Chapter 1

General introduction
Epidemiology

With over 1.2 million newly diagnosed patients and more than 608,000 deaths annually, colorectal cancer (CRC) is one of the most frequent malignancies and one of the leading causes of cancer-related deaths in the world\(^1\). This also accounts for the Netherlands, in which each year 12,000 patients are diagnosed with CRC and 6,000 patients die from this disease\(^2\). Several factors influence the incidence of CRC, such as gender, age, and geographic region. Both men and women are affected at high rates, however, a higher incidence and mortality can be seen in men\(^3,4\). Like for many cancers, age is a significant risk factor to develop CRC. More than 90% is diagnosed in persons over age 50\(^5\). Although CRC occurs in all races, the incidence is highest in economically developed countries, such as North America, Europe, Australia and New Zealand. The incidence is lowest in developing countries such as Africa and South-Central Asia\(^5\).

Etiology

The etiology of CRC is complex and heterogeneous. Besides environmental factors, host-related factors, inherited and somatic (epi)genetic alterations contribute to the development of CRC.

Environmental and host-related factors

CRC incidence varies between economically developed and developing countries, and increases in populations that migrate from developing to developed countries\(^6\). Hence, environmental factors such as diet and lifestyle play important roles in the development of CRC. Risk factors include smoking, obesity, high intake of unsaturated fat and red/processed meat, excessive alcohol consumption and reduced physical activity\(^7,8\). Factors that decrease the risk of CRC include high intake of fruit, vegetables and fibers, the use of aspirin or non-steroidal anti-inflammatory drugs (NSAID), calcium, folate, and physical activity\(^7-9\). Furthermore, endogenous factors such as estrogens and androgens are associated with reduced risk of CRC, and insulin-related hormones with increased risk\(^10-12\). A recent model proposed inflammation as an underlying factor of the association with insulin, estrogen, and energy-related factors (obesity, physical activity, and energy intake) with CRC\(^11\). A role for inflammation as risk factor for CRC development is strengthened by an increased risk of CRC for patients with inflammatory bowel diseases such as long-standing ulcerative colitis and Chron’s disease\(^13\).

Hereditary factors

It is estimated that hereditary factors play a role in 15-30% of all CRCs. Most of these CRCs are considered familial, with low-penetrant genetic variations or single nucleotide polymorphisms (SNPs) associated with the increased risk\(^14\). In 5% of all cases, CRC arises in the setting of well-defined inherited syndromes, of which Familial Adenomatous Polyposis (FAP) and Lynch syndrome are most common. FAP is caused by a germline mutation of the adenomatous polyposis coli (APC) gene. Somatic inactivation of the second allele by mutation or less frequently by allelic loss (loss of heterozygosity; LOH) causes polyp formation. Patients with FAP develop hundreds to thousands polyps in the colon, which can develop into carcinomas if not removed\(^8,14-16\). In Lynch syndrome, unlike...
FAP, most patients do not have an unusual number of polyps. Lynch syndrome, previously known as hereditary nonpolyposis colon cancer (HNPCC), is mainly caused by germline mutation in the mismatch repair (MMR) genes MLH1, MSH2 and less frequent in MMR genes MSH6, PMS1 or PMS2. Somatic inactivation of the second allele, mostly by LOH, subsequently causes the formation of tumors. Recently, it has been discovered that in a sub-part of Lynch tumors MSH2 can be inactivated by another mechanism as well. Deletion of the last exons of the upstream gene EpCAM leads to gene-fusion with MSH2 and consequently allele-specific epigenetic gene silencing of MSH2. The disturbance of the MMR machinery causes an accumulation of DNA mutations throughout the genome (mutator phenotype) that leads to the formation of a tumor cell. Lynch tumors are mostly located in the proximal colon and characterized by microsatellite instability (MSI). Microsatellites are repeating units of DNA that occur normally throughout the genome and are susceptible to mutations due to replication errors by DNA polymerase slippage. MMR deficient cells fail to repair those errors. MSI is therefore used as a marker for a defect in the MMR machinery.

**Somatic alterations**

Somatic (epi)genetic alterations accumulate in the genome of all dividing cells as a result of DNA replication errors or exposure to mutagens. Some somatic alterations provide a selective growth advantage to the cells and can cause cancer (‘driver’ genes). Other alterations are biologically neutral and are expected not to contribute to tumorigenesis (‘passenger’ genes). The prevalence and signature of somatic alterations are highly variable per tumor, but irrefutable includes the activation of proto-oncogenes and inactivation of tumor suppressor genes, providing tumor cells the required biological capabilities to survive. These biological capabilities can be summarized into the so-called hallmarks of cancer: self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. Common somatic (epi)genetic alterations that occur in colorectal cancer are described in more detail in the section ‘Molecular pathogenesis of sporadic colorectal cancer.’

**Histology, pathology and the adenoma-carcinoma sequence**

The normal colon mucosa consists of three main components; the epithelium, a single layer of columnar cells that creates a barrier between the host and the colon lumen; the lamina propria; and the muscularis mucosae. The epithelial cells make finger-like invaginations into the underlying tissue, which are called crypts of Lieberkühn. The basis of a crypt contains ongoing dividing precursor cells that subsequently differentiate and migrate towards the luminal surface (Figure 1). Eventually the cells are exfoliated in the colonic lumen. This process of renewing the epithelium takes 3-8 days.

It is well established that many, if not all, carcinomas are preceded by pre-malignant precursor lesions called adenomas. The development of carcinomas from adenomas is referred to as the adenoma-carcinoma sequence. This does not imply that all adenomas will eventually become a carcinoma. In fact, it is estimated that only ~5% of all adenomas will ever progress to a carcinoma.
Chapter 1

Adenomas

A disturbed balance of cell proliferation and differentiation in the crypt epithelial cells leads to the formation of polyps. Polyps are quite common in persons over age 60 in countries with a Western lifestyle\textsuperscript{29-32}. The two most common histologic types are hyperplastic and adenomatous. Hyperplastic polyps, as the name says are hyperplasias, do not show dysplastic features and are considered to have no malignant potential\textsuperscript{33}. However, individuals with hyperplastic polyposis syndrome have an increased risk to develop CRC\textsuperscript{34}. Adenomatous polyps, or simply adenomas, show dysplastic features, such as nuclear hyperchromatism, stratification, or atypia\textsuperscript{33}.

The macroscopic appearance of adenomas can be grossly divided in polypoid and nonpolypoid. Polypoid adenomas protrude above the surrounding mucosa, showing a pedunculated morphology, i.e. with stalk, or sessile morphology. Nonpolypoid adenomas include lesions that are slightly elevated, completely flat, or depressed compared to the surrounding mucosa\textsuperscript{35-37}. The prevalence of nonpolypoid adenomas varies per study (8% to 42%), but is overall lower compared to polypoid adenomas. The progression to carcinoma is believed to be faster and more frequent in nonpolypoid adenomas\textsuperscript{25,36,38}. Yet, it is not clear whether these lesions have a more aggressive biology, or that their progression rate is higher because they are more easily missed by colonoscopy.

Besides shape, adenomas have more variable characteristics including size, histology, and dysplasia. These three pathologic characteristics are currently established as the most important determinants for malignant progression. From autopsies and endoscopic studies it has been shown that 70-95% of adenomas are less than 1 cm in diameter\textsuperscript{39-42}. The size of an adenoma has been linked to risk of progression with adenomas $\geq 1$ cm more likely to progress to carcinoma compared to adenomas $< 1$ cm\textsuperscript{43}. Adenomas can be classified in 3 major histological types: tubular, villous and tubulovillous. Adenomas with villous histology have the highest tendency to progress. Dysplasia is used to describe...
structural and cytological alterations of the epithelial cells. These abnormalities are classified as low-grade or high-grade dysplasia, of which the latter is linked to increased malignant potential. From this knowledge, literature has adopted the term ‘advanced adenoma’, which describes adenomas of ≥ 1 cm and/or villous histology (tubulovillous or villous) and/or high-grade dysplasia, and are considered as having high potential to progress. This classification, however, does not fully reflect the (epi)genetic heterogeneity and complexity of the disease. Because progression of colorectal cancer is driven by an accumulation of (epi)genetic abnormalities, molecular alterations might be more specific to discriminate adenomas with high-risk to progress. For example, chromosomal aberrations of progressed adenomas are highly comparable to that of cancers, independent of degree of dysplasia. The natural history of adenomas remains uncertain and difficult to study, however, since all polyps detected at colonoscopy are removed.

**Carcinomas**

The vast majority (~95%) of colorectal malignancies are epithelial adenocarcinomas, which develop from the epithelial cells of the colon mucosa. As mentioned above, it is well established that carcinomas are preceded by an adenoma, which can be any type of adenoma. It takes, on average, 8 to 12 years for an adenoma to evolve into a carcinoma. A lesion is considered malignant when the neoplastic cells pass through the muscularis mucosae and infiltrate the submucosa. Carcinomas occur throughout the colon, but about two-third are located in the left-side of the colon, i.e. distal from the splenic flexure (figure 2). Like adenomas, carcinomas are graded based on architecture and cytologic features, distinguishing low-grade (well and moderately differentiated) and high-grade (poorly differentiated) tumors. Besides, 10% to 20% of adenocarcinomas show an accumulation of mucin.

**Staging and survival**

Staging of colorectal cancer knows different systems. The oldest staging system is the Dukes classification, which was later modified by Astler and Coller. Dukes A tumors have invaded the submucosa, but do not extent the muscular wall of the colon. Dukes B tumors have grown through the muscularis propria into the serosa. Dukes C tumors have at least one regional lymph node metastasis and Dukes D tumors show distant metastasis. Although the Dukes classification is still being used, the tumor, node, metastasis (TNM) staging system of the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) is now the advised standard for colorectal cancer staging. In the TNM system, the “T” refers to the depth of invasion of the primary tumor at time of diagnosis. The “N” refers to the presence of local lymph node metastases, and “M” refers to the presence of distant metastatic disease. Table 1 shows the different staging systems and their classifications.

Survival of colorectal cancer is inversely related to the stage at diagnosis, with 5-years survival rates of more than 90% for stage I colorectal cancer to up to ~20% for stage IV colorectal cancer. At time of diagnosis less than 40% of all colorectal cancer patients present with localized disease (stage I or II), and about 20% of all colorectal cancer patients have metastatic disease (stage IV). A subgroup of stage II colorectal
cancer patients, which do not have metastasis at time of surgery, eventually relapse from recurrent tumors after months or even years after treatment. It is on forehand not predictable which patients will relapse and which will not. Tumor cells mainly spread via the blood and lymphatic system, with liver as the most common site of metastasis. Other sites of metastases are the lungs, peritoneum, pelvis and adrenals, which often become involved only after hepatic or lymphatic metastases.

Treatment

Treatment is dependent on the tumor stage at time of diagnosis. Surgical resection is the first treatment option for most tumors, and is the only treatment with curative intent. Subsequently the resected specimen is used for pathological examination to determine the stage of the tumor and to assess prognosis. Adjuvant chemotherapy is used to eradicate potential micrometastasis in patients with stage III tumors and in a subset of patients with stage II tumors, and thereby reduces the risk of relapse. In patients with advanced disease (i.e. stage IV), chemotherapy is mainly palliative, with the aim to prolong overall survival and maintain quality of life for as long as possible. However, metastases restricted to the liver are potentially curable with surgery as well. Rectal cancers are often treated with neoadjuvant chemoradiation before surgically resected.

5-Fluorouracil (5-FU) has been the only chemotherapeutic regimen for years.
availability of irinotecan and oxaliplatin last decade has improved patients’ outcome substantially and standard 1st line chemotherapy now includes fluoropyrimidine-based therapy combined with either irinotecan or oxaliplatin. Novel targeted therapies against angiogenesis (bevacizumab; a monoclonal antibody that targets vascular endothelial growth factor-A (VEGF-A),) and epidermal growth factor receptor (EGFR) (cetuximab and panitumumab) have further improved response rates. Bevacizumab combined with fluoropyrimidine-based chemotherapy is frequently used as first-line treatment for metastatic colorectal cancer.

### Early detection by screening

Patients can be cured from CRC when the tumor is detected and removed in an early stage. Since CRC does not often cause symptoms before it has reached an advanced stage, only a minority of patients is diagnosed while the tumor is still localized. The only realistic way to decrease the number of deaths from CRC is by means of screening. Indeed, incidence and mortality rates of CRC have declined in countries where screening has been introduced. For screening to be beneficial, a disease should meet certain criteria as developed by Wilson and Jungner. Accordingly, CRC is a suitable candidate for screening since it has a high prevalence, it has a well-defined precursor lesion (adenoma) and it has good treatment options. A variety of screening methods are available. Because screening involves healthy individuals, the screening method should be simple, safe, precise and validated.

Colonoscopy is the gold standard for detecting colorectal tumors, with a sensitivity of >90% for lesions ≥1 cm in size. During the procedure, adenomas can be removed and biopsies of larger lesions can be obtained for pathological examination. Colonoscopy is not without risk of complications, is an expensive procedure, and needs well trained specialists to be performed, and would therefore not be the first choice for mass-screening. Other available screening modalities are flexible sigmoidoscopy, CT-colonography, and fecal occult blood testing (guaiac-based fecal occult blood test (gFOBT) or immunochemical fecal immunochemical test (FIT)). These methods need follow-up by colonoscopy when a lesion is found or suspected. FIT is currently the most widely used screening method.

### Table 1: The different classification systems in colorectal cancer and corresponding survival rates

<table>
<thead>
<tr>
<th>Stage</th>
<th>AJCC/UICC TNM</th>
<th>Modified Astler Coller</th>
<th>Dukes</th>
<th>5-year survival rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Tis, N0, M0</td>
<td>N/A</td>
<td>N/A</td>
<td>96%</td>
</tr>
<tr>
<td>I</td>
<td>T1, N0, M0</td>
<td>Stage A</td>
<td>A</td>
<td>97%</td>
</tr>
<tr>
<td>II</td>
<td>T2, N0, M0</td>
<td>Stage B1</td>
<td>A</td>
<td>97%</td>
</tr>
<tr>
<td>IIIA</td>
<td>T3, N0, M0</td>
<td>Stage B2</td>
<td>B</td>
<td>88%</td>
</tr>
<tr>
<td>IIB</td>
<td>T4, N0, M0</td>
<td>Stage B3</td>
<td>B</td>
<td>72%</td>
</tr>
<tr>
<td>IIIA</td>
<td>T1, T2, N1, M0</td>
<td>Stage C1</td>
<td>C</td>
<td>88%</td>
</tr>
<tr>
<td>IIIB</td>
<td>T3, T4, N1, M0</td>
<td>Stage C2, C3</td>
<td>C</td>
<td>65%</td>
</tr>
<tr>
<td>IICC</td>
<td>Any T, N2, M0</td>
<td>Stage C1, C2, C3</td>
<td>C</td>
<td>55%</td>
</tr>
<tr>
<td>IV</td>
<td>Any T, Any N, M1</td>
<td>Stage D</td>
<td>N/A</td>
<td>11-20%</td>
</tr>
</tbody>
</table>

AJCC/UICC, American Joint Committee on Cancer/ International Union Against Cancer; TNM, tumor, node, metastasis; N/A, not available
tool in practice. Unfortunately, only half of individuals with a positive FIT test actually have an advanced adenoma (40%) or CRC (8%) in their colon\textsuperscript{65}. Consequently, half of individuals with a positive FIT test are offered futile colonoscopy. On the other hand, among individuals with a negative FIT test there is still a significant number of individuals that do have a colorectal tumor. Especially high-risk precursor lesions, i.e. advanced adenomas, are mostly left undetected by FIT\textsuperscript{66-68}, but also around 30\% of cancers are missed\textsuperscript{66,69}. Non-invasive stool- and blood-based molecular tests are in development and are expected to be more sensitive, specific, and informative compared to current non-invasive screening tools, but are not yet used in clinical practice (for a comprehensive overview see reference \textsuperscript{70} and \textsuperscript{71} (chapter 2 of this thesis).

The opinion on which screening approach is the best differs between the USA and Europe. In Europe there is a stronger preference for organized, programmed screening using a non-invasive test, such as the FOBT, whereas in the USA sporadic screening using colonoscopy is most applied\textsuperscript{72}. The overall effectiveness of CRC screening will depend upon the screening test characteristics (sensitivity and specificity), upon the compliance of individuals to the offered screening and to further evaluations when tested positive, and upon the quality of care delivered when a lesion is detected\textsuperscript{72}. In the Netherlands, population-wide screening to CRC will be implemented in 2013, using FIT as pre-selection for colonoscopy. All individuals between 55 and 75 of age will be invited to participate once every two years. With a compliance of 60\%, the number of prevented deaths from CRC is estimated to be 1400 per year\textsuperscript{66}.

Molecular pathogenesis of sporadic colorectal cancer

On the molecular level, colorectal cancer should be considered as a collective term for a heterogeneous group of diseases. Not only do hereditary types form a distinct class of tumors, sporadic CRC evolve through multiple pathways as well. In general the molecular pathogenesis of CRC is associated to the accumulation of genetic and epigenetic alterations, leading to alterations in tumor suppressor genes and oncogenes that transform cells and promote tumor progression. For CRC, the two main and best recognized pathways to tumor progression are the chromosomal instability (CIN) pathway, comprising \~85\% of all CRCs, and the microsatellite instability (MSI) pathway, comprising \~15\% of all CRCs\textsuperscript{17,73,74}.

Chromosomal instability

Chromosomal instability (CIN) is characterized by numerical and/or structural chromosomal aberrations, resulting in losses and gains of whole or parts of chromosomes. These chromosomal aberrations show a non-random distribution and are associated to different times of occurrence in colorectal tumorigenesis. Common observed aberrations are losses of chromosome 1p, 4q, 5q, 8p, 14, 15q, 17p, 18, 20p and 22q and gains of chromosome 1q, 7, 8q, 13q, 19q and 20\textsuperscript{45,46,75-78}. The observation that these aberrations are already present in progressed adenomas implicates their involvement in the progression from adenomas to carcinomas\textsuperscript{45,46}. Losses of 14q and gains of 1q, 11, 12p, and 19 appear to be late events\textsuperscript{76}.

The fact that these chromosomal aberrations show a non-random distribution indicates
that these changes provide a growth advantage and lead to clonal expansion of the altered cells. The affected chromosomal regions therefore may harbor genes that play a key role in colorectal carcinogenesis. Indeed, tumor suppressor genes are often located in frequently deleted chromosomal regions, such as APC (5q) and SMAD4 (18q) and chromosomal gains can increase the expression of proto-oncogenes, such as c-MYC (8q) and EGFR (7p)\textsuperscript{77-80}. Identifying the relevant cancer-related genes in affected chromosomal regions is not always straightforward because many chromosomal alterations span large regions that harbor multiple genes and more than one gene may be important. Gain of chromosome 20q, for example, is strongly associated with colorectal adenoma to carcinoma progression and is achieved by gain-of-function of multiple cancer-related genes at chromosome 20q, rather than affecting a single driver gene\textsuperscript{45,81}. On the contrary, many genes in the same chromosomal region might be affected, but may not be causally involved in the neoplastic process.

**Microsatellite instability**

Microsatellite instable (MSI) tumors are characterized by a high mutation rate because of a defective mismatch repair (MMR) system. In contrast to the germline mutation of the MMR genes seen in Lynch syndrome, somatic mutations in MMR genes have rarely been found in sporadic colorectal MSI tumors. Rather, the majority of sporadic MSI tumors have their MMR system inactivated by somatic epigenetic silencing of MLH\textsuperscript{178,82}. The consequence is equivalent, however, with many unrepaired replication errors throughout the genome. Especially repetitive nucleotide sequences, known as microsatellites, are prone to these mutations and form the markers to measure the defective MMR machinery\textsuperscript{83}. The actual cause of tumorigenesis is the increased mutation rate of target genes that function in various critical cell functions including DNA repair (RAD50, MSH3, MSH6, BLM), apoptosis (APAF1, BAX, BCL10), signal transduction (TGFβRII, ACTRII, IGFIIR, WISP3), cell cycle regulation (PTEN, RIZ), transcription factors (TCF4), and other putative ‘driver’ genes\textsuperscript{83,84}. These target genes have in common that they have microsatellite repeat regions in their coding regions, which are prone to mutation in MMR deficient cells.

MSI tumors are phenotypically and clinically different compared to CIN tumors. MSI tumors are characterised by proximal location, mucinous histology, poor differentiation, and lymphocytic infiltration. In addition, patients with MSI tumors have a better prognosis\textsuperscript{51}.

**Genetic alterations and dysregulated pathways**

Colorectal cancer is one of the best characterized tumor types in terms of genetics. Fearon and Vogelstein have established a genetic model for colorectal tumorigenesis in 1990, describing the adenoma-carcinoma sequence in which genetic changes in APC, KRAS, DCC, and TP53 occur in a preferred -but not essential- order\textsuperscript{85}. Research has built upon this model since and has added substantial knowledge to the biology of CRC and the key cellular signaling pathways that are dysregulated.

APC plays a prominent role in colorectal tumors and is mutated early in tumorigenesis. Up to 80% of colorectal tumors, both adenomas and carcinomas, have mutations that lead to a truncated APC protein\textsuperscript{86}. APC is a multi-functional protein, but the most
prominent role in colorectal tumorigenesis is involving the Wnt signaling pathway. The Wnt signaling pathway functions by regulating the amount of the transcriptional coactivator beta-catenin, which in turn controls gene expression programs involved in development and homeostasis e.g. cell proliferation and differentiation. Wnt signaling is well controlled in normal cells. In the absence of Wnt, cytoplasmic beta-catenin is continuously targeted for ubiquitin-mediated proteasomal degradation by a complex of proteins, including AXIN, APC, CK1 and GSK3. When Wnt binds to cell-surface receptors of the Frizzled and LRP families, AXIN is recruited to the receptor complex leading to disruption of the beta-catenin degradation complex. As a consequence, beta-catenin accumulates, enters the nucleus, and forms a complex with TCF to activate transcription of Wnt target genes. When APC is mutated, the degradation complex cannot form, and beta-catenin accumulates in the absence of Wnt. Beta-catenin subsequently translocates to the nucleus where it acts as a transcriptional co-activator of target genes including oncogenes like c-MYC and Cyclin-D, eventually leading to neoplastic transformation.

Besides mutations in APC, allelic loss (LOH) or epigenetic inactivation of APC by DNA promoter hypermethylation occurs, as well as mutations in other Wnt signaling genes, such as beta-catenin and AXIN1/2.

Other pathways commonly affected in CRC are those of TGF-β (mutations in TGF-β receptor I and II, mutations and deletions of SMAD2, SMAD4), EGF/MAPK (mutations in KRAS, BRAF, amplification of EGFR), PI3K (mutations in PIK3CA and PTEN), Notch (consistently activated, no mutations described in CRC, but cross-talk with RAS signaling exists), and Hedgehog (enhanced activation, no mutations described in CRC).

Obviously, these dysregulated pathways lead to altered processes in a cell, but also cause an accumulation of genetic alterations that do not necessarily all play a role in the neoplastic process. Systematic analyses of genetic alterations in CRC have identified recurrent mutations and chromosomal aberrations that are likely to play a functional important role during colorectal carcinogenesis and considered as ‘driver’ or candidate cancer (CAN) genes. This does not imply, however, that low-frequency aberrations are not biologically or clinically important. These low-frequency aberrations may represent a different mechanism for tumor initiation or progression in a small sub-set of colorectal cancers.

Epigenetics

Besides genetic alterations in cancer, it is increasingly clear that epigenetic alterations play an equal important role as well. Epigenetics refers to changes in gene expression, which are maintained through cell division, without alterations in the actual genomic sequence. Basically, epigenetic mechanisms are responsible for the different phenotypes of cells throughout our body whereas they all contain the same genetic information. The three main epigenetic mechanisms are DNA methylation, histone modifications, and small non-coding RNAs.

DNA methylation

DNA methylation is the covalent binding of a methyl group (CH3) to the 5’-position of a cytosine in the context of a CpG dinucleotide. Overall, these CpG dinucleotides are
underrepresented in the genome, but occur in higher density in short stretches of DNA (500–4000 bp) called CpG islands. These CpG islands only represent ~1% of the human genome, but cover more than half of all genes\textsuperscript{100-102}. The addition of methyl groups to the DNA is carried out by a group of enzymes called DNA methyl transferases (DNMTs). DNMT3a and DNMT3b are responsible for \textit{de novo} DNA methylation, which means methylating DNA that is unmethylated for both strands. DNMT1 is responsible for \textit{maintenance} DNA methylation, meaning that DNA methylation is copied to the newly synthesized DNA strand after replication\textsuperscript{103,104}. Established DNA methylation consequently involves transcriptional repression. This occurs either direct by inhibiting transcription factors to bind to their binding sites, or indirect by recruiting methyl-binding domain proteins (MBDs) and histone modifying enzymes resulting in chromatin condensation\textsuperscript{100,105}.

DNA methylation is important in several physiological processes, such as imprinting (parent-of-origin specific allele silencing), X chromosome inactivation in females, silencing of transposons and repetitive elements, and ageing\textsuperscript{100,106}. In normal cells the CpGs throughout the genome are methylated, whereas CpG islands remain unmethylated. In cancer cells DNA methylation patterns are disturbed and the main observed events are global, genome-wide, hypomethylation and at the same time regional hypermethylation of CpG islands. Functionally, global hypomethylation can cause chromosome instability, activation of transposons, and loss of genomic imprinting. However, it has also been described that proto-oncogenes can become hypomethylated, leading to increased expression. DNA hypermethylation mainly occurs in promoter regions of (tumor suppressor) genes, thereby causing gene silencing\textsuperscript{100,107-109}. DNA methylation thus represents an alternative mechanism for genetic alterations that lead to chromosomal instability, activation of proto-oncogenes and inactivation of tumor suppressor genes. Changes in DNA methylation can affect both alleles of a certain gene, but can also be added to the Knudson’s two-hit model in which DNA hypermethylation can be seen as the second inactivating change besides DNA mutation or chromosomal deletion\textsuperscript{100,110}.

Aberrant DNA methylation is a well-recognized phenomenon in colorectal cancer. Many (tumor suppressor) genes are described to be hypermethylated in CRC and affect the same signaling pathways that are targeted for mutational events\textsuperscript{111,112}. A good example is MLH1, which is mutated in hereditary colorectal cancer, but is bi-allelically methylated in sporadic colorectal cancers and causes MSI in these tumors\textsuperscript{82}. Another example is the transcriptional inactivation of negative regulators of the Wnt signaling pathway by DNA methylation, such as \textit{SFRP} family members 1, 2, 4, and 5\textsuperscript{113,114}, \textit{DKK1-4}\textsuperscript{115,116}, \textit{DACT3}\textsuperscript{117}, \textit{AXIN2}\textsuperscript{118,119}, \textit{APC}\textsuperscript{119,120}, \textit{SOX7} and -17\textsuperscript{121,122}, \textit{WIF-1}\textsuperscript{114} and \textit{Wnt5A}\textsuperscript{123}.

A subset of CRCs shows a widespread presence of cancer-specific hypermethylated genes, a phenomenon which is termed CpG Island Methylator Phenotype (CIMP)\textsuperscript{124,125}. The definition of CIMP, that is, which genes should be methylated and how many genes should be examined to define CIMP, is still under debate although the existence of CIMP is widely accepted\textsuperscript{125,126}. CIMP tumors are believed to be molecularly and biologically distinct. They are associated with older age, female sex, family history of CRC, proximal location, mucinous cell differentiation, specific precursor lesions (serrated adenomas), smoking, MSI, \textit{BRAF} and \textit{KRAS} mutations\textsuperscript{112}. To make it more complicated, within CIMP different subclasses have been described, recognizing CIMP-low or
CIMP-high, but also distinct methylation patterns that associate to specific molecular characteristics\textsuperscript{127,128}.

Genome-wide methylation analyses are currently adding substantial knowledge on differential methylated DNA in (colorectal) cancer. The well recognized concept of linking promoter CpG island methylation to gene silencing needs some adaptations. For many genes, the location of hypermethylation within the promoter CpG island varies per gene but also per tissue-type. Hence, the location of DNA methylation is crucial for the functional effect\textsuperscript{129}. Also DNA methylation outside the classical promoter CpG island methylation, such as CpG island shores\textsuperscript{130}, first exons\textsuperscript{131} and throughout a gene (gene-body methylation)\textsuperscript{132} have functionally important effect on gene regulation.

**Histone modifications**

In its natural state DNA is packaged into chromatin, which is a highly structured and dynamic complex, consisting of histones and non-histone proteins. The fundamental unit of chromatin is the nucleosome, which is composed of an octamer of four core histones H2A, H2B, H3, and H4, around which 147 base pairs of DNA are wrapped. Histone H1, the linker protein, is bound to DNA between nucleosomes. The N-terminal tails of the core histones protrude out of the nucleosomal core structure and are subject to posttranslational modifications, including methylation, acetylation, ubiquitination, phosphorylation, deamination, sumoylation, ADP-ribosylation and proline isomerisation. Histone modifications regulate chromatin structure and function, such as gene transcription, DNA repair, DNA replication, and DNA recombination. Hence, altered histone modification patterns affect gene expression of crucial genes involved in crucial cellular pathways, as well as genome integrity and/or chromosome segregation, which eventually may result in neoplastic transformation\textsuperscript{133-135}. Histone (de)acetylation and (de)methylation are the best characterized histone modifications. Acetylation of histones is a hallmark of open chromatin structure and transcriptional activity, and takes place on lysine residues of all four core histones. Methylation of histones are associated with either transcriptional active or transcription silent genes depending on their position, and takes place on lysine and arginine residues on histone H3 and H4\textsuperscript{134}. Altered histone modification patterns in cancer are the consequence of deregulated activity of proteins responsible for ‘writing’, ‘reading’ and ‘erasing’ histone modifications\textsuperscript{135,136}. In colorectal cancer, this includes increased expression of lysine acetyltransferases (CBP and P300)\textsuperscript{137}, increased expression of lysine deacetylases (HDAC1 and -2)\textsuperscript{137,138}, increased expression of methyltransferase SUV39H1 (responsible for the repressive mark H3K9 methylation)\textsuperscript{139,140} and dysregulated expression of polycomb group proteins (PcGs) (EzH2, Suz12, BMI1)\textsuperscript{141-143}. Polycomb-group proteins are a family of proteins that interact with chromosomal elements to repress genes important for developmental pathways by the repressive mark H3K27 triple methylation\textsuperscript{142}. In addition, by cross-talk to DNMTs, PcG-mediated transcriptional silencing might predispose their target genes to promoter CpG island hypermethylation, since many genes reported to be hypermethylated in CRC are PcG target genes\textsuperscript{132,144}.

**MicroRNAs**

Besides epigenetic mechanisms involving DNA and histones that modify gene expression profiles by changing the accessibility for several factors to the DNA, gene expression
is also post-transcriptionally controlled by small non-coding RNAs, such as microRNAs (miRNAs)\textsuperscript{143}. MiRNAs are ~22 nt non-coding RNAs that can hybridize to the 3’ untranslated regions of mRNAs from protein-coding genes and repress their expression. They do so mainly by mRNA cleavage, when a perfect complementarity exists between the miRNA and the target sequence and, for a smaller part, by translation inhibition, when no perfect complementarity exists between the miRNA and the target sequence\textsuperscript{146,147}. Each miRNA can target multiple different mRNAs and each mRNA can on its turn be targeted by more than one miRNA\textsuperscript{148}. The regulatory network of miRNAs is therefore very complex. miRNAs are highly conserved across different species, are highly specific for tissue and developmental stage, and play crucial functions in the regulation of important processes, such as development, proliferation, differentiation, apoptosis, and stress response\textsuperscript{149}.

Aberrant miRNA expressions contribute to the pathogenesis of all cancers, including CRC. miRNAs can function either as potential oncogenes or tumor suppressor genes, depending on the target genes they regulate\textsuperscript{150}. Dysregulation of miRNA expression in cancer occurs through mutations, epigenetic silencing, and dysregulation of transcription factors, and miRNAs are also frequently located in fragile or amplified chromosomal regions\textsuperscript{151}. As an example to the latter, the expression of miRNA-17-92 cluster has been shown to have an increased expression in association with 13q copy number gain, one of the major chromosomal aberrations in colorectal adenoma to carcinoma progression\textsuperscript{152}. Other examples of miRNAs involved in CRC are miRNAs involved in epithelial differentiation (miRNA-141 and miRNA-200c), WNT signaling (miRNA-145, miRNA-135a and miRNA-135b), and migration and invasion (miRNA-373 and miRNA-520c)\textsuperscript{112}.

**Interplay between genetics and epigenetics**

It is obvious that cancer is not only a genetic disease, and that epigenetic alterations have at least an equal important contribution. It is likely that these alterations do not occur as independent mechanisms in a tumor cell but rather are influenced by, or related to one another. They co-evolve in the same cell, accumulating changes in gene expression during tumorigenesis, thereby creating the ideal balance in different gene expression networks, so called ‘pathways’, for a cancer cell to survive. Several aspects of a relation between genetics and epigenetics have already been mentioned above, such as epigenetically silenced hMLH1 leading to MMR deficiency and accumulation of genetic mutations, DNA hypomethylation leading to chromosomal instability, the combination of DNA hypermethylation and mutation or deletion in Knudson’s two-hit model for inactivating tumor suppressor genes, or the increased expression of miRNAs due to DNA copy number gain. Other examples are germline sequence variants (single nucleotide polymorphisms; SNPs) that make gene promoters prone for being methylated in cancer cells, e.g. \textit{MGMT} in lung cancer and \textit{MLH1} in CRC\textsuperscript{153,154}, or double strand breaks that initiate gene silencing and DNA methylation\textsuperscript{155}. The emerging field of genome-wide (epi) genetic and expression profiling, and integration of these different data, will provide us with a better insight of the complex interactions of CRC genetics and epigenetics, which will help us to better understand the disease and hence to develop novel clinical applications\textsuperscript{112}. 
From bench to bedside - the use of molecular markers for early detection and predicting therapy response

The increasing knowledge on the molecular characteristics of CRC provides a large window of opportunities to improve current clinical practice. Molecular changes related to the neoplastic process can serve as highly specific markers for early detection, prognosis, therapy response prediction, and surveillance. Molecular markers can be measured in body excrements and fluids such as stool and blood, as well as in the tumor tissue itself, depending on the clinical purpose. A well-known example is serum carcinoembryonic antigen (CEA), which is widely used as surveillance marker following curative resection for primary CRC.

Molecular markers for early detection

Molecular markers have potential to improve current screening tests for the early detection of CRC. The fecal immunochemical test (FIT) is a widely used screening tool and detects small traces of human blood in stool. FIT is actually an example of a molecular test, because it detects the human protein hemoglobin or its early degradation products. Yet, improvement is expected by testing for molecules in stool or blood that are more directly related to the disease process. Research in this field mainly focuses on the development of new biomarkers with better test-characteristics, large-scale validation of (combinations of) known markers, and optimization of test methods. A variety of molecules have been tested as candidate screening markers in stool and blood. Most emphasis currently is on DNA and protein markers. Among DNA markers there is trend to move away from mutation markers in favor of methylation markers, and to use a panel of markers instead of using single markers. The recent boost in proteomics research leads to many new candidate protein markers. Although in initial studies, some of these DNA or protein markers showed a better performance than the FIT or gFOBT, for which large-scale validation studies are required. Chapter 2 in this thesis provides a comprehensive overview of the performances of DNA, RNA, and protein markers for CRC detection in stool and blood.

Molecular markers predicting therapy response

Although CRC is a heterogeneous disease and several systemic therapeutic regimens are available, treatment protocols are mainly based on a one-size-fits-all concept. Only a subset of patients actually responds, and consequently many patients unnecessarily suffer from the toxic effect. Hence, there is a need for markers to select patients for treatment, in order to reduce toxicity as well as costs, and ultimately to develop individualized-therapy. Markers that correlate with outcome are either prognostic or predictive. Prognostic markers correlate with outcome independent of treatment, and predictive markers correlate to the response of a specific treatment. A good example of a predictive marker is KRAS mutation, which causes resistance to treatment with cetuximab or panitumumab (anti-EGFR therapeutic agents). Another example is the lack of benefit from 5-FU for patients with MSI tumors. Several prognostic and predictive markers have been described so far, but most are either not validated, or show inconsistent results among different studies. Given the molecular heterogeneity of CRC, it is expected that several sub-groups of patients exist that differ in their
prognosis and response on certain treatments, depending on their molecular profile. One single biomarker will therefore not be relevant for all CRC patients. Nevertheless, the development of biomarkers that result in response prediction for even a small group of CRC patients may already be of great value.

Aim and outline of the thesis

Survival from colorectal cancer (CRC) has improved over the past few decades, for which earlier detection and improved treatment options have played major roles. However, colorectal cancer remains one of the most common types of malignancies and still half of the patients diagnosed with colorectal cancer will eventually die. The advantages of early detection clearly highlight the ongoing need to develop cost-effective, highly accurate methods to detect colorectal lesions in an early stage. For patients that do present with advanced CRC several treatment options are available. However, most available drugs are registered as one size fits all whereas response rates vary substantially, highlighting the need for methods to identify a priori those patients that will benefit from a specific treatment.

Because colorectal cancer biologically is a heterogeneous disease we hypothesized that molecular biomarkers can contribute to these two clinical needs, i.e. improved early detection and improved therapy response prediction of advanced CRC. In this thesis we therefore aimed to identify molecular biomarkers (i.e. DNA methylation and proteins) that improve current non-invasive screening methods (chapter 2-5) and molecular biomarkers (i.e. DNA methylation) that can predict therapy response (chapter 6). In addition, we investigated genetic and epigenetic events in CRC and aimed to gain more insight in the complex interactions of CRC genetics and epigenetics, which will help us to better understand the disease and hence to develop novel clinical applications (chapter 7-8).

Chapter 2 presents an overview of the literature concerning molecular markers in stool and blood for the early detection of CRC. In chapter 3, the technical possibilities of detecting DNA methylation markers in stool were explored by investigating the analytical sensitivity of a stool-based DNA methylation assay and the rate of DNA degradation in stool over time. In chapter 4 we describe the identification of a novel hypermethylated gene, PHACTR3, which has potential to serve as a biomarker for early detection of colorectal cancer (CRC) in stool. In addition, we evaluated its complementary value to a Fecal Immunochemical Test (FIT). Chapter 5 presents a large proteome profiling study on stool samples from CRC patients and control subjects, in order to identify novel human protein biomarkers in stool that outperform or complement FIT. In chapter 6, we describe the identification of a novel hypermethylated gene in CRC, DCR1, which predicts response to treatment with irinotecan in patients with metastatic CRC. In chapter 7 we investigated epigenetic silencing of 5 genes, MBD1, CXXC1, SMAD4, DCC and MBD2, located on a 4 Mb region on chromosome 18q21; a chromosomal region which becomes frequently lost during colorectal carcinogenesis. Finally, in chapter 8, we tested the hypothesis that DNA promoter methylation is one of the mechanisms involved in controlling gene expression in chromosomal gained regions by integrating genome-wide DNA methylation, DNA copy number and expression data from CRC cell lines.
Chapter 1

References

General introduction

Chapter 1

103. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247-257.
General introduction


