INTRODUCTION

Chapter I

Epigenetics Overview

Essentials of Molecular Biology

On the simplest level, genes can be defined as segments of deoxyribonucleic acid (DNA) that encode for proteins or ribonucleic acid (RNA) products with biological functions. DNA is a biological substance that carry genetic information and is a polymer of nucleotides. Genetic information is reproduced from parent to daughter cells during cell division through process of DNA replication. When genes are expressed (switched on), the DNA sequence is transcribed into RNA. RNA molecules are polymers of ribonucleotides that exist in a number of functional forms. RNA molecules that act as intermediates for protein production are termed messenger RNA (mRNA). RNA molecules that serve a direct biological function without coding for a protein are collectively termed non coding RNA (ncRNAs). [1] mRNA is in turn translated into a protein, which is a polymer of amino acids. Each amino acid is encoded by a triplet nucleotide code termed a codon. The human genetic code compromises 64 codons encoding for 21 amino acids and 3 stop codons. mRNA codons are read by anticodon regions of transfer RNA (t RNA) molecules, which are small RNAs that bring the corresponding amino acid to the growing polypeptide chain.The polypeptide chain is synthesized by ribosomes which are macromolecular complexes containing ribosomal RNA (r RNA). [2,3]

Most human cells contain two full copies of haploid human genome, which is organized and packaged into 23 pairs of chromosomes (diploid genome). A chromosome is highly ordered structure of a single DNA molecule with specialized structural features, namely,one centromere and two telomeres. Every individual inherits one copy of human genome from his or her father and another from the mother. Thus, the human genome contains two copies, termed alleles, of each autosomal gene. Although a gene sequence may encode for a specific protein or RNA with defined functions, alleles of genes may demonstrate sequence variations that in turn contribute to variations in functional characteristics of gene product between individuals. The primary nucleotide sequences of the two gene alleles form the genotype, whereas the expressed function or biological effect of gene product is termed the phenotype. Thus one could study human disease or trait at the genetic level through determination of the allelic sequence of a gene (genotyping) or at the functional level (phenotyping) for example; ABO blood group, electrophoretic mobility of hemoglobin variants or investigation of enzyme concentration. [1]

Definition of Epigenetics

Historically, the word “epigenetics” was used to describe events that could not be explained by genetic principles. Conrad Waddington (1905–1975), who is given credit for coining the term, defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being”. [4]

Epigenetics, in a broad sense, is a bridge between genotype and phenotype, a phenomenon that changes the final outcome of a locus or chromosome without changing the underlying DNA sequence. For example, even though the vast majority of cells in a multicellular organism share an identical genotype, organismal development generates a diversity of cell types with disparate, yet stable, profiles of gene expression and distinct cellular functions. [5] Thus, cellular differentiation may be considered an epigenetic phenomenon. Epigenetics refers to control mechanisms that are layered onto the primary nucleotide sequence to produce on/off states of gene expression [4]

Also, the ability of epigenetic marks to persist during development and potentially be transmitted to offspring may be necessary for generating the large range of different phenotypes that arise from the same genotype. For instance, cloned animals generated from the same donor DNA are not identical and they develop diseases with different penetrance from, their donor. Human clones that arise spontaneously; monozygotic twins are identical at the DNA sequence level, but have different DNA methylation and histone modification profiles that might affect the penetrance of several diseases, this phenomenon is also observed at a single cell level, for example stem cells can develop into any type of cell. [6, 7]

Epigenetics is a somatically heritable status of gene expression resulting from changes in chromatin structure without alterations in the DNA sequence, [8] Over the past decades, epigenetic studies have been focused mainly on embryonic development, aging, and cancer. Presently, epigenetics is highlighted in many other fields, such as inflammation, obesity, insulin resistance, type 2 diabetes mellitus, cardiovascular diseases, neurodegenerative diseases, and immune diseases. In addition, epigenetics is exceptionally important in the nutritional field, because nutrients and bioactive food components can modify epigenetic phenomena and alter the expression of genes at the transcriptional level.[9]

Increasing interest in epigenetics has been accompanied by technological breakthroughs that now make it possible to undertake large-scale epigenomic studies. These allow the mapping of epigenetic marks. In parallel, the role of epigenetics in editing primary genomic sequence and its impact on our understanding of human health and disease have been rapidly emerging.[10]

Epigenetic modifications can be grouped into four main categories:histone modifications, chromatin remodeling ,non coding RNA and DNA methylation, .[11]

1- Histone modification

Histones are key players in epigenetics together with the DNA, form the chromatin. Chromatin is the complex of chromosomal DNA associated with proteins in the nucleus. DNA in chromatin is packaged around histone proteins, in units referred to as nucleosomes. A nucleosome has 147 base pairs of DNA associated with an octomeric core of histone proteins, which consists of 2 H3-H4 histone dimers surrounded by 2 H2A-H2B dimers. N-terminal histone tails protrude from nucleosomes into the nuclear lumen. Histone H1 is called the linker histone. It does not form part of the nucleosome but binds to the linker DNA (that is, the DNA separating two histone complexes), sealing off the nucleosome at the location where DNA enters and leaves.[12]

Histones can be modified by methylation, acetylation, phosphorylation, biotinylation, ubiquitination, sumoylation, and ADP-ribosylation. The location of the histone

modifications is at the histone tails that consist of 15–38 amino acids. Lysine residues in the histone tails can be either methylated (mono-, di-, and tri-) or acetylated, and arginine residues can be mono- or di-methylated. [13]

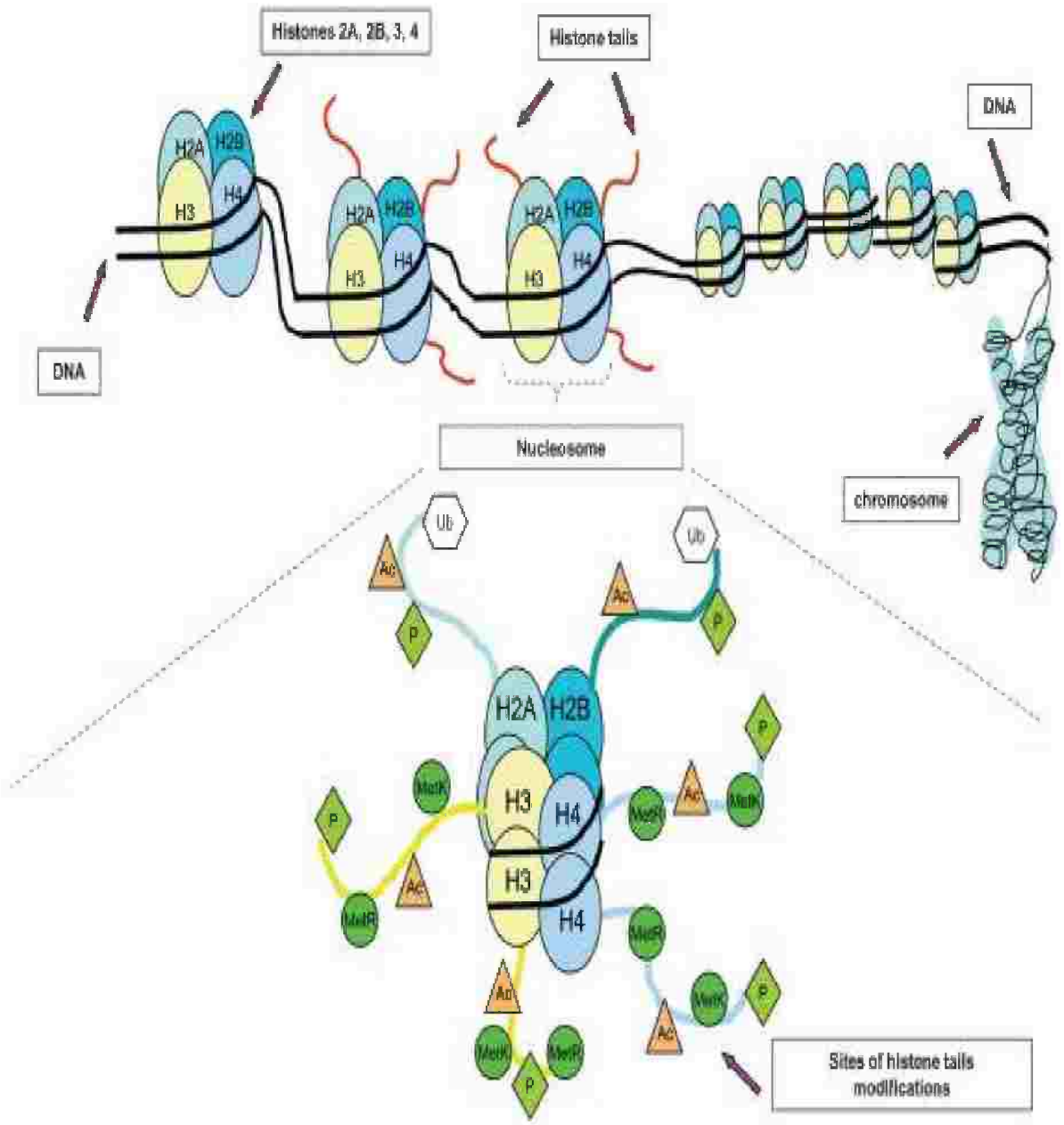

Figure (1): Nucleosome and histone modifications. [9]
Each nucleosome comprises an octamer of histone molecules and double-stranded DNA. The amino termini of histones, which are called histone tails, can be post-translationally modified and function as signal integration platforms. The main epigenetic modifications at histone tail sites are: lysine and arginine methylation (MetK and MetR respectively), phosphorylation (P), acetylation (Ac), and ubiquitination (Ub).

Histones can be modified at different sites simultaneously. The core histones forming the nucleosome can each have several modifications. Thus, a single histone mark does not determine the outcome alone; instead, it is the combination of all marks in a nucleosome or

region that specifies the outcome. Many enzymes that catalyze covalent post-translational modifications have been described. Because the modifications are dynamic, enzymes to remove these post-translational modifications have also been reported. However, the list of histone modifications, its writers and erasers, might not yet be completed. Of the enzymes that modify histones, methyltransferases, histone demethylases and kinases are the most specific to individual histone subunits and residues. Conversely, most of the histone acetyltransferases (HATs) and histone deacetylases (HDACs) are not highly specific and modify more than one residue.[14]

Surprisingly, it has recently been reported that HDACs and HATs are both targeted to transcribed regions of active genes by phosphorylated RNA polymerase II. Thus, most HDACs in the human genome function to reset chromatin by removing acetylation at active genes, whereas HATs, by contrast, are mainly linked to transcriptional activation.[12]

Depending on histone modifications, these N-terminal histone tails are recognized by other chromatin proteins that activate or repress transcription, and therefore certain histone modifications can establish, and potentially maintain active or silent epigenetic states.[15]

2- Chromatin remodeling:

Noncovalent mechanisms such as chromatin remodeling and the incorporation of specialized histone variants provide the cell with additional tools for introducing variation into the chromatin template. ATP-dependent chromatin remodeling complexes are thought to modify chromatin accessibility by altering histone- DNA interactions, perhaps by sliding or ejecting nucleosomes. [16,17] Histone variants regulate nucleosome positioning and gene expression. For example, the incorporation of the histone variant H2A.Z protects genes against DNA methylation. Thus, the interplay among different epigenetic partners becomes evident. The nucleosome remodeling machinery is influenced by DNA methylation and has been linked with specific histone modifications. [18,19]

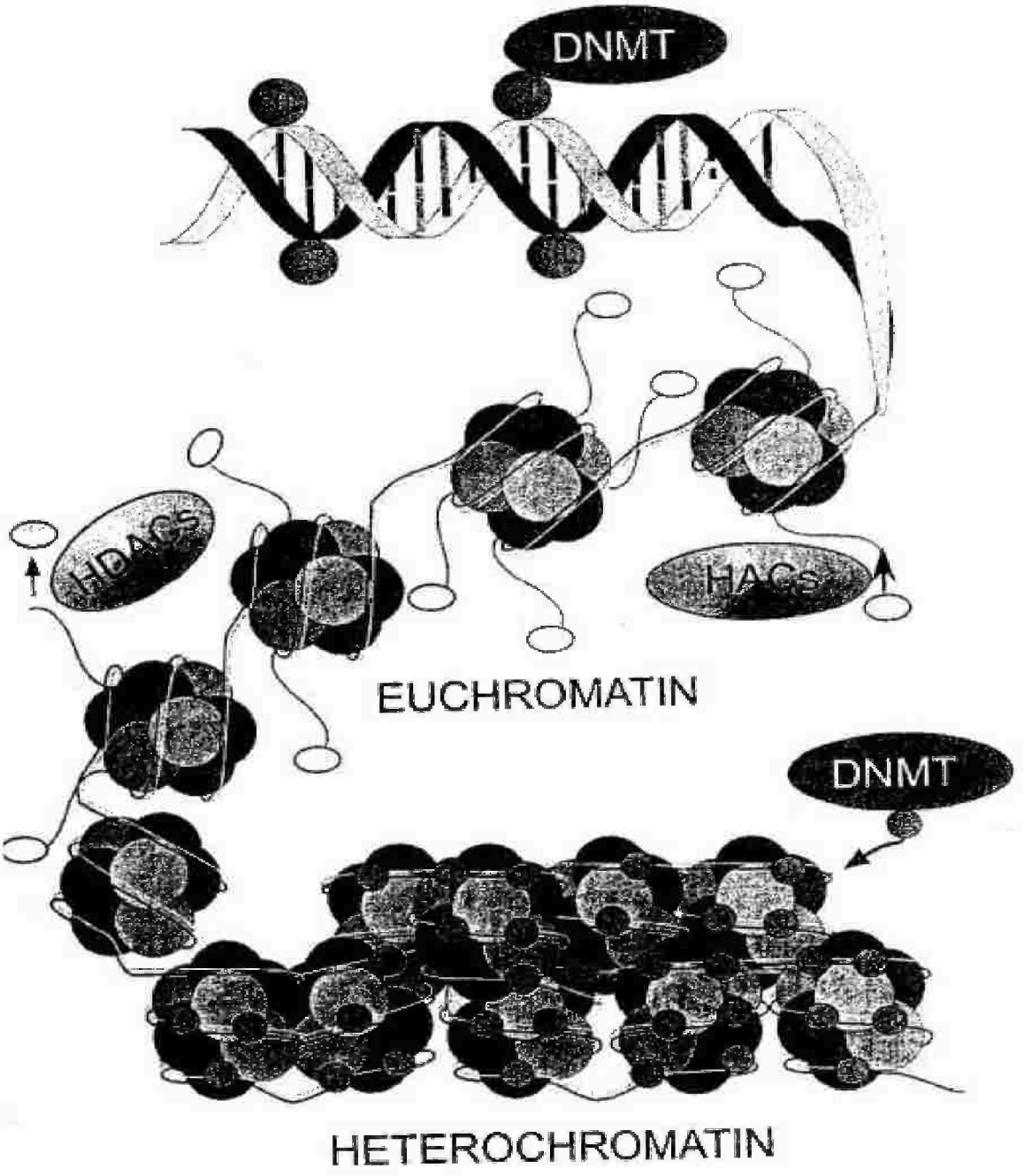

Figure (2): **Structure of DNA & Chromatin:** [20]Chromatin regulation involves high-order conformational changes-relaxation or tightening of this thread of DNA-histone complexes-such that the transcription of some regions is enhanced and that of others diminished or silenced. Condensed states of chromatin (heterochromatin) inhibit transcription, while relaxed states (euchromatin) are conducive to transcription. DNMT: DNA methyltransferase, HACs: histone acetylase, HDACs: histone deacetylase.

3- Non coding RNA

Non-coding RNAs are important modulators of chromatin regulation and gene expression. Whole genome and transcriptome sequencing demonstrated that at least 90% of the genome is actively transcribed, although less than 2% represent protein-coding genes. Thus, the non-coding part of the transcriptome became a new focus in gene expression and regulation. Currently, two major groups of non-coding RNA players can be distinguished: small non coding RNAs and long non coding RNAs. Small non coding RNAs are processed from longer precursors and comprise, in addition to transfer RNAs and ribosomal RNAs, the well-studied microRNAs, piwi interacting RNAs, small nuclear RNAs (snoRNAs) and other less well-characterized RNAs.[21]

MicroRNA is a new class of noncoding, endogenous, small RNA that regulates gene expression by translational repression, representing a new important class of regulatory molecules. MicroRNA can play important roles in controlling DNA methylation and histone modifications, creating a highly controlled feedback mechanism. Interestingly, epigenetic mechanisms such as promoter methylation or histone acetylation, can also modulate microRNA expression. A connection between epigenetic phenomena and microRNA has been described in several physiological processes and an altered balance between them represents one of the mechanisms leading to pathological conditions such as cancer. An aberrant expression of microRNA has been associated with the development or progression of human cancers by altering cell proliferation and apoptosis processes.[9, 22]

Long non coding RNAs are a heterogeneous class of m RNA like transcripts . They have regulatory roles in gene expression by assembling protein complexes and localizing them to their genomic target DNA sequence .[21]

4- DNA Methylation.

Chapter II

DNA Methylation

DNA methylation is perhaps the best characterized chemical modification of chromatin. Mammalian cells possess the capacity to epigenetically modify their genomes via the covalent addition of a methyl group to the 5 position of the cytosine ring within the context of the cytosine phospho guanine CpG dinucleotide (adjacent nucleotide parings are referred to in a NpN fashion, where N is the base and p is the phosphodiester bond; thus CpG is the designation given to cytosine phosphoguanine). [23]

It is becoming clear that CpG methylation does not act alone as a gene-silencing mechanism, and that many components of a multidimensional network can be involved in setting epigenetic states of gene expression in cancer cells. In some situations, DNA hypermethylation is a mark that appears late (e.g., after repressive histone modifications) but once established, it is a powerful signal to prevent mRNA transcription suggesting that DNA methylation and histone deacetylation may cooperate in silencing the expression of some genes in cancer cell lines. [24]

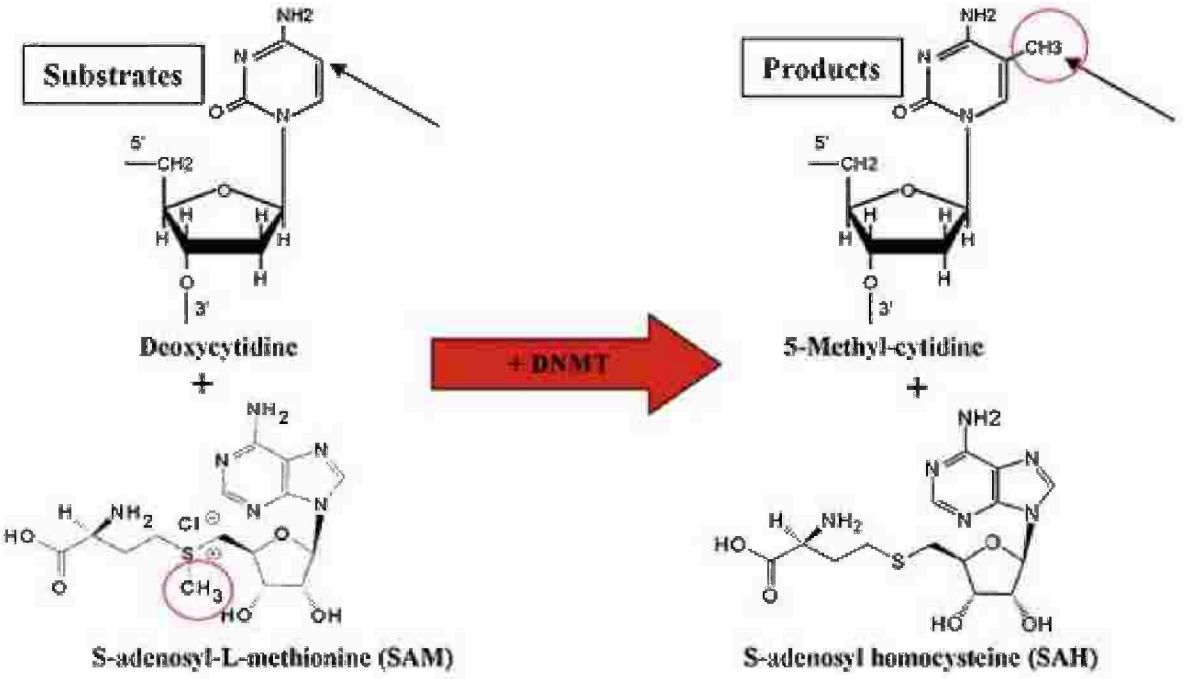

Figure (3): DNA methylation mediated by DNA methyltransferases (DNMTs).[20]

In human somatic cells, 5 methyl cytosine (m^5C) accounts for about 1% of total DNA bases and affects 70%–80% of all CpG dinucleotides in the genome. The most striking feature of vertebrate DNA methylation patterns is the presence of CpG islands, that is, unmethylated GC-rich regions that possess high relative densities of CpG and are positioned at the 5′ ends of many human genes. [25] Computational analysis of the human genome sequence predicts 29,000 CpG islands. Earlier studies estimated that approximately 60% of human genes are associated with CpG islands, of which the great majority is unmethylated at all stages of development and in all tissue types. A small but

significant proportion of all CpG islands become methylated during development, and when this happens the associated promoter is stably silent. [6,26]

In mammals, nearly all DNA methylation occurs on cytosine residues of CpG dinucleotides and DNA methylation of these islands correlates with transcriptional repression.[4,27] DNA methylation alters gene expression levels primarily through regulating methylation state dependent interactions with transcriptional activators or repressors, and chromatin remodeling enzymes. Several classes of methyl-DNA binding proteins, including factors such as methyl-CpG-binding protein 2 (MeCP2), methyl binding domain protein (MBD) bind to methylated DNA and repress transcription, either by directly disrupting the formation of the RNA polymerase complex and associated factors at the transcriptional start site, or by recruiting other chromatin modifiers that result in impaired transcription.[28]

Methylation can be classified into de novo and maintenance methylation. The de novo methylation events occur in germ cells or the early embryo, suggesting that de novo methylation is particularly active at these stages. There is evidence, however, that de novo methylation can also occur in adult somatic cells. A significant fraction of all human CpG islands are prone to progressive methylation in certain tissues during aging, or in abnormal cells such as cancers and permanent cell lines. On the other hand, maintenance methylation describes the processes that reproduce DNA methylation patterns between cell generations. The simplest conceivable mechanism for maintenance depends on semiconservative copying of the parental strand methylation pattern onto the progeny DNA strand (Figure 4).[6,20]

DNA methylation is established and maintained by a family of DNA methyltransferases (DNMTs). Five different DNMTs are known: DNMT1, DNMT2, DNMT 3aDNMT3b and DNMTL. DNMT1 is a maintenance DNMT , where it adds methyl groups to hemimethylated DNA during DNA replication ,while DNMT 3a,3b and L are de novo which act after DNA replication. The function of Dnmt2 is not yet clear. DNA methylation is considered a dynamic process between active methylation, mediated by DNMT using S adenosylmethionine (SAM) as the methyl donor and removal of methyl groups forming 5-methylcytosine residues by both passive and active mechanisms. After transfer of the methyl group, SAM is converted to S-adenosylhomocysteine (SAH), a potent inhibitor of most SAM-dependent methyltransferases. [6]

Genomic patterns of cytosine methylation, whether donated by de novo or maintenance DNMTs, play a critical role in gene regulation and chromatin organization during embryogenesis and gametogenesis.[4,18] Patterns of DNA methylation can be propagated to daughter cells during tissue growth, making this biochemical modification a very effective way to maintain epigenetic states.[18] Thus , DNA methylation provides a heritable and critical component of epigenetic regulation.

Table (I): DNA methyltransferases [20]

DNMT	Catalytic activity	Substrate preference
1	Yes	CpG, hemimethylated
2	Not demonstrated in vivo	CpT, CpA, physiologic substrate uncertain
3a	Yes	CpG, de novo
3b	Yes	CpG, de novo
3L	No	—

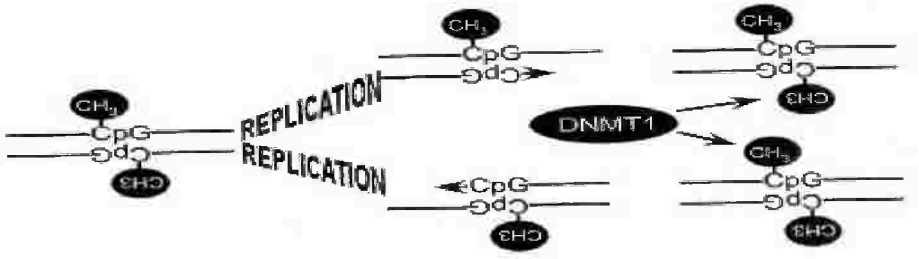

Figure (4): Propagation of methylation pattern to daughter cells. [20]

CpG= cytosine phospho guanine

Role of DNA methylation in normal cells and disease states:

DNA methylation plays a role in many cellular processes including silencing of repetitive and centromeric sequences; X chromosome inactivation in female mammals; and mammalian imprinting, all of which can be stably maintained. [29,30]

Genomic imprinting:

Normally, DNA methylation has its primary effects on gene regulation during embryonic development, when it can affect organism-wide gene expression, and in differentiation, where its role is restricted to specific tissue types. With regard to the effect of DNA methylation on organism-wide gene expression, this occurs mainly in the context of the germline transmission of heritable epigenetic states. In the germline, a major role of DNA methylation in mammals is to ensure the expression of only one allele for specific genes based on the parent of origin, which is a phenomenon known as imprinting.[31] In mammals, failure to methylate and 'imprint' the correct parental allele leads to a spectrum of diseases, including Silver- Russell, Beckwith-Wiedemann and Prader-Willi syndromes. When the normally active gene is lost or turned off, a loss-of-function phenotype may result; if the normally inactive copy of the gene is turned on, overproduction of a gene product may also produce a disease phenotype. For example, loss of a growth-promoting factor may lead to growth restriction while an excess of such a factor may contribute to overgrowth and malignancy. Beckwith-Wiedemann is a prototype for understanding epigenetic dysregulation and its role in tissue growth and cancer. This syndrome is characterized by somatic overgrowth, including macroglossia and visceromegaly;

congenital malformations, including omphalocele; and, in a subset of cases, a predisposition to embryonal pediatric tumors, including Wilms tumors of the kidney.[20]

Another disease of the best-characterized imprinting related disease is Prader-Willi syndrome (PWS), which is a neurobehavioral disorder that results from paternal deficiency of the same chromosomal fragment. The 15q11-q13 region contains a number of imprinted genes that are regulated by an imprinting center, which contains functional elements. The loss of either the maternal or paternal imprinting center leads to abnormal DNA methylation in this locus and abnormal expression of the affected genes. In addition to developmental disorders, the deregulation of a number of imprinted loci has been implicated in cancer. [20]

Examples of this phenomenon include loss of imprinting (LOI) of PEG1/MEST in lung cancer, CDKN1C in pancreatic cancer and IGF2 in colorectal cancer.[28,32 ,33]

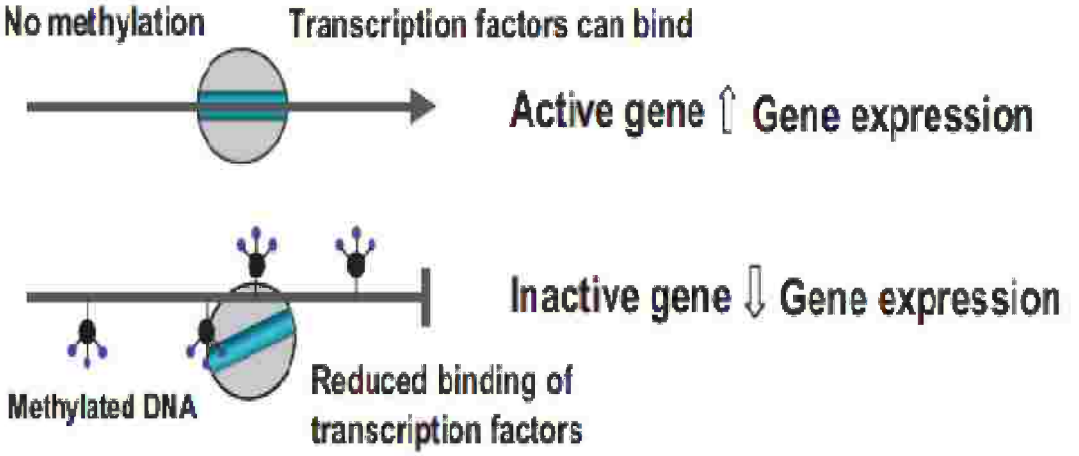

Figure (5): Effect of DNA methylation on gene expression. [34]

As methylation level increases at the prompter region , it reduces the binding of transcription factors which leads to decrease in gene expression

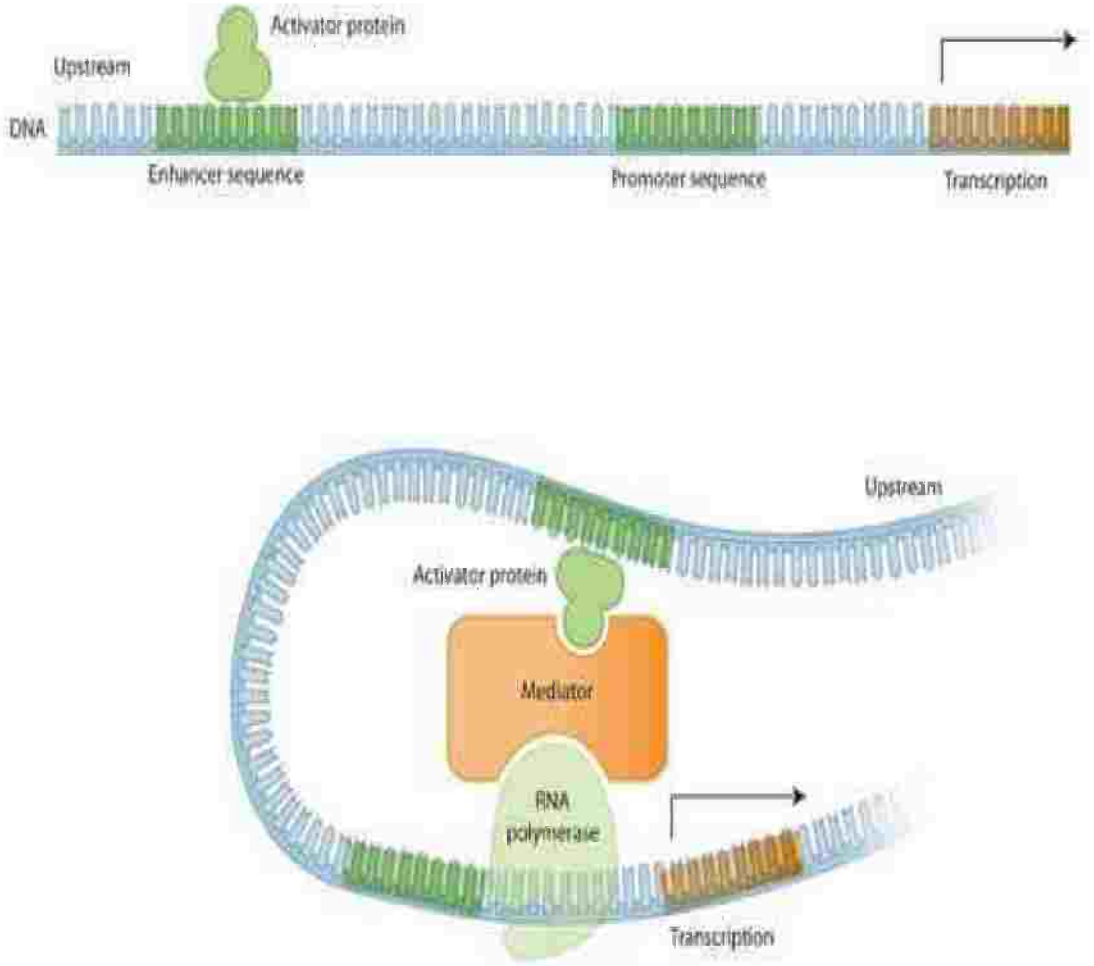

Figure (6): Diagram illustrating the gene regions.[20]

An activator protein bound to DNA at an upstream enhancer sequence can attract proteins to the promoter region that activate RNA polymerase (green) and thus transcription. The DNA can loop around on itself to cause this interaction between an activator protein and other proteins that mediate the activity of RNA polymerase

Effect of aging and diet on DNA methylation:

A growing body of research has reported associations between age and the state of the epigenome. In particular, DNA methylation associates with chronological age over long time scales, and changes in methylation have been linked to complex age-associated diseases such as metabolic diseases and cancer. ''Epigenetic drift,'' is a phenomenon whereby the DNA methylation marks in identical twins increasingly differ as a function of age. Thus, the idea of the epigenome as a fixed imprint is giving way to the model of the epigenome as a dynamic landscape that reflects a variety of chronological changes.[35]

If individuals indeed age at different rates, it might be expected that their individual methylomes should diverge over time. This is based on that the methylomes of the very young share certain similarities and that these similarities diminish as individuals accumulate changes over time. [36,37]

The epigenetic pattern may change during the course of life, affecting key genes in the respiratory chain Cytochrome c oxidase polypeptide 7A1(COX7A1), which is part of complex 4 of the respiratory chain, is a target of age-related DNA methylation. Whereas DNA methylation of the COX7A1 promoter is increased in skeletal muscle of elderly compared with young, the opposite pattern is found for COX7A1 gene expression.[36]

On the contrary, many mammalian tissues demonstrate global hypomethylation of DNA and decreased methyltransferase (DNMT1 and DNMT3a) expression with increased age. Global hypomethylation of DNA is seen in repetitive sequences and may promote genomic instability during aging. Increased age is also associated with hypomethylation of specific genes, e.g., proto-oncogenes, thereby increasing susceptibility to cancer, especially if combined with hypermethylation of tumor suppressor genes. [34,38]

S-Adenosylhomocysteine (SAH), an inhibitor of multiple transmethylation reactions including DNA methylation, also increases with age relative to S-adenosylmethionine (SAM), which may contribute to the progressive hypomethylation.[39]

Endogenous and exogenous mechanisms can modify DNA methylation. A diet deficient in methyl donors (choline, folate, methionine, and vitamin B-12), can promote reversible hypomethylation of DNA. Similarly, folate restriction results in lymphocyte DNA hypomethylation, suggesting that dietary factors might be important. Exogenous agents such as UV light and medications such as procainamide and hydralazine promote DNA hypomethylation. [40]

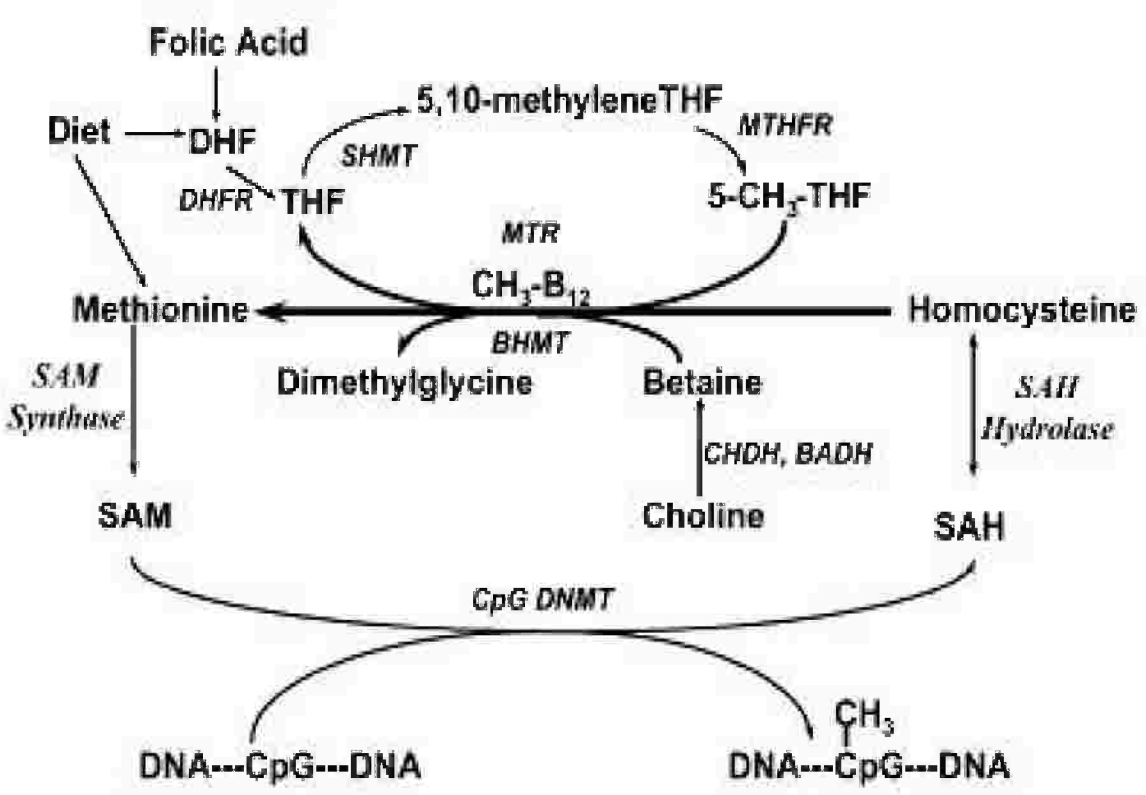

Figure (7): **Simplified scheme of the role of folate in DNA methylation.** B12, vitamin B-12; BADH, betaine aldehyde dehydrogenase; BHMT, betaine: homocysteine methyltransferase; CH_3, methyl group; CHDH, choline dehydrogenase; DHF, dihydrofolate; DHFR, dihydrofolate reductase; MTHFR, methylenetetrahydrofolate reductase; MTR, methionine synthase; SHMT, serine hydroxymethyltran sferase; THF, tetrahydrofolate; SAH, S-Adenosylhomocysteine;SAM,S-adenosylmethionine.[40]

DNA methylation and Cancer:

DNA Hypermethylation:

Deregulation of the mechanisms that control DNA methylation in somatic tissues, and subsequent alteration of DNA methylation patterns, is a common phenomenon in cancer. The aberrant methylation of DNA in cancer appears to be regional and results in both the hypermethylation of CpG islands in the promoter regions of genes, and global DNA hypomethylation (generally in pericentric heterochromatin). [29] The aberrant methylation of gene promoters that is observed in many cancers is believed to contribute to the darwinian evolution of the tumors by favoring the clonal expansion of cells that have acquired the aberrantly methylated genes. This is because DNA methylation in specific promoters can silence tumor suppressor genes such as CDKN2A, DAPK1 and CDH1. Methylation of tumor suppressor genes can promote the acquisition of tumorigenic behaviors, like increased proliferation, escape from apoptosis and enhanced invasiveness. [41,42]

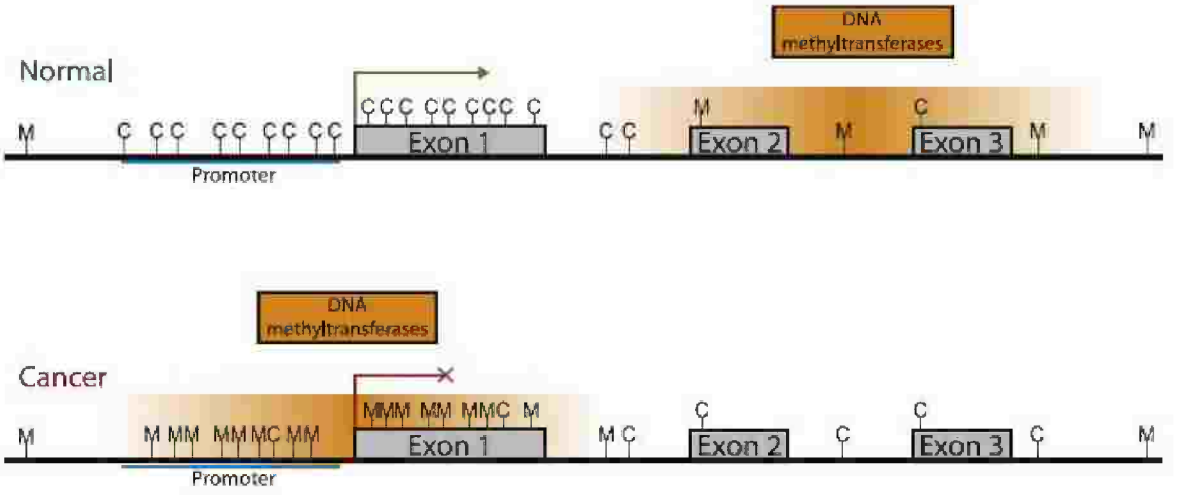

Figure (8): DNA methylation in normal and cancer cell. In a normal cell CpG islands (C) in actively transcribed promoter regions are not methylated allowing transcriptional activity (green arrow). CpG islands elsewhere, within genes and intergenic spaces are more often methylated (M). In a cancer cell the reverse is observed. Red line with X: repressed transcription.[41]

Hypermethylated promoter DNA is associated with virtually every type of human tumor, with each type of tumor having its own signature of methylated genes, such as the methylation of GSTP1 in prostate cancer, the von Hippel-Landaui syndrome gene VHL in renal cancer, the mismatch repair gene MLH1 in colon and endometrial cancers, and APC in esophageal cancer. [43]

There is growing evidence to suggest that cancers may originate from alterations of adult stem cells rather than of more differentiated cells. Alternatively, a differentiated cell may be induced to de-differentiate and return to a stem cell-like state through alterations in the epigenetic state of these cells. Adult stem cells are thought to function as long-lived, multipotent and self-renewable cells that serve to replace cells in a given tissue as needed. One of the best-studied adult stem cell populations is the hematopoietic stem cells (HSCs).[44] These cells reside in the bone marrow and are capable of becoming all types of hematopoietic cells, and some evidence suggests that they can occasionally differentiate into other cell types as well. In HSCs, differentiation requires the interplay of a large number of transcriptions factors, and one of these, PU.1(Purine rich), is itself regulated by DNA methylation and can complex with DNMT3a and DNMT3b to methylate a number of targets genes. Consequently, altering the methylation status of the PU.1 promoter can lead to both lymphoma and leukemia. Thus, epigenetic alterations appear to play a significant role in the abnormal regulation of cancer stem cells and may be a dominant mechanism governing this aspect of tumor biology. [45]

DNA Hypomethylation:

In contrast to hypermethylation of a selected subset of promoters, hypomethylation of heterochromatin can impair the cells ability to silence parasitic DNA elements, such as endogenous retroviral elements, whose expression or movement throughout the genome may lead to cancer formation. Indeed, global DNA hypomethylation has been correlated

with genetic instability, and in somatic cells, this instability can contribute to both mutator phenotypes, which increase the mutation rates in tumors cells, and to large-scale chromosomal abnormalities such as translocations. These phenotypes can ultimately result in cancer by favoring the accumulation of gene mutations that lead to oncogene activation or the inactivation of tumor suppressor genes. Similar to hypomethylation of parasitic elements in heterochromatin, hypomethylation of some gene promoters may increase their expression and cause them to function as oncogenes. [46]

Hypomethylation of DNA sequences is often observed during the early stages of tumorigenesis or in abnormal non-neoplastic tissue, such as hyperplasia. For example, age-related hypomethylation of certain DNA sequences appears to precede aneuploidy in a subset of gastrointestinal cancers. [47]

DNA Demethylation:

In mammals, DNA demethylation also plays an important role in development and tumorigenesis. DNA demethylation, occurring in primordial germ cells (PGCs) and in early embryos, is essential for cells to return to a pluripotent state. The greatest loss of methylation in PGCs is observed within introns, intergenic regions and repeats, followed by exons, and then promoters. Demethylation may be carried out by cytosine deaminases, converting 5mC to thymine, followed by T-C mismatch repair that specifically replaces thymine with cytosine. [48,49]

Carcinogenesis and tumor progression lead to the initiation of cancer-associated DNA demethylation. This demethylation then may spread in *cis* by both additional rounds of active demethylation and by passive demethylation involving failures in classical maintenance methylation and replication-associated repair methylation. [50] The net result of some of this cancer-associated DNA demethylation could be abnormal modulation of transcription and even some aberrant posttranscriptional processing of transcripts as well as increases in DNA recombination, thereby contributing to tumor formation and progression.[51,52,53] (Fig.9)

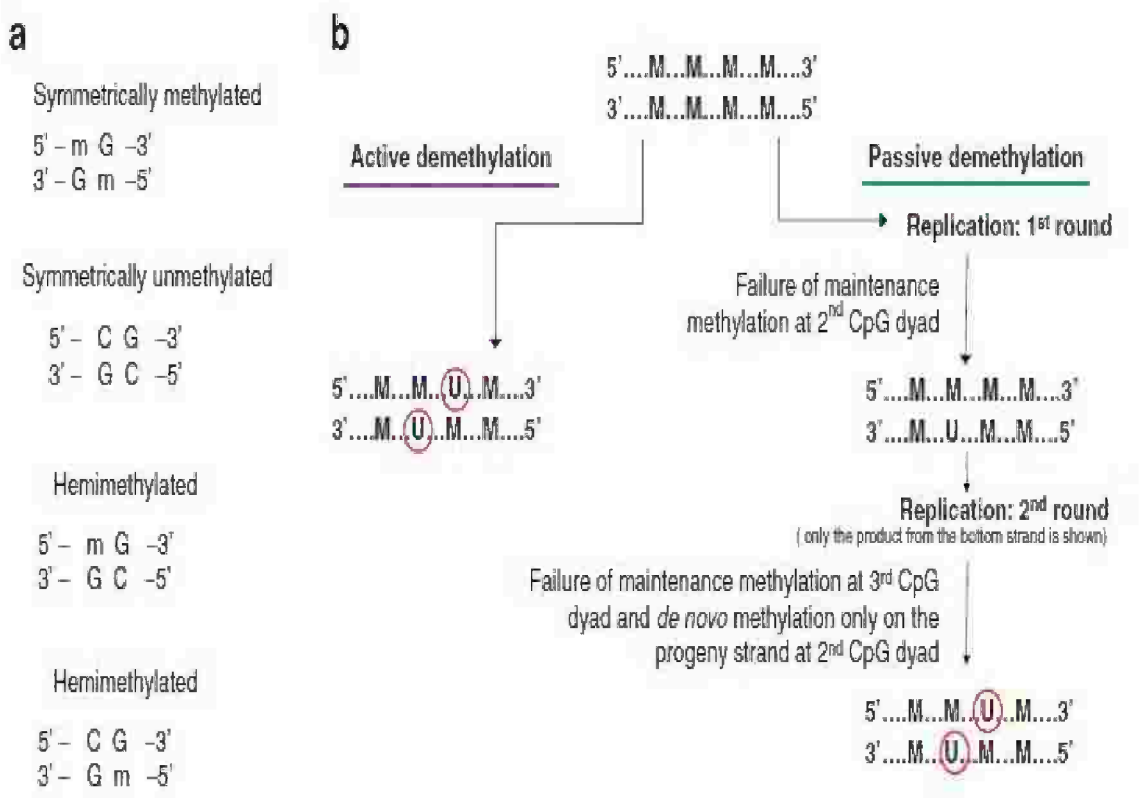

Figure (9): Consecutive hemimethylated dyads of opposite orientation in normal and cancer cells are best explained by active demethylation. (a) m, 5mC; C, unmethylated cytosine. (b) M, 5`-5mCpG-3`; U, 5`-CpG-3`. The generation of hemimethylated dyads of opposite orientation by passive demethylation would involve improbable changes in the second round of replication.[53]

Techniques applied in DNA methylation studies :

Chromatographic techniques:

Several different analytical techniques have been used in the determination of the major and minor base composition of DNA. One of the earliest techniques used was paper chromtography of DNA digests with quantitation by ultraviolet absorption. Paper and thin layer chromatographic methods have also been applied to digests of DNA which were labeled in vivo with radioactive precursors. Later, reversed-phase high performance liquid chromatographic method had been used for the analysis of major and minor deoxyribonucleosides in DNA.[54]

Molecular techniques:

The earliest method, which is still the most direct and least subject to artifacts, is Southern blotting of genomic DNA digested with methylation-sensitive restriction enzymes. Certain restriction enzymes (e.g., *Hpa*II, *Sma*I, *Not*I) with CpG as part of their recognition sequences do not cut that site when the C is methylated, the enzyme product is conveniently measured by Southern blot hybridization using a suitable radioactive probe.[20]

The standard molecular biology techniques used for examination of genetic information have traditionally not been applicable to study DNA methylation. Methylation analysis requires an initial step to select for methylated DNA and this DNA can then be combined with standard techniques such as Polymerase Chain Reaction (PCR), array hybridisation and sequencing. Earlier studies of methylation typically focused on small numbers of candidate genes on a gene-by-gene basis. [55]

While numerous methods have been developed to examine DNA methylation, most of these are based on an initial DNA selection step using one of three basic methods:

1- Examination of bisulphite converted DNA
2- Cutting of DNA by methylation-sensitive restriction enzymes
3- Immunoprecipitation using antibodies against methylation-related proteins. [56]

Bisulphite conversion

One of the key events in the progression of methylation studies was the application of bisulphite conversion to methylation analysis in the early 1990s. This chemical reaction is a three-step process that converts unmethylated cytosine bases to uracil while methylated cytosines remain unchanged.[57](Fig.10,11).

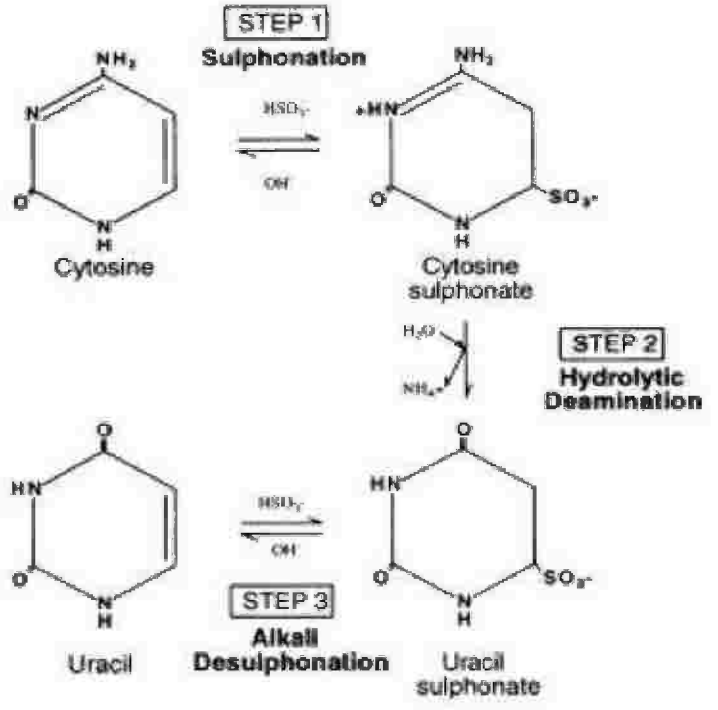

Figure (10): Chemistry of bisulphite conversion.

In this method, DNA is first denatured with sodium hydroxide prior to bisulphite-modification. Unmethylated cytosines undergo deamination by sodium bisulphite by the following three steps: 1- Sulphonation of cytosine to form cytosine sulphonate 2- Hydrolytic deamination to give a uracil-bisulphite derivative 3- Removal of the sulphonate group by a subsequent alkali treatment to form uracil. The initial two steps are performed at acidic pH while the final step needs alkali treatment. [57]

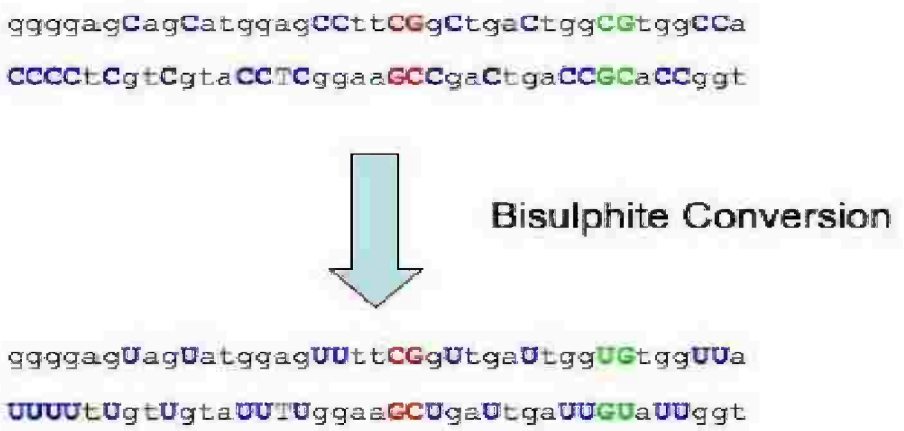

Figure (11): Effect of bisulphite conversion on DNA sequence. While unmethylated cytosines (highlighted in blue or green) are converted to uracil, methylated cytosines (highlighted in red) remain unchanged following bisulphite conversion. The sequence difference thus created between methylated and unmethylated alleles facilitates design of subsequent PCR-based assays that can discriminate between methylated and unmethylated DNA.[57]

Methylation specific PCR:

The discovery of the effects of bisulphite conversion led to a revolution in DNA methylation analysis in the subsequent years. An epigenetic modification of DNA could now be analysed as a genetic difference, as an effect of a single nucleotide polymorphism (SNP), allowing numerous standard molecular biology techniques to be coupled with the initial bisulphite conversion step. Among the first of these techniques was **Methylation-specific PCR (MSP)** in which two primer sets are designed to bisulphite-converted DNA. One primer set is complementary to methylated DNA (M-MSP reaction) while the other primer set is specifically complementary to unmethylated DNA (U-MSP reaction). Bisulphite-converted DNA is amplified in two separate reactions and the products are analysed in parallel by agarose gel electrophoresis. [58]

MSP is useful as a screening tool as a negative M-MSP reaction suggests absence of methylation. However, the method is at best semi-quantitative with a positive M-MSP reaction generally being unable to distinguish between samples with high and low levels of methylation.[58] This has been one of the most widely used techniques for methylation analysis of single genes given its flexibility to interrogate almost any CpG site, its relative ease and its cost efficiency. Subsequent years have seen modifications of basic MSP such as the **Methy Light** technique that provides quantitative values through a methylation-specific modification of the real-time PCR based mechanism. [59]

Bisulphite sequencing:

Bisulphite converted DNA has also been coupled with **sequencing** techniques. In the early 1990s, this involved Sanger sequencing of cloned PCR products from individual loci or direct sequence analysis. This method was applied on a large scale to analyse the methylation patterns of human Chromosomes 6, 20 and 22 as well as the coupling of bisulphite-converted DNA with sequencing-by-synthesis methods. These methods involve sequential addition of nucleotides and their incorporation into a primer-directed polymerase extension. As each nucleotide is incorporated they release pyrophosphate which is enzymatically converted to give a measurable light signal. This allows quantitative measurement of methylation status at each locus by examining the ratio of T and C signals. Such methods are not only applicable to single gene analysis but also can be applied in wide genome methylation assays as with **Pyrosequencing.**[56]

Methyl specific PCR VS Sequencing:

Furthermore, with MSP, simultaneous detection of unmethylated and methylated products in a single sample confirms the integrity of DNA as a template for PCR and allows a semiquantitative assessment of allele types.(Fig.12) MSP can provide similar information and has the following advantages over sequencing. First, MSP is much simpler and requires less time than genomic sequencing, with a typical PCR and gel analysis taking 4-6 hr. In contrast, for genomic sequencing, amplification, cloning, and subsequent sequencing may take days. Second, MSP avoids the use of expensive sequencing reagents and the use of radioactivity. [60, 61]

MSP eliminates the frequent false positive results due to partial digestion of methylation-sensitive enzymes inherent in previous PCR methods for detecting methylation. MSP also allows examination of all CpG sites, not just those within

sequences recognized by methylation-sensitive restriction enzymes. This markedly increases the number of such sites that can be assessed and will allow rapid, fine mapping of methylation patterns throughout CpG-rich regions.[58]

Figure (12): Methyl specific PCR product on agarose gel. [58]

Restriction enzymes

Restriction enzymes are enzymes that cut DNA at specific sequences known as restriction sites. DNA may be cut into methylated and unmethylated genomic DNA by methylation-sensitive restriction enzymes that can distinguish between methylated and unmethylated recognition sites. This is used as an initial selection step that is then coupled with standard genetic techniques. These initially included techniques such as Southern Blotting which required large amounts of starting DNA and while PCR amplification (Restriction enzyme-related PCR)[62] assisted with this issue, the technique still suffers from the fact that it can only detect methylation changes at sites where the enzyme had recognition sites. Restriction enzyme method has also been used in tandem with bisulphite conversion methods for locus–specific studies in a technique called Combined Bisulphite Restriction Analysis (COBRA).[63]

Restriction enzyme based analysis was one of the first methods used to examine methylation on a more global scale in technique called Restriction Landmark Genomic Scanning (RLGS) in an important study of methylation in multiple human tumours.[64] Another technique using restriction enzymes that has been adapted for multiple gene analysis is Differential Methylation Hybridisation. This technique uses methylation-sensitive restriction enzymes to cut genomic DNA into fragments containing CpG islands. These fragments are then amplified by PCR and hybridised onto probes on a nylon membrane or glass slide microarray.[65]

Immunoprecipitation

An alternative study method is to treat genomic DNA with an antibody specific for a component of methylated DNA allowing the methylated fragments to be immunoprecipitated. The best-known application of this is Methylated DNA Immunoprecipitation (MeDIP). Following sonication of genomic DNA, anti-5´-methylcytosine antibody is used to precipitate methylated DNA. This can then be compared with input DNA by differentially labelling both samples and hybridising to a CpG array. [66] This has been coupled with array Comparative Genomic Hybridisation and more recently with high-throughput sequencing techniques so that fragments captured by 5-methylcytosine can be sequenced and mapped to a reference genome sequence to more specifically identify sites of methylation (MeDIP-seq). [67]

CHAPTER III

Hepatitis C virus and Hepatocellular Carcinoma

Hepatitis C virus

HCV is a small-sized spherical enveloped RNA virus that belongs to the family Flaviviridae. It was discovered by Choo and coworkers in 1989. [68]HCV consists of a single-stranded RNA genome encoding a single polyprotein, which is post-translationally processed into single known proteins, 4 structural (C, E1, E2 and p7) and 6 non-structural (NS2, NS3, NS4A, NS4B, NS5A and NS5B).[69] Some of these proteins have a role in starting and maintaining chronic inflammation. NS5A, for instance, promotes inappropriate upregulation of cyclooxygenase-2 (COX-2), which is an inducible COX isozyme able to contribute to chronic inflammation and fibrosis through production of various prostaglandins.[70]

HCV is divided among six genotypes with numerous subtypes. These genotypes can differ up to 30% from each other in nucleotide sequence. [71]Although all genotypes may be found throughout the world, a marked geographical distribution has been noted. In North America, genotype 1a predominates followed by 1b, 2a, 2b and 3a. In Europe, genotype 1b is predominant followed by 2a, 2b, 2c, and 3a. Genotypes 4 and 5 are found almost exclusively in Africa. Genotype 4a accounts for most infection in Egypt, [72] while genotype 6 can be found in south-east Asia. [73]

It has been well documented that Egypt has one of the highest prevalence rates of HCV infection in the world to be around 15-20 %.[74] Infection with the hepatitis C virus (HCV) is a leading cause of chronic liver disease world-wide. Pathogenesis of liver injury is not fully understood. There is a complex relationship between HCV and its host. The infection leads to viral persistence and chronic disease in a very high proportion of cases, despite broad humoral and cellular immunological responses to viral proteins. High rate of mutation leads to the generation of quasispecies, which continuously evolve in infected individuals.[75, 76]

The HCV infection is self limited with minimal liver affection in a minority of cases, while the majority develop persistent infection. The fate of persistent infection could be; chronic hepatitis, progressive fibrosis and even hepatocellular carcinoma. However, the rate at which fibrosis develops varies between individuals with some progressing to cirrhosis and others never progress to any significant liver disease within their natural life span.[77]

Hepatocellular carcinoma:

Epidemiology

Hepatocellular carcinoma (HCC) is one of the most common cancer worldwide. It shows a wide geographical variation with low incidence areas in North America and Europe and high incidence areas in Africa and Asia. 70%-80% of hepatocellular carcinoma occurs in cirrhotic liver. In high incidence areas, such as Asia and Africa, HCC is strongly associated with chronic viral hepatitis B and C and liver cirrhosis. Nutritional factors, toxins and metabolic diseases contribute also to hepatocarcinogenesis. [78]

Hepatocellular Carcinoma (HCC) is the fifth most common solid tumor worldwide and the third leading cause of cancer-related death, accounting for approximately 600,000 deaths per year worldwide. Infection with either HBV or HCV, is the major risk factor for HCC worldwide, with at least one of the two viruses present in over 80% of HCC cases. More than 80% of HCC cases occurring in developing countries are due to HBV infection, which are preventable through effective childhood HBV vaccination. [78]

On the other hand, the recent increase in the incidence of HCC in Western countries is largely due to the HCV endemic. About 200 million people are infected with Hepatitis C Viruses (HCV) worldwide. More than two thirds of people with acute HCV infection will develop persistent HCV infection, leading to chronic hepatitis, liver cirrhosis and ultimately HCC It has been estimated that age adjusted HCC incidence rates have doubled between 1985 and 2002. [79]

Surgical resection or liver transplantation remains the most effective treatment options for HCC; however, very few patients are suitable for these treatments. Despite recent advances in HCV antiviral therapies, these therapies are only effective to prevent HCC in a small proportion of highest risk patients, as sustained viral clearance is difficult to accomplish among patients with liver cirrhosis.Therefore, it is imperative to elucidate the molecular mechanisms underlying HCV caused hepatocarcinogenesis, in order to identify early detection biomarkers as well as effective targeted therapies and thus improve clinical outcomes of HCV associated HCC. [78]

The burden of hepatocellular carcinoma (HCC) has been increasing in Egypt with a doubling in the incidence rate in the last decade . The natural history of HCV infection and disease progression, however, are influenced by additional factors such as duration of infection, age at infection, sex, co-infection with HBV, the level of HCV viraemia and its genotype. [80]

Molecular aspects of Hepatocarcinogenesis

Central to the concept of molecular carcinogenesis are mutations of oncogenes and tumor suppressor genes as well as genetic instability of cellular DNA, including mismatch repair deficiency and impaired chromosomal segregation. [81]

In hepatocarcinogenesis, these genetic events occur in the setting of liver cell injury and necrosis associated with an increased rate of hepatocyte regeneration and mitosis. Any exogenous agent, viral or other, that contributes to chronic low-grade liver cell damage and mitosis potentially increases the risk of HCC development, rendering liver cell DNA susceptible to additional genetic alterations. Overall, there is a variety of molecular mechanisms by which environmental and viral carcinogens may play a role in HCC development.[81]

Cirrhosis was found to be a common pathway by which several risk factors exert their carcinogenic effect and it can be considered as preneoplastic condition. Several mechanisms at the cirrhosis stage appear to accelerate cancer formation including: telomere dysfunction, chromosomal instability and alterations of the microenvironment and macroenvironment stimulating cellular proliferation. [82]

The HCC is thought to evolve along a multi-step process, where dysplastic nodules develop in the cirrhotic liver, progressing to HCC with an accumulation of genetic mutations. The emergence of neoplasia requires several elements, including self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, unlimited replicative potential, tissue invasion and metastasis, and sustained angiogenesis.[83]

The precise molecular mechanism underlying HCV infection caused hepatocarcinogenesis is not fully understood. Because HCV is an RNA virus that does not involve a DNA intermediate, it has been proposed that HCV infection leads to hepatocarcinogenesis indirectly through viral induced inflammation and oxidative stress. Subsequently, this microenvironment sets the stage for malignant transformation of hepatocytes through accumulation of both genetic and epigenetic changes. However, inflammation alone could not fully explain HCV induced hepatocarcinogenesis, as patients with autoimmune hepatitis rarely develop HCC despite the presence of persistent liver inflammation and cirrhosis. More recent studies have suggested that HCV might play a more direct role in HCC carcinogenesis through interaction between viral and cellular proteins. [79] A detailed understanding of how these alterations select for distinct molecular pathways during the development of HCC could improve the screening, prevention, and treatment of HCC in patients with chronic liver disease and cirrhosis.

Epigenetic mechanisms are not only essential for the dynamic transcriptional regulation in embryonic and somatic stem cells, but are also actively involved in tumorigenesis: genes important for pluripotency are epigenetically regulated and aberrant epigenetic changes have been detected in virtually all human malignancies studied, including HCC. [79]

Chapter IV

DNA Methyaltion and Liver Diseases

Currently, liver fibrosis is known to be a part of the dynamic process of continuous extracellular matrix (ECM) remodeling in chronic liver injury. In liver fibrosis, a liver injury activates the Kupffer cells–resident macrophages of the liver sinusoids–thereby inducing inflammation. This inflammatory response triggers the activation of hepatic stellate cells, which play a key role in fibrogenesis by transdifferentiating into myofibroblasts. The proliferation of myofibroblasts and stimulation of ECM synthesis, ultimately results in liver fibrosis. [84]

The DNA methylation status of well-known genes associated with fibrosis progression, such as Rasal1, Fli1, and Thy1, has been reported to increase along with fibrosis progression, which induces proliferation of fibroblasts and the production of collagen. [84]

As for many other tumors, development of HCC is understood as accumulation of genetic and epigenetic alterations in regulatory genes, leading to activation of oncogenes and inactivation or loss of tumor suppressor genes (TSG). In the last decades, in addition to genetic alterations, epigenetic inactivation of (tumor suppressor) genes by promoter hypermethylation has been recognized as an important and alternative mechanism in tumorigenesis.[85] In HCC, aberrant methylation of promoter sequences occurs not only in advanced tumors, but it also has been observed in premalignant conditions such as chronic viral hepatitis B or C and cirrhotic liver.[79]

Aberrant epigenetic changes have been detected in virtually all human malignancies studied, including HCC. Infection with HCV is a major risk factor for the development of HCC. As viral proteins may actively participate in epigenetic regulation of hepatic cancer stem cell phenotypes and induce HCC-specific epigenetic changes. Identification of host epigenetic alterations induced by HCV infection and epigenetic differences between hepatic cancer stem cells and the bulk non-tumorigenic cancer cells, may yield potential biomarkers for early detection, as well as therapeutic targets for HCV associated HCC. [86]

Overexperession of DNMT in HCC was reported by several studies,it was observed that there was higher expression level of DNMT 1,3a,3b in HCC, also in cirrhosis and dysplastic liver nodules. Besides, the progression of liver fibrosis has been reported to be associated with hypermethylation of DNA.HCC from patients without precancerous conditions or risk factors, respectively, showed significantly lower levels of methylation than HCC arising from patients with chronic hepatitis B and C or patients with cirrhosis.[79]

Several candidate gene approach studies have reported the role of DNA methylation of various panels of genes during the stepwise progression of HCC. Methylation of several genes occurred not only in HCC and its precursor lesions, but also in chronic hepatitis and liver cirrhosis, suggesting that these changes are early events during HCC progression.[87,88,89] DNA methylation of four genes (Collagen, type I, alpha 2(Col1A2), Insulin like growth factor binding protein 2 (IGFBP2) , Connective tissue growth factor (CTGF) and fibronectin) increased from normal liver, chronic hepatitis, liver cirrhosis to hepatoma. Frequency of E-cadherin promoter methylation increased

from dysplastic nodules to early stage and late stage HCCs. Similarly, methylation of p16, p15 and SFRP1 was not only present in HCC, but was also present at low frequencies in chronic hepatitis and liver cirrhosis samples. Further, methylation analysis in various liver tissues demonstrated that the number and frequency of genes methylated progressively increased in liver cirrhosis, dysplastic nodules and HCC, supporting the hypothesis that CpG island methylation of tumor-related genes is an early and frequent event and methylation changes accumulate during a multistep hepatocarcinogenesis.[78,90]

Chapter V

Serine Protease Inhibitor, Kunitz type 2, Gene

The official name of this gene is "Serine protease inhibitor, Kunitz type 2, *SPINT2* is the gene's official symbol. This gene encodes a transmembrane protein with two extracellular Kunitz domains that inhibits a variety of serine proteases. The protein inhibits hepatocyte growth factor HGF activator which prevents the formation of active hepatocyte growth factor. This gene is a putative tumor suppressor, and mutations in this gene result in congenital sodium diarrhea. Alternative names are Serine peptidase inhibitor, Kunitz type 2, Hepatocyte growth factor activator inhibitor type 2 (HAI-2), placental bikunin and inhibitor of HGF activator. It inhibits other proteins like plasmin, plasma and tissue kallikreins and factor XIa. [91]

***SPINT2* gene location:**

Cytogenetic Location: 19q13.1, The SPINT2 gene is located on the long (q) arm of chromosome 19 at position 13.1. Molecular Location on chromosome 19: base pairs 38,264,457 to 38,292,613. [91] (Fig.13)

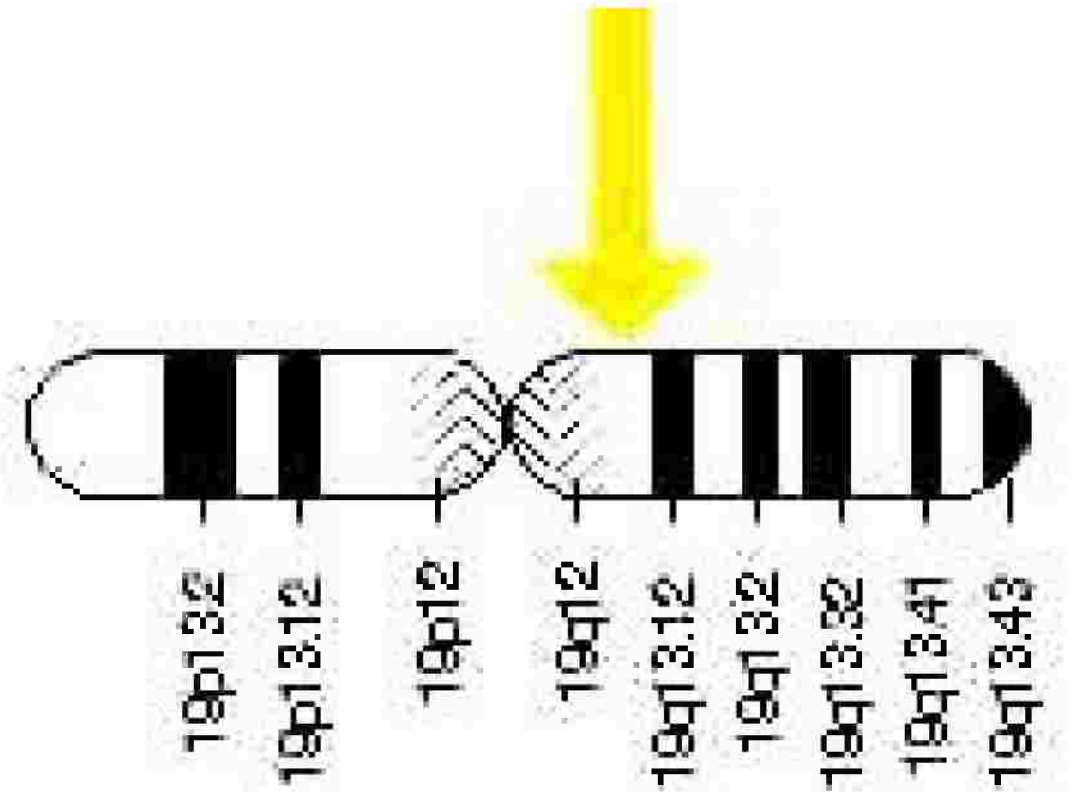

Figure (13): Chromosomal location of SPINT 2 gene. [91]

SPINT 2 gene is expressed in placenta, kidney, pancreas, prostate, testis, thymus, trachea and liver.[92]SPINT2 has been implicated in the pathogenesis of ovarian and hepatocellular carcinoma. [93]

Hence, inactivation of SPINT 2 might promote tumorigenesis by multiple mechanisms.[94] During development, HGF is a potent growth factor for epithelial cells and has a critical role in regulating cellular motility . Dysregulation of the HGF/MET signaling pathway is frequent in human cancer and may result from a variety of mechanisms . Thus, up-regulation of HGF and c-MET expression can activate c-MET signaling by a paracrine effect. Tumor formation and metastasis is a complex multistep process that requires the acquisition of a variety of properties (e.g., proliferation, invasion, angiogenesis, and antiapoptosis) that are associated with MET activation. [95]

In studies of ovarian cancer, SPINT2 had been implicated as an inhibitor of tumor cell invasion and metastasis. In addition to inhibiting serine proteases, SPINT2 can also bind to high-affinity cell surface receptors and down-regulate urokinase plasminogen activator (uPA) and its receptor (uPAR),which is responsible for generation of plasmin from plasminogen to induce extracellular matrix degradation and promote tumor cell migration and metastasis.[96,97]

Several studies had been conducted in different types of cancer in a trial to find a relation between the methylation status SPINT 2 gene promoter and carcinogenesis; Dong et al, 2010[98] concluded that epigenetic inactivation of SPINT2 is a common event contributing to gastric carcinogenesis and may be a potential biomarker for gastric cancer in a study performed on Chinese population. In addition, Morris et al (2005),[97] found frequent loss of expression and promoter region methylation of SPINT2 in renal cell carcinoma cell lines and primary tumors.

A number of studies have shown that c-MET was overexpressed in HCC and correlated with poor prognosis. [99,100,101]HGF levels in the liver and serum were also demonstrated to be higher in cirrhosis and correlated with the occurrence of HCC. [102] Taken together, these studies suggest that the HGF/c-MET signaling pathway might be involved in the development of HCC. Because plasminogen converted to plasmin by urokinase type plasminogen activator induces the degradation of extracellular matrix and the migration of tumor cells, the urokinase-type plasminogen activator/plasmin system has been assumed to have important roles in the invasive process of many types of cancer, including HCC.[101]

Moreover, it has been demonstrated that the HGF-induced invasion of HCC is mediated by the up-regulation of urokinase-type plasminogen activator.[103] Together with the inhibitory activity of SPINT2 against plasmin, SPINT2 might have a suppressive role in the progression of HCC through the regulation of HGF signaling.

Serine peptidase inhibitor, Kunitz type 2 (SPINT2) seems to play an important role in hepatocarcinogenesis,promising a high specificity for methylation patterns of SPINT2 circulating in the bloodstream. Therefore, epigenetic changes in preneoplastic or early neoplastic stages may serve as indicator or "biomarker" for screening of patients with an increased risk for HCC.[79, 104]

DNA methylation inhibitors, including, 5-aza-2′-deoxycytidine (decitabine), 5-azacytidine, 5,6-dihydro-5-azacytidine, zebularine and RNA interference (RNAi) and antisense inhibitors of DNMT1 are being used as cancer chemotherapeutic agents. 5-azacytidine and decitabine have had the most success as cancer chemotherapeutic agents for patients with myelodysplastic syndromes (MDS), acute myeloid leukemia and chronic

myelogenous leukemia. Clinical efficacy for this type of therapy applied to MDS patients has been increased by using optimized treatment schedules that take into account the mechanism of action of demethylating agents. Combination therapies involving DNA demethylating agents have been tried for various cancers and are being refined with differing success rates.[47]

The lack of reports concerning methylation status SPINT2 gene promoter in chronic hepatitis-C infection makes it noteworthy to study it in Egyptian chronic hepatitis-C infected patients as potential epigenetic marker for evaluating the risk of development of hepatocellular carcinoma.