THE EFFECT OF ETHANOLIC BEE POLLEN EXTRACT FROM *GENIOTRIGONA THORACICA* ON DNA DAMAGE AND DNA METHYLATION IN HT-29 COLORECTAL CELL LINE

NURDIANAH BINTI HARIF FADZILAH

UNIVERSITI SAINS MALAYSIA

2022

THE EFFECT OF ETHANOLIC BEE POLLEN EXTRACT FROM *GENIOTRIGONA THORACICA* ON DNA DAMAGE AND DNA METHYLATION IN HT-29 COLORECTAL CELL LINE

by

NURDIANAH BINTI HARIF FADZILAH

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

September 2022

ACKNOWLEDGEMENT

In the name of Allah, The Most Gracious and The Most Merciful. Peace and blessings upon the beloved Prophet Muhammad S.A.W., his family and his companions. All Praise to ALLAH S.W.T the Almighty for giving me the blessing, strength, chance, and endurance to complete this study. La hawla wala quwwata illa billah. Alhamdulillah, I have received great support and encouragement throughout my Ph.D. journey.

First and foremost, I am extremely grateful to my supervisor, Dr. Wan Adnan Wan Omar for his invaluable supervision, support, advice, and patience. Thank you so much for guiding me throughout the long process of doctoral research and always being available and enthusiastic to propose ideas and projects. Jazakallahu Ahsanal Jaza, may Allah bless you and your family with barakah, happiness, and good health.

I also thank Prof. Dr. Narazah Mohd Yusoff for her treasured support and mentorship. In addition, I would like to thank laboratory staff in the Integrative Medicine Laboratory, Oncological and Radiological Laboratory, and Regenerative Medicine Laboratory for their kind help and valuable support. My gratitude extends to all IPPT lecturers and staff for their helping hands and sweet smiles every time I came to ask them for any help.

My appreciation also goes out to my awesome research team – Dalila, Hakimah, Haziq, and other lab mates for a cherished time spent together in the lab and social settings. All of you have made my study a wonderful time. Additionally, I would like to acknowledge all my colleagues in ARC, IPPT, who provided happy distractions to rest my mind outside of the research.

I would like to express special gratitude to my dearest family, my beloved parents – Dr. Harif Fadzilah Che Hashim & Pn. Nuriah Hj. Hassan; my loving husband – Ahmad Kamil Azizan; my sweet siblings – Hazirah, Hanif & Nofa, Dr. Hidayah & Dr. Fauzi, Dr. Hafiz & Dr. Nabilah. Thank you for your wise counsel, love, comfort, and always be there for me. My beloved family, you are my Alhamdulillah. Without your tremendous understanding and encouragement in the past few years, it would be impossible to complete my study.

And finally, to my own self, always remember that seeking knowledge and wisdom is a virtue and a lifelong journey to become a better person. “O Allah, please benefit me with what You have taught me, and teach me that which benefit me, and grant me knowledge which will benefit me.” Ameen ya Rabb.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

BPE	Bee pollen extract
CA	Caffeic acid
QUER	Quercetin
H_2O_2	Hydrogen peroxide
ROS	Reactive oxygen species
CRC	Colorectal cancer
DDR	DNA damage response
DNMT	DNA methyltransferase
SAM	S-adenosylmethionine
5mC	5-methyl cytosine
TE	Transposable element
ALU	*Arthrobacter luteus* restriction endonuclease
LINE-1	Long interspersed nuclear element-1
PCR	Polymerase chain reaction
TPC	Total phenolic content
TFC	Total flavonoid content
DPPH	1,1-diphenyl-2-picrylhydrazyl radical
HPLC	High-performance liquid chromatography
GC-MS	Gas chromatography-mass spectrometry
MCF-7	Human breast adenocarcinoma cell line
MCF-10A	Human mammary epithelial cell line
HT-29	Human colorectal adenocarcinoma cell line
TI	Therapeutic index
EC_{50}	Minimum effective concentration

IC_{50}	Minimum inhibitory concentration

LIST OF APPENDICES

KESAN DEBUNGA LEBAH *GENIOTRIGONA THORACICA* YANG DIEKSTRAK MENGGUNAKAN ETANOL TERHADAP KEROSAKAN DAN METILASI DNA DALAM SEL USUS HT-29

ABSTRAK

Spesies Oksigen Reaktif (ROS) boleh mengganggu metabolisme sel serta merosakkan biomolekul sel, yang mana berupaya menyebabkan kerosakan DNA. Kerosakan DNA boleh menyebabkan perubahan epigenetik seperti hipometilasi global dan menggalakkan ketidakstabilan genom. Sebatian semulajadi termasuk debunga lebah kelulut mengandungi antioksida yang mempunyai kesan perlindungan terhadap ROS. Debunga lebah dianggap sebagai makanan lengkap yang mengandungi pelbagai nutrien yang komprehensif dan mempunyai kesan terapeutik yang berupaya menyingkirkan radikal bebas di dalam badan. Kajian ini bertujuan untuk mengkaji kesan perlindungan debunga lebah kelulut terhadap kerosakan DNA dan mengkaji kesannya terhadap metilasi DNA global. Ekstrak etanol debunga lebah (BPE) disediakan dari tiga spesies lebah kelulut Malaysia, iaitu *Tetrigona apicalis*, *Heterotrigona itama*, dan *Geniotrigona thoracica*. Metodologi yang digunakan dalam kajian ini adalah kaedah spektrofotometri untuk aktiviti kimia dan antioksida, teknik kromatografi HPLC dan GC-MS untuk pengenalpastian sebatian fenolik, ujian penyisihan *trypan blue* untuk antiproliferasi sel, ujian komet untuk aktiviti kerosakan DNA, dan PCR-Pyrosequencing untuk analisis metilasi DNA global. Hasil daripada ujian antioksida menunjukkan bahawa BPE dari spesies *G. thoracica* memiliki kapasiti tertinggi untuk meneutralkan radikal DPPH, dengan nilai EC_{50} adalah 0.98 ± 0.18 mg/mL. Dalam analisis HPLC, empat sebatian fenolik dikenal pasti terdapat pada *G. thoracica* yang tidak dikesan dalam spesies lain. Analisis GC-MS menunjukkan

variasi beberapa kumpulan kimia yang berlainan dalam setiap spesies. Dalam ujian antiproliferasi sel, *G. thoracica* menunjukkan indeks terapeutik tertinggi (TI = 3.12); dengan nilai EC_{50} adalah 0.5 mg/mL dalam sel HT-29. Analisis aktiviti kerosakan DNA dalam tempoh rawatan selama 24 jam menggunakan ekstrak *G. thoracica* menunjukkan penurunan yang signifikan terhadap kerosakan DNA yang disebabkan oleh H_2O_2 berbanding dengan sel yang tidak dirawat (**63.82%** $\pm$ 2.46 berbanding **90.86%** $\pm$ 0.68), ($p < 0.01$). Penurunan yang signifikan juga dilihat dengan pengambilan asid kafeik (**49.05%** $\pm$ 4.23) dan quercetin (**43.98%** $\pm$ 3.77), ($p < 0.01$). Pengurangan kerosakan DNA yang lebih signifikan diperhatikan dalam tempoh rawatan selama 72 jam menggunakan ekstrak *G. thoracica* (**20.49%** $\pm$ 0.73), asid kafeik (**5.65%** $\pm$ 0.35), dan quercetin (**7.58%** $\pm$ 0.32), ($p < 0,01$) terhadap sel HT-29 yang terdedah dengan H_2O_2. Dalam kajian metilasi DNA, rawatan dengan ekstrak *G. thoracica* dapat meningkatkan metilasi Alu secara signifikan pada 24 jam (**26.00%** $\pm$ 0.58 berbanding **21.50%** $\pm$ 0.96), ($p = 0.007$) dan pada 48 jam (**31.75%** $\pm$ 1.89 berbanding **20.25%** $\pm$ 0.95), ($p = 0.004$). Purata metilasi LINE-1 di ketiga-tiga kawasan juga jauh lebih tinggi dan signifikan pada 24 jam (**79.92%** $\pm$ 1.26 berbanding **72.83%** $\pm$ 2.21, $p = 0.050$) dan 72 jam (**95.25%** + 2.46 berbanding **74.83%** + 0.44, $p = 0.001$) dibandingkan dengan sel yang tidak dirawat. Disebabkan faktor antioksida yang tinggi, ekstrak *G. thoracica* berupaya memberi kesan perlindungan terhadap kerosakan DNA yang disebabkan oleh H_2O_2 melalui peningkatan metilasi Alu dan LINE-1, seterusnya berupaya membina kestabilan genetik.

THE EFFECT OF ETHANOLIC BEE POLLEN EXTRACT FROM *GENIOTRIGONA THORACICA* ON DNA DAMAGE AND DNA METHYLATION IN HT-29 COLORECTAL CELL LINE

ABSTRACT

Reactive Oxygen Species (ROS) can disturb cellular metabolism and damage cellular biomolecules, which could lead to DNA damage. The damage can alter epigenetic changes such as global hypomethylation and promote genome instability. Natural compounds including stingless bee pollen contain nutrient antioxidants that have a protective effect against ROS. Bee pollen is considered a complete food which contains comprehensive nutrients and therapeutic properties that work by eliminating free radicals in the body. This study aims to investigate the protective effect of stingless bee pollen against DNA damage and to measure the effect of bee pollen on global DNA methylation. Bee pollen ethanolic extracts (BPE) were prepared from three stingless bee species native to Malaysia: *Tetrigona apicalis, Heterotrigona itama*, and *Geniotrigona thoracica*. The methodologies used in this study were spectrophotometric method for chemical and antioxidant activities, HPLC and GC-MS chromatographic techniques for phenolic compounds identification, trypan blue exclusion assay for antiproliferation test, comet assay for DNA damage activities, and PCR-Pyrosequencing analysis for global DNA methylation analysis. In antioxidant assay, the result showed that *G. thoracica* BPE possessed the highest capacity to neutralize DPPH radicals, with the EC_{50} of 0.98 ± 0.18 mg/mL. In HPLC analysis, four phenolic compounds were identified in *G. thoracica* which were not detected in other species. GC-MS analysis showed variations of chemical groups among each species. In antiproliferation assay, *G. thoracica* exhibited the highest therapeutic index

(TI=3.12), with the EC_{50} of 0.5 mg/mL in HT-29 cells. Analysis of DNA damage activity at 24 h of *G. thoracica* treatment showed a significant decrease of H_2O_2-induced DNA damage compared to the untreated cells (**63.82%** $\pm$ 2.46 vs. **90.86%** $\pm$ 0.68), ($p < 0.01$). Similarly, a significant reduction was also seen with the supplementation of caffeic acid (**49.05%** $\pm$ 4.23) and quercetin (**43.98%** $\pm$ 3.77), ($p < 0.01$). More significant reduction of DNA damage was observed at 72 h of *G. thoracica BPE* (**20.49%** $\pm$ 0.73), caffeic acid (**5.65%** $\pm$ 0.35), and quercetin (**7.58%** $\pm$ 0.32), ($p < 0.01$) treatment on H_2O_2-exposed HT-29 cells. In DNA methylation study, *G. thoracica* BPE treatment was able to significantly increased Alu methylation at 24 h (**26.00%** $\pm$ 0.58 vs. **21.50%** $\pm$ 0.96), ($p=0.007$) and at 48 h (**31.75%** $\pm$ 1.89 vs. **20.25%** $\pm$ 0.95), ($p=0.004$) compared with untreated cells. The means methylation of LINE-1 at all three sites were also significantly higher at 24 h (**79.92%** $\pm$ 1.26 vs. **72.83%** $\pm$ 2.21, $p=0.050$) and 72 h (**95.25%** $\pm$ 2.46 vs. **74.83%** $\pm$ 0.44, $p=0.001$) compared with untreated cells. Due to its strong antioxidant capacity, it showed that the ethanolic extract of *G. thoracica* BPE gave a protective effect towards H_2O_2-induced DNA damage in HT-29 cell lines through increasing Alu and LINE-1 methylation and finally could promote genetic stability.

CHAPTER 1

INTRODUCTION

1.1 Bee Pollen

1.1.1 Stingless Bee and Bee-collected Pollen

There are more than 500 stingless bee species, and the most common species producing honey are classified under two main genera, *Melipona* and *Trigona* (Al-Hatamleh *et al*., 2020). In Malaysia, 45 stingless bee species from 14 genera were documented (Mohd *et al*., 2010; Samsudin, Mamat and Hazmi, 2018), including *Tetrigona apicalis, Heterotrigona itama* and *Geniotrigona thoracica,* which are commonly domesticated for honey production (Fig. 1.1, 1.2, 1.3).

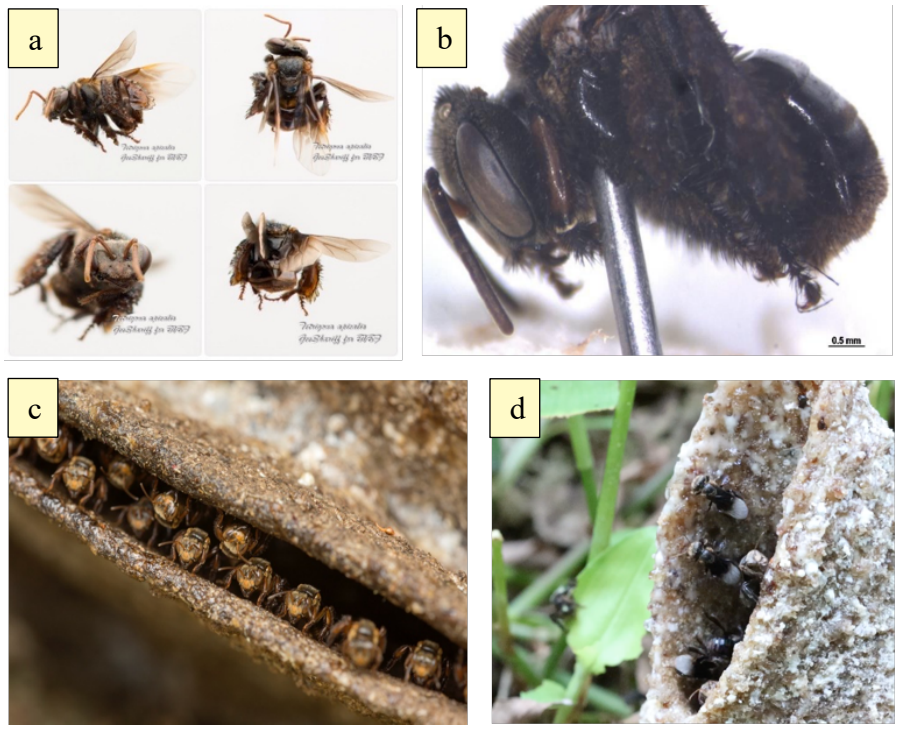

Figure 1.1 **(a)(b)** *Tetrigona apicalis* stingless bee, **(c)(d)** the flattened shape hive

*Source of image: (a)https://www.flickr.com/photos/geeshariff07/20593662789, (b)(Samsudin, Mamat and Hazmi, 2018), (c)https://www.flickr.com/photos/lonesomecrow/50844993726, (d)https://www.flickr.com/photos/budak/33755435218

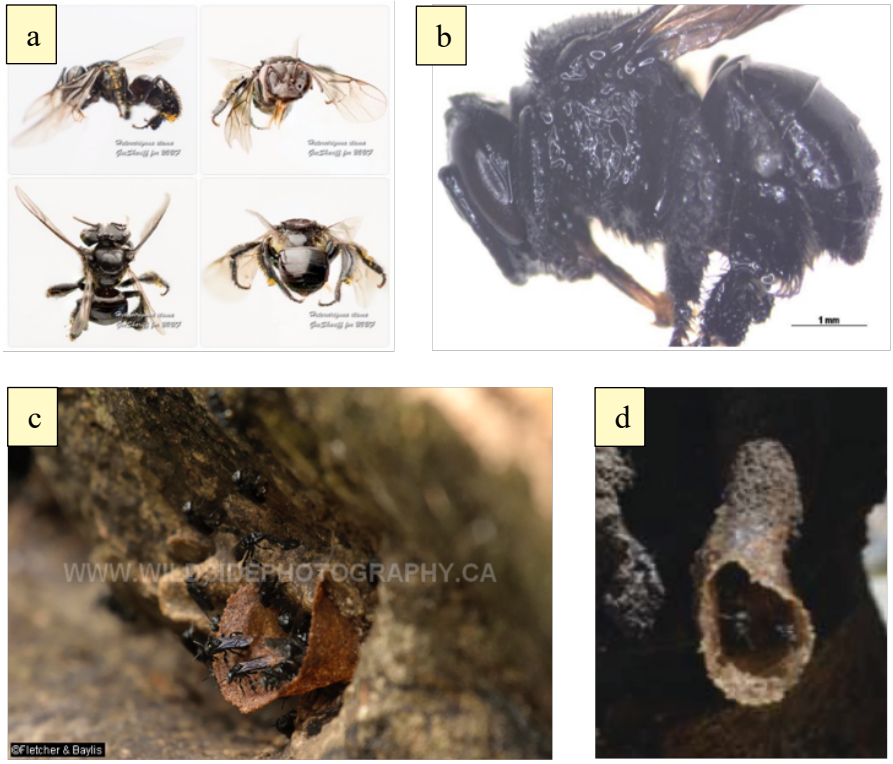

Figure 1.2 **(a)(b)** *Heterotrigona itama* stingless bee, **(c)(d)** funnel-shaped entrance of the hive

*Source of image: (a)https://www.flickr.com/photos/geeshariff07/20787399651, (b)(Samsudin, Mamat and Hazmi, 2018), (c)https://www.flickr.com/photos/wildside-photography/39600620520, (d)(Kelly *et al.*, 2014)

Stingless bee is a well-known pollinator in tropical rainforests. Various tropical fruits in Malaysia such as starfruit, mango, durian, rambutan, watermelon, guava, honeydew and coconut are pollinated by stingless bees. The beehives are made of a mixture of wax, resin and gum, and the entrance is species-specific to the bee (Klakasikorn *et al*., 2005). Stingless bees have different body sizes; the smallest is 2 mm, while the biggest is 14 mm (Mohd *et al*., 2010). *Geniotrigona thoracica* (mean body size: 8.44 mm) is among the species that possess a larger body size (Samsudin, Mamat and Hazmi, 2018). The most abundant domesticated Malaysian stingless bee species are *Heterotrigona itama and Geniotrigona thoracica,* where log hives of *H*.

itama can be more easily found in Malaysian forests (Fatima *et al*., 2018; Shamsudin *et al*., 2019)

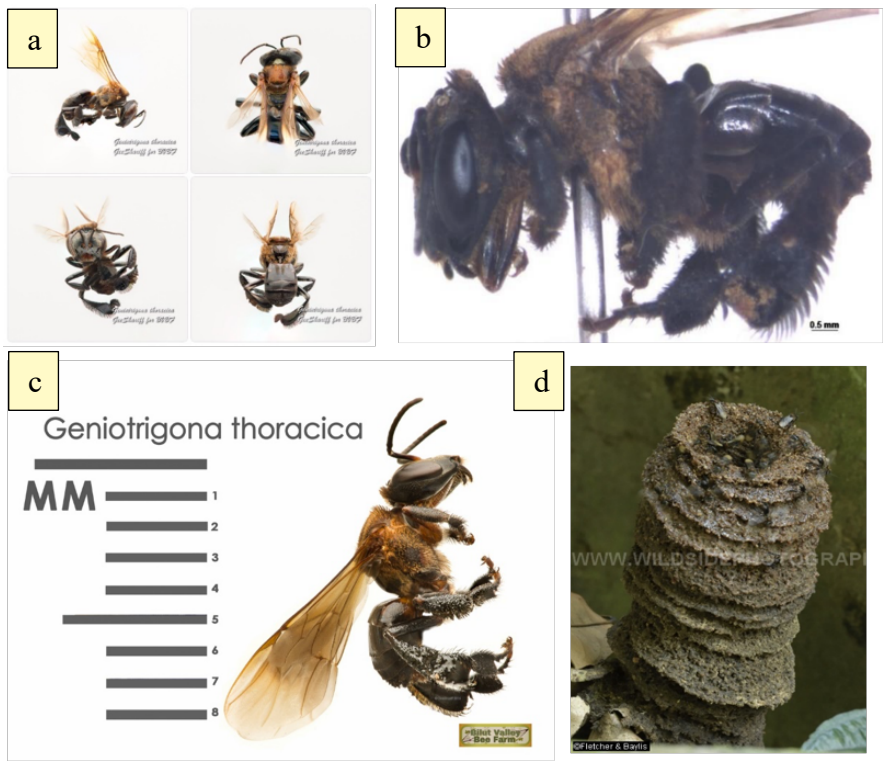

Figure 1.3 **(a)(b)(c)** *Geniotrigona thoracica* stingless bee, **(d)** the hill-shaped hive

*Source of image: (a)https://www.flickr.com/photos/geeshariff07/20780423955, (b)(Samsudin, Mamat and Hazmi, 2018), (c)https://www.flickr.com/photos/geeshariff07/26988047594, (d)https://www.flickr.com/photos/wildside-photography/49616874182

Stingless bee produces honey, propolis and bee pollen (Figure 1.4). Stingless bee-collected pollen composes flower pollen that preserved with floral nectar and the bee digestive enzymes (Takeshi Nagai, Reiji Inoue, Nobutaka Suzuki, Takao Myoda, 2005; Silva *et al*., 2009). Pollen is used as a source of food for the growth and survival of bees in the hive (Aličić *et al*., 2014). It was found that the pollen load of a single bee can weigh about 35% of the bee's body weight (Ellis *et al*., 2013).

Bee pollen serves as a source of nutrients for both adult bees and larvae. It is known as a complete food since the food energy produced is relatively high, ranging from 396.4 to 411.1 kcal/100 g of pollen (Kocot *et al*., 2018). Bee pollen is also a popular health supplement for maintaining health and longevity.

Studies had demonstrated 70% of bee pollen compositions are biologically active and exhibit numerous benefits including nutrition, cardioprotection, hepatoprotection, antioxidation, anticarcinogen, antibacteria, antiosteoporosis, antiprostatitis, anti-anemia, anti-aging, anti-inflammatory, and immunostimulant (Campos *et al*., 2010; K. Yang *et al*., 2013; Rzepecka-Stojko, Stojko, Kurek-Górecka, Górecki, Kabała-Dzik, *et al*., 2015).

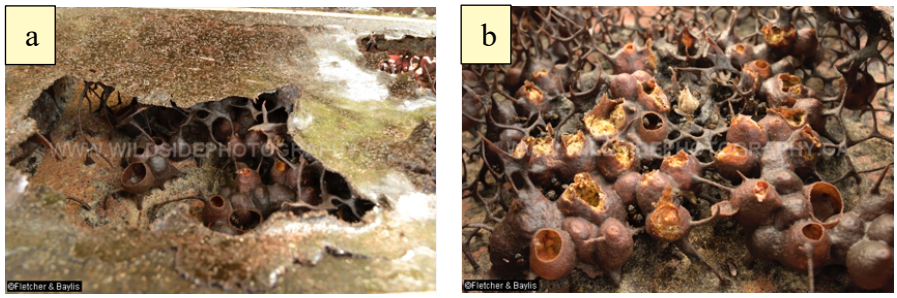

Figure 1.4 **(a)** Honey and pollen are stored beneath a rubbery sheet the bees make from plant resin and beeswax (propolis), **(b)** Pollen storage pots inside the beehive

*Source of image: (a)https://www.flickr.com/photos/wildside-photography/38504077984, (b)https://www.flickr.com/photos/wildside-photography/27435609569

1.1.2 Phenolic Compounds and Antioxidant Activity of Bee Pollen

Bee pollen that originates from floral pollen has a strong antioxidant activity. Few studies showed when bee pollen was supplemented with antitumor drugs, it could significantly enhance the immune system and inhibit tumor growth (X. Yang *et al*., 2007; Omar *et al*., 2016). In China, bee pollen is used as a food supplement to

strengthen body resistance against cancer and it was found to induce apoptosis in PC-3 cells, the human prostate cancer cells (Wu and Lou, 2007). Bee pollen extracts also inhibit proliferation and induce apoptosis in some other cancer cell lines (Wu and Lou, 2007; Kustiawan *et al*., 2014; Omar *et al*., 2016).

Its health beneficial effect is due to the composition of polyphenols and flavonoids, which are found abundantly in bee pollen and contribute to its high antioxidant activity (Carpes *et al*., 2009; Graikou *et al*., 2011). Polyphenols are ubiquitously distributed in plants as secondary metabolites – organic compounds that involve in protection against ultraviolet radiation and pathogens (Pandey and Rizvi, 2009). Polyphenols are regarded as effective antioxidants that work by eliminating free radicals and preventing DNA damage in the body and thus play important roles in the prevention of several diseases such as cancer, arteriosclerosis, diabetes, and cardiovascular diseases (J. Serra Bonvehí, Soliva Torrentó and Centelles Lorente, 2001; Bahadoran, Mirmiran and Azizi, 2013).

1.2 DNA Damage

DNA is highly susceptible to chemical modification, which can cause various damages and subsequently instability to the genome (Hoeijmakers, 2009). Endogenous DNA damage generated spontaneously during cellular metabolism can be due to oxidation, SAM-induced methylation, bases deamination, depurination and alkylation. Exogenous DNA damage that is caused by environmental agents includes carcinogenic and genotoxic factors such as ultraviolet (UV) from sunlight, ionizing radiation (IR), cigarette smoke, chest and dental X-rays, mammography, coronary angioplasty and tumor PET scan (Ciccia and Elledge, 2010).

Free radicals of oxygen or reactive oxygen species (ROS) are byproducts of normal cellular metabolism that are endogenously produced in living cells as part of the physiological processes, metabolic, and other biochemical reactions (Valavanidis, Vlachogianni and Fiotakis, 2009). ROS are also extracellularly produced from exposure to environmental agents such as superoxide radical (O2•–), hydroxyl radical (•OH), and non-radical hydrogen peroxide (H_2O_2) (Evans, Dizdaroglu and Cooke, 2004).

ROS caused oxidative damage to DNA, proteins, cellular membranes' lipids; and are considered etiopathogenic factors in various human diseases (Valavanidis, Vlachogianni and Fiotakis, 2009). ROS triggers oxidative stress and chronic inflammatory process in human cells (e.g., in the colonic mucosa), inducing oxidative damage, genomic instability and genetic mutations to the DNA (Ribeiro *et al*., 2008). Mutations may occur in the genes that regulate cell cycle and DNA repair, thus affecting differentiation, adhesion, proliferation, and apoptosis process in the cell (Ribeiro *et al*., 2008).

1.2.1 DNA Damage Event in Cancer

Oxidative DNA damage was observed in cancerous, pre-cancerous and also non-cancerous pathological conditions such as cardiovascular diseases (e.g., atherosclerosis and hypertension), neurodegenerative diseases (e.g., Parkinson's disease and Alzheimer's disease), and inflammation (e.g., rheumatoid arthritis and systemic lupus erythematosus), (Evans, Dizdaroglu and Cooke, 2004). Genomic instability associated with the accumulation of DNA damage is the hallmark of many cancers (Hoeijmakers, 2009).

Epigenetic events and oxidative damage are critical mechanisms that mainly contribute to the initial stages of carcinogenesis, such as colorectal cancer (CRC). Oxidative DNA damage modulates epigenetic regulation by global hypomethylation and regional hypermethylation of normally unmethylated CpG islands, thus leading to genome instability (Franco *et al*., 2008).

In colorectal cancer carcinogenesis, continuous exposure of ROS in the intestinal lumen would promote oxidative damage to the epithelial cells' DNA, triggering the occurrence of genetic mutations and epigenetic changes in CRC (Ehrlich, 2002; Ribeiro *et al*., 2008). Epigenetic changes such as global DNA hypomethylation induced by oxidative damage are common in cancer, promoting genome instability. Reduce methylation at repeat elements (which are abundant in the human genome) had been implicated in poor prognosis and survival outcomes in colorectal cancer (Mima *et al*., 2016).

High levels of oxidative stress and DNA damage also showed high mutation rates of the *p53* tumor suppressor gene, suggesting that the action of ROS could cause mutations to the control mechanism for programmed cell apoptosis, leading to carcinogenesis (Ribeiro *et al*., 2008).

1.2.2 DNA Damage Response and Repair

Numerous DNA damage incidents occur daily in ~10^{13} cells in the human body, which could block normal genome replication and transcription (Jackson and Bartek, 2009). DNA damage could cause abnormal replication and gene transcriptions, leading to gene mutations or wider-scale genome aberrations if they are not repaired correctly. DNA must be protected from damage; thus, cellular repair mechanisms have been developed to limit mutagenesis, cytotoxicity, cytostasis (inhibition of cell growth), and

maintain genomic integrity (Evans, Dizdaroglu and Cooke, 2004; Ciccia and Elledge, 2010).

When the DNA is damaged, repair enzymes fix the damage while not interfering with the DNA base-pair sequence. Epigenetic alterations are needed to modify the compact structure of DNA by acetylation or methylation process of histones and DNA CpG sequences (Bernstein and Bernstein, 2018).

A network of enzymes and molecules that regulate and repair DNA damage is generally known as DNA damage response (DDR). The intercellular and intracellular signaling events of DDR can regulate DNA replication, DNA damage repair, and cell-cycle arrest (O'Connor, 2015). Numerous enzymes that precisely regulate DNA damage are polymerases, nucleases, ligases, helicases, kinases, topoisomerases, glycosylases, recombinases, phosphatases, and demethylases (Ciccia and Elledge, 2010).

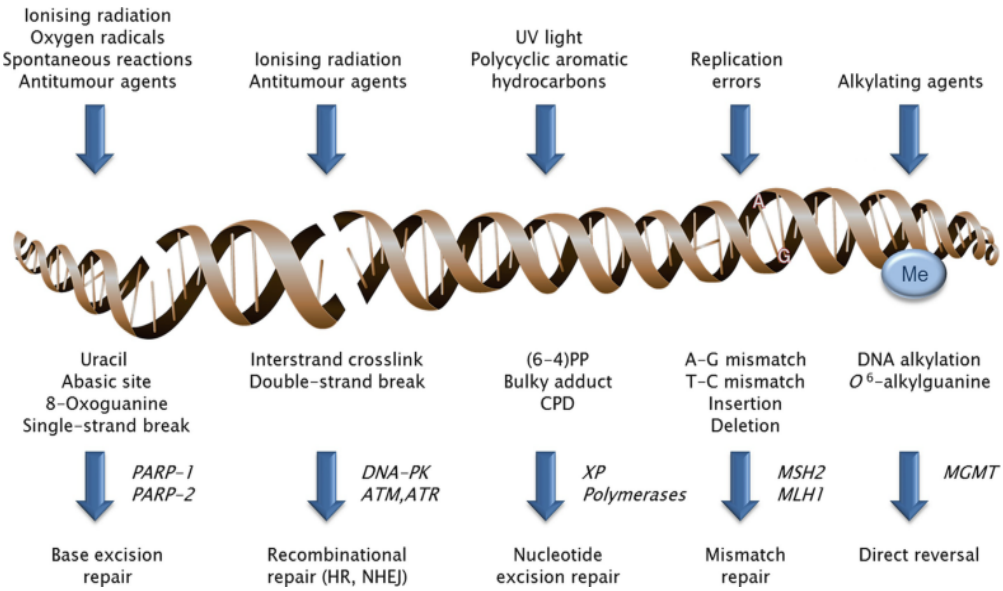

Figure 1.5 Mechanisms of DNA damage repair

*Source of image: (Harnor, Pickles and Cano, 2017)

Pathways for DNA repair mechanisms comprise base excision repair (BER), homologous recombination repair (HRR), non-homologous end joining (NHEJ), nucleotide excision repair (NER), and mismatch repair (MMR) (Mirza-Aghazadeh-Attari *et al*., 2018) (Figure 1.5). Minor DNA bases alterations like single-strand breaks (SSBs) are fixed by BER pathway through excision of the damaged base. At the same time, complex lesions such as pyrimidine dimers are repaired by NER pathway by means of damaged bases removal (Ciccia and Elledge, 2010). Both HRR and NHEJ pathways fix DNA double-strand breaks (DSBs), while MMR pathway repairs replication errors such as nucleotide insertions/deletions and mismatch base-pairing (O'Connor, 2015).

ROS that caused oxidative damage to the DNA could affect the epigenetic regulation by global hypomethylation and regional hypermethylation. The "access-repair-restore" model was developed to describe the roles of epigenetic modifiers in DNA repair (Polo and Almouzni, 2015). To serve the DNA repair process, epigenetic alterations act by relaxing or condensing certain chromatin regions to repress the transcription process, and finally return to the state before damage occurred when the repair process is completed (Bernstein and Bernstein, 2018). In a nutshell, DNA damage and repair process may have an impact and influence the DNA methylation print on that particular sequence.

1.3 Epigenetics

1.3.1 What is Epigenetics?

The term epigenetics was first established by Conrad H. Waddington from the Institute of Edinburgh in the year 1942, where he described epigenetics as the

interaction between genes and their surroundings. He introduced epigenetics as "the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being" (Waddington, 2012; Deans and Maggert, 2015).

Numerous studies were done on the relationship between cellular differentiation, DNA methylation and gene expression – they found that changes in gene expression also occurred during the adult stage instead of only during the development of an organism. Epigenetics was later redefined in a more specific way as "the study of changes in gene function that are mitotically or meiotically heritable and that do not entail change in the DNA sequence" (Wu and Morris, 2001; Deans and Maggert, 2015).

Basically, epigenetics is the study of biological mechanisms that switch genes 'on' and 'off', and how the genes work. In contrast to genetic changes, epigenetic modifications are reversible and do not alter the DNA sequence but can affect the transcription process (Eng, Herman and Baylin, 2000). Types of epigenetic changes that can influent gene expression are DNA methylation, histone modification and non-coding RNAs (microRNAs) (Sharma, Kelly and Jones, 2009).

Epigenetic events, predominantly DNA methylation, regulate levels of gene expression and maintaining genomic stability. Failure of the correct epigenetic process can result in improper activation or inhibition of numerous signaling pathways and can alter gene function and malignant cellular transformation (Sharma, Kelly and Jones, 2009).

1.3.2 Epigenetic Events in Early Tumor Development

Genetic alterations and epigenetic abnormalities involve the initiation and progression of a benign neoplasm, for example polyp; to malignant tumors (Sharma, Kelly and Jones, 2009). Genetic alterations include the mutations in several genes such as oncogenes (e.g., *KRAS*, *c-MYC*, *c-KIT*), tumor suppressor genes (e.g., *APC*, *TP53*, *SMAD4*) or DNA mismatch repair genes (e.g., *MLH1*, *MSH2*, *PMS2*). (Pappou and Ahuja, 2010; Nguyen and Duong, 2018; Xavier *et al*., 2019). Apart from the genetic alterations, epigenetic mechanisms also could initiate cancer development and progression.

In the initiation and progression of cancer, epigenetic mechanisms particularly DNA methylation, histone modifications, and non-coding RNA activity play significant pathophysiological roles (Villota-Salazar, Mendoza-Mendoza and González-Prieto, 2016). Non-coding RNA (ncRNA) such as microRNA (miRNA) and long ncRNA are important regulators of gene expression that are involved in many cancer-related pathways. DNA methylation and non-coding RNA regulators are closely interacting between each other and play important roles in determine the functional expression of the genes.

Epigenetic changes are reversible and biomarkers based on DNA methylation can be developed as potential future epigenetic drug/therapy for cancer treatment (Sharma, Kelly and Jones, 2009; Jung *et al*., 2020). Discussion in this thesis will emphasize mainly on DNA methylation patterns which include hypomethylation and hypermethylation.

1.4 DNA Methylation

One of the most broadly studied epigenetic alterations is DNA methylation. DNA methylation takes place when a methyl group (–CH3) is added by DNA methyltransferases (DNMTs) to the 5th carbon of the cytosine ring, where S-adenosylmethionine (SAM – the universal methyl donor) acts as a cofactor (Weisenberger *et al.*, 2005a) (Figure 1.6). In humans, DNA methylation predominantly takes place at CpG sites – regions of linear DNA bases along its 5'→3' direction with cytosine is preceded by guanine nucleotide; where 5-methyl cytosine (5mC) is formed (Cui and Xu, 2018). Its occurrence is spread across the genome.

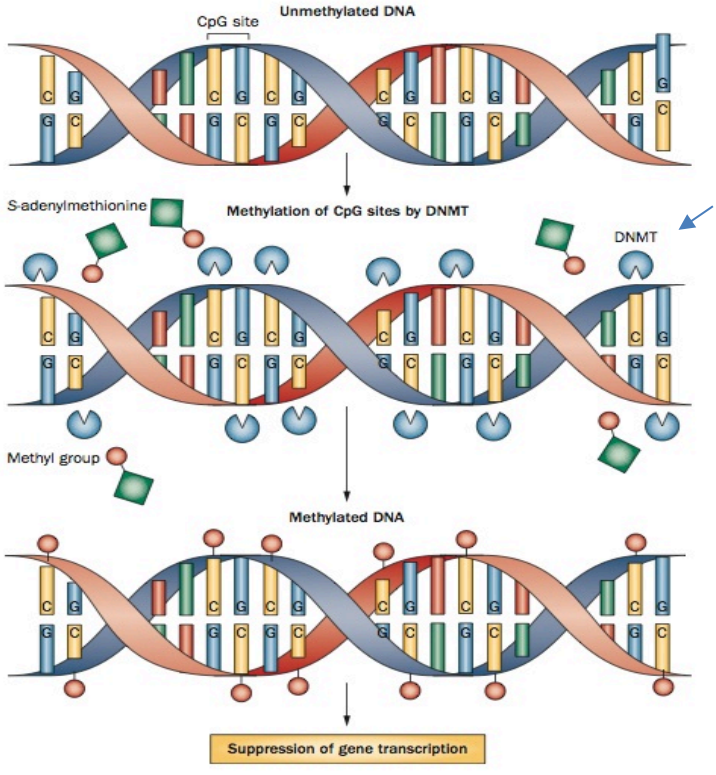

Figure 1.6 Mechanism of DNA methylation. DNA methyltransferases (DNMTs) attach methyl groups to the 5' position of cytosine nucleotides to methylate DNA

*Source of image: (Koch, Metz and Kovalchuk, 2013)

DNA methylation is a significant molecular mechanism involving genomic imprinting, x-inactivation, and tissue-specific gene expression. Aberrant DNA methylation in promoter and non-promoter regions has been associated with different diseases including cancer, neurodegenerative, cardiovascular, metabolic and autoimmune disorders (Pajares *et al*., 2020).

The DNA methylation patterns in tissue can be characterized or profiled in early cancer development to distinguish many types of tumour and cancer. The most commonly used techniques to assess specific DNA methylation are bisulfite conversion-based methods, biological/restriction enzyme-based approaches, and affinity enrichment-based assays (Pajares *et al*., 2020). Detailed methods for DNA methylation analysis are shown in Figure 1.7.

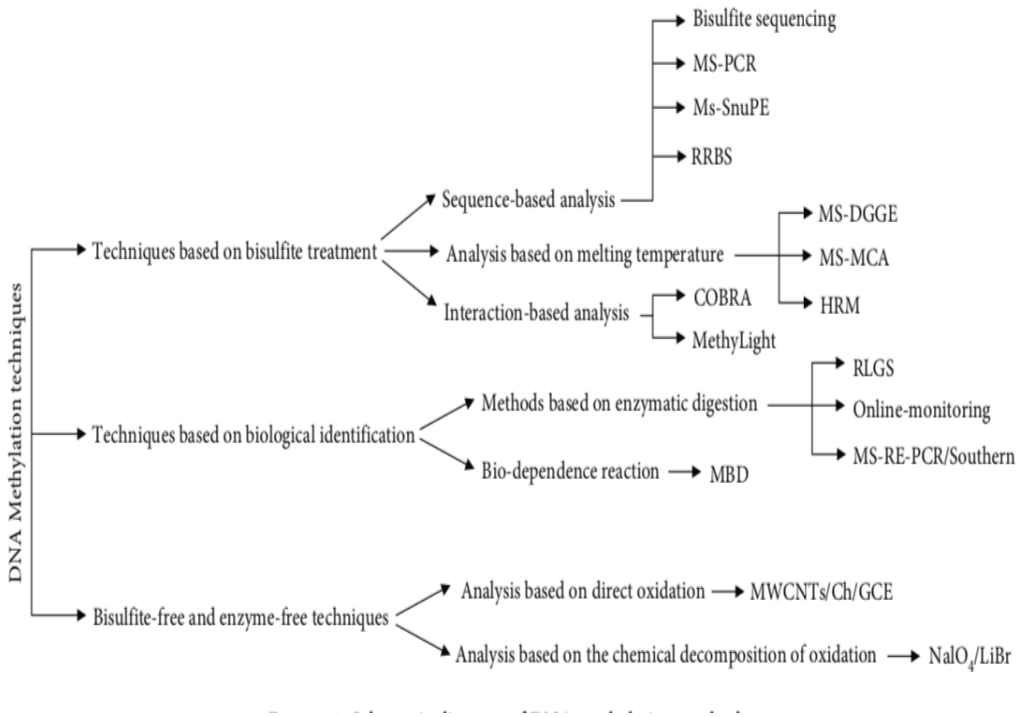

Figure 1.7 Detection methods of DNA methylation

*Source of image: (Khodadadi *et al.*, 2021)

DNMTs mediate the DNA methylation process. There are three classes of the DNMTs family of enzymes, i.e., DNMT1, DNMT2, and DNMT3A/3B/3L. DNMT1 regulates the maintenance of methylation, while DNMT2 is however poorly investigated. DNMT3A/DNMT3B regulate the *de novo* methylation, and DNMT3L (DNMT3-like) does not possess any fundamental enzymatic activity but interacts with DNMT3A2 (– an isoform of DNMT3) (Bilian and Robertson, 2013; Cui and Xu, 2018; Hervouet *et al*., 2018).

DNMTs are essential for transcriptional silencing of diverse sequence classes, particularly repetitive sequences, and perform a significant role in genomic integrity and chromosome stability. To regulate cell/tissue differentiation, DNMTs level changes in different developmental stages. Aberrant DNMTs activity may result in chromosome instability and tumor progression (Bilian and Robertson, 2013).

DNA methylation plays a role in maintaining chromosomal integrity, and regulation of gene expression and DNA recombination. Late replication of heavily methylated DNA forms inactive chromatin, which promotes transcriptional silencing of non-coding regions (Peinado, 2011), while methylation in a gene promoter region makes the gene to be silent. DNA methylation occurs in 70–80% of CpGs in mammalian DNA. It involves the development of an embryo, imprinting of parental genes, silencing of transposon, inactivation of X-chromosome, development of carcinogenesis, and aging (Hervouet *et al*., 2018).

In cancer, three classes of altered DNA methylation are hypomethylation, hypermethylation, and loss of imprinting (LOI) (Kim, Lee and Sidransky, 2010). Imprinting genes are inherited from the parents and are expressed as what have been expressed by the parent (parental allele-specific). LOI indicates the loss of

monoallelic gene regulation, normally caused by DNA methylation (Jelinic and Shaw, 2007).

CpG-islands' regional hypermethylation that occurs in the promoter and global DNA hypomethylation that occurs in repetitive elements are crucial DNA methylation changes that are associated with various cancers, including breast, head and neck, renal, bladder, and colorectal cancer (Hsiung *et al*., 2007; Moore *et al*., 2008; Choi *et al*., 2009; Liao *et al*., 2011; Walters *et al*., 2013).

DNA methylation is a dynamical age-related biochemical process regulating cell/tissue development (Cui and Xu, 2018). DNA methylation level also can be used to reflect the biological aging of human cells/tissues. A study by Horvath (2013) had identified 353 CpG sites to form an epigenetic age clock where he demonstrated that cancer tissues appeared to be more than 30 years older than healthy tissues (Horvath, 2013).

1.4.1 CpG Islands

Approximately one per 80 nucleotides are CpG dinucleotides (Cytosine followed by Guanosine dinucleotides, with 'p' indicates the phosphate link between two nucleotides) that are outspread throughout the genome and comprise 1–2% of the human genome (Baba *et al*., 2018). 'CpG islands' are isolated CpG-rich regions (short DNA sequences rich in CpG sites) that consist of more than 55% C-G content and the sequence is longer than 500 bp (Wang and Leung, 2004). Around 45,000 CpG islands generally overlap with promoter regions of various genes, while nearly 40% of the promoter regions contain CpG islands (Weisenberger *et al*., 2005a; Baba *et al*., 2018).

Therefore, CpG islands play an important role in the regulation of gene expression. Changes in methylation, either reduce genomic methylation globally or regional CpG-islands hypermethylation may cause specific tumors or tumor subtypes (Eng, Herman and Baylin, 2000). An average of 600 out of 45,000 CpG islands in the genome are aberrantly methylated in different types of tumors (Costello *et al*., 2000). While 70% of CpG sites are methylated in the entire mammalian genome, regional CpG islands are mostly unmethylated in normal cells (Cui and Xu, 2018). Thus, methylation of CpG islands that normally contain gene promoters or exons, is related to delayed replication, condensed chromatin, inhibition of transcription, and long-term gene silencing (Costello *et al*., 2000).

1.4.2 DNA Hypomethylation

DNA hypomethylation refers to the loss of the methyl group in the 5-methylcytosine (5mC) nucleotide of most CpG sites, either in a specific sequence or in the bulk of the genome that is usually methylated (Peinado, 2011). This event substantially decreased the proportion of methylated versus unmethylated cytosines relative to the normal tissue-specific pattern. In 1983, DNA hypomethylation was the first epigenetic change reported in human cancer (Nishiyama *et al*., 2005; Ehrlich, 2009).

Global DNA hypomethylation is a genome-wide event of the overall decrease in 5-methylcytosine in the DNA, leading to activation of previously silenced genes and contributing to cancer development and progression (Weisenberger *et al*., 2005a). A remarkable global DNA hypomethylation was reported in the cancer cell, with 20–60% reduction of genomic 5mC than its normal condition (Esteller and Herman, 2002).

Global DNA hypomethylation also affects and associates with repeated DNA elements and has multiple ways of contributing to carcinogenesis (Weisenberger *et al*., 2005a; Ehrlich, 2009; Walters *et al*., 2013). The methylation loss of repetitive sequences causes silenced areas of the genome (such as retrotransposons) to become active. This event would lead to genetic instability contributing to cancer development and progression (Baba *et al*., 2018).

Global hypomethylation was found in almost every type of cancer, as early occurrence or late phases. However, global hypomethylation was higher in metastatic tumors compared to primary tumors (Ehrlich, 2002; Frigola *et al*., 2005). Global hypomethylation promotes cancer progression through chromosomal stability (chromosome structure and integrity) and causing genomic instability (Eden *et al*., 2003). Genomic instability plays a role in activating cancer gene expression (Feinberg and Tycko, 2004; Frigola *et al*., 2005).

The reduction of genome methylcytosine content could serve as a marker in colorectal cancer and other tumor types (Peinado, 2011). Global hypomethylation is associated with both cancer phenotype (patient survival) and cancer genotype (genomic disruption) (Suzuki *et al*., 2006).

There are three mechanisms of carcinogenesis that are contributed by global hypomethylation – chromosomal instability, reactivation of transposable elements especially in LINE-1 and Alu repeats, and loss of imprinting/ affect the imprinted genes (genes that are supposed not to be expressed) (Esteller and Herman, 2002).

1.4.3 DNA Hypermethylation

While global DNA hypomethylation occurred at the genome level, increased DNA methylation or hypermethylation frequently occurred at the gene-specific level. Hypermethylation that involves the promoters of various genes can cause the genes to be inactivated in different malignant tumors, including breast, lung, prostate and colon cancers (Moore *et al*., 2008; Dumitrescu, 2012). Gene locus-specific hypermethylation of several genes such as tumor suppressor genes, homeobox genes, and genes involved in DNA repair and apoptosis are characteristics of the cancer genome (Ehrlich, 2009; Kim, Lee and Sidransky, 2010). The expression of certain tumor suppressor genes are repressed in metastatic tumor cells, avoiding cells from apoptotic cell death and thus acquire cells with the metastatic phenotype (Kim, Lee and Sidransky, 2010).

1.4.4 Global DNA Hypomethylation and Hypermethylation of Specific Genes in Colorectal Cancer

Colorectal cancer (CRC) is the second most common cancer in Malaysia, the most common cancer in men (16.3%), and the second most common in women (10.7%) (Chandran *et al*., 2020). Sixty-five percent of the CRC cases are detected at late stages (stage three and four), and 5-year survival rates for all ethnic groups are 41% to 53% (Veettil *et al*., 2017; Chandran *et al*., 2020). In Malaysia, 90% of the CRC cases are detected in people more than 40 years old, while in younger adults, the incidence increased by 2% per year (Veettil *et al*., 2017).

The gradual genetic alterations and epigenetic abnormalities in colonic epithelial cells are the key factors that occur during neoplastic transformation, which are influenced by multifactorial etiology, nutrigenomics, dietary and lifestyle factors (Kim, Lee and Sidransky, 2010). Genetic and lifestyle factors that generally contribute

to the increased risk of CRC are obesity (4.5%), smoking (9%), high-fat diet (9%), age (11%), low fiber diet (16%), and family history (24%) (Ghee, 2014).

Various tumor suppressor genes and oncogenes are associated with CRC, such as *APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *TP53*, *TGFBR2*, *KRAS*, *SMAD4*, *PTEN*, and *BRAF* (Munteanu and Mastalier, 2014). In the CRC model of carcinogenesis, the transition to malignancy starts with inactivation of tumor suppressor gene *adenomatous polyposis coli* (*APC*) that causes failure to control cell proliferation, followed by oncogenic *KRAS* mutations, and finally deletion of chromosome 18q and inactivation of *TP53* tumor suppressor gene on chromosome 17p (Nguyen and Duong, 2018).

The major pathways of CRC pathogenesis are chromosomal instability (CIN), microsatellite instability (MSI) and aberrant DNA-methylation (Munteanu and Mastalier, 2014). In CIN, chromosome changes occur in number, structure, rearrangements and loss of heterozygosity (LOH), which lead to loss of the wild allele of suppressor genes that function in blocking the development of malignant phenotype (Munteanu and Mastalier, 2014). In MSI, the changes occur in microsatellites, the nucleotides that typically repeat in genomic DNA or protein transcription. MSI is also related to the loss of function of DNA mismatch repair genes.

Another change during CRC carcinogenesis is aberrant DNA methylation, including the loss of global methylation level, especially during aging; and hypermethylation of CpG islands near gene promoters (Munteanu and Mastalier, 2014). The development of CRC was contributed by both global hypomethylation and hypermethylation of CpG islands near promoter sites of specific genes (Figueiredo *et al*., 2009).

Both global DNA hypomethylation and site-specific CpG island promoter hypermethylation are cancer-linked phenomena that involve the development of tumorigenesis, i.e., in early tumorigenesis and increase with tumor progression (Ehrlich, 2009; Baba *et al*., 2018). There are differences in hypomethylation and hypermethylation patterns based on cancer-type specific, where some DNA sequences can be either extensively hypomethylated or hypermethylated in certain cancer (Costello *et al*., 2000; Nishiyama *et al*., 2005).

Most CRC evolves in a stepwise progression from benign neoplasms into adenocarcinomas in the adenoma-carcinoma sequence. Significant differences in gene methylation status are observed between adenomas and carcinomas, suggesting that both hypomethylation and hypermethylation play crucial roles in the initiation and progression of CRC (Kim, Lee and Sidransky, 2010). In colorectal neoplasia, global DNA hypomethylation is connected with an increased risk of colorectal carcinogenesis and metastatic stages of CRC (Bariol *et al*., 2003; Kim, Lee and Sidransky, 2010).

Jordi Frigola et al. (2005) suggested that global hypomethylation (in early CRC progression) and hypermethylation (in advanced stages) are regulated by different mechanisms and independently play different roles in tumorigenesis of CRC (Frigola *et al*., 2005).

DNA hypomethylation that occurs at the repetitive sequences is associated with mechanisms that could cause neoplastic progression (Bariol *et al*., 2003), usually the loss of genomic imprinting that is seen in ~40% of CRC tissue, for instance, at the *IGF2/H19* region (Kim, Lee and Sidransky, 2010). In DNA hypermethylation, aberrant methylation of CpG islands near the promoter sites may silence the gene expressions and alter specific oncogenic pathways leading to cancer progression. For

example, hypermethylation of *hMLH1* and *MGMT* DNA repair genes promoter may reflect the early carcinogenic process of CRC (Menigatti *et al*., 2009).

DNA hypomethylation is consistent with DNA hypermethylation but occurs in different sequences (Ehrlich, 2009). Hypomethylation patterns could activate the previously silenced genes and may subject chromosomes to breakage, therefore predispose chromosomes to CIN and aneuploidy (Kim, Lee and Sidransky, 2010). Increased DNA hypomethylation and hypermethylation are associated with increased genomic damage in CRC, including aneuploidy in CRC cell lines and high mutation rate (Ehrlich, 2002; Feinberg and Tycko, 2004; Suzuki *et al*., 2006; Peinado, 2011). In a nutshell, suppressor genes are inactivated by hypermethylation of CpG island in the promoter, resulting in their transcriptional silencing; while genomic instability is associated with genome-wide hypomethylation and thus expedites tumor progression.

1.4.5 Repetitive and Transposable Elements

Genome-wide global hypomethylation is an event that influences repetitive and transposable DNA elements (Park *et al*., 2014). Approximately half of the human genome comprises repetitive elements subdivided into two principal types; interspersed repeats and tandem repeats (or satellite DNA) (Weisenberger *et al*., 2005a; Gemayel *et al*., 2010).

Interspersed repeats derive from autonomous (can move on their own; protein-coding; e.g., LINE-1) or non-autonomous (needs other TEs to move; non-protein-coding; e.g., Alu) transposable elements (TEs). Another repetitive element is tandem repeats that consist of complex or simple sequences and are located adjacent to one another (Weisenberger *et al*., 2005a; Pray, 2008; Gemayel *et al*., 2010; Walters *et al*., 2013; Baba *et al*., 2018). About ~45% of the human genome are interspersed repeat

sequences/TEs, and ~3% of the genome was made up of tandem repeats (Huda *et al*., 2009).

TEs are DNA sequences that can transpose to new locations in the genome. Retrotransposons (or RNA transposons) are Class 1 TEs that mobilize through a 'copy-and-paste' mechanism, and DNA transposons are Class 2 TEs that are mobilized through a 'cut-and-paste' mechanism using DNA intermediate (Bourque, Kathleen H. Burns, *et al*., 2018). Most human elements are Class 1 Retrotransposons, including short interspersed nuclear elements (SINEs; e.g., Alu) and long interspersed nuclear elements (LINEs; e.g., LINE-1) (Burns, 2017). Thirty-four percent of the human genome consists of SINEs and LINEs that invade new genomic sites using RNA intermediates (Weiner, 2002).

Repeated sequences are highly methylated at their CpG sites in the genome (Weisenberger *et al*., 2005a). The methylated repetitive sequences are integral to maintaining genome stability. Alu and LINE-1 are the main repetitive transposable DNA elements and the most regularly studied DNA cancer-hypomethylated repeats (Hoffmann and Schulz, 2005; Rodriguez *et al*., 2008; M. J. Kim *et al*., 2009). CpG sites within Alu and LINE-1 are commonly methylated in normal cells, contributing to the inactivation of transcription and suppression of retrotransposition (Park *et al*., 2014).

Because of the abundance of Alu and LINE-1 levels in the genome, they are beneficial to be used as a surrogate marker for genome-wide methylation status. Methylation of Alu and LINE-1 had been shown to strongly associated with global DNA methylation, measured by HPLC, which had been considered as a gold standard in measuring global DNA methylation (Lisanti *et al*., 2013).

Activation of transcription in the event of Alu and LINE-1 hypomethylation results in the transposable element retrotransposition, chromosome alteration and finally genomic instability (Saito *et al*., 2010; Bae *et al*., 2012). Alu and LINE-1 hypomethylation was observed as an early incident in the multistep carcinogenesis of CRC and other types of human cancers (Chalitchagorn *et al*., 2004; B. H. Kim *et al*., 2009; Lee *et al*., 2009; Kwon *et al*., 2010; Sunami *et al*., 2011; Bae *et al*., 2012; van Hoesel, A. Q. *et al*., 2012).

1.4.6 Alu

Alu, a ~282 bp non-Long Terminal Repeat (non-LTR) DNA sequence, is the most abundant SINE element in the human genome. Alu elements are presented in ~1 million copies per haploid genome, comprising ~10% of the human genome (Weisenberger *et al*., 2005a; Ye *et al*., 2020). There are three major subfamilies of Alu, i.e., AluJ, AluS, and AluY; where AluY express the highest rate of retrotransposition in the human genome (Luo, Lu and Xie, 2014).

Alu hypomethylation is most significant at certain ages (34 to 68) and associated with an increased risk of various tumors (Luo, Lu and Xie, 2014). Alu hypomethylation influences genomic instability by gene recombination, chromosome translocation, nucleosome formation and genome evolution (Ye *et al*., 2020).

1.4.7 LINE-1

Long interspersed nuclear elements (LINEs) are the most transpositionally active elements in the human genome which encode for two proteins: open reading frame (ORF)-1 that acts as RNA-binding protein, and ORF-2 that functions as an endonuclease and reverse transcriptase (Dunker *et al*., 2017). The only abundant

LINEs in human is LINE-1, a ~6kb non-LTR sequences that comprise approximately 17-20% of the human genome (Weisenberger *et al*., 2005a; Baba *et al*., 2018).

LINE-1 elements are presented at over 500,000 copies in the human genome; however, due to the accumulation of random mutations over time, only 30–100 are active retrotransposons (Weisenberger *et al*., 2005a). LINE-1 elements are protein-coding retrotransposons that self-propagate themselves through RNA intermediate. About 2,700 copies of LINE-1 protein in the human genome act as reverse transcriptase and can be integrated into the new site in the genome (Burns, 2017).

Gastrointestinal tract tumors, particularly CRC, are prone to somatic LINE-1 activity (Burns, 2017). LINE-1 hypomethylation is strongly related to poor prognosis in different cancers, including CRC and other gastrointestinal cancers (Baba *et al*., 2018). LINE-1 hypomethylation associates with heterozygosity losses on different chromosomal loci and precede genomic damage in tumor cells (Peinado, 2011). In CRC, LINE-1 insertions at the *APC* tumor suppressor gene could harm the gene function by disrupting coding exons or interfering with mRNA splicing close to the exons (Burns, 2017). LINE-1 hypomethylation demonstrates a more aggressive progression of CRC and could activate the methylation-silenced proto-oncogenes in CRC metastasis (Hur *et al*., 2014). Inevitably, LINE- 1 methylation could be used as a surrogate marker of global DNA methylation and could act as a potential prognostic biomarker of CRC.