EFFECTS OF CpG ISLANDS DNA METHYLATION ON THE HUMAN CHOLINE KINASE ALPHA PROMOTER ACTIVITY

SITI AISYAH FATEN BT MOHAMED SA'DOM

UNIVERSITI SAINS MALAYSIA

2021

EFFECTS OF CpG ISLANDS DNA METHYLATION ON THE HUMAN CHOLINE KINASE ALPHA PROMOTER ACTIVITY

by

SITI AISYAH FATEN BT MOHAMED SA'DOM

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

May 2021

ACKNOWLEDGEMENTS

Firstly, praise to Allah S.W.T for giving me the strength, health and patience in completing my research in PPSK. I would like to express my deepest appreciation to my supervisor, Assoc. Prof. Dr Few Ling Ling for her continuous guidance, encouragement and patience from the early stage of this research. I would like to thank my co-supervisor, Assoc. Prof. Dr See Too Wei Cun for the commitment and guidance you have given me. Without their guidance and persistent help, this thesis would not have been possible. I would like to thank my co-supervisor, Prof. Shaharum for his kind assistance. I would sincerely like to thank my friend and labmate, Sweta Raikundalia Pradeep who was always with me and support me through thick and thin. I also want to extend my thanks and acknowledge all staff in PPSK, especially Pn Siti Kurunisa Mohd Hanafiah, as well as students in PPSK especially for their contributions and constant support. My deepest gratitude of love and appreciation is dedicated to my parents, Mr. Mohamed Sa'dom Daud and Mrs. Siti Rakiah Mamat for their endless love and support throughout my study period. Not forgotten my sister, Nur Zulaikha and my brothers. I thank them for always being there for me. Last but not least, many thanks to Universiti Sains Malaysia for the financial assistance funded by the Bridging Grant Scheme (304/PPSK/6316334) and Research University Individual (RUI) Grant Scheme (1001/PPSK/8012239) from Universiti Sains Malaysia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

$(NH_4)_2SO_4$	Ammonium sulphate
°C	Degree celcius
µg	Microgram
µl	Microliter
µM	Micromolar
5-azaCdR	5-aza-2'-deoxycytidine
5-azaCR	5-azacytidine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
APS	Ammonium persulfate
ATP	Adenosine Triphosphate
AZA	5-azacytidine
BRE	B recognition element
BSA	Bovine serum albumin
CAGE	Cage Analysis of Gene Expression
CCD	Charge-coupled device
CCT	Cytidyltransferase
cDNA	Complementary DNA
CER	Cytoplasmic Extraction Reagent
CG	Cytosine Guanine
CH_3	Methyl group
ChIP	Chromatin Immunoprecipitaion
CK	Choline kinase
CMP	Cytidine monophosphate
CO_2	Carbon dioxide
CpG	C-phosphate-G
CPT	Cholinephosphotransferase
CRM	*cis*-regulatory module
CTSB	Cathepsin B gene

DAG	Diacylglycerol
DBTSS	DataBase of Human Transcriptional Start Site
DEC	Decitabine
DHAC	Dihydro-5-azacytidine
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxynucleotide triphosphate
DPE	Downstream core promoter element
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGCG	Epigallocatechin 3-gallate
EGF	Epidermal Growth Factor
EGFR	Estimated Glomerular Filtration Rate
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic Mobility Shift Assay
ER	Estrogen receptor
ES	Embryonic stem
Exp_{CpG}	Expected-CpG
FAS	Fatty acid synthase
FBS	Fetal bovine serum
FDA	Food and Drug Administration
g	Times gravity
GTF	General transcription factor
HC-3	Hemicholinium-3
HDAC	Histone deacetylase
HIF	Hypoxia-inducible Factor
HRE	Hypoxia response element
HT-SELEX	High throughput-SELEX
Inr	Initiator element

IRS2	Insulin Receptor Substrate 2
kb	Kilobase
LB	Luria Bertani
luc	Luciferase
MAPK	Microtubule associated protein kinase
MBD	Methyl-binding domain
mC	Methylcytosine
MCF-7	Michigan Cancer Foundation-7
^{m}CG	Methylcytosine-guanine
MeCP1	Methyl-CpG binding protein 1
MeCP1	Methyl-CpG binding protein 2
MeDIP	Methylated DNA Immunoprecipitation
mg	Milligram
$MgCl_2$	Magnesium chloride
ml	Mililiter
mM	Milimolar
MRS	Magnetic Resonance Spectroscopy
MTase	Methyltransferase
MTE	Motif ten element
Mw	Molecular weight
MZF1	Myeloid zinc-finger 1
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanogram
Obs_{CpG}	Observed-CpG
OD	Optical density
PAH	Phenylalanine hydroxylase
PBS	Phosphate buffered saline
PC	Phosphatidylcholine
PCho	Phosphocholine

PCR	Polymerase Chain Reaction
PC-TP	Phosphatidylcholine Transfer Protein
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol 800
pH	Potential of hydrogen
PH	Purinyl-6-histamine
PI	Phosphatidylinositol
PIC	Preinitiation complex
POPC	1-palmitoyl-2-oleoyl-*sn*-glycerol-3-phosphocholine
PPAR	Peroxisome Proliferator-Activated Receptor
PS	Phosphatidylserine
psi	Pound per square inch
RLGS	Restriction Landmark Genomic Scanning
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
SAM	S-adenosyl-methionine
SD	Standard deviation
SDM	Site-directed mutagenesis
SELEX	Systemix Evolution of Ligands by Exponential enrichment
siRNA	Small interfering RNA
SRA	SET and RING associated
TAE	Tris-acetate-EDTA
TET	Ten-eleven translocation
TF	Transcription factor
TMA-NP	Tetramethylaammonium-based Nanopore
TSS	Transcription start site
v/v	Volume per volume
w/v	Weight per volume
XRE	Xenobiotic response element

LIST OF APPENDICES

KESAN METILASI DNA KEPULAUAN CpG TERHADAP AKTIVITI PROMOTER KOLINA KINASE ALPHA MANUSIA

ABSTRAK

Kolina kinase (CK) adalah enzim sitosolik yang merupakan pemangkin yang terlibat dalam fosforilasi kolina kepada penghasilan fosfokolina (PCho) dalam proses biosintesis fosfatidilkolina (PC), komponen utama dalam fosfolipid membran. Di sebalik kepentingan CK dalam biosintesis PC, pertumbuhan sel dan karsinogenesis, maklumat berkaitan pengawalaturan transkripsi gen *ckα* masih terhad. Kewujudan kepulauan CpG di bahagian promoter gen *ckα* mencadangkan penglibatan metilasi DNA dalam pengawalaturan transkripsi gen *ckα*. Oleh itu, kajian ini bertujuan untuk mengkaji kesan metilasi DNA kepulauan CpG terhadap aktiviti promoter gen *ckα*. Promoter gen *ckα* bersaiz 2009 bp telah diklonkan ke dalam vektor pelapor, firefly luciferase (pGL4.10) untuk menghasilkan plasmid rekombinan, pGL4.10-*ckα* (-2000/+9). Kemudian, satu siri mutasi penghapusan kepulauan CpG telah dihasilkan dengan kaedah mutagenesis berpandu tapak PCR, dan diklon ke dalam vektor pGL4.10 untuk dikaji dalam sel adenokarsinoma payudara manusia, MCF-7. Status metilasi selepas rawatan dengan menggunakan agen pengurangan metilasi, 5-azasitidina dan agen penambahan metilasi, budesonida menunjukkan peranan metilasi DNA di bahagian promoter gen *ckα* yang lebih ketara dalam sel kanser MCF-7 berbanding sel normal MCF10A. Sebanyak empat kepulauan CpG telah dikenalpasti di dalam kawasan promoter ini menggunakan perisian MethPrimer dan EMBOSS CpGPlot. Penyingkiran pada kawasan -225 ke -56 bp dalam kepulauan CpG keempat menunjukkan peningkatan aktiviti promoter berbanding dengan

promoter berkepanjangan penuh. Ini menunjukkan adanya unsur pengawalseliaan negatif yang penting yang mungkin dimodulasikan oleh metilasi DNA. Analisis *in vitro* menunjukkan metilasi promoter berkepanjangan penuh menghasilkan aktiviti yang lebih rendah jika dibandingkan dengan metilasi promoter yang terhapus kepulauan CpG keempat. Ini menggambarkan bahawa kepulauan CpG ini mungkin mengandungi tapak pengikatan untuk faktor transkripsi penghalang. Mutasi tapak pengikatan MZF1 menunjukkan peningkatan yang signifikan dalam aktiviti promoter *ckα* berbanding dengan promoter *ckα* berkepanjangan penuh sekaligus menunjukkan sifatfungsi perencatan elemen jujukan ini. Analisis EMSA menunjukkan terdapat pengikatan faktor transkripsi pada bahagian tapak perlekatan MZF1, dan metilasi sitosina pada bahagian ini menunjukkan peningkatan terhadap pengikatan faktor MZF1 jangkaan ini di bahagian -181 sehingga -175 pada promoter *ckα*. Tambahan lagi, mutasi pada tapak perlekatan MZF1 menghapuskan pembentukan kompleks protein-DNA. Ini menunjukkan bahawa metilasi DNA mengurangkan aktiviti promoter *ckα* dengan cara mempromosikan pengikatan faktor transkripsi MZF1 di kepulauan CpG keempat pada bahagian -225/-56. Sebagai kesimpulan, kajian ini memberi perspektif mengenai penglibatan kepulauan CpG dan metilasi DNA dalam pengawalaturan transkripsi gen *ckα*.

EFFECTS OF CpG ISLANDS DNA METHYLATION ON THE HUMAN CHOLINE KINASE ALPHA PROMOTER ACTIVITY

ABSTRACT

Choline kinase (CK) is a cytosolic enzyme catalyzing the phosphorylation of choline to phosphocholine (PCho) in the biosynthesis of phosphatidylcholine (PC), a major component of membrane phospholipid. Despite the importance of CK in PC biosynthesis, cell growth and carcinogenesis, little is known about the transcriptional regulation of *ckα* gene. The presence of CpG islands on the promoter region of *ckα* gene suggests the involvement of DNA methylation in its transcriptional control. Therefore, this study aimed to investigate the effects of CpG islands DNA methylation on *ckα* gene promoter activity. A 2009 bp promoter region of the human *ckα* gene was cloned into a firefly luciferase reporter vector (pGL4.10) to create a recombinant plasmid, pGL4.10-*ckα* (-2000/+9). Then, a series of CpG island deletion mutants were constructed using PCR site-directed mutagenesis method and cloned into pGL4.10 vector and studied in human breast adenocarcinoma, MCF-7 cells. The methylation status after treatment with a demethylating agent, 5-azacytidine and re-methylating agent, budesonide showed a prominent role of DNA methylation of *ckα* gene promoter in MCF-7 cancer cells compared to the corresponding normal cells MCF10A. A total of four CpG islands were identified within the promoter region by using MethPrimer and EMBOSS CpGPlot software. Deletion of the region between -225 to -56 bp in the fourth CpG island showed an increased promoter activity as compared to the full-length promoter indicating the presence of important negative regulatory elements which could be modulated by DNA

methylation. An *in vitro* methylation analysis showed the methylated full-length promoter activity was significantly lower than the methylated fourth CpG island deletion suggesting that this CpG island contains elements for the binding of suppressor transcription factors. Mutation of MZF1 binding site in the fourth CpG island caused a significant increase in the *ckα* promoter activity, suggesting a repressive role of this sequence element. EMSA analysis showed that there is a binding of transcription factor to the MZF1 binding site, and the cytosine methylation at this site showed an increase of the binding of this putative MZF1 transcription factor at -181 to -175 *ckα* promoter region. Furthermore, mutation of MZF1 binding site abolished the protein-DNA formation complex. These results suggest that DNA methylation decreased the *ckα* promoter activity by promoting the binding of MZF1 transcription factor to the fourth CpG island located at -225 to -56 region. In conclusion, this study provides a perspective on the involvement of CpG island and DNA methylation in the transcriptional control of *cka* gene.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Choline kinases (CK) (EC 2.7.1.32) are cytosolic enzymes catalyzing the phosphorylation of choline to phosphocholine (PCho) in the biosynthesis of the phosphatidylcholine (PC) (Wu *et al.*, 2008). PC is the primary phospholipid of eukaryotic cellular membranes and has crucial roles in the structure and function of those membranes (Gibellini and Smith, 2010). Human CK is encoded by two separate genes named *ckα* and *ckβ. ckβ* codes for a single protein (CKβ) while *ckα* undergoes alternative splicing to produce CKα1 and CKα2 isoenzymes (Gallego-Ortega *et al.*, 2011). Increased activities of CK and PCho have been implicated in human carcinogenesis where CK overexpression increases the invasiveness and drug resistance of breast cancer cells (Shah *et al.*, 2010). Many researchers have focused on the abnormal expression of *ckα* in various human cancers such as colorectal, lung, and prostate adenocarcinomas (Nakagami *et al.*, 1999; Ramirez de Molina, 2002; Rizzo *et al.*, 2021) and the potential of *ckα* inhibition as anticancer therapy. Yet, the regulation of CK gene expression at the transcriptional level, particularly by epigenetic mechanism, has never been explored.

Epigenetics is defined as a heritable process that alters gene activity without changing the DNA sequence (Weinhold, 2006). Epigenetic processes are natural and vital to many organism functions, and abnormal epigenetic changes often lead to dysregulation of developmental activities (Hon *et al.*, 2012). DNA methylation is the most well-studied

epigenetic mechanism that involved in diverse cellular function, including silencing of transposable elements, inactivation of viral sequences, maintenance of chromosomal integrity, X-chromosome inactivation, and transcriptional suppression of a large number of genes (Lister *et al.*, 2009; Olkhov-Mitsel and Bapat, 2012). In somatic cells, DNA methylation occurs at cytosine in any context of the genome but predominantly in a cytosine-phosphate-guanine (CpG) dinucleotide context (Jin *et al.*, 2011). Methylated CpGs augment transcription repression by a number of processes, including the direct blockage of transcription initiation complexes from binding to DNA promoter regions and recruitment of transcriptional repressor complexes, including methyl CpG binding proteins (MBPs) that bind at methylated DNA sequence (Sasai *et al.*, 2010). Aberrant methylation levels have been postulated to inactivate tumor suppressors and activate oncogenes, which lead to carcinogenesis (Gal-Yam *et al.*, 2008).

In mammals, methylation occurs predominantly at the CpG dinucleotides, which are extremely depleted in the genome except at a short stretch genomic region termed as CpG islands, which are usually located at gene promoters (Deaton and Bird, 2011). Roughly about 50% of mammalian gene promoters are associated with one or more CpG islands, making this the most common promoter type in the vertebrate genome (Ioshikhes and Zhang, 2000). While the CpG dinucleotides in the genome are heavily methylated, the CpG dinucleotides in these islands remain unmethylated. Inactivation of numerous numbers of genes has been associated with the increased CpG island methylation in tumors such as *hMTLH1*, *p16*, *MGMT*, *BRCA1*, and *CCDN2* (Lian *et al.*, 2012). Hence, methylation of CpG islands is an important mechanism for gene inactivation in the prevention of tumor growth and development.

1.2 Rationale of the study

Despite the importance of CK in PC biosynthesis, embryogenesis, muscular dystrophy and tumorigenesis, literature describing transcriptional regulation of *ckα* gene is still lacking. Higher level of *ckα* is a common feature in many types of cancer. Over the years, enormous efforts have been focused on investigating the expression of CK in different cancer cells which led to the use of CK inhibitors as potential anticancer agents (Trousil *et al.*, 2016; Zimmerman and Ibrahim, 2017; Khalifa *et al.*, 2020). Unfortunately, less attention has been given to the intracellular regulation of choline kinase gene expression including by epigenetic mechanism. DNA methylation of CpG islands especially on the promoter of a gene is one of the mechanisms that regulate the gene expression at transcriptional level.

Analysis of 5' flanking region of *ckα* gene showed that it possesses characteristics of a housekeeping gene which are: absence of TATA box in close proximity to the transcription start site and containing several proximal CCAAT boxes as well as Sp1 binding sites (Aoyama *et al.*, 2004). The TATA-less and high GC-rich sequence promoters are typically characterized as CpG island promoter, generally associated with DNA methylation. The presence of numerous Sp1 binding sites indicates that the *ckα* promoter contains high GC contents which led to the assumptions that transcriptional regulation of *ckα* gene might be controlled through DNA methylation at the promoter region. Based on the presence of several CpG islands on the promoter region, we hypothesize that the levels of DNA methylation in the *ckα* promoter could be affected by epigenetic drugs such as 5-azacytidine, a demethylating agent and budesonide, a methylating agent. MCF-7 cell line was used for the analysis of DNA methylation as it

showed the highest promoter activity compared to the other cell lines and could activate the transcription of firefly luciferase for promoter study (Kuan *et al.*, 2014).

DNA methylation is suggested to modulate the binding of transcription factors to DNA (Héberlé and Bardet, 2019). Our previous studies have identified important transcription factor binding sites in the promoter region of *ckα* gene. Hence, this study aimed to investigate the correlation between DNA methylation of CpG island and transcription factor binding based on the overlaps between methylation sites and transcription factor binding motifs. From this study, the involvement of CpG island and DNA methylation in the transcriptional control of *ckα* gene would be elucidated.

1.3 Objectives of the study

1.3.1 General objective

To study the effect of DNA methylation on *ckα* CpG islands promoter activity.

1.3.2 Specific objectives

1. To identify putative CpG islands of human *ckα* promoter by *in silico* analysis.
2. To determine the level of methylation on the methylation-prone CpG island of *ckα* promoter.
3. To identify important CpG islands that regulate the activities of *ckα* promoter by site-directed mutagenesis.
4. To investigate the effects of 5-azacytidine and budesonide on *ckα* promoter activity in MCF-7 cell lines.

5. To confirm the binding of transcription factors on methylation-prone CpG island *ckα* promoter using EMSA.

CHAPTER 2

LITERATURE REVIEW

2.1 Phospholipids

Back in 1915, membranes isolated from red blood cells were found to be composed of lipids and proteins (Campbell and Reece, 2005). Lipid constitutes approximately 50% of most animal cell membranes in which phospholipids are the most abundant membrane lipids. A phospholipid molecule consists of a polar head group and two fatty acids tails in which one tail contains one or more *cis*-double bonds (unsaturated) which create a small kink in the tail, while the other tail does not (Alberts *et al.*, 2002). A glycerol molecule is attached to one end of two fatty acids and to the other end of a phosphate group linked to an organic compound such as choline (Figure 2.1). The fatty acids tails are hydrophobic and not soluble in water whereas the hydrophilic polar head group is ionized and readily water soluble to enable interaction with the environment (Solomon *et al.*, 2004). Due to its amphiphilic properties, phospholipids are spontaneously arranged in lipid bilayers in aqueous solution and aggregated into membranous structures (Alberts *et al.*, 2002). These fundamental components make them uniquely suited to form membranes of living cells (Marinetti, 1990).

Phospholipids are categorized into two major classes namely glycerophospholipids and sphingolipids based on their alcohol structure. Glycerophospholipids and sphingolipids contain glycerol and sphingosine respectively as the alcohol group (Newsholme and Leech, 2011). These phospholipid constituents play specific roles in the physiological

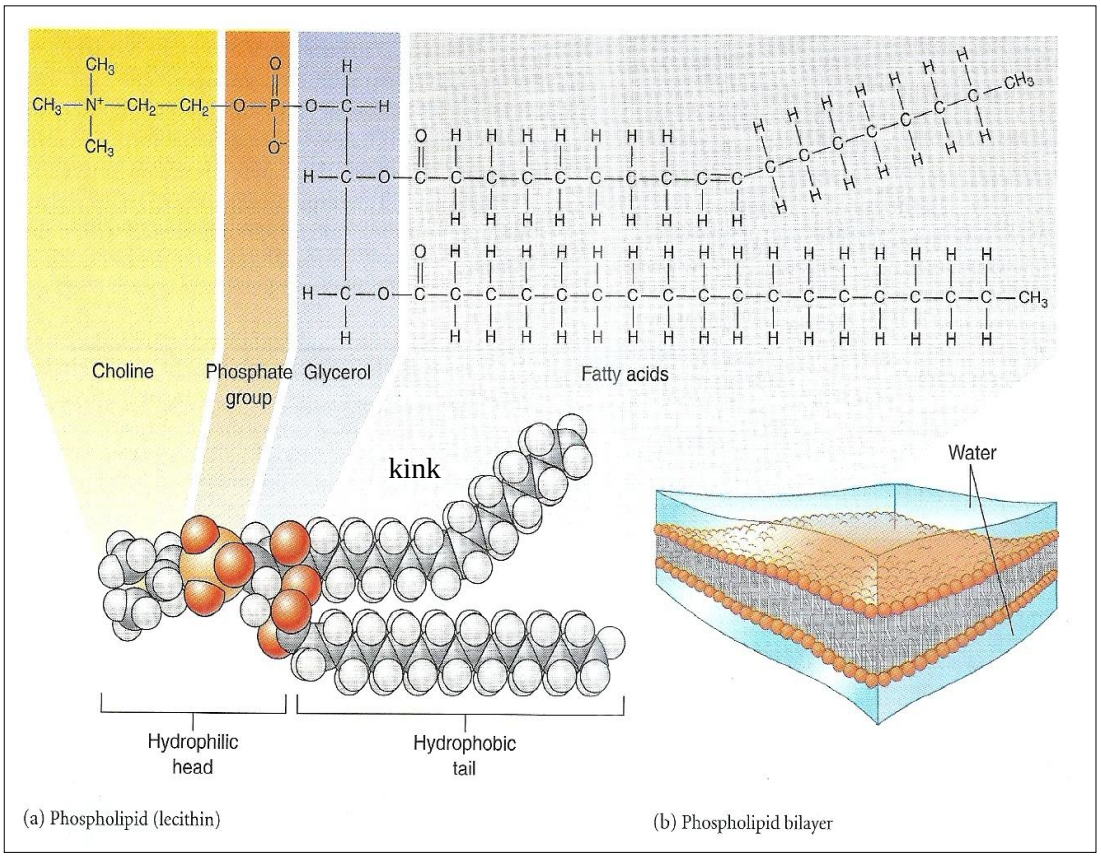

Figure 2.1 Structure of phospholipid and a phospholipid bilayer. a) A phospholipid consists of a hydrophobic tail made up of two fatty acids and a hydrophilic head consists of a glycerol bonded to a phosphate group, which in turn bonded to an organic group, choline. The fatty acid at the top contains one double bond that produces a kink in the chain. b) Phospholipids form lipid bilayers where the hydrophilic head interacts with water whereas the hydrophobic tails are arranged in bilayers. Adapted from Solomon *et al*. (2004).

functions depending on their chemical structure. Glycerophospholipids mainly act as structural components of cell membranes while sphingolipids are often used as part of a signaling cascade (Lim and Kwan, 2018).

2.1.1 Phosphatidylcholine

Phosphatidylcholine (PC) is the major glycerophospholipid, accounting for 40-50% of total phospholipids in all eukaryotic membranes. This is followed by phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidylserine (PS), sphingomyelin, cardiolipin and its phosphorylated derivatives which are also predominant in plasma membrane (Table 2.1) (Vance, 2015). PC plays a vital role in maintaining the cells and is found in all the subcellular components of the nervous system (Ansell, 1972). A study by Chakravarthy *et al*. (2009) discovered an isoform of PC, known as 1-palmitoyl-2-oleoyl-*sn*-glycerol-3-phosphocholine (POPC) that serves as endogenous ligand for Peroxisome Proliferator-Activated Receptor (PPARs) in hepatocytes. PPAR plays regulatory roles in gene expression and has been used as drug target to treat human disorders of lipid metabolism. PPARα-dependent gene expression is reduced with inactivation of fatty acid synthase (FAS) in the hypothalamus, which is required for the presence of POPC. However, injection of POPC into the hepatic veins of mice for several days induced PPARα-dependent gene expression and decreased hepatic steatosis. These data suggest that POPC is able to influence gene expression and acts as signaling molecule in mammals (Chakravarthy *et al.*, 2009).

PC also plays a distinct role in insulin transduction (Furse and De Kroon, 2015). Phosphatidylcholine transfer protein (PC-TP) is a phospholipid-binding protein that

catalyzes the intermembrane exchange of phosphatidylcholine *in vitro* (Wirtz, 1991; Kang *et al.*, 2010). Elsoy and colleagues (2013) reported that PC-TP inhibits Insulin Receptor Substrate 2 (IRS2), which is an effector of insulin signaling that is impaired in diabetes, suggesting the functional role of PC-TP as a sensor of membrane phosphatidylcolines (Ersoy *et al.*, 2013).

2.2 CDP-choline pathway

PC biosynthesis in all mammalian cells is synthesized mainly via the CDP-choline pathway, also known as Kennedy pathway (McMaster, 2018). This pathway consists of three steps: the first reaction of choline phosphorylation to form phosphocholine (PCho) is catalyzed by choline kinase (CK) using ATP and Mg^{2+} as cofactor. This is followed by the formation of CDP-choline from PCho which is catalyzed by cytidyltransferase (CCT), and final condensation of CDP-choline with a lipid anchor, diacylglycerol (DAG) to PC catalyzed by cholinephosphotransferase (CPT) (Figure 2.2) (Aoyama *et al.*, 2004). During the biosynthesis of PC, the conversion of choline into PC accounts for approximately 95% of the total choline embedded in most animal tissues, whereas the remaining 5% consists of free choline, phosphocholine, glycerophosphocholine, CDP-choline and acetylcholine (Li and Vance, 2008).

2.3 Choline kinase

The first step of PC biosynthesis involves choline kinase. Choline kinase (CK, ATP:choline phosphotransferase) was discovered in 1953 in Brewer's yeast by Wittenberg and Kornberg (Wittenberg and Kornberg, 1953). This cytosolic enzyme

Table 2.1 Lipid composition of a typical nucleated mammalian cell. Adapted from Vance (2015).

	Percentage of total lipids[a]
Phosphatidylcholine	45 – 55
Phosphatidylethanolamine	15 – 25
Phosphatidylinositol	10 - 15
Phosphatidylserine	5 – 10
Phosphatidic acid	1 – 2
Sphingomyelin	5 – 10
Cardiolipin	2 – 5
Phosphatidylglycerol	<1
Glycosphingolipids	2 – 5
Cholesterol	10 - 20

[a] Data are averaged from several sources

present in various tissues in which the enzymatic activity has been observed to occur in liver, brain, intestine and kidney of several species (Wittenberg and Kornberg, 1953). CK is the first enzyme in the CDP-choline pathway for the *de novo* biosynthesis of PC (Farine *et al.*, 2015) and changes in CK can influence the rate of PC synthesis (Gibellini and Smith, 2010). Until its purification in 1984, subsequent cloning and expression of cDNA of CK from yeasts, mammals and plants have been characterized which led to the description of the gene structure (Wu and Vance, 2010).

In mammalian cells, CK exists in three isoforms namely CKα1 (50 kDa, 435 amino acids), CKα2 (52 kDa, 453 amino acids) and CKβ (45 kDa, 394 amino acids) which are encoded by two separate genes that are *ckα* and *ckβ*, located on chromosomes 11q13.2 and 23q13.33, respectively (National Center for Biotechnology Information (NCBI). Available from: https://www.ncbi.nlm.nih.gov/). The CKα1 and CKα2 functional isoforms are the results of alternative splicing of CKα transcript which differ in an additional 54 bp extra internal nucleotide sequence, yielding 18 amino acids insertion starting at nucleotide 155 for CKα2. On the other hand, protein sequence of CKβ shares approximately 60% sequence identity with CKα1 and CKα2. CK isoform is active only in either homo or heterodimeric form but not in monomeric form in which α/α homodimer is the most active form, followed by α/β heterodimer and β/β homodimer which is the less active phenotype (Aoyama *et al.*, 2004; Arlauckas *et al.*, 2016).

2.3.1 Expression and regulation of CK activity

CKα and CKβ are both ubiquitously expressed in mammalian cells, yet the distribution of CK is reported to be tissue-specific (Aoyama *et al.*, 2002). The analysis of expression

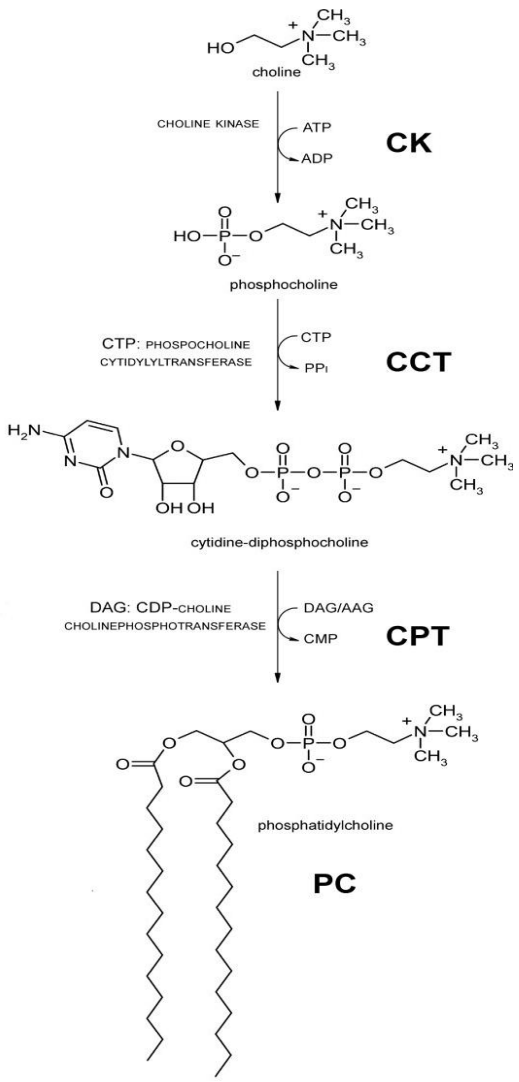

Figure 2.2 The CDP-choline pathway. CK, choline kinase; CCT, CTP:phosphocholine cytidyltransferase; CPT,cholinephosphotransferase; PC, phosphatidylcholine; DAG, diacylglycerol; CMP, cytidine monophosphate; PCho, phosphocholine; CDP-Cho, cytidinediphosphocholine. Adapted from Gibellini and Smith (2010).

and distribution of CK isoforms in mouse tissue using both Northern blot and Western blot analysis shows the expression of CKα isoform is the highest in the testis, whereas that of CKβ isoform is comparatively high in the heart and liver (Aoyama *et al.*, 2002; Arlauckas *et al.*, 2016). Further investigation was carried out to estimate each CK isoform activity in the mouse tissue by immunoprecipitation with each isoform-specific antiserum. They found out that the addition of anti-CKα and anti-CKβ antisera mixture in mouse tissue cytosols resulted in complete inhibition of CK activity (Aoyama *et al.*, 2002). This finding indicates that each CK isoform plays a distinct function in the expression of mammalian cells.

In addition to its involvement in the biosynthesis of PC, CK also has other functions in regulating the cell signaling pathway. Downregulation of *ckα* expression with small interfering RNA (siRNA) silencing decreased the phosphatidylcholine, phosphatidic acid and signaling through the MAPK and P13/AKT pathway, which has been associated with cell proliferation (Yalcin *et al.*, 2010). In another study, a group of researchers discovered that CKα forms a complex with EGFR in a c-Src dependent manner in which overexpression of EGFR and c-Src ultimately increases the total cellular activity and protein levels of CKα (Miyake and Parsons, 2012). EGFR and c-Src has been shown to have a synergistic effect in the tumorigenesis of breast as well as other cancers. Mutations of *ckα* introduced at Y197 and Y333 resulted in reduced complex formation, EGFR-dependent activation of CKα enzyme activity and EGF-dependent cell proliferation (Miyake and Parsons, 2012).

Several studies discovered the regulation of CK activity at the transcriptional level (Uchida, 1994; Aoyama *et al.*, 2000; Glunde *et al.*, 2008). Characterization of human putative promoter region of *ckα* gene (-2.3 kb region upstream of translation start site) shows that hypoxic environment regulates the expression of CKα and consequently increasing cellular PC and total choline levels (Glunde *et al.*, 2008). The binding of hypoxia-inducible factor (HIF-α) on the HRE sites was shown to suppress *ckα* mRNA levels in a human prostate cancer model as shown through chromatin immunoprecipitation assay (Glunde *et al.*, 2008).

2.3.2 CK and carcinogenesis

Cancer is characterized by uncontrolled cell growth due to uncontrolled proliferation and decreased apoptosis which is capable of invading adjacent tissues and organs. It is postulated that cancer is derived from the accumulation of mutated genes including tumor suppressor genes, oncogenes as well as invasion/metastasis related genes, where certain mutation may lead to development of malignant changes in their enzymatic activities (Han *et al.*, 2019). Aberrant lipid metabolism has been observed in many types of cancer in which as tumor cells and tumor progresses, phospholipid biosynthesis become greater than in normal tissue (Szachowicz-Petelska *et al.*, 2013; Sola-Leyva *et al.*, 2019). Elevated activities of CK and its product, PC has been implicated in carcinogenesis as demonstrated by a large number of magnetic resonance spectroscopy (MRS) studies in cancer cells and solid tumors (Negendank, 1992; Nakagami *et al.*, 1999; Ronen and Leach, 2000). This elevation has been observed in most cancer types and can be targeted as an endogenous biomarker of cancer (Ackerstaff *et al.*, 2003).

Overexpression of *ckα* gene has been reported in a number of human tumor-derived cell lines and in biopsy samples of colon, lung, ovarian and prostate carcinomas when compared with normal tissue (de Molina *et al.*, 2007; Granata *et al.*, 2014; Bagnoli *et al.*, 2016). This indicates that *ckα* is crucial in PC biosynthesis and is required to control the development of cancer cells (Glunde *et al.*, 2011). In contrast, there was no evidence to implicate *ckβ* in carcinogenesis as no changes of *ckβ* expression was detected in breast, lung and ovarian cancer cell lines (Eliyahu *et al.*, 2007; Gallego-Ortega *et al.*, 2009).

An increased activity of *ckα* was shown in human breast cancers where a significant increase of *ckα* activity was observed in approximately 38.5% tumor samples compared to the corresponding normal tissue (de Molina *et al.*, 2002; Rizzo *et al.*, 2021). Ovverexpression and increased activity of *ckα* correlated with histological tumor grade suggesting that *ckα* dysregulation might be associated with prognosis and malignancy of the disease. However, no significant correlation was observed with age, tumor size or progesterone receptor status in these studied breast tumors. These findings suggest that *ckα* activity is directly associated with increased breast cancer proliferation making it a potential marker for breast prognosis (de Molina *et al.*, 2002).

The involvement of *ckα* in carcinogenesis suggests that *ckα* inhibition could be an effective cancer therapy. Early discovery of CK inhibitors includes the study of choline phosphorylation in the presence of the thiol group inhibitors that leads to CK inhibition by N-ethylmaleimide (Arlauckas *et al.*, 2016). A preliminary study targeting the inhibition of choline kinase using purinyl-6-histamine (PH), which is selectively cytotoxic against tumor cells demonstrated the inhibition of choline phosphorylation,

reflecting its anti-tumor activity (Mayer and Werner, 1974). Hemicholinium-3 (HC-3), a well-known CK inhibitor is shown to reduce PC levels and reduce the growth factor-induced DNA synthesis *in vitro* (Arlauckas *et al.*, 2016). Glunde *et al*. (2005) reported a molecular approach by RNA interference (RNAi) to inhibit the expression of specific targeted genes in mammalian cells. RNAi knockdown of CK reduced proliferation and promoted differentiation of breast cancer cells as detected by MRS (Glunde *et al.*, 2005). Specific inhibition of *ckα* selectively induces apoptosis in several cancer cell lines while the normal cell is not affected (Bañez-Coronel *et al.*, 2008).

2.4 Promoter and transcriptional regulation of gene expression

The expression of a gene is regulated at different stages from transcription initiation to post-translational modification of protein. However, the key factor for proper functioning of regulatory elements occur at the level of transcription initiation, particularly gene promoter which is crucial for coordinated transcription within a cell (De Vooght *et al.*, 2009). Till date, the structure of regulatory DNA sequences remains poorly understood. With a variety of DNA regulatory elements present within promoter region, the identification and characterization of these elements are crucial for the understanding of the human gene regulation.

Promoters are stretches of genomic sequence typically located upstream of a gene. Core promoter is a promoter region typically 60-120 bp, surrounding the transcription start site (TSS) that recruits a complex of general transcription factors for the initiation of transcription (Haberle *et al.*, 2014). This minimal promoter region is sufficient to direct the accurate initiation of transcription. Sequence motifs commonly found within the core

promoter region includes TFIIB recognition element (BRE), initiator (Inr), TATA box and downstream core promoter element (DPE) (Butler and Kadonaga, 2002). Each of these motifs specifically involves in the initiation of transcription process, though these elements are not necessarily present in all core promoters. The core promoter provides a docking site for RNA Polymerase II transcriptional machinery in a tightly regulated manner for a proper level of gene expression (Kumar and Bansal, 2018). RNA Polymerase II requires specific core promoter element to initiate transcription through the assembly of transcription preinitiation complex (PIC). This process requires general transcription factors (GTFs) that recognize and bind core promoter motifs and subsequently direct RNA Polymerase II to the TSS and starts the transcription of a gene. The common GTFs bind to the core promoter in the following order: TFIID, TFIIB, RNA Polymerase II-TFIIF complex, TFIIE, followed by TFIIH (Héberlé and Bardet, 2019).

In addition to basal transcriptional regulation of core promoter, transcriptional activity is greatly stimulated by a concerted action of other elements including proximal promoter elements such as enhancers, silencers and insulators (Figure 2.3) (Butler and Kadonaga, 2002; Hernandez-Garcia and Finer, 2014). Proximal promoter elements such as CAAT box, *cis*-regulatory module (CRM) and GC box which are located immediate upstream of core promoter, contain recognition sites for specific consensus elements that involved in transcriptional regulation (Kumar and Bansal, 2018). Proximal promoter elements which are present in the distal promoter region are mainly act as connecting element for enhancers, silencers and insulators.

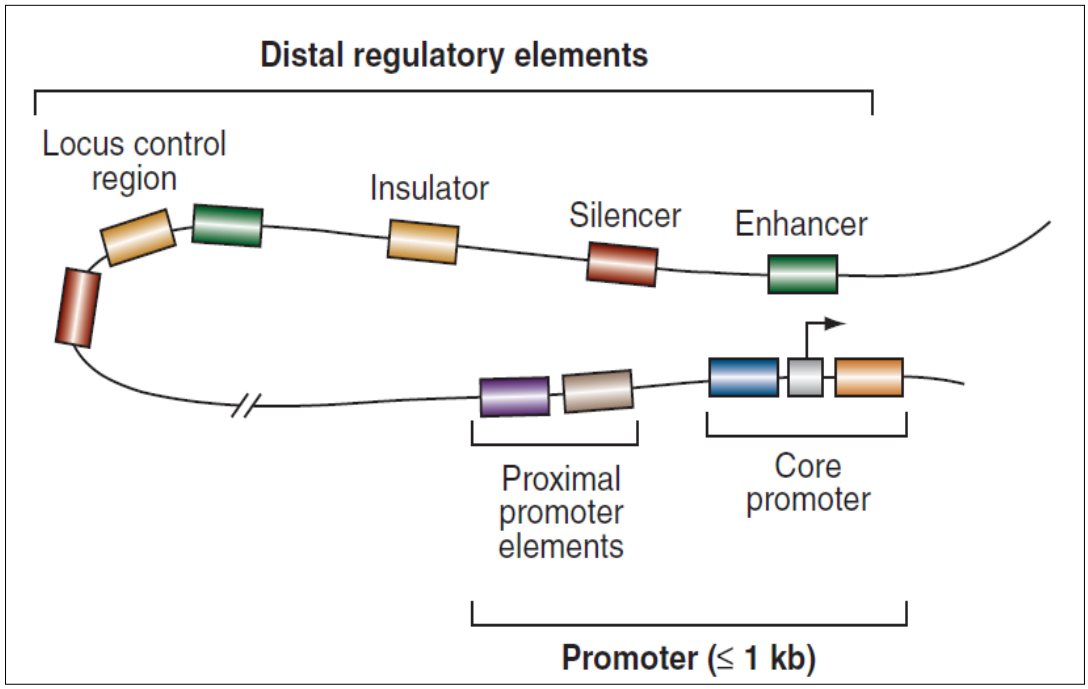

Figure 2.3 Schematic structure of a gene promoter region. The promoter composed of core promoter and proximal promoter elements typically span less than 1 kb pairs. Distal promoter elements located upstream of the promoter includes enhancers, silencers and insulators. These distal elements may contact the core promoter or proximal promoter by looping out the intervening DNA. Adapted from Maston *et al*. (2006).

Cis-regulatory elements are regions of non-coding DNA which regulate the transcription of neighboring genes, whereas trans-regulatory elements regulate the expression of distant genes. Transcription initiation is a strictly controlled process that involves both *cis*-acting and *trans*-acting factors (Das and Singal, 2004). The presence of both positive and negative regulatory elements within the promoter provides regulatory control of a unique gene expression pattern (Maston *et al.*, 2006).

The upstream *trans*-acting DNA binding transcription factors such as activators and coactivators, interact with the regulatory element within core promoter, proximal promoter elements and distal promoter to enhance the efficiency of transcription initiation. On the other hand, transcription can be inhibited by *trans*-acting repressors which directly or indirectly bind to DNA binding motif and negatively regulate gene transcription. A study using full-length cDNA sequence for the identification of TSS in the transcriptional human promoters revealed that putative negative regulatory elements were located at -1000 to -500 bp upstream of the TSS for 55% genes tested (Cooper *et al.*, 2006).

Activators or repressors regulate gene transcription mostly through coregulators, even though they can bind directly with PIC complex associated with core promoter (Fuda *et al.*, 2009). These processes are important in a mediation of precise controlled patterns of gene expression (Maston *et al.*, 2006). A study of the 5' flanking sequence of mouse *ckα* gene by the promoter-reporter assay reveal the presence of two putative promoter regions which are proximal and distal promoter. Various Sp-1 consensus sequences are identified within the proximal region indicating the criteria of housekeeping gene for *ckα*

gene. Meanwhile, distal promoter consists of responsive elements such as XRE and AP-1 boxes which demonstrated a high expression of *ckα*. AP-1 binding element responds to carbon tetrachloride (CCl_4) which resulted in increased expression level of *ckα* mRNA and CK activity in murine liver. Deletion of 9 base pair (bp) sequence corresponding to AP-1 binding element resulted in the loss of promoter activity whereas the duplication insertion of this 9 bp element caused an increase in promoter activity. These results indicated that *ckα* gene expression is positively regulated by AP-1 or together with other transcription factors that could be involved in the promoter activity (Aoyama *et al.*, 2004). In contrast, no distal promoter sequence has been found in 5' flanking region of *ckβ* gene indicating the absence of any responsive elements in its regulatory region (Figure 2.4) (Aoyama *et al.*, 2004).

2.5 Epigenetics

Epigenetics is a study of heritable changes in gene expression that occur without any changes in DNA sequence (Bird, 2007). The term epigenetics was first coined by Conrad Waddington in 1942 to describe the influence of internal and external interactions between genes and the microenvironment towards the development of phenotype (Goldberg *et al.*, 2007). Epigenetic modifications are required for normal development and are involved in a variety of cellular differentiation, morphogenesis and variability of an organism. This process influences gene activity at the transcriptional and post-trasncriptional level as well as at the translational and post-translational protein level (Halušková, 2010). Dysregulated epigenetics processes have been found to be involved in various diseases, particularly cancers, immune disorders and mental retardation associated disorders.

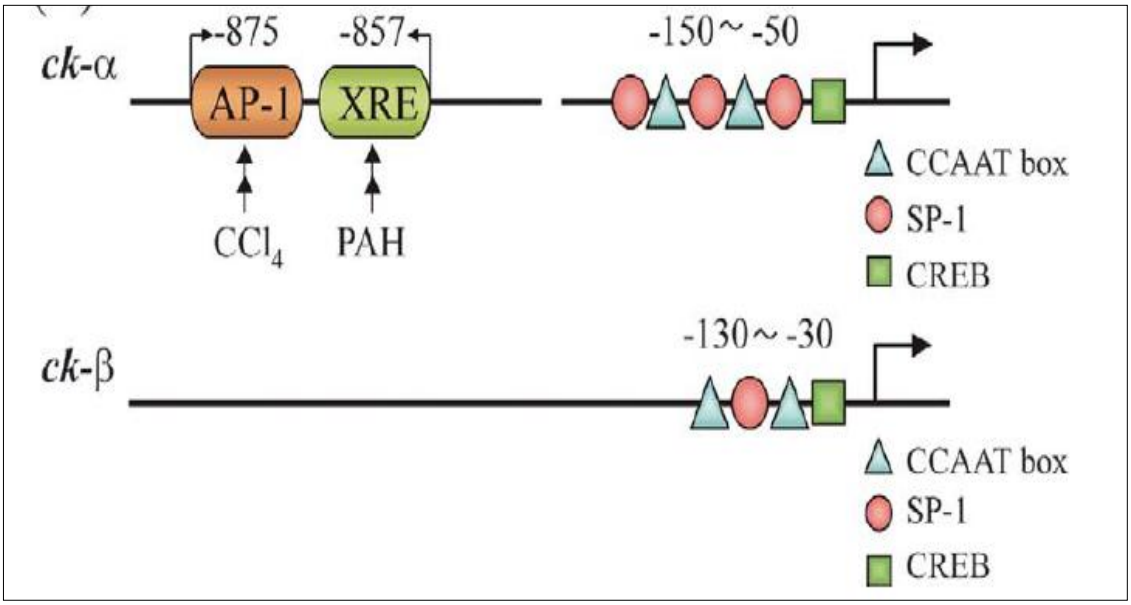

Figure 2.4 Schematic structure of murine *ckα* and *ckβ* promoters. The predicted contribution of AP-1 and XRE sites of *ckα* gene in CCl_4 and PAH-induced in mouse liver. Adapted from Aoyama *et al*. (2004).

Back in 1983, cancer was the first human diseases to be linked to epigenetics (Feinberg and Vogelstein, 1983). Cytosine methylation of *hMLH1* promoter was reported in four colorectal tumor cell lines but absent in adjacent normal tissue that expressed *hMLH1* which results in silencing of the gene encoding MLH1 (Kane *et al.*, 1997). The most characterized epigenetic modifications include DNA methylation, chromatin remodeling, modifications of histones, non-coding RNA mechanisms and positioning of nucleosome along the DNA (Kulis and Esteller, 2010). These epigenetic signals work synergistically to ensure proper transcriptional activity and repression by chromatin-modifying activity.

2.6 DNA methylation

DNA methylation is the most common epigenetic modifications in vertebrates and is originally proposed as a silencing epigenetic mark in 1970s (Holliday and Pugh, 1975; Riggs, 1975). In mammals, DNA methylation occurs exclusively at cytosine residues that precede a guanine nucleotide or known as CpG sites (Feltus *et al.*, 2003). The 'p' indicates cytosine (C) and guanine (G) are connected by a phosphodiester bond. Approximately 5 x 10^7 of total cytosines are methylated per diploid nucleus. Although all methylated cytosines are present within CpG dinucleotides, only 70-80% of these potentially methylated sites are actually in a methylated form (Antequera and Bird, 1993).

DNA methylation involves the covalent addition of a methyl group (CH_3) at the 5-carbon of the cytosine ring which results in the conversion of cytosine to 5-methylcytosine (5-mC). The methyl groups protrude into the major groove of DNA and

provides molecular interactions within major grooves of DNA double helix (Fatemi and Wade, 2006). The modified cytosine was first discovered during the separation of DNA nucleosides by paper chromatography (Hotchkiss, 1948). However, it was not until two decades later that DNA methylation was demonstrated to be involved in cellular differentiation and regulation of gene expression at the transcriptional level (Holliday and Pugh, 1975; Compere and Palmiter, 1981).

DNA methylation patterns are established during early embryonic development and stably maintained throughout an individual's life. Several hours after conception, sperm DNA is exposed to methylation in the single-celled embryo. The cells begin to differentiate into various tissue types as the embryo started to develop and divide, gradually establishing the methylation pattern. However, an active demethylation occurs mostly in paternal genomes during the early steps of embryo development immediately after fertilization and in pre implantation embryos (Geiman and Robertson, 2002). This process is followed by the establishment of global *de novo* methylation patterns following implantation (Almouzni and Cedar, 2016) that will be maintained predominantly in somatic tissues (Chen and Riggs, 2011).

2.6.1 DNA methyltransferases family

DNA methylation is regulated by a group of DNA methyltransferase (DNMT) protein family; DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L (Espada and Esteller, 2007; Cheng and Blumenthal, 2008). These enzymes work synergistically for the establishment, recognition and removal of DNA methylation throughout the genome (Moore *et al.*, 2013). DNMT3A and DNMT3B are *de novo* methyltransferases that are

highly expressed in developing embryo, responsible for the establishment of DNA methylation profile during embryonic stage (Heerboth *et al.*, 2014). On the other hand, DNMT1 is a maintenance methyltransferase that is found abundantly in somatic cells, and has 30 to 40~folds preference to methylate hemimethylated DNA and maintaining methylation pattern from the parental to the daughter strand during DNA replication (Jeltsch, 2006; Espada and Esteller, 2007). A strong preferential binding to hemimethylated CG sites is shown by a multidomain protein UHRF1 as it interacts and colocalizes with DNMT1 for stable association of DNMT1 to chromatin. This particular protein contains a methyl DNA binding domain, SRA (SET and RING associated) domain which involved in the recruitment of DNMT1 to hemimethylated DNA in order to facilitate efficient maintenance of DNA methylation (Bostick *et al.*, 2007). In some cases, *de novo* methyltransferases, DNMT3A and DNMT3B act as maintenance of DNA methylation patterns by methylating the hemimethylated CG dinucleotides (Chen and Riggs, 2011).

Unlike the aforementioned DNMT family members, another member of DNMT3 family, DNMT3L lacks conserved motif and is catalytically inactive. It has been postulated that DNMT3L functions as regulatory factors in germ cells by recruiting DNMT3A isoforms to nucleosome that contain unmethylated H3K4 to trigger *de novo* DNA methylation (Chen and Riggs, 2011). Owing to its role as the only DNA methyltransferase family that is expressed in germ cells, DNMT3L is crucial for the establishment of methylation patterns in both male and female germ cells (Bourc'his *et al.*, 2001).