CHAPTER 1

GENERAL INTRODUCTION
INTRODUCTION ON COLORECTAL CANCER

Epidemiology and Etiology

Cancer is a leading cause of death with an estimated six million deaths worldwide in 2008\textsuperscript{1,2}. Currently, colorectal cancer (CRC) is one of the most common cancers, with an estimated incidence of 1.2 million new cases in 2008\textsuperscript{1}. In women CRC is the second most common cancer following breast cancer while in men CRC ranks third following lung and prostate cancer\textsuperscript{2}. The incidence of CRC shows large variation worldwide with the highest incidence rates found in the more industrialized countries such as North America, Australia/New Zealand and Europe and the lowest incidence rates in Africa and South-Central Asia. This worldwide variation of CRC incidence suggests a role for environmental factors such as eating habits and lifestyle in the etiology of CRC. Indeed, especially red and processed meat consumption\textsuperscript{3}, smoking\textsuperscript{4} and alcohol consumption, a sedentary lifestyle\textsuperscript{5} and overweight or obesity, seems to be associated to a higher risk of colorectal cancer\textsuperscript{6}. Overall incidence of CRC is still rising in countries where previously CRC incidence was low, such as Spain and the Czech Republic, most likely due to an increase in cancer-related factors such as smoking and obesity\textsuperscript{2}.

Besides these environmental and lifestyle factors the main risk factors for CRC are age and family history. More than 90\% of all CRC cases occur in individuals older than 50 years\textsuperscript{7} and the risk of ‘sporadic’ CRC increases with 2-3 fold if first-degree relatives were also affected with the disease\textsuperscript{8,9}. In fact, an estimated 30\% of all colorectal cancers have some familial origin, of which only 2-5\% arise in the context of a well-defined hereditary syndrome. The most common CRC syndromes are familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC)\textsuperscript{10}. FAP is an autosomal dominant disorder in which patients develop hundreds to thousands of colonic adenomas during their teenage years, which will invariably lead to CRC. FAP is caused by inactivating germ-line mutations in the tumor suppressor \textit{adenomatous polyposis coli (APC)} gene\textsuperscript{10,11}. In the case of HNPCC one of the DNA mismatch repair genes (\textit{hMLH1}, \textit{hMSH2}, \textit{hMSH6}, \textit{hPMS1} and \textit{hPMS2}) is affected by a (germ-line) mutation. This will then result in defective mismatch repair, which will give rise to tumors at an early age often between 20-30 years old in multiple organs, including colon cancer\textsuperscript{12}.

Pathology

The large intestine or colon is lined with a single layer of epithelial cells, which are structured as invaginations into the underlying tissue, called crypts of Lieberkühn\textsuperscript{13}. At the base of these crypts stem cells reside that give rise to transit-amplifying cells, which also divide and will change into fully differentiated cells. These post-mitotic differentiated cells consist of absorptive cells or enterocytes which function in absorption of water and nutrients, and two types of secretory cells; goblet (secretion of mucus) and entero-endocrine cells (secretion of gut hormones)\textsuperscript{13,14}. The differentiated cells move up to the top of
the crypt where they will die by apoptosis and are finally shed into the colon lumen, in this way the whole intestinal epithelium is renewed every 4-5 days.

The majority of colon malignancies originate from the epithelial lining of the colon, resulting in adenocarcinomas. This development starts with the formation of a colon adenoma, which is a benign lesion resulting from a failure in the normal process of proliferation, differentiation and apoptosis. Although most colon adenomas have a polypoid structure, non-polypoid adenomas also exist and these can be flat, slightly depressed or elevated. These colon adenomas display mild to severe dysplasia but do not have the potential to invade through the basement membrane into the surrounding tissue or to metastasize. Based on histology adenomas are subdivided into three types; tubular, villous or tubulo-villous. Risk of progression of an adenoma is traditionally based on the histopathological features; size, villous architecture and severity of dysplasia. An adenoma larger than 1 cm and/or with a villous component and/or with high-grade dysplasia is called an advanced adenoma, which is generally regarded to have a higher risk of progression. Formal evidence for that risk is scarce, since a longitudinal study would be ethically difficult to perform.

The concept of colon adenomas that progress into carcinomas has been around for decades, firstly proposed by Morson and colleagues. Based on the incidence of finding a focus of cancer in adenomas it is estimated that only approximately 5% of all adenomas progress into a carcinoma. This process is estimated to take up to 10-20 years although there are also indications for a more rapid progression. Macroscopically CRC typically has a polypoid or ulcerating crater-like appearance. Histologically, mucinous or non-mucinous tumors can be distinguished as well as the grade of tumor differentiation. In addition to the neoplastic (epithelium) cells a large proportion of tumor mass consist of tumor stroma, which is a tumor-specific microenvironment that contributes to tumor progression. During colon adenoma-to-carcinoma progression the tumor stroma expands, a process that is hallmarked by formation of fibrotic tissue, ingrowth of blood vessels, (myo)fibroblast accumulation, and lymphocytic infiltration. The interplay between cancer cells and the surrounding tumor stroma results in production of growth signals as well as survival signals to evade apoptosis, facilitate migration and metastasis through remodeling of the extracellular matrix (ECM), and to provide oxygen and nutrients through angiogenesis.

**Staging and Prognosis**

Extent of tumor growth is staged according to the (TNM) classification of the American Joint Committee on cancer (AJCC) and the International Union against Cancer (UICC). The tumor classification (T) represents the degree of growth into the bowel wall, the node (N) the presence and number of lymph node metastases and the metastasis (M) the presence of distant metastases. The TNM classification was preceded by the Dukes staging (table 1).
The prognosis of a patient is mostly based upon the stage in which the tumor is detected\textsuperscript{26}. In an early stage when the tumor has not yet spread outside the colon the five-year survival is high (90\%) while for late stage tumors the five-year survival drops dramatically (12-19\%) (table 1)\textsuperscript{2,27,28}. Unfortunately, still for the majority of patients (57\%) the tumor has already spread outside the colon at the time of diagnosis (table 1).

Treatment recommendations are based upon the stage, whereby surgical removal of the tumor is the first treatment possibility followed by adjuvant chemotherapy for stage III and high-risk stage II patients. High-risk stage II patients are those where less than twelve lymph nodes were found (staging may not be accurate), the tumor is poorly differentiated, when there is vascular, lymphatic or perineural invasion, presence of obstruction or perforation at tumor presentation or when the tumor is classified as T4\textsuperscript{25}.

A typical adjuvant chemotherapy regime consists of fluorouracil (5-FU, intravenous or orally as capecitabine) in combination with leucovorin and oxaliplatin, also known as the FOLFOX regime\textsuperscript{29}. Targeted drugs for CRC treatment include the anti-EGFR antibody Cetuximab and the anti-VEGF antibody Bevacizumab\textsuperscript{30}.

**Clinical Needs**

Firstly, since early detection of the disease seems the most realistic approach to reduce CRC mortality, screening programs are being implemented in various countries\textsuperscript{31}. These screening programs will reduce CRC mortality although the available methods leave room for improvement\textsuperscript{20}. Screening methods are frequently based on stool tests (detecting haem or hemoglobin), although in some countries primarily colonoscopy-based screening is provided. There are obvious drawbacks of hemoglobin testing; since hemoglobin originates from red blood cells it is therefore not a specific marker for neoplastic cells. Moreover, not all tumors bleed or they bleed intermittently. Therefore a need for tumor-specific molecular biomarkers exists since such markers are expected to be more sensitive, specific and informative, and could improve the currently available tests for early detection of CRC\textsuperscript{32}.

<table>
<thead>
<tr>
<th>Dukes</th>
<th>TNM</th>
<th>UICC Stage</th>
<th>Stage at first diagnosis (%)\textsuperscript{**}</th>
<th>5-year survival (%)\textsuperscript{**}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Tis</td>
<td>NO MO</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>T1-2</td>
<td>NO MO</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>T3</td>
<td>NO MO</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>NO MO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>T1-2</td>
<td>N+ MO</td>
<td>III</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>N+ MO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>N+ MO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Tx</td>
<td>Nx MI</td>
<td>IV</td>
<td>20</td>
</tr>
</tbody>
</table>

\*Adapted from www.oncolline.nl

\**2001-2007 Siegel et al.27
Secondly, prognosis and treatment regimens are mainly dependent on tumor stage. However, still 20-30% of stage II patients will get a disease relapse, while on the other hand only about 15% of stage III patients actually benefit from adjuvant chemotherapy. This illustrates the unmet clinical need for biomarkers to aid clinical decision making in stage II and III CRC.

Thirdly, imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI) and contrast agents such as 18-fluorodeoxyglucose (FDG-PET) play an important role in clinical management of CRC. For example, imaging is routinely applied for assessment of disease spread and treatment response. The role of imaging modalities could be extended by application of (more) molecular markers in MRI and PET to better predict prognosis and treatment outcome. Therefore, molecular biomarkers such as proteins or small molecules, which can be applied for imaging are needed.

Fourthly, in patients in whom the disease has spread to distant organs, drug therapy is a major component of treatment and, with the recent introduction of targeted therapies, companion diagnostics for predicting response to these therapies are needed. For example, (lack of) response to anti-EGFR therapy can partially be predicted by KRAS mutation status, although other downstream molecules in the EGFR pathway (such as BRAF) can also result in lack of therapy response. For several other targeted drugs against CRC such as anti-VEGF therapy no reliable predictive biomarkers exist. The possible adverse effects of these targeted treatments for CRC patients and the high costs of these drugs demands better patient stratification based on predictive biomarkers.

Lastly, disease and treatment response monitoring is currently based on measurement of protein biomarkers in serum. In CRC these include CEA and CA19.9, with CEA being the marker of choice for monitoring response to systemic therapy in metastatic disease. If CEA and CA19-9 levels drop at least 50%, the positive predictive value for treatment response would be 60% and 52.2%, respectively. Obviously, additional more sensitive biomarkers could support clinical treatment decision-making.

Molecular biomarkers can potentially be used clinically to aid early detection, diagnosis and disease monitoring or to guide therapy selection for CRC patients. These different applications of protein biomarkers in several disease stages will be addressed in more detail in chapter two of this thesis.

Molecular Changes in Adenoma-to-Carcinoma Progression

Adenoma-to-carcinoma progression is a multistep process, in which a combination of genetic and epigenetic alterations is required for a cell to become malignant, a genetic model has been generated by Vogelstein et al. Underlying these alterations are genomic instabilities that result in an accumulation of gene mutations, DNA copy number changes, epigenetic modifications, post-transcriptional and post-translational changes. The two most common
forms of genomic instability are chromosomal instability (CIN) in about 85% of CRC cases and microsatellite instability (MSI) in about 15% of CRC cases\textsuperscript{12,41}.

**Chromosomal instability**

Chromosomal instability is characterized by an accumulation of multiplication, deletion or translocation of whole chromosomes or chromosome arms\textsuperscript{41}. The causes for CIN are only partly understood. There are indications that defects in the regulation of the mitotic checkpoint or in chromosome segregation can lead to CIN\textsuperscript{42}. Another potential cause is abnormal centrosome number and function, as the presence of extra centrosomes may lead to the formation of multiple spindle poles during mitosis, resulting in the unequal distribution of chromosomes\textsuperscript{41}. A key player in this process is the \textit{AURKA} gene that is involved in centrosome organization and multiplication, for which several studies have shown overexpression and a functional driving role in carcinogenesis\textsuperscript{43,44}. Other mechanisms that may have a causal role in the development of CIN are shortened telomere lengths and faulty DNA damage response, for example by \textit{TP53} mutations\textsuperscript{40}.

Ultimately, CIN will result in specific patterns of copy number aberrations (CNAs) whereby gains and losses in (parts of) chromosomes accumulate during tumor development, a process that already starts in adenomas\textsuperscript{45}. Gains of chromosome 8q, 13q, and 20q and losses of 8p, 15q, 17p and 18q are strongly associated with adenoma-to-carcinoma progression. These CNAs are therefore thought to contribute to adenoma-to-carcinoma progression\textsuperscript{20,43,45} and colon adenomas harboring such CNAs have been regarded as high-risk adenomas\textsuperscript{46}. Chromosomal gains and losses have also been linked to progression of disease\textsuperscript{47} and clinical outcome\textsuperscript{48,49}.

**Microsatellite instability**

The main function of mismatch repair is to recognize and repair DNA mismatches, short insertion loops and deletion loops that occur during DNA replication\textsuperscript{11}. Malfunctioning of mismatch repair results in an accumulation of mutations mostly in repetitive DNA sequences of non-coding regions, although some of them affect coding regions of genes such as \textit{TGF\betaRII}\textsuperscript{50}. Whereas germ-line mutations in the \textit{hMLH1}, \textit{hMSH2} and \textit{hMSH6} genes are responsible for defective mismatch repair in HNPCC patients\textsuperscript{51}, non-familial MSI tumors are frequently caused by hypermethylation of the promoter region of \textit{hMLH1}. While the proportion of MSI among all CRC cases is estimated at 15%, only 2-3% arises in the context of HNPCC, therefore the majority of colorectal tumors with MSI are sporadic\textsuperscript{12}.

Patients with MSI tumors usually have a better prognosis than patients with CIN tumors\textsuperscript{52}. MSI tumors are more frequently mucinous and can contain large amounts of tumor infiltrating lymphocytes. This may result from an immune reaction against neoantigens formed as a result of frameshift mutations in protein-coding sequences\textsuperscript{53}.
Alterations in key molecular and biological pathways

Cancer cells acquired the ability to sustain proliferation, enable immortality, sustain the supply of nutrients and oxygen (angiogenesis and reprogramming of cell metabolism) and become invasive to form metastases (figure 1). On other hand cancer cells also need to evade growth suppressors, cell death and immune destruction\textsuperscript{23}. In order to reach a malignant phenotype multiple genes and molecular pathways need to become altered. The most important affected molecular pathways are shortly described below.

Activation of the Wnt signaling pathway is an early event that occurs in about 80% of all colon adenomas, and is present in 93% of all CRCs. This is most often the result of a mutation in the \textit{APC} tumor suppressor gene\textsuperscript{40,54}. An \textit{APC} mutation will result in the activation of $\beta$-catenin which drives the transcription of genes involved in tumor growth and invasion\textsuperscript{55}. Other frequent genetic alterations include activating mutations in \textit{EGFR}, \textit{KRAS} and \textit{BRAF}, which will lead to constitutive activation of the RAS signaling pathway resulting in tumor growth. Ras signaling is normally activated by growth factors through e.g. EGFR, and stimulates cell growth, differentiation and survival. Ras signaling is altered in 59-80% of CRCs, and \textit{KRAS} mutations are already present in 40% of adenomas making it an early event in progression\textsuperscript{15,38,40}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Adenoma-to-carcinoma progression. The top panel shows the activation of Wnt signaling and genomic instability as well as some of the key biological pathways, which are altered during adenoma-to-carcinoma progression. The lower panel, displays hematoxylin and eosin stained examples of different stages of colon tumor development. Colorectal cancer develops from normal colon through a benign precursor lesion, the colon adenoma, via a high-risk adenoma or progressed adenoma to a carcinoma. Adenomas have undergone several molecular alterations compared to normal colon tissue, of which activation of the Wnt signaling pathway is most prevalent. The far majority of approximately 95% of adenomas will not progress into carcinomas. Genomic instability is a main characteristic of CRC, in most cases chromosomal instability (CIN) and in a minority microsatellite instable tumors (MIN) which have a lower frequency of metastases.}
\end{figure}
Also the well-known tumor suppressor gene \textit{TP53} is often inactivated. Normally the P53 protein is activated through cellular stress such as DNA damage or aberrant proliferative signals. When activated it initiates cell cycle arrest possibly followed by cell death through induction of apoptosis\textsuperscript{56}. Inactivation of P53 therefore enables proliferation of damaged cells. \textit{TP53} mutations have been identified in 50\% to 75\% of sporadic CRC and in a small subset of colorectal adenomas, indicating that \textit{TP53} mutation is important in adenoma-to-carcinoma progression\textsuperscript{38,40}.

Transforming growth factor beta (TGF-\(\beta\)) is involved in many cellular processes including regulation of cell adhesion, cell growth, differentiation and apoptosis. The receptors for TGF-\(\beta\) can activate the Smad proteins that regulate transcription of downstream genes. The TGF-\(\beta\) pathway is altered in 27\%-87\% of CRC tumors, mostly through inactivating mutations in \textit{TGF}\(\beta\)RII and \textit{ACVR2A}\textsuperscript{40}.

Insight in altered biological processes has also been generated by pathway analysis of genome-wide mRNA profiling. This has revealed up regulation of pathways involved in chromosomal instability, proliferation, differentiation, angiogenesis, stroma activation, invasion, cell cycle regulation/mitosis and old age whereas fatty acid metabolism was found to be down regulated\textsuperscript{57,58}.

\section*{PROTEIN BIOMARKERS}
\subsection*{Proteomics based biomarker discovery}
Proteomics aims at examination of the whole proteome, i.e. “the entire complement of proteins expressed in a specific state of an organism or a cell population”\textsuperscript{59}. In contrast to changes in DNA and mRNA expression, studies describing (near) genome-wide protein expression changes are scarce\textsuperscript{60}. The field of proteomics is still evolving and only in the last few years technical developments have allowed for such studies to be carried out. The methods available to profile these changes in protein expression are mass spectrometry (MS)-based proteomics and antibody-based proteomics\textsuperscript{59,61}.

In short, protein identification by mass spectrometry entails the ionization of proteins or peptides, the separation of those peptides based on their mass relative to their charge and finally detection of the intensity of these peptides. In a tandem mass-spectrometer the first MS cycle typically determines the peptide masses after which the peptides are fragmented and a second MS cycle detects the masses of the fragments. In this manner peptide masses and peptide sequences can be mapped against theoretical or empirically obtained mass spectra for identification. For protein identification the currently most widely applied approach is peptide separation via nanoliquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).

Protein biomarker identification based on mass spectrometry holds great promise for the identification of novel biomarkers with the potential to be used for clinical application, this topic is discussed in more detail in chapter two of this thesis\textsuperscript{62}.
**Tools for protein biomarker verification and validation**

Following protein biomarker discovery, a biomarker should be verified and finally validated in an increasing number of clinical samples\(^63\). A potential bottleneck in this process is the availability of suitable assays for such a verification and/or validation.

A widely used technique for visualization of protein expression in cells or tissues is immunohistochemistry. Hereby an epitope is microscopically visualized by the binding of an antibody to that epitope.

Especially the application of immunohistochemistry on tissue microarrays (TMAs) has enabled large-scale protein expression analysis with this technique\(^64\). Another major advance in this field has been the digitizing of tissue slides and TMAs to facilitate and standardize scoring of the results. The ongoing efforts of the Human Protein Atlas in antibody production as well as tissue characterization by immunohistochemistry are a major resource of information for protein biomarker research today\(^65\). Other antibody-based techniques include ELISA; mostly applied to detect antigens in liquid samples such as a serum, and western blotting to detect specific proteins in e.g. a cell lysate or tissue homogenate.

Selected reaction monitoring-MS (SRM) is an emerging novel technique in which protein quantification is possible in an antibody-independent and a multiplex manner, allowing for the validation of a large number of proteins in a large number of samples while taking up minimal instrument time\(^66\). This technique can support the verification of candidate protein biomarkers, so that large-scale validation efforts (usually antibody-based) have a higher chance of success.

**AIMS AND OUTLINE OF THIS THESIS**

CRC is one of the most prevalent and deadliest forms of cancer, whereby most patients already have late stage disease at the time of diagnosis. Currently, mortality is expected to decrease as a result of implementation of population-wide screening programs. However, there is still an evident need for specific biomarkers to improve the current available tests for early detection, prediction of prognosis, patient stratification for adjuvant drug therapy or disease monitoring. Molecular biomarkers that reflect the processes underlying adenoma-to-carcinoma progression and later development of metastases would be of high value in this respect. Proteins are responsible for many processes in the cancer cell and (antibody-based) protein detection would be easy to implement in routine clinical setting. Therefore, protein biomarkers have great potential for various applications in clinical practice. The aim of the studies described in this thesis was to identify protein biomarkers to address the various clinical needs for CRC patients.

Firstly an overview of the colon cancer proteomics field with a focus on studies directed at identification of biomarker proteins that carry potential for
a clinical application is provided in chapter 2. In this review we describe the status of the field and identify current bottlenecks for the implementation of protein biomarkers in clinical use. Aiming at the identification of biomarkers for early-stage CRC potentially applicable in a blood-based or stool-based test a mouse model for sporadic CRC was applied to identify proteins in mouse normal colon- and tumor-secretomes by LC-MS/MS in chapter 3. A similar study in which human colon tumor and normal colon tissue secretomes were collected and analyzed by LC-MS/MS is described in Chapter 4. The aim of this study was the identification of biomarkers potentially applicable in early diagnosis, prognosis, therapy prediction and/or disease monitoring of CRC. Following the same line, we investigated the actual biological fluid in which these biomarkers should be detectable in chapter 5. Here stool samples from CRC patients and healthy controls were analyzed by LC-MS/MS. This study aimed at the identification of novel protein biomarkers in stool that may outperform or complement hemoglobin as a test for the early detection of CRC. Finally a subset of candidates was verified in independent stool samples of healthy controls, adenoma, advance adenoma and CRC patients by SRM. Focusing on another type of clinical application we set out to identify cell surface protein biomarkers that are of particular interest for application as molecular markers for imaging approaches in chapter 6. The aim of this study was to identify cell surface protein biomarkers for colorectal adenoma-to-carcinoma progression that can distinguish clinically harmless low-risk adenomas from clinically relevant high-risk adenomas and CRC.

Tumor stroma plays an important role in the progression and metastasis of CRC. The glycoproteins versican and lumican were previously found to be overexpressed in colon carcinomas and are associated with the formation of tumor stroma. Chapter 7 describes the expression of lumican and versican in the tumor stroma and epithelium of adenomas and carcinomas and the potential association with commonly occurring genomic alterations in adenoma-to-carcinoma progression. In chapter 8 the potential prognostic value of lumican and versican expression in the epithelial and stromal compartment of stage II and III colon cancer was investigated. Chapter 9 describes the identification of the special AT-rich sequence-binding protein 2 (SATB2), a nuclear matrix-associated transcription factor and epigenetic regulator, which is a tissue type-specific protein. The aim of this study was to further validate the specific expression pattern of SATB2 by immunohistochemistry as a clinical biomarker for the differential diagnosis of carcinoma of unknown primary origin. Lastly in chapter 10 a summary of the results and a general discussion of those results is provided, aiming to review the results obtained in this thesis and to provide future directions for the field.
References