PLASMA MicroRNAs AS A BIOMARKER FOR DETECTION OF COLORECTAL CANCER IN HOSPITAL UNIVERSITI SAINS MALAYSIA

SYARAH SYAMIMI BINTI MOHAMED

UNIVERSITI SAINS MALAYSIA

2023

PLASMA MicroRNAs AS A BIOMARKER FOR DETECTION OF COLORECTAL CANCER IN HOSPITAL UNIVERSITI SAINS MALAYSIA

by

SYARAH SYAMIMI BINTI MOHAMED

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

September 2023

ACKNOWLEDGEMENT

Finally, it comes to this little masterpiece. I am grateful and blessed that Allah has assisted me along the way. He gave me the courage and determination to complete this adventurous journey. I want to thank Allah for bestowing me with a support system for the people around me. I received unending support from my parents, Mr Mohamed bin Junoh and Mrs Hurul Atini binti Ab Manan, as well as my siblings (Najjwa Nidana and Alif Akram), who made significant sacrifices to allow me to pursue my studies and patience that restored me. All of my supervisors, Professor Dr. Zaidi bin Zakaria, Professor Dr. Andee Dzulkarnaen bin Zakaria, Associate Professor Dr. Md Salzihan bin Md Salleh, and Dr. Tengku Ahmad Damitri Al-Astani bin Tengku Din, who have always been my mentors. Furthermore, I would like to thank all of my postgraduate colleagues in the Pathology Postgraduates Room from 2019 to 2022 for assisting me and the Pathology staff for always being kind towards me and my friend. I also want to thank En Mad bin Ros, Science Officer of USM's Genome Department, for teaching me some fundamental and technical skills in the microarray technique. Dr. Ahmad Aizat bin Abdul Aziz and Dr. Shazana Hilda binti Shamsuddin willingly came to my research lab freely to observe my qPCR process and showed me how to analyze the results thoroughly. Next, Associate Professor Dr. Nurul Syakima binti Ab Mutalib from UMBI, UKM Cheras, is eager to spend her time and provide feedback on my study idea. Not to mention Associate Professor Dr. Ariffin bin Kaderi, one of our research team, and Dr. Azmir bin Ahmad from IIUM Kuantan, whom I view as my field mentor who has guided me and been patient with me anytime I bombarded him with questions. Thanks!.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

µL	Microlitre
mL	Millilitre
rpm	Rate per minute
℃	Degree celsius

LIST OF ABBREVIATIONS

AUC	Area under the curve
EDTA	Ethylenediaminetetraacetic acid
CRC	Colorectal cancer
DEM	Differential expressed miRNA
MiRNA	MicroRNA
ROC	Receiver operating characteristic
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction

PLASMA MikroRNA SEBAGAI PENANDA BIO BAGI MENGESAN KANSER USUS DI HOSPITAL UNIVERSITI SAINS MALAYSIA

ABSTRAK

Kanser kolorektal (CRC) merupakan salah satu penyakit dengan kadar survival yang rendah dan kebiasaannya berakhir dengan kematian. Mutakhir ini, kaedah kolonoskopi merupakan kaedah piawaian standard dalam mengesan penyakit kolorektal. Walaubagaimanapun, disebabkan prosedur yang kurang selesa, kaedah ini kurang menjadi pilihan ramai. Semenjak penemuannya, mikroRNA (miRNA) telah menjadi perintis bagi era kaedah diagnosis menggunakan molekul. MiRNA adalah molekul regulator pendek pasca-transkripsi yang mensasarkan RNA pengutus (mRNA), mengakibatkan proses degradasi protein atau mengubah translasi protein. Pengekspresan miRNA didapati terderegulasi dalam kanser tisu kolorektal. Kehadiran miRNA dalam plasma dapat menyediakan alat saringan invasif yang minima untuk mengesan penyakit serta dapat membantu mengesan penyakit di tahap awal. Meskipun banyak miRNA telah diprofil dalam CRC, namun masih ada kekaburan bagi miRNA yang digunakan untuk mengesan CRC secara tepat. Oleh itu, kajian tentang pengekpresan miRNA telah dijalankan mengunakan populasi kajian dari Hospital Universiti Sains Malaysia (HUSM) untuk mencari potensi miRNA dalam sistem peredaran sebagai penanda-bio bagi saringan CRC. Senarai miRNA yang signifikan telah diprofil dari tisu CRC dengan membandingkan pengekpresan mereka terhadap tisu normal. Beberapa miRNA yang terkawal dari tisu (hsa-miR-20a-5p, hsa-miR-21-5p, and hsa-miR-210-3p) telah disahkan dalam plasma mereka serta dalam beberapa kumpulan kajian mandiri yang lain (termasuk kumpulan CRC dan individu sihat yang lain). Secara relatifnya, antara tiga miRNA yang dipilih ini, hsa-miR-21-5p (FC=3.87)

mempunyai perubahan lipatan (FC) dalam plasma kumpulan CRC yang paling tinggi berbanding individu sihat. Kemudian diikuti dengan hsa-miR-20a-5p (FC=1.23) dan hsa-miR-210-3p (FC=-0.21). Walaubagaimanapun, pengekpresan hsa-miR-210-3p dalam plasma didapati tidak konsisten berbanding tisu. Menurut analisis korelasi Spearman, kami mendapati hsa-miR-21-5p mempunyai korelasi positif yang terkuat (r=0.84) berbanding hsa-miR-20a-5p (r=0.20). Sementara itu, hsa-miR-210-3p menunjukan korelasi negative antara tisu dan plasma (r=-0.75). Akhir sekali, kami menilai prestasi diagnosis miRNA yang terpilih secara individu dan juga kombinan menggunakan analisis lengkuk Receiver Operating Charateristics (ROC) berdasarkan peringkat penyakit CRC. Hasil dapatan analisis mendapati, hsa-miR-20a-5p (AUC: 0.82, 86% kepekaan dan 88% kespesifikan) berpotensi sebagai panel penanda-bio bagi mengesan CRC ditahap awal, manakala hsa-miR-21-5p dan hsa-miR-210-3p (AUC: 1.0, 100% kepekaan dan kespesifikan bagi kedua-dua panel) berpotensi sebagai penanda-bio bagi mengesan CRC ditahap akhir. Manakala, kombinasi ketiga-tiga panel miRNA memberikan kuasa diskriminasi (AUC=0.98) antara pesakit CRC dan individu sihat dengan kadar kepekaan 96% dan kespesifikan 75%. Sementara itu, kuasa diskriminasi bagi kombinasi dua panel miRNA pula, AUC=0.97 bagi kombinasi hsa-miR-21-5p dan hsa-miR-210-3p, dan AUC= 0.91 bagi kombinasi hsa-miR-20a-5p dan hsa-miR-21-5p. Kajian ini mendapati tiga pengedar miRNA, hsa-20a-5p, hsa-miR-21-5p, dan hsa-210-3p, yang berpotensi menjadi penanda-bio invasif yang minima untuk saringan CRC. Hsa-miR-21-5p dan has-miR-210-3p mempunyai korelasi positif dan negatif yang kuat antara tisu dan plasma. Kesimpulannya, mengikut prestasi individu, hsa-miR-21-5p mempunyai kuasa diskriminatif tertinggi sementara jika mengikut prestasi kombinan, kombinasi ketiga-tiga panel miRNA memberikan kombinasi terbaik dengan kuasa diskriminatif tertinggi.

PLASMA MicroRNAs AS A BIOMARKER FOR DETECTION OF COLORECTAL CANCER IN HOSPITAL UNIVERSITI SAINS MALAYSIA

ABSTRACT

Colorectal cancer (CRC) is one of the cancer with a low survival rate and usually end with death. Colonoscopy is currently the gold standard in detecting CRC disease. However, due to the inconvenience of the procedure, it becomes less appealing. Since its discovery, microRNA (miRNA) has pioneered a new era of molecular diagnosis. MiRNA is a short post-transcriptional regulatory that targets messenger RNA (mRNA), either causing degradation or altering the protein translation. MiRNA expression is dysregulated in CRC tissue. The presence of miRNAs in plasma is thought to provide non-invasive screening tools for detecting disease and may aid early detection. Even though many miRNAs have been profiled in CRC, there is still ambiguity on the robustness of the reported miRNAs in detecting CRC. Thus, a miRNA expression study was conducted using a population sample in Hospital Universiti Sains Malaysia to search for potential circulating miRNAs as CRC screening biomarkers. The significant miRNA profile in CRC tissue was obtained by comparing their expression to its adjacent normal tissue. A few upregulated miRNAs from tissue (hsa-miR-20a-5p, hsa-miR-21-5p, and hsa-miR-210-3p) were validated in their plasma and other independent study groups (CRC and healthy individuals). Relatively, among these three miRNAs, hsa-miR-21-5p (FC=3.87) has the highest fold change (FC) in the plasma of CRC patients compared to healthy individuals, followed by hsa-miR-20a-5p (FC=1.23) and hsa-miR-210-3p (FC=-0.21). However, expression of hsa-miR-210-3p in plasma was inconsistent as compared to tissue. According to Spearman correlation analysis, hsa-miR-21-5p has a strong positive correlation

(r=0.84) compared to hsa-miR-20a-5p (r=0.20). Meanwhile, hsa-miR-210-3p showed a negative correlation (r=-0.75) between tissue and plasma. Lastly, we evaluated the diagnostic performance of individual and combined miRNAs using receiver operating characteristics curve analysis according to stages. Our studies revealed that hsa-miR-20a-5p (AUC: 0.82, 86% sensitivity and 88% specificity) potentially act as a biomarker panel for early CRC, whereas both hsa-miR-21-5p and hsa-miR-210-3p (AUC: 1.0, 100% sensitivity and specificity simultaneously) are for advanced CRC. Combining all three miRNAs also gives good discriminative power (AUC:0.98) between CRC patients and healthy individuals with 96% sensitivity and 75% specificity. Meanwhile, the combination of two miRNAs, namely hsa-miR-21-5p and hsa-miR-210-3p, and hsa-miR-20a-5p and hsa-miR-21-5p, resulting in discrimination power of 0.97 and 0.91, respectively. The study discovered three circulating miRNAs, hsa-20a-5p, hsa-miR-21-5p, and hsa-210-3p, which may be minimally invasive diagnostic biomarkers. Hsa-miR-21-5p and hsa-miR-210-3p have strong positive and negative correlations with their tissue counterparts. In conclusion, individual hsa-miR-21-5p has the highest discriminative power, while combined miRNAs showed that combining all three was the best combination with the highest discriminatory power.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Data shown by the Global Cancer Observatory (GLOBOCAN), a platform that provides statistics globally, colorectal cancer (CRC) is the third leading cancer cause of death (Asia et al., 2019). In the United States, the American Cancer Society reported that in 2020, the CRC individual incidence number reached 147 950. One-third of the cases would end in death for individuals aged below 50 years (Siegel et al., 2020). Meanwhile, in Malaysia, CRC is one of the ten most common cancers with incidence rates of 13.5% over the 100 000 population (Abu Hassan et al., 2016).

For the last 5 years, males face higher lifetime risk than females (Abu Hassan et al., 2016). Most CRC patients are diagnosed at late stages, possibly contributing to mortality (Azizah et al., 2019). The large number of CRC incidences and mortality has prompted researchers and physicians to develop a screening strategy to identify the early stages of CRC, which may aid survival. This would not be easy when individuals are identified at an advanced age and advancements in health technology cannot assist owing to various circumstances such as comorbidity, immunological level, and treatment (Allemani et al., 2013).

There are several tools available that are currently used for screening for CRC. Some are invasive, such as fecal occult blood test, endoscopy, and colonoscopy (Helsingen et al., 2019). Each of the tools may have its limitations to access. Apart from the invasive issue, some of the screening tests are expensive, making them not accessible to everyone (Zhang et al., 2019). Hence, developing a new screening tool that aims for early detection of CRC is needed. Ideally, this new method must be accurate, minimally invasive, and not labour intensive. The study of circulating

microRNA (miRNA) has recently emerged as a new focus of research interest, giving a minimally invasive diagnostic tool for biomarker research (Masuda et al., 2017).

Genome profiling has demonstrated that the expression of microRNAs plays an essential role in carcinogenesis and progression since it is associated with the type and grade of tumor and a variety of clinical outcomes. Since then, they have been designated potential biomarkers and therapeutic target candidates. MiRNAs simultaneously target several genes and regulate multiple signaling pathways that promote cell differentiation, proliferation, and survival (Pereira et al., 2013; Peng and Croce, 2016). The robustness of miRNAs has impacted several methods for identifying further evidence on miRNA analysis over the last two decades. Transcriptome study has shown the targeted network of miRNAs. Various omics studies also contribute to an abundance of data expression, which contributes to identifying miRNAs for specific pathways (Vafaee et al., 2018). Microarray or sequencing research on different cell types, tissues, and disease states has aided in developing cell-by-cell data on miRNA effects (Lai, 2015). The tissue-specific miRNAs identified by the sequencing study also serve as a criterion for the reliability of the specificity of biomarkers (Friedländer et al., 2014; Kryuchkova-Mostacci and Robinson-Rechavi, 2017).

MiRNAs are conserved across biological processes, either normal physiological or pathological, such as carcinogenesis (Li & Zhang, 2019). Since then, many researchers have found that CRC-miRNA might help predict CRC biomarkers from various research approaches (Sabry et al., 2019; Zhang et al., 2019). The study of miRNA regulation in plasma and serum has been initiated since their first discovery by Mitchell and team (2008). Meister (2013) stated that extracellular miRNAs were shielded from RNases, keeping them stable as they were released from cells through the microvesicle, exosomes or bound to Argonaut proteins. Because of their abundance

in other physiological fluids, such as blood, saliva, and urine, makes them conveniently available without tedious procedures (Mitchell et al., 2008; Cortez et al., 2011).

Fundamentally, miRNAs have been known to be cancer causes. They could act as oncoMir or tumor suppressor Mirs (TS-Mir). Therefore, the dysregulation of miRNAs pathologically was disclosed with carcinogenesis (Svoronos et al., 2016). The discovery of circulating miRNAs in body fluid has laid the foundation for minimally invasive biomarker research related to miRNAs. The differential expression of circulating miRNAs in CRC and healthy individuals were reported in various study (Wang et al., 2015). Canonically, miRNA transport is being proposed for the study of minimally invasive biomarkers because it is assumed that miRNAs as one of the causes of cancer development and are abundant in body fluids (Wang & Chen, 2014). However, there is an argument on the source and function of dysregulated miRNAs. According to Carter and team (2017), miRNAs that are dysregulated in a particular disease could not be easily pointed out as solid biomarkers unless there is a value for comparing dysregulated miRNAs in healthy samples.

1.2 Problem statement

Despite having several screening tools, the case occurrence is still out of control and unpredictable. Even though colonoscopy may be chosen because of its high sensitivity in detecting and removing colorectal polyps, a concern regarding this invasive detection is the cost of treatment, and patient bowel preparation may cause poor compliance. Colonoscopy diagnosis is only reliable with the presence of an experienced expert. Therefore, this method is unsuitable for mass screening (Zhang et al., 2019). In addition, many individuals are hesitant to undergo a colonoscopy due to embarrassment, time-consuming preparation, fear of pain, and discomfort due to the

procedure’s invasiveness. This approach is also unable to cover the entire colon for the diagnosis (Arnold et al., 2017).

Aside from that, a radiological approach known as double-contrast (BE) is employed in the general population for CRC screening. However, the clinical literature does not support the conclusion because of the low accuracy and the inability to do a biopsy like a colonoscopy. Following that, fecal occult blood test (FOBT) and carcinoembryonic antigen (CEA) screening provide a simple and inexpensive test, but their low sensitivity and specificity make them less than clinically needed. Thus, it is relatively unsuitable for asymptomatic patients (He et al., 2015).

1.3 The rationale of the study

Despite having many comprehensive instruments, we still need a more ‘friendly tool’ called a specific biomarker. The biomarker must compromise the ideal characteristics to be appointed. A biomarker must be particular to a specific disease, susceptible during screening, easy to assess, minimally invasively towards the patient, and possibly less expensive. Therefore, molecular research is suggested for the progress of the period in the scientific method in detecting CRC and ruling outpatient management.

The overexpression of the CEA was first detected in 1975 by Gold and Freedman (Sidransky, 2002). In CRC cells, this protein expression is used as a serum-based marker. Over the last five years, CEA and Carbohydrate Antigen 19-9 (CA19-9) have been the only known biomarker for gastrointestinal cancer. Both markers’ expression levels are still not significantly sensitive and valid for early CRC-stage detection (Fang et al., 2015).

Thus, it is essential to develop a more sensitive and specific biomarker, such as microRNA (miRNA). Recently, blood screening has been more favourable, as it is expected to be low-cost and minimally invasive compared to other techniques (Carter et al., 2017). No specific molecular biomarker is yet to be established as the sole biomarker in detecting CRC. This discovery could lead to an invention in clinical management for disease screening in the general population (Wang et al., 2015).

Since the early 20s, many researchers started to look for potential miRNAs as biomarker studies. Even though many miRNAs have been profiled, there is still an ambiguous finding on a solely valid biomarker, particularly for CRC (Cortez et al. 2011; Hur 2015; Carter et al. 2017; Condrat et al. 2020). In the case of CRC, despite advances in the field and the discovery of a growing number of miRNAs as potential biomarkers, their translation to the clinic has been hampered in part by inconsistent results from independent research and a lack of uniformity among biomarker panels (Fonseca et al., 2022).

Thus, developing a plasma-based miRNA screening assay is expected to be an alternative strategy for screening CRC that is minimally invasive, low-cost, and accessible to all. MiRNA research is considered to be in its early stages. Despite numerous studies of miRNAs in plasma, there are differences in identifying dysregulated miRNAs that are only associated with CRC. It could be due to the diverse patient profiles recruited or the study design. Therefore, obtaining a specific miRNA profile on CRC may hopefully contribute to fundamental research discoveries.

1.4 Research question

In this study, the research questions are formulated as;

i. What is the significantly differential expression pattern of miRNA in the tumoral tissue compared to normal adjacent tissue of CRC?

ii. Does the presence of the miRNA in the CRC tissue also detect in the blood plasma, and how well do they correlate with each other?

iii. What is the diagnostic performance of circulating miRNAs in predicting CRC cases?

1.5 Objective

1.5.1 General objective

To study the differential expression of circulating miRNAs in CRC and their potential as biomarkers.

1.5.2 Specific objective

1) To determine the significance of different miRNAs profiles between tumor tissue and normal adjacent tissue of CRC correspond to their clinic-pathological characteristics.
2) To examine the correlations between the potential differential miRNA expressed in tissue and plasma of CRC patients.
3) To examine the performance of diagnostic value by specific expressed miRNAs in CRC as potential biomarkers.

1.6 Hypothesis

1) There are significant miRNAs dysregulated in tumor tissue compared to normal tissue.

2) The expressed miRNAs in tissue are also presented in blood plasma within the individual.
3) The miRNAs are expressed differently according to the stages of CRC.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of colorectal cancer (CRC)

Colorectal cancer (CRC) is a type of gastrointestinal cancer that begins as a polyp in the colon or rectum. It is known as colon or rectal cancer, depending on the cancer growth lesion area (*What Is Colorectal Cancer? | How Does Colorectal Cancer Start?*, no date). Because of similarities in biological and clinical features, these two parts are frequently merged (Society, 2017) and are known as CRC. CRCs mostly come from abnormal sequential tissue lining growth in the base of the colonic crypts. An adenomatous tissue evolved from a polyp that started with an aberrant crypt foci (ACF) into a neoplastic lesion. The polyp turns into adenocarcinoma within 10-15 years (Dekker et al., 2019). They assumed that the origin of the CRC cell is likely from the stem cell or stem-cell-like cell underlying colonic crypts. This cell is vital in starting and maintaining the tumor progression (Medema, 2013; Nassar and Blanpain, 2016).

The anatomy of the colon is divided into two lesions: right (comprised of caecum, ascending colon, and hepatic flexure) and left (comprised of splenic flexure, descending colon, sigmoid, and rectosigmoid). On the other hand, the rectum belongs to another embryological origin (Dekker et al., 2019). Refer to Figure 1.1(A). CRC manifestation relies on the regional lesion, where 25% of patients are diagnosed with colon cancer and 18% with rectal cancer. The first morphology that appeared as a cancer sign was hyperplasia, caused by a genetic mutation that failed to control the cell cycle properly. Adenomatous polyp, a benign cancer that may be a precursor to CRC, develops from hyperplasia. When the cell grew uncontrollably and penetrated the lining wall of the colon,

it progressed to the adenocarcinoma stage before frantically metastasizing to all potential parts of the body (Dekker et al., 2019; Ng et al., 2019). Refer to Figure 1.1(B). Figure 1.1 shows (A) the schematic diagram of colon anatomy, including the caecum, ascending colon, transverse colon, descending colon, sigmoid colon, and rectum. These areas are divided into two lesions, the right and left side. Meanwhile, (B) is the initiation and progression of carcinogenesis on the colon wall.

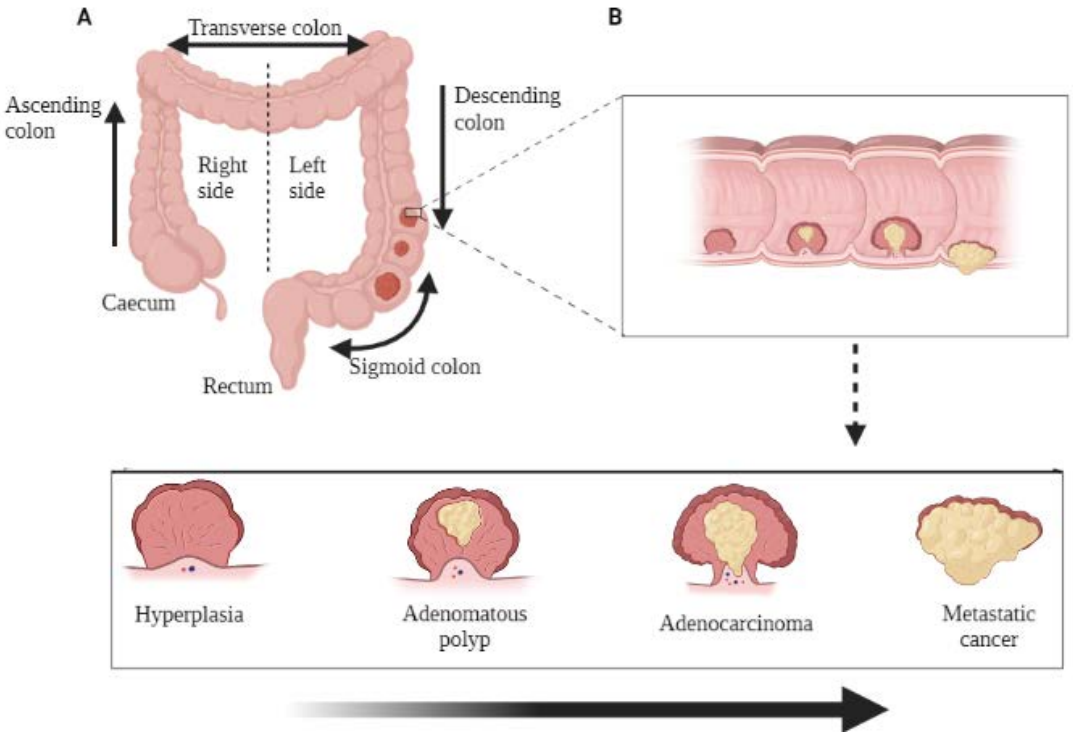

Figure 1. 1 The overview of colorectal cancer localization and the progression

2.1.1 The histological and morphological of CRC

CRC is developed from a polyp, a group of abnormal cells within the intestinal mucosa from a benign to metastasis state, as illustrated in Figure 1.1. The genetic changes in the polyp cell can invade and protrude into the intestinal lumen, which spread to the lymph nodes and gain the hallmark of the cancer state. Adenoma and sessile serrated polyps (SSPs) are two potential polyps related to CRC. These polyps typically can be developed throughout the entire length of the colon and rectum. Theoretically, polyp-tumor development involves a series of histological, morphological, and genetic alterations.

Both adenoma polyp and SSP polyp result from genetic progression that is required or inherited. CRC that developed from the adenoma pathway is from the APC mutation, KRAS oncogene, and the destruction of the p53 function. The serrated path is from the mutated BRAF gene and hyper-methylation of the aberrant gene promoter region. However, another genetic cause that can occur in both two pathways is microsatellite instability (MSI), where the DNA repair genes are disrupted in the germline state (Simon, 2016).

2.1.2 Genetic Predisposition of CRC

Carcinogenesis is developed by altering important gene-protein encoded and malfunctioning epigenetic events that promote apoptosis or angiogenesis, cell adhesion, chromosome instability, and impaired DNA Genetics. CRC is a heterogeneous disease that transforms the normal colonic mucosa to proliferate hyperactively, forming benign adenoma into an invasive tumor that metastasized to all body parts (Vogelstein et al., 1988).

In CRC cases, only a minority (5-7%) is diagnosed to inherit the cancer from the upper generation and about 10-30% have the chance of developing the disease through familial predisposition (Burt, 2000). Accordingly, a CRC patient will likely present a hereditary nonpolyposis colorectal cancer (HNPCC) and familial adenomatous polyposis (FAP) syndromes for congenital cases (Rustgi, 2007). CRC histology is divided into two types based on germline mutation: adenomatous polyposis and non-polyposis structured. APC mutations are frequent in FAP, and STK11 mutation is in hamartomatous polyposis (usually caused Peutz-Jeghers syndrome, PJS). Moreover, Lynch syndrome (LS) is caused by a DNA mismatch repair (MMR) gene mutation in non-polyposis individuals (Stoffel et al., 2015).

De Rosa and team (2015) collectively listed that the heterogeneity of CRCs at the molecular level is due to several mechanisms. For instance, chromosomal instability (CIN), microsatellite instability (MSI), aberrant DNA methylation, and mismatched DNA repair are taking parts, transforming the epithelial cell into an invasive tumor (De Rosa et al., 2015). There are two basic types of genetic-cancer mutation: acquired and germline. The mutated genes may behave against nature, such as inactivating the gene suppressor, promoting the oncogenes, and mismatched DNA repair genes (Cancer.Net, 2015). These mutations may lead to hereditary CRC (De Rosa et al., 2015). According to a review of research conducted by Keum and Giovannucci (2019), the increasing number of CRC cases is primarily due to aetiological genetic impairment in the individual (Mutation in MLH1 and APC germlines) and dominantly present as sporadic diseases. In CRC, altered RAS and RAF are specific mutation genes practically tested in clinical diagnosis (Zarkavelis et al., 2017).

Many studies have attempted to link CRC carcinogenesis to genetic causes. Notable oncogenic mutations include KRAS, EGFR, MYC, ERBBA, etc. While TP53, PTEN, WNT, SMAD2, SMAD4, etc., are gene suppressors and mutational genes from DNA repair genes including bSMH2, bMLH1, bPSM1, bPSM2, and MSH6. Epigenetic modifications, such as changes in DNA methylation status, both global hypo-methylation and promoter hyper-methylation, are also involved in colorectal carcinogenesis. Global hypo-methylation is related to genomic instability in CRC relative to normal colonic tissue. Still, promoter hyper-methylation represses various tumor suppressor genes and their particular pathways (e.g., TP53, PI3K/PTEN/AKT/mTOR, TGFb/SMAD) (Balacescu et al., 2018). The identification of genetic pathways to neoplasia aids in our understanding of how to identify families at risk and provide timely clinical intervention, including chemoprevention (Rozen et al., 2005).

2.1.3 Descriptive epidemiology of CRC

The International Agency for Research on Cancer (IARC) from the World Health Organization (WHO) reported that CRC is the third most frequent disease with malignancy after lung and breast cancer (Asia et al., 2019). Nearly half of all CRC cases result in mortality, particularly in Western countries such as North America, Europe, Australia, and New Zealand (Abu Hassan et al., 2016; Asia et al., 2019). In 2016, the United States recorded 36% of 134 490 cases ending in death (Marley and Nan, 2016). In the next two years, Asians scored the highest incidence cases, with 51.8% over 100 000 population, and half of the patients also end up with death (Onyoh et al., 2019). Japan, Thailand, Iran, Saudi Arabia, the Philippines, China, South Korea, and Singapore are now

facing diseases with a higher frequency (Marley and Nan, 2016). It indicates that this disease has become a global issue without concern about population or country status.

In the last decade, over 1 million individuals have been diagnosed with this disease in both sexes. Research on gender factors showed males face a higher risk of getting the disease than females (Abu Hassan et al., 2016). As per a cross-sectional study conducted in the United Kingdom, males occupied the incidence of statistical cases of CRC based on age-standardized rate (ASR) with a ratio of 1.7:1 towards females. According to the staging analysis, although women had more early-stage diagnoses, men still outnumbered women in average cases. These gender factors are primarily because of differences in disease type, disease location, survival rate, lifestyle adaptation, self-awareness, and screening test results (White et al., 2018). In 2016, Public Health England announced that men were mainly diagnosed when checked by a bowel screening program. Meanwhile, women always come up with emergency presentation cases. Men are vulnerable to developing bowel cancer because they tend to consume a red and processed meat diet, alcohol intake, heavy smokers, and visceral fat (White et al., 2018).

Most diagnosed patients are people over 50 years old, with late-stage and poor prognosis (Lim et al., 2020). The younger (below 50 years old) also present with CRC. Obesity and diabetes are not the only causative factors in the aetiology of young-onset CRC (Weinberg and Marshall, 2019). The familial factor potentially contributes as well. Veettil and team (2017) have analyzed that, recently, the incidence of younger increased by over 2% per year, and most of them were also diagnosed with poor prognosis. The statistical test showed that in the 2003-2005 range, from 8077 patients enrolled, 14.6% were patients younger than 50 years and 7% were below 40 years, a similar trend found back in 1988 (Veettil et al., 2017).

2.1.4 Risk factor

Based on a retrospective study conducted, the researcher concluded that the 5-year survival rate, which corresponds to statistics reported by the National Cancer Patient Registry-Colorectal Cancer (NCPR-CC), two significant factors that attributed to important survival causal are age and stage of the disease (Lim et al., 2020). These concern numbers of individuals are also related to family background, red or processed meat consumption, alcohol intake, heavy smoking, and poor dietary intake (Marley and Nan, 2016). Based on Keum & Giovannucci (2019), the impact of assimilating to Western culture has led to an increase in weight, less participation in physical activity, consumption of alcohol and a smoking habit that raises concerns about the occurrence of the disease.

Another risk factor is genetic inheritance. Those who experienced CRC patients in their lineage may inherit the causative genes. Despite this, the finding on familial cases is still limited and progress. Kampman (2007) reported that those with first-degree relatives face a 2-fold increased risk of inheriting cancer, while those with two or more relative degrees face a 4-fold increased risk. Individuals who have a familial history of CRC, Lynch syndrome, ethnic backgrounds, and underlying type 2 diabetes are classified as a person at risk with non-modifiable factors (Simon, 2016b).

2.1.5 Screening and diagnosis of CRC

Cancer survival data is an essential indicator of the cancer system's ability to detect and treat the disease. The 5-year survival rate cannot be generalised to all patients because the cumulative survival rate varies depending on the type of cancer. Patients with localised cancer had the highest survival rate of 90%, followed by those with regional cancer at 70%, and patients with distant/malignant disease had the lowest survival rate of 15%

(Mattiuzzi et al., 2019). The survival rate varies according to the diagnosis and staging. Those diagnosed with early-stage cancer have a better chance of recovery (Radzi and Dee, 2010).

Veettil and colleagues (2017) reported that Malaysian cases were in the late primary stage. As previously reported by Allemani and colleagues (2013), CRC diagnosed with Dukes Stage C and D in Malaysia is higher than in the United States and Europe. This finding is enough to emphasize the urge to develop an effective screening program, which is thought to reduce the CRC incidence and mortality rate. The disparities in Malaysian survival rates as compared to those of Singapore, the United Kingdom, and the United States are primarily because of a lack of practice guidelines for CRC management, as well as inadequate technologies to meet the growing demand for diagnostics, therapeutics, and care interventions in the country (Veettil et al., 2017).

Pathological observation is commonly used to identify cancer. Pathologists must consider the tumor staging, surgical margin, and prognostic parameters to accurately determine a histopathological diagnosis, including lymphovascular and perineural invasion. Future patient care, prognosis assessment, and family therapy depend on the pathological examination of biopsy, resection specimens, and polypectomy samples (Fleming et al., 2012). CRC diagnosis are currently graded according to TNM staging by the American Joint Cancer Committee (AJCC). This classification, based on the depth of local invasion (T), involvement of lymph node (N), and distance of metastases (M) is combined and characterized as an overall TNM stage. However, in reality, the TNM staging is quite challenging to be related to prognosis due to the heterogeneity of CRC (Gunderson et al., 2010).

Various screening tests exist, like stool-based, imaging, and endoscopic examinations. The rise of different tests is because of some limitations endured in one stool while confirming the accuracy of the test. The fecal occult-based test (FOBT) is a tool to test hem or haemoglobin in patient stool, known as a fecal immunochemical test (FIT). FIT seems to have higher sensitivity and requires no diet restriction compared to FOBT. Then, those who are positive for FIT are usually referred for colonoscopy. Besides, the general practitioner also applied the double-contrast barium enema (DCBE), computed tomographic colonography (CTC), and colon capsule endoscopy (CCE) for the imaging test. These methods need an intensive bowel preparation, but no biopsy can be run. Variation of screening methods helps to reduce and create early awareness of cancer. It also helps to diagnose the disease; thus, patient management can be done earlier (Lim et al., 2020).

Colonoscopy is always the most preferable diagnosis choice as it covers 97-98% of the specificity and sensitivity of the test. Its capability to detect the cancerous lesion in the entire large bowel and the distal part of the small bowel makes it better at work than flexible sigmoidoscopy (FS). However, the restriction of colonoscopy tests is costly and requires expert supervision (Ferlizza et al., 2021). Colonoscopy is prioritized for individuals aged 45 years and above. However, potential youngsters with causative factors such as family history, blood-stool mix, aggressive weight loss, and unusual bowel habits are more prone to undergo colonoscopy tests (Dekker et al., 2019).

A meta-analysis of findings on the effectiveness of disease screening showed that FIT, FS, colonoscopy, and FOBT reduced incidence and mortality with effective rates of 59%, 33%, 61%, and 38%, respectively. By including the data recruited from 1992 to 2016 from 44 types of studies regarding CRC screening, they suggested that colonoscopy

is better at covering CRC from deaths (Zhang et al., 2017). There is no definite tool to be appointed as the best tool. Countries may encounter different strategies based on country-cost provided and patients' adherence to commit (Maida et al., 2017).

An innovation in research tools has proposed using biomarker analysis, such as protein and genetic molecular (DNA or RNA), to screen the disease. Although it seems to need more validation tests, this tool has the potential to be applied in clinical practice because it provides higher sensitivity and less cost (Binefa et al., 2014). Individuals at high risk should consult with their healthcare provider to determine when to screen and how frequently to repeat the screening (Simon, 2016).

Screening using these tools may not be an option because it is expensive, lacks public awareness, and is inaccessible to all, whether in urban or rural areas. Participation in screening programs varied differently across the country. For example, over 50% of Netherlanders chose to join, but only 10% of Canadians were up to it. Decent countries like South Korea and Australia prefer colonoscopy for screening. Based on gender analysis, women have the desire to opt for the screening test compared to men willingly. However, the positive FOBT and FIT results were mainly positive in men (Navarro et al., *2*017).

2.1.6 Patient management and treatment

People with CRC will have hematochezia, occult bleeding, anemia, and fatigue, but tenesmus is only for those with specified rectal cancer. Rectal bleeding is a common symptom for everyone, regardless of the stage (De Rosa et al., 2015). CRC can be considered an asymptomatic disease, as it shows no early signs. When a severe symptom exists, it is already in an advanced stage.

Surgery is the famous choice for cancer patients with no systemic disease or metastasis. Complete mesocolic excision (CME), mainly used by ligating the main vascular trunk and colonic segmental resection, is performed based on the site of the tumor (right hemicolectomy transverse colectomy or left hemicolectomy). However, surgery could be risky for elderly patients with an acute malignant colonic obstruction. Meanwhile, conventional laparoscopic surgery (CLS) with CME is a safe technique and is being used around the globe. This robotic surgery may result in fast recovery but intensively increase the cost. Thus, only a health system with high-end technology can be practised (De Rosa et al., 2015).

5% -10% of CRC cases are developed by inheritance through polyposis syndrome. For instance, adenomatous polyposis syndromes, serrated polyposis syndrome, and hamartomous polyposis. Patients with this syndrome demand a multidisciplinary team because of the disease presentation, diagnosis, and management complexity. An analysis of genetics can commonly be referred to follow-up patients and at-risk relatives for further precaution (Patel and Hyer, 2019).

2.1.7 Prevention of CRC

Various strategies like health campaigns, different screening technologies, and the development of personalized medicine gradually reduced CRC incidence. High incidences in men are primarily due to low awareness of screening. In this digitalization era, campaigning on mass media is the most effective in improving knowledge of cancer (Schliemann et al., 2018). Until now, for such an asymptomatic disease, colonoscopy is suggested to prevent tumorigenesis. The removal of an adenoma-carcinoma polyp from the colon and the ability to provide a biopsy is the advancement of colonoscopy technique,

particularly for at-risk individuals over 50. Individuals younger than the age range are ineligible unless they have a cancer-related family history (Ambe et al., 2017).

Besides, disruption of gut microbiota may lead to the development of colon diseases like dysplasia, clonal expansion, and malignant transformation. Therefore, it is crucial to maintain and pay more attention to the results of the mucosal immune system and actions to protect against pathogenic microorganisms (Eslami et al., 2019). Other precautions that can be taken to avoid the recurrence of cancer is having physical exercise. Compared to different preventive strategies such as lifestyle modification, diet, and smoking, physical activity saves 15% of lives. However, the specific type of exercise, dose, and intensity is still in lacuna (Oruc and Kaplan, 2019).

A study by Ho and the team focused on two modifiable variables (diet and physical activity) was the first study on CRC survivors in the Asian population. Although there are limited findings on the effective way to change the lifestyle of CRC survivors, this finding promotes a change in individuals' diet and physical activity (Ho et al., 2013). In conclusion, Asians need to stop adopting the Western diet, reduce red or processed meat intake, and practise more physical activities.

2.2 Overview of microRNA (miRNA)

2.2.1 What is miRNA

MicroRNA (miRNA) is a post-transcriptional gene regulator produced endogenously and comprises a sequence of 22 nucleotides (nt) of single-stranded RNA. MiRNA was first discovered in *Caenorhabditis elegans* (*C.elegans*) in 1993 by Victor Ambros and colleagues (Barbu et al., 2020). While studying the *lin-4* gene in the

roundworm *C.elegans,* which is supposed to encode a protein that regulates larvae development, they discovered a new gene instead. The mutation of *lin-4* has altered postembryonic developmental events by negatively regressing LIN-14 protein production during the first larval stage (L1). The *lin-4* gene does not encode for LIN-14 protein in all four tested of Caenorhabditis species; instead, they transcript two small fragments of molecular RNA, 22 and 61nt, that were believed to have a complement on LIN-14 mRNA by an antisense RNA-RNA interaction (Lee et al., 1993).

Another researcher supports this discovery by creating a model to prove the *lin-4* RNA attached to 3'UTR of LIN-14 messenger RNA (mRNA), which regulates the LIN-14 protein production during the transition of the first stage to the second stage of larval development. According to Wightman and team, during the development of *C.elegans,* a temporal gradient of the LIN-14 protein was generated, proving the post-transcriptional mechanism. The LIN-14 3'UTR is conserved between *C.elegans* and *C.Briggsae* and the sequences were figured out to be complemented with the *lin-4* RNA (Wightman et al., 1993).

After seven years of this novel discovery, other researchers found another miRNA called let-7, another gene in the *C.elegans* that encoded the LIN-29 mRNA in nematodes (Reinhart et al., 2000; Slack et al., 2000). Like the lin-4 gene in *C. elegans,* the let-7 gene is mapped to the same function. Numerous researchers cloned tiny RNAs from worms, drosophila, and human cells to pinpoint the precise roles of these RNAs based on these two fundamental miRNA results. On top of that, 20 novel genes were found in drosophila, 30 in people, and 60 in worms. Different miRNA origins demonstrated various miRNA functions. This thought process began when these newly discovered genes were not considerably expressed and expressed as lin-4 and let-7 during the developmental stage

as opposed to being expressed differently in specific cell types (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee & Ambros, 2001).

2.2.2 The biogenesis of miRNA

In 2004, Bartel synthesized the complete finding on the biogenesis of miRNA. To simplify the process, we can divide the miRNA canonical into several phases: transcription, cleavage, maturation, translational repression, and target recognition. The miRNAs reside in introns of the primary messenger RNA (mRNA) that will be transcripted as poly-adenylated in the stem-loop form structure by RNA polymerase II and III and is recognized as primary microRNA (pri-miRNA) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee & Ambros, 2001).

The first cleavage is the nuclear cleavage step performed by Drosha and DiGeorge syndrome critical gene 8 (DGCR8). These two endonucleases (known as Pasha in fly and worm organisms) will cleave the pri-miRNA (~70-100 nt) with intermediate stem-loop structure at sites near the base of the primary stem-loop forming precursor miRNA (pre-miRNA) with 5' phosphate and 3' overhang. The pre-miRNAs are then transported out from the nucleus to the cytoplasm site through a protein receptor called Exportin-5 with the help of Ran-GTP (Bartel, 2004; Bohnsack et al., 2004) and experience a second cleavage by an enzyme called Dicer into ~22nt incomplete double-stranded miRNA duplexes (Bohnsack et al., 2004; Lee et al., 2003; Yi et al., 2003).

Dicer acts as a helicase, unwinding the double strand of miRNA and producing two single-strand miRNA. The miRNA single strand undergoes the maturation process by joining Argonaute protein and assembling with RNA-induced Silencing Complex (RISC) with the help of transactivation responsive RNA-binding protein (TRBP) (Khoury & Tran,

2015; Kim et al., 2016). Dicer, also an RNase III endonuclease, was first recognized for its role in generating the small interfering RNAs (siRNAs) mediating RNA interference (RNAi). It was later shown to play a role in miRNA maturation. It first recognizes the double-stranded portion of the pre-miRNA, perhaps with a particular affinity for a 5' phosphate and 3' overhang at the base of the stem loop.

Then, at about two helical turns away from the base of the stem-loop, it cuts both strands of the duplex. This cleavage by Dicer lops off the terminal base pairs and loop of the pre-miRNA, leaving the 5' phosphate and ~2 nt 3' overhang characteristic of an RNase III and producing a siRNA-like imperfect duplex that comprises the mature miRNA and similar-sized fragment derived from the opposing arm of the pre-miRNA. The nuclear cut by Drosha defines one end of the mature miRNA, while the other end is generated by the enzyme Dicer in the cytoplasmic region (Bartel, 2004). One strand of the duplexes remains as mature miRNA (5') released by Argonaute, and the other strand is passenger miRNA strand (3'). In Figure 2.1, the canonical process of miRNA biogenesis is explained in a word flow (A), illustrated in the diagrammatic version (B), and the phases involved were also listed (C).

2.2.3 MiRNA nomenclature

The duplexed miRNA strands will produce two types of arms: the guide strand (5') and the passenger strand (3'). The 5' component is named for miR. Meanwhile, miR* is for the other 3' arm. Before this, it is believed that only the 5' strand will react upon the mRNA by binding to mature RISC, and the 3' will be degraded in the cytoplasm. However, results from deep sequencing showed that minor of them also persist and control

the functional gene regulator. Hence, the miRNA naming is changed from miR/miR* to miR-5p and miR-3p, respectively (Czech & Hannon, 2011; Griffiths-Jones et al., 2011; Yang et al., 2011).

MiRNA's naming keeps changing throughout the literature findings. According to Bernardo and team (2012), the difference between the first three names resembles different species, where 'has' for human, 'mmu' for mouse, and 'rno' for rat species. (eg: hsa-miR-XX, mmu-miR-XX and rno-miR-XX). MiRNA's names were explicitly written based on arm localization. It could be either 5p or 3p. For Example, hsa-miR-XX-5p and hsa-miR-XX-3p. The capitalization of 'R' in 'miR' is essential to differentiate between mature miRNA against precursor-miRNA, mir.

Mature miRNAs can transcribe from genes with different precursor sequences while generating the same series. A numerical suffix is typically used to distinguish this type of miRNA. Examples are hsa-miR-XX-1 and hsa-miR-XX-2. Meanwhile, hsa-miR-XXa and hsa-miR-XXb were named for similarly related miRNAs that differ only by one or two nucleotides. Yet, there are certain exceptions to the naming method for plants and viral miRNAs. Because of historical reasons, the names frequently use non-numerical names, such as let-7 and lin-4. Each miRNA may have multiple names throughout the study, but they were all logged with a unique ID that was always the same for future reference.

2.2.4 MiRNA target recognition

The incorporation of mature miRNA with AGO protein, the component of RISC, will bind to 3' UTR target mRNA and trigger the silencing process (Gregory et al., 2005) by inhibiting the protein synthesis through translational repression, mRNA deadenylation,

and decay or transcript cleavage (Gustafson et al., 2016). MiRNA will bind to complementary regions in the 3’ UTR of mRNA according to the sequence of complementarity (Bartel, 2009; Shin et al., 2010; Ameres and Zamore, 2013).

The specificity of miRNA targets is defined by Watson-Crick complementarities between positions 2-8 from the 5’ miRNA with the 3’ untranslated region of their target mRNA. RISC will induce degradation in high-specificity attachment, while imperfect miRNA-mRNA complements will block the protein translation (Bartel, 2009). As Svoboda (2015) mentioned, the action taken by AGO-miRNA loading will depend on whether endonucleolytic cleavage or translational repress activity is required. The complementarity of miRNA-mRNA binding will determine it. Before this, Yekta, Shih, and Bartel (2004) proved that mir-196's complete binding with HoxB8 mRNA caused the endonucleolytic cleavage and was attributed to RNA interference (RNAi) pathway. On the contrary, imperfect binding of miRNA-mRNAs resulted in translational repression and tended to degrade (Svoboda, 2015). Despite this, both activities of miRNA-mRNA are reducing the amount of protein encoded (Garzon et al., 2010).

Additionally, miRNA-5'UTR binding and activity have been demonstrated. Lim and colleagues transfected miR-124 and miR-1 into human cells and used microarray analysis to examine how the mRNA profile changed. They discovered that the brain has an extensive miR-124 profile, whereas the muscle has an miR-1 profile. The shift of these two miRNAs toward the organ or specific organ where they liked to express themselves led to the hypothesis that miRNAs are likely tissue-specific regulators (Lim et al., 2005). Since miRNA genes involved up to 30% of protein regulation, any changes in coding repression would cause unexpected effects. Although miRNA regulation of the human