THE ROLE OF FOXP3 IN RELATION TO BRAFV600E AND MMR PROTEIN STATUS IN EARLY-ONSET COLORECTAL CANCER

ISAH RILWANU TSAMIYA

UNIVERSITI SAINS MALAYSIA

2025

THE ROLE OF FOXP3 IN RELATION TO BRAFV600E AND MMR PROTEIN STATUS IN EARLY-ONSET COLORECTAL CANCER

by

ISAH RILWANU TSAMIYA

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

June 2025

ACKNOWLEDGEMENTS

In the name of the Most Gracious and Merciful Allah, I am incredibly grateful to Allah for providing me with the opportunity and fortitude to complete this study. I could not have made it this far without the blessing and help of Allah SWT. I would like to express my sincere gratitude to my main supervisor, Dr. Siti Norasikin Mohd Nafi, for her tireless efforts in providing supervision, guidance, and constructive feedback throughout my study period. I appreciate your patience and support in making this study a success. I sincerely thank my Co-supervisors, Associate Prof. Dr. Anani Aila Mat Zin, Associate Prof. Dr. Nur Asyilla Che Jalil, and Associate Prof. Dr. Sarina Sulong, for their scholarly contributions and encouragement at every stage of my research work. I acknowledge the assistance and support of the laboratory staff in the Department of Pathology, the Human Genome Centre, and the Central Research Laboratory of the School of Medical Sciences during the conduct of the laboratory research work. Among them, notable ones were Cik Normayazi Binti Mohammed Ali, Encik Mohd Zaki Bin Hussein, and Puan Mazira Binti Mohammed Ghazali. I am very grateful to you. I will also recognise Mishaal Elbasheer, Ng Phei Ying, and Siti Noresah Mohd Zaim, who were close friends during my study period. I must equally acknowledge Usmanu Danfodiyo University, Sokoto, Nigeria, for giving me study leave, and the Tertiary Education Trust Fund (TETFund) of Nigeria for providing me with financial support. I am grateful to my father, Isah Abubakar Tsamiya; my mother, Hadiza Suleiman Tsamiya; my wife, Zainab Isah Yola; my daughters, namely Khadija and Fatima; and all my friends for their unwavering patience, support, and prayers while I am far away from them pursuing my studies. May Almighty Allah reward you abundantly, amen.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **ii**

TABLE OF CONTENTS ..... **iii**

LIST OF TABLES ..... **x**

LIST OF FIGURES ..... **xii**

LIST OF ABBREVIATIONS ..... **xix**

LIST OF APPENDICES ..... **xxiii**

ABSTRAK ..... **xxiv**

ABSTRACT ..... **xxvi**

CHAPTER 1 INTRODUCTION **1**

1.1 Background of the Study 1

1.2 Problem Statement of the Study 2

1.3 The Rationale of the Study 4

1.4 Research Questions 5

1.5 Research Hypotheses 5

1.6 Aim and Objectives of the Study 6

1.6.1 General Objective/Aim of the Study 6

1.6.2 Specific Objectives of the Study 6

CHAPTER 2 LITERATURE REVIEW **7**

2.1 Epidemiology of CRC 7

2.1.1 Global Incidence of CRC and Economic Burdens 7

2.1.2 Incidence of the CRC in Asia 8

2.1.3 Incidence of the CRC and Economic Burdens in Malaysia 9

2.2 Epidemiology of EOCRC 10

2.2.1 Global Incidence of EOCRC 10

2.2.2 Incidence of EOCRC in Asia 11

2.2.3 Incidence of EOCRC in Malaysia 11
2.3 Anatomy of the Large Intestine 12
2.4 Risk Factors of the EOCRC 14
2.4.1 Non-modifiable Risk Factors of the EOCRC 14
2.4.2 Modifiable Risk Factors of the EOCRC 15
2.5 Symptoms of EOCRC 16
2.6 Types of Histology in EOCRC 17
2.6.1 Grading of Adenocarcinoma 17
2.6.2 Mucinous Adenocarcinoma 17
2.6.3 Signet Ring Cell Adenocarcinoma/Carcinoma 18
2.6.4 Medullary Carcinoma and Other Less Commonly Occurring Histologies 18
2.7 Pathogenesis and Molecular Basis of EOCRC 18
2.7.1 Types of CRC based on Heredity 19
2.7.1(a) Hereditary CRC 19
2.7.1(b) Familial CRC 21
2.7.1(c) Sporadic CRC 21
2.7.2 Molecular Pathways of CRC Carcinogenesis 23
2.7.2(a) Chromosomal Instability (CIN) Pathway 24
2.7.2(b) Microsatellite Instability (MSI) Pathway 25
2.7.2(c) The CpG Island Methylator Phenotype (CIMP) Pathway 27
2.8 DNA Mismatch Repair System and Its Role in CRC 28
2.8.1 Composition of the DNA MMR System 29
2.8.2 The Working Mechanism of the DNA MMR System 29
2.8.3 The Implications of the Loss of DNA MMR Proteins 29
2.8.4 Various Methods of Detecting dMMR/MSI in CRC 30
2.8.4(a) MMR Protein Expression Detection by IHC 30

2.8.4(b) MSI Testing using Polymerase Chain Reaction (PCR) Assay 32

2.8.4(c) Next-Generation Sequencing (NGS) 33

2.8.5 Types of CRC with dMMR 33

2.8.5(a) Sporadic dMMR CRC 33

2.8.5(b) Hereditary dMMR CRC 33

2.8.6 Clinical Relevance of Assessing the MMR Status 33

2.9 BRAFV600E Mutation and Its Relationship with dMMR in EOCRC 34

2.9.1 The Role of BRAFV600E Mutation in CRC Pathogenesis 35

2.9.2 Types of BRAF Mutations 36

2.9.3 Clinical Importance of Testing for BRAFV600E Status 37

2.10 Tregs as a Component of the Immune System 38

2.10.1 Types of Immune System Response 38

2.10.1(a) Innate immune system response 39

2.10.1(b) Adaptive immune system response 39

2.10.2 Types of T-lymphocytes and their Functions 40

2.10.3 Development and Activation of Tregs 42

2.10.4 Classification of FOXP3CD4^{+} Tregs and Their Implications for CRC Prognosis 44

2.10.5 The Tumour-Infiltrating Lymphocytes (TILs) as Components of the CRC Tumour Microenvironment (TME) 45

2.11 FOXP3 Structure and Function 47

2.11.1 FOXP3 Gene and Protein Structure 47

2.11.2 Functions of FOXP3 Protein Domains 48

2.11.2(a) N-terminal repressor domain 48

2.11.2(b) Zinc-finger domain 49

2.11.2(c) Leucine-zipper domain 49

2.11.2(d) Forkhead (FKH) DNA-binding domain 49

2.11.3 Isoforms of the *FOXP3* 49

2.11.3(a) *FOXP3FL* isoform 50

2.11.3(b) *FOXP3△2* isoform 50

2.11.3(c) *FOXP3△2△7* isoform 50

2.11.4 FOXP3 Expression and Regulation 51

2.11.5 The Suppressive Function of FOXP3Tregs in CRC 52

2.11.6 Prognostic Role of FOXP3 Expression in CRC 53

2.12 Gene Expression using RT-qPCR 54

2.12.1 Principle of RT-qPCR using SYBR Green 55

2.12.2 Steps in RT-qPCR 56

2.12.2(a) Denaturation 57

2.12.2(b) Annealing 57

2.12.2(c) Extension 57

2.12.3 The Major Terms Used in RT-qPCR Analysis 57

2.12.4 Types of Quantification of the RT-qPCR Target Product 59

2.12.4(a) Absolute quantification 59

2.12.4(b) Relative quantification 59

CHAPTER 3 MATERIALS AND METHODS 61

3.1 Materials of the Study 61

3.1.1 Materials for Epidemiology and Clinicopathological Studies 61

3.1.2 Materials for IHC Technique 61

3.1.3 Materials for RT-qPCR 62

3.1.3(a) Total RNA Extraction using NucleoSpin Kit 62

3.1.3(b) cDNA Synthesis using Sensifast cDNA Kit 64

3.1.3(c) Materials for RT-qPCR 64

3.2 Methodology of the Study 66

3.2.1 Ethical Clearance 66

3.2.2 Study Design 66

3.2.3 Subject Eligibility Criteria 67

3.2.4 Sample Collection 68

3.2.5 Retrieval of EOCRC Epidemiology and Clinicopathological Features 68

3.2.6 Immunohistochemistry Determination of FOXP3, BRAFV600E, and MMR Protein in EOCRC 69

3.2.7 Determination of *FOXP3* Gene Expression using RT-qPCR 74

3.3 Data Analysis 89

3.3.1 Data Analysis for Epidemiological and Clinicopathological Data 89

3.3.2 Data Analysis for FOXP3, BRAFV600E, and MMR Proteins from IHC 89

3.3.3 Data Analysis for *FOXP3* Gene Isoforms Expression from RT-qPCR 89

CHAPTER 4 RESULTS 91

4.1 Incidence and Clinicopathological Characteristics of EOCRC 91

4.1.1 Overview 91

4.1.2 Prevalence of EOCRC in HUSM from 2013 to 2021 91

4.1.3 Demographic and Clinical Characteristics of EOCRC 92

4.1.4 Histopathological Characteristics of EOCRC 95

4.1.5 Correlation between Demographic and Clinical Characteristics of EOCRC 97

4.1.6 Summary of the Findings 100

4.2 Expression of BRAFV600E and MMR protein status in EOCRC patients . 101

4.2.1 Overview 101

4.2.2 Results of BRAFV600E Staining Optimisation 101

4.2.3 Results of MMR Antibodies Staining Optimisation 102

4.2.4 BRAFV600E and MMR Proteins Status in EOCRC 106

4.2.5 Summary of the Findings 113

4.3 FOXP3 Expression and Its Correlation with BRAFV600E and MMR 115

4.3.1 Optimisation of FOXP3 Staining in the IHC Technique 115

4.3.2 Determination of the Cut-off Value of FOXP3 Expression into Low and High using the ROC Curve 116

4.3.3 Expression of FOXP3 and comparison with MMR and BRAFV600E Status 118

4.3.4 The Density of FOXP3 Expression with Demographic and Clinicopathological Characteristics of EOCRC Patients 120

4.3.5 Summary of the Findings 124

4.4 The Gene Expression Pattern of *FOXP3* in EOCRC Patients 126

4.4.1 Overview 126

4.4.2 Quantification of total RNA extracted from the FFPE archival EOCRC tissues 127

4.4.3 Optimisation of *GAPDH*, *FOXP3FL*, and *FOXP3⊿2* Primers Annealing Temperature 129

4.4.4 Temperature Gradient RT-qPCR of *FOXP3FL* Primers and the products gel electrophoresis 131

4.4.5 Optimisation of *GAPDH*, *FOXP3FL*, and *FOXP3⊿2* Primers at 62^0C using 2-Step and 3-Step Cycling 133

4.4.6 Expression of *FOXP3FL* and *FOXP3⊿2* Isoforms 137

4.4.7 Summary of the Findings 140

CHAPTER 5 DISCUSSION 142

5.1 The Incidence and Clinicopathological Characteristics of EOCRC 142

5.1.1 Prevalence, Age Range, and Gender Occurrence of EOCRC 142

5.1.2 Demographic and Clinical Data of the EOCRC 144

5.1.3 Clinicopathological Characteristics of EOCRC 146

5.2 Expression of BRAFV600E and MMR Protein Status 148

5.2.1 MMR and BRAFV600E Status in EOCRC Patients 148

5.2.2 EOCRC demographic and clinicopathological characteristics between pMMR and dMMR 151

5.3 FOXP3 Expression and Its Correlation with BRAFV600E and MMR 152

5.3.1 FOXP3 Protein and its association with BRAFV600E and MMR Status in EOCRC Patients 152

5.3.2 Comparison of Demographic and Clinicopathological Characteristics of EOCRC with FOXP3 Protein Expression..... 154

5.4 The Gene Expression Pattern of *FOXP3* in EOCRC Patients 155

5.4.1 RNA Extraction and Optimisation of the RT-qPCR..... 155

5.4.2 Expression of *FOXP3* Isoforms in EOCRC..... 157

CHAPTER 6 CONCLUSION AND FUTURE RECOMMENDATIONS.... 160

6.1 Conclusions 160

6.2 Limitations and Recommendations for Future Research 161

REFERENCES..... 163

APPENDICES

LIST OF TABLES

Page

Table 2.1: Structure and Molecular Weight of the 3 *FOXP3* Isomers 50

Table 2.2: The level of FOXP3 expression in lymphocytes 51

Table 3.1: Monoclonal Antibodies Used for the IHC Technique 62

Table 3.2: Components of the NucleoSpin RNA Extraction Kit 63

Table 3.3: Primer Design for *FOXP3FL*, *FOXP3⊿2*, and the *GAPDH* (Control) 64

Table 3.4: Details of Manufactured Primers from Integrated DNA Technologies 65

Table 3.5: Reconstitution of the Lyophilised Primers to 100µM Volume 66

Table 3.6: Primary antibodies used in the IHC staining, dilution, and incubation time 71

Table 3.7: Immunoreactive Score (IRS) of IHC Staining (FOXP3) 73

Table 3.8: Contents and Volume of the Sample for Gel Electrophoresis 78

Table 3.9: Contents and Volume of the cDNA Synthesis Tube 79

Table 3.10: Protocol of 2-step cycling used for the optimisation of the primers .. 84

Table 3.11: Protocol of 3-step cycling used for the optimisation of the primers .. 85

Table 3.12: Triplicates qPCR Set-up for Each Test Sample 86

Table 3.13: RT-qPCR Reaction Mixture Composition 87

Table 4.1: Demographic and clinical data of EOCRC (n = 65) 94

Table 4.2: Histopathological Characteristics of EOCRC 96

Table 4.3: Correlation of Demographic and Clinical Characteristics of EOCRC within age groups 97

Table 4.4: Correlation of Histopathological Characteristics of EOCRC with age groups 99

Table 4.5: Comparison of EOCRC demographic and clinicopathological characteristics between pMMR and dMMR 111

Table 4.6: Area under the ROC Curve of the FOXP3 Expression 117

Table 4.7: Coordinates of the FOXP3 Expression ROC Curve 117

Table 4.8: FOXP3Treg expression in EOCRC cases with MMR and BRAF V600E status 120

Table 4.9: Comparison of Demographic Characteristics between the Low and High FOXP3 Expression 121

Table 4.10: Comparison of Clinicopathological Characteristics between Low and High FOXP3 Expression 123

Table 4.11: Extracted RNA Concentration and Purity for FOXP3 Gene Expression 127

Table 4.12: Comparison of FOXPFL gene expression with some selected parameters 139

Table 4.13: Comparison of *FOXP⊿2* gene expression with some selected parameters 140

LIST OF FIGURES

Page

Figure 2.1: Global Incidence of CRC new cases in 2020 (yellow colour) for both sexes and all age groups (Ferlay et al., 2020). 8

Figure 2.2: Asian incidence rate of CRC (green colour) for both sexes and all ages (Ahmad et al., 2021). 9

Figure 2.3: Stages of CRC diagnosis in Malaysia among males and females (Ministry of Health Malaysia, 2021). 10

Figure 2.4: The anatomy and classification of large intestine parts (Lin et al., 2016). 13

Figure 2.5: CRC incidence rates for hereditary, familial, and sporadic types (Tan, 2018). 19

Figure 2.6: Molecular pathways of CRC pathogenesis (Modified from Ballester et al., 2016) 23

Figure 2.7: Molecular pathways of CRC involving the conventional pathway (CIN), serrated pathway (CIMP), and MSI pathway with associated affected genes (Huang & Yang, 2022). 24

Figure 2.8: CRC progression along the chromosomal instability pathway and affected genes connected to microsatellite instability (Cerrito & Grassilli, 2021). 25

Figure 2.9: The carcinogenesis of the MSI pathway mechanism in colorectal cancer (Pastor & Schlom, 2021). 26

Figure 2.10: A diagram showing the potential serrated pathway through which CIMP-high CRCs with microsatellite instability grow through methylation of the MLH1 gene (Rex et al., 2012). 28

Figure 2.11: The schematic representation of pMMR and dMMR complex mechanisms during DNA replication 30

Figure 2.12: The principle of IHC staining in demonstrating targeted protein 31

Figure 2.13: Schematic representation of CRC diagnostic flow based on BRAF V600E and MMR protein expression. *Keys - CIN: Chromosomal Instability; CIMP: CpG Island Methylator Phenotype; +: Positive status/intact; -Negative status/loss*.....32

Figure 2.14: MAPK signaling pathway in BRAFV600E CRC, where RAS activates the RAF family proteins (BRAF). Activated RAF proteins cause MEK1/2 proteins to be phosphorylated and activated, which in turn leads to the phosphorylation and activation of ERKs, promoting cellular proliferation and tumour growth (Caputo et al., 2019).....36

Figure 2.15: Classification and signalling pathways of BRAF mutations (Kono et al., 2020).....37

Figure 2.16: The innate immune system and adaptive immune response show the differentiation of T-cells (Kumar, 2021).....39

Figure 2.17: Different types of T-lymphocytes after the dendritic cell (DC) presents the antigen to the naïve cell, which then becomes activated and differentiated into cytotoxic T-cells (CD8 cells), T-helper cells (CD4 cells), and Tregs (with FOXP3 as their marker). Finally, a few T-cells will be converted to memory T-cells to retain the memory of the previously exposed antigen (Jia et al., 2019). 41

Figure 2.18: Developmental stages of T-cells from the progenitor in the bone marrow to differentiation and maturation in the thymus (Sapkota, 2022). 43

Figure 2.19: Development of pTregs from naïve $CD4^{+}$Tregs in the peripheral tissues triggered by encountering foreign antigens, suppressor cytokines, and other specific conditions (Pereira et al., 2017). 44

Figure 2.20: FOXP3Tregs phenotypes in CRC progression showing (A) nTregs in a normal colon, (B) Less suppressive FOXP3Tregs in the adenoma stage associated with protection against tumour formation, (C) Highly suppressive FOXP3Tregs in the carcinoma stage (Olguín et al., 2020).....45

Figure 2.21: The colorectal cancer tumour microenvironment showing the components of tumour infiltrating lymphocytes and other lymphoid structures (Bai et al., 2022). 46

Figure 2.22: The structure of the FOXP3 gene and protein on the human X-human chromosome (Pereira et al., 2017). 48

Figure 2.23: Role of FOXP3+ Tregs in CRC immune suppression (Grimmig et al., 2013). 53

Figure 2.24: The working principle of SYBR green (A) and probe (B) during the RT-qPCR procedure. Melt curve analysis is also used in RT-qPCR to measure the amplicon product's dissociation and determine the primers' specificity (Adams, 2020). 56

Figure 2.25: The qPCR amplification curve showing initiation, exponential, and plateau phases (Adams, 2020)..... 59

Figure 3.1: Detailed contents of the Sensifast cDNA synthesis kit (extracted from the manufacturer's manual)..... 64

Figure 3.2: Flow chart of the study, indicating how the study's objectives were achieved and the methodologies used. 67

Figure 3.3: An IHC staining of FOXP3 indicating the grading of the staining intensity with (A) Negative staining with score of 0 (B) Mild staining with score of +1 Moderate staining with score of +2 (C) Moderate staining with score of +2 (D) Strong staining with score of +3 (Wang et al., 2010). 72

Figure 3.4: Workflow chart of *FOXP3* gene expression from RNA extraction to cDNA synthesis, used as a template for the RT-qPCR assay, and fluorescence was detected using SYBR Green. The data was quantified using relative quantification to determine the quantity of *FOXP3* gene expression..... 74

Figure 3.5: Assessment of extracted RNA concentration and purity using the ThermoScientific MultiSkan SkyHigh spectrophotometer. 77

Figure 3.6: Synthesis of cDNA using a Bio-Rad thermal cycler..... 80

Figure 3.7: RT-qPCR tube with cDNA template and other contents of the master mix (extracted from the manufacturer's manual).....81

Figure 3.8: Melt curve RFU on the Y-axis and temperature on the X-axis 82

Figure 3.9: Melt peak with -d(RFU)/dT on the Y-axis and temperature on the X-axis 83

Figure 3.10: Temperature gradient RT-qPCR showing uniform temperature among the wells in all the steps except in the annealing step, where the temperature was varied along the column of the wells as indicated by the arrow. 83

Figure 3.11: Conduction of 2-step cycling method and melt curve analysis using CFX Opus 96 BIO-RAD Real-Time PCR System 88

Figure 4.1: Incidence of EOCRC, showing the number of cases per year 92

Figure 4.2: Number of EOCRC cases according to age range and gender.....93

Figure 4.3: BRAF IHC optimization at pH 6.0 overnight. (A) Dilution of 1:100 with mouse linker; (B) Dilution of 1:100 without mouse linker; (C) Dilution of 1:200 with mouse linker; (D) Dilution of 1:200 without mouse linker. (X200) 102

Figure 4.4: MLH1 IHC optimization at pH 9.0 (A) Dilution of 1:50 without mouse linker for 1 hour (B) Dilution of 1:50 with mouse linker overnight (C) Dilution of 1:100 without mouse linker for 1 hour (D) Dilution of 1:100 with mouse linker overnight. (X200).....103

Figure 4.5: MSH2 IHC optimisation at pH 9.0 (A) Dilution of 1:50 with mouse linker for 1 hour (B) Dilution of 1:100 with mouse linker for 1 hour (C) Dilution of 1:200 with mouse linker for 1 hour (D) Dilution of 1:300 with mouse linker for 1 hour. (X200). 104

Figure 4.6: MSH6 IHC optimisation at pH9.0 for 1 hour (A) Dilution of 1:100 with rabbit linker (B) Dilution of 1:100 without rabbit linker (C) Dilution of 1:200 with rabbit linker (D) Dilution of 1:200 without rabbit linker. (X200).....105

Figure 4.7: PMS2 IHC optimisation at pH9.0 (A) Dilution of 1:50 without linker for 1 hour (B) Dilution of 1:50 with linker overnight (C) Dilution of 1:100 without linker for 1 hour (D) Dilution of 1:100 with linker overnight. (X200) 106

Figure 4.8: Colonic tumour cells with negative expression of MLH1 (blue arrow) in the presence of nuclear staining of lymphocytes serving as an internal control (black arrow). (B) Colonic tumour cells with nuclear positive expression of MLH1 (blue arrow) (C) Colonic tumour cells with negative expression of PMS2 (blue arrow) in the presence of stained lymphocytes serving as internal controls (black arrow). (D) Colonic tumour cells with nuclear positive expression of PMS2 (blue arrow). (X200) 107

Figure 4.9: (A) Colonic tumour cells with negative expression of MSH2 (blue arrow) in the presence of nuclear staining of lymphocytes serving as an internal control (black arrow). (B) Colonic tumour cells with nuclear positive expression of MSH2 (blue arrow) (C) Colonic tumour cells with negative expression of MSH6 (blue arrow) in the presence of stained lymphocytes serving as an internal control (black arrow). (D) Colonic tumour cells with nuclear positive expression of MSH6 (blue arrow). (X200) 108

Figure 4.10: Pie chart showing the percentage of MMR protein negativity 109

Figure 4.11: BRAF V600E expression pattern with (A) Negative BRAF, (B) Mild expression, (C) Moderate expression, and (D) Strong expression. (X200) 110

Figure 4.12**:** FOXP3 IHC optimisation at pH 6.0 for 1 hour using mouse linker. (A) Dilution of 1:50 (B) Dilution of 1:100 (C) Dilution of 1:200 (D) Dilution of 1:300. (X200) 116

Figure 4.13: The ROC curve of FOXP3 expression with specificity on the x-axis and sensitivity on the y-axis. 117

Figure 4.14: Nuclear FOXP3 protein expression (blue arrows) by the FOXP3Tregs in EOCRC (A) Negative FOXP3 expression (B) Mild FOXP3 expressing by the FOXP3Tregs (C) Moderate

FOXP3 expressing by the FOXP3Tregs (D) Marked FOXP3 expressing by the FOXP3Tregs. (X200). 118

Figure 4.15: FOXP3 IHC staining with (A) Low nuclear expression of FOXP3 by the lymphocytes (arrow) (B) High nuclear expression of FOXP3 by the lymphocytes (arrow). (X200) 119

Figure 4.16: Gel electrophoresis of extracted RNA from the FFPE of the selected samples 129

Figure 4.17: Amplification curve and melt peak of *GAPDH* at 60^0C 129

Figure 4.18: Amplification curve and melt peak of *FOXP3Δ2* at 60^0C 130

Figure 4.19: Amplification curve and melt peak of *FOXP3* full length at 60^0C . 131

Figure 4.20: Temperature gradient RT-qPCR of *FOXP3FL* using five varying temperatures 132

Figure 4.21: Gel electrophoresis of the *FOXP3FL* temperature gradient PCR products 132

Figure 4.22: Amplification curve of *GAPDH* using 2-step with Cq of 31.25 and 3-step with Cq of 32.03 at 62°C 133

Figure 4.23: A single peak observed in the 2-step and 3-step Melt Curve of the *GAPDH* at 62°C 134

Figure 4.24: Amplification curve of *FOXP3⊿2* using 2-step with Cq of 32.71 and 3-step with Cq of 36.65 at 62°C 134

Figure 4.25: A single peak observed in the 2-step and 3-step cycling of the FOXP3△2 at 62°C 135

Figure 4.26: Amplification curve of *FOXP3FL* using 2-step with Cq of 32.91 and 3-step with Cq of 35.58 at 62°C 136

Figure 4.27: Multiple peaks observed in the 2-step and 3-step cycling of *FOXP3FL* at 62°C 136

Figure 4.28: Melting curve analysis of *FOXP3FL* represented by (A), and gel electrophoresis of the *FOXP3FL* product indicating a single band, represented by (B). 137

Figure 4.29: Relative expression of the *FOXP3* mRNA in relation to the normal colon control free of EOCRC from young patients, showing higher expression of *FOXP3FL* isoform than *FOXP3⊿2* isoform by the EOCRC patients. Data was presented using the mean of relative expression ±SEM 138

LIST OF ABBREVIATIONS

AP	Activator Protein
APC	Adenomatous Polyposis Coli
BMI	Body Mass Index
bp	base pair
BRAF	v-raf murine sarcoma viral oncogene homolog B1
CD	Cluster of Differentiation
cDNA	Complementary Deoxyribonucleic Acid
CEA	Carcinoembryonic Antigen
CI	Confidence Interval
CIMP	CpG Island Methylator Phenotype
CIMP-L	CIMP-Low
CIMP-H	CIMP-High
CIN	Chromosomal Instability
CRC	Colorectal Cancer
Ct	Cycle of threshold
DAB	Diaminobenzidine
DC	Dendritic Cell
DM	Diabetes Mellitus
dMMR	Deficient Mismatch Repair
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded DNA
EGFR	Epidermal Growth Factor Receptor
EOCRC	Early-onset Colorectal Cancer
FAP	Familial Adenomatous Polyposis
FDRs	First Degree Relatives
FFPE	Formalin-Fixed Paraffin-Embedded
FKH	Fork-head
FOXP3	Forkhead box Protein 3
FOXP3△2	FOXP3 with deletion of exon 2
FOXP3△7	FOXP3 with deletion of exon 7
FOXP3△2△3	FOXP3 with deletion of exon 2 and exon 3

FOXP3△2△7	FOXP3 with deletion of exon 2 and exon 7
FOXP3FL	Forkhead box Protein 3 Full Length
g	gramme
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GLOBOCAN	Global Cancer Observatory
H&E	Haematoxylin and Eosin
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
HRP	Horseradish Peroxidase
HUSM	Hospital Universiti Sains Malaysia
IHC	Immunohistochemistry
IL	Interleukin
IPEX	Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked
IRS	Immunoreactive Scoring
kDa	kilo Dalton
LIS	Laboratory Information System
LOCRC	Late-onset Colorectal Cancer
LZ	Leucine Zipper
MAPK	Mitogen-Activated Protein Kinase
MHC	Major Histocompatibility Complex
MIQE	Minimum Information for Publication of RT-qPCR Experiment
mL	millilitre
MLH1	MutL Homolog 1
MMR	Mismatch Repair
mRNA	messenger Ribonucleic Acid
MSH2	MutS Homolog 2
MSH6	MutS Homolog 6
MSI	Microsatellite Instable
MSI-L	MSI-Low
MSI-H	MSI-High
MSS	Microsatellite stable
MYR	Malaysian Ringgit
NCBI	National Centre for Biotechnology Information
NCCN	National Comprehensive Cancer Network

NES	Nuclear Export Signal
NFAT	Nuclear Factor of Activated T-cells
ng	nanogramme
NGS	Next Generation Sequencing
nm	namometres
NK	Natural Killer
NTC	Non-Template Control
No-RT	No-Reverse Transcriptase
PCR	Polymerase Chain Reaction
pMMR	Proficient Mismatch Repair
PMS2	Postmeiotic Segregation 2
pTregs	Peripheral regulatory T-cells
rDNase	Recombinant DNase
RFU	Relative Fluorescent Unit
RG	Reference Gene
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristics
rpm	Revolution per minute
RT	Room Temperature
RT-qPCR	Reverse Transcription quantitative real-time PCR
SD	Standard Deviation
TAM	Tumour-Associated Macrophages
TBE	Tris-Borate-EDTA
TBS	Tris Buffered Saline
TCR	T-Cell Receptor
TILs	Tumour Infiltrating Lymphocytes
TGF-β	Transforming Growth Factor - β
TME	Tumour Microenvironment
T_a	Temperature of annealing
TNM	Tumour, Node, Metastasis
Tregs	Regulatory T-lymphocytes cells
USD	United States Dollar
µg	microgramme
µL	microlitre

µM	micromolar
WHO	World Health Organisation
ZF	Zinc finger

LIST OF APPENDICES

Appendix A JEPeM Ethical Approval Letter

Appendix B List of Publications

Appendix C List of Conference Presentations

Appendix D List of Awards

PERANAN FOXP3 BERKAITAN DENGAN STATUS PROTEIN BRAFV600E DAN MMR BAGI KANSER KOLOREKTAL BERMULA AWAL

ABSTRAK

Peningkatan global dalam EOCRC adalah membimbangkan, kerana patogenesisnya masih dalam siasatan. Di Malaysia, EOCRC dalam pesakit CRC di bawah umur 50 tahun biasanya didiagnosis dalam peringkat lanjut, selalunya menunjukkan ciri histologi yang agresif seperti pembezaan yang lemah dan histologi mucinous atau signet. Perkembangan EOCRC dipengaruhi oleh penindasan imun sel perumah, dengan sel pengawalseliaan T (Tregs) memainkan peranan penting. Protein kotak forkhead 3 (FOXP3) ialah faktor transkrip dan penanda kritikal untuk CD4+ dan CD25+ Treg. Kajian itu mengkaji protein FOXP3 dan ekspresi gen bersama-sama dengan v-raf murine sarcoma virus homolog B1 (BRAF), khususnya mutasi BRAFV600E yang menggantikan valine (V) dengan asid glutamik (E) pada kedudukan 600, dan status protein pembaikan tidak sepadan (MMR) dalam kalangan pesakit EOCRC. Kajian itu menganalisis pesakit EOCRC bawah 50 tahun di HUSM antara 2013 dan 2021, berdasarkan parameter umur daripada rekod perubatan dan Sistem Maklumat Makmal (LIS). Kajian itu menggunakan antibodi monoklonal untuk mengotorkan protein FOXP3, BRAFV600E, dan MMR secara imunohistokimia, yang kemudiannya dijaringkan menggunakan pemarkahan imunoreaktif (IRS). Jumlah RNA tisu FFPE telah ditukar kepada cDNA, dan ekspresi mRNA FOXP3 diukur menggunakan RT-qPCR untuk panjang penuh FOXP3 (FOXFL) dan FOXP3 dengan exon 2 dipadamkan (FOXP3△2). Antara 2013 dan 2021, HUSM mendiagnosis 65 pesakit EOCRC, dengan prevalens 20.4%, terutamanya ditemui pada kolon kiri dan kerap pada individu tanpa sejarah keluarga CRC. Adenokarsinoma yang dibezakan sederhana (81.5%) adalah histologi yang paling biasa didiagnosis pada peringkat

lanjut, diikuti oleh adenokarsinoma mucinous (15.38%) dan karsinoma cincin tanda (6.2%). Kajian itu mendapati perkaitan yang signifikan (p = 0.02) dalam penyusupan limfosit tumor yang diwarnai dengan haematoxylin dan eosin (H&E) merentasi kumpulan umur yang berbeza. Limfosit yang menyusup tumor telah diiktiraf oleh nukleus berwarna biru tua, kecil dan bulat dengan sejumlah kecil sitoplasma berwarna merah jambu dalam stroma tumor EOCRC. 53.8% daripada pesakit EOCRC mempunyai MMR (pMMR) yang mahir, dengan positif dalam keempat-empat protein MMR yang diuji, manakala 46.2% mempunyai kekurangan MMR (dMMR), dengan negatif dalam satu atau lebih daripada empat protein MMR yang diuji. Protein BRAFV600E telah diekspresikan secara berlebihan dalam 69.2% daripada kes EOCRC. Protein FOXP3 dinyatakan oleh 93.8% pesakit EOCRC, manakala 6.2% adalah negatif. Pesakit dengan pMMR dan BRAFV600E positif menunjukkan ekspresi protein FOXP3 yang lebih tinggi (54.2%) berbanding mereka yang mempunyai dMMR dan BRAFV600E positif (22.9%). Kajian itu mendedahkan korelasi yang signifikan antara ekspresi protein FOXP3 yang tinggi, jenis histologi, gred tumor, MMR, dan status BRAFV600E ($p \leq 0.05$). FOXP3FL ialah varian nyata utama FOXP3, dengan ungkapan relatif min 14.86 ± 6.5. Sebaliknya, FOXP3$\triangle$2 kurang dinyatakan, dengan ungkapan relatif min 1.03 ± 0.66. Kedua-dua varian FOXP3 diperhatikan dinyatakan oleh EOCRC. Ekspresi varian FOXPFL oleh EOCRC telah meningkat dengan ketara (p = 0.034), kerana kebanyakan pesakit mempunyai CRC sporadis dan bukannya keturunan. Kajian ini menyerlahkan kepentingan ekspresi gen dan protein FOXP3, BRAFV600E, dan status protein MMR dalam memahami EOCRC dalam pesakit HUSM. Kajian itu mengesyorkan penyelidikan lanjut menggunakan teknik lain untuk menjelaskan peranan imunosupresif FOXP3 dalam EOCRC.

THE ROLE OF FOXP3 IN RELATION TO BRAFV600E AND MMR PROTEIN STATUS IN EARLY-ONSET COLORECTAL CANCER

ABSTRACT

The global rise in EOCRC is concerning, as its pathogenesis is still under investigation. In Malaysia, EOCRCs in CRC patients below the age of 50 are typically diagnosed in advanced stages, often exhibiting aggressive histologic features like poor differentiation and mucinous or signet histology. EOCRC progression is influenced by the immune suppression of host cells, with T regulatory cells (Tregs) playing a crucial role. Forkhead box protein 3 (FOXP3) is a transcriptional factor and a critical marker for CD4+ and CD25+ Tregs. The study examined FOXP3 protein and gene expression in conjunction with v-raf murine sarcoma viral oncogene homolog B1 (BRAF), specifically the BRAFV600E mutation that replaces valine (V) with glutamic acid (E) at position 600, and mismatch repair (MMR) protein status among EOCRC patients. The study analyzed EOCRC patients under 50 at HUSM between 2013 and 2021, based on age parameters from medical records and the Laboratory Information System (LIS). The study utilised monoclonal antibodies to immunohistochemically stain FOXP3, BRAFV600E, and MMR proteins, which were then scored using immunoreactive scoring (IRS). FFPE tissues' total RNA was converted to cDNA, and FOXP3 mRNA expression was measured using RT-qPCR for FOXP3 full length (FOXFL) and FOXP3 with exon 2 deleted (*FOXP3△2*). Between 2013 and 2021, HUSM diagnosed 65 EOCRC patients, with a 20.4% prevalence, primarily found on the left colon and frequently in individuals without a family history of CRC. Moderately differentiated adenocarcinoma (81.5%) was the most common histology diagnosed at the advanced stage, followed by mucinous adenocarcinoma (15.38%) and

signet ring carcinoma (6.2%). The study found a significant association ($p = 0.02$) in the infiltration of tumour lymphocytes stained with haematoxylin and eosin (H&E) across different age groups. The tumour-infiltrating lymphocytes were recognized by their dark blue-stained, small, and round nuclei with a small amount of cytoplasm stained pink within the tumour stroma of the EOCRC. 53.8% of the EOCRC patients had proficient MMR (pMMR), with positivity in all four MMR proteins tested, while 46.2% had deficient MMR (dMMR), with negativity in one or more of the four MMR proteins tested. The BRAFV600E protein was overexpressed in 69.2% of the EOCRC cases. The FOXP3 protein was expressed by 93.8% of EOCRC patients, while 6.2% were negative. Patients with pMMR and BRAFV600E positive showed higher FOXP3 protein expression (54.2%) than those with dMMR and BRAFV600E positive (22.9%). The study revealed a significant correlation between high FOXP3 protein expression, histological types, tumour grade, MMR, and BRAFV600E status ($p \leq 0.05$). *FOXP3FL* was the major expressed variant of the *FOXP3*, with a mean relative expression of 14.86±6.5. In contrast, *FOXP3△2* was less expressed, with a mean relative expression of 1.03±0.66. The two variants of the *FOXP3* were observed to be expressed by the EOCRC. The *FOXPFL* variant expression by the EOCRC was significantly increased ($p = 0.034$), as most patients have sporadic rather than hereditary CRC. In conclusion, the study provided valuable insights into the characteristics and molecular mechanisms of EOCRC in HUSM patients. The findings highlight the importance of FOXP3 protein and gene expression, as well as BRAFV600E and MMR protein status, in understanding EOCRC and its potential prognostic implications. This study recommends further research using other molecular techniques to elucidate the anti-tumour and immunosuppressive roles of FOXP3 in EOCRC.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Colorectal cancer (CRC) is the third most prevalent cancer globally and the second most lethal malignancy (Mauri et al., 2019; Sharma et al., 2022; Syed et al., 2019; Xi & Xu, 2021). Malaysia is among the Asian countries with the highest CRC prevalence; CRC was ranked the second leading cancer in the nation (Ferlay et al., 2020; Magaji et al., 2014; Muhamad et al., 2023). At a time when late-onset CRC (LOCRC) occurrence in patients of 50 years and above was observed to be decreasing in many parts of the world, the global incidence of Early-onset CRC (EOCRC) occurring in patients below the age of 50 is seen to be rapidly increasing (Alyabsi et al., 2021; Azar et al., 2021; Kim et al., 2021; Gausman et al., 2020). The high incidence rate of EOCRC is particularly concerning because its pathogenesis remains under investigation, and it affects young people who currently lack a well-established population-based screening programme in Malaysia (Chandran et al., 2020; Karikalan et al., 2021; Azzani et al., 2019). Some of the CRC screening methods include stool-based tests such as faecal immunological test (FIT) and DNA testing (FIT-DNA); other screening methods involved using visual techniques such as colonoscopy, sigmoidoscopy and computed tomography (CT) colonography (Chandran et al., 2020, 2022; Lin et al., 2016; Venugopal & Carethers, 2022).

EOCRC progression is associated with immunosuppression of host cell-mediated immune responses (Ganapathi et al., 2014; Strasser et al., 2019). Regulatory T-cells (Tregs) play a crucial role in the immunosuppression of the immune system in EOCRC patients, and FOXP3 is a key marker for them (Allan et al., 2005; Jia et al., 2019; Nam et al., 2018). Tregs are primarily characterised by FOXP3 as their specific

marker, including the CD4+ and CD25+ subsets (Alessandra et al., 2020; Masugi et al., 2017). FOXP3 induces immunosuppressive functions through either direct contact with the cells or by secreting transforming growth factor-β (TGF-β) and interleukin-10 (IL-10), respectively (Abd-Allah et al., 2022). *FOXP3* has variants with different biological properties in the tumour microenvironment (Nam et al., 2018; Wozniakova et al., 2022).

The extent of the FOXP3Tregs infiltration into the colon mucosa of EOCRC has been shown to depend on the status of the mismatch repair (MMR) protein and BRAFV600E (Bupathi & Wu, 2016; Mei et al., 2022; Yambert et al., 2022). There is also growing evidence that deficient MMR (dMMR) of EOCRC differs from proficient MMR (pMMR) of EOCRC in both molecular and clinicopathological characteristics (Mei et al., 2022).

1.2 Problem Statement of the Study

Despite breakthroughs in understanding CRC pathogenesis and the development of new treatment modalities, CRC remains one of the world's significant public health burdens (Karikalan et al., 2021; Mannucci et al., 2019; Wong et al., 2019). Malaysia is experiencing an increasing prevalence of EOCRC; Ibrahim et al. (2020) reported an EOCRC prevalence of 14.5% in Northern Malaysia, with 893 cases of CRC diagnosed in individuals under the age of 50 from 2007 to 2017. Among individuals with EOCRC, 20% have a familial history of CRC. In comparison, 30% have mutations in genes such as Breast Cancer Gene 1/2 (BRCA1/2), retinoblastoma transcriptional corepressor 1 (RB1), adenomatous polyposis coli (APC), phosphatase and tensin homolog deleted on chromosome 10 (PTEN), tumour protein p53 (TP53), MutL Homolog 1 (MLH1), MutS Homolog 2 (MSH2), MutS Homolog 6 (MSH6) and

Postmeiotic Segregation 2 (PMS2) that cause hereditary cancer-predisposing syndromes (Constantinou & Constantinou, 2024; Eun Kim et al., 2021; Mauri et al., 2019; Perna et al., 2021; Xu et al., 2020). The other half of EOCRC patients (50%) is a bigger research problem because they don't have familial or hereditary syndromes. Instead, they have sporadic CRC (Campos, 2017; Maloberti et al., 2022; Mauri et al., 2019). EOCRC differed from LOCRC by exhibiting differential clinical, pathologic, and molecular features (Akimoto et al., 2021; Perea et al., 2021).

The EOCRC was primarily observed on the left side of the colon, typically diagnosed at an advanced stage with more aggressive histological features, including mucinous or signet ring histology with poor differentiation (Jiang et al., 2020; Dharwadkar et al., 2021; Orsini et al., 2015; Yeo et al., 2017). Another notable feature of EOCRC was the prevalence of dMMR, which is higher than in LOCRC and is associated with hereditary CRC or MLH1 hypermethylation, as well as other epigenetic changes (Saraiva et al., 2023; Yambert et al., 2022). The dMMR is caused by the loss of one or more MMR proteins, including MLH1, MSH2, MSH6, and PMS2, respectively (Guyot et al., 2021; Maloberti et al., 2022; Saizul et al., 2021). EOCRCs with dMMR having a loss of MMR protein(s) are highly immunogenic through their ability to elicit a rapid immune response with massive production of Tregs, other forms of antibodies, and immune cells that will infiltrate the tumour mucosa compared to pMMR having all the MMR proteins (Evrard et al., 2019; Yoon et al., 2012). Previous studies have documented variations in FOXP3+ Treg infiltration between dMMR and pMMR (Ling et al., 2018; Xu et al., 2020; Yoon et al., 2012).

Currently, there is a lack of studies highlighting the role of FOXP3 in relation to v-RAF murine sarcoma viral oncogene homolog B1 (BRAF) V600E and MMR protein status among EOCRC patients at Hospital Universiti Sains Malaysia (HUSM).

There is also an absence of literature highlighting the epidemiology and clinicopathological characteristics of EOCRC patients in HUSM. Therefore, this study aimed to address these gaps.

1.3 The Rationale of the Study

The study aims to investigate the relationship between FOXP3 expression, the BRAFV600E mutation, and MMR protein status in patients with EOCRC. This was achieved by exploring the interplay between FOXP3 expression, BRAFV600E, and MMR proteins. The study offers insights into the immune microenvironment and the molecular mechanisms that drive the initiation and progression of the EOCRC. The study will help further understand the role of FOXP3 in EOCRC by examining the protein expression of FOXP3 using immunohistochemistry. The study has determined the low and high expression of the FOXP3 protein, as well as its relationship to the demographics and histopathological characteristics of EOCRC patients.

The study investigated the impact of MMR protein status on FOXP3 expression, examining whether the absence or presence of MMR proteins affects FOXP3 expression, and whether these impacts are associated with distinct demographic or clinicopathological features in EOCRC. The study also investigated the relationship between FOXP3 expression and the BRAFV600E mutation. This will help determine whether the presence or absence of the mutant BRAFV600E can influence FOXP3 expression and whether this relationship contributes to the development or progression of EOCRC.

Additionally, the combination of assessing FOXP3 protein expression using immunohistochemistry and FOXP3 gene expression using real-time quantitative polymerase chain reaction (RT-qPCR) methodologies enabled the examination of

FOXP3 expression at both the protein and messenger ribonucleic acid (mRNA) levels, providing a more comprehensive understanding of FOXP3 expression in EOCRC.

Therefore, the study findings may contribute to an improved understanding of the molecular mechanisms underlying EOCRC, particularly concerning the tumour immune microenvironment. The study findings may suggest potential biomarkers for EOCRC through the use of FOXP3, BRAFV600E, and MMR proteins, which can aid in early diagnosis, prognosis, treatment decisions, and other therapeutic modalities that focus on modulating the immune microenvironment or targeting specific molecular pathways.

1.4 Research Questions

1. What is the prevalence of EOCRC?
2. What are the demographics and clinicopathological characteristics of EOCRC?
3. What are the FOXP3, BRAFV600E, and MMR protein expression patterns among the EOCRC?
4. Is there a correlation between FOXP3, BRAFV600E, and MMR proteins expressed in the EOCRC?
5. What are the variants of the *FOXP3* gene expressed by the EOCRC?

1.5 Research Hypotheses

1. There is an association between prevalence and FOXP3 expression in EOCRC.
2. There is a relationship between demographic and clinicopathological characteristics with FOXP3 expression in EOCRC

3. There is an association between FOXP3 expression and EOCRC histological subtypes.
4. There is a correlation between FOXP3 expression with BRAFV600E and MMR status in EOCRC.
5. There is a relationship between the gene expression of FOXP3 with BRAV600E and MMR status in EOCRC

1.6 Aim and Objectives of the Study

1.6.1 General Objective/Aim of the Study

To study the role of FOXP3 in relation to BRAFV600E and MMR protein status of the EOCRC patients.

.

1.6.2 Specific Objectives of the Study

1. To determine the demographic and clinicopathological characteristics of patients with EOCRC.
2. To evaluate the expression of BRAFV600E and MMR protein status in the EOCRC.
3. To determine the FOXP3 protein expression using the immunohistochemistry technique in EOCRC.
4. To determine the *FOXP3* gene expression variants using RT-qPCR in EOCRC.
5. To correlate the FOXP3 protein and gene expression with BRAFV600E and MMR protein status in the EOCRC.
6. To associate the FOXP3 protein and gene expression with the demographic and clinicopathological characteristics of EOCRC.

CHAPTER 2

LITERATURE REVIEW

2.1 Epidemiology of CRC

The CRC is a malignant neoplasm that develops from the glandular epithelial cells of the digestive system in the colon or rectum (Fleming et al., 2012; Keum & Giovannucci, 2019; Perna et al., 2021). Colon and rectal cancer can occur individually but are mostly referred to as CRC because of their similarities (Ballester et al., 2016). The global, regional, and Malaysian epidemiology of CRC is presented in this section:

2.1.1 Global Incidence of CRC and Economic Burdens

CRC has accounted for 10% of global cancer incidence, with 1,931,590 new cases (Figure 2.1), and 9.4% of global cancer mortality, with 935,173 deaths in 2020 (Mauri et al., 2019). The number of new CRC cases internationally is predicted to reach 3.2 million by 2040 (Xi & Xu, 2021). Men are more likely than women to get CRC as a result of biological and gender-related variables such as smoking, drinking alcohol, consuming a lot of red meat, and having a greater visceral fat deposit (White et al., 2018). Studies have shown that developed countries have a 3-4 times higher incidence of the disease compared to developing countries (Akkoca et al., 2014; Alyabsi et al., 2021; Rawla et al., 2019). Figure 2.1 presented the global incidence of new cases of CRC in 2020 for both sexes and all age groups below:

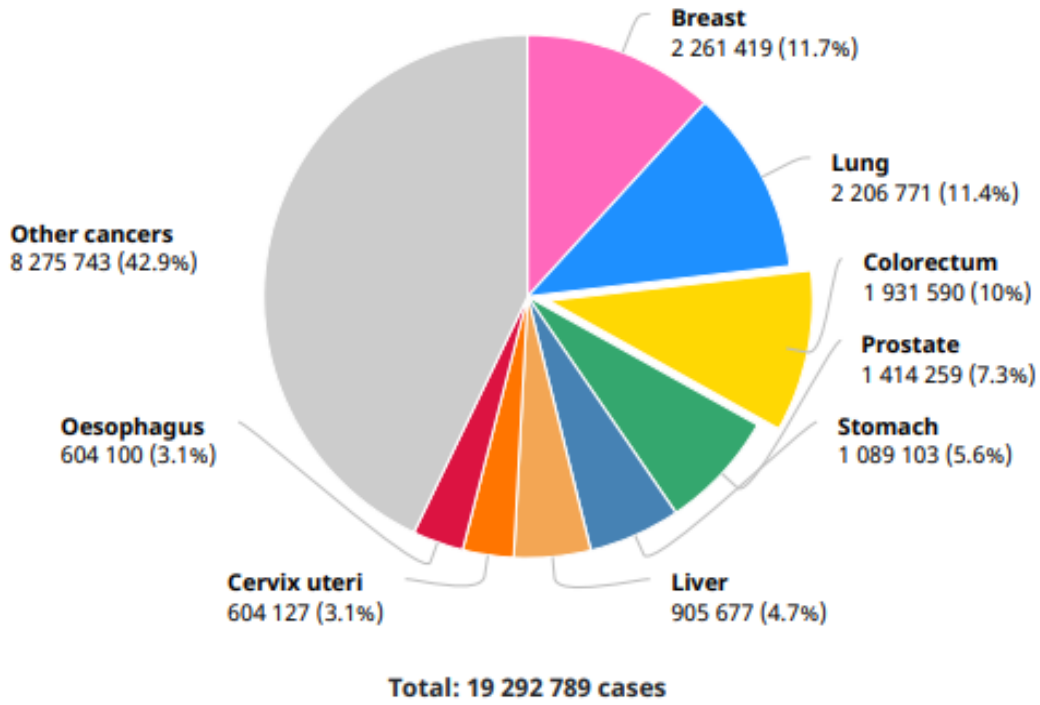

Figure 2.1: Global Incidence of CRC new cases in 2020 (yellow colour) for both sexes and all age groups (Ferlay et al., 2020).

CRC causes numerous deaths worldwide and imposes economic burdens on countries through budgetary allocations for diagnosis and treatment (Ibrahim et al., 2020; Nawawi et al., 2021; Tze et al., 2017). According to Azzani et al. (2019), the annual cost of managing CRC worldwide is 39 billion United States dollars (USD).

2.1.2 Incidence of the CRC in Asia

Across all genders and ages combined, Asia had the highest incidence of CRC cases (51.8%), as shown in Figure 2.2, and the highest mortality rates (52.4%) per 100,000 people worldwide in 2018 (Onyoh et al., 2019). According to the Global Cancer Observatory (GLOBOCAN), Asia reported 9,503,710 new cancer cases in 2020, of which 1,009,400 were CRC cases (Huang et al., 2022). The 5-year prevalence rates of CRC in China, Japan, Korea, Malaysia, Singapore, and Turkey were more than 46.5% cases per 100,000 people compared to other Asian nations (Wong et al., 2019). Previous studies have attributed the high incidence of CRC in Asia to rapid changes in socioeconomic and lifestyle habits, lack of physical activity (sedentary lifestyle), and

smoking (Pardamean et al., 2023; Pourhoseingholi, 2014). Figure 2.2 presented the CRC incidence in Asia for both ages and sexes below:

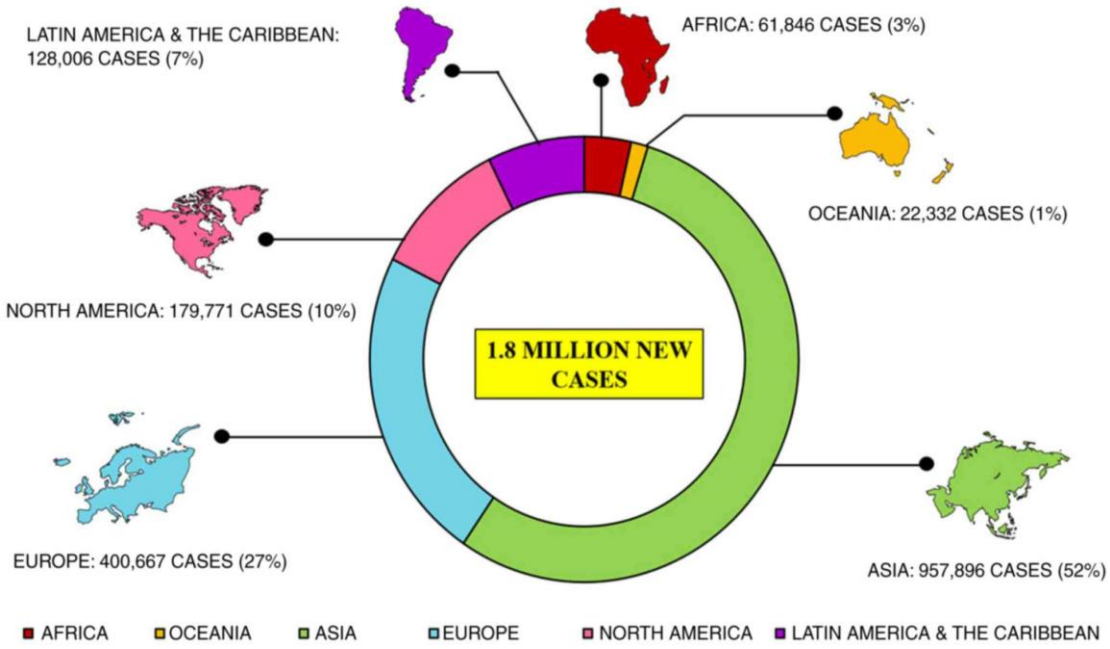

Figure 2.2: Asian incidence rate of CRC (green colour) for both sexes and all ages (Ahmad et al., 2021).

2.1.3 Incidence of the CRC and Economic Burdens in Malaysia

Malaysia is among the Asian countries with a high prevalence of CRC and is currently ranked the second most common cancer in the country (Chandran et al., 2020; Radzi et al., 2016; Veettil et al., 2017; Wong et al., 2021). According to the Malaysian National Cancer Registry Report of 2012-2016, CRC is the most prevalent cancer in men (16.9%) and women's second most common cancer (10.7%) of all cancers diagnosed (Nawawi et al., 2021). In 2020, Malaysia had 6,597 (13.6%) new cases and a mortality rate of 3,420 (11.6%) for both sexes and all ages (Ferlay et al., 2020). Most CRCs in Malaysia (around 70%) were diagnosed late, with a tendency toward poor prognosis, as indicated by Figure 2.3 (Che Jalil et al., 2022; Azzani et al., 2019; Wong et al., 2021).

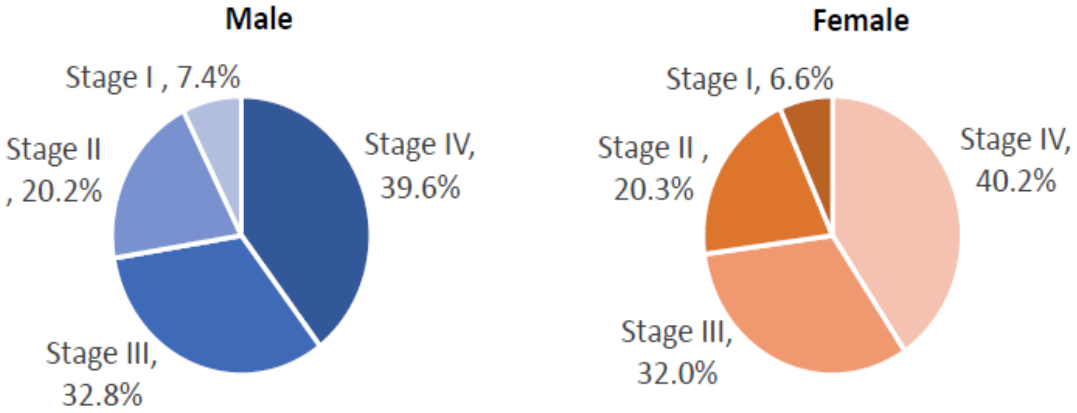

Figure 2.3: Stages of CRC diagnosis in Malaysia among males and females (Ministry of Health Malaysia, 2021).

The late diagnosis of CRC in Malaysia inevitably increased the financial burden on patients and the government due to the more expensive cost of treatments and lower quality of life (Veettil et al., 2017). The average cost of CRC treatment annually per person in Malaysia ranges from Malaysian Ringgit (MYR) 13,622 for stage 1 to MYR 27,377 for stage 4 (Nawawi et al., 2021). The total cost of managing new cases of CRC in Malaysia is estimated at MYR 108 million annually (Nawawi et al., 2021; Tze et al., 2017).

2.2 Epidemiology of EOCRC

2.2.1 Global Incidence of EOCRC

Contrary to the declining rates of CRC among adult individuals, the incidence of CRC in young people is increasing worldwide (Azar et al., 2021; Hofseth et al., 2020; Lamprell et al., 2023; Syed et al., 2019). The worldwide incidence of EOCRC has more than doubled so far, from 95,737 in 1990 to 226,782 in 2019 (H. Pan et al., 2022). The EOCRC incidence is expected to increase by 90% in colon carcinoma and 124% in

rectal carcinoma worldwide by 2030 (AlZaabi et al., 2022; Gu et al., 2022; Hofseth et al., 2020; Vuik et al., 2019).

Previous studies have revealed that EOCRC differs from LOCRC in terms of epidemiology, anatomical location, pathology, and molecular perspectives (Arriba et al., 2019; Hofseth et al., 2020; Wei et al., 2020). The EOCRCs are mostly diagnosed at an advanced stage and are primarily located in the distal colon and rectum (De Campos et al., 2017; Dharwadkar et al., 2021; Mannucci et al., 2019). Other reported EOCRC features include more aggressive histologic characteristics such as poor differentiation and mucinous or signet ring histology (Daniel et al., 2018; Li et al., 2020; Orsini et al., 2015).

2.2.2 Incidence of EOCRC in Asia

Studies have shown that the incidence of EOCRC varies worldwide, with nearly 20% of cases reported in Asia (Hoseini et al., 2022). There has been a sharp rise in the prevalence of EOCRC in some Asian countries, including China, Japan, India, South Korea, Indonesia, and Singapore (Dharwadkar et al., 2021; Wong & Sung, 2020). According to institutional research, the percentage of EOCRC in Asia varies by country, from 6.7% in Taiwan to 39% in India (Wong et al., 2021). Like other regions of the globe, EOCRC is more likely to appear in the distal colon or rectum in Asia (Ballester et al., 2016; Chandran et al., 2020; Chong et al., 2015; Valan et al., 2021).

2.2.3 Incidence of EOCRC in Malaysia

Malaysia also has a high rate of EOCRC, similar to other countries; the total age-standardized incidence rate of EOCRC was 25.23 per 100,000 persons (Ibrahim et al., 2020). A study conducted in Northern Malaysia reported an EOCRC prevalence of 14.5%, with 893 cases of CRC cancer diagnosed in individuals under the age of 50 out

of a total of 6,172 CRC cases observed from 2007 to 2017 (Ibrahim et al., 2020). Another study also reported an EOCRC prevalence of 10.7%, conducted at the University of Malaya Medical Centre, with 206 CRC cases in individuals under 50 out of the total number of 1921 CRC cases seen from 2002 to 2016 (Wong et al., 2021).

Most of these EOCRC patients in Malaysia were diagnosed with cancer at an advanced stage, generally on the left side, and with adenocarcinoma ranging from moderately to poorly differentiated type (Prabhakaran et al., 2022; Siegel et al., 2023; Zaborowski et al., 2021). Various countries across the world are at different stages of implementing the population-based national screening programme for CRC, countries with implementation of the programme include Italy, Netherland, Ireland, Croatia, , Czech Republic, Switzerland, Slovenia, Finland, Belgium, Denmark, Sweden, Estonia, Canada, Hong Kong, China, America, Australia, England, France, South Korea, Thailand, Taiwan and Chile among others (Cenin et al., 2018; Lin et al., 2016; Navarro et al., 2017; Veettil et al., 2017). In population-based screening programmes, the most widely used extensive screening methodologies are faecal occult blood, faecal immunochemical test (FIT), and visual methods such as colonoscopy or sigmoidoscopy (Bilal et al., 2020; Chandran et al., 2020, 2022; Onyoh et al., 2019; Yusoff et al., 2021). Malaysia is yet to implement a population-based national screening programme for CRC fully ; this could be one of the factors responsible for the country's advanced diagnosis of both EOCRC and LOCRC (Wong et al., 2021).

2.3 Anatomy of the Large Intestine

The caecum, colon, rectum, and anus make up the large intestine, which measures around 1.5 metres in length (Farraj et al., 2019; Khan and Ismat, 2019; Nigam et al., 2019). The large intestine has a very similar anatomy to the small intestine, except

it has no villi in its mucosa (Nigam et al., 2019). The colon is divided into two (2) major parts:

I. The proximal colon, which is classified as proximal to the splenic flexure, is comprised of the caecum, ascending colon, hepatic flexure, and transverse colon (Ghanipour et al., 2017; Murphy et al., 2019; Quintanilla-Guzman et al., 2018).

II. The distal colon that is distal to the splenic flexure is comprised of the descending colon and sigmoid colon (Lee et al., 2017a; Lin et al., 2016).

Figure 2.4 shows the anatomy and classification of the large intestine into proximal and distal parts as below:

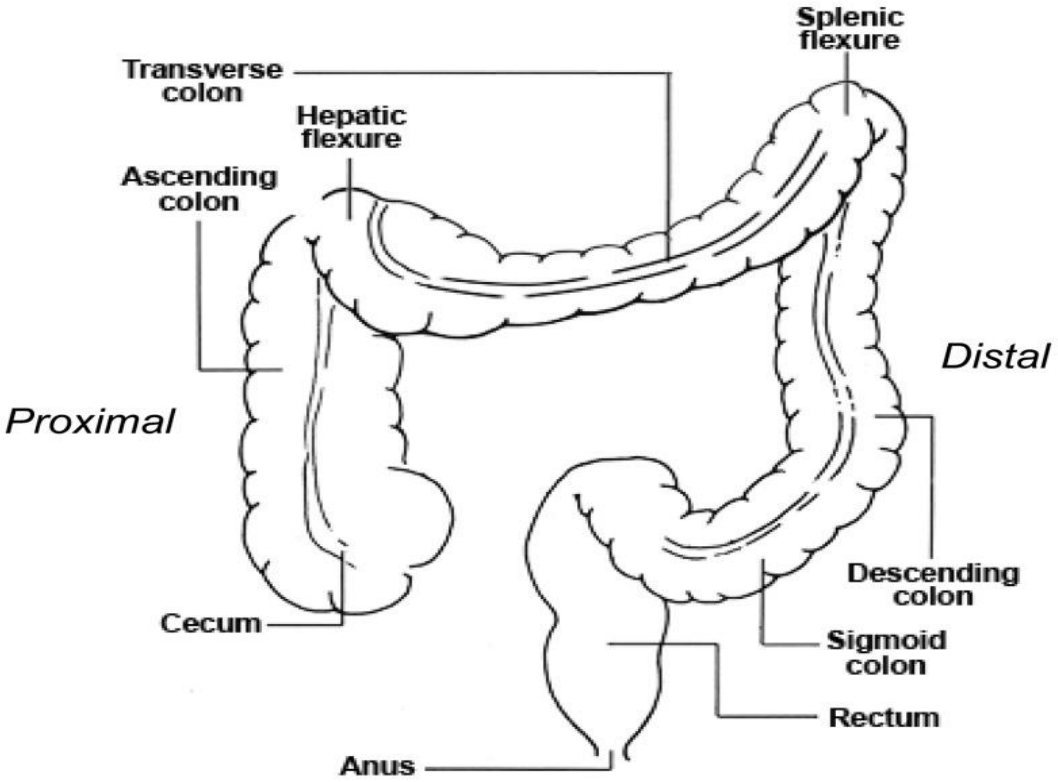

Figure 2.4: The anatomy and classification of large intestine parts (Lin et al., 2016).

2.4 Risk Factors of the EOCRC

Risk factors for EOCRC are those factors that increase the chances of acquiring the disease (Ahmad et al., 2021). The risk factors are divided into non-modifiable and modifiable factors, as below:

2.4.1 Non-modifiable Risk Factors of the EOCRC

Non-modifiable risk factors are those factors that the patient cannot change; these factors include:

I. Age: The most significant risk factor for the onset of CRC is age; age increase, especially between 40 and 50 years, is an associated risk factor contributing to EOCRC development (Garrett et al., 2022; Saraiva et al., 2023; Wu et al., 2022).

II. Presence of family history: Young adults with a family history of CRC, especially first-degree relatives, have been associated with an increased risk of EOCRC development (Danial et al., 2022; Juhari et al., 2015; Syed et al., 2019).

III. Hereditary syndrome and other conditions: Adenomatous polyps and inflammatory bowel disease (Crohn's disease and ulcerative colitis) have also been linked to a higher risk of EOCRC development (Ahnen et al., 2014; Hubbard & Grothey, 2013; Weinberg & Marshall, 2019). Genetic factors play a significant role in the development of EOCRC, examples include Lynch Syndrome (LS), Familial Adenomatous Polyposis (FAP), MUTYH-Associated Polyposis (MAP), other rare genetic syndromes such as Cowden syndrome, Peutz-Jeghers syndrome (PJS), Juvenile polyposis syndrome (JPS) among others (Chang et al., 2012; Hubbard & Grothey, 2013; Khairunnisa et al., 2020; Peltomäki et al., 2023). LS is caused by germline mutations in DNA MMR genes of MLH1, MSH2, MSH6 and PMS2 (Al-Sohaily et al., 2012; Kawakami et al., 2015; Moreira et al., 2012). FAP is as a result of germline mutation in the

APC gene, this hereditary syndrome is characterised by development of numerous adenomatous polyps in the colon and rectum (Armelao & De Pretis, 2014; Chang et al., 2012; Hubbard & Grothey, 2013; Willauer et al., 2019). Other genetic variations or mutations that increase the risk of developing EOCRC are APC mutation (APC variant CRC), particularly in patients with a family history of CRC and KRAS mutation (KRAS variant CRC) especially in individuals with history of smoking (Fan et al., 2021; Gausman et al., 2020; Willauer et al., 2019; Xu et al., 2020; Ye et al., 2015).

IV. Gender: The male gender is associated with an increased risk of EOCRC development and mortality rate than the female counterpart (White et al., 2018; Wu et al., 2022).

V. Ethnicity: In Malaysia, people of Chinese ancestry had the highest incidence of CRC (27.35%), followed by Malay ancestry (18.95%) and Indian ancestry (17.55%), respectively (Wong et al., 2021).

2.4.2 Modifiable Risk Factors of the EOCRC

Modifiable risk factors for EOCRC are behaviours and exposures that can increase the risk of developing CRC; however, these factors are changeable. These risk factors are grouped into 3:

I. Westernised diet or dietary factors: Consumption of a diet that is high in red meat or processed meat, high in sugary food or drinks, low in fibre, and other vital nutritional contents has been associated with an increased risk of EOCRC (Danial et al., 2022; Keum & Giovannucci, 2019; Shen et al., 2021).

II. Behavioural factors: Studies have shown that alcohol intake, smoking, physical inactivity, or a sedentary lifestyle prone younger adults to a higher risk of EOCRC development (Dekker et al., 2019; Onyoh et al., 2019; Venugopal & Carethers, 2022). There are increased cases of alcohol-related liver disease among youths due to heavy alcohol consumption, with a 1.71 relative risk of CRC development (Venugopal & Carethers, 2022). Cigarette smoking, especially in a dose-dependent manner, increases the risk of CRC development, with a mean relative risk of 1.18 for current and past smokers (Murphy et al., 2019; Venugopal & Carethers, 2022).

III. Metabolic factors: Obesity, increased body mass index (BMI), and diabetes are related to an increased risk of developing EOCRC (Lu et al., 2021; Venugopal & Carethers, 2022; Weinberg & Marshall, 2019). These have also been shown to be significant in the risk of developing EOCRC among young adults (Elangovan et al., 2021; Vekic et al., 2021; Zaborowski et al., 2021). According to Garrett et al. (2022), younger adults who are overweight (≥25) or obese (≥30) have approximately 32% and 88% higher risk of developing CRC than those with average weight, respectively. In terms of BMI, a young patient with a BMI ≥30 was associated with a nearly two-fold rise in the relative risk of EOCRC (Venugopal & Carethers, 2022). Type II diabetes, also closely related to obesity, is a risk factor for CRC development, with a relative risk of 1.3 (Venugopal & Carethers, 2022).

2.5 Symptoms of EOCRC

EOCRC is mainly asymptomatic until it reaches an advanced stage; hence, many EOCRC diagnosed with a symptomatic disease are already at the advanced stage

(Dekker et al., 2019; Hubbard & Grothey, 2013). The sigmoid colon and rectum (rectosigmoid colon) are the most common sites affected by the EOCRC, so symptoms of abdominal pain are more likely to show up on the left side of the colon (Hubbard & Grothey, 2013; Zbuk et al., 2009). The three (3) most common presenting symptoms of EOCRC patients are abdominal pain, per-rectal bleeding, and altered bowel habits (Kaur Sindhu et al., 2019; Lamprell et al., 2023; Wu et al., 2022). Other presenting symptoms in EOCRC patients include fatigue, unexplained weight loss, anaemia, nausea, and vomiting (Schliemann et al., 2020).

2.6 Types of Histology in EOCRC

The following are the different types of histology in the EOCRC:

2.6.1 Grading of Adenocarcinoma

The majority of the EOCRC histology belongs to adenocarcinoma, which is characterised by glandular formation and divided into:

1. Well-differentiated adenocarcinoma has more than 95% of adenocarcinoma made up of glands by the epithelial cells of the CRC mucosa (Ahmad et al., 2021; Fleming et al., 2012; Perna et al., 2021).
2. Moderately differentiated adenocarcinoma has 50-95% glandular formation by the epithelial cells of the CRC mucosa (Perna et al., 2021; Saraiva et al., 2023).
3. Poorly differentiated adenocarcinoma has <50% glandular formation by the epithelial cells of the CRC mucosa (Feng et al., 2019; Fleming et al., 2012).

2.6.2 Mucinous Adenocarcinoma

Mucinous adenocarcinoma is defined by the World Health Organisation (WHO) classification as having >50% extracellular mucin in the lesion (Li et al., 2020).

However, when the mucinous component is <50%, it is called adenocarcinoma with mucinous features or mucinous differentiation (Fleming et al., 2012). Compared to LOCRC patients, young patients' tumours more frequently exhibit adverse histologic characteristics, such as mucinous adenocarcinoma or mucinous differentiation (Ballester et al., 2016; Saraiva et al., 2023).

2.6.3 Signet Ring Cell Adenocarcinoma/Carcinoma

Signet ring cell adenocarcinoma or carcinoma has >50% of the tumour lesion exhibiting signet ring cell features, with a prominent intracytoplasmic mucin vacuole that pushes the nucleus to the periphery (AlZaabi et al., 2022; Fleming et al., 2012; Young et al., 2015). Signet ring cell carcinoma is high-grade and poorly differentiated by description, and it has a worse prognosis than typical adenocarcinoma (Fleming et al., 2012). Another distinguishing feature of EOCRC is the presence of more frequent signet ring features in the histology than LOCRC (Farraj et al., 2019; Mauri et al., 2019; Willauer et al., 2019).

2.6.4 Medullary Carcinoma and Other Less Commonly Occurring Histologies

Medullary carcinoma is one of the less commonly occurring histologies in CRC (Fleming et al., 2012; Luévano-González et al., 2011). Sheets of epithelioid neoplastic cells with large vesicular nuclei, prominent nucleoli, and a lot of cytoplasm can be seen in medullary carcinoma (Fleming et al., 2012). Other less common occurring (rare) types of CRC include neuroendocrine, squamous cell, spindle cell, adenosquamous, and undifferentiated carcinomas (Fleming et al., 2012).

2.7 Pathogenesis and Molecular Basis of EOCRC

EOCRC is a heterogeneous disease with a solid connection to the hereditary or familial component, accounting for 10-20% of the occurrence; however, most of the

EOCRC cases (80%) are sporadic (Alvarez et al., 2021; Armelao & De Pretis, 2014; Arriba et al., 2019; Perea et al., 2021). The heterogeneous nature of EOCRC is due to diverse genetic and epigenetic molecular alterations (Perea et al., 2021). The variations can be distinguished through their histopathologic or molecular characteristics (Silla et al., 2014). Figure 2.5 indicates the incidence of hereditary, familial, and sporadic CRC.

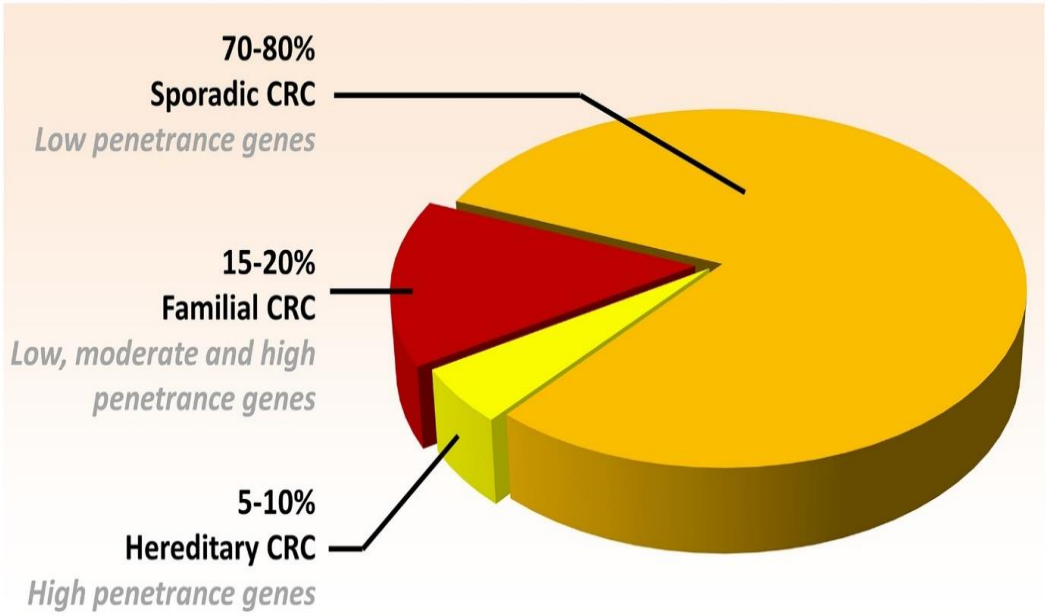

Figure 2.5: CRC incidence rates for hereditary, familial, and sporadic types (Tan, 2018).

2.7.1 Types of CRC based on Heredity

The following are the three (3) types of CRC:

2.7.1(a) Hereditary CRC

Hereditary CRC is a combination of inherited syndromes due to germline mutations of high penetrance genes responsible for the carcinogenesis of this CRC; the CRC accounted for 5-10% of CRC cases (Tan, 2018). The syndromes in hereditary CRC are known and are detectable using germline testing (Evrard et al., 2019; Ramdzan et al., 2021; Vos et al., 2020).

The most common hereditary CRC syndromes are familial adenomatous polyposis (FAP) and Lynch syndrome (LS), all of which have an autosomal dominant disorder (Al-Sohaily et al., 2012; Chao et al., 2019). The FAP is an autosomal dominant gene condition due to a germline mutation of the adenomatous polyposis coli (APC) gene (Al-Sohaily et al., 2012; Willauer et al., 2019; Xu et al., 2020). The LS, which was formerly called hereditary nonpolyposis CRC (HNPCC), is the result of an autosomal dominant disorder as a result of a germline mutation in the DNA MMR genes (Khairunnisa et al., 2020; Lee et al., 2017b).

Apart from LS and FAP, germline mutations in POLE and POLD1 are also associated with the development of EOCRC (Eun Kim et al., 2021). POLE is a gene that encodes for the DNA Polymerase epsilon catalytic subunit, and POLD1 is a gene that encodes the catalytic and proofreading subunit of DNA Polymerase delta (Eun Kim et al., 2021; Puccini et al., 2018).

Germline mutations in the SMAD4 gene can also increase the EOCRC development as this gene acts as a tumour suppressor and a transcriptional factor (Malki et al., 2021; Xu et al., 2020). SMAD4 is a key component of the TGF-β signalling pathway that regulates cellular processes such as growth, differentiation and apoptosis (Choi et al., 2020). Hence, germline mutations in SMAD4 will affect regulation growth, differentiation and apoptosis, thereby increasing the tendency of cellular proliferations out of control (Malki et al., 2021).

STK11 (also known as LKB1) mutations are germline mutations associated with an increased risk of EOCRC development (Koveitypour et al., 2019). Additionally, other forms of germline mutations that increase the risk of developing EOCRC are MAP mutation, NTHL1 mutation, AXIN2 mutation, PTEN mutation and BMPR1A mutation,

respectively (Al-Sohaily et al., 2012; Hagland et al., 2013; X. Li et al., 2020; Shia et al., 2012).

2.7.1(b) Familial CRC

Among the EOCRC patients, up to 20% have familial CRC, which probably may be due to a combination of the interplay between inherited genetic factors such as low or moderate penetrance genetic alterations and shared exposure to agents such as mutagen or carcinogens found in the environment (Ahnen et al., 2014; Alvarez et al., 2021; Armelao & De Pretis, 2014). Patients with hereditary disorders that have not been recognised and those with seemingly sporadic forms of the disease that cluster in families are included in this heterogeneous group of CRC known as familial CRC (Armelao & De Pretis, 2014).

Previous studies have shown that first-degree relatives (FDRs) of people with CRC have a two to four times greater chance of also having this type of cancer than the general population (Arriba et al., 2019; Che Jalil et al., 2022; Tantoğlu et al., 2018). A first-degree relative is a family member (parents, children, and siblings) who shares at least 50% of their DNA with a specific family member (Armelao & De Pretis, 2014; Ghanipour et al., 2017; South et al., 2009). The familial risk is inversely correlated with the age of the youngest FDRs and directly related to the number of afflicted FDRs (Armelao & De Pretis, 2014).

2.7.1(c) Sporadic CRC

CRCs that develop from the colorectum and have no known genetic origins, a strong family history, or other risk factors, such as inflammatory bowel disease, are

classified as sporadic CRCs (Carethers & Jung, 2015). Many EOCRC cases are sporadic (70%-80%), resulting from the accumulation of multiple acquired somatic genomic and epigenetic alterations affecting low-penetrance genes (Huang & Yang, 2022; Tan, 2018). Examples of somatic mutations that affect low-penetrance genes, thereby increasing the risk of developing EOCRC, are TP53, KRAS, PIK3CA, and BRAF somatic mutations (Deschoolmeester et al., 2010; Fan et al., 2021; Fleming et al., 2012; Perna et al., 2021; Venugopal & Carethers, 2022; Yan, 2014).

Epigenetic alterations that can occur in EOCRC include DNA methylation (Huang & Yang, 2022). Methylation of the DNA can cause the silencing of tumour suppressor genes, such as MLH1, which contributes to the development of EOCRC (Maloberti et al., 2022; McCarthy et al., 2019; Yan, 2014). Another form of epigenetic alteration in EOCRC is histone alteration; changes in histones through methylation or acetylation can also contribute to EOCRC (Pagè et al., 2018; Grover et al., 2021; Puccini et al., 2018). MicroRNA dysregulation is another form of epigenetic alteration that can contribute to EOCRC through targeting tumour suppressor genes or oncogenes (Ahmad et al., 2021; Lopes et al., 2006).

Most sporadic colorectal cancers (CRCs) are microsatellite stable (MSS) because they exhibit chromosomal instability (CIN) and lack features associated with microsatellite instability (Akkoca et al., 2014; Daniel et al., 2018; Gelsomino et al., 2016). Sporadic CRC patients lack known genetic alterations or a strong family history of CRC; hence, they are frequently diagnosed at advanced stages of the disease because they are not included in screening programmes (Mauri et al., 2019).

2.7.2 Molecular Pathways of CRC Carcinogenesis

The "adenoma-carcinoma sequence" describes the progression of CRC from normal colonic epithelium to an intermediate adenomatous state and finally to an adenocarcinoma as indicated by Figure 2.6 (Bilal et al., 2020; Pino & Chung, 2010).

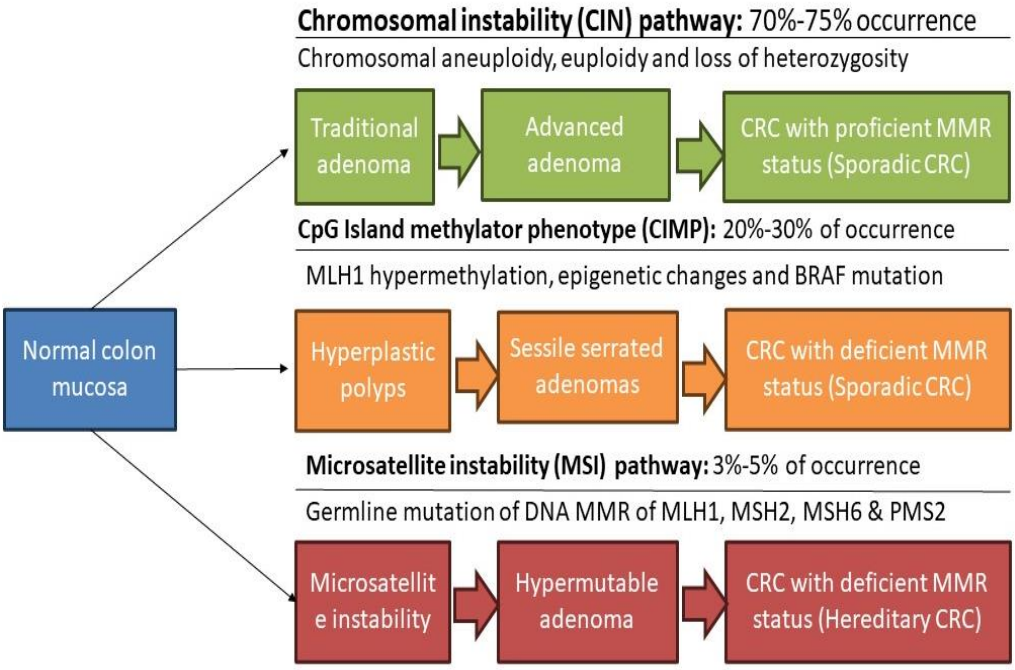

Figure 2.6: Molecular pathways of CRC pathogenesis (Modified from Ballester et al., 2016)

The loss of genomic stability causes many mutations that make the growth and development of CRC easier (Ewing et al., 2014). Genomic instability, therefore, generates a favourable environment where a prospective cancer cell can accumulate sufficient mutations to turn into a cancer cell (Li et al., 2021). The molecular pathways of CRC based on genomic instability are classified into three (3) major groups, as indicated by Figures 2.6 and 2.7. Namely, chromosomal instability (CIN), microsatellite instability (MSI), and CpG island methylator phenotype (CIMP), respectively (Nguyen & Duong, 2018; Ishaque et al., 2021; Fred & Yan, 2014).

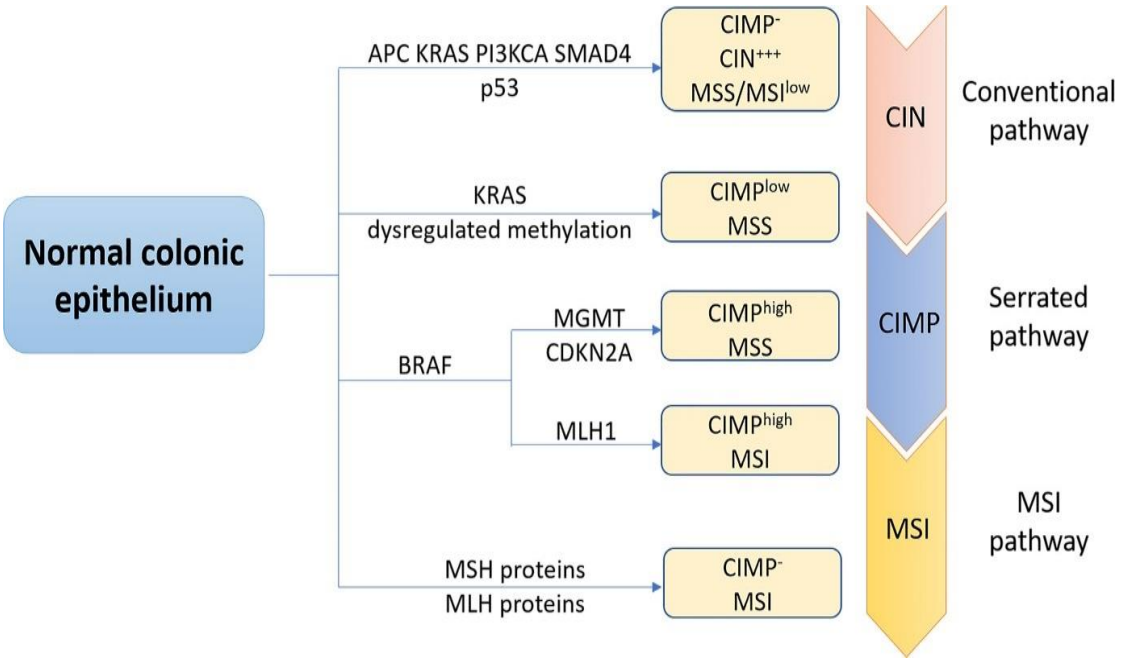

Figure 2.7: Molecular pathways of CRC involving the conventional pathway (CIN), serrated pathway (CIMP), and MSI pathway with associated affected genes (Huang & Yang, 2022).

2.7.2(a) Chromosomal Instability (CIN) Pathway

Among the three molecular pathways of colorectal cancer (CRC), the CIN was the first described pathway and is the most common one, affecting 70%-85% of sporadic CRCs (Carethers & Jung, 2015; Malki et al., 2021). The pathway of CIN is characterised by alterations in chromosomal number, known as aneuploidy, or in chromosomal structure in the forms of insertions, deletions, focal gene amplification, and a high frequency of loss of heterozygosity (Hagland et al., 2013; Li et al., 2021; Pino & Chung, 2010).

The CIN has been linked to the loss of function of tumour-suppressor genes, specifically the adenomatous polyposis coli (APC) gene, whose regular role is to resist carcinogenesis (Ewing et al., 2014). Other characteristics of CIN include activation of the Wnt signalling pathway and deletion of TP53, resulting in the loss of function of the p53 suppressor gene (Li et al., 2021). The progression of CRC from normal epithelium to carcinoma is indicated by Figure 2.8.