# CHAPTER II
REVIEW OF RELATED LITERATURES

1. CARCINOGENESIS IS A MULTISTAGE, MULTISTEP PROCESS

Cancer cells manifest six essential alterations in cell physiology that collectively dictate malignant growth: self-sufficiency in growth signals, insensitivity to growthinhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis[8]. Carcinogenesis is a multistage, multistep process consisting of at least three experimentally defined stages: initiation, promotion, and progression[9, 10]. The terms initiation and promotion refer to modes of action. The first stage, initiation, involves a heritable alteration to the genome that facilitates the clonal expansion of initiated cells in response to a promotion stimulus. Operationally, the next stage, promotion, entails the clonal expansion of initiated cells, i.e., cells that may proliferate in response to promoter treatment. In experimental models, it is possible to demonstrate that a chemical exerts a carcinogenic effect by acting at a particular stage(s). For example, many (but not all) genotoxic chemicals have the potential to act as mutagens and are effective initiating agents. Information is abundant on mechanisms of action of genotoxic carcinogens and the role that mutations play in carcinogenesis. However, carcinogenesis is more than mutagenesis; many carcinogens are not mutagens and vice versa. It is instructive to note that a mismatch repair deficiency has been identified in phenotypically normal human cells. The people who donated these cells had numerous mutations in a variety of tissues but very few tumors were evident[11]. Promoting agents are capable of facilitating the clonal expansion of initiated cells, potentially leading to a tumor by increasing cell proliferation and/or by decreasing apoptosis[12, 13]. The potential exists for the development of spontaneous mutations (some of which may result in initiation) during periods of cell proliferation, e.g., compensatory hyperplasia in response to necrosis[14, 15] and following treatment with promoting agents. Finally, the progression stage of carcinogenesis is characterized by changes in ploidy and autonomous clonal expansion. The experimentally based description of the three stages of carcinogenesis should not, in our view, be read as connoting that mutation only occurs during initiation, that mutation is the one and only basis for initiation, or that promotion only involves stimulating proliferation of and/or

inhibiting apoptosis of initiated cells[12]. In fact, heritable alterations to the genome (e.g., point mutation, rearrangements, deletions, chromosome loss and altered methylation) take place at multiple points in the carcinogenic process[16, 17]. Three hallmarks of carcinogenesis serve to keep our focus on the biology of the process: (a) the clonal evolution of tumor cell populations involves a stepwise selection of sublines that are increasingly abnormal and have a selective growth advantage over adjacent normal cells—most of the variants are eliminated[18]; (b) operationally, the promotion stage is reversible[19]; and (c) tumors arising in a single organ in response to treatment with a particular carcinogen frequently exhibit different phenotypes, which indicates that multiple pathways may lead to carcinogenesis[20]. The numerous heritable alterations to the genome involved in the sequential clonal expansion of precancerous cells that lead to a frank malignancy are illustrated in Figure 2-1. This is a modification of an earlier scheme[21] that pictured mutation as the genetic mechanism underlying carcinogenesis. We employ the term "critical event" to connote the possibility of heritable epigenetic changes, e.g., altered methylation, being involved in addition to mutation.
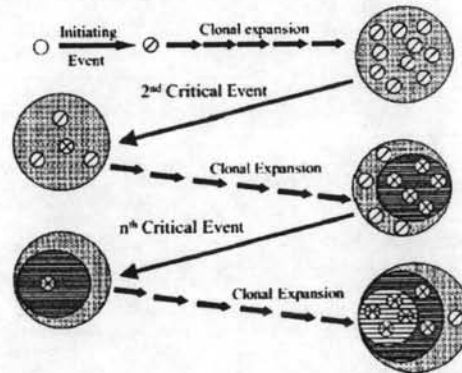
Figure 2-1: Initiation and cell proliferation in multistage carcinogenesis. The critical events referred to involve heritable changes in the genome. This diagram has been adapted from Swenberg et al.[21] to illustrate that epigenetic changes such as altered DNA methylation, in addition to mutation, may play a key role in carcinogenesis. Each line through a cell represents a critical event. Altered DNA methylation may be a mechanism underlying selective clonal expansion, i.e., hypomethylation may facilitate an aberrant increase in expression of oncogenes and/or hypermethylation may silence tumor suppressor genes. Either of these events could provide a cell with a selective growth advantage over the surrounding cells.

2. THE IMPORTANCE OF EPIGENETIC MECHANISMS

2.1 Inheritance Should be Considered on a Dual Level

Inheritance must be considered on a dual level. That is, we should distinguish the transmission of genes from generation to generation or in the somatic sense (i.e., inheritance ofDNAbase sequence) from the mechanisms involved in the transmission of alternative states of gene activity following cell division. Epigenetics is the term used to describe the latter. It may be defined as the study of mechanisms responsible for the temporal and spatial control of gene activity, e.g., changes in gene expression during development, segregation of gene activities such that daughters of an individual cell have different patterns of gene expression, and mechanisms to permit the somatic inheritance of a specific set of active and quiescent genes. DNA methylation (5-methylcytosine content of DNA) is one epigenetic mechanism by which gene activity may be regulated[22].

2.2 Epigenetics and Carcinogenesis

There is, particularly among toxicologists, an excessive focus upon mutagenesis as the (read the one and only) mechanism underlying carcinogenesis. An increased emphasis upon epigenetic mechanisms of carcinogenesis is appropriate[23]. Epigenetic regulation of gene expression is based upon a modulation of transcription by heritable mechanisms superimposed on that conferred by the primary DNA sequence. DNA methylation is an example of such a mechanism, and altered DNA methylation may play a key role in carcinogenesis, as an epigenetic nongenotoxic mechanism[24].

2.3 Initiation of Carcinogenesis May Have an Epigenetic Basis

It is usually assumed that mutagenesis provides the basis for initiation. However, there could be an epigenetic basis as well. Despite the fact that many initiators of carcinogenesis are capable of acting as mutagens under certain experimental circumstances, one does not have to assume that all of their biological effects stem from mutagenesis. Additionally, it is possible that nonmutagens may act as initiators. Hypermethylation-induced silencing of a tumor suppresser gene(s) and/or hypomethylation-facilitated aberrant increase in expression of an oncogene(s)[25] are plausible mechanisms that could underlie initiation. The involvement of epigenetics in initiation is not mutually incompatible with a role for mutation. Indeed, one or the other, or

perhaps both, may play key roles depending upon the particular circumstances, e.g., causative agent, dose, target organ, and species.

2.4 Increased Gene Expression Without Mutation May Play a Key Role in Carcinogenesis

The role that mutation plays in carcinogenesis by activating proto-oncogenes to oncogenes and silencing tumor suppresser genes is appreciated widely. It is axiomatic that, in order to affect the phenotype of a cell, a mutated oncogene must be expressed. However, aberrant increased expression of nonmutated genes plays a key role as well. Proto-oncogene overexpression may be a mechanism of activation of the ras pathway, alternative to point mutation. Overexpression of myc as well as K-ras can contribute to transformation. Furthermore, overexpression of HER2/c-erbB2 receptor tyrosine kinase induces the transformed phenotype of NIH3T3 cells and is required for tumor formation and progression in nude mice. In this context, it is important to note that the U.S. Environmental Protection Agency's proposed Cancer Risk Assessment Guidelines include a section (section 2.3.5.3) entitled "Nonmutagenic and Other Effects," which refers explicitly to a role for altered DNA methylation as a basis for the altered gene expression involved in carcinogenesis[26].

3. CELL PROLIFERATION AND CARCINOGENESIS

Cell proliferation is a fundamental component of the multistage process of carcinogenesis[15]. It plays a key role in expanding clones of initiated cells and, in addition, cell replication may contribute to carcinogenesis by facilitating mutagenesis. This can occur either by causing the fixation of promutagenic DNA damage prior to repair or as a consequence of a normal error occurring during DNA replication. In addition, during periods of cell replication, there exists the possibility for heritable decreases in DNA-5MeC (i.e., hypomethylation) owing to a limitation in the capacity for and/or fidelity of DNA maintenance methylation. This event is expected to exhibit a threshold and could result in a heritable epigenetic increase in the potential for gene expression. We do not anticipate a simple one-to-one relationship between the level of aberrant cell proliferation and the possibility for altered methylation of DNA. The ability to maintain the nascent pattern of methylation is dependent on a complex relationship between the capacity and fidelity of DNA maintenance methylase (including the

accessibility of CpG regions to the enzyme), the amount of S-adenosylmethionine, and the level of cell proliferation. A role for hypomethylation leading to increased oncogene expression in tumorigenesis is not mutually exclusive with a role for mutation. In this context, it should be noted that spontaneous deamination of 5MeC to thymine often results in C:G to T:A transitions[27]. While methylated CpG islands can be hot spots for these point mutations, they are not all expected to function equally well in this capacity because chromatin structure can result in resistance to deamination. In addition, enzymatic methylation of cytosine in DNA can be altered by DNA alkylation, e.g., adjacent O6-methylguanine residues. Loss of methylation may occur if O6-methylguanine residues occur in CpG doublets of the newly synthesized daughter strand opposite parental hemimethylated sites during DNA replication[26].

3.1 Alterations in DNA Methylation Play a Variety of Roles in Carcinogenesis

For a number of years there has been considerable interest in the role that DNA methylation plays in both normal development[28] and carcinogenesis[29]. However, at first glance there may seem to be conflicting reports concerning the role of DNA methylation in carcinogenesis: Hypomethylation facilitates aberrant oncogene gene expression in tumorigenesis[30]; hypermethylation leads to inactivation of tumor suppressor genes and marking chromosome regions for deletion[31]. Other investigators have downplayed the importance of alterations in gene expression and favor mutation playing the key role. However, carcinogenesis is more than mutagenesis. Further, the traditional view that the key mutations in cancers stem from carcinogen-DNA adducts is too narrow. The current literature provides a compelling basis for suggesting that mutations arising secondary to deamination of 5MeC and/or C are an important source of critical point mutations. Mutation, altered gene expression, hypomethylation, and hypermethylation may all play roles in carcinogenesis; they are not mutually exclusive. We do not anticipate a simple one-to-one relationship between DNA methylation and cancer, mutation and cancer, or cell proliferation and cancer, nor do we anticipate all tissues to have identical mechanisms operative. In some situations hypomethylation may be most important, in others hypermethylation, and in others mutation. Actually there is more harmony than discord here. Focusing attentionsingly on one mechanism may impede an overall understanding of carcinogenesis, e.g., both hypo- and hypermethylation appear to play key roles in carcinogenesis and which of the two predominates can depend upon

species, target organ, and the carcinogen being evaluated. Thus, it becomes important to take a holistic approach. It is instructive to recognize that apparently disparate views can be reconciled in a fashion that provides insight regarding mechanisms underlying carcinogenesis. When we juxtapose the view that carcinogenesis is a multistep/multistage process that occurs in a whole animal with the notion that carcinogenesis is more than mutagenesis, it becomes apparent that one should expect DNA methylation to play multiple roles in the transformation of a normal cell into a frank malignancy. There is the potential to undo alterations in methylation through the actions of the de novo methylases and/or demethylation not linked to replication. Thus, changes in methylation status could be involved in tumor promotion as operational reversibility is a key feature of this stage of carcinogenesis. It is important to understand that there are multiple steps that must be traversed in order to affect a change in DNA methylation, and this is true for increased methylation leading to C:G to T:A transitions as well. Therefore, it appears likely that factors altering normal methylation patterns (e.g., carcinogen treatment) would exhibit thresholds. However, this would have to be demonstrated experimentally for individual chemicals of interest[26]. The multiple factors that combine to regulate DNA methylation are illustrated in Figure 2-2, and the different ways altered methylation may facilitate carcinogenesis are illustrated in Figure 2-3.
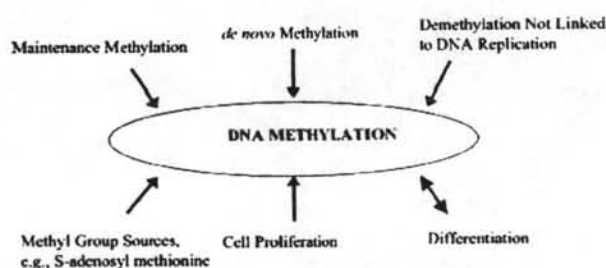
Figure 2-2 : Multiple factors controlling DNA methylation. A particular pattern of DNA methylation is the product of multiple, interdependent factors. Alteration of one or more of these can lead to major changes in methylation status. The state of differentiation can affect methylation, and methylation status can influence the state of differentiation; thus, the double-headed arrow between methylation and differentiation[26].

13
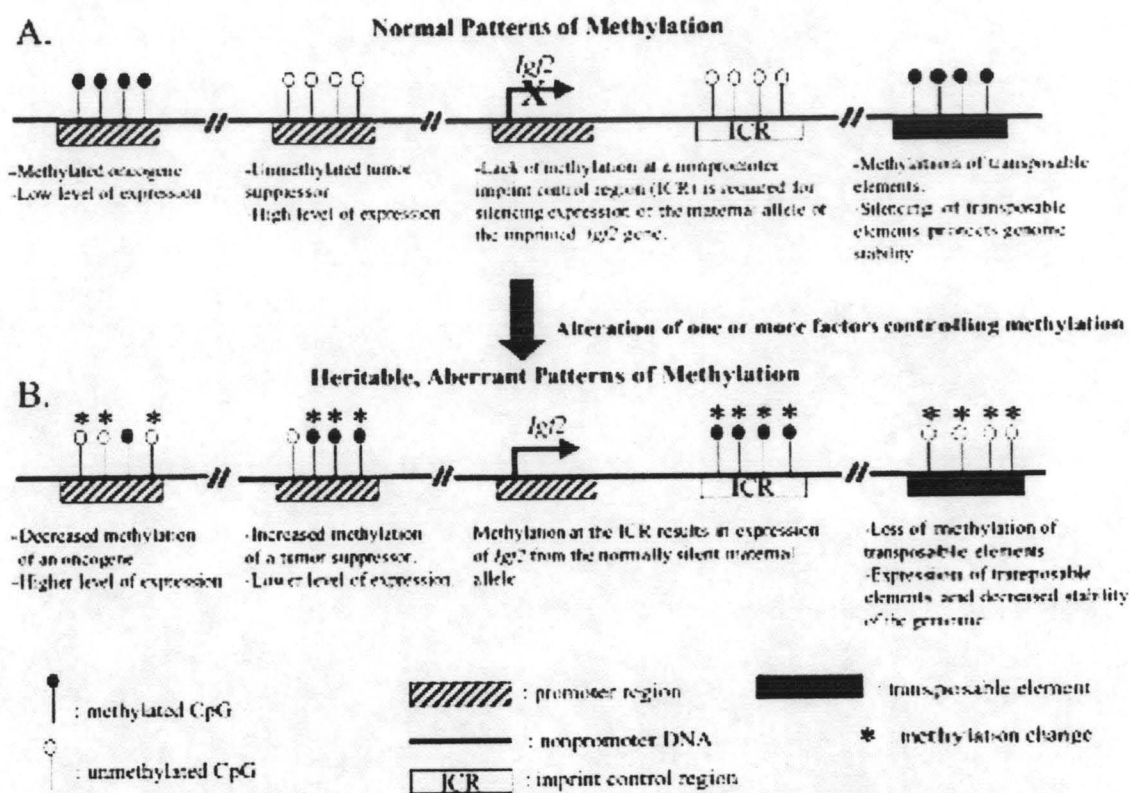
Figure 2-3 : Illustration of four possible alterations in normal patterns of methylation that may facilitate tumorigenesis. (*A*) Normal methylation patterns. (*B*) Aberrant methylation. A decrease in methylation of the promoter region of an oncogene can result in increased expression, whereas an increased methylation of the promoter region of a tumor suppressor gene can silence its expression. Altered methylation can affect imprinting. An increase in methylation of the imprint control region (ICR) of the imprinted oncogene *Igf2* may lead to expression of *Igf2* from the normally silenced maternal allele in addition to the expression that occurs normally from the paternal allele. Furthermore, a decrease in methylation of transposable elements can lead to their expression and, thus, contribute to genetic instability. Additionally, 5-methylcytosine may deaminate spontaneously to thymine, resulting in a C:G to T:A point mutation[26].
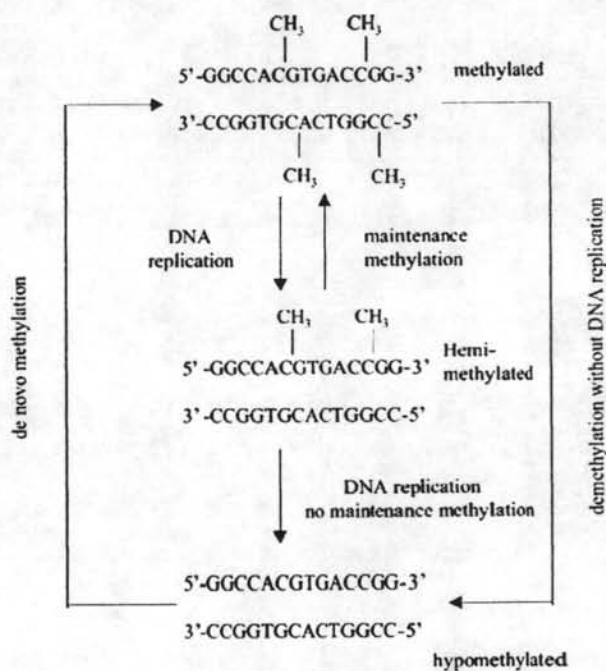
4. DNA METHYLATION

DNA methylation (i.e., the 5MeC content of DNA) is an important determinant of gene activity. In contrast to mutation, this does not involve a change in DNA base-coding sequence, i.e., both cytosine and 5-methylcytosine base pair with guanine. Altered DNA methylation which leads to aberrant gene expression due, in part, to affecting the ability of methylatedDNA-binding proteins to interact with their cognate cis

elements[32], may underlie some of the crucial changes in gene expression involved in carcinogenesis. There is a persuasive body of evidence indicating that differential methylation of DNA is a determinant of higher order chromatin structure[33] and that the methyl group provides a chemical signal recognized by transacting factors. Binding or lack of binding of these factors regulates transcription, e.g., by interfering with the ability of transcription factors to access their cognate cis elements. Thus,DNAmethylation is a mechanism whereby cells can control the expression of genes with similar promoter regions in the presence of ubiquitous transcription factors[34]. Transcriptional repression is dependent upon the density of methylation; it is not simply a case of being on or off[32, 35]. Furthermore, there is a direct causal relationship between DNA methylation-dependent transcriptional silencing and modification of chromatin. A particular methylated DNA binding protein, MeCP2, recruits histone deacetylase, facilitating the remodeling of chromatin and transcriptional repression. In light of the enzymatic steps involved, it may be expected that modifications in DNA methylation would result from threshold-exhibiting events, though this would have to be determined experimentally for each particular chemical of interest. Under normal conditions, DNA is methylated symetrically on both strands. When DNA replication occurs, 5-methylcytosine is not incorporated directly into the newly synthesized DNA strand. Consequently, the new double-stranded DNA contains hemimethylated sites that provide the signal for DNA maintenance methylase to transfer a methyl group from its cofactor, S-adenosylmethionine, to a cytosine residue on the newly synthesized strand. DNAmethyl transferase (Dnmt1) is the maintenance methylase responsible for propagating the parental pattern of methylation in daughter cells following cell replication[36]. If maintenance methylation does not occur (e.g., owing to a decrease in capacity or fidelity of DNAmaintenance methylase activity and/or decreased levels of S-adenosylmethionine) and cell division followed by a second round of DNA replication takes place, then that daughter strand will give rise to double-stranded DNA that has lost a methylated site. This epigenetic change is heritable. Demethylation without DNA replication and de novo methylation may also occur. Additionally, demethylation not linked to DNA replication may occur through the action of a 5-methylcytosine- DNA glycosylase and/or a demethylase that transforms 5-methylcytosine to cytosine. Dnmt 3a and 3b are the DNA methylases responsible for de novo methylation. Thus, maintaining

a normal methylation pattern depends on the interplay between maintenance methylation following DNA replication—demethylation not linked to DNA replication and de novo methylation. A disruptionof the normal methylation pattern can disrupt development and affect the phenotype in a fashion that may contribute to carcinogenesis (e.g., silencing of suppressor genes and/or facilitating increased expression of oncogenes). It is important to note the potential to undo alterations in methylation through the actions of the de novo methylases and/or demethylation not linked to replication may provide a mechanism for reversal of aspects of tumor promotion. Operational reversibility is a key feature of the promotion stage of carcinogenesis. The multiple factors involved in maintaining the normal methylation status of DNA are illustrated in Figure 2-4.

$$CH_3 \quad CH_3$$
5'-GGCCACGTGACCGG-3' methylated
3'-CCGGTGCACTGGCC-5'
$$CH_3 \quad CH_3$$

DNA replication   maintenance methylation

de novo methylation   demethylation without DNA replication

$$CH_3 \quad CH_3$$
5'-GGCCACGTGACCGG-3' Hemi-methylated
3'-CCGGTGCACTGGCC-5'

DNA replication
no maintenance methylation

5'-GGCCACGTGACCGG-3'
3'-CCGGTGCACTGGCC-5'

hypomethylated

Figure 2-4 : Maintenance of DNA methylation. Newly synthesized DNA is not methylated. Shortly after DNA replication, an S-adenosylmethionine (SAM)-requiring maintenance methylase recognizes hemimethylated sites and methylates cytosine at the 5-position to reestablish the original methylation pattern. A failure to maintenance methylate (e.g., due to decreased levels of SAM and/or inhibition of maintenance methylase during periods of cell proliferation) can result in daughter cells that contain hemimethylated DNA sites. The next round of replication can lead to cells containing

hypomethylated DNA, and this is heritable. Furthermore, there are opportunities for demethylation that are not linked to DNA replication and de novo methylation, which does not require a hemimethylated signal. Adapted from Hergersberg[34]. Methylcytosine residues are represented as C-CH3 [26].

4.1 DNA methylation and development

Dramatic changes in overall methylation of DNA occur at different periods of embryogenesis, development, and differentiation to adult cells [37]. The genome of the primodial germ cells of the embryo are not methylated to any extend. After glonadal differentiation and as the germ cell begin to develop, de novo methylation occurs leading to substantial methylation of the DNA of mammalian sperm and egg cells in Fig.2-4.The sperm genome is more heavily methylated than the egg genome. The genome of the fertilized oocyte is an aggregate of the sperm and egg genome and so it and the very early embryo are substantially methylated with methylation differences at paternal and maternal alleles of many genes. Later on, at the morula and early blastula stages in the preimplantation embryo, genome-wide demethylation occurs. A wave of demethylation initially erases presetmethylation patterns in the first days of embryogenesis. This is followed by several waves of de novo methylation that eventually establish adult patterns of gene methylation. In differentiated cells, methylation patterns change relatively little and are perpetuated after DNA replication through the high affinity of DNA methyltransferase for hemimethylated DNA

Evidence that DNA methylation is important in development comes from the observation that disrupting both *DNMT1* alleles in embryonic stem cells results in embryonic death[38]. Dnmt3a and Dnmt3b are also essential for mammalian development; homozygous Dnmt3a deficiency causes running and death at 4 weeks of age, while Dnmt3b deficiency is embryonic lethal[39].
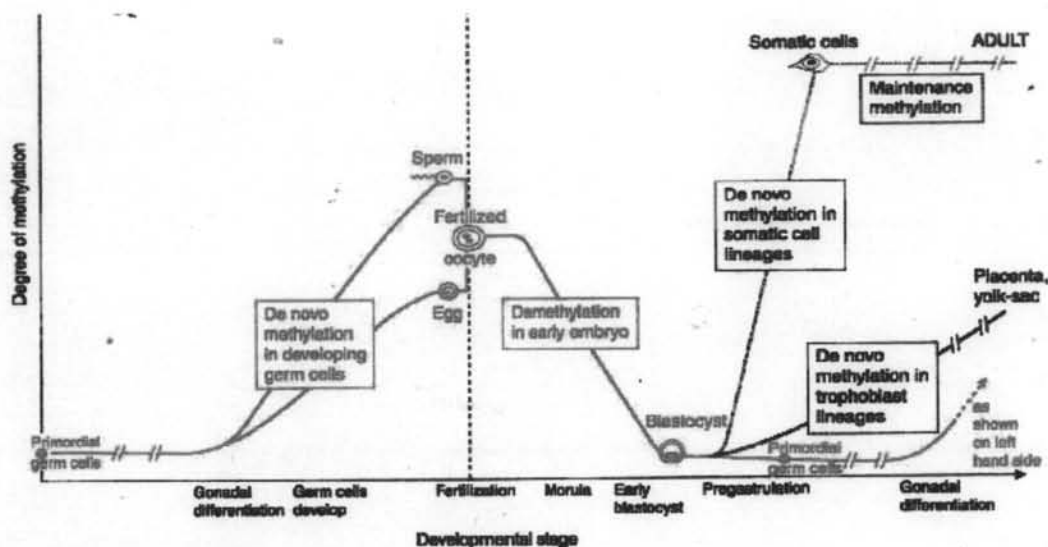
Figure 2-5. Changes in DNA methylation during mammalian development

5. DNA METHYLATION AND CARCINOGENESIS

5.1 Hypomethylation

Changes in DNA methylation are a consistent finding in cancer cells. Hypomethylation is an early event in carcinogenesis[40] and is observed very frequently. In addition, most metastatic human neoplasms have significantly lower genomic 5MeC than benign neoplasms or normal tissue, and the percentage of primary malignancies with hypomethylated DNA is intermediate between those of metastases and benign neoplasms. Furthermore, there is a direct relationship between DNA methylation of the promoter region and gene silencing. Hypomethylation of a gene is necessary, but not sufficient, for its expression. Therefore, a hypomethylated gene can be considered to possess an increased potential for expression as compared to a hypermethylated gene. Hypomethylation may facilitate the aberrant increased gene expression that plays a role in the transformation of a normal cell into a frank malignancy. One must keep in mind the fact that, with the exception of tumor suppressor genes, a mutated gene must be expressed in order to affect phenotype. It is interesting to note two recent publications indicate that marked hypomethylation can inhibit tumorigenesis[41, 42]. Since DNA methylation plays a critical role in development and differentiation, we believe it is reasonable to suggest that hypomethylation at an intermediate level does play a critical role in carcinogenesis, while excessive hypomethylation may not be compatible with the

life of the affected cells (e.g., owing to exceptional deregulation of gene expression). It then follows that the dead cells cannot lead to the formation of tumors[26].

5.2 The DNA methylation setting of healthy cells

The inheritance of information based on gene expression levels is known as epigenetics, as opposed to genetics, which refers to information transmitted on the basis of gene sequence. The main epigenetic modification in humans is themethylation of the cytosine located within the dinucleotide CpG. 5mC in normal human tissue DNAs constitutes 0.75%–1% of all nucleotide bases, and about 4%–6% of all cytosines are methylated in normal human DNA. CpG dinucleotides are not randomly distributed throughout the vast human genome. CpG-rich regions, known as CpG islands, are usually unmethylated in all normal tissues and frequently span the 5_-end region (promoter, untranslated region, and exon 1) of a number of genes; they are excellent markers of the beginning of a gene. If the corresponding transcription factors are available, the histones are in an acetylated and unmethylated state, and if the CpG island remains in an unmethylated state, then that particular gene will be transcribed. Of course, there are exceptions to the general rule.We can find certain normally methylated CpG islands in at least four cases: imprinted genes, X-chromosome genes in women, germline-specific genes, and tissue-specific genes. Genomic or parental imprinting is a process involving acquisition of DNA hypermethylation in one allele of a gene early in the male and female germline that leads to monoallelic expression. A similar phenomenon of gene-dosage reduction can also be invoked with regard to the methylation of CpG islands in one X-chromosome in women, which renders these genes inactive to avoid redundancy. Finally, although DNA methylation is not a widely occurring system for regulating "normal" gene expression, sometimes it does indeed accomplish this purpose, as with the genes whose expression is restricted to the male or female germline and not expressed later in any adult tissue, such as the MAGE gene family. Finally, methylation has been postulated as a mechanism for silencing tissue-specific genes in cell types in which they should not be expressed. However, it is still not clear whether this type of methylation is secondary to a lack of gene expression owing to the absence of the particular cell type–specific transcription factor or whether it is the main force behind transcriptional tissue-specific silencing. What is the significance of the presence of DNA methylation outside the CpG islands? One of the most exciting

possibilities for the normal function of DNA methylation is its role in repressing parasiticDNAsequences. Our genome is plagued with transposons and endogenous retroviruses acquired throughout the history of the human species.We can control these imported sequences with direct transcriptional repression mediated by several host proteins, but our main line of defense against the large burden of parasitic sequence elements (more than 35% of our genome) may be DNA methylation. Methylation of the promoters of our intragenomic parasites inactivates these sequences and, over time, will destroy many transposons. The perfect epigenetic equilibrium of the previously described normal cell is dramatically transformed in the cancer cell. The epigenetic aberrations observed can be summarized as falling into one of two categories: transcriptional silencing of tumor suppressor genes by CpG island promoter hypermethylation and global genomic hypomethylation[43].

5.3 Genomic hypomethylation of transformed cells

At the same time that certain CpG islands become hypermethylated, as discussed below, the genome of the cancer cell undergoes dramatic global hypomethylation. The malignant cell can have 20%–60% less genomic 5mC than its normal counterpart. The loss of methyl groups is accomplished mainly by hypomethylation of the "body" (coding region and introns) of genes and through demethylation of repetitive DNA sequences , which accounts for 20%–30% of the human genome. The degree of genomic DNA hypomethylation increases through all the tumorogenic steps, from the benign proliferations to the invasive cancers. How does global DNA hypomethylation contribute to carcinogenesis? Three mechanisms can be invoked: chromosomal instability, reactivation of transposable elements, and loss of imprinting. Undermethylation ofDNAmight favor mitotic recombination, leading to loss of heterozygosity as well as promoting karyotypically detectable rearrangements. Additionally, extensive demethylation in centromeric sequences is common in human tumors and may play a role in aneuploidy. It has been reported that patients with germline mutations in DNA methyltransferase 3b (DNMT3b) have numerous chromosome aberrations. Hypomethylation of malignant cell DNA can also reactivate intragenomic parasitic DNA, such as L1 (long interspersed nuclear elements, LINES) and Alu (recombinogenic sequence) repeats. These, and other previously silent transposons, may now be transcribed and even "jump" to other genomic regions where

they can disrupt normal cellular genes. Finally, the loss of methyl groups can affect imprinted genes. The best-studied case concerns the effects of the H19/IGF-2 locus on chromosome 11p15 in certain childhood tumors. However, we still know very little about the real role of DNA hypomethylation in the development of cancer cells. Is it really a "causative" factor? Or just a "modulator of cancer risk?" Or only a "bystander passenger?" The studies in mouse models are extremely interesting but puzzling: When the mouse deficient in DNA methylation owing to a defect in DNMT1 is crossed with the colon adenomaprone Min mouse (with a genetic defect in APC), the resulting mouse has fewer tumors ; but another DNMT1 defective mouse may have an increased risk of lymphomas. This paradox is an important question that needs to be addressed in the near future[43].

5.4 Hypomethylation of highly repeated, interspersed DNAsequences

Hypomethylation has been observed very often in DNA repeats in diverse cancers. The phenomenon of repeat induced gene silencing, which has been seen in mammals as well as plants and fungi[44], is probably related to the finding that mammalian DNA repeats tend to be highly methylated in postnatal somatic tissues. The repeats that display tumor-associated hypomethylation include endogenous retrotransposons. Retrotransposons or retroviral-derived elements can have their transcription upregulated in vivo by DNA demethylation. This was concluded from studies of Dnmt1 knockout mouse embryos, interspecies mammalian hybrids, and mice with an inherited epigenetically controlled phenotype whose expression is regulated by a genetically linked retrotransposon (IAP)[45] Also, there is evidence for frequent activation of expression of full-length transcripts from retrotransposons in certain types of murine cancer[46].

6. RETROTRANSPOSON : LONG INTERSPERSED NUCLEAR ELEMENTS, *LINES*
6.1 Overview of Retroelements

Although once thought of as "junk" DNA, the importance of interspersed elements in the genome has become increasingly appreciated in recent years. It has been estimated that at least one third of the mammalian genome consists of these elements in various forms[47]. In a broad sense they are collectively referred to as transposable elements, which encompass both transposons and retrotransposons .

Transposons have inverted terminal repeats, encode a transposase activity, and move from one site to another through a "cut and paste" mechanism[48]. Retrotransposons, which move by a "copy and paste" mechanism, proceed through an RNA intermediate largely dependent on their encoded reverse transcriptase activity. However, they may utilize the host's reverse transcriptase[49]. In this manner a copy of the original can be integrated into a new genomic location. Therefore, stability of the genome depends upon keeping these movable and amplifiable elements transcriptionally repressed. DNA methylation plays a key role in the regulation of gene expression overall, including keeping transposable elements transcriptionally silent [50].

6.2 Retrotransposable Elements: LINE-1

Retrotransposable elements are categorized as either autonomous or nonautonomous elements, where autonomous refers to the property of self-sufficiency for mobility. There are two classes of autonomous elements: long terminal repeat (LTR) and non-LTR retrotransposons. Similar in structure to retroviruses, although lacking a functional env gene, LTR retrotransposons encode proteins necessary for retrotransposition. Likewise, non-LTR retrotransposons also encode a reverse transcriptase and endonuclease that play a role in their ability to mobilize themselves and other non-autonomous elements. A basic difference between the LTR and non-LTR retrotransposons is their method of recombination. LTR retrotransposons move by first being transcribed into RNA, followed by reverse transcription leading to a DNA copy that recombines with genomic DNA. Non-LTR retrotransposons move through a somewhat different RNA-mediated event, discussed below. Up to several kb in length, the non-LTR retrotransposons are commonly referred to as long interspersed nuclear elements (LINEs). LINE-1 repeats constitute about 15% of the human genome, but of the about 4×10^5 copies of LINE-1 elements in the human genome, only about 30-60 are estimated to be competent for transposition Structurally, they contain an internal promoter for RNA polymerase II, a 5' untranslated region (UTR), two open reading frames (ORFs), and a 3' terminal polyadenylation site. The ORF1 protein is an RNA binding protein, while ORF2 encodes both a reverse transcriptase and a DNA endonuclease (Fig 2-5.)[49].
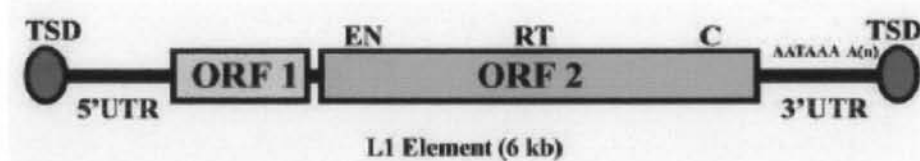
L1 Element (6 kb)

Figure 2-6. Non-LTR elements contain an internal promoter for RNA polymerase II, a 5' untranslated region (UTR) and a 3' deoxyadenosine (A)-rich tract(49).
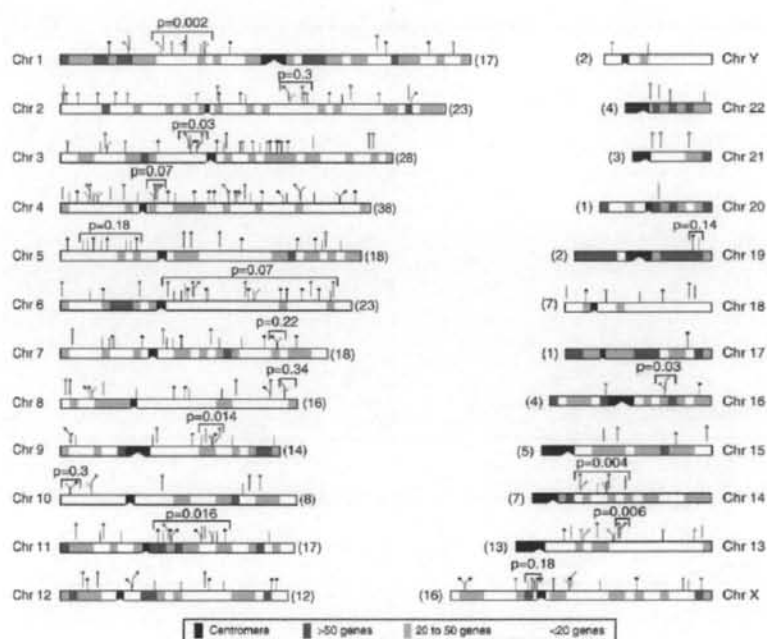

Figure 2-7 : Insertion sites of LINE-1 Ta-1 elements in the human genome. Ta-1 integration sites are shown as tick marks above each chromosome. Tick marks with thinner lines impinging directly on the chromosome diagrams are those identified in the human genome database. Tick marks with heavier lines offset from the chromosomes are the ones that we cloned. Solid circles indicate polymorphic inserts, open circles indicate fixed ones, and tick marks without either are indeterminate (not amenable to PCR, see Methods and Table 1). The number in parentheses indicates the number of Ta-1 insertion sites on each chromosome. Chromosome 15 only shows the positions of 4 of the 5 Ta-1 elements on this chromosome because one of them was located on an unassigned segment. The shaded boxes indicate the number of known genes per 5 Mb segments[50].

6.3 Insertion of active LINE-1 retrotransposon in humans

Figure 2-6 shows the chromosomal distribution of 295 LINE-1 Ta-1 elements. They were found on every chromosome and their number per chromosome ranges from 38 on chromosome 4 to one on chromosomes 17 and 20. Large chromosomes have on average more LINE-1 Ta-1 inserts than smaller ones, and the number of inserts on each chromosome is nearly proportional to the length of the chromosomesHowever, chromosome 4 is a clear exception to these generalizations as it contains significantly more LINE-1 Ta-1 inserts than expected for either its length or gene density. The enrichment of LINE-1 Ta-1 elements on chromosome 4 is not because it is uniquely hospitable to L1 insertions. In addition, examination of Figure 2-6 suggests that some LINE-1 Ta-1elements may be clustered on some chromosomes. However, even random insertion of LINE-1 Ta-1 elements could yield apparent clusters of inserts that could bemisinterpreted as insertional "hot spots." Therefore, we assessed the statistical significance of any apparent cluster by testing it against the random (uniform) insertion model. Seemingly over-long gaps (i.e., genomic stretches without insertions) may also be observed, and were also tested for statistical significance[50].

6.4 Genomic Consequences of Long Interspersed Nuclear Element-1 (LINE-1)

At the forefront of genomic consequences due to retrotransposon expression and movement is insertional mutagenesis. Insertion of these elements, whether random or targeted, represents a mutation, and therefore, retrotransposition poses a clear risk to the stability of the genome. Not only is movement of these elements critical but also their capability to transduce surrounding DNA sequences. At times this may promote genomic diversification(exon shuffling)[51], but more apparent is the possible contribution to mutagenicity. On a larger scale, fully LINE-1 and their transduced sequences can result in chromosomal rearrangements[52]. Medically, muscular dystrophy, characterized by a progressive loss of muscle strength in humans, has been associated with a LINE-1 insertion within exon 48 of the dystrophin gene[53]. These findings supported recent LINE-1 retrotransposition activity and directly demonstrated the consequential toxicity associated with the aberrant regulation of these elements. Additionally, altered regulation of gene expression by insertion of LINE-1 elements as a direct mutation has been documented numerous times. cited twenty-one examples of sequence element inclusion from Drosophila, sea urchin, human, and mouse genomes that serve a function in terms of transcriptional competency. Counterpart to insertions are deletions and

duplications, which can arise from unequal crossing-over and mispairing of homologous LINE-1 sequences . As much as a 3% frequency of DNA deletions due to LINE-1 retrotransposition has been proposed. Gilbert et al. (2002) observed a large deletion of the genomic DNA following the retrotranspositional event. A common mechanism preceding this deletion, among other alterations, was shown to involve cleavage of the genomic top strand. Variations of this model also suggest that chimaeric LINE-1s, large deletions, and long duplications are also possible[54]. Clinically, inactivating mutations arising from LINE-1-mediated recombination can lead to the accumulation of mutations in specific target genes during cancer and development[55]. This highlights the fact that LINE-1 elements are capable of reshaping the genome through direct mutation.
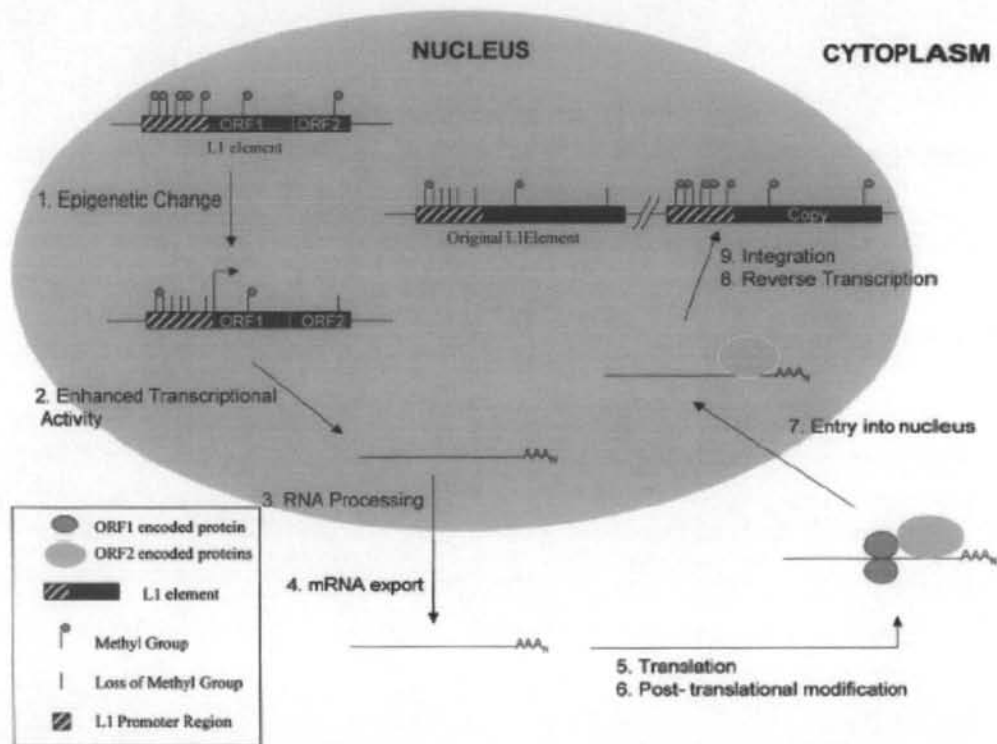

Figure 2-8 : Schematic representation of an epigenetic change as a precursor to expression and movement of retrotransposable elements. (1) An epigenetic change, e.g., hypomethylation of the retrotransposable element allows for (2) enhanced transcriptional activity. (3) RNA processing and (4) mRNA export ensue. (5) Translation and (6) posttranslational modification precede the formation of a ribonucleoprotein particle in which ORF1 and ORF2 encoded proteins are associated with the original

mRNA. (7) Once entry into the nucleus has occurred, (8 and 9) reverse transcription and integration are achieved via the encoded reverse transcriptase and endonuclease through a mechanism termed target primed reverse transcription (TPRT)[56].

6.5 Relationship of DNA methylation and LINE-1

It is instructive to consider the role of altered methylation as an epigenetic mechanism for the activation of retrotransposable elements leading to their expression and possible retrotransposition (Figure 2.7)[56]. LINE-1 sequences in the human has been characterized structurally as having a promoter region which controls the expression of the two open reading frames, ORF1 and ORF2. Expression of these regions, encoding an RNA binding protein, a reverse transcriptase, and an endonuclease, is required for integration of a new copy of the original element into the genome. Given the distribution of these elements, both their movement and expression can lead to unstable conditions within the genome. Therefore, an important aspect of DNA methylation is its connection to the host-defense system, which acts to offset the threats from these largely parasitic sequences by maintaining them in a methylated, transcriptionally silent state. Genome instability is a common feature of tumorigenesis[57]. There have been occasional reports of cancer-associated retrotransposition-like insertions involving LINE-1 sequences [58], and they may mobilize cellular RNAs at low frequencies[59]. Their activation can also lead to transcriptional interference involving neighboring genes[60]. However, retrotransposition of endogenous elements is implicated in disease much less frequently for humans than for mice[61]. Extensive hypomethylation results in genome instability reflected by an increase in mutation frequency[62]. Hypomethylation-induced transcriptional activation of LINEs contributes to this instability[63]. Furthermore, hypomethylation of LINE-1 sequences has been observed in various cancers. LINE-1 hypomethylation was observed in chronic lymphocytic leukemia vs normal mononuclear blood cells[64], urinary bladder carcinomas compared to normal bladder (4), hepatocellular carcinomas vs non-tumorous 'normal' or cirrhotic tissue[63], and prostate carcinomas vs normal prostate and other normal tissues[65]. In a chemically induced mouse hepatoma, LINE-1 hypomethylation was also seen[66]. That study involved methylation-sensitive representational difference analysis (MS-RDA) to survey HpaII-sensitive fragments in the tumor vs normal liver, with conformation of the results by Southern blotting[66]. LINE-1 hypomethylation was the only alteration seen

repeatedly in the MS-RDA banding patterns from different tumors. Hypomethylation, including of single-copy DNA sequences, was observed more frequently than hypermethylation. In these studies of murine and human tumors, it was not demonstrated that hypomethylation of LINE-1 sequences increased the use of these repeats as transcription units in cancer, but it has been hypothesized that this hypomethylation might sometimes lead to the retrotransposition of the very small percentage of retrotransposition-competent copies of LINE-1[65]. This led to the speculation that demethylation of LINE-1 sequences may promote genomic instability and facilitate tumor progression.

6.6 Examples of TE influences on gene regulation

Over the last 15 years, an abundance of experimental evidence has accumulated that directly points to the contribution of repetitive DNA to gene regulation. This evidence consists largely of examples where TEs have been shown to contribute to the regulation of a host gene by providing cis-regulatory sequences that interact with host trans factors. Interestingly, the vast majority of these cases were uncovered fortuitously in the sense that the investigators were not out to assess the role of TEs in gene regulation but rather were seeking to understand the molecular basis of the regulatory properties of the particular system that they were working on. The first example of this kind came from the study of the sex-limited protein (Slp) encoding gene in mouse [67]. Slp is one of two tandem genes and is closely related to the adjacent C4 gene that encodes the fourth component of complement. Apparently, after the duplication of these two genes an endogenous retrovirus (ERV) inserted upstream of the Slp gene and this insertion resulted in an altered expression pattern for Slp which in turn drove the functional divergence of the protein [68]. Unlike the C4 gene, Slp is expressed only in males due to androgen dependence conferred by androgen response elements found in the long terminal repeat of the ERV [69]. Pursuant to his interest in the relationship between repetitive DNA and gene regulation, Britten reviewed a number of such cases where insertions of TEs have resulted in fixed novel regulatory patterns and established four criteria for the identification of convincing examples: 1 – the presence of a known TE sequence in the gene region, 2 – evidence that the insertion has been present long enough to be fixed, 3 – evidence that part of the TE sequence participates in the regulation of the nearby gene and 4 – evidence that the gene encodes some

function [70]. By 1997, Britten was able to find more than 20 examples that conformed to all four of these criteria and many more similar examples have been uncovered since that time. For instance, a number of cases where human TEs can be shown to serve as promoters for adjacent genes have recently been identified [71, 72]. The most extensive literature survey to date of TE contributions to host gene regulation identified almost 80 cases where regulatory elements of vertebrate genes are derived from TEs [73]. In addition to serving as promoter and enhancer sequences for nearby genes, TE insertions have also been shown to influence host gene expression by providing alternative splice sites[74, 75] and polyadenylation sites[76, 77]. L1 forms a component of most mammalian transcription units, but the effects of these primarily intronic inserted sequences have not been studied carefully. About 79% of human genes are estimated to contain at least one segment of L1 in their transcription unit, and L1 segments from pre-existing and newly derived insertions usually contain L1 ORF2. As these sequences are mostly intronic, it has been assumed that the extra sequences are spliced out and do not affect target gene expression. Through a combination of transcriptional elongation inhibition and premature polyadenylation, L1 insertions in either orientation can affect the RNA production of endogenous genes, both qualitatively and quantitatively. In human genome model in which L1 has led to numerous subtle but potentially significant transcriptome alterations[78].

7. Background of the experiment approach

This study has designed a new method to be able to study specific LINE-1 loci methylation status from carcinogenesis sample group. This PCR uses principle of COBRA (combined bisulfite restriction analysis) method[79]. Previously, reports of LINE-1 genomic DNA methylation status were performed by COBRA LINE-1, Southern blotting and methylation sensitive enzyme (MS-PCR)

Disadvantages of COBRA LINE-1, Southern blot analysis and MS-PCR

- COBRA LINE-1 can't identify each LINE-1 methylation status.
- Southern blot analysis required a large amount of DNA.
- Southern blot analysis required a high quality of DNA.
- Mix type of tissue from fresh specimen.

- Incompatible with DNA isolated from paraffin sections.

- Methylation-sensitive restriction enzyme digestion followed by PCR is prone to false-positive results since even low levels of spurious incomplete digestion can result in a PCR product. This problem is exacerbated in samples derived from paraffin sections.

Advantages COBRA unique to LINE-1 method

- compatibility with paraffin sections.

- quantitative accuracy.

- applicability to large numbers of samples.

- COBRA circumvents incomplete digestion by restriction digestion of a purified PCR product, rather than of the original genomic DNA

7.1.1 COBRA unique to LINE-1(CU-L1)

Methylation-dependent sequence differences are introduced into the genomic DNA by the standard sodium bisulfite treatment and then PCR amplified . This combination of bisulfite treatment and PCR amplification results in the conversion of unmethylated cytosine residues to thymine and methylated cytosine residues to cytosine. This sequence conversion can lead to the methylation-dependent creation of new restriction enzyme sites or it can lead to the methylation-dependent retention of pre-existing sites such as *BstUI* (CGCG). The general outline of the method is depicted in Fig 2-7. COBRA unique to LINE-1 develop from COBRA LINE-1 by replace 5' oligo of LINE-1 with 5' upstream unique sequence of LINE-1s. The COBRA unique to LINE-1 PCR technique is designed to evaluate CpG nucleotide at 270 and 285 of LINE 1.2, an active LINE-1 locus. The sequence at position 267-270 will be AAT (T/C) and 284-287 will be T (T/C) GA. Thus TasI will digest hypomethylation sequence at 270 and TaqI cut de novo methylation at 285 in figure 2-8. The summation of proportion of TaqI and TasI digested fragments is always close to 100%. From preliminary study interestingly, this COBRA unique to LINE-1 technique has low failure rate of PCR from paraffin embedded tissue in which the little amount of DNA usually degrades. The underlining reasons of this improvement should be due to the shorter of the amplicon size and the significantly

larger in number of LINE-1approximately 3-4,000 full-length copies per cell, as the DNA template.

COBRA - *Co*mbined *B*isulfite *R*estriction *A*nalysis

| Bisulfite-PCR | ⟹ | Quantitation |

$$C \rightarrow U \rightarrow T$$
$$^{5m}C \rightarrow {}^{5m}C \rightarrow C$$

Restriction Digestion
PAGE Gel
Oligo Hybridization
Phosphorimager Quantitation

Example:

Methylated *Bst* UI Site	Unmethylated *Bst* UI Site
$^{5m}CG^{5m}CG$	CGCG
↓	↓
CGCG	TGTG
Bst UI Site Retained	*Bst* UI Site Lost

Sample 1   Sample 2   Sample 3
- BstUI   - BstUI   - BstUI

Methylation: 0 %   50 %   100 %

$$\% \text{ Methylation} = 100 \times \left(\frac{C}{B + C} \right)$$
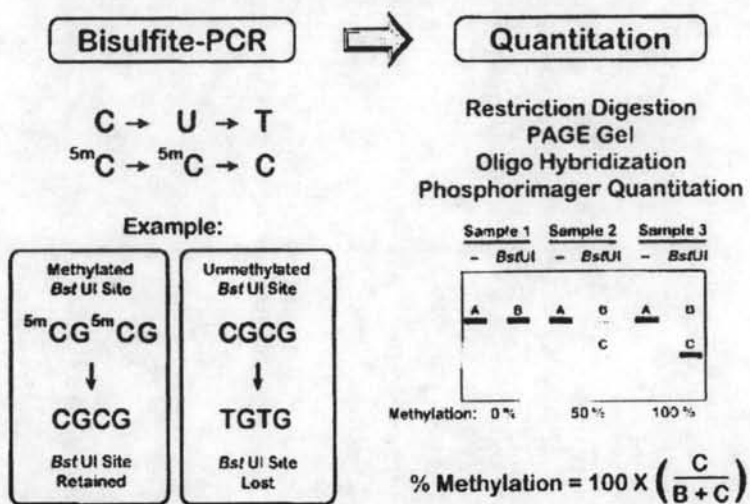
Figure 2-9. Outline of the COBRA procedure. COBRA consists of a standard sodium bisulfite PCR treatment followed by restriction digestion and quantitation