejuni, Campylobacter jejuni subsp. jejuni serotype O:6, Campylobacter jejuni subsp. doylei, Campylobacter jejuni subsp. jejuni serotype O:23/36, Campylobacter lari	thrS
2	C3L0F7;C1FSN0; B1KWP9;B1II90;A 7GG48;A7FW24;A 5I4N7	Acetate kinase	Clostridium botulinum, Clostridium botulinum, Clostridium botulinum, Clostridium botulinum, Clostridium botulinum, Clostridium botulinum, Clostridium botulinum)	ackA
3	Q64P83;Q5L923; A6KXL2	Triosephosphate isomerase	Bacteroides fragilis, Bacteroides fragilis, Bacteroides vulgatus	tpiA
4	Q8PLC6;Q3BUB6; Q2P1U6;B2SV97	Lysine--tRNA ligase	Xanthomonas axonopodis pv. Citri, Xanthomonas campestris pv. vesicatoria, Xanthomonas oryzae pv. Oryzae, Xanthomonas oryzae pv. oryzae	lysS
5	P32617	Chromatin-remodeling complex subunit IES6	Saccharomyces cerevisiae	IES6
6	Q72JT2;Q5SJG0	UPF0365 protein TT_C0686	Thermus thermophilus, Thermus thermophilus	TT_C0686; TTHA1048
7	Q6C2Q7	Nucleolar protein 12	Yarrowia lipolytica	NOP12
8	Q9K9Y6	Methionyl-tRNA formyltransferase	Bacillus halodurans	fmt
9	Q9PPV1	Energy-coupling factor transporter ATP-binding protein EcfA1	Ureaplasma parvum serovar 3	ecfA1

Microbial proteins shared between Stage III and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	P47077;C8ZBK1;C7GV42;B3LQ92;A6ZPV9	Nucleolar protein 9	Saccharomyces cerevisiae, Saccharomyces cerevisiae, Saccharomyces cerevisiae, Saccharomyces cerevisiae, Saccharomyces cerevisiae	NOP9
2	P36000	AP-1 complex subunit beta-1	Saccharomyces cerevisiae	APL2
3	Q8PZ26	Ketol-acid reductoisomerase (NADP(+))	Methanosarcina mazei	ilvC

Microbial proteins shared between Stage II and Stage IV in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	P53937	37S ribosomal protein SWS2, mitochondrial	Saccharomyces cerevisiae	SWS2

Microbial proteins shared between Stage I, Stage II and Stage III in colorectal cancer (CRC)

No.	Protein IDs	Protein names	Microbial species	Gene names
1	Q758C9	Mitochondrial import inner membrane translocase subunit TIM54	Ashbya gossypii (strain ATCC 10895 / CBS 109.51 / FGSC 9923 / NRRL Y-1056)	TIM54