



Universiteit Utrecht

Ittai Bernardo Muller

Student no. 3257886

**DNA methylation: An important
mechanism in colorectal carcinogenesis
and pretumor progression**

June 2011

Biomedical Sciences
University Utrecht

Thesis

This thesis will give a description of the pathology of colorectal cancer and the implications of epigenetics in colorectal cancer development and pretumor progression.

Introduction

Colorectal cancer (CRC) is the third most common cancer diagnosed in both men and women in the United States, afflicting more than 100.000 people every year. It is also the third highest cause of death by cancer, killing over 50.000 people a year¹. At least 50% of the Western Population develops a colorectal tumor by the age of 70 and progression to malignancy ensues in 10 percent of these individuals².

As with all cancers, CRC is a disease of the genes primarily caused by mutations. In the last decade, however, studies suggest an increasing role for epigenetic mechanisms underlying the pathology of CRC. Especially DNA methylation, the addition of a methyl-group to cytosines in the DNA, seems to play an important part in colorectal tumorigenesis. It has been proposed that CRC arises from stem cells, since these are the only cells in the colorectum with a long enough lifetime to collect mutations. It has been suggested that tumorigenesis starts with accumulating phenotypically invisible alterations. This process is called pretumor progression and its theory is supported by the fact that CRC is usually observed in patients over 50 years of age, without phenotypic changes prior to that³. This thesis will elaborately describe the pathology of CRC and the recent advances in gastrointestinal research concerning DNA methylation and stem cell dynamics.

Carcinogenesis

Colorectal cancer, like all cancers arises from the accumulation of nonlethal genetic damage. Four classes of regulatory genes are the principal targets of this genetic damage: the growth-promoting proto-oncogenes, the growth-inhibiting

tumor suppressor genes, genes that regulate programmed cell death (apoptosis), and genes involved in DNA repair. Proto-oncogenic mutant alleles are considered dominant because they transform cells despite their normal counterpart. However, both alleles of the tumor suppressor genes must be mutated for transformation to occur. There are exceptions: some tumor suppressor genes lose their function when a single allele is inactivated or lost. This “two-hit hypothesis” was first proposed by Knudson *et al.*⁴ and supported by the fact that patients with germline mutations of the *adenomatous polyposis coli (APC)* gene do not necessarily develop colorectal cancer. APC is considered a gatekeeper gene for the initiation of colorectal cancer and a second “hit” is necessary for APC to lose its gatekeeper function².

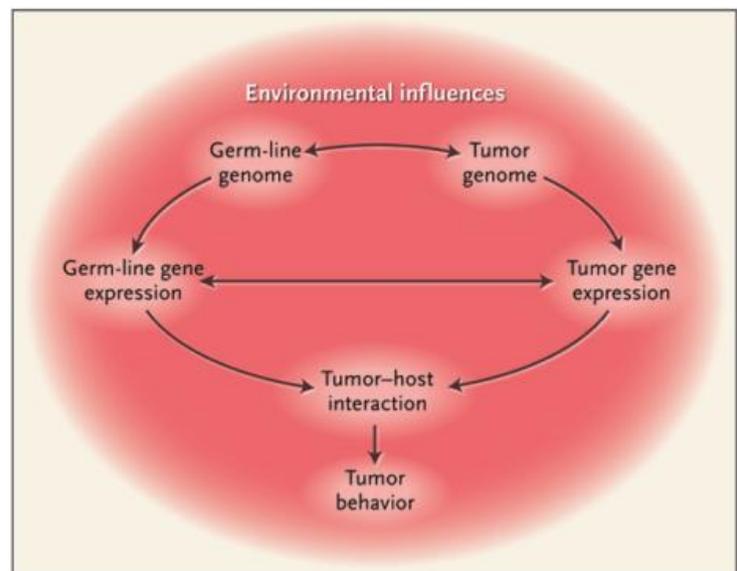


Figure 1 adapted from reference 5.

Molecular events that drive colorectal carcinogenesis. This process involves interactions among environmental influences, germline genomic factors and accumulated somatic changes in colorectal epithelium

The initiation, promotion and progression of CRC occur on many interrelated levels. While germ-line mutations are among the best described ones, somatic mutations as well as environmental influences play a large role (Figure 1). This multifactorial process lies at the basis of CRC, emphasizing that germline mutations alone do not necessarily cause CRC and that inactivation of corresponding second alleles and accumulating somatic mutations must first occur.

Colorectal carcinogenesis can be driven by genomic instability. Genomic instability is the affliction where the integrity of the genome is compromised.

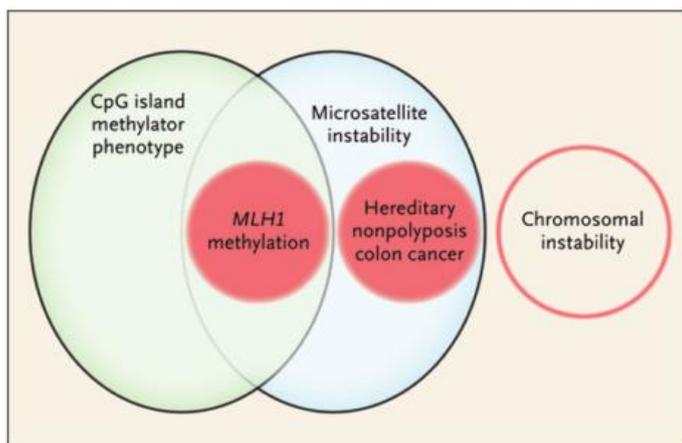


Figure 2 adapted from reference 5.

Genetic instability pathways that drive colorectal carcinogenesis.

Shown are the different pathways that underlie colorectal neoplasia: chromosomal instability, microsatellite instability due to defects in DNA mismatch-repair genes that are either caused by inheritance (hereditary nonpolyposis colon cancer) or somatically acquired (aberrant methylation and epigenetic silencing of MLH1), and the CpG island methylator phenotype.

Three distinct mechanisms have been proposed: Chromosomal instability (CIN), DNA mismatch-repair defects or microsatellite instability (MSI), CpG island methylator phenotype (CIMP)⁵. These

examples of genomic instability can be found in most colorectal cancers and are often the cause of the required “second hit”. It is important to emphasize that CRC is not a single disease but a heterogeneous complex of diseases and that each CRC patient has a unique disease caused by distinctive genetic and epigenetic backgrounds. While this distinction implies that no two CRC are the same, the molecular classifications described earlier can be made to similar CRC phenotypes to try to predict the pathogenesis and behaviour of similar cancers⁶.

Chromosomal instability

Approximately 70%–85% of CRCs develop via the CIN pathway⁷. Genetic aberrations occur through the accumulation of numerical or structural chromosomal abnormalities. These chromosomal abnormalities consist of losses or gains of whole chromosomes and chromosome translocations. Alterations in chromosome numbers (aneuploidy) are found in nearly all major tumors. An example of chromosome translocation is the Philadelphia chromosome in chronic myelogenous leukemias; the carboxy terminus of the *c-abl* gene on chromosome 9 is fused with the amino terminus of the *BCR* gene on chromosome 22.⁸

The molecular basis of CIN is fairly unknown. Aneuploidy (and an underlying CIN) is nearly ubiquitous in cancers which could mean that CIN results simply from the abnormal structure and growth properties of the cancer cell. However, studies have shown that CIN generally does not occur in cancer cells exhibiting the MSI phenotype, and that two fused diploid MSI cell lines, resulting in a tetraploid cell, remain chromosomally stable. This suggests a specific origin for

MSI other than a characteristic of the cancer itself⁹.

CIN is suggested to be an early step in colorectal carcinogenesis, and together with the evidence that specific patterns of loss and gains of chromosomes occur in this process, is consistent with the idea that CIN is pathogenic in CRC¹⁰. The aforementioned APC gatekeeper gene is suggested to be a cause of CIN. Even though some well-characterized colon cancer cell lines with APC mutations are diploid and show chromosomal stability, it is thought that APC inactivation allows a permissive state that allows the tumor to tolerate aneuploidy. APC is thought to aid in stabilizing microtubules during proliferation, which might explain why APC affects chromosomal instability⁷. Subsequently, it has been reported that aneuploidy is indeed the result of an abnormally high rate of CIN, persisting throughout the life of the tumor. Studies with fluorescence *in situ* hybridization (FISH) of interphase cells with a panel of centromeric probes on colon cancer cell lines revealed a 10-100 fold increase in the rate of chromosome loss or gain per generation of aneuploid cell lines compared to diploid colon cancer cell lines. Lengauer also showed that CIN is not caused by aneuploidy and that CIN is an autosomal-dominant trait, suggesting it arose from a gain-of-function mutation⁹. The mechanisms responsible for CIN are, however, still not fully understood.

The clue to understanding the molecular basis of CIN was provided by studies of chromosome instabilities in unicellular organisms. A large number of genes that, when altered, gave rise to the CIN phenotype in these organisms. Among these genes are genes that influence chromosome condensation, sister-chromatid cohesion, kinetochore structure and function and centrosome/microtubule formation and

dynamics, as well as “checkpoint” genes that monitor the progression of the cell cycle.⁸ Several of these mitotic checkpoint genes have been identified in breast cancer and leukemia, however similar discoveries are still to be found in CRC¹¹. However, a panel of 19 colorectal cancer cell lines have not yet been proven to exhibit mutations in any of these genes¹².

Recent studies have suggested an important role for hCDC4 in chromosomal instability. *hCDC4* mutations have been identified in 22 chromosomally unstable colorectal lesions and targeted disruption of this gene in karyotypically stable colorectal lesion causes a CIN-like phenotype, implying its role as a possible cause for chromosomal instability.¹³

Microsatellite instability.

A different form of genomic instability arises from defects in the DNA mismatch-repair system. The DNA mismatch repair system consists of several proteins that recognize and repair base-pair mismatches in the DNA. MSI occurs when genetic or epigenetic inactivation impairs this system. It is characterized as frameshift mutations in microsatellite repeats throughout the genome⁷. Sporadic tumors with MSI account for 13% of the total colorectal cancers in the Western world whilst 2-3% are assigned to the hereditary form called Hereditary non-polyposis colorectal cancer(HNPCC).

Discovery of the MMR system started with research in prokaryotes, where the MMR system consists of a family of enzymes that detect S-phase DNA replication errors. Sporadically, DNA polymerase makes errors when incorporating the correct number of bases during DNA replication. These errors can be recognized by the MMR system. When these faulty DNA sequences are not repaired during the second round of replication, the original strand, the

parental strand is copied and the daughter strand will contain a mutation. The errors can consist of single base-pair mutations or an insertion-deletion loop. The latter will result in a downstream nonsense mutation and the corresponding gene will produce a nonfunctional protein¹⁴.

Bacterial genes *mutS* and *mutL* were identified as the cause of MSI in prokaryotes and subsequently, yeast homologues became the object of interest, as they were hypothesized to be the cause of MSI in eukaryotes. The yeast homologues of the bacterial *mutS* and *mutL* genes were cloned and given the names Mut S homologue (MSH) and Mut L homologue (MLH). The yeast *MLH* gene has 4 mammalian homologues: *MLH1*,

MLH3, *PMS1*, and *PMS2*. Mammalian homologues of the yeast gene *Mut S* are: *MSH2*, *MSH3* and *MSH6*¹⁴.

In contrast to the prokaryotic MMR system, the 7 mammalian MMR proteins form specific heterodimers instead of homodimers. These heterodimers all have specific functions. MSH2–MSH6 (MutS α) recognizes single base-pair mismatches and repairs them by binding MLH1 and PMS2 subsequently. The same heterodimer initiates a different reaction where it binds exonuclease-1, proliferating cell nuclear antigen, and DNA polymerase and excises the daughter strand straight back to the mismatch site. The third mechanism involves MutS β , consisting of MSH2 and MSH3, which can

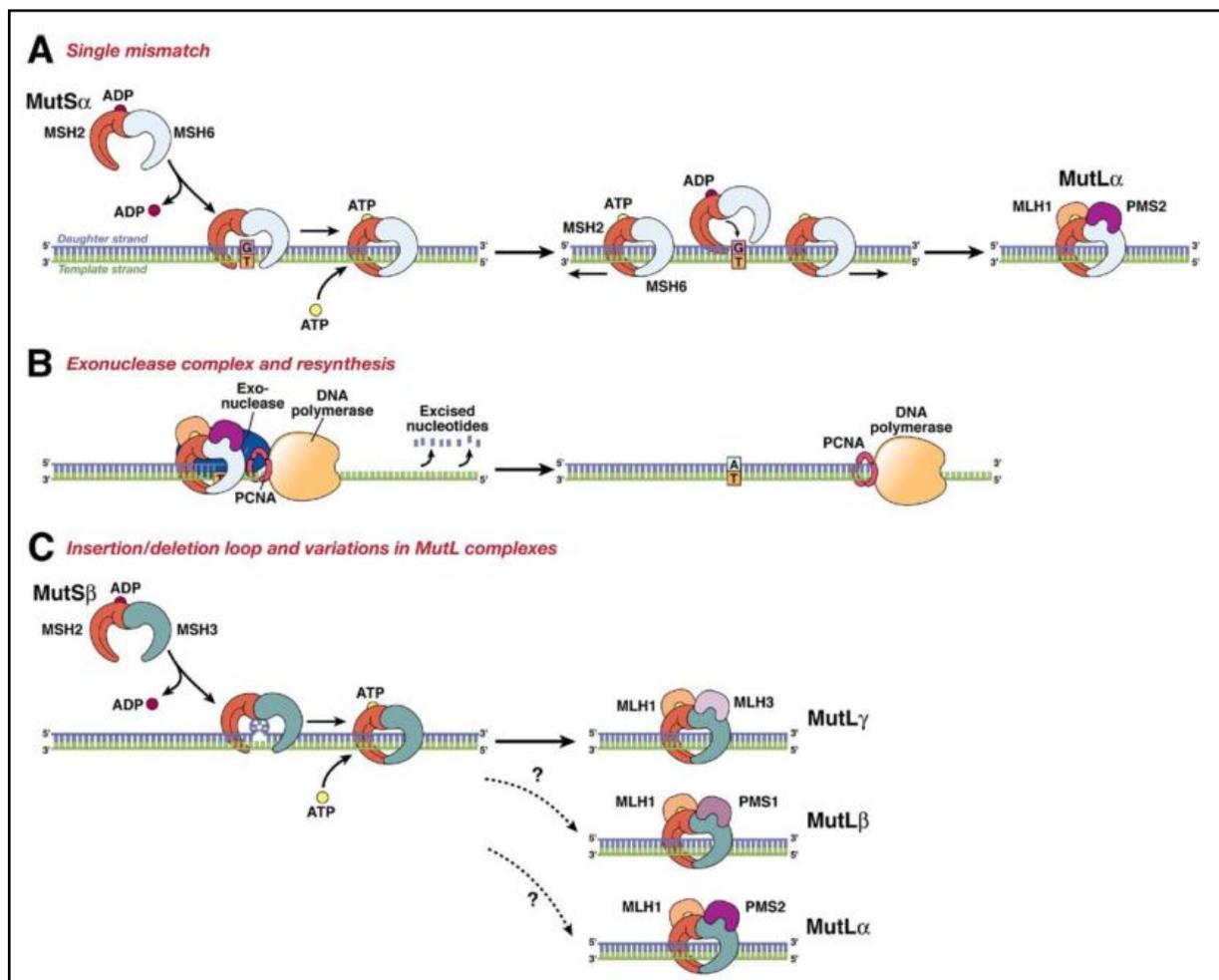


Figure 3. Adapted from reference 14

Three different types of MMR mechanisms are shown: (A) single mismatch repair, (B) exonuclease mechanisms and resynthesis and (C) insertion-deletion loop repair. MutL γ , MutL β and MutL α are heterodimers of mutL homologues MLH1, MLH3, PMS1 and PMS2. Question marks indicate mechanisms not yet reported but likely similar to MutL γ function.

repair larger insertion-deletion loops (Figure 3)^{14,15}.

Assigning MSI to colorectal cancers consists of recognizing frameshift mutations for at least 2 of 5 selected loci which are: two mononucleotide (BAT25 and BAT26) and three dinucleotide microsatellites (D5S346, D2S123, and D17S250). These loci are internationally accepted by a National Cancer Institute consensus conference. MSI-High or MSI-H is assigned when 2 or more loci exhibit MSI. MSI-Low or MSI-L when 1 locus exhibits MSI. Microsatellite stable or MSS colon cancers show no MSI at any of the loci¹⁶. The distinction between MSI-L and MSS is still under debate. Studies have shown that virtually all CRCs display some degree of MSI when tested against a large enough number of markers¹⁷.

Whilst HNPCC causes the pure form of MSI, characterized by germline mutations in *MLH1*, *MLH2*, *PMS2* and *MSH6*, the majority of CRCs that have the MSI-H phenotype are caused by an inactivating promoter methylation of the *MLH1* gene and consequently transcriptional silencing of *MLH1* expression^{2,15}. This inactivating promoter methylation provides MSI-like phenotypes without germline mutations. The underlying process is not yet fully understood. However, promoter methylation as a means to silence genes has been researched thoroughly in the last decade, suggesting a novel mechanism in colorectal carcinogenesis.

CpG island methylator phenotype.

CIN and MSI have long been recognized as pathogenic factors for colorectal carcinogenesis but recently a third pathway was hypothesized. The CpG island methylator phenotype, or CIMP, is a subclass of colorectal cancers where the “second hit” in Knudson’s model of tumor suppressor inactivation is provided by promoter hypermethylation. Studies have

shown that a distinction can be made between age-related and cancer-related methylation, suggesting a novel cancer phenotype¹⁸. This cancer phenotype is characterized by a high degree of methylation and displays “epigenetic instability”¹⁹.

While DNA primarily consists of four bases – adenine, guanine, cytosine and thymine – a covalent modification produces a “fifth base”. A methyl group is added to the cytosine ring by enzymes called DNA methyltransferases (DNMTs) using S-adenosyl-methionine as a methyl donor (Figure 1). This results in a 5-methyl cytosine base and this reaction is only possible on cytosines that precede a guanine base in the DNA sequence (the CpG dinucleotide)²⁰.

The CpG dinucleotides are divided asymmetrically throughout the genome

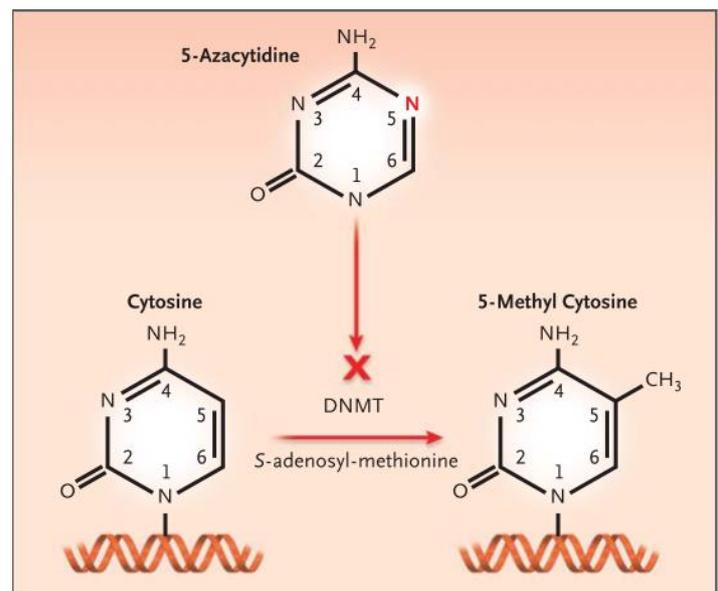


Figure 4. Adapted from reference 20

DNA methylation. DNA methyltransferases use S-adenosyl-methionine as a methyl-donor to covalently bond a methyl group to a cytosine base. The resulting molecule is 5-methyl cytosine. 5-Azacytidine is capable of inhibiting this reaction.

and most of the DNA methylation occurs in the noncoding regions of the DNA. The bulk of the CpG dinucleotides in the

genome (80%) reside in the noncoding regions and are methylated while the CpG dinucleotides associated with CpG islands are usually unmethylated. Methylation in the noncoding regions is thought to prevent binding of transcription factors in areas of repetitive DNA repeats outside of the exons. The importance of DNA methylation as a silencing agent is proven by the fact that promoter regions of several transcriptionally silenced genes on maternal X-chromosomes are fully methylated. This promoter methylation causes the “imprinting” of maternal alleles so that only one allele is expressed in normal tissue²⁰.

Methylation of DNA promoter regions results in silencing of genes. Clusters of CpG nucleotides, called CpG islands, are usually found in the promoter region of a gene. Regional hypermethylation of these CpG islands results in transcriptional

block, possibly through chromatin condensation, which prevents binding of transcription factors and causes gene silencing(Figure 5)²⁰⁻²².

It is also reported that not only does hypermethylation occur in the CpG-rich promoter regions of the DNA, global DNA hypomethylation in the noncoding regions occurs as well. This global effect may cause an increase in chromosomal instability, overexpression of a variety of proto-oncogenes(possibly due to *loss of imprinting (LOI)*), and increased mutation rates, suggesting that hypomethylation of the noncoding regions also aids in tumorigenesis²²⁻²⁴.

While CIMP is mostly considered a cancer phenotype, much of these aberrant CpG methylation events can be assigned to aging. In colonic adenomas and carcinomas, the *ER* gene, mostly associated with breast cancers, was

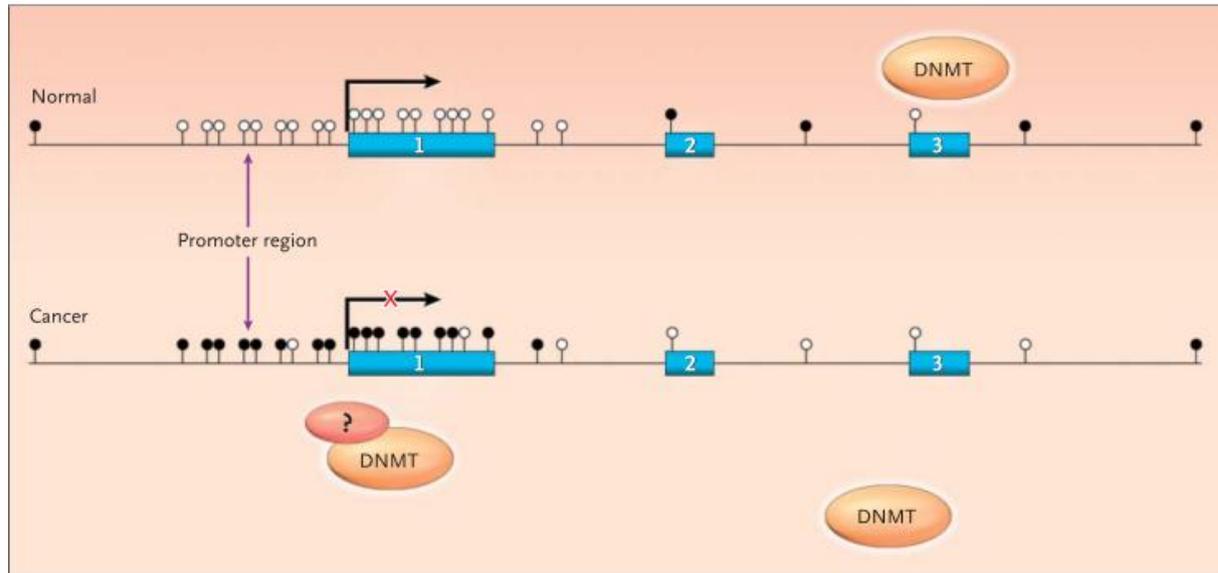


Figure 5. Adapted from reference 20

Distribution of CpG island in normal cells and cancer

Normal cells (above) have a multitude of unmethylated (white spheres) CpG sites (CpG islands) at the promoter and methylated CpG sites outside of the CpG islands. In cancer, DNA methylation patterns are shifted. Numbers 1,2 and 3 depict exons 1,2 and 3 of a fictive gene Exon 1 harbors a CpG island. DNA methyltransferase(DNMT) are mostly present outside the CpG island in normal tissue, whereas in cancer tissue they reside at the CpG island.

reported to be methylated in all cell lines, adenomas and carcinomas examined. This in contrast to normal colonic tissue, where methylation was rare. Furthermore, methylation of *ER* was not seen in young individuals and became progressively more prominent with age²⁵. More genes were then recognized as targets of age-related methylation. Among these genes are IGF2, MYOD and N33. Further investigation showed that in a panel of 25 CpG islands, selected for their differential methylation in colon cancer when compared to normal colon tissue, 18 of these islands were also methylated as a function of age²⁶. These data suggest that aging can account for hypermethylation events in colorectal carcinogenesis.

Age-related methylation can result in methylational inactivation of genes. Studies suggest that *de novo* methylation occurs as a crucial age-dependant mechanism. Cell divisions cause patterns of methylation to extend from the borders of genes towards the CpG islands. Protection against methylation is strong at these islands, due to protective proteins. Some genes exhibit weak protection and age-related methylation might suffice for spread of methylation and gene-silencing in normal en neoplastic tissue. Other genes might only become hypermethylated at the promoter region through an genetic event, possibly overactivation of DNMTs(Figure 6). This might explain why methylation is an age-related event, as genetic events underlying CIMP might not affect young tissues, in which methylation is not yet built up at the island borders, ready to spread and inactivate the gene^{19,27}.

When age-related methylation is filtered out, a distinct subset of colorectal cancers can be classified. These CRCs carry the CIMP phenotype. It is important to emphasize the distinction between age-

related and cancer-related methylation. Age-related methylation is an indication of cell age and cell lineage. Methylation is thought to increase with age as a random event. Cancer-related methylation is a process that inactivates tumor suppressor genes by methylation of CpG islands in the promoter region of a gene¹⁹.

Tumors can be classified as CIMP-positive/high(CIMP-H), CIMP-low(CIMP-L) and CIMP-negative(CIMP-0). CIMP-positive CRCs are currently defined by a panel of CpG island methylation markers, that are classified as having or not having DNA methylation on the basis of certain thresholds.¹⁵ The CIMP-L phenotype was suggested later due to the discovery of certain mutations often present in CIMP-H tumors. The CIMP-high phenotype has been consistently associated with *BRAF* mutations and a specific subtype of CIMP-status was proposed: the CIMP-low phenotype, which is associated with *KRAS* mutations⁶.

Even though target genes for promoter hypermethylation are known, the underlying mechanisms are still under thorough research. In order to give a better understanding, a series of arguments must first be considered. First, not all genes are equal in regards to selectivity of DNA methylation in cancer. While mutations in *MLH1* and *MSH2* both give rise to similar hereditary forms of colorectal cancer, only *MLH1* seems targeted for promoter methylation, despite *MSH2* having a considerable CpG island²⁸. Second, gene selection occurs. Tumor suppressor gene methylation is clearly selected for, and oncogenes are selected against DNA methylation²⁹. Third, the discovery of CIMP implies that an important event can cause a dysfunctional methylation system, raising methylation rates by several fold in selected cancers¹⁹. These insights give clues as to how the CIMP-phenotype is established and how

the underlying mechanisms work. Not all arguments provide a distinct answer but the question whether gene selectivity is plausible was solved by studies on the formation of DNA methylation patterns in embryogenesis³⁰. It was proposed that methylation is caused by the formation of “methylation centers” – short sequences that attract DNMTs – and that these centers mostly consist of repetitive elements in the DNA, like transposons. Regardless of function, if a gene does not have such a “methylation center” near a promoter, it will not be susceptible to DNA methylation and transcriptional silencing will not occur.

Many of the genes modified by promoter hypermethylation show tumor suppressor function. Examples are the VHL gene in renal cancer, the cell-cycle-control gene p16 in several types of cancer and the MMR gene MLH1 in CRC. Demethylating drugs such as 5-azacytidine can reactivate the affected genes and improve expression of methylated genes, restoring protein production in cultured

cancer cells²⁰, implying clinical relevance to early recognition of the methylator phenotype.

CIN, MSI and CIMP correlates

It is important to emphasize that CRCs arise in a multifactorial manner, where accumulation of genetic and epigenetic aberrations underlie its pathogenesis. The earlier described CRC-phenotypes are a method for classification of CRCs and for pathogenic understanding. However, the relations between the phenotypes are also of crucial importance. CIN+ tumors gain genetic aberrations through earlier described mechanisms like aneuploidy and chromosomal translocations, while most CIN- tumors gain these defects from CIMP and/or MSI mechanisms⁹.

Although it is often assumed that all colorectal carcinomas arise from the initial APC mutation through CIN-based loss of tumor suppressor genes, studies have shown that this pathway only accounts for 60% of CRCs. The majority of the remaining 40% (approximately 35%) can

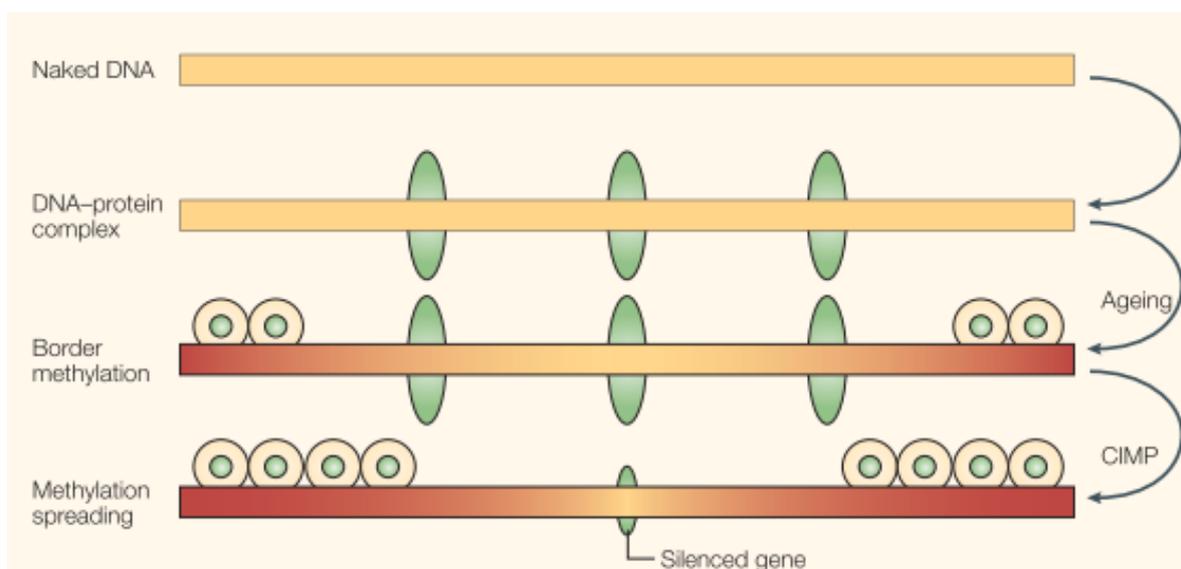


Figure 6. Adapted from reference 19

Hypermethylation model in cancer. Naked CpG island DNA (top) is unmethylated and harbors methylation-protecting protein (green ovals). Ageing might cause borders of the DNA to become methylated (yellow and green circles). This border methylation may cause methylational pressure and spreading, resulting in loss of protective proteins and methylational silencing of genes (bottom).

be accounted to the CIMP pathway. The other 5% describes the MSI-pathway in HNPCC/Lynch-syndrome³¹.

Ogino *et al.* proposed a model for classification of CRCs based on CIMP and MSI status and is based on an earlier model by J. Jass³²(Figure 7). This model suggests 6 subtypes of CRC: MSI-H CIMP-H(Group 1), MSI-H CIMP-low/0(Group 2), MSI-L/MSS CIMP-H(Group 3), MSI-L CIMP-Low(Group 4), MSS CIMP-Low(Group 5), and MSI-L/MSS CIMP-0(Group 6). Note that CIN+ status is only present in Group 6⁶.

Group 1 is generally known as sporadic MSI-H and associated with *hMLH1* promoter hypermethylation caused by CIMP. Group 2 includes HNPCC/Lynch syndrome as well as sporadic MSI-H unrelated to CIMP. Group 3 is composed of CRCs that shows CIMP-H phenotype without defects in MSI. *BRAF* mutations are commonly found in this group. Group 4 is associated with *KRAS* mutations, like Group 5. The significance of these mutations will be discussed further in this thesis. Last, Group 6 consists of CIN+ tumors, lacking CIMP-H/L and MSI-H phenotypes and, in combination with Group 4 and 5, accounts for the majority of CRCs(75-80%)⁶. The distinction between these groups according to molecular and genetic phenotypes provides insights in pathogenic relationships. Two distinct pathways can be concluded, a CIN+ pathways, where chromosomal aberrations are the main cause of mutational defects in colorectal carcinogenesis, and a CIN- pathway, which consists of several CRC phenotypes, mostly caused by MSI and CIMP-related events. An inverse correlation can be suggested for MSI-H and CIMP-H. A mere 10% of all CRCs are MSI-H CIMP-H, which suggests that MSI-H status is often combined with lack of CIMP phenotype and *vice versa*. Also, *BRAF* and *KRAS*

mutations are inversely correlated in regards to CIMP-status. CIMP-H CRCs mostly carry a *BRAF* mutation while MSI-L/0 CRCs often have a *KRAS* mutation³³. Both mutations have similar effects through the MAPK-pathway⁵.

Classifying CRCs according to CIN, MSI and CIMP status has important clinical implications. MSI-H CRCs tend to have better prognosis that MSI-L CRCs³⁴. It is thought that MSI-H tumors, due to the abundance of mismatch repair defects throughout the genome, generate many aberrant proteins, triggering a strong anti-

	CIMP-H	CIMP-L	CIMP-0
MSI-H	Gr. 1 10%	Gr. 2 5%	
MSI-L	Gr. 3 5-10%	Gr. 4 5%	Gr. 6 40%
MSS		Gr. 5 30-35%	

Figure 7. Adapted from reference 6
Classification of CRCs according to MSI and CIMP status

cancer immune-response³⁵. However, the favorable prognosis might also be due to the confounding effect of its inverse relationship with CIN, as CIN is generally an indicator of worse survival, independent of MSI-status. Also, the correlations between CIMP-status and *BRAF/KRAS* mutations have shown that conventional classification of CIMP, dependent on methylation markers might become analogous to classification

according to the presence of *BRAF/KRAS* mutations³⁶.

The three distinct mechanisms of genetic instability, underlying CRC, have given great insights in the general pathology of colorectal cancer. In order to implement these findings in our current knowledge, their importance in the well-known adenoma-carcinoma sequence must be illustrated³⁷.

Adenoma-Carcinoma sequence

The first implication of a genetic model for colorectal tumorigenesis came from research by Fearon *et al.*. It was proposed that normal colorectal tissue went through a series of specific mutations, transforming it into adenomatous tissue and, subsequently, CRC³⁷. Research has since then uncovered a multitude of other significant molecular events and their role in the AC-sequence.

Activation of the Wnt signaling pathway is regarded as the initiating step in CRC. Normal Wnt signaling causes β -catenin to localize to the cellular adherens

junction, where it is degraded by other proteins, preventing transcription and proliferation of the cell. One of these proteins is *adenomatous polyposis coli*(APC). The APC gene was identified more than 23 years ago through its association with a hereditary form of CRC called Familial Adenomatous Polyposis(FAP)³⁸. Sporadic mutation of the APC gene occurs in 50% and 80% of sporadic adenomas and adenocarcinomas, respectively². Also, mutations in the APC gene are the earliest known genetic alteration in colorectal tumorigenesis due to the fact that they are found in the smallest detectable adenomas. They are also found in one of the earliest detectable colorectal lesions, the aberrant crypt foci(ACF)³⁹. Mutations that render β -catenin resistant to degradation are also known and are clinically similar².

FAP is characterized by a multitude (hundreds to thousands) of colorectal adenomas at a young age (before 20 years). These adenomas are initially not life-threatening but their massive

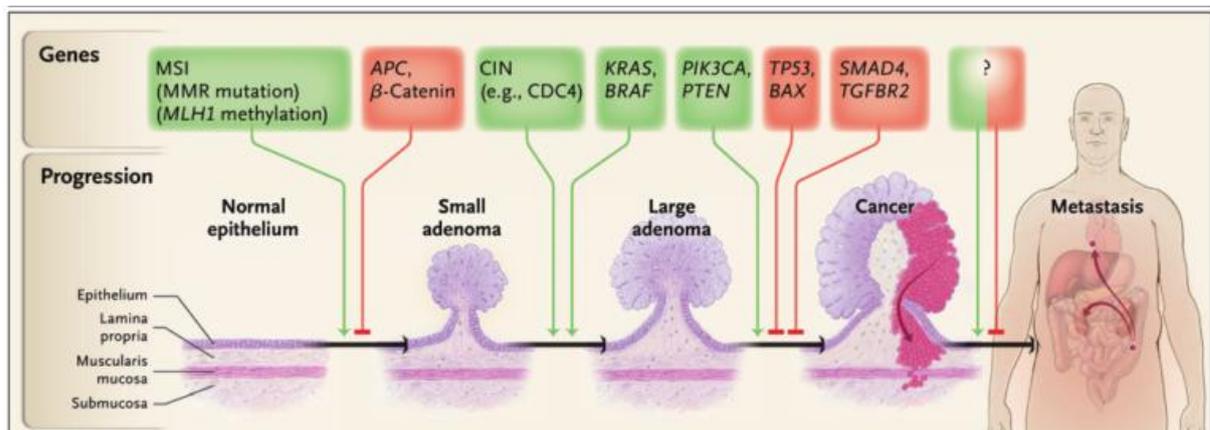


Figure 8. Adapted from reference 5

The adenoma-carcinoma sequence. In the progression of cancer, genetic alterations target the genes identified at the top of the diagram. MSI = microsatellite instability due to mismatch repair (MMR) gene mutations or MLH1 promoter methylation. The question mark indicates unidentified genetic changes specific to metastatic processes. CIN denotes chromosomal instability.

numbers virtually guarantee that some will develop into malignant lesions or CRC. The risk of developing colorectal cancer by the age of 40 is almost 100%².

Figure 8 is a version of the Kinzler-Vogelstein model, implementing recent findings like MSI-based colorectal carcinogenesis, *MLH1* promoter methylation and the CIN pathway. Note that, even though a specific timeline is suggested, not all aberrations mentioned are present in each of the different phenotypes. For example, mutations in *BRAF* are most common in CIMP-H tumors while mutations in *KRAS* are not. When a *BRAF* mutation is identified, the CIMP pathway is suggested rather than a CIMP-I/O phenotype.

MSI, through mutations in MMR-genes, or through methylation of *MLH1* is regarded as the initiating step, providing genetic instability. Mutations in *APC* or *B-catenin*, somatically or inherited, are the first key mutations in the AC-sequence. Both of these genetic alterations provide the means for transforming normal epithelial cells into small adenomatous polyps.

APC mutations might also cause a permissive state for cells to acquire chromosomal aberrations. This, in combination with a possible mutation in *hCDC4* often initiates the CIN-pathway, causing multiple losses of heterozygosity at multiple loci, like *TP53* and *SMAD4*^{9,40}.

Oncogenic mutations of *KRAS/BRAF* are often a next step in the adenoma-carcinoma sequence. Both of these genes play an important role in the MAPK-pathway, which stimulates proliferation, causing small adenomas to transform into larger adenomas. As mentioned earlier, aberrant *MLH1* methylation and *BRAF* mutational silencing are often associated with the CIMP-phenotype^{6,36}.

Mutations in the phosphatidylinositol 3-kinase (PI3K) pathway occur in less than

one third of colorectal cancers. Stimulation of this pathway results in cell survival signaling and apoptosis suppression. *PI3CKA*, as well as, *PTEN*, can become mutated and cause promotion of PI3K signaling⁴¹

Wildtype p53 mediates cell-cycle control and a cell-death checkpoint. Inactivation of *TP53* is regarded as the second step in colorectal carcinogenesis and also considered a crucial one as it often coincides with the transition of large adenomas into invasive carcinomas⁴². Many MSI-H colorectal carcinomas exhibit a wildtype *TP53*, although the activity of the p53 pathway is probably attenuated by mutational inactivation of *BAX*, an inducer of apoptosis⁵.

Inactivation of the transforming growth factor- β (TGF- β) signaling pathway occurs later in the AC-sequence, and also coincides with transition of large adenomatous tissue into carcinomas. *TGFBR2* can be silenced through somatic mutations in *TGFBR2* itself, or its downstream target *SMAD4*, of which the latter is associated with familial juvenile polyposis. *TGFBR2* can also be inactivated by a frameshift mutation caused by MMR-defects⁴³. A general approach to classifying CRCs according to MSI, CIMP and CIN-status as well as common mutations during the adenoma-carcinoma sequence is suggested in Table 1. Both the models of Ogino and Goel⁶ and Jass³² are combined. This stratification might provide further clinical relevance. Histologically, groups 1, 3, 4, and 5 are difficult to find and the cancers associated with these groups are often viewed collectively as “serrated adenocarcinomas”. Hyperplastic polyps, often rendered harmless, have recently been proven to be able to develop into these serrated adenocarcinomas which suggests that relatively inconspicuous

polyps are the precursors of a subset of colorectal cancers⁴⁴.

The adenoma-carcinoma sequence³⁷ launched the investigation of colorectal tumorigenesis as a sequence of somatic mutations rather than a fully random event. The discovery of genetic instability, as a different means to aid in tumorigenesis gave rise to distinct CRC phenotypes. Instability through loss/gain of chromosomes, microsatellite instability and aberrant methylation (resulting in a CpG island methylator phenotype) have all been indicated as mechanisms promoting CRC. As CRC is a disease mostly found in the elderly, a question arises: Are there any genetic or epigenetic events that occur prior to the adenoma-carcinoma sequence of tumor progression? In order for that question to be answered, one must first assess the origin of CRC, namely the colorectal crypt. In this crypt, stem cells maintain the crypt and it is with this colorectal stem cell, that CRC is thought to arise. Investigation in colorectal stem cells is abundant and, due to the long life-span might provide important insights in the causes and origin of CRC, prior to the adenoma-carcinoma sequence.

Stem cells and pre-tumor progression

A colorectal stem cell is defined as a cell that must be of a relatively undifferentiated type capable of proliferation and self-maintenance, producing a variety of cell lineages and capable of tissue regeneration following injury. They reside at the base of the crypt and are responsible for maintaining it.⁴⁵ It is thought that the phenotype of a colorectal cancer cell progenitor for the majority of its lifetime is a colorectal stem cell because mutations in nonstem cells will not accumulate⁴⁶. A cancer cell in a 70-year-old male may have 1 year with a cancer phenotype, 10 years with an adenoma phenotype and 59 years with a stem cell phenotype. Despite this, mutations only seem to accumulate in tumor cells and not stem cells. Hence, the adenoma-carcinoma sequence does not apply in these stem cells so a different approach to early cancer development must apply to colorectal stem cells. Also, visible tumorigenesis is not evident in colorectal stem cells due to the lack of mutations that alter morphology. The question arises: does anything important happen before visible tumorigenesis? The mutations and events that precede visible tumorigenesis are called pretumor progression³, and recent studies have

Table 1: Classification of CRCs according to genetic instability and common mutations. MSI, microsatellite instability; CIMP, CpG island methylator phenotype; CIN, chromosomal instability; H, high; S, stable; L, low; +, positive/presence of mutation; -, negative/absence of mutation.

Feature	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
MSI status	H	H	S/L	L	S	S
CIMP	H	L	H	L	L	0
CIN	-	-	-	-	-	+
APC	+/-	++	+/-	+++	+++	+++
KRAS	-	++	+	++	++	++
BRAF	+++	-	++	-	-	-
TP53	-	+	+	+++	+++	+++

suggested an important role for age-related methylation events.

Histologically visible genetic alterations, like the earlier described adenoma-carcinoma sequence, provide important insights in the understanding of colorectal carcinogenesis. Pretumor progression implies the existence of morphologically invisible alteration, aiding in carcinogenesis. If pretumor progression

stem cells. Immortal stem cells always divide asymmetrically while niche stem cells sometimes produce two identical daughter cells. Asymmetric division produces a stem cell and a differentiated cell, of which the latter does not divide further and lineage is lost. Symmetric division produces either two differentiated non-stem cells or two stem cells. Niche stem cells have both

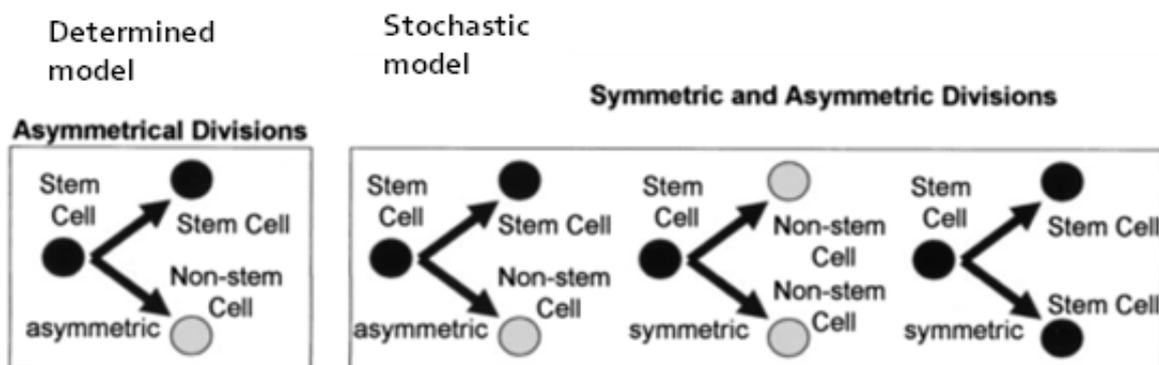


Figure 9 adapted from reference 47

Immortal stem cells and a stem cell niche . Left panel: Immortal stem cells divide asymmetrically into a stem cell and a differentiated non-stem cell, never to become extinct. Right panel: asymmetric and symmetric division of a stem cell niche. Niche succession suggests a possibility of a ‘bottle-neck’, eradicating all colorectal stem cells except one.

eventually contributes to carcinogenesis, mutations or alterations must occur in colorectal stem cells and the progeny of these cells must persist. Stem cell lineage is an important aspect in pretumor progression.

Two models of stem cell succession in crypts have been hypothesized: succession through a determined model (immortal stem cell lineage) and succession through a stochastic model(stem cell niche)(Figure 9)⁴⁷. Both models describe ways of inheriting alterations or mutations after stem cell division. Immortal stem cells only divide asymmetrically whereas a stem cell niche has both asymmetric and symmetric division. Niches are specialized regions that influence cells within them to act as

symmetric and asymmetric division. Most of the time, the crypt will be maintained due to production of a daughter stem cell. However, it is possible that, from all the stem cells in one crypt, only a single daughter stem cell will be produced. This random stem cell loss with replacement will eventually lead to extinction of all lineages except one: clonal succession⁴⁷. Clonal succession suggests that all crypt cells have a similar progenitor due to “bottleneck-like” extinction.

Proof of niche succession came with research in age-related methylation events. The theory behind this is that age-related methylation is an accumulating random event. Such random somatic errors resemble genetic drift during species evolution and, as such, can be



used to reconstruct somatic ancestry because methylation exhibits somatic inheritance. Therefore, somatic methylation patterns might illustrate numbers of divisions since birth, and ancestral relationships among cells within a crypt. Increased overall percent methylation should resemble cell ageing while methylation diversity can illustrate cell ancestry⁴⁸.

Further investigation by Kim *et al.*(2005) supported this theory. Both colon and small intestine crypt are thought to be maintained by clonal expansion of stem cells. Small intestinal villi represent cellular mixtures of different crypts which suggests that, through somatic inheritance of methylation patterns, they should be more diverse than colonic and intestinal crypts. Although the villi should show more diverse methylation patterns than crypts, they should have the same number of divisions since birth. Indeed, age-related methylation patterns showed that villus and crypt percent methylation were not significantly different, while villi contained significantly more methylation diversity than crypts⁴⁹. In conclusion, the existence of colorectal cancer stem cells together with the somatic inheritance of methylation events through ageing suggest that pretumor progression can possibly be visualized in the DNA using methylation events in colonic stem cells.

Using methylation pattern diversity, other studies suggest that *APC* mutations alter niche stem cell survival. Results showed that enhanced stem cell survival, allowing for accumulation of mutations and increased risk of cancer, is associated with *APC* mutations in human colon tissue. Both FAP and non-FAP colons are normal at birth and tumors appear decades later. In FAP, *APC* mutations result in enhanced stem cell survival and a higher mutational burden within each

crypt. In non-FAP, somatic mutation of *APC* allows a selective niche “sweep”, transforming the entire crypt into stem cells with mono-allelic inactivation of *APC*. Similarly to FAP, the crypt would have an enhanced stem cell survival and higher mutational burden. Loss of the second *APC* allele subsequently ends pretumor progression and begins tumor progression⁵⁰

This pretumor progression model might explain why *APC* mutations are one of the key initiating steps in CRC. Stem cells that acquire a somatic mutation anywhere on the *APC* locus that do not enhance stem cell survival will be frequently lost during normal cell turnover. Somatic mutations in regions of the *APC* locus associated with FAP however, improve stem cell survival, allowing for a crypt niche “sweep” to occur. This sweep would increase *APC* mutations in the crypt by division and not by mutation, generating a FAP-like crypt⁵⁰. Second hits, and maybe even third hits for *APC* can occur similarly, implying a distinct role for pretumor progression in crypt niches in early colorectal carcinogenesis.

Colorectal cancer stem cells

The concept of a small subpopulation of cells driving tumorigenesis, rather than proliferation and growth in every tumor cell has recently been hypothesized. These cells called “cancer stem cells” were suggested in research in hematologic malignancies where it was discovered that only a small fraction of the cellular population were able to form colonies. Two models supporting this concept can be suggested: the stochastic model predicts that all cancer cells can form colonies, but with a low probability of doing so. The stem cell model suggests that only a defined heterogeneous subpopulation of cells can proliferate and form new tumors. The latter is supported

by research demonstrating that a specific group of leukemia cells (CD34+CD38-) were able to produce new malignancies after transfection in nonobese, immunodeficient mice^{51,52}.

The presence of such a subpopulation of cells raises the question whether this is caused by a monoclonal cancer stem cell population or whether there are multiple different cancer stem cells within the same tumor. Studies have shown that single cell clones obtained from human colon cancers grown in culture are able to differentiate into multiple lineages in nude mice, expressing markers of enterocytes, goblet cells and neuroendocrine cells⁵³.

Further indication of the presence of colon cancer stem cells arises from the fact that normal stem cells are the only cells capable of self-renewal and neoplasia is essentially dysregulated self-renewal. Also, normal stem cells are the only cells that live long enough to acquire the necessary amount of genetic alterations needed to transform normal tissue into malignant tissue. If, for example, an intestinal cell acquired mutations that allowed it to become malignant, this alteration would become redundant in 5-7 days because by then the cell would leave the crypt altogether⁴⁶.

The earlier described adenoma-carcinoma sequence coincides with the theory that early colorectal lesions originate from mutant cells at the surface epithelium of the colon, where they spread laterally and downward to form new crypts⁵⁴. This theory was challenged with an alternative model proposing that adenomas grow initially in a bottom-up pattern, where the origin of the adenomas lie in the intestinal stem cells(Figure 10)^{46,55}. As the existence of the cancer stem cells is evident, the latter should apply to colorectal carcinogenesis. This process is however not well defined,

and there is a continuing debate whether the origin of colorectal adenoma lie in the intestinal stem cells, or in the surface epithelium. However, if the 'top-down' model is true, it would imply that the stem cell zone is at the top of the crypt, which is in disagreement with current literature⁴⁶.

Somatic inheritance of methylation events in colorectal cancer stem cells provide insights in stem cell aging, crypt diversity, and possibly increased risk of

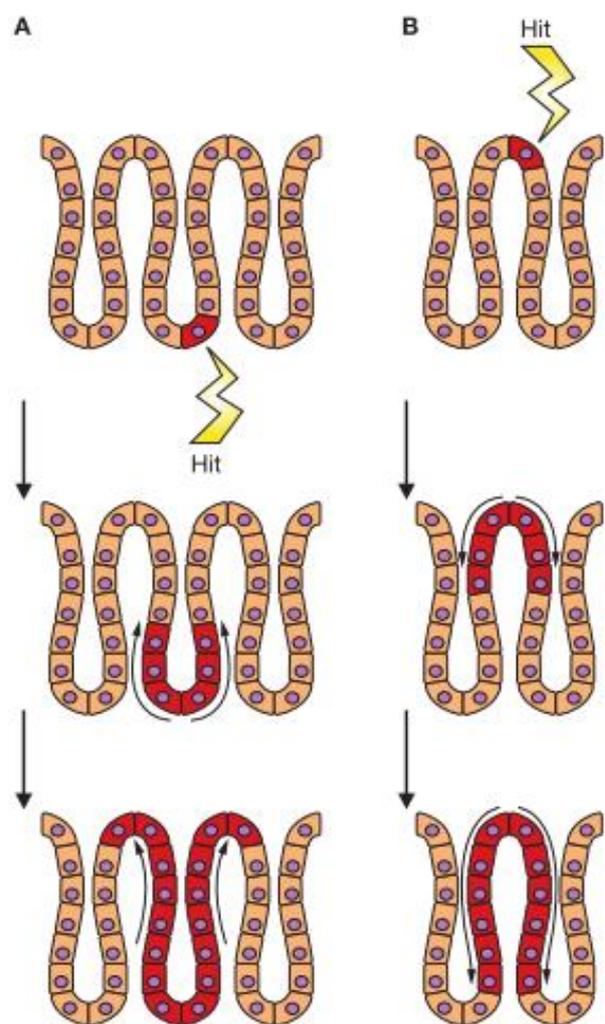


Figure 10 Adapted from reference 49
Top-down and bottom-up models of colorectal carcinogenesis. The 'bottom-up' model depicts dysplastic cells spreading from the base of the crypt within the stem cell zone and continuing upwards so that the entire crypt is dysplastic. (B) An opposing theory described as the 'top-down' model, which suggests a mutation occurs near the top of the crypt, resulting in a dysplastic cell, spreading downward to the base of the crypt.

cancer. More research is now being done in order to determine whether these methylation events can differentiate between normal colonic tissue and patients with hereditary CRC like: FAP, Peutz-Jeghers Syndrome, Juvenile Polyposis and even patients with Barret's esophagus, a syndrome that transforms

normal esophageal epithelium in to colon-like crypts. The following research was performed by the author at the UMC Utrecht Pathology Research Laboratory.

DNA methylation patterns in Barret's esophagus

Abstract

Individuals with Barret's esophagus have esophageal tissue resembling small intestinal epithelium. To determine stem cell age and division of esophageal cells, epigenetic patterns were used as cell fate markers. Methylation exhibits somatic inheritance and random changes that potentially record lifelong stem cell division history as tags in CpG sites. Methylation tags were samples with bisulfite sequencing at the cardiac specific homeobox locus. Samples from the Barret's esophagus patient did not have a significant increase in average percent methylation or number of unique methylation tags. Further investigation needs to be done in order to truly determine whether Barret's esophagus predisposes patients to increased age-related methylation, implying a role in pretumor-progression

Introduction

Barrett's esophagus (BE) is the premalignant lesion of the esophageal adenocarcinoma, originating from gastroesophageal reflux disease (GERD). Esophageal adenocarcinoma incidence increased in the Western world by 600% in 25 years⁵⁶ It was first described by Norman Barret in 1950. Barret stated that BE resembled gastric epithelium in the lower part of the esophagus⁵⁷. The current definition is the endoscopic appearance of a columnar epithelium in the tubular esophagus and a biopsy demonstrating specialized intestinal metaplasia (SIM) on histological examination. SIM creates a distinct predisposition to the development of dysplasia and, subsequently, adenocarcinoma. This is not restricted to the lower parts of the esophagus however⁵⁶. The morphologic similarity of

BE to colon tissue and the increased chance of esophageal cancer implies a role for age-related methylation, simultaneously researched by the staff of the UMC Utrecht Pathology Research Laboratory.

The purpose of these experiments is to determine whether stem cells of patients with Barret's esophagus are predisposed to increased age-related methylation compared to control patients. Stem cell hierarchy has been previously shown in colorectal stem cells and this stem cell crypt niche model implies the heredity of crypt stem cells and their methylation patterns in Barret's esophagus as well⁵⁸.

In order to investigate the variability in stem cell methylation patterns in Barret's esophagus (BE), crypt samples of BE patients were treated with bisulfite and sequencing followed. The materials and methods will only apply to samples researched by the author. Results and

discussion will include the samples investigated by the Pathology Research Lab at UMC Utrecht, the Netherlands.

Methods

Isolation of single crypts

Esophageal biopsies were previously taken from patients with BE and tissue samples were formalin fixed and embedded in paraffin. Three groups of samples can be distinguished: (1) BE patients with incidence of tumors, (2) BE patients without incidence of tumors and (3) control group consisting of normal small intestinal epithelium. Patients groups were age-matched in order to distinguish between age-related methylation events and 10 crypts were cut per patient.

Single crypts were cut out using PALM[®] Robot Microbeam laser microdissection system and covered with oil. Hematoxylin and eosin(HE) stained slides were used as reference. 7 crypts were used from patient no. 30, a 69 year old male with earlier presentation of adenocarcinomas in the esophagus.

Arcturus[®] PicoPure[®] DNA Extraction Kit(Applied Biosystems[™]) was used to extract DNA from the samples according to protocol. DNA was extracted in a 20- μ l solution of PicoPure DNA Extraction mix (as defined by protocol) overnight at 56°C, followed by boiling for 10 min.

Bisulfite treatment

Epitect[®] Bisulfite Kit (Qiagen[®]) was used for bisulfite conversion and cleanup. This step deaminates cytosine into uracil, which will then be converted into thymine. Methylated cytosine (5-methylcytosine) will not be affected. Thus, a distinction can be made between methylated and unmethylated sequences. After the bisulfite treatment, a nested PCR was performed with CSX specific primers(Appendix I/II). Subsequently, purification of the PCR-products was done using GE Healthcare illustra[™] GFX[™] PCR DNA and Gel Band Purification Kit, according to protocol.

pGEM-T Cloning.

An A-overhang was first created in order to improve the efficacy of pGEM-T cloning. The pGEM-T vector has a single 3'-terminal thymidine at both ends. This T-overhang will prevent recircularization of the vector and provide a compatible overhang for PCR products. The A-overhang was created by adding 30 μ l purified PCR-product to 4 μ l NEB Buffer, 3.2 μ l 25mM MgCl₂, 0.32 μ l 25mM dNTPs, 2.33 μ l H₂O, and 0.15 μ l Taq Polymerase. The mixture was then heated to 72°C for 10 minutes.

pGEM-T cloning was performed to separate the CSX from the different cells within the crypt. The vector system allows blue/white screening for insertional inactivation of the enzyme β -

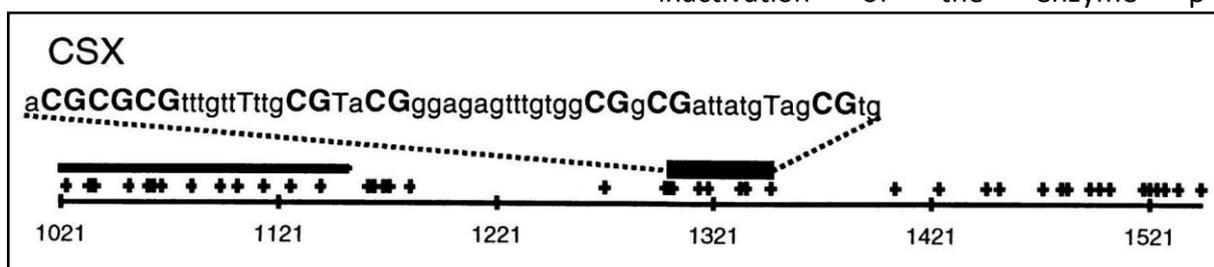


Figure 11 adapted from reference 58. Methylation tags on CSX locus. CpG sites are indicated by + marks and transcription start sites are indicated by arrows. Bisulfite-treated sequences are illustrated with converted Cs indicated by capital Ts and CpG in bold.

galactosidase. White colonies will have the CSX-PCR product from a single cell whereas blue colonies will not. 12 white colonies were scraped off incubated agar plates and used for colony PCR(Appendix II). Afterwards, BigDye® Terminator Cycle Sequencing was performed.

Methylation analysis

For each of the 12 samples per crypt, 8 CpG sites were analyzed for the cardiac-specific homeobox gene(CSX; GenBank accession no. NM004387; nucleotides 1,272-1,377). Age-related methylation has been previously demonstrated for CSX⁵⁹. Its lack of function in the colon implies that CSX can be used as a marker for random methylation. Part of the CSX nucleotide sequence, with corresponding CpG islands are depicted in Figure 11. A figure illustrating the primer locations can be seen in Appendix I.

Calculations

Interpretation of the results was done by using two statistical numbers. Percent methylation describes the average percent of methylated sites, which is an indication of cell age, as random methylation due to ageing increased percent methylation^{48,49}. Number of unique methylation patterns is the amount of different methylation patterns found in each crypt, indicating cell ancestry. Multiple crypt tags likely represent multiple long-lived stem cells, where diversity increases with the lifespan of the stem cell lineage⁶⁰.

Results

After the nested PCR, agarose gel electrophoresis was done. All 7 samples were positive for CSX which indicated DNA extraction and nested PCR were both successful(Figure 12).

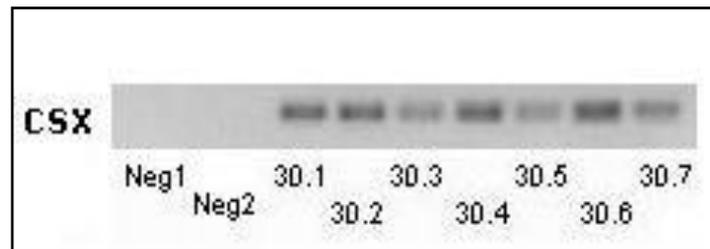


Figure 12. Gel electrophoresis for nested PCR samples 30.1-30.7. Numbered bands shown are CSX positive. Neg1 and Neg2 are negative controls.

A second gel electrophoresis was done after PGEM-T cloning and subsequent colony PCR to determine if the colonies were scraped off properly and if the correct CSX PCR-product was incorporated into the vector(Figure 13).

Of the 12 colonies, gel electrophoresis-positive samples were used for Big Dye® Terminator Sequencing. 10 DNA sequences were analyzed per crypt. Each cytosine base on a CpG site was determined methylation-positive, tyrosine bases were considered methylation-negative. Several control tyrosine bases were investigated to make sure that bisulfite treatment was CpG site-specific. Control tyrosine bases are tyrosines that do not reside with a CpG site and as such, cannot be methylated through bisulfite treatment. Aspecific bisulfite converted samples were rejected from the results.

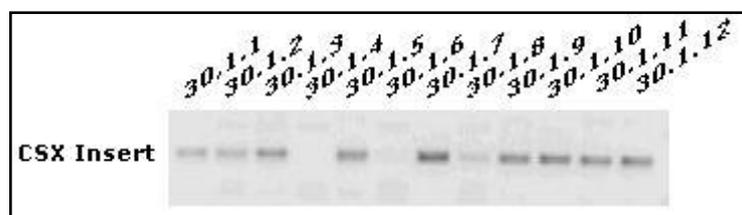


Figure 13. Gel electrophoresis post-Colony PCR. 12 colony samples from sample 30.1. Samples 30.1.4 and 30.1.6 are excluded from sequencing.

Table 2. Percent methylation and unique methylation patterns in Barret's esophagus patients and control group.

	Patient no.(age)	Average percent methylation(std)	Average no. of unique methylation patterns(std)
Barret's Esophagus	30(69)	57.9(23.8)	2.4(0.98)
Control	20(61)	10.5(10.1)	1.8(0.84)
	21(52)	37.5(20.2)	3.5(1.2)
	25(81)	38.1(27.0)	3.0(1.4)
	26(74)	37.8(23.5)	3.0(2.2)
	27(68)	25.9(23.1)	1.6(0.89)
	Average	27.6(22.3)	2.58(1.44)

Patient 30 displayed a total average percent methylation of 57.9% with a standard deviation(std) of 23.8%. The average number of unique methylation patterns was 2.4(std: 0.98). Results gathered by the staff of the UMC Utrecht Pathology Research lab included control patients 20, 21, 25, 26 and 27 with total average percent methylation ranging from 10.5 - 38.1%(std: 12.03) and average number of unique methylation patterns range of 1.6 – 3.5(std: 1.44) Total control group average percent methylation was therefore 27.6%. Control group average number of unique methylation patterns was 2.58(Table 2).

The 27.9% elevated percent methylation of the BE patient compared to the average percent methylation of the control group was tested for significance using a Student's T-test(95% confidence interval, p=0.26203). The same calculation was done for the average number of unique methylation patterns(95% confidence interval, p=0.99914). The higher average percent methylation was deemed not significant similar to the slightly lower unique methylation patterns in BE, which were also not significant. Figure 14 illustrates average percent methylation values of control group and BE patient no. 30.

Discussion

The results gathered originate from a single BE patient. Hence, interpretation of the results is difficult. The average increase in percent methylation compared to the control group is not significant, possibly due to lack of further BE patient testing. Also, both control group and patient no. 30 percent methylation have a relatively high standard deviation. This renders conclusions of results difficult and inaccurate. Similar arguments can be said for methylation tag diversity. Even though average methylation tag diversity is slightly lower in the BE patient than in the control group, conclusions cannot be made due to lack of significance.

One major flaw in this research is the sampling of the crypts. Research suggests that crypts, proximal to tumor location could display higher methylation tag diversity than crypts distal to tumor location. Distinction between these crypts has not been made due to lack of useable crypts in BE dissection material. Data on crypt location and distance is unknown and intercrypt or intracrypt variation in regards to crypt location cannot be identified.

Conclusion

Even though average percent methylation is higher in BE patient no. 30 than in the control group, conclusions cannot be made whether this is a stochastic event or BE-related. One would expect a significant increase in average percent methylation in BE patients due to their increased risk of esophageal cancer and the possible role

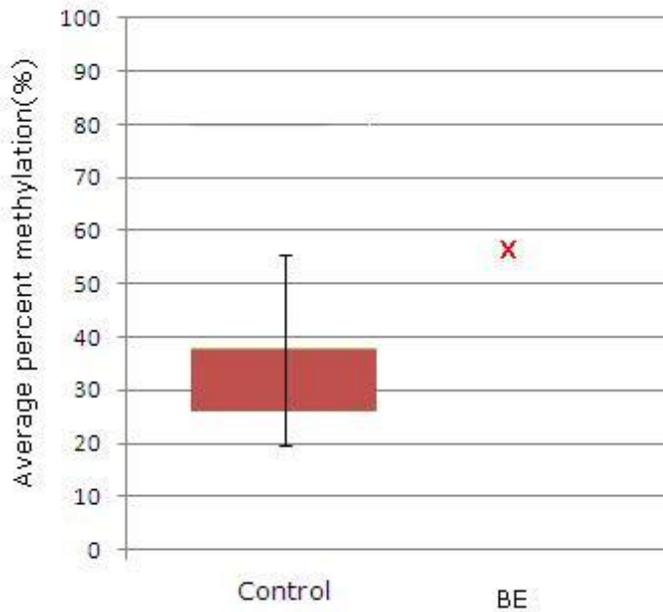


Fig 14. Boxplot of average percent methylation of control group and Barret's esophagus patient no. 30(BE). BE value is 57.9%.

of age-related methylation in carcinogenesis. Average number of unique methylation patterns within a crypt can possibly give an indication of the time since the last niche succession. A low number of unique methylation patterns suggests that niche succession was recent. An increase in this number suggests multiple long-lived stem cells which coincides with enhanced stem cell survival and multi-cellular origin of samples. In both cases, this is not seen in the results and further investigation is needed to elucidate this question.

Future insights

Age-related methylation provides a read-out for the age and lineage of a cell due to its random and age-dependant nature. Using bisulfate sequencing, it is possible to illustrate this process and assign a corresponding value. The idea that an increase of age-related methylation predisposes patients to tumorigenesis, by ways of increasing cancer-related methylation is difficult to assess. If an increase of age-related random methylation is seen in patients with BE or patients with other hereditary syndromes like FAP, it might be a clue as to how a visually normal cell acquires epigenetic alterations, aiding in possible tumorigenesis. The possible link between age-related methylation and cancer-related methylation, through methylational pressure at the borders of genes, seems plausible and might provide a mechanism for this theory. It is however not yet clear whether age-related methylation is increased in these patients and if this methylation is the cause for increased tumorigenesis prior to visible genetic alterations. Further investigation in methylation events is necessary to establish the mechanisms for pretumor progression and its implications in colorectal carcinogenesis.

Acknowledgements

I thank Wendy de Leng and her colleagues at the UMC Utrecht Pathology Research Lab for their assistance and guidance during the experiments and the making of this thesis.



References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. *CA: a cancer journal for clinicians*. 2010;(November 1950).
2. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell*. 1996;87(2):159-70.
3. Calabrese P, Tavaré S, Shibata D. Pretumor progression: clonal evolution of human stem cell populations. *The American journal of pathology*. 2004;164(4):1337-46.
4. Knudson AG. Hereditary cancer: two hits revisited. *Journal of cancer research and clinical oncology*. 1996;122(3):135-40.
5. Kouzminova N, Lu T, Lin AY. Molecular basis of colorectal cancer. *The New England journal of medicine*. 2010;362(13):1245-6; author reply 1246-7.
6. Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. *The Journal of molecular diagnostics : JMD*. 2008;10(1):13-27.
7. Grady WM. Genomic instability and colon cancer. *Cancer and Metastasis Reviews*. 2004;23(1/2):11-27.
8. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998;396(6712):643-9.
9. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature*. 1997;386(6625):623-7.
10. Stoler DL, Chen N, Basik M, et al. The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(26):15121-6.
11. Jin DY, Spencer F, Jeang KT. Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell*. 1998;93(1):81-91.
12. Cahill DP, Costa LT da, Carson-Walter EB, et al. Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics*. 1999;58(2):181-7.
13. Rajagopalan H, Jallepalli PV, Rago C. Inactivation of hCDC4 can cause chromosomal instability. *Nature*. 2004;428(March):77-81.
14. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138(6):2073-2087.e3.
15. Worthley DL, Leggett B a. Colorectal cancer: molecular features and clinical opportunities. *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*. 2010;31(2):31-8.
16. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition : Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer A National Cancer Institut. *Cancer Research*. 1998.
17. Tomlinson I, Halford S, Aaltonen L, Hawkins N, Ward R. Does MSI-low exist? *The Journal of pathology*. 2002;197(1):6-13.
18. Toyota M. CpG island methylator phenotype in colorectal cancer.

Proceedings of the National Academy of Sciences. 1999;96(15):8681-8686.

19. Issa JP. CpG island methylator phenotype in cancer. 2004;4(December):988-993.

20. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *The New England journal of medicine*. 2003;349(21):2042-54.

21. Herman JG. Hypermethylation of tumor suppressor genes in cancer. *Seminars in cancer biology*. 1999;9(5):359-67.

22. Suzuki H, Toyota M, Sato H, et al. Roles and causes of abnormal DNA methylation in gastrointestinal cancers. *Asian Pacific journal of cancer prevention : APJCP*. 2006;7(2):177-85.

23. Cui H, Cruz-Correa M, Giardiello FM, et al. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science (New York, N.Y.)*. 2003;299(5613):1753-5.

24. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983;6(301):89-92.

25. Ahuja N, Issa JP. Aging, methylation and cancer. *Histology and histopathology*. 2000;15(3):835-42.

26. Toyota M, Issa JP. CpG island methylator phenotypes in aging and cancer. *Seminars in cancer biology*. 1999;9(5):349-57.

27. Velicescu M, Weisenberger DJ, Gonzales FA, et al. Cell Division Is Required for de Novo Methylation of CpG Islands in Bladder Cancer Cells. *Cancer Research*. 2002;2378-2384.

28. Cunningham JM, Christensen ER, Tester DJ, et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer research*. 1998;58(15):3455-60.

29. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nature reviews. Genetics*. 2002;3(6):415-28.

30. Turker MS. Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene*. 2002;21(35):5388-93.

31. Snover DC. Update on the serrated pathway to colorectal carcinoma. *Human pathology*. 2011;42(1):1-10.

32. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*. 2007;50(1):113-30.

33. Goel A, Nagasaka T, Arnold CN, et al. The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. *Gastroenterology*. 2007;132(1):127-38.

34. Sinicrope FA, Rego RL, Halling KC, et al. Prognostic impact of microsatellite instability and DNA ploidy in human colon carcinoma patients. *Gastroenterology*. 2006;131(3):729-37.

35. Phillips SM, Banerjea A, Feakins R, et al. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. *The British journal of surgery*. 2004;91(4):469-75.

36. Curtin K, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future.

Pathology research international. 2011;2011:902674.

37. Fearon E, Vogelstein B. A Genetic Model for Colorectal Tumorigenesis. *Cell*. 1990;61:759-767.

38. Goss KH, Groden J. Biology of the adenomatous polyposis coli tumor suppressor. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2000;18(9):1967-79.

39. Jen J, Powell SM, Papadopoulos N, et al. Molecular Determinants of Dysplasia in Colorectal Lesions Molecular Determinants of Dysplasia in Colorectal Lesions '. *Cancer Research*. 1994;55:23-5526.

40. Barber TD, McManus K, Yuen KWY, et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(9):3443-8.

41. Parsons DW, Wang T-L, Samuels Y, et al. Colorectal cancer: mutations in a signalling pathway. *Nature*. 2005;436(7052):792.

42. Baker SJ, Preisinger AC, Jessup JM, et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer research*. 1990;50(23):7717-22.

43. Grady WM, Myeroff LL, Swinler SE, et al. Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. *Cancer research*. 1999;59(2):320-4.

44. Jass JR. Molecular heterogeneity of colorectal cancer: Implications for cancer control. *Surgical oncology*. 2007;16 Suppl 1:S7-9.

45. Bach SP, Renehan a G, Potten CS. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis*. 2000;21(3):469-76.

46. McDonald SAC, Preston SL, Lovell MJ, Wright NA, Jankowski JAZ. Mechanisms of disease: from stem cells to colorectal cancer. *Nature clinical practice. Gastroenterology & hepatology*. 2006;3(5):267-74.

47. Kim K-M, Shibata D. Methylation reveals a niche: stem cell succession in human colon crypts. *Oncogene*. 2002;21(35):5441-9.

48. Bird A. DNA methylation patterns and epigenetic memory. *Genes & development*. 2002;16(1):6-21.

49. Kim JY, Siegmund KD, Tavaré S, Shibata D. Age-related human small intestine methylation: evidence for stem cell niches. *BMC medicine*. 2005;3:10.

50. Kim K-M, Calabrese P, Tavaré S, Shibata D. Enhanced stem cell survival in familial adenomatous polyposis. *The American journal of pathology*. 2004;164(4):1369-77.

51. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine*. 1997;3(7):730-7.

52. Yeung TM, Mortensen NJ. Colorectal cancer stem cells. *Diseases of the colon and rectum*. 2009;52(10):1788-96.

53. Vermeulen L, Todaro M, Sousa Mello F de, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(36):13427-32.



54. Shih IM, Wang TL, Traverso G, et al. Top-down morphogenesis of colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(5):2640-5.

55. Preston SL, Wong W-man, Chan AO-on, et al. Bottom-up Histogenesis of Colorectal Adenomas : Origin in the Monocryptal Adenoma and Initial Expansion by Crypt Fission. *Cancer Research*. 2003:3819-3825.

56. Gilbert EW, Luna R a, Harrison VL, Hunter JG. Barrett's Esophagus: A Review of the Literature. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract*. 2011:708-718.

57. Barrett NR. Chronic peptic ulcer of the oesophagus and "oesophagitis". *The British journal of surgery*. 1950;38(150):175-82.

58. Siegmund KD, Marjoram P, Woo Y-J, Tavaré S, Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(12):4828-33.

59. Yatabe Y, Tavaré S, Shibata D. Investigating stem cells in human colon by using methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(19):10839-44.

60. Yatabe Y, Tavaré S, Shibata D. Investigating stem cells in human colon by using methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(19):10839-44.

Appendix I: Nested PCR primers and corresponding CSX locus

PCR 1	Primer sequences
Forward primer	ggggggagaa ggggctcca acat
Reverse primer	tagctgtcccc caggagtgcc ctc
PCR 2	Primer Sequences
Forward CSX	ggagacctag gaacttttc tgtccc
Reversed CSX	caatgagtga tcctgcagcc tgggtg

CSX locus

CSX locus

```

1021 agccggccac tgcgcgcgcc aacaacaact tctggaactt cggcgtcggg gacttgaatg
1081 cggttcagag ccccgggatt ccgcagagca actcgggagt gtccacgctg catggtatcc
1141 gagcctggta ggaagggac ccgcgtggcg cgaccctgac cgatcccacc tcaacagctc
1201 cctgactctc ggggggagaa ggggctcca acatgacct gagtcccctg gattttgcat
1261 tcaactctgc ggagacctag gaacttttc tgtcccacgc gcttttgttc ttgcacgj
1321 gagagtttgt ggcgcggatt atgcagcgtg caatgagtga tcctgcagcc tgggtgctta
1381 gctgtcccc caggagtgcc ctccgagagt ccatgggcac ccccggttg aactgggact
1441 gagctcgggc acgcagggc tgagatctgg ccgccattc cgcgagccag ggccgggcgc
1501 ccgggccttt gctatctcgc cgtcgcgcgc ccacgcacc acccgtattt atgtttttac
1561 ctattgctgt aagaaatgac gatcc

```

Forward and reversed primers PCR1
Forward and reversed primers PCR2
CpG sites

Appendix II. PCR solutions and programmes

PCR 1	solution	volume(μl) per sample
	10x MSP Buffer*	3.41
	15mM MgCL2	2.71
	25mM dNTPs	0.27
	Forward Primer	3.41
	Reverse Primer	3.41
	H2O	1.43
	Platinum Taq	0.34
	Sample	8
	Total	23

Programm		
Duration in minutes	Temper ature in °C	
5:00	95	
0:30	95	
1:30	61	5x 4
1:00	72	
5:00	72	
∞	10	

PCR II	solution	volume(μl) per sample)
	10x PCR buffer	2
	25mM MgCL2	2
	25mM dNTPs	0.3
	Forward CSX Primer	2
	Reverse CSX Primer	2
	H2O	10.2
	Platinum Taq	0.5
	Diluted PCR-I sample	1
	Total	20

Duration in minutes	Temper ature in °C	
5:00	95	
0:30	95	
1:30	61	0x 3
1:00	72	
5:00	72	
∞	10	



Colony PCR	solution	volume(μ l) per sample
	10x PCR buffer	2
	25mM MgCL2	1.2
	25mM dNTPs	0.16
	Forward Colony Primer	0.8
	Reverse Colony Primer	0.8
	H2O	9.89
	NEB enzym	0.15
	Colony sample	5
	Total	20

Duration in minutes	Temperature in °C	
2:00	94	
0:15	94	
0:15	55	0x 4
0:45	72	
5:00	72	
∞	10	

*MSP Buffer: 10x MSP buffer, $(\text{NH}_4)_2\text{SO}_4$ 166 mM, Tris pH 8.8 670 mM, MgCl_2 67 mM, 6-mercaptoethanol 100 mM