NEXT GENERATION SEQUENCING BASED TARGETED GENE MUTATIONAL PROFILES OF FORMALIN-FIXED PARAFFIN EMBEDDED SPECIMENS FROM COLORECTAL CARCINOMA CASES IN HOSPITAL PAKAR UNIVERSITI SAINS MALAYSIA

HAFEEZ ABIOLA AFOLABI

UNIVERSITI SAINS MALAYSIA

2025

NEXT GENERATION SEQUENCING BASED TARGETED GENE MUTATIONAL PROFILES OF FORMALIN-FIXED PARAFFIN EMBEDDED SPECIMENS FROM COLORECTAL CARCINOMA CASES IN HOSPITAL PAKAR UNIVERSITI SAINS MALAYSIA

by

HAFEEZ ABIOLA AFOLABI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

August 2025

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to Allah, the one and only Almighty God who makes this easy for me. It is my pleasure to acknowledge the following people who made this Ph.D. dissertation possible: My supervisor, Prof Madya Dr Md Salzihan Md Salleh for your inspiration and knowledge impact, your relentless support was instrumental in keeping me focused and beyond what may be expected of supervisor. My sincere appreciation also goes to my co-supervisors for their unending effort and direction towards achieving a successful dissertation. I am extremely humbled and thankful for your efforts and time, your wisdom and skills availing to students are exemplary, glad to have you as my supervisor. Profound acknowledgement goes to the Universiti Sains Malaysia for availing me the opportunity to come here for my PhD program and also to the entire staff of the Department of Pathology for their support. A very big appreciation to Cik Maya and the department staffs. Special appraisal goes to Dr Afolabi Lukman, a brother and inspiration for his immeasurable assistance ranging from his prudent advice and selfless guidance towards ensuring a successful and well-detailed dissertation. Big thanks also go to Dr Ahmad Irekeola and Dr Sameer Al-Mhanna, a brother and idea-sharing companion, a very big thanks to you. To every member of my noble family, Alhaja M. Afolabi, sisters and brothers, thank you all for your support and care, and for never stop believing in me. Thanks to my peers on the health campus as well as my fellow nationality mates whose support was important throughout the periods. You constantly compelled me to keep brisk and up-to-date with the current literature. To Kelantan, the beautiful city for keeping my moods afloat and providing the calmness and atmosphere conducive to learning.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APPENDICES

LIST OF ABBREVIATIONS

HPUSM:	Hospital Pakar Universiti Sains Malaysia
NGS:	Next Generation Sequencing
FFPE:	Formalin-Fixed Paraffin-Embedded
CRC:	Colorectal Cancer
DiH_20:	Deionised Water

PROFIL MUTASI GEN BERSASAR BERASASKAN PENJUJUKAN GENERASI HADAPAN BAGI SPESIMEN YANG DIFIKSASI DALAM FORMALIN DAN DIBENAM DALAM PARAFIN DARIPADA KES KARSINOMA KOLOREKTAL DI HOSPITAL PAKAR UNIVERSITI SAINS MALAYSIA

ABSTRAK

Karsinoma kolorektal (CRC) merupakan kanser ketiga paling lazim dan penyebab ketiga paling umum kematian berkaitan dengan kanser di seluruh dunia. Ia merupakan proses pelbagai langkah yang melibatkan pengumpulan pelbagai perubahan genetik. Melalui pemprofilan molekular komprehensif dan analisis CRC menggunakan penjujukan generasi seterusnya (NGS), mutasi genetik dalam kanser kolorektal dapat dikenal pasti dengan tepat untuk potensi saringan CRC. Kajian ini menjalankan analisis molekular terhadap spesimen parafin yang difiksasi dengan formalin (FFPE) daripada 30 kes CRC di Hospital Pakar Universiti Sains Malaysia (HPUSM) menggunakan NGS. DNA genomik (gDNA) yang diekstrak daripada sampel FFPE menggunakan kit pengekstrakan QIAgen FFPE telah digunakan untuk penyediaan perpustakaan, dan seterusnya dijujuk menggunakan mesin Illumina MiniSeq. Daripada jumlah 30 sampel FFPE [60% lelaki dan 40% perempuan] yang dijujuk, 22 sampel berjaya disempurnakan dalam proses penjujukan, mewakili 73.3% daripada jumlah keseluruhan sampel: 20 sampel CRC dan 2 tisu normal. Berdasarkan data demografik, data tertinggi dicatatkan bagi kategori "kurang daripada 65 tahun" iaitu 66.7%, "kumpulan pesara" sebanyak 70%, "tahap 3" sebanyak 53.3%, "gred sederhana terdiferensiasi" sebanyak 76.7%, dan pesakit dengan "lebih daripada 2 komorbiditi" sebanyak 36.7%. Secara keseluruhannya, sebanyak 552 mutasi yang melibatkan 29 gen dan 11 kromosom telah dikesan. Gen yang paling banyak

mengalami peningkatan ekspresi ialah KIT:68(12.3%), FGFR4:61(11.1%), EGFR:60(10.9%), ALK:53(9.6%), DCUN1D1:41(7.4%), PDGFR:40(7.2%), KRAS:33(6.0%), CDK4:92, dan CDK4:92. FGFR3:26(4.7%). Mutasi gen yang paling banyak menurun ialah ESR1, FGFR1, CCND1, HRAS, AR. Lima kromosom yang paling terlibat ialah chr4:(24.3%), chr7:(15.2%), chr12:(12.2%), chr5:(11.6%), dan chr2:(11.1%). Selepas penapisan data berdasarkan kriteria tertentu, 105 mutasi melibatkan 15 gen melepasi tapisan. Model Ramalan Dalam Siliko mengenai kesan klinikal mutasi yang telah ditapis mengenal pasti 21 mutasi patogenik (daripada 9 gen), 57 mutasi jinak yang diterima (daripada 5 gen), dan 17 varian tidak ditentukan (daripada 4 gen). Mutasi gen patogenik yang dikenal pasti termasuk APC, NRAS, ALK, PIK3CA, KRAS, IDH1, FGFR1, ERBB2, dan ESR1. Pengesahan terhadap mutasi patogenik ini menggunakan pangkalan data dalam talian seperti dbSNP, Ensembl, dan Franklin oleh pangkalan data dalam talian Genoox turut mengenal pasti gen-gen ini sebagai patogenik. Ini adalah profil genetik NGS komprehensif yang pertama bagi kes CRC di Malaysia dan HPUSM. Keunikan kajian ini terletak pada kejayaan mengenal pasti mutasi gen patogenik dalam kes kanser kolorektal di Malaysia dan HPUSM. Penemuan ini akan menjadi asas penting kepada potensi rawatan sasaran bagi kes CRC.

NEXT GENERATION SEQUENCING BASED TARGETED GENE MUTATIONAL PROFILES OF FORMALIN-FIXED PARAFFIN EMBEDDED SPECIMENS FROM COLORECTAL CARCINOMA CASES IN HOSPITAL PAKAR UNIVERSITI SAINS MALAYSIA

ABSTRACT

Colorectal carcinoma (CRC) is the third most prevailing cancer and the third most common cause of cancer-related deaths worldwide. It is a multistep process involving the accumulation of multiple genetic alterations. Through comprehensive molecular profiling and analysis of CRC using NGS, genetic mutations in colorectal cancer can be accurately identified for potential CRC screening. The present study carried out molecular analysis of formalin-fixed, paraffin-embedded (FFPE) specimens from 30 CRC cases in Hospital Pakar Universiti Sains Malaysia using NGS. Extracted gDNA from FFPE samples using a QIAgen FFPE extraction kit was used for library preparation, then sequenced in an Illumina MiniSeq machine. From a total of 30 FFPE samples [60% male and 40% female] sequenced, 22 completed the sequencing run, representing 73.3% of the samples: 20 CRC samples and 2 normal tissues. From demographics results, the highest data was recorded for “less than 65 years old” at 66.7%, “Retired group” at 70%, “Stage-3” at 53.3%, “Moderately differentiated grade” at 76.7%, and lastly, patients with “More than 2 comorbid” at 36.7%. Overall, 552 mutations involving 29 genes and 11 chromosomes were detected. The most upregulated were *KIT*:68(12.3%), *FGFR4*:61(11.1%), *EGFR*:60(10.9%), *ALK*:53(9.6%), *DCUN1D1*:41(7.4%), *PDGFR*:40(7.2%), *KRAS*:33(6.0%), *CDK4*:27(4.9%), and *FGFR3*:26(4.7%). Most downregulated gene mutations were ESR1, FGFR1, CCND1, HRAS, AR. The five most involved chromosome were chr4:(24.3%), chr7:(15.2%), chr12:(12.2%), chr5:(11.6%), and chr2:(11.1%)

respectively. After applying filter criteria, 105 mutations involving 15 genes passed the filter. In-Silico Prediction model of clinical consequences of filtered mutations identified 21 pathogenic mutations (from 9 genes), 57 tolerated benign mutations (from 5 genes), and 17 variants unspecified (from 4 genes). Identified pathogenic gene mutations are APC, NRAS, ALK, PIK3CA, KRAS, IDH1, FGFR1, ERBB2, and ESR1. Validation of the pathogenic mutations on dbSNP, Ensembl, and Franklin by Genoox online databases also identified them as pathogenic genes. This is the first comprehensive NGS genetic profiling of CRC cases in Malaysia and HPUSM. The study's novelty is based on the successful identification of pathogenic gene mutations in colorectal cancer cases in Malaysia and HPUSM. These findings will serve as a leadway for potential targeted therapy for CRC cases.

CHAPTER 1

INTRODUCTION

1.1 Study Background

Colorectal carcinoma (CRC) is the third most prevailing cancer and the third most occurring cause of cancer-related deaths worldwide, with almost 700,000 deaths per annum globally (Barik et al.,, 2023; Todua et al., 2015). According to the Malaysian National Cancer Registry Report in 2007- 2011, CRC is the third most frequently occurring cancer among men and second among women (Azizah et al., 2019), but in the latest report, CRC is the second most common cancer in Malaysia at 13.5% (in 2012-2016) s 14.1% (in 2017-2021) respectively (Azizah et al., 2019; Ike et al.,, 2021). It is also the most common cancer among males, and the incidence increases with age and the increased rate is seen in all age groups (Ike et al., 2021). Approximately 25% of patients have distant metastases at the time of diagnosis, and the most common site of metastasis, which develops in 50% of patients with CRC is the liver because the circulating blood from the intestines harbouring the primary CRC tumours is first filtered via the hepatic system via the portal vein (Yang et al., 2023). The poor prognosis of metastatic CRC has been the driving force for the ongoing efforts to develop precision-targeted therapy approaches that can improve patient outcomes (Khan et al., 2018).

In today's omics world, we are edging to an era of personalised and precision medicine where cancer diagnosis and treatment will be tailored to each patient's disease status depending on their unique genetic signatures. For example, receptor tyrosine kinase and epidermal growth factor receptor (EGFR) play an important role in the growth and survival of many solid cancers, including colorectal cancer (CRC) (Binefa et al., 2014), proven gene mutation profiling to be an effective screening and therapeutic strategic

in tumourigenesis (Maida et al., 2017). The successful completion of the Human Genome Project in 2003 marked the era of genomic medicine breakthroughs. the core goal of this project is to utilise genomic profiling knowledge for better treatment, prevention, and overall reduction in health costs in cancer management (Bentley et al., 2000; Collins et al., 1998). Colorectal cancers (CRCs) represent a group of molecularly heterogeneous diseases that are characterized by a range of genomic and epigenomic alterations (Wang et al., 2020), our increasing understanding of the molecular, pathological and epidemiology of CRC will enable clinicians and researchers to develop personalized therapy to improve the management of patients with CRC. Thus, boosting the development of potential targeted therapies based on molecular-specific subtypes.

Colorectal tumorigenesis is a multistep process that involves the accumulation of multiple genetic alterations, including chromosomal abnormalities, gene mutations, and/or epigenetic changes, that transform normal colonic epithelium to colorectal carcinoma (Thomas et al., 2024). Recent data regarding the comprehensive molecular characterizations of CRCs have increased our understanding of the genomic and epigenomic landscapes of CRCs, and have enabled their classification into various subtypes according to their distinct molecular pathologies and clinical features (Wang et al., 2020), thence, opening the knowledge of genetic mutations in metastatic CRC (mCRC). The prognostic and predictive implications of certain mutations, including RAF, RAS, and deficient mismatch repair (dMMR), are well-established culprits in the CRC pathway and are now routinely assessed as a component of clinical care (Kim et al., 2024). Many genes are known to be involved in carcinogenesis and this was based on the understanding of how a normal cell can transform its phenotype following genotypic changes into malignant cells.

Reliable and practicable screening of certain gene mutations has been incorporated into clinical practice in many advanced countries but still remains as an under-practised approach in many developing countries and low-income nations (Katsanis et al., 2013). Hence, there is a need for the development of a rapid and low-cost detection of mutations with high accuracy. Though colonoscopy is the gold standard for CRC screening, DNA sequencing is still an understudy approach for the clinical screening of CRC especially in developing countries, including Malaysia. Through comprehensive molecular profiling and analysis of CRC, early screening of cancer and CRC can be accurately done in addition to the classical histopathological data. This ensures a precision medical treatment geared towards achieving a good prognosis via the use of a targeted approach chemotherapy combination.

Next-generation sequencing (NGS) has revolutionized the study of cancer for the detection of somatic variants ranging from sequencing entire tumour genomes to targeted clinical diagnostic gene panels. NGS can sequence hundreds and thousands of genes or whole genomes in a short period (Satam et al., 2023). The information generated by targeted gene panels or sequence variants/mutations detected by NGS can be widely used for disease diagnosis, therapeutic decisions, follow-up of patients and/or provide prognostic insights for a particular tumour (Zalis et al., 2024). Also, advances in sequencing reliability, pipeline analysis, accumulation of relevant data, and the reduction of costs are rapidly increasing the feasibility of NGS-based clinical decision-making (McCombie et al., 2019). Over the years, NGS has become more versatile because of its capability to allow simultaneous analysis of many genomic loci while revealing the exact sequence changes (Del-Vecchio et al., 2017). This advancement in technology has reduced the turnaround time (TAT) for cancer

screening which makes it more feasible for routine diagnostic laboratory testing (Han et al., 2013).

Routine genomic testing of tumours in clinical settings in many countries and Malaysia remains very limited. The AmpliSeq for Illumina focus panel targets hundreds of mutations across 52 driver genes associated with solid tumours including CRC. These panels can be used to investigate single nucleotide variants (SNVs), insertion/deletions (indels), and copy number variants (CNVs) in DNA samples, or gene fusions in RNA samples (Almuzzaini et al., 2021). The targeted gene content includes *ERBB2-ERBB4, PIK3CA, AKT1, NRAS, BRAF, mTOR, MYC, PDGFRA, EGFR, KIT, KRAS* etc. This will enable us to focus resources on relevant genes that are most likely to play a role in CRC tumourigenesis. Especially, as USM prides itself as a citadel of academic excellence and a tertiary medical centre, it is time to move up the ladder to the next advanced level of advanced research in NGS by procuring such a facility.

1.2 Problem statement

Colorectal cancer (CRC) is reported in many countries, such as in United State and United Kingdom as a common cause of cancer-related death including the Asia Pacific region (Veettil et al., 2017). In Malaysia, many articles regarding CRC mortality have been published as well as its substantial economic burden which is likely to increase over time in Malaysia, owing to the current trend in colorectal cancer incidence (Veettil et al., 2017) which is increasing in Malaysia, and in other developing Asian countries (Muhamad et al., 2023). In Malaysia, increasing affluence and an increased prevalence of risk factors for colorectal cancer, such as westernized-style of diet, obesity, and smoking could be associated with an increasing incidence of colorectal cancer in the country (Center et al., 2009). Malaysia is undergoing an ageing of its population, and

this ageing trend may further increase the prevalence of CRC (Zulkifli et al., 2014) because, 80% of colorectal cancer cases in Malaysia are diagnosed in people older than 50 years (Gcc-Lim et al., 2008).

While the incidence and mortality rates for colorectal cancer in Japan, South Korea, and Singapore have been stable and are even declining in a trend similar to the pattern seen in the USA and the UK (Sung et al., 2015), this decrease is majorly attributed to efficient colorectal cancer screening programs which have reduced the prevalence of risk factors, and/or improved treatments in these countries (Shin et al., 2013). Here in USM, the facility is yet to commence early molecular analysis or molecular screening of CRC patients at the level of NGS which is considered the most advanced molecular testing method in the management of cancer (Satam et al., 2023; Tarawneh et al., 2022; Zalis et al., 2024).

In Malaysia, most patients with colorectal cancer are usually diagnosed at a late stage when the prognosis is poor, with the 5-year relative survival at this stage being lower than that in developed Asian countries (Natrah et al., 2012). This late stage is associated with a poor prognosis, increased health burden due to the higher treatment cost and poor quality of life in the late stages, and at present, there is no formal/structured national colorectal cancer screening program in Malaysia (Veettil et al., 2017). Early screening of the rising incidence and early detection of colorectal cancer oncogenes or molecular markers through advanced techniques such as NGS is required to mitigate these burdens. Ages from 45 years old were delineated as the screening age level for average-risk individuals with a family history of colorectal cancer (Cheney et al., 2024).

This present study analytically investigated the molecular alteration in genetic mutations associated with CRC as well as to determine clinical consequencies by

compiling data on gene profiling of CRC in Malaysia. Despite this rising CRC incidence and the promising result from NGS application, and the lack of early NGS screening facilities, molecular gene profiling for CRC cases is still insufficient in HPUSM and Malaysia, hence the need for this study and its contribution to early screening, diagnosis, and targeted treatment of CRC.

While NGS application for gene mutation detection has been incorporated into clinical practice for routine laboratory testing in many advanced countries, it remains an under-practised approach in Malaysia and other developing nations (Lee et al., 2019). HPUSM is a tertiary hospital, Apex University, ranked among the best in the country, with robust medical facilities and medical professionals. Presently, HPUSM is yet to commence its own molecular NGS profiling of cancer cases. The management of CRC is via the standard conventional medical approach based on medical history and imaging investigation, including the colonoscopy and CT-scan/MRI. The general use of NGS for CRC screening and outcome prognosis in Malaysia is very minimal (Hasbullah et al., 2023). The latter is available in few private facilities or institutions and less available in government hospitals throughout Malaysia (Othman et al., 2023). This limited adoption is largely attributed to factors such as high costs and the lack of properly established infrastructure and experienced staff required for NGS testing (Oldoni et al, 2023; Zauber et al, 2015). Molecular screening for CRC using NGS is described as still being "in its infancy" in Malaysian hospitals (Afolabi et al., 2025; Othman et al., 2023).

1.3 Justification of Study

This study highlights the significance of NGS in gene mutation profiling among CRC cases in Malaysia by providing rapid detection of hot-spot mutations or driver mutations in CRC from Formalin-Fixed Paraffin-Embedded (FFPE) samples. Also,

this study identifies gene mutations that can serve as hot-spot mutations for targeted therapy in colorectal cancer, and validates the mutated genes with published CRC gene profiling in places such as in China (N.Zhang et al., 2020), India (Jauhri et al., 2016; Mayank.J et al., 2017), South Korea (Lim et al., 2015), the United States (Russo et al., 2014; Telysheva et al., 2022), and Netherlands (Schweiger et al., 2018).

The present study is carried out in line with the study's goal of identifying crucial mutations in patients with CRC who have tumour tissues available for molecular analysis. This study's results highlight the significance of NGS as a prudent investigating tool to contribute to detection of various gene mutations in patients with CRC, portraying NGS as an efficient tool to identify uncommon somatic mutations with low amplification. Also, providing potential for a precision-targeted approach for cancer management.

The discovery of crucial mutations in CRC will present Hospital-Pakar USM (HPUSM) as a front-runner in molecular testing and targeted approach for CRC patients in Malaysia, because the available information or data on the molecular mutations will contribute to excellent prognosis. Through the success of this study, Hospital-Pakar USM will be able to provide a reliable and accurate test to carry out mutational screenings of tumour specimens in the clinical practice and diagnostic laboratory. Because NGS identification of pathogenic mutated genes promotes precision target therapy approach. Identifying the pathogenic mutations will accord the patient a targeted therapy option rather than non-specific chemotherapy, which produces deleterious side effects because of attacks on all rapidly dividing cells to produce adverse effects such as nausea and vomiting, alopecia (hair loss), mouth Sores (Stomatitis/Mucositis), bone marrow suppression (Myelosuppression) (Domingo et al., 2018; Schweiger et al., 2018).

This study's findings will boost clinicians' decision-making in opting for targeted therapy options. In doing so, we will identify pathogenic mutations that contribute to the early screening of colorectal cancer. This latter approach will reduce morbidity of the disease, improve quality of life for the patients, and proffer a good prognosis. Although there is ongoing research on the detection of CRC in Malaysia, however, there are still significant data gaps and a universal lack of comprehensive surveillance programs that include early molecular testing, and this has made it difficult to make early CRC detection.

Using the NGS technique, we can significantly detect various types of mutations that could not be detected by traditional PCR methods, and these results will contribute to therapeutic decisions in administering early cancer diagnosis and treatment. Furthermore, the clinical implication of the minor mutant allele in whole tumour cells can be successfully investigated in larger clinical cohorts in shorter duration with high sensitivity and specificity ratios. Lastly, because data on the effectiveness of early molecular screening through NGS in Hospital-Pakar USM are not available, our study will be the first of its kind to be carried out in Hospital-USM on molecular gene profiling on CRC using NGS.

To the best of our knowledge, this is the first comprehensive study on molecular profiling of colorectal cancer using NGS in Malaysia. Our study's novelty identifies eight pathogenic gene mutations in colorectal cancer disease.

1.4 Hypothesis

1. There is an association between specific gene mutations and colorectal carcinoma using available probes that detect 50 common genes known to be altered in tumours
2. There is an association between the genetic changes and the identified patterns

3. The changes, specific patterns or profiling can be made organ-specific
4. The pathogenic status and mutational effect of the variants can be identified using the In-Silico Prediction model for validation with dbSNP, Ensembl GRCh37, and Franklin by Genoox databases.
5. There is an association between the detected variant's status and clinicopathological parameters

1.5 Research Questions

1. What are the specific gene mutations profiling in colorectal carcinoma cases in HUSM using available probes that detect 50 common genes known to be altered in tumours?
2. Can the genetic changes be analysed and whether or not there is any identified pattern?
3. If there are changes, is there any specific pattern or profiling that can be made organ-specific?
4. Can the pathogenic mutations identified using the In-Silico Prediction model be validated on dbSNP, Franklin by Genoox, and Ensembl GRCh37 databases?
5. What is the association of the detected variant's status with clinicopathological parameters in CRC cases in HUSM?

1.6 Study Objectives

1.6.1 General Objectives

The main objective of the study is to identify the molecular gene profiling of colorectal carcinoma cases from formalin-fixed embedded paraffin specimens at Hospital Universiti Sains Malaysia (HUSM) using Next-Generation Sequencing (NGS).

1.6.2 Specific Objectives

1. To identify the specific gene mutations associated with colorectal carcinoma using available probes that detect 50 common genes known to be altered in tumours.
2. To analyse these genetic changes and whether or not there is any specific pattern.
3. To analyse the patterns observed for organ-specific and prognostic use for CRC
4. To validate the pathogenic mutation genes identified using the In-Silico Prediction model on dbSNP, Franklin by Genoox, and Ensembl GRCh37 databases.
5. To determine the association of the detected variant's status with clinicopathological parameters.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of Colorectal Carcinoma (CRC)

Colorectal carcinoma, also known as colon cancer or colorectal cancer (CRC) is one of the most prevailing large intestinal cancers globally (Abedizadeh et al., 2024). CRC is typically classified as either sporadic, hereditary or familial, based on the genetics and aetiology of the illness (Lv et al., 2024; Sameer et al., 2013). CRC arises from benign adenomatous polyps which can undergo genetic mutations to change into malignant tumours over time (Abedizadeh et al., 2024). The tumour arises in the rectum or colon, although more commonly in the rectal part of the large intestine, hence, the name rectal cancer (Aljabiry et al., 2024). Colorectal cancer is described as a group of molecularly heterogeneous diseases characterized by a range of genomic and epigenomic alterations (Liu et al., 2024), that typically originate from the epithelial cells lining of the large intestine, which transform from the normal epithelium cell to cancerous cells due to episodes of genomic aberrations such as in the *APC, KRAS, BRAF*, and *TP53* genes (Su et al., 2024; Shi et al., 2024; Yang et al., 2024), coupled with risk factors such as advancing age, familial history, lifestyle existences (e.g., diet high in red meat, sedentary activity), and some diseases such as inflammatory bowel disease and in Lynch syndrome (Su et al., 2024; Shi et al, 2024). Worldwide, among all cancers, CRCs are the third most diagnosed tumour among the cancers, third most common cancer globally, (Abboud et al., 2024; Constantinou et al, 2024; Dunne et al, 2024; Simon et al., 2024), and are generally preventable via regular screening, especially after age 45 (Cheney et al., 2024; Leong et al., 2024)

2.2 Global Epidemiology of CRC

Colorectal cancer (CRC) is a serious pandemic health menace with a significantly high prevalence rate, presently affecting billions worldwide, with disastrous socio-clinical setbacks (Demb et al., 2024; Radzi et al, 2008). CRC represents 10% of all cancers, and it is the third most common cancer worldwide (Cheney et al., 2024; Simon et al., 2024), and the second most common cause of cancer-associated deaths globally, accounting for approximately 9.5% of cancer deaths (Morgan et al., 2023). The World Health Organisation WHO health release states that the majority of CRC cases occur in adults >50 years old, recording over 1.9 million new cases and over 930,000 deaths globally in 2020 (Alzahrani et al., 2021). Nonetheless, the Global Burden of Colorectal Cancer in the GLOBOCAN findings revealed geographical variations in incident rates, with the highest rates recorded in Europe, Australia, and New Zealand, while mortality was highest in Eastern Europe (Constantinou et al., 2024; Morgan et al., 2023). The lowest rate was reported in African regions and Southern Asia (Ionescu et al., 2023; Morgan et al., 2023).. It is predicted that by 2040, the CRC burdens will rise to 3.2 million new cases annually (63% increment) and 1.6 million mortality per year (73% increment), and most of the increment will be associated with advancing age and lifestyle changes, especially in developing nations (Morgan et al., 2023).

2.3 The Regional Incidence of Colorectal Cancer

Colorectal cancer (CRC) is among the most diagnosed cancers in the US and Canada, mostly linked to dietary-lifestyle alterations and fold increase in obesity (Li et al., 2021). Also, due to early detection and improved outcomes, but not without differences among numerous population regions (Siegel et al., 2020). In the US and fellow European nations, obesity and physical inactivity in the Westernised style of living are a major risk factor for the risen rates of somatic CRC (Siegel et al., 2020). In contrast,

Asia, Africa, and South America showed low incidence rates. CRC projection in 2024 in the US accounts for 10% of all cancers in the ratio of 1 in 23 for men and 1 in 25 for women (ACS et al., 2024; Siegel et al., 2024). In Canada, CRC represents 10% of all cancers with over 247,100 cancer cases, CRC recorded 25,200 cases and 9,400 fatalities (Cancer et al, 2024; Fitzgerald et al., 2024). While global incidence rates for colorectal cancer (CRC) differ, some countries in the American continent such as Mexico, Guatemala, Honduras, Nicaragua, Ecuador and Peru have consistently reported lower rates than the US and Canada, however, this is believed to be associated with less screening and less statistical updates (Organization et al, 2022; Wong et al., 2019)

In Europe, Western and Northern Europe have the highest rate of CRC incidence with nations such as Germany recording 493,200 cancer cases in 2020, among these 54,770 were CRC, 30,530 in males and 24,240 in females with a 5-year survival at 62% (Pardamean et al., 2023; Robert et al., 2020). In 2022, the WHO Global Cancer Observatory Agency report showed that Denmark had 48,480 new cancer cases and 17,204 deaths, 14.1% (6,878) were CRC with 1992 fatalities. As of 2023, Denmark recorded the highest rate of CRC globally with an age-standardized incidence rate estimated at 46.7 per 100,000 people (Pardamean et al., 2023; IARC et al, 2022). The lowest rate was noted in Eastern Europe in countries such as Moldova, Albania, and Bosnia and Herzegovina (IARC et al, 2022). One main contributing factor to these variations was related to less industrialized settings, remote lifestyle practices, and the standard of varying national screening programs being practiced and/or less advanced healthcare organisations which might underreport the incidences due to limited access to screening.

As of present, CRC is the third most prevailing cancer in Asia in both genders (Park et al., 2024; Joseph et al., 2005), with an increased incidence rate in Eastern Asian nations including China, Japan, South Korea, and Singapore. There have also been reports of a growing prevalence in Taiwan, Malaysia, and Thailand (Ashari et al., 2023; Wee et al., 2024; Yang et al., 2004; Yiu et al., 2004). Higher rates in these countries are reported to be majorly associated with increasing urbanization, dietary changes, and advancing ageing (Rawla et al., 2019). In contrast, countries like India, Indonesia, Vietnam, and Iran have relatively reported low prevalence (Wong et al., 2020), The death rate from CRC has been rising in Asian nations over the past decade. Singapore, Taiwan, and Japan have the highest rates (Shin et al., 2013).

In Malaysia, the gender ratio was reported at 2.2 for male to female ratio in the 2012-2016 national cancer registry report (Azizah et al., 2019; Muhamad et al., 2023; L. Xu et al., 2023; J Zhang et al., 2016). Lifestyle changes, sedentary daily routine, heightened taste, and preference for obesity-related fast food are the major reasons for the increment (Al-Ahwal et al., 2013; Azizah et al., 2019; Ismail et al., 2013; Shi et al., 2019).

In summary, universal differences in colorectal cancer incidence can be tied to differences in dietary intake, daily life activities, epigenetics and genetics, and the degree of readily accessible screening programs and facilities (Afolabi et al., 2022; Morgan et al., 2023; Xu et al., 2023). Advanced nations with higher occurrence rates recorded gains from more extensive and good screening, leading to CRC early detection and better aftermath outcomes (Azadeh et al., 2008; Rawla et al., 2019; Zacharakis et al., 2023). In contrast, developing countries reported mounting incidence rates and challenges associated with late-stage diagnosis and limited accessibility to medical care (Arnold et al., 2017; Hossain et al., 2022).

2.4 The histological and morphological presentations of colorectal carcinoma

Polyps are CRC precursors and aberrant cells in the intestinal mucosa that can progress from benign to metastatic condition, they represent altered nuclei acid that infiltrate and extend into the intestinal lumen, infiltrate lymphatic nodes, and acquire the characteristics of a cancerous condition (Shin et al., 2023; Tariq et al., 2016). Two main forms of polyps associated with CRCs include adenomatous and sessile serrated polyps. Theoretically, several histological, morphological, and genetic changes are involved in the development of precancerous polyps (Gui et al., 2020). The histological changes originate from the innermost mucosa layer of the colon and spread outwardly to the outermost (serosa) layer, after which, the spread is to nearby regions and finally other organs and tissues (Figure 2.1).

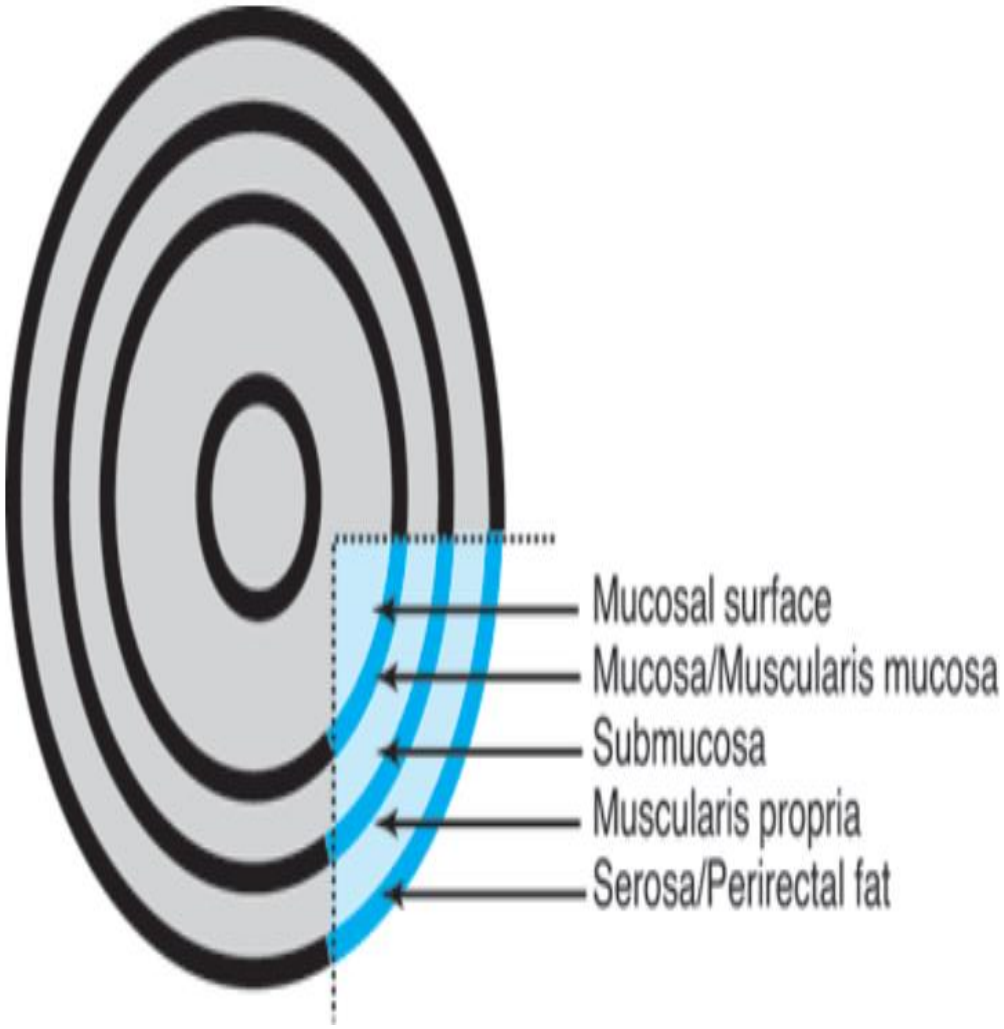

Figure 2. 1 Morphology layers of the colon

CRC primarily arises from the epithelial cells of the mucosal surface lining the colon and rectum with various histological features. Adenocarcinoma is primarily the most common histological subtype of CRC characterized by glandular structures formed by malignant cells (Gui et al., 2020), other forms include mucinous adenocarcinoma, signet-ring cell carcinoma, and medullary carcinoma (Huang et al., 2021). The gross appearance of the tumour differs and is dependent on the location and the stage of the tumour in describing tumour morphology (Rijsemus et al., 2024). The tissue lesion appearance can be described as polypoid and exophytic lesions (tumours are raised and protrude into the colon lumen as polyp-like masses), ulcerative lesions, annular, constricting lesions, and infiltrative lesions (Ono et al., 2024).

2.5 Aetiology and risk factors for colorectal carcinoma

Even though there is noteworthy progress attained in the diagnostic and medical treatment of patients with colorectal cancer, recently, there has also been a significant rise in the occurrence of colorectal cancer among people under 50 years old (Ionescu et al., 2023). The distribution of CRC is not even across the globe, subpopulation-wise; there is a marked difference in CRC incidence by race and ethnicity. In particular, in the US, the black Americans show the highest incidence, and have the highest mortality among major U.S. racial and ethnic groups at an overall incidence of 41.9 per 100,000, as compared to that of White Americans of 37.0 per 100,000 (Rawla et al., 2019). The Caucasians have a higher overall CRC incidence rate compared to Hispanics, with some studies showing a higher rate of KRAS mutations in Hispanics, the overall incidence and mortality rates are higher for Caucasians (Verma et al., 2025). The two main aetiology genetic pathways of CRC are 75-85% associated with the inactivation of the tumour suppression genes (TSG) and adenomatous polyposis coli

that are mutated in the germline of those with Familial Adenomatosis Polyps (FAP), while the rest 25-15% of sporadic CRC involves mutational inactivation of a protein lineage in DNA mismatch repair (MMR), including MSH2, MLH1, and PMS2 which accounted for the HNPCC syndrome (Hisamuddin et al., 2004). As low as 5-7% of CRC is diagnosed as inherited, while about 10-30% are from familial predisposition (DeRosa et al., 2015)

As seen in several organ tissue tumourigenesis, the pathogenesis of CRC involves adenoma-carcinoma changes, involving bio-genetic changes of the tissue and epithelial gradual changes from dysplasia to carcinoma in situ (Saraiva et al., 2023), and chromosomal instability (CIN) leading to amplifications and deletions of large DNA segments (Hassan et al., 2010; Harrold et al., 2024). The schematic pathogenesis pathway of CRC is depicted in Figure 2.2, illustrating how precancerous polyps' cells undergo cell variations and transformation into cancer cells.

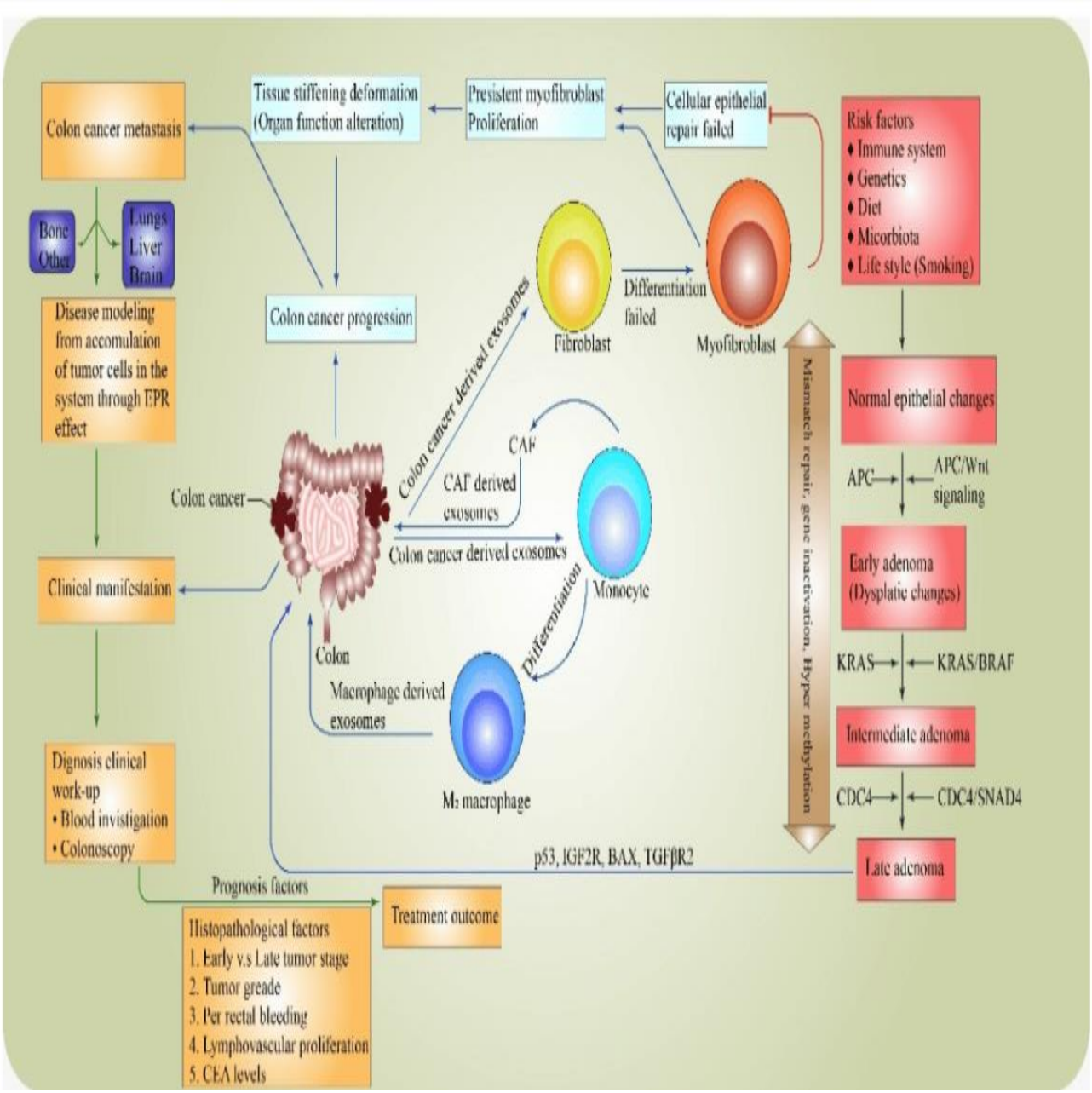

Figure 2. 2. Mechanism and prognosis factors influence the pathogenesis of colorectal cancer

The morphological alteration and gene aberration in the growth of preinvasive lesions/pre-cancerous lesions (adenomatous polyps) involving chromosomal instability in CRC are reported as allelic losses more on chromosomes 1p, 5q, 8p, 17p, 18p, and 18q, and less commonly as gains on chromosome 7p, 1q, 8q, 12q, 13q and 20q (Harrold et al., 2024; Kibriya et al., 2024; Muzny et al., 2012; Zarzour et al., 2015). In CRC tumourigenesis, CIN in the form of gains or losses of chromosomal regions, chromosomal rearrangements, and heterozygosity loss results in the alteration of tumour-regulatory genes or at cell cycle checkpoints leading to CRC initiation and propagation (Grady et al., 2014; Pino et al., 2010; Tsang et al., 2014). Chromosome loss in 5p is common in heterozygosity loss in the *APC* gene (crucial in early-stage CRC), 17p in heterozygosity loss in the *TP53* gene responsible for cell cycle and apoptosis control, 18q is seen as allelic losses causing deletion in CRC in *SMAD4* genes which is associated with the rapid advancement of adenomas to neoplastic lesion (Strubberg et al., 2017).

However, there is no single mutagen responsible for CRC aetiology because it is a complex disease with multi-factorial contributing factors (Afolabi et al., 2022; Murphy et al., 2019). On this, the exact causes are largely unknown (Murphy et al, 2024; Raskov et al., 2014), however, several environmental influences such as dietary intake of red and processed meat, and low-fibre diet (Cáceres-Matos et al., 2024; Ismail et al., 2024), inflammation such as chronic inflammatory bowel diseases (Katayama et al., 2024; Wang et al., 2024), hormones especially postmenopausal hormone replacement therapy (Liu et al., 2024; Tian et al., 2024), and lifestyle such as excessive alcohol consumption, smoking, including e-cigarette vapor from vaping (because e-cigarette aerosol contains chemicals like formaldehyde and acrolein, heavy metals, and volatile organic compounds, some of which are known carcinogens when inhaled deep

into the lungs can potentially cause lungs parenchyma damage) (Heywood et al., 2025), and physical inactivity (Lasabova et al., 2024) are the main causative risk factors.

The contribution of genetic susceptibility from genetic mutations and immune system vulnerability involving cellular pathways and gene alterations is an area of ongoing research that concentrates on how dissimilarities in mutated genes impact vulnerability to CRC advancement, and the immune system's incapability to identify and terminate cancer cell progression.

2.6 Screening for CRC

Several health agencies proposed appropriate screening guidelines for CRC based on age groups, and they include the American Cancer Society (ACS), the U.S. Preventive Services Task Force (USPSTF), the American College of Physicians (ACP), and the American College of Gastroenterology (ACG). ACP 2023 standard procedures suggest screening for CRC at age 50 for those with average risk, although ACS 2018, USPSTF 2021, and ACG 2021 guidelines advocate screening for CRC starting at age 45 for those with average risk (Lin et al., 2021; Qaseem et al., 2023; Shaukat et al., 2021; Wolf et al., 2018). Guidelines from ACS 2018, USPSTF 2021, and ACG 2021 advise everyone between the ages of 50 and 75 to get screened for CRC (Lin et al., 2021; Qaseem et al., 2023; Shaukat et al., 2021). According to the ACP 2023 recommendations, screening should be stopped if the average lifespan is 10 years or less. The ACS 2018, USPSTF 2021, and ACG 2021 guidelines all stated that screening should be carefully examined in people aged 75 and 85 years based on the risk parameters (Lin et al., 2021; Qaseem et al., 2023; Shaukat et al., 2021; Wolf et al., 2018). Unfortunately, there is presently no population-based NGS molecular screening

for CRC in many countries and Malaysia (Ghee et al., 2014; Loong et al., 2023; Pang et al., 2023)

CRC is a heterogeneous disease with at least three major molecular pathways to tumourigenesis and four consensus molecular subtypes (Fessler et al ., 2016; Wang et al., 2019). The purpose is to identify hereditary CRC syndromes, forecast response to focus therapies and investigate microsatellite instability (MSI) status in patients with CRC (Andrei et al., 2022; Reece et al., 2019). According to recent data, the five-year survival rate of stage I CRC is 91%, but drops to 72% for locally advanced CRC and even to 14% for stage IV, suggesting that early screening can credibly improve the morbidity and mortality of CRC (Chetroiu et al., 2021; Levin et al., 2018; Shaukat et al., 2022). CRC is the most occurring gastrointestinal tract malignancy worldwide, with approximately 150,000 new cases being diagnosed yearly in the United States with over 52,000 mortality each year, designating CRC as the second most fatal carcinoma in the US (Baumgartner et al., 2024; Poh et al., 2007). The frequency is similar in both genders and has been so over the past two decades, age rather than gender is a more important factor for screening selection (Messina et al., 2012; Wong et al., 2013; Zapka et al., 2002). The extensive implementation of national screening programs will significantly decrease the incidence of CRC, while early detection together with advances in clinical care will help to reduce CRC mortality.

Over the past decades, tumour screening to detect CRC at an early stage for better treatment outcomes and to prevent CRC by detecting and removing precancerous lesions such as polyps is growing significantly (Chang et al., 2024; Jacob et al., 2024; Toth et al., 2024). However, colonoscopy is still the gold standard for CRC screening through direct visualization of the colon with simultaneous therapeutic possibilities to

remove the polyps (Cordero et al., 2024; Tamraz et al., 2024; Xu et al., 2024). Although there are other supporting screening methods which include faecal occult blood tests (FOBT) and faecal immunochemical tests (FIT) that detect hidden blood in the stool and sigmoidoscopy which is a lesser version of colonoscopy because only the lower part of the colon and rectum can be view (Lee et al., 2024; Maida et al., 2024). FIT (Faecal Immunochemical Test) offers significant benefits in cancer screening, particularly for colorectal cancer. It's highly sensitive for detecting human blood in stool, indicating potential cancer or advanced polyps. Its ease of use and non-invasive nature lead to better patient participation in screening programs. FIT is also a cost-effective primary screening tool, helping to direct more invasive procedures like colonoscopies to those who truly need them. Ultimately, by enabling earlier detection, FIT programs contribute to a reduction in colorectal cancer mortality. FIT's main drawbacks include test variability between systems, false negatives, where cancers or polyps are missed due to inconsistent bleeding, leading to delayed diagnoses. False positives caused by benign bleeding, result in unnecessary colonoscopies, increased costs, and patient anxiety.

Omics science today introduces molecular research such as gene sequencing to identify potential biomarkers or aberrant genes associated with CRC (Li et al., 2024), especially as it is non-invasive and can also be done in the early stage of CRC propagation (J. Xu et al., 2024). Figure 2.3 below briefly illustrates the schematic investigative workout procedure for colorectal cancer wherein both physical patient details, biopsied approach and precision medicine are constitutively employed for maximum clinical results/findings.

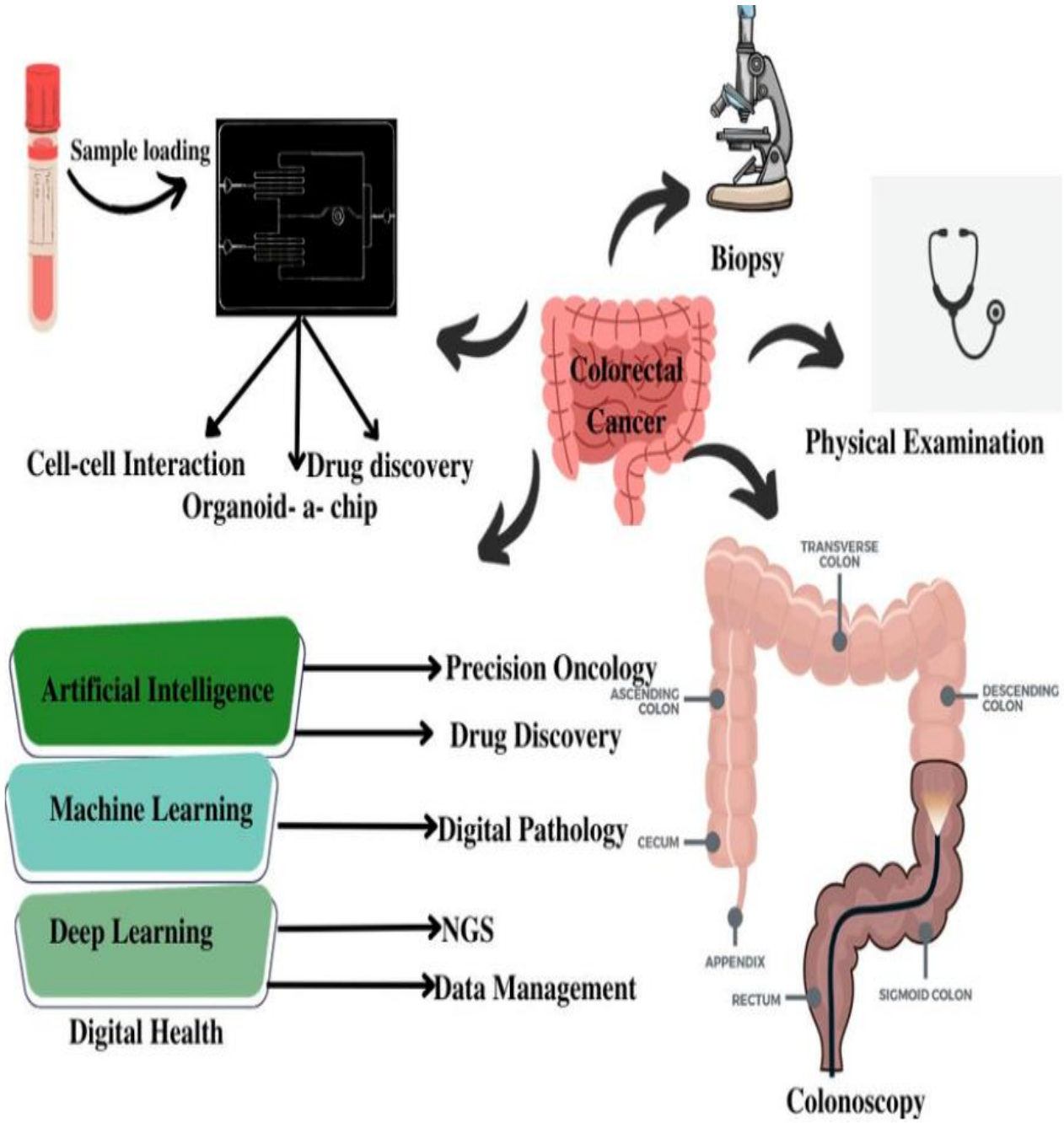

Figure 2. 3 Schematic illustrations of different diagnostic modalities for CRC