Aus dem Pathologischen Institut
der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. Thomas Kirchner

Identification of novel prognostic and predictive biomarkers in colorectal cancer

---

Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät
der Ludwig Maximilians Universität München
vorgelegt von

Manal Elmasry

aus Ägypten

2020

Supervisors:	Prof. Dr. med. David Horst
	Prof. Dr. rer. nat. Andreas Jung
Third expert:	Prof. Dr. rer. nat Christine Sers
Fourth expert:	PD Dr. med. Dominik Modest
Dean:	Prof. Dr. med. dent. Reinhard Hickel
Date of oral defense:	23.03.2020

Affidavit

I hereby declare, that the submitted thesis entitled

„ Identification of novel prognostic and predictive biomarkers in colorectal cancer ”

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich, 04.12. 2019

Manal Elmasry

My Family

PUBLICATIONS

Parts of this thesis have been published in:

- **Elmasry, M.,** Brandl, L., Engel, J., Jung, A., Kirchner, T. & Horst, D. RBP7 is a clinically prognostic biomarker and linked to tumor invasion and EMT in colon cancer. *J Cancer* **10**, 4883-4891, (2019) [1].
  - Author contribution: E.M. performed all the experiments in this publication.

In addition, I contributed to the following publications, which are not further described in this thesis:

- Spartalis, C., Schmidt, E. M., **Elmasry, M.,** Schulz, G. B., Kirchner, T. & Horst, D. In vivo effects of chemotherapy on oncogenic pathways in colorectal cancer. *Cancer Sci* **110**, 2529-2539, (2019).

- Schulz, G. B., Grimm, T., Sers, C., Riemer, P., **Elmasry, M.,** Kirchner, T., Stief, C. G., Karl. A. & Horst, D. Prognostic value and association with epithelial-mesenchymal transition and molecular subtypes of the proteoglycan biglycan in advanced bladder cancer. *Urol Oncol* **37**, 530.e539-530.e518, (2019).

CONTENTS

1 INTRODUCTION

1.1 Epidemiology of Colorectal cancer

1.1.1 Incidence and mortality

Cancer is a considerable public health problem, and is one of the main causes of death [2]. In 2018, an estimated 18.1 million new cancer cases and 9.6 million cancer deaths occurred worldwide, and among them 1.8 million colorectal cancer (CRC) cases causing 881,000 deaths. Moreover, Globally CRC is the second most common cancer in women and the third in men, and it is the third and fourth leading cause of cancer-related deaths in women and men, respectively [3,4]. Furthermore, it is predicted that by the year 2035, deaths from colon cancer would have increased by 60 % [5].

Despite advances in CRC early detection and treatment, metastasis remains the main cause of cancer related death, and nearly half of all patients with CRC present with metastasis either at the time of diagnosis (20 %) or in the fellow up period (20-30 %) [6,7]. Therefore, there is an urgent need to identify robust and informative biomarkers for risk stratification of CRC patients which could improve therapeutic decision making in routine clinical practice and CRC outcome.

1.1.2 Survival

Survival rates for colorectal cancer can differ based on a variety of factors, but it is highly dependent on stage of disease at diagnosis. The 5-year survival rate of patients with localized stage I and II CRC is up to 90 %. If the cancer has spread to the regional lymph nodes (stage III), the 5-year survival rate is about 71 %. If the cancer has metastasized to distant parts of the body (stage IV), the 5-year survival rate is only around 14 % [2].

1.1.3 Risk factors

The Risk of developing CRC can be increased by environmental and hereditary factors [8]. Some of the environmental risk factors are modifiable such as lifestyle-related factors including obesity, diet rich in red and processed meat, physical inactivity, smoking and moderate to heavy alcohol consumption [8-10]. Other risk factors are non-modifiable such as increasing age (50 years and over), male sex, inflammatory bowel disease, diabetes mellitus, and family history of CRC which accounts for 20 % of all CRC cases [11]. Hereditary factors accounts for about 3-5 % of all CRC cases [12]. The two most common forms of hereditary colorectal cancers are hereditary nonpolyposis colorectal cancer (HNPCC or Lynch syndrome) and familial adenomatous polyposis (FAP) [13]. Indeed, the majority of CRC cases, approximately 75 %, occurs sporadically without family history of CRC or apparent evidence of having inherited disorders [12].

1.2 Development of sporadic colorectal cancer

1.2.1 Molecular background of colorectal carcinogenesis

The development of CRC is a multistep process during which multiple genetic and epigenetic alterations sequentially accumulate in normal colonic epithelial cells and

transform them into malignant neoplastic cells [14-16]. Some of these molecular alterations are called ‘‘drivers’’, conferring selective growth advantages to the affected cells via activating proto-oncogenes, inactivating tumor suppressor genes and abnormalities in DNA repair genes, and ultimately give rise to malignant transformation of the affected cell by disturbing key signaling pathways that regulate normal cell proliferation, differentiation, and survival [17-19]. The remainder alterations are non-oncogenic and called “passengers” [20]. It was proposed that at least four or five genetic alterations (mainly *APC, KRAS, SMAD4,* and *TP53*) are required to drive the transformation of normal colonic epithelial cells into invasive carcinoma [14]. Afterwards, it was suggested that three driver gene alterations are sufficient for invasive carcinoma formation [21].

Furthermore, the advances in understanding the molecular genetics and epigenetics of CRC have enabled the identification of two pathways of histological/molecular development of colorectal cancer, beginning with two different precursor lesions: the adenoma-carcinoma sequence and the serrated neoplasia pathway; each pathway is associated with specific sequence of genetic and epigenetic alterations, standard clinical and histological characteristics and resulting in diverse phenotypes of CRC, which explains the heterogeneity of CRC [22].

1.2.2 Pathways of colorectal carcinogenesis

1.2.2.1 The adenoma-carcinoma sequence

The adenoma-carcinoma sequence is the classical pathway, proposed by Fearon and Vogelstein and accounts for approximately 60-70 % of sporadic CRCs [17]. It is a multistep mutation pathway, describing the gradual progression of normal colorectal epithelium to benign precursor lesions, called conventional adenomas (Adenomatous polyps), and ultimately to adenocarcinoma [14].

The initial step transforming normal colorectal epithelium into microscopic adenomas composed of a few dysplastic glands is often caused by inactivation of the tumor suppressor gene *adenomatous polyposis coli* (*APC*) [18,23,24]. *APC* mutations are found in about 70-80 % of colorectal adenomas and sporadic CRCs, and are essential for initiating adenoma formation and sustained activation of Wnt/β-catenin signaling pathway [17].

The second step involves activating mutations of the *KRAS* oncogen at codons 12, 13, 61 which lead to enlargement of preexisting adenomas to become more than 1cm [17,25]. *KRAS* mutations are present in about 40 % of colorectal adenomas and carcinomas [26], and lead to activation of GTPase enzyme that increase RAS signaling and promote tumor progression by MAPK signaling pathway activation (RAS/RAF/MEK/ERK/MAPK)[18]. Subsequently, malignant transformation is driven by loss of heterozygosity (LOH) of chromosome 18q in about 70 % of CRCs that is usually associated with mutational inactivation of tumor suppressor gene *SMAD4* which facilitates cell proliferation and malignant progression of CRC by inactivation of transforming growth factor β (TGF-β) pathway [27,28]. Additionally, inactivating mutations of the tumor suppressor gene *PTEN* or activating mutations of the oncogene *PIK3CA* may occur [29,30], and result in inhibition of apoptosis and promotion of tumor cell survival by activating PI3K pathway [31].

The final step is inactivation of the tumor-suppressor gene TP53 which encodes for p53, the main cell-cycle checkpoint which causes an uncontrolled entry in the cell cycle and is mainly detectable in colorectal carcinomas with an estimated frequency of approximately 60 % [24,25]. A small group of conventional adenomas develops through DNA mismatch repair (MMR) deficiency which result in CRC with high level microsatellite instability, termed (MSI-H) phenyotype [32] (Figure 1A).

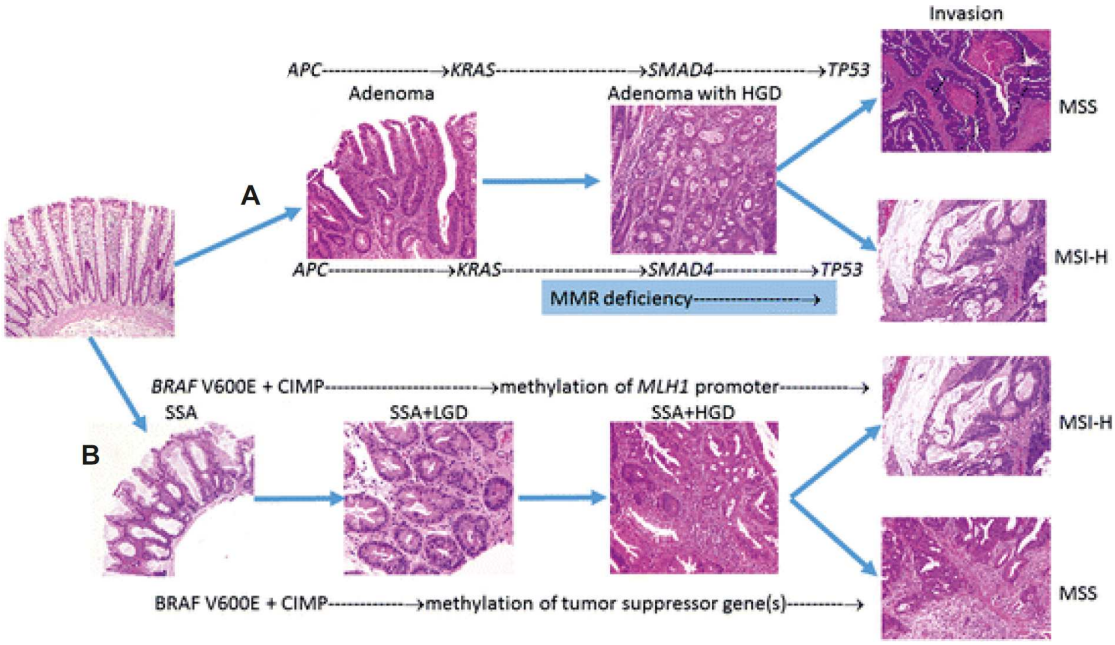

Figure 1. Pathways of colorectal carcinogenesis

(A) The classical pathway involves the progression of conventional adenomas to invasive adenocarcinomas with MSS or MSI phenotypes if MMR deficiency occurs. **(B)** The serrated neoplasia pathway involves serrated polyps and their progression to either MSI-H cancers if methylation occurs in MLH1 promoter, or MSS cancers if methylation occurs in tumor suppressor genes. The mutated or epigenetically altered genes during the progression sequence are shown. Figure from reference [33].

1.2.2.2 The serrated neoplasia pathway

The serrated neoplasia pathway is an alternative pathway for CRC development, described by Jass and Smith and accounts for approximately 30 % of all CRCs [34]. In the serrated pathway, serrated polyps including sessile serrated lesions (SSL) and traditional serrated adenomas (TSA) are recognized precursors for colorectal carcinoma and are histologically characterized by a “serrated” (or saw-toothed) appearance of the epithelial glandular crypts [35,36].

The development of serrated polyps is often initiated by activating *BRAF* oncogene mutations, more common at codon 600 (V600E), which promote cell proliferation through MAPK signaling pathway and found in about 5-15 % of CRC cases, mainly in SSL and TSA, or activating mutations in KRAS, specifically in TSA [37,38]. Then serrated polyps proceed to evident epithelial dysplasia and eventually serrated adenocarcinoma through Wnt signaling pathway activation and progressive CPG island methylation

which frequently inactivates DNA mismatch repair (MMR) gene, mutL homologue 1 (*MLH1)* and give rise to MMR-deficient CRCs with MSI-H phenotype, or TP53 mutations which result in MMR-proficient CRCs with microsatellite stable (MSS) phenotype [39-41] (Figure 1B).

1.2.3 Patterns of genomic Instabilities in colorectal cancer and hallmarks of cancer

1.2.3.1 Chromosomal instability

Chromosomal instability (CIN) is the most common form of genomic instability in CRC, observed in 85 % of sporadic CRC and adenomas, which are derived from the adenoma-carcinoma sequence [42,43]. CIN refers to widespread gains or losses of whole or large portions of chromosomes during cell division that can result from defects in chromosomal segregation, centromere function, telomere stability and DNA damage response, giving rise to karyotypic variation between cells [44-46]. Moreover, CIN is characterized by changes in chromosome copy number (aneuploidy) and frequent loss of heterozygosity (LOH) [47,48], which can drive CRC initiation and progression by inducing ongoing copy number alterations (CNAs) in large groups of genes, such as tumor suppressor genes, oncogenes and apoptotic genes [49,50].

1.2.3.2 Microsatellite instability

Microsatellite instability (MSI) is found in about 15 % of sporadic CRCs [42], and occurs as a consequence of defective DNA MMR system which is caused by inactivation of DNA MMR genes, most commonly *MLH1,* by promoter hypermethylation [51-53]. MSI is characterized by multiple mutations, mainly insertions and deletions, in repetitive sequences of DNA called microsatellites [54]. Clinically, based on the MSI status, CRCs can be categorized into MSI-high (MSI-H), if 30 % or more of the repeats are unstable; MSI-low (MSI-L), if fewer than 30 % of repeats are unstable, and microsatellite stable (MSS), if no repeats are unstable [55].

Furthermore, MSI-H encourages malignant transformation by allowing the accumulation of mutations in microsatellites located in DNA coding regions of specific genes that are implicated in tumor progression, such as *TGFβRII* and *BAX*, then generating frameshift mutations [56,57], and subsequently leading to sporadic CRCs which frequently carry *BRAF* (V600E) mutations [58].

1.2.3.3 CpG island methylator Phenotype (CIMP)

CIMP is found in about 20 % of CRC, and is characterized by widespread hypermethylation of promoter CpG islands of tumor suppressor genes or DNA repair genes such as *CDKN2A*, which encodes p16, or *MLH1* respectively, resulting in transcriptional silencing [51,59-61]. On the basis of aberrant CpG island methylation, CRCs can be classified into three subclasses: CIMP-1 (CIMP-high) which is more frequently associated with the MSI-H phenotype and with the presence of *BRAF* (V600E) mutations, CIMP-2 (CIMP-low), and CIMP-negative [39,62]. Additionally, CIMP-high facilitates the initiation and progression of CRC mainly from serrated polyps through the serrated neoplasia pathway [40,41].

To sum up, genomic instability has been recognized as an enabling characteristic of cancer as it accelerates the accumulation of genetic and epigenetic alterations that enable selective growth advantages and acquisition of essential functional capabilities which in turn critically drive colorectal tumors development and cancer progression [63] (Figure 2). These capabilities are not limited to the development of CRC, but are acquired during the development of almost all human cancer types, and were highlighted and termed “Hallmarks of cancer” by Hanahan and Weinberg [16,63].

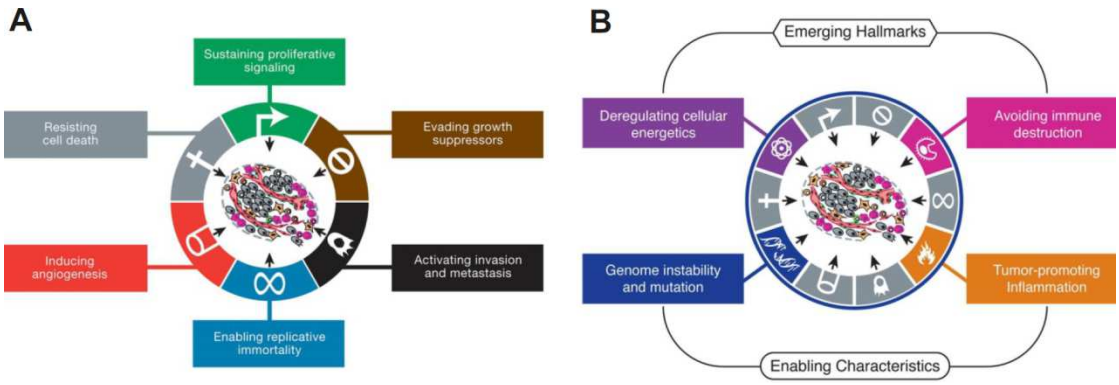

Figure 2. The hallmarks of cancer.

(A) The original hallmark capabilities acquired during the development of human cancers. **(B)** Emerging Hallmarks and Enabling Characteristics. Figure from reference [63].

The originally proposed hallmark capabilities of cancer are self-sufficiency in signals, insensitivity to growth inhibitory signals, escaping from normal programmed cell death, allowing limitless replication potential by maintaining telomere length, triggering angiogenesis, and activating invasion and metastasis [16] (Figure 2A). Moreover, the original list has been expanded by emerging hallmarks such as disrupting cellular energetics and evading immune response [63] (Figure 2B).

1.3 Metastasis of colorectal cancer

1.3.1 The metastatic cascade

Following the malignant transformation of normal colonic epithelial cells, the neoplastic cells continue to evolve by acquiring the ability to invade and metastasize, which is one of the hallmarks of malignancy [16,63]. During metastatic progression, each cancer cell

must accomplish a number of sequential and interrelated steps, known as the invasion-metastasis cascade [64]. This cascade starts with the local invasion of primary tumor cells into adjacent tissue, followed by dissemination of cancer cells and formation of secondary tumors at distant sites [65] (Figure 3).

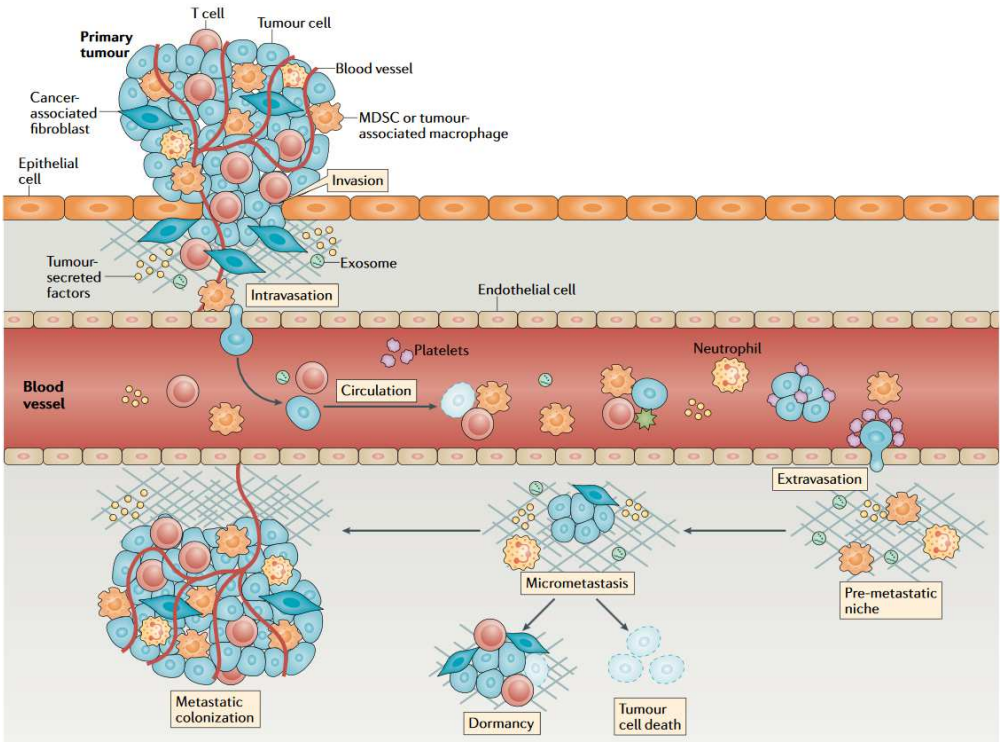

Figure 3. Invasion-metastasis cascade

Metastasis is a complex multistep process. Tumor cells detach from the primary tumor site, migrate and invade through the BM and the surrounding stroma, enter the blood or lymphatic vessels (intravasation), survive in the circulation, leave the blood or lymphatic vessels after identifying a premetastatic niche (extravasation), adhere and grow as micro-metastasis which at end may die or become dormant or form metastatic colonization. Figure from reference [66].

The initial steps of this cascade including cancer cell local invasion and migration are achieved by loss of cell-cell adhesion that allows cancer cells to detach from the primary tumor mass and changes in cell-matrix interaction that enable the cells to pass through the basement membrane and invade the surrounding stroma. These changes are strongly driven by a cellular program termed the epithelial-mesenchymal transition (EMT) [67-69]. Once cancer cells have invaded the surrounding stroma, it becomes reactive and promotes intravasation of these cells into blood or lymphatic vessels where they can survive in the blood or lymphatic circulation, and are known as circulating tumor cells (CTCs) [70]. Once CTCs reach organs which provide a suitable

stromal environment for secondary growth termed pre-metastatic niche, the cells may extravasate through vascular walls into the parenchyma of distant tissues [71]. Some of these cells that survive may form micrometastatic colonies in this parenchyma, and finally proliferation of these microscopic colonies may lead to macroscopic clinically detectable metastases [72] (Figure 3).

In metastatic CRC cases, the most common first site of metastasis is the liver, whereas the lung and bones are considered the second most common target organs [73], this may be because the venous drainage of the colon and upper rectum is via the portal vein; therefore those cancer cells dissociate from primary tumors within the colon usually reach the capillary network of the liver as a first station for colonization [74,75].

Notably, CRC metastasis is a multistep process by which cancer cells spread from primary tumors and form new tumor colonies at distant tissues. The initial and last steps of the metastatic cascade require the support of a process termed epithelial-mesenchymal transition (EMT) and its reverse program mesenchymal-epithelial transition (MET).

1.3.2 Epithelial-mesenchymal transition

EMT is a complex cellular and molecular program that plays an essential role in enhancing the migratory and invasive properties of cancer cells during metastasis and controlled by various families of transcriptional regulators through different signaling pathways [63,76]. EMT occurs not only during cancer progression but also is essential for embryonic development and many pathological processes such as wound healing and tissue fibrosis [77-79].

During EMT, The polarized epithelial cancer cells are subjected to loss of apical–basal polarity, dissolution of adhesion forces between cells and reorganization of cytoskeletal architecture, which enable these cells to acquire mesenchymal phenotype with spindle-like morphology, increased cellular motility, cancer stem cell activity, elevated

resistance to apoptosis, and ability to degrade and penetrate the basal extracellular matrix (ECM) by matrix metalloproteinases (MMPs) secretion [80-82]. Within CRC, These mesenchymal characteristics are more predominant in cancer cells at the invasive front of colorectal tumors, whereas those towards the center often exhibit a more epithelial-like phenotype [78,83].

Furthermore, EMT is characterized by a high level of plasticity and thus is a reversible process. During metastatic colonization, the mesenchymal-like cancer cells revert again into the epithelial phenotype to restore proliferative capacity and form macrometastasis by undergoing Mesenchymal-epithelial transition (MET) [84].

The initiation and progression of EMT is orchestrated by a network formed of multiple molecular signaling pathways and regulators [85,86]. The induction of EMT is mediated by many extracellular stimuli and corresponding intracellular signalling pathways including TGFβ-SMAD signalling, Wnt/b-Catenin, growth factor-receptor tyrosine kinase, and Notch pathway. These different signaling pathways converge on the activation of a group of transcription factors called EMT-inducing transcription factors (EMT-TFs), such as SNAI1, SNAI2 (Slug), zinc-finger E-box-binding homeobox1 (ZEB1) and ZEB2, and Twist-related protein 1 (TWIST1) [87,88] (Figure 4).

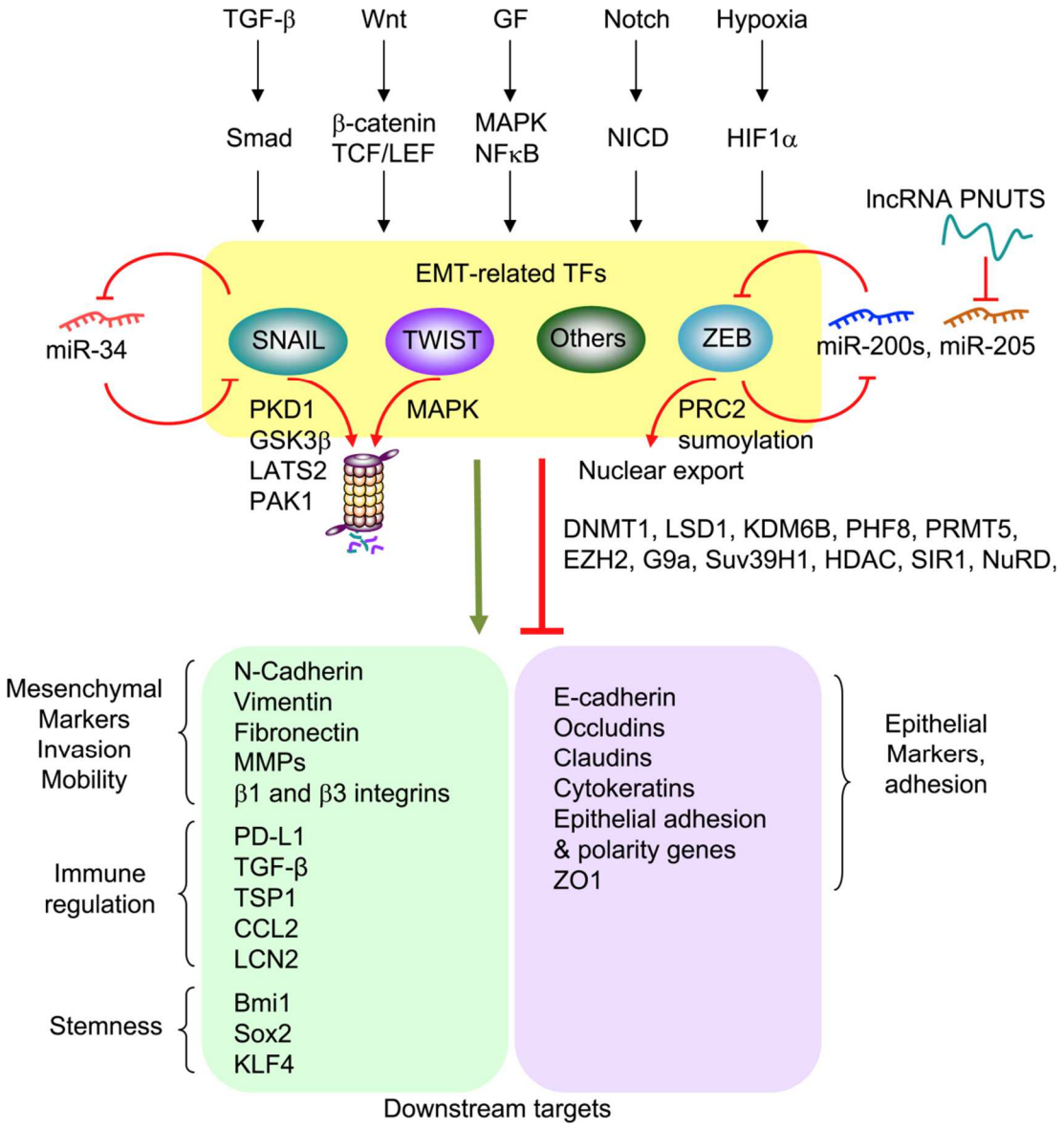

Figure 4: Regulatory molecular network of EMT

A number of signaling pathways activate EMT-related transcription factors such as the Snail, Twist, and Zeb families which can be further modified by epigenetic, post-transcriptional, and post-translational regulators such as microRNAs and other non-coding RNAs to organize the expression of epithelial and mesenchymal markers during EMT. Figure from reference [89].

The activated EMT-TFs drive alterations in the expression of EMT-associated genes which in turn regulate the shift to a mesenchymal state by suppressing expression of epithelial markers, mainly E-cadherin, which is a key protein for epithelial cell-cell adhesion. It also induces expression of other markers associated with the mesenchymal state such as, vimentin, fibronectin, and N-cadherin [90,91]. The loss of E-cadherin expression is considered a crucial event in EMT and can be mediated by its

transcriptional repression through the binding of EMT-TFs such as SNAI1, SNAI2, ZEB1, and ZEB2 to E-box sequence in the promoter region of *CDH1* [88,92] (Figure 4).

In addition, the expression and functions of EMT-TFs are controlled at multiple levels by different regulatory mechanisms, including epigenetic modifications, post-transcriptional modifications by miRNAs, translational control, and post-translational modifications [89,93,94] (Figure 4). All theses mechanisms together regulate the activation of EMT.

In summary, EMT and its reverse program MET play crucial roles in the metastatic progression of CRC. When CRC metastasizes in patients, the prognosis of these patients become worse, and according to the extent of metastasis, CRC stage and treatment are determined.

1.4 Prognostic and predictive determinants in colorectal cancer

1.4.1 Pathological factors

1.4.1.1 Tumor stage

Pathologic stage of tumor after surgical resection is by far the most important prognostic predictor of postoperative outcome for CRC patients [95]. However, there is sometimes considerable stage-independent variability in clinical outcome. The most common staging system for CRC is the TNM system, of the combined American Joint Committee on Cancer (AJCC)/Union for International Cancer Control (UICC) [96,97].This staging system is based on three parameters. "T" describes the depth of tumor invasion in the colorectal wall, "N" the extent of lymph node involvement, and "M" the extent of metastasis [97] (Table 1, Figure 5).

Most CRCs present as localized tumor at time of initial diagnosis, involving the colorectal wall (stages I and II) and regional lymph nodes (stage III), and these tumors are amenable to potentially curative surgical resection [98]. While stage I colorectal cancers are mostly cured by surgery alone, stage II and III disease more frequently recurs [99], and especially for stage III adjuvant chemotherapy is an additional treatment option after surgery and can reduce disease recurrence by up to 30 % [100]. Yet, for node negative stage II colorectal cancer the use of adjuvant chemotherapy remains controversial, since only patients with high-risk stage II disease may benefit from this treatment, while others are unnecessarily harmed by toxic side effects of chemotherapy [101,102]. Therefore, additional stage independent risk stratification is needed to identify high-risk stage II CRC patients who should receive adjuvant therapy. In this context, indicators of tumor aggressiveness, such as histologic tumor grade, vascular invasion, perforation, and DNA MMR status may guide the decision for adjuvant therapy [103-105].

Table 1. CRC TNM staging AJCC/ UICC 8th edition [96,97]

Primary tumor	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ, intramucosal carcinoma
T1	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades through muscularis propria into pericolorectal tissues
T4	Tumor invades visceral peritoneum or invades or adheres to adjacent organ or structure
T4a	Tumor invades through visceral peritoneum
T4b	Tumor directly invades or adheres to adjacent organs or structures
Nodal status (N)	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	1-3 regional lymph nodes are positive (tumor in lymph nodes ≥0.2 mm), or any number of tumor deposits are present and all identifiable lymph nodes are negative
N1a	1 regional lymph node is positive
N1b	2 or 3 regional lymph nodes are positive
N1c	No regional lymph nodes are positive, but there are tumor deposits in the: ▪ Subserosa ▪ Mesentery ▪ Nonperitonealized pericolic, or perirectal/mesorectal
N2	4 or more regional nodes are positive
N2a	4-6 regional lymph nodes are positive
N2b	7 or more regional lymph nodes are positive
Distant metastases(M)	
M0	No distant metastases
M1	Metastasis to 1 or more distant sites or organs or peritoneal metastasis is identified
M1a	Metastasis to one site or organ is identified without peritoneal metastasis
M1b	Metastasis to two or more sites or organs is identified without peritoneal metastasis
M1c	Metastasis to the peritoneal surface is identified alone or with other site or organ metastases
Stage	**TNM Classification**
0	Tis, N0, M0
I	T1,T2, N0, M0
IIA	T3, N0, M0
IIB	T4a, N0, M0
IIC	T4b, N0, M0
IIIA	T1-T2, N1/N1c, M0
	T1, N2a, M0
IIIB	T3-T4a, N1/N1c, M0
	T2-T3, N2a, M0
	T1-T2, N2a, M0
IIIC	T4a, N2a, M0
	T3-T4a, N2b, M0
	T4b, N1-N2, M0
IVA	T (any), N (any), M1a
IVB	T (any), N (any), M1b
IVC	T (any), N (any), M1c

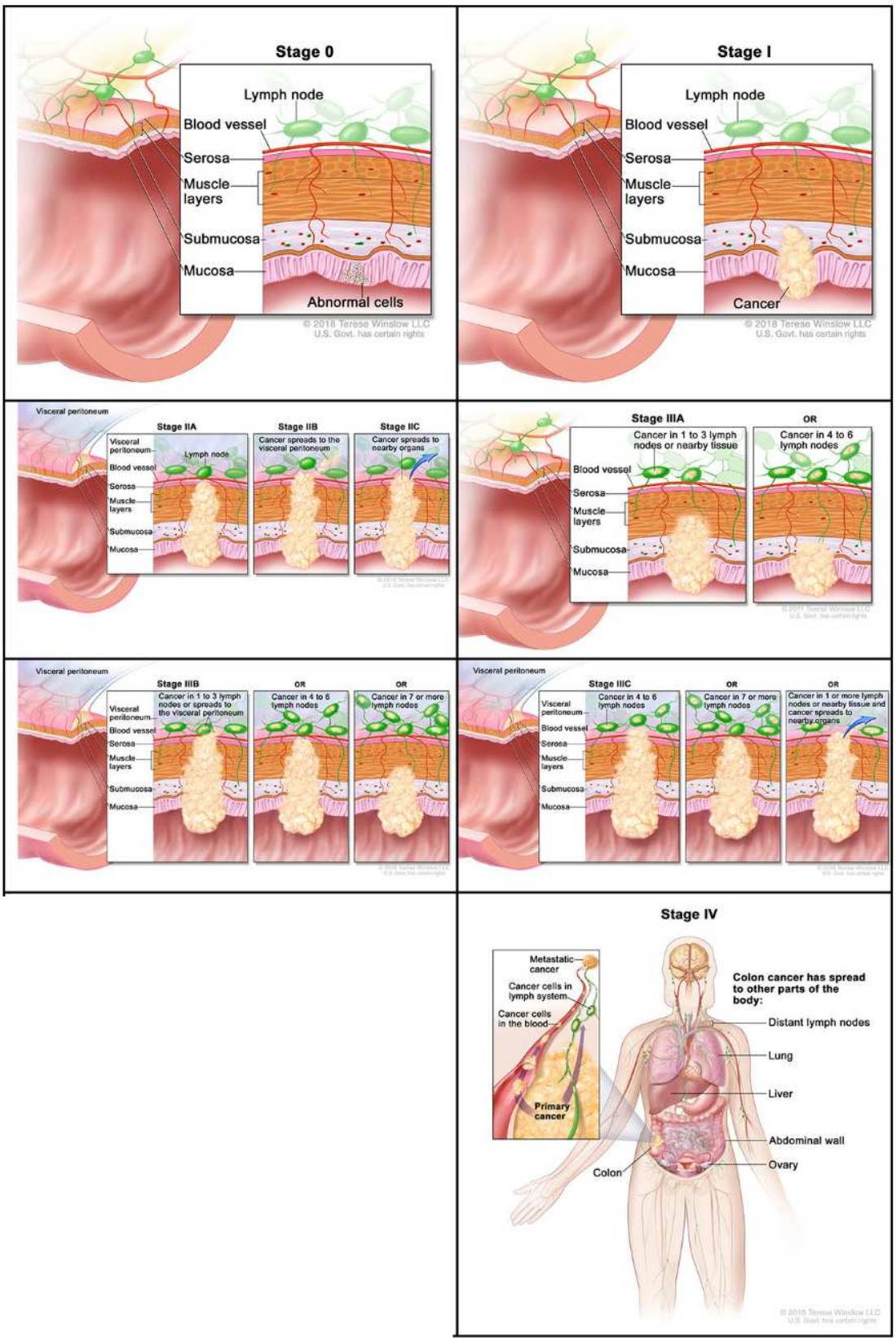

Figure 5: CRC TNM staging AJCC/ UICC, 8th edition

The diagram illustrates the depth of tumor invasion in the bowel wall, the extent of lymph node involvement, and the extent of metastasis in CRC stages from 0 to IV. For the National Cancer Institute © 2018 Terese Winslow LLC, U.S. Govt. has certain rights.

1.4.1.2 Tumor grade

Tumor grade is a traditional prognostic parameter in CRC, which can be assessed on hematoxylin and eosin (H&E) stained slides [106]. Tumor grading describes the degree of tumor cell differentiation and the resemblance of the tumor to the tissue of origin. Grading of CRC can only be applied on adenocarcinomas, not otherwise specified (NOS) which represent 90 % of all CRCs. Moreover, grading should be based upon the least differentiated component of carcinomas which are sometimes heterogeneous and should not include the leading front of tumor invasion [107].

According to the current WHO classification, tumor grading, depending on the extent of glandular appearance, classifies tumors into low grade, showing 50 % or more gland formation (formerly well- to moderately differentiated) and high grade, showing less than 50 % gland formation (formerly poorly differentiated) [107]. Notably, combining well and moderately differentiated into low grade reduces interobserver variation in the interpretation of tumor grade and improves its prognostic significance [103]. Tumor grading has a prognostic value in CRC, since the loss of differentiation during tumor progression has repeatedly been associated with tumor aggressiveness, thereby indicating poor patients' outcome. However, it is not always true because some CRC tumors may show high grade morphology but behave as low grade tumors because of their DNA mismatch repair status/MSI status [108].

1.4.1.3 Lymphovascular invasion (LVI) and perineural invasion (PNI)

Tumor invasion into veins or lymphatics or the space surrounding nerves is an important prognostic determinant for CRC [109,110]. Moreover, lymphovascular and perineural invasion are included in the definition of "high-risk" stage II colon cancer [111,112], and usually are associated with poor prognosis [113,114]

1.4.1.4 Other pathological features

There are other factors that could affect CRC prognosis, including status of the surgical resection margins, tumor border configuration, host immune response, Peritumoral fibrosis and tumor location [115-119].

1.4.2 Clinical factors

Poor prognostic clinical factors at diagnosis include bowel obstruction or perforation and high preoperative carcinoembryonic antigen (CEA) level [120].

1.4.3 Molecular factors

1.4.3.1 Microsatellite instability

MSI reflects deficiency of mismatch repair enzymes, and it has been demonstrated to be an independent positive prognostic factor after curative resection of CRC [121,122]. In Patients with localized CRCs (stages I and II), tumors that are MMR-deficient/MSI-H are associated with longer survival, compared with MMR-proficient /MSI-L or MSS tumors [123,124]. Additionally, MSI in combination with BRAF status seems to be predictive of a lack of response to chemotherapy [125]. Furthermore, it has been suggested as a predictive marker for response to immunotherapy such as anti-programmed-cell-death protein 1 (PD-1) therapy [126]

1.4.3.2 *RAS* and *BRAF* mutations

RAS and *BRAF* mutational status has a significant prognostic and predictive value [127], and has become part of routine pathological evaluation for CRC greater than stage I, since it may alter treatment strategy [128,129]. *RAS* mutations, including *KRAS* and *NRAS* codons 12 and 13 of exon 2; 59 and 61 of exon 3; and 117 and 146 of exon 4, are used to identify CRC patients who will not benefit from anti-epidermal growth factor receptor (EGFR) treatment [130]. *BRAF* mutations, in particular the V600E mutation, are

demonstrated to be associated with worse survival in CRC patients with microsatellite stable (MSS) tumors compared with those with microsatellite instable (MSI) tumors [131].

1.4.3.3 Other molecular markers

The prognostic value of other potentially clinically applicable molecular markers has been studied in CRC. However, many individual markers linked to CRC outcome were not selected due to best performance in outcome prediction but based on certain tumor cell characteristics or phenotypic traits, such as markers for putative cancer stem cells and budding colon cancer cells, respectively [132,133].

1.4.3.4 Consensus Molecular Subtypes

In addition to predictive biomarkers for CRC risk stratification, owing to recent advances in understanding the molecular characterization of CRC, a number of gene expression based classifications has been suggested and may improve clinical risk stratification [134,135]. Recently, classification of CRCs into four consensus molecular subtypes (CMSs) with distinguishing molecular and clinical characteristics has been proposed. In this classification, CRCs with subtype CMS1 involve MSI-H tumors, and also tumors with a CIMP and mutations in the BRAF oncogene, whereas CSM2 CRCs have high CIN as well as activation of the Wnt signalling. Additionally, CRCs with subtype CMS3 include tumors with KRAS mutations and show disruption of metabolic pathways, and CSM4 CRCs have a mesenchymal phenotype and frequently, CIMP phenotype[134]. Moreover, it was demonstrated that CMS1 tumors are associated with a good prognosis, whereas the CMS4 tumors are associated with a poor prognosis, and the CMS2 and CMS3 tumors have an intermediate prognosis [134,136,137] .

Finally, Molecular classification is not yet ready for incorporation into available staging systems or prognostic stratifications, and it is difficult to be reproducible since it requires analysis of a large number of genes. Therefore further robust and potential prognostic and

predictive genetic markers are needed for CRC risk stratification; hence, improving CRC patients' outcome and clinical decision making regarding therapeutic strategies.

2 AIMS OF THE STUDY

The present study had the following aims:

- Identify robust prognostic biomarkers for patient outcome in colorectal cancer through exploitation of the publicly available TCGA dataset.

- Explore the distribution and expression levels of RBP7 and Annexin A9 on the protein level using tissue collections of colorectal cancer patients with long survival follow-up data.

- Evaluate the prognostic power of RBP7 and Annexin A9 for patients with colorectal cancer.

- Investigate the functional relevance of RBP7 in colorectal cancer using biological assays.

3 MATERIALS

3.1 Chemicals and reagents

Table 2. Chemicals and reagents used in this study

Chemical/Reagent	Manufacturer
4x Laemmli Sample Buffer	Bio-Rad, Munich, Germany
Acetic acid 100 %	Carl Roth GmbH, Karlsruhe, Germany
Agarose Biozym LE	Biozym Scientific, Hessisch Oldenforf, Germany
All-purpose Hi-Lo DNA Marker	Bionexus Inc., Oakland, CA, USA
Ammonium peroxodisulfate	Carl Roth GmbH, Karlsruhe, Germany
Ampicillin sodium salt	Sigma-Aldrich, St. Louis, MO, USA
Bovine serum albumin 25 % (BSA)	Thermo Fisher Scientific Inc., Waltham, MA, USA
BSA (Albumin Faktor V)	Carl Roth GmbH, Karlsruhe, Germany
β-Mercaptoethanol	Bio-Rad, Munich, Germany
Crystal violet	Carl Roth GmbH, Karlsruhe, Germany
Dimethylsulfoxide	Carl Roth GmbH, Karlsruhe, Germany
DMEM	Biochrom, Berlin, Germany
Deoxycholic acid sodium salt	Carl Roth GmbH, Karlsruhe, Germany
Ethylenediaminetetraacetic acid	Sigma-Aldrich, St. Louis, MO, USA
Ethidium bromide solution 1 %	Carl Roth GmbH, Karlsruhe, Germany
Fetal Bovine Serum	Biochrom, Berlin, Germany
Fugene 6	Promega GmbH, Mannheim, Germany
Glycine	Carl Roth GmbH, Karlsruhe, Germany
Immobilon-P PVDF Membrane	Merck Millipore, Billerica, MA, USA
Immobilon Western Chemiluminescent HRP Substrate	Merck Millipore, Billerica, MA, USA
LB Broth (Luria/Miller)	Carl Roth GmbH, Karlsruhe, Germany
Matrigel	Corning Life Sciences, Tewksbury, MA, USA

Chemical/Reagent	Manufacturer
Methanol	Carl Roth GmbH, Karlsruhe, Germany
Opti-MEM	Thermo Fisher Scientific Inc., Waltham, MA, USA
Penicillin/Streptomycin	Biochrom, Berlin, Germany
SDS ultra pure	Carl Roth GmbH, Karlsruhe, Germany
Skim milk powder	Sigma-Aldrich, St. Louis, MO, USA
Sodium chloride	Carl Roth GmbH, Karlsruhe, Germany
TEMED	Carl Roth GmbH, Karlsruhe, Germany
TRIS	Carl Roth GmbH, Karlsruhe, Germany
TWEEN 20	Sigma-Aldrich, St. Louis, MO, USA

3.2 Enzymes

Table 3. Enzymes used in this study

Enzymes	Manufacturer
Restriction endonucleases	Thermo Fisher Scientific Inc., Waltham, MA, USA
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific Inc., Waltham, MA, USA
T4 DNA Ligase	Thermo Fisher Scientific Inc., Waltham, MA, USA

3.3 Kits

Table 4. Kits used in this study

Kits	Manufacturer
DC Protein Assay	Bio-Rad, Munich, Germany
ImmPRESS HRP Anti-Rabbit IgG (Peroxidase) Polymer Detection Kit	Vector Laboratories Inc., Burlingame, CA, USA
mi-Plasmid Miniprep Kit	Metabion International AG, Planegg, Germany
PureYield Plasmid Midiprep System	Promega GmbH, Mannheim, Germany
QIAamp DNA Micro Kit	Qiagen GmbH, Hilden, Germany

Kits	Manufacturer
Rapid DNA Ligation Kit	Thermo Fisher Scientific Inc., Waltham, MA, USA
ThinCert cell culture inserts	Greiner Bio-One, Kremsmünster, Austria
UltraView Universal DAB Detection Kit	Ventana Medical Systems, Inc., Tucson, AZ, USA
Wizard SV Gel and PCR Clean-Up System	Promega GmbH, Mannheim, Germany

3.4 Bacterial strain and vectors

Table 5. Bacterial strain and vectors used in this study

Strain	Source
Escherichia coli DH5α	Invitrogen GmbH, Karlsruhe, Germany
Vector name	Source
pcDNA3.1	Invitrogen GmbH, Karlsruhe, Germany
RBP7 gene fragment	Integrated DNA Technologies, Inc., Coralville, Iowa, USA

3.5 Antibodies

Table 6. Antibodies used in this study

Antibody	Source/Clone	Manufacturer/Catalog #	Dilution used	
			IHC	WB
Primary antibodies				
ANXA9	Mouse/F9	Santa Cruz- Biotechnology/sc-374288	1:00	-
RBP7	Rabbit	Sigma-Aldrich/HPA034749	1:100	1:250
Tubulin	Mouse/DM1A	Sigma-Aldrich/T6199	-	1:50000
Secondary antibodies				
Anti-Mouse HRP	Goat	Promega GmbH/W4021	-	1:30000
Anti-Rabbit HRP	Goat	Sigma-Aldrich/12-342	-	1:5000
Immunohistochemistry (IHC), Western Blot (WB)				

3.6 Buffers

Table 7. Buffers used in this study

Buffer	Components
4x Lower gel buffer	0.4 % SDS, 1.5 M TRIS, pH 8.8
4x Upper gel buffer	0.4 % SDS, 500 mM TRIS, pH 6.8
10x Running buffer	1.92 M Glycine, 1 % SDS, 250 TRIS, pH 8.5
10x TBS buffer	150 mM NaCl, 20 mM TRIS, pH 7.6
1x TBST buffer	10x TBS buffer, 0.1 % Tween 20
10x Transfer buffer	1.92 M Glycine, 20 % Methanol, 1 % SDS, 250 mM TRIS, pH 8.5
50x TAE buffer	20 mM Acetic acid, 1 mM EDTA, 40 mM TRIS, pH 8.0
RIPA buffer	1 % NP 40, 150 mM NaCl, 0.1 % SDS, 0.5 % Deoxycholic acid sodium salt, 50 mM TRIS hydrochloride, pH 8.0

3.7 Laboratory equipment

Table 8. Equipment used in this study

Equipment	Manufacturer
BenchMark XT	Ventana Medical Systems, Inc., Tucson, AZ, USA
Centrifuge 5415R	Eppendorf AG, Hamburg, Germany
Heracell 240i CO_2 Incubator	Thermo Fisher Scientific Inc., Waltham, MA, USA
Heraeus Megafuge 40R Centrifuge	Thermo Fisher Scientific Inc., Waltham, MA, USA
Herasafe KSP Class II Biological Safety Cabinet	Thermo Fisher Scientific Inc., Waltham, MA, USA
HTU SONI-130G. Heinemann Ultraschall- und labortechnikMini-PROTEAN Tetra Cell	Bio-Rad, Munich, Germany
NanoDrop 1000 Spectrophotometer	Thermo Fisher Scientific Inc., Waltham, MA, USA

Equipment	Manufacturer
Odyssey Fc Imaging system	LI-COR Biosciences, Bad Homburg, Germany
PerfectBlue 'Semi-Dry'-Blotter, Sedec	Peqlab Biotechnologie GmbH, Erlangen, Germany
T100 Thermal Cycler	Bio-Rad, Munich, Germany
Theromixer comfort	Eppendorf AG, Hamburg, Germany
Varioskan Flash Multimode Reader	Thermo Fisher Scientific Inc., Waltham, MA, USA

3.8 Software and databases

Table 9. Software and database used in this study

Software/Datebase	Reference
cBioPortal	https://www.cbioportal.org
Endnote X9	Clarivite analytics, https://endnote.com
GraphPad Prism version 5.0	GraphPad Software, www.graphpad.com
GSEA	Broad Institute, http://software.broadinstitute.org/gsea/index.jsp
Microsoft Office Excel 2010	Microsoft Corporation
Morpheus	Broad Institute, https://software.broadinstitute.org/morpheus
PANTHER version 10.0	www.pantherdb.org
Qupath	https://qupath.github.io
SPSS version 25.0	IBM, https://www.ibm.com
TCGA	https://cancergenome.nih.gov

4 METHODS

Parts of the methods presented in this section are from Elmasry et al., 2019 [1].

4.1 Clinical samples collection

All tumor samples used in this work were from CRC patients who underwent curative surgical resection between 1994 and 2007 at the hospital of the Ludwig-Maximilians-Universität München (LMU), had long survival follow-up data and had no history of receiving adjuvant therapies. Follow-up data of these patients were registered by the Munich Cancer Registry, all information regarding their personal identification was removed from samples and follow-up data, and the need for consent was waived by the institutional ethics committee of the Medical Faculty of the LMU. All Samples were collected from the archives of the Institute of Pathology of the Ludwig-Maximilians-Universität München (LMU) in the form of Formalin-fixed, paraffin-embedded (FFPE) tissue blocks. From these blocks Tissue microarrays (TMAs) with representative 1 mm cores were constructed, including tumor edges and tumor centers of each case. Furthermore, information about tumor stage and grade of tumor differentiation in each case was reviewed.

For evaluation of RBP7 expression, a collection of tumor samples from 219 cases with localized colorectal adenocarcinomas and without lymph node involvement or distant metastasis at the time of diagnosis (UICC stage I and II [96,97]) was used. During the follow-up period of these patients, 42 cases (19 %) have died from CRC.

Furthermore, for validating ANXA9 for survival and metastasis prediction, survival and metastatic collections of colon cancer patients were analysed. The survival collection constituted of 244 tumor samples from cases with UICC stage II CRC at the time of diagnosis. The median follow-up period for these cases was 4.9 years. Additionally, the metastatic collection had a case- control design and comprised of tumor samples

from 90 patients. No tumor tissue from these patients was used in the survival collection. 50 % of the patients used in the metastatic collection had colon cancer with synchronous liver metastases (UICC stage IV), detected via clinical imaging or liver biopsy. Moreover, this collection included a group of patients with colon cancer, no distant metastasis at the time of diagnosis (UICC stages I–III) and at least 5 years disease-free survival after primary surgical resection as controls. Criteria used for matching cases and controls were right sided location of tumor, T-category and tumor grade (according to WHO 2010), giving rise to 46 matched pairs.

4.2 Immunohistochemistry

4.2.1 Immunohisochemical staining

5 µm sections were cut from constructed colon cancer TMAs, deparaffinized with xylene and rehydrated with graded ethanol. Then, staining was carried out on a BenchMark XT autostainer (Ventana Medical Systems) by using primary antibodies which were listed with their dilutions in section 3.5 (Table 6). Subsequently, staining was visualized by ultraView DAB detection kits (Ventana Medical Systems).

4.2.2 Immunohistochemical scoring

For RBP7 staining, the expression intensities were categorized at first by semi-quantitative scoring into barely detectable, weak, moderate or strong expression, based on the extent of positive staining, followed by quantitative scoring, slides were scanned using a Panoramic Desk digital slide scanner (3D Histech), and analyzed using the QuPath digital image analysis software [138]. All scanned images of immunohistochemically stained TMA sections were imported into QuPath to be dearrayed, and computational color deconvolution was applied to separate haematoxylin and 3,3'-diaminobenzidine (DAB) stains, as previously described [139]. An automated detection algorithm was used in QuPath to differentiate tumor and non-

tumor cells. After calibration of RBP7 immunopositivity thresholds, H-scores were calculated based on the extent and intensity of RBP7 nuclear staining by adding 3x % of strongly stained tumor cells, 2x % of moderately stained tumor cells, and 1x % of weakly stained tumor cells [138]. For regional differences, H-scores of RBP7 staining were separately determined in tumor cells at the tumor stroma interface (tumor edge), and 100 µm or more away from the tumor stroma interface (tumor center), for each case. All analyses were conducted in a blinded fashion from clinical outcome.

For ANXA9 staining, semi-quantitative scoring was used to assess the degree of expression and a numerical score was given: score 0 (no tumor cells stained), score 1 (<10 % tumor cells stained), score 2 (10 %-50 % tumor cells stained), or score 3 (> 50 % tumor cells stained).

4.3 TCGA colon cancer data analysis and GSEA

The colon cancer related mRNA expression profile and corresponding clinical and mutational data of 457 cases were downloaded from the TCGA database and cBioPortal (https://www.cbioportal.org/). To explore the functional relevance of RBP7, a ranked gene list was generated by calculating Pearson correlations of *RBP7* expression and the expression of 20,531 genes within the TCGA dataset. Then, correlations between this ranked gene list and curated gene sets from the Molecular Signatures Database v5.0 [140] were searched for by gene set enrichment analysis (GSEA) [141]. The default parameters of GSEA using gene lists of 15 to 500 genes were applied, and analyses were run with 1,000 permutations. Heat maps and clustering for *RBP7* mRNA expression and individual EMT regulators were generated by using Morpheus software (Broad Institute, https://software.broadinstitute.org/morpheus/). For *RAS* and *BRAF* status, activating mutations in codons 12, 13, 61, 117, and 146 of *KRAS* and *NRAS*, and in codon 600 of *BRAF* were considered, respectively. Furthermore, the 194 putative genes for survival and metastasis prediction were

characterized according to the classes of proteins they encode by using PANTHER version 10.0 (www.pantherdb.org).

4.4 RBP7 cloning in vector pcDNA3.1

A synthetic sequence containing *RBP7* mRNA transcript variant 1 (accession number NM_052960.2) was created via Integrated DNA Technologies. The mammalian expression vector, pcDNA3.1 (Invitrogen) was used as an insert plasmid target. pcDNA3.1 was digested using Eco32I restriction enzyme, dephosphorylated, then loaded on a 1 % agarose gel to be separated by electrophoresis at 120 V and extracted at the end from gel using WizardSV Gel and PCR Clean-up System. Next, *RBP7* synthetic sequence was inserted between BamH1 and Xba1 sites of pcDNA3.1 vector using T4 DNA ligase. Successful cloning was finally verified by restriction analysis and Sanger sequencing (GATC Biotech).

4.5 Bacterial cell culture

For cloning the *RBP7* expression vector and replicating plasmids with an ampicillin resistance, the *Escherichia coli* DH5α strain was used. The bacterial cells were grown overnight at 37 °C in LB-medium containing 100 µg/ml ampicillin for the selection of antibiotic-resistant cells. For transforming bacterial cells, plasmid DNA and competent *E. coli* DH5α were mixed and incubated on ice for 30 minutes and then a heat shock in a 42 °C water bath for 45 seconds was performed, followed by cooling the bacteria on ice for 2 minutes and incubation in 500 µL LB medium without antibiotics at 37 °C for 45 minutes. Afterwards, bacteria were cultured on LB agar plates containing ampicillin and let to grow at 37 °C overnight. For multiplying the transformed bacteria, a single bacterial colony was picked, inserted in LB medium with added ampicillin and let to grow at 37 °C on a shaker for 12 hours. Finally, Plasmid DNA was extracted and purified from bacteria by using the mi-Plasmid Miniprep Kit (Metabion) or the PureYield Plasmid Midiprep System (Promega) according to the manufacturer's instructions.

4.6 Cancer cell culture

4.6.1 Preparation of human colon cancer cell lines

HCT116 and SW1222 human colon cancer cell lines were purchased from the American Type Culture Collection. Cells were grown in DMEM media supplemented with 10 % FBS, 100 U/ml penicillin, and 0.1 mg/ml streptomycin (Biochrom) and kept in an incubator at 37 °C and 5 % CO2. Cryopreservation of culture cells was performed by slowly cooling cells to -80 °C using 90 % FCS and 10 % DMSO (Sigma) and then cells were transferred into liquid nitrogen for long term preservation.

4.6.2 Transient transfection of cells by plasmids

For transient RBP7 overexpression, HCT116 and SW1222 colon cancer cells were seeded in 35-mm dishes. 24 hours after seeding, 36 µl FuGENE 6 (Promega) was added to Opti-MEM (Thermo Fisher) and incubated to 5 minutes, followed by adding 1 ug of constructed plasmid pcDNA3.1-RBP7 or, as control, empty plasmid pcDNA3.1 (Invitrogen) and incubated for another 15 minutes. Subsequently, the transfection mixture was added to the plated cells and then they were incubated. 48 hours after transfection, transfected cells were harvested for further analysis.

4.7 Western blot analysis

Transfected HCT116 and SW1222 colon cancer cells were lysed in ice cold RIPA buffer containing protease and phosphatase inhibitors (Roche), followed by sonication twice for 10 seconds (HTU SONI-130). Next, samples were centrifuged at 13000 rPM for 30 min at 4°C and supernatants were transferred to new tubes. The Protein concentration of the lysates was determined by DC protein assay kits (Bio-Rad) as stated in the manufacturer's protocol and measured by using a Varioskan Flash Multimode Reader with SkanIt Software 2.4.4 (Thermo Fisher Scientific). For preparing samples for electrophoresis, 30 µg of the diluted protein samples with 4x Laemmli

sample buffer (Biorad) and added 10 % β-Mercaptoethanol (Biorad) were denaturated at 95 C for 5 minutes. Next, samples were loaded on a 10 % SDS-PAGE gel and separated by molecular weight after running the gel in Mini-PROTEAN Tetra Cell (Biorad) filled with TRIS/glycine/SDS running buffer. The gel was run at 90 volt for 20 minutes followed by 120 volt for 60 minutes. After electrophoresis, the separated proteins are blotted onto PVDF membranes, Immobilon-P (Millipore) soaked up in transfer buffer using the PerfectBlue 'Semi-Dry'-Blotter, Sedec (Peqlab). Electrotransfer was performed at 100 mA and 10 volt for 20 minutes. The membranes were then placed in 5 % blocking solution (5 % non fat dried milk diluted in TBST) for 1 hour on a shaker then washed once in TBST and incubated with primary antibodies in 5% bovine serum albumin (Carl Roth) diluted in TBST overnight in 4°C on a shaker. Next, the membranes were washed with TBST for 10 minutes on a shaker, repeated 3 times. Afterwards, the membranes were incubated in secondary antibody in 5% non fat dried milk diluted in TBST for 1 hour then washed with TBST for 10 minutes on a shaker, repeated 3 times. Finally, the membranes were incubated with a chemiluminescent HRP substrate (Millipore) for 1-2 minutes then protein bands were detected using the Odyssey Fc imaging system (Li-COR). Primary and secondary antibodies used are listed in section 3.5 (Table 6).

4.8 Cell migration and invasion assays

To analyze the effect of RBP7 overexpression on cell migration and invasion, a cell suspension containing $1x10^5$ transfected HCT116 and SW1222 colon cancer cells/ml in serum free medium was prepared, and then 250 µl of this cell suspension solution was added to the inside of each upper chamber of ThinCert cell culture inserts with 8 µm pore size (Greiner Bio-One). To carry out cell invasion assays, 100 µl of 1 mg/ml growth factor depleted Matrigel (Corning) were added to the inserts and solidified in a 37°C incubator for 30 minutes before adding the transfected colon cancer cells. Next, 500 µl serum-free medium was added to the lower chamber of the inserts, and replaced

after 24 hours by DMEM containing 10 % FBS. After incubation in a cell culture incubator for 48 hours for migration and 96 hours for invasion, inserts were removed; cells were fixed by adding 4 % paraformaldehyde then Methanol into both sides of inserts, and then stained by 0.1 % crystal violet. Cells that have not migrated or invaded were removed carefully from the upper surface of the filters by cotton swabs and photomicrographs of migrated or invaded cells were taken. To quantify cell migration and invasion rates, the culture inserts were incubated in 250 µl of 30 % acetic acid for 30 seconds on a shaker, the stained migrated or invaded cells will be lysed by the acetic acid and the Crystal Violet will be liberated. Finally, the optical density of the 30% acetic acid was read using a Varioskan Flash Multimode Reader with SkanIt Software 2.4.4 (Thermo Fisher Scientific).

4.9 Statistical analysis

Receiver operated characteristic (ROC) curve analysis was used for binary classification of CRC cases. Moreover, the optimal cutoff values for RBP7 protein and mRNA expressions were determined by Youden's index, whereas for mRNA reads of *TCHH*, *ANXA9*, *HOTAIR*, *CCT6B* and *MCU* were determined manually. Cancer-specific survival involved the time period from the date of tumor resection to the date of death from colon cancer, deaths of a cause other than colon cancer were censored. The cancer specific survival was analyzed using the Kaplan-Meier method, and the log-rank test was used to compare the differences between groups. The univariate analysis and multivariate analysis were carried out using Cox proportional hazards model, and the hazard ratio (HR) and the associated 95 % confidence interval were calculated. Cases with missing data were excluded from respective analyses. t-tests were used to compare the differences between some groups. z-scores for survival and metastasis were generated using Microsoft Office Excel 2010. SPSS v.25.0 (IBM) and GraphPad Prism v.5.00 were used. For all analyses, a p value < 0.05 was considered statistically significant.

5 RESULTS

5.1 Identification of RBP7 as a prognostic biomarker in colon cancer

The results presented in this section are part of Elmasry et al., 2019 [1].

5.1.1 RBP7 is expressed in colon cancer cell subpopulations

To learn about the distribution and expression of RBP7 in colon cancer, we examined tissue specimens of a collection of 219 colon adenocarcinomas. RBP7 protein was located in the tumor cell nuclei of colon cancers. The number of RBP7 positive tumor cells and expression intensities varied greatly, ranging from barely detectably in few, to strong expression in most tumor cells (Figure 6A). Interestingly, within individual cancers RBP7 expression was not evenly distributed but instead labelled tumor cell subsets, which was most apparent in cases with weak to moderate expression (Figure 6A). Next, in order to assess RBP7 expression objectively, we applied a digital quantitative scoring approach to determine H-scores [138] that integrated the frequency (range 0 %-100 %) and staining intensity (range 0-3) of RBP7 positive tumor cells for each case (Figure 6B). In line with our initial semiquantitative analysis, H-scores ranged widely among different colon cancers, with a minimum of 0 and a maximum of 184.27 in our case collection (Figure 6B-C). We then analyzed different regions within each tumor, and observed that tumor cells close to the tumor edge showed significantly higher RBP7 expression scores when compared to tumor cells that were located in the tumor center (Figure 6D)

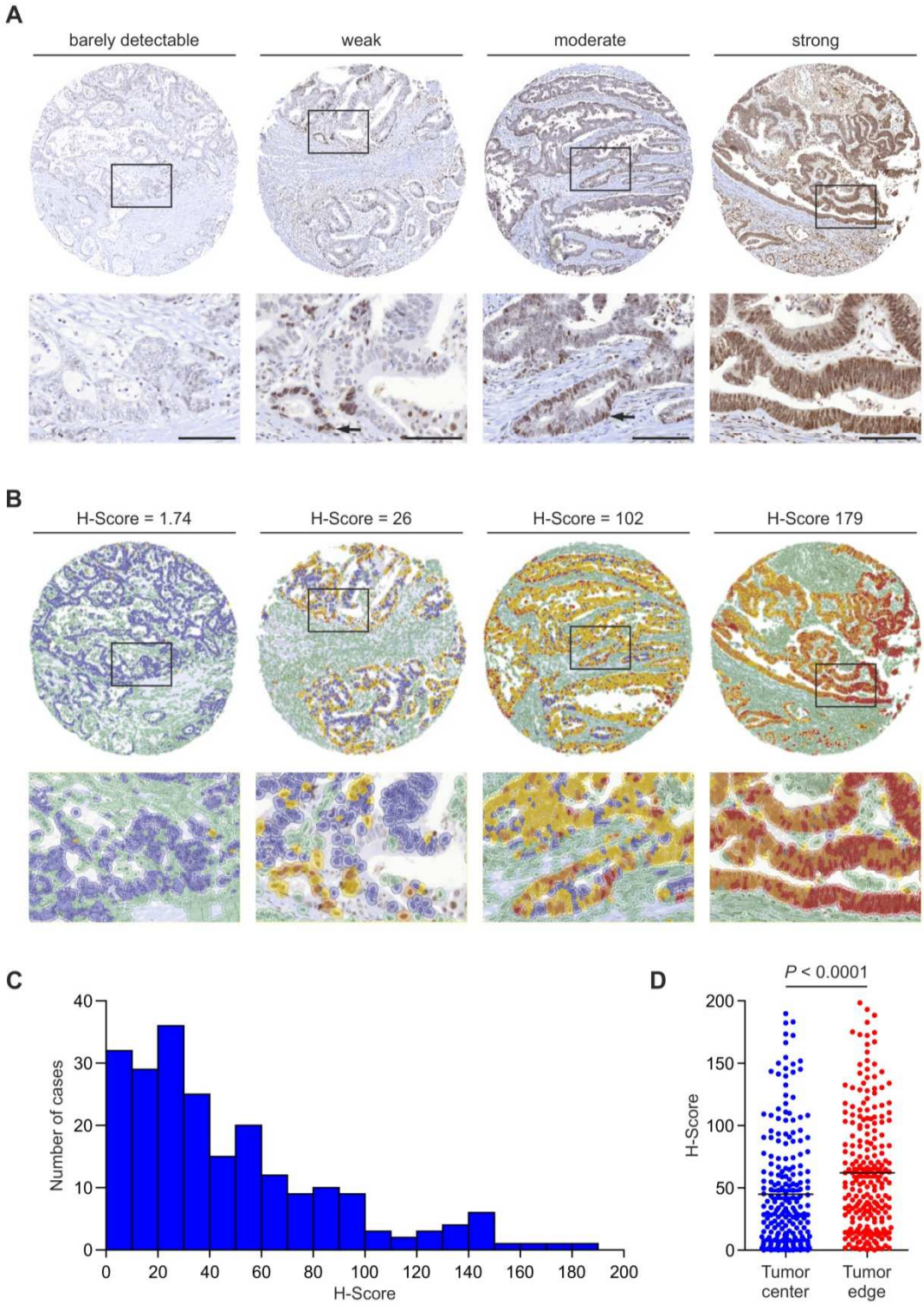

Figure 6. RBP7 protein expression and distribution in colon cancer.

(A) Detection of RBP7 by immunostaining in primary human colon cancers. Tumors were assigned semi-quantitative categories from barely detectable to strong expression of RBP7. Arrows indicate positively stained tumor cells in cases with weak or moderate expression. Lower panel images are magnifications of areas boxed in upper panel images. Scale bars, 100 µm. **(B)** Representative images showing digital quantitative scoring of RBP7 protein expression on the same cases as in (A). Detected cells were color-coded according to their classification. Green, non-tumor cells. Blue, negative tumor cells. Yellow, weakly stained tumor cells. Orange, moderately stained tumor cells. Red, strongly stained tumor cells. H-scores are indicated. Lower panel images are magnifications of areas boxed in upper panel images. **(C)** Histogram showing the distribution of H-score values in n = 219 colon cancer cases. **(D)** Distribution of H-scores, when separately measured in tumor cells at the tumor stroma interface (tumor edge), and 100 µm or more away from the tumor stroma interface (tumor center). Horizontal bars indicate mean and *P*-value is t-test result. Figure from Elmasry et al., 2019 [1].

These data indicated that RBP7 is expressed in tumor cell nuclei of most colon cancers, increases in expression towards the tumor edge, and can be quantitatively assessed in tumor tissue specimens.

5.1.2 High RBP7 expression indicates poor outcome in patients with early stage colon cancer

In order to determine the clinical significance of RBP7 expression in colon cancer, we tested for associations with clinicopathological variables and patient follow-up in our collection of 219 cases, which included UICC stages I and II. Using ROC curve analysis and Youden's index for cancer specific survival, we identified an optimal cut-off H-score of 32.5 for dichotomal classification into cases with high or low RBP7 expression, respectively (Figure 7A). Indeed, Kaplan-Meier analysis and log-rank testing demonstrated significantly poorer cancer specific survival of patients whose tumors were RBP7 high when compared to RBP7 low cases (*P* = 0.003; Figure 7B).

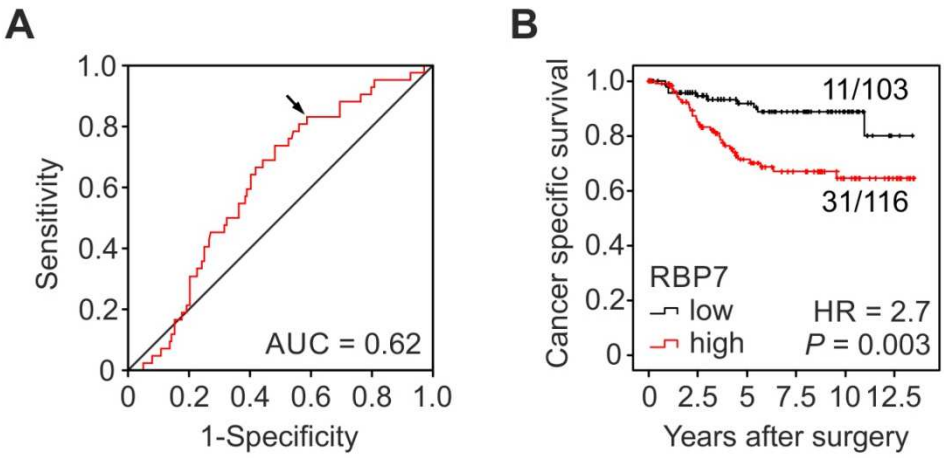

Figure 7. High RBP7 expression indicates poor survival in colon cancer patients.
(A-B) Analysis of RBP7 protein expression and cancer specific survival in a case collection of n = 219 UICC stage I and II colon cancer cases. **(A)** ROC curve for determining best discrimination thresholds of RBP7 H-scores for tumor specific survival prediction. *Arrow* indicates chosen value for binary classification. AUC, area under curve. **(B)** Kaplan-Meier plot for tumor specific survival of cases with low or high H-scores. *P*-value indicates a log-rank test result. Ratios on curves indicate the number of events over the number of patients per group. HR, hazard ratio. Figure from Elmasry et al., 2019 [1].

Next, we evaluated correlations of RBP7 high and low expression with other clinicopathological variables by Chi-square testing. High RBP7 expression marginally

significantly correlated with high tumor grade (P = 0.05), whereas we found no correlations with age, gender, T-category, or UICC-stage (Table 10).

Table 10. Clinical data and RBP7 protein expression in UICC stage I and II colon cancer
Table from Elmasry et al., 2019 [1].

Characteristics	Total	RBP7		*P*
		Low	High	
All patients	219 (100.0)	103 (47.0)	116 (53.0)	
Age (y, Median 69)				
≤ 68	108 (49.3)	51 (47.2)	57 (52.8)	0.956
≥ 69	111 (50.7)	52 (46.8)	59 (53.2)	
Gender				
Male	121 (55.3)	59 (48.8)	62 (51.2)	0.569
Female	98 (44.7)	44 (44.9)	54 (55.1)	
T-category				
T1	1 (0.5)	0 (0.0)	1 (100.0)	0.467
T2	35 (16.0)	17(48.6)	18 (51.4)	
T3	175 (79.9)	84 (48.0)	91 (52.0)	
T4	8 (3.7)	2 (25.0)	6 (75.0)	
UICC-stage				
I	36 (16.4)	17 (47.2)	19 (52.8)	0.989
II	183 (83.6)	86(47.0)	97 (53.0)	
Tumor grade (WHO)				
low grade	197 (90.0)	97 (49.2)	100 (50.8)	0.05
high grade	22 (10.0)	6 (27.3)	16 (72.7)	

Values in parentheses indicate column and row percentage for total and RBP7 low or high cases, respectively.

Moreover, proportional hazards regression analysis demonstrated that high RBP7 expression was an independent predictor of poor tumor specific survival in this case collection (HR = 2.54; P = 0.009; Table 11).

Table 11. Multivariate analysis of cancer specific survival in UICC stage I and II colon cancer. Table from Elmasry et al., 2019 [1].

Variables	Cancer specific survival		
	HR	(95% confidence interval)	*P*
Age ≥ median (69 y)	2.04	(1.06-3.93)	0.032
Female vs. male	0.76	(0.40-1.46)	0.408
T-category	3.32	(1.43-7.70)	0.005
High tumor grade	1.20	(0.52-2.78)	0.669
RBP7 high	2.54	(1.26-5.10)	0.009

These findings suggested that RBP7 is a prognostic marker in early stage colon cancer.

5.1.3 High *RBP7* expression is an independent predictor of poor survival in colon cancer

For further validation, we next tested for clinical correlations of *RBP7* mRNA levels using publicly available gene expression data of 457 colon cancer cases from TCGA , 379 of which had information on clinical follow-up. ROC curve analysis and Youden's index identified an optimal cutoff score of 21.01 *RBP7* normalized mRNA reads for dichotomal classification of cases (Figure 8A). Also in this data set, Kaplan-Meier analysis and log-rank testing demonstrated a strong positive correlation of high *RBP7* expression and poor cancer specific survival when compared to tumors with low RBP7 levels (*P* = 0.00007; Figure 8B).

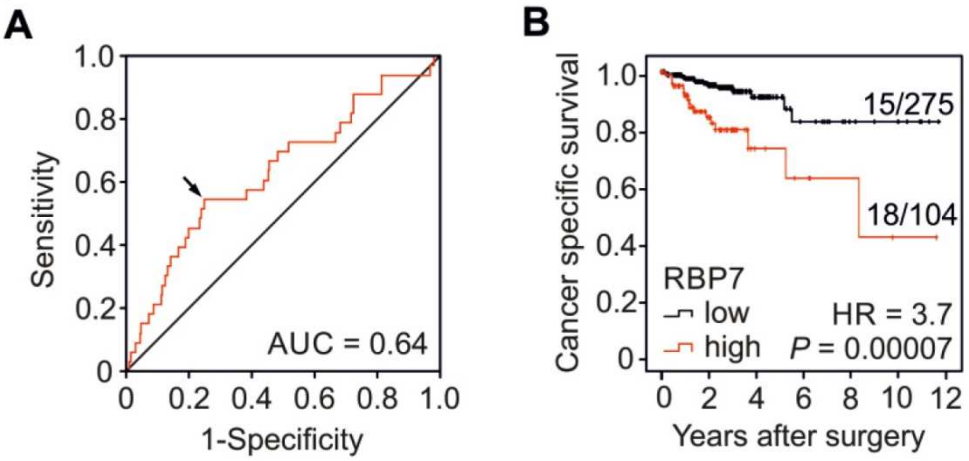

Figure 8. High *RBP7* expression indicates poor survival in colon cancer patients.
(A-B) Analysis of *RBP7* mRNA expression and cancer specific survival in n=379 colon cancer cases from TCGA **(A)** ROC curve for determining best discrimination thresholds of *RBP7* mRNA reads for survival prediction. Arrow indicates chosen value for binary classification. AUC, area under curve. **(B)** Kaplan-Meier plot for cases with low or high *RBP7* mRNA expression. *P*-value indicates a log-rank test result. Ratios on curves indicate the number of events over the number of patients per group. HR, hazard ratio. Figure from Elmasry et al., 2019 [1].

We then tested for associations with other core clinical variables, and found that the frequency of high *RBP7* expression increased with increasing T-category and also was higher in tumors that had metastasized to lymph nodes. Other variables including microsatellite instability as well as *RAS* and *BRAF* mutation status were not associated with *RBP7* (Table 12).

Table 12. Clinical data and *RBP7* mRNA expression in colon cancer cases from TCGA.

Table from Elmasry et al., 2019 [1].

Characteristics	Total	*RBP7*		*P*
		Low	High	
All patients	457 (100.0)	332 (72.6)	125 (27.4)	
Age (y, Median 68)				
≤ 67	211 (46.9)	155 (73.5)	56 (26.5)	0.875
≥ 68	239 (53.1)	174 (72.8)	65 (27.2)	
Gender				
Male	235 (52.2)	171 (72.8)	64 (27.2)	0.201
Female	215 (47.8)	158 (73.5)	57 (26.5)	
T-category				
T1	11 (2.5)	11 (100.0)	0 (0.0)	0.00002
T2	77 (16.8)	67 (87.0)	10 (13.0)	
T3	306 (67.0)	222 (72.5)	84 (27.5)	
T4	53 (11.5)	27 (50.9)	26 (49.1)	
Nodal Metastasis				
Negative	263 (58.8)	206 (78.3)	57 (21.7)	0.003
Positive	184 (41.2)	121 (65.8)	63 (34.2)	
Distant Metastasis				
Negative	326 (83.8)	249 (76.4)	77 (23.6)	0.104
Positive	63 (16.2)	42 (66.7)	21 (33.3)	
MSI status				
MSS/MSI-low	340 (81.3)	248 (72.9)	92 (27.1)	0.838
MSI-high	78 (18.7)	56 (71.8)	22 (28.2)	
RAS status				
Mutated	174 (38.1)	130 (74.7)	44 (25.3)	0.691
Wild type	216 (47.3)	153 (70.8)	63 (29.2)	
Unknown	67 (14.7)	49 (73.1)	18 (26.9)	
BRAF (V600E)				
Mutated	47 (10.3)	35 (74.5)	12 (25.5)	0.948
Wild type	343 (75.1)	248 (72.3)	95 (27.7)	
Unknown	67 (14.7)	49 (73.1)	18 (26.9)	

Values in parentheses indicate column and row percentage for total and *RBP7* low or high cases, respectively.

Furthermore, proportional hazards regression analysis including key clinical variables demonstrated independent prognostic power of high *RBP7* mRNA expression (HR = 2.5, $P = 0.038$; Table 13).

Table 13. Multivariate analysis of cancer specific survival in colon cancer cases from TCGA. Table from Elmasry et al., 2019 [1].

Variables	Cancer specific survival		
	HR	(95% confidence interval)	*P*
Age ≥ median (68 y)	0.65	(0.28-1.50)	0.314
Female vs. Male	1.09	(1.09-2.54)	0.847
T-category	2.06	(0.82-5.18)	0.124
Nodal metastasis	2.01	(0.56-7.22)	0.286
Distant metastasis	15.85	(4.66-53.91)	0.00001
MSI-high	2.00	(0.36-10.86)	0.439
RBP7 high	2.50	(1.05-5.88)	0.038

Collectively, these data provided additional evidence on the mRNA level that *RBP7* is linked to advanced tumor stages and colon cancer progression.

5.1.4 *RBP7* is linked to invasion and EMT in colon cancer

To gain insights into the functional role of RBP7 in colon cancer, we conducted Gene Set Enrichment Analyses (GSEA) using the TCGA dataset. Interestingly, when we tested for associations with curated gene sets (n = 4.762), we found a top enrichment for a multicancer invasiveness gene signature [142], while *RBP7* itself was not part of this gene set (Figure 9A).

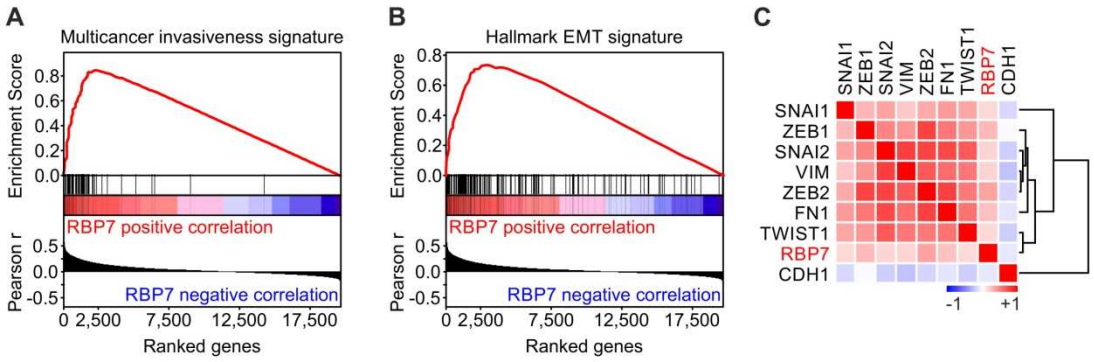

Figure 9. *RBP7* expression is linked to invasion and EMT in colon cancer.
(A-B) Gene Set Enrichment Analyses for genes ranked by Pearson correlation (Pearson r) of expression to *RBP7* indicates enrichment for **(A)** multicancer invasion and **(B)** hallmark EMT gene signatures. $P < 0.001$. **(C)** Heat map indicates clustering and positive correlation of RBP7 expression with colon cancer relevant EMT markers and negative correlation with *CDH1*. Colors indicate Pearson r from -1 (blue) to 1 (red). Figure from Elmasry et al., 2019 [1].

We then further tested for associations with hallmark gene sets (n = 50) [143], and found the strongest enrichment for genes linked to epithelial mesenchymal transition. Moreover, individual markers that indicate or drive EMT in colon cancer showed a significant overexpression in tumors with high *RBP7* expression, including *ZEB1* (r =

0.27, $P < 0.0001$) and *ZEB2* (r = 0.36, $P < 0.0001$) (Figure 9C). In contrast, the epithelial differentiation marker *CDH1* negatively correlated with *RBP7* (Figure 9C). Importantly, *RBP7* itself again was not part of this EMT gene set. These findings suggested a previously unknown functional link of RBP7, invasion and EMT in colon cancer cells.

5.1.5 Overexpression of RBP7 enhances migration and invasion of colon cancer cells

Finally, due to its link with EMT and cancer invasion, we tested for a functional relevance of RBP7 for invasion and migration of colon cancer cells. We constructed a vector for transient overexpression of RBP7. Transfection of HCT116 and SW1222 colon cancer cells with RBP7 encoding vector caused strong ectopic expression in both cell lines, when compared to empty control vector (Figure 10A).

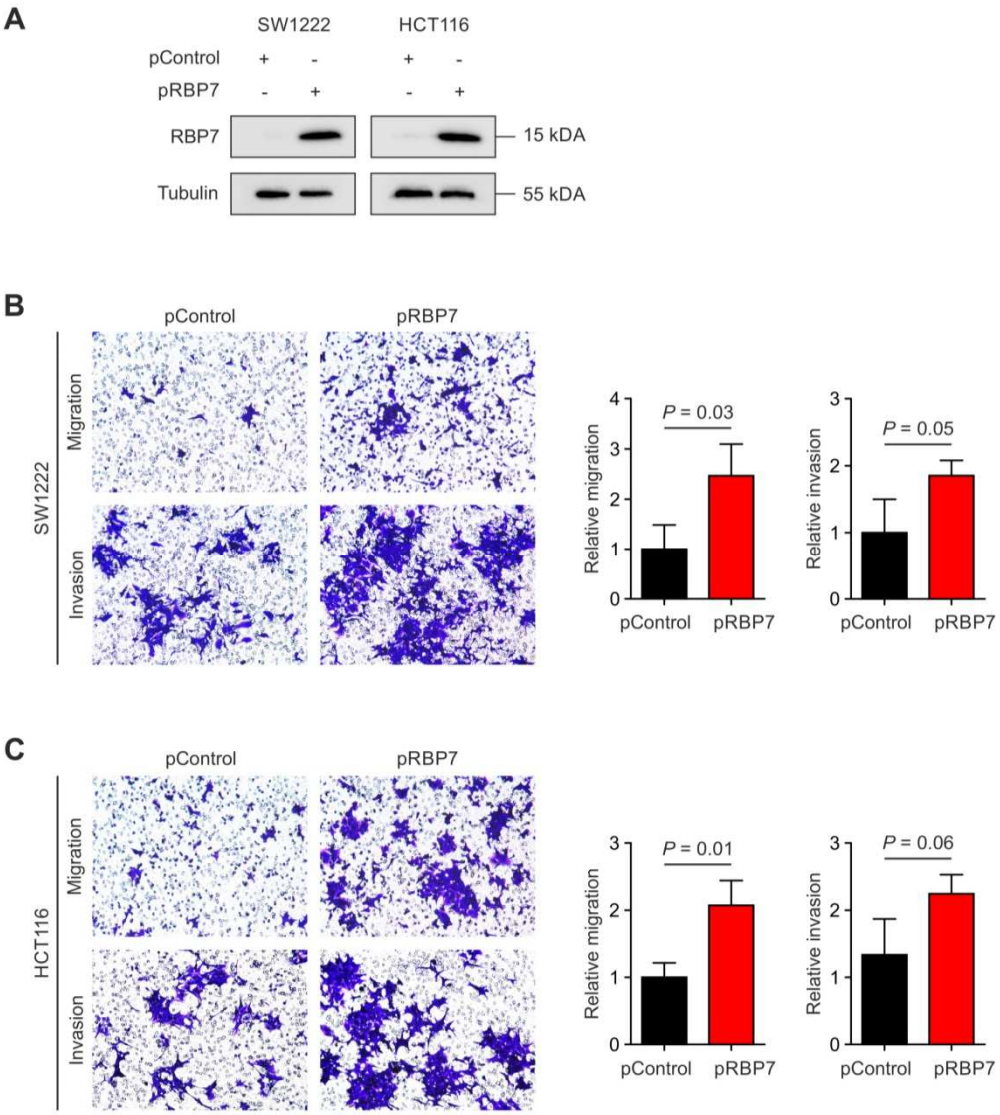

Figure 10. Overexpression of RBP7 enhances migration and invasion of colon cancer cells.

(A) Immunoblotting for indicated proteins on whole cell lysates of SW1222 and HCT116 colon cancer cells harvested 48 h after transfection with pcDNA3.1-RBP7 (pRBP7) or empty pcDNA3.1 (pControl) vector. **(B-C)** Representative micrographs (left panels) and quantification (right panels) of migrated or invaded **(B)** SW1222 and **(C)** HCT116 colon cancer cells in transwell assays. Data are mean ± SD, n ≥ 3, *P*-values are t-test results. Figure from Elmasry et al., 2019 [1].

We then seeded both cell lines in Boyden chamber assays that were coated with matrigel for invasion. Importantly, ectopic expression of RBP7 increased the number of migrated and invaded tumor cells, and these effects were comparable in both cell lines (Figure 10B-C). These findings supported the idea that RBP7 is a regulator of invasion and migration, which are malignant traits of colon cancer progression.

5.2 Unbiased screening for prognostic biomarker identifies Annexin A9 for independent risk stratification in colon cancer

5.2.1 Identification of prognostic indictors in colon cancer

To identify predictors of survival and tumor metastasis in colon cancer patients in an unbiased approach, we used gene expression and clinical data of 457 colon cancer cases from The Cancer Genome Atlas network (TCGA). For each of the 20,253 genes represented in this data set, we plotted receiver operator characteristics (ROC) curves for sensitivity and specificity in discriminating cancer survival and tumor metastasis, and then calculated the area under the curve (AUC) for each gene (Figure 11)

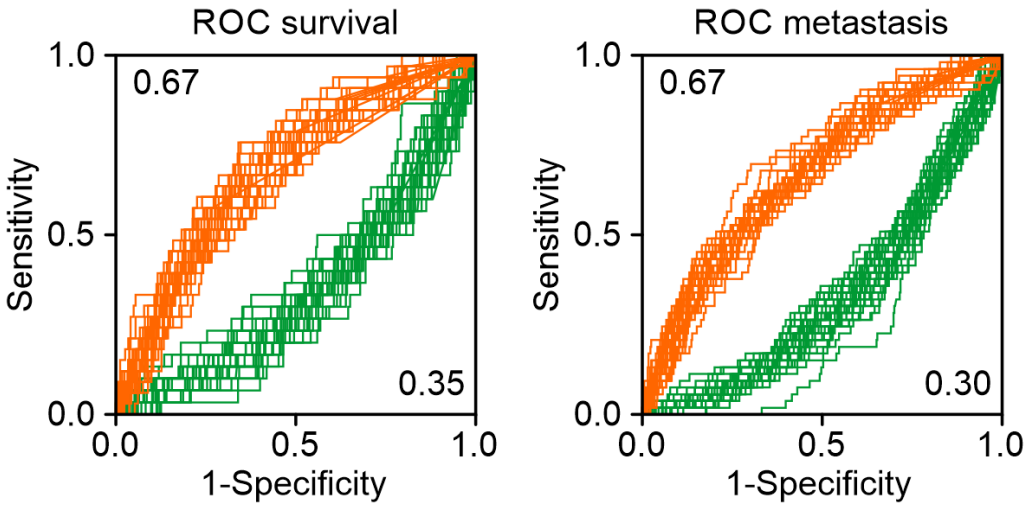

Figure 11. ROC curves for sensitivity and specificity in discriminating cancer survival and tumor metastasis
Using mRNA expression level of each gene in TCGA data. Orange curves indicate genes with AUC values > 0.5, while green curves indicate genes with AUC values < 0.5.

To equally weigh both AUC-survival and AUC-metastasis values for further analyses, we then transformed them into normalized z-scores for survival (z^{surv}) and metastasis (z^{met}). Direct comparison of z^{surv} and z^{met} for all genes confirmed the expected linkage of survival and metastasis prediction in this data set (Figure 12A), and allowed us to combine both into a single average $z^{surv/met}$ score. Next, to determine significance levels for outcome prediction, we classified colon cancer cases into those with low and high expression for each gene by expression median, and calculated *P*-values for survival and metastasis prediction by Log-Rank and Chi-square testing, respectively. Moreover, Plotting average *P*-values of both tests against $z^{surv/met}$ illustrated that genes with high significance levels in survival and metastasis prediction were among those with lowest and highest $z^{surv/met}$ scores and validating that our approach identified potential predictors of cancer outcome in this dataset (Figure 12B). Of note, when comparing median gene expression levels to $z^{surv/met}$ scores, genes that predicted outcome showed rather average expression levels, while genes with low or extremely high expression levels were less likely linked to patient outcome (Figure 12C).

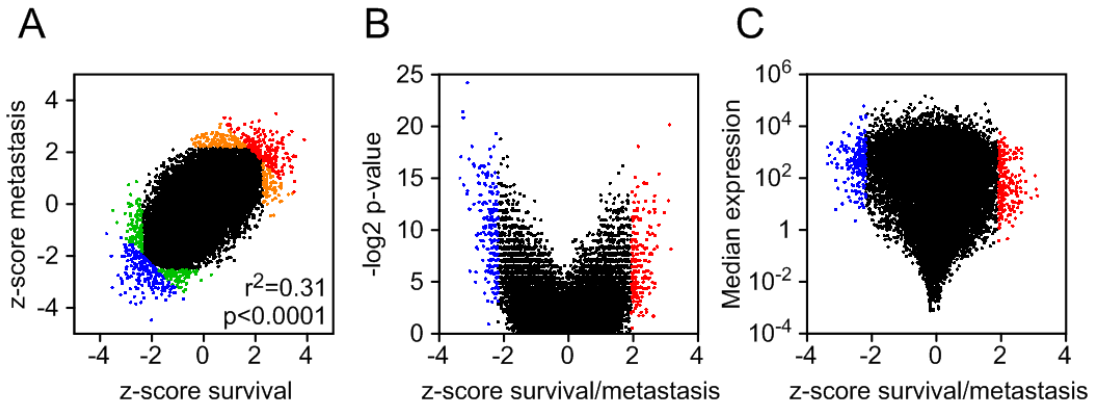

Figure 12. Volcano plots of normalized z-scores for survival (z^{surv}) and metastasis (z^{met}) for all TCGA genes.

(A) Direct comparison of z^{surv} and z^{met} for all genes confirm the expected linkage of survival and metastasis in TCGA data. **(B)** Plotting average -*P* values for survival and metastasis prediction against $z^{surv/met}$ shows that genes with high significance levels in survival and metastasis prediction were among those with lowest and highest $z^{surv/met}$ scores. **(C)** Comparing median gene expression levels to $z^{surv/met}$ scores shows that genes with low or extremely high expression levels were less likely linked to patient outcome. Colored dots denote genes that are significantly upregulated (red & orange) or downregulated (green & blue) $P = 0.0001$, $r^2 = 0.31$.

Ranking all genes based on $z^{surv/met}$ and filtering for those with significant *P*-values for survival and metastasis prediction resulted in a final list of 194 putative candidates with prognostic power in colon cancer.

5.2.2 Characterization of identified indictors according to their encoded Protein and survival association

To further characterize genes in this list, we added information on classes of proteins they encode. Of note, we found several candidates indicating poor survival and metastasis encoded for known or potential drug targets, transporters, transcription factors, or G-protein coupled receptors (Figure 13A).

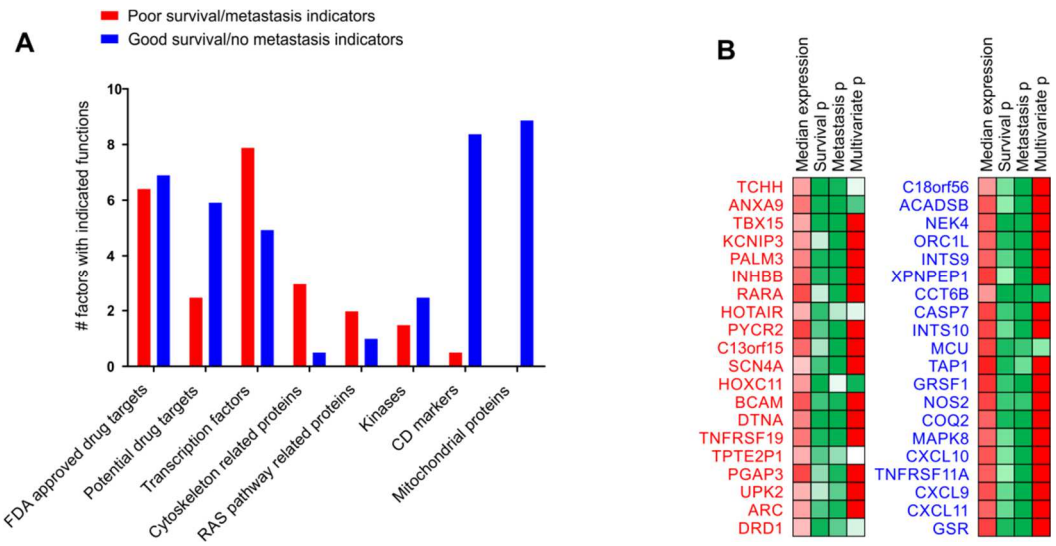

Figure 13. Characterization of identified prognostic candidate genes according to classes of proteins they encode and survival association.

(A) The result of PANTHER analysis of 194 putative prognostic genes showing the classes of proteins they encode. Blue bars indicate genes associated with good survival and absence of metastasis in CRC, whereas Red bars indicate genes associated with poor survival and metastasis in CRC. **(B)** Heat maps of the top 35 candidate genes for CRC prognosis indicators showing the correlation of each gene expression median and p values of survival, metastasis and main clinically relevant covariates prediction. Genes written in red color are associated with poor CRC outcome, while genes written in blue color are associated with good CRC outcome. Colors indicate Pearson r from -1 (red) to 1 (green).

On the contrary, mitochondrial proteins were solely encoded by genes whose expression indicated good survival and absence of metastasis. Additionally, to determine which of these candidates may be potentially independent predictors of cancer survival, we calculated proportional hazards regression analyses for survival association for each gene, including age, gender, T-stage, nodal status, and microsatellite instability (MSI) as main clinically relevant covariates. 35 genes significantly passed this analysis, suggesting independent prognostic power in colon cancer, with *TCHH* and *ANXA9* ranking top in poor survival prediction (Figure 13B).

5.2.3 Identification of prognostic potential of the top 5 independent indicators of colon cancer outcome

To test for prognostic usability of genes identified by this method, we focused on *TCHH*, *ANXA9*, and *HOTAIR* as the top independent genes whose expression indicated poor outcome, and *CCT6B* and *MCU* as top candidates indicating good outcome. Using ROC curve analyses, we manually determined best cutoff scores for dichotomal

classification of colon cancers by expression of these genes for survival prediction (Figure 14A). As expected, Kaplan-Meier plots indicated strong separation into groups with good and poor cancer survival by these genes (Figure 14B), and significantly selected for cases with high or low proportions of metastasized tumors, respectively (Figure 14C). Among these markers, *ANXA9* performed best in multivariate analysis and high expression most significantly indicated tumors that had metastasized. Collectively, we therefore suggest that our list of genes derived from TCGA data provides a useful resource to identify not only predictors of outcome in colon cancer but also new potential therapeutic targets, and functionally relevant factors for cancer progression.

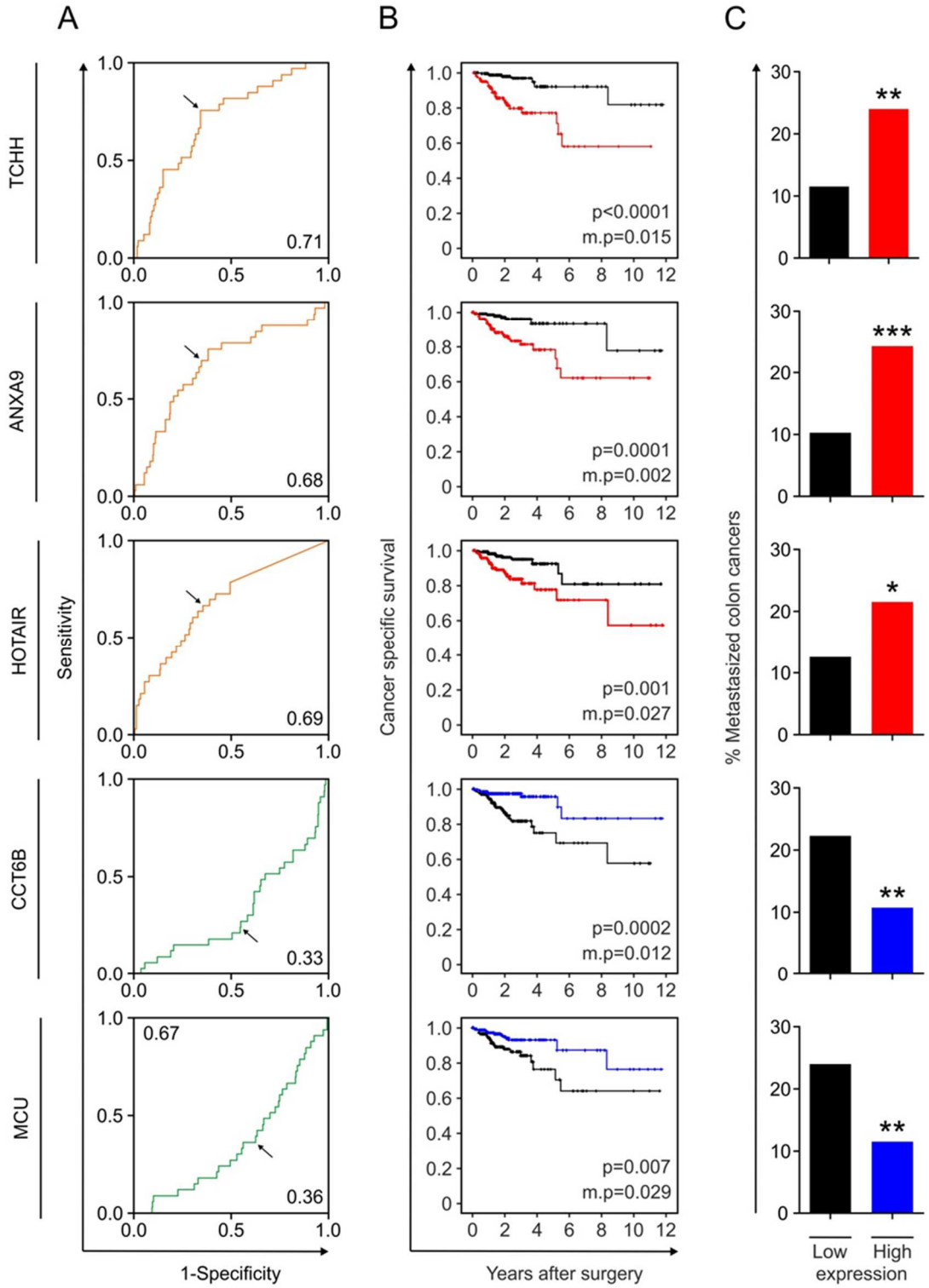

Figure 14. The prognostic potential of the top 5 independent indicators of colon cancer outcome.

(A) ROC curves for determining best discrimination thresholds of the top 5 prognostic genes mRNA reads for survival prediction. Arrows indicate chosen value for binary classification. AUC = Area under curve. **(B)** Kaplan-Meier statistics for binary (low and high) classified top 5 prognostic genes expression indicates shorter tumor specific survival for *TCHH*, *ANXA9*, and *HOTAIR* high expression and longer tumor survival for *CCT6B* and *MCU* low expression. Significance p-values indicate log-rank test results. Ratios on curves indicate the number of events over the number of patients per group. m.p = malignant potential. **(C)** Cases with high *TCHH*, *ANXA9*, and *HOTAIR* expression are significantly (t-test) associated with distant metastasis when compared to cases with low expression of these genes. While cases with high *CCT6B* and *MCU* expression are significantly (t-test) not associated with distant metastasis. *P*-values are log-rank test results. *, $P \leq 0.05$, **, $P \leq 0.01$, ***, $P \leq 0.001$.

5.2.4 Validation of ANXA9 as a strong prognostic marker on the protein level

Due to the highest significance in multivariate survival and metastasis prediction (Figure 15), we selected ANXA9 to test if our findings from the TCGA data set may be translated on the protein level in independent tissue collections of colorectal cancer. We used two collections, one stratified for survival analysis of cases with UICC stage II colon cancer (n = 244) and long term follow-up data, and a second matched case control collection of colon cancer patients with and without synchronous liver metastasis (n = 90).

We analyzed both collections for ANXA9 protein expression by immunostaining and found differential expression among these cases, ranging from complete absence (2.4 %, score 0), through weak (65.3 %, score 1) and moderate (29.0 %, score 2) expression, to strong expression (3.3 %, score 3, Figure 15).

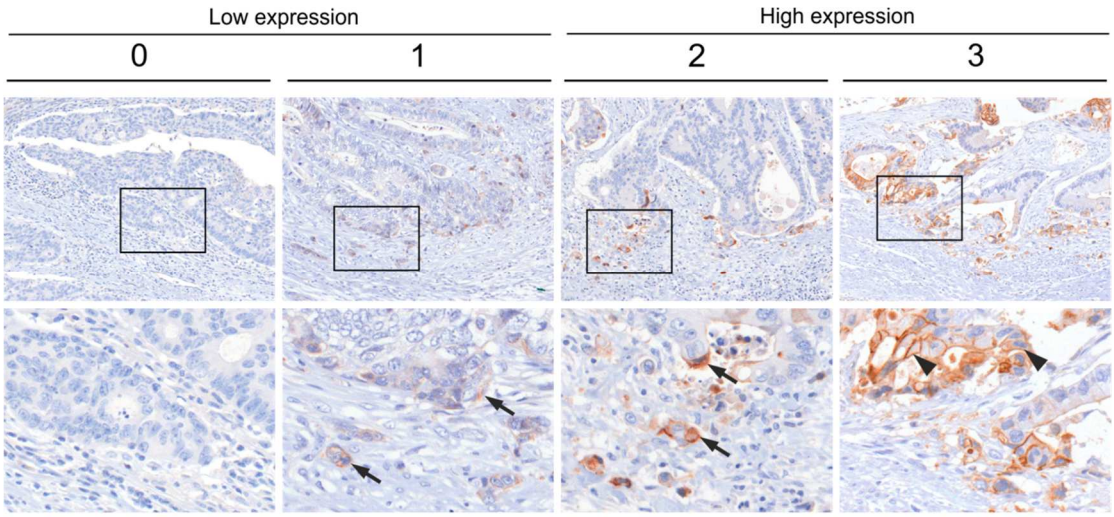

Figure 15. Immunostaining of ANXA9 in colorectal cancer tissues.
Assessment of ANXA9 staining in a collection of 244 cases with stage II colon cancer. ANXA9 staining is restricted to the cytoplasm of colorectal cancer cells. According to intensity of ANXA9 staining, tumors were given scores from 0 (no ANXA9 staining) to 3 (strong ANXA9 staining in most tumor cells), and moreover categorized as ANXA9 low (score 0–1) and ANXA9 high (score 2–3). Arrows indicate strong ANXA9 staining in tumor cells at the infiltrative edge. Lower panel images are magnifications of areas boxed in upper panel images (20X magnification).

Furthermore, ANXA9 expression was heterogeneous within the tumors and was most strong in tumor cells at the leading tumor edge, including cells that invaded the surrounding stroma by apparently detaching from the gland forming tumor mass (Figure 15). In cases with weak and moderate expression, ANXA9 was absent in gland forming tumor cells of the tumor center, while in cases with strong ANXA9 expression,

staining extended to this compartment (Figure 15). Interestingly, stromal cells surrounding the tumor were negative for ANXA9, indicating that tumor cells exclusively or predominantly contributed to ANXA9 expression levels.

5.2.5 High expression of ANXA9 predicts poor survival in colon cancer

To test for clinical relevance of ANXA9 protein expression, we then looked at cancer survival in the collection of 244 colon cancer cases with clinical follow-up data. In Kaplan-Meier curves, ANXA9 expression scores significantly separated patients into groups with different cancer specific survival (Figure 16).

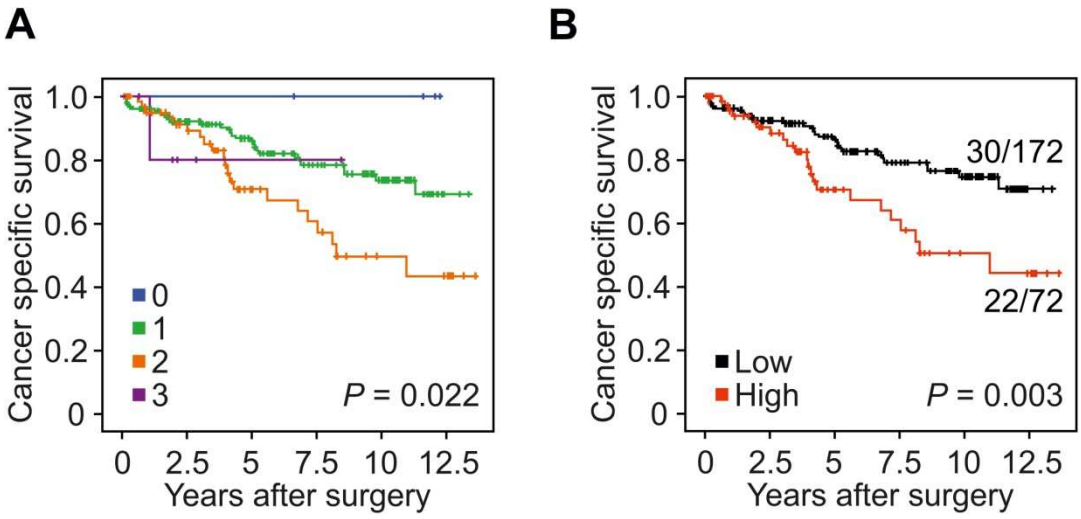

Figure 16. High ANXA9 indicates poor survival in colorectal cancer.

(A) Kaplan-Meier plots of cancer specific survival for different ANXA9 expression scores indicate significant poorer outcome with increasing ANXA9 expression. **(B)** Kaplan-Meier analysis of cancer specific survival for dichotomal ANXA9 expression (low and high). *P*-values indicate log-rank test results. Ratios on curves indicate the number of events over the number of patients per group.

Importantly, all patients whose colon cancers were negative for ANXA9 (score 0) fully survived their follow-up period, while moderate or strong ANXA9 expression indicated worse outcomes (Figure 16A). Based on these observations and due to low prevalence of cases with ANXA9 expression scores 0 and 3, we then reclassified cases into two categories of low (scores 0 and 1) and high (scores 2 and 3) ANXA9 expression only. Kaplan-Meier curves indicated highly significantly different patient outcome of these two categories, with ten year survival rates of 75 % vs. 51 % of cases with low and high

ANXA9 expression, respectively (Figure 16B). Next, we determined associations of ANXA9 expression with other clinical and pathological variables. High ANXA9 expression was slightly more frequent in female patients, while no associations with age, tumor grade or T-stage were found in this collection (Table 14).

Table 14. Clinical data and ANXA9 protein expression in UICC stage II colon cancer.

Characteristics	Total	ANXA9		*p*
		Low	High	
All patients	244 (100)	172 (70.5)	72 (29.5)	
Age (y, Median 69)				
≤ 68	122 (50.0)	92 (75.4)	30 (24.6)	0.092
≥ 69	122 (50.0)	80 (65.5)	42 (34.5)	
Gender				
Male	131 (53.7)	100 (76.3)	31 (23.7)	0.031
Female	113 (46.3)	72 (63.7)	41 (36.3)	
Tumor stage (UICC)				
T3	201 (82.4)	144 (71.6)	57 (28.4)	0.394
T4	43 (17.6)	28 (65.1)	15 (34.9)	
Tumor grade (WHO)				
low grade	220 (90.2)	153 (69.5)	67 (30.5)	0.326
high grade	24 (9.8)	19 (79.2)	5 (20.8)	

Values in parentheses indicate column and row percentage for total and ANXA9 low or high cases, respectively.

Importantly, multivariate proportional hazards regression analyses demonstrated that ANXA9 indicated poor cancer specific survival, independent of other clinical and pathological variables (Table 15).

Table 15. Multivariate analysis of cancer specific survival in UICC stage II colon cancer

Variables	Cancer specific survival		
	HR	(95% confidence interval)	*P*
Age ≥ median (69 y)	1.32	(0.76-2.31)	0.316
Female vs. male	0.94	(0.54-1.63)	0.828
T-stage	2.52	(1.39-4.57)	0.374
High tumor grade	0.59	(0.18-1.89)	0.002
ANXA9 high	2.00	(1.13-3.52)	0.017

Taken together, ANXA9 protein expression robustly predicted poor outcome in colorectal cancer patients in univariate and multivariate analyses.

5.2.6 High ANXA9 expression is strongly correlated with metastasis in colon cancer

Finally, we tested for associations of ANXA9 expression and tumor metastasis in a case control collection of 90 colon cancer patients, half of which had metastasized to the liver at the time of diagnosis. Indeed, in this collection high ANXA9 expression strongly correlated with synchronous liver metastasis (Figure 17A, Table 16), while we also observed a weak association with nodal metastasis (Figure 17B, Table 16).

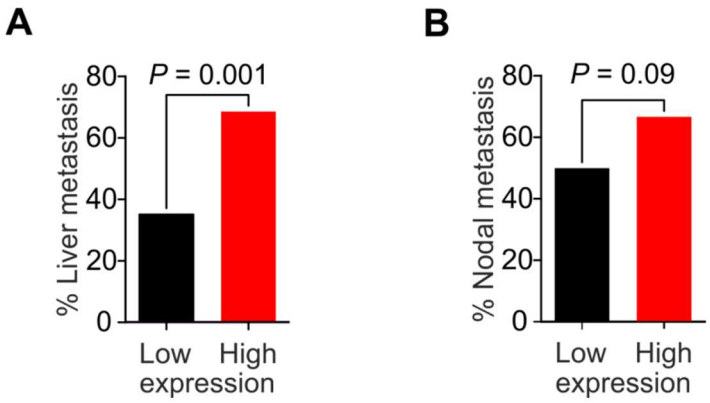

Figure 17. High ANXA9 expression is significantly correlated with metastasis in colon cancer.

(A) Cases with high ANXA9 expression in a matched case control collection of 90 colon cancer patients are significantly associated with liver distant metastasis when compared to cases with low ANXA9 expression. **(B)** The same cases as in (A) show a weak association with nodal metastasis. P-values are t-test results.

In addition, and contrary to the survival collection, high ANXA9 expression was slightly more frequent in male patients, suggesting that associations of ANXA9 and gender where at random (Table 16).

Table 16. Clinicopathological variables and correlation with ANXA9 protein expression in a case control collection of 90 colon cancer patients

Characteristics	Total	ANXA9		*P*
		Low	High	
All patients	90 (100)			
Age (y, Median 68)				
≤ 68	46 (51.1)	24 (52.2)	22 (47.8)	0.121
≥ 69	44 (48.9)	30 (68.2)	14 (31.8)	
Gender				
Male	43 (47.8)	21 (48.8)	22 (51.2)	0.039
Female	47 (52.2)	33 (70.2)	14 (29.8)	
Tumor size (UICC)				
T2	8 (8.9)	5 (62.5)	3 (37.5)	0.844
T3	67 (74.4)	41 (61.2)	26 (38.8)	
T4	15 (16.7)	8 (53.3)	7 (46.7)	
Nodal status				
N0	38 (42.2)	27 (71.1)	11 (28.9)	0.0673
N+	52 (57.8)	27 (52)	25 (48)	
Metastasis (Liver)				
M0	45 (50)	35 (77.8)	10 (22.2)	0.0006
M1	45 (50)	19 (42.2)	26 (57.8)	
Tumor grade (WHO)				
G neu 2	30 (33.3)	18 (60)	12 (40)	0.502
G neu 3	58 (64.4)	34 (58.6)	24 (41.4)	
G neu 4	2 (2.3)	2 (100)	0 (0)	

Values in parentheses indicate column and row percentage for total and ANXA9 low or high cases, respectively

Similarly to our findings in the survival collection, no correlations of ANXA9 with age, tumor grade or T-stage were found. Collectively, our findings demonstrate that ANXA9 is a strong and independent marker for poor outcome and tumor progression in colorectal cancer, derived from the TCGA data set and validated in independent colorectal cancer case collections.

6 DISCUSSION

Parts of the discussion presented in this section are from Elmasry et al., 2019 [1]

6.1 RBP7 is a clinically prognostic biomarker and linked to tumor invasion and EMT in colon cancer

RBP7 is one of the cellular retinol-binding proteins (CRBPs), which play an important role in vitamin A stability and metabolism [144]. Vitamin A and its metabolites act as regulators of multiple biological activities, such as epithelial cell proliferation, differentiation, and apoptosis [145]. Previous studies suggested that some CRBP members and Vitamin A signaling might promote CRC progression and regulate colon cancer stem cell functions [146,147]. In addition, it was recently reported that RBP7 expression is significantly increase in renal cell carcinomas and suggested that it may contribute to the progression of some types of cancer [148]. However, nothing is known about RBP7 expression and function in colon cancer.

This work demonstrates that RBP7 is a prognostic biomarker in colon cancer. Using a collection of 219 stage I and II colon cancer cases with long-term follow-up data, we show that high levels of RBP7 protein expression were strongly linked to poor cancer specific survival. This finding is important when considering that the clinical management of patients with colorectal cancer is mainly guided by disease stage. While patients with early and localized stage I and II disease generally have the best prognosis, and in most cases can be curatively treated by surgical resection alone, patients with advanced and metastatic disease may benefit from primary or adjuvant chemotherapy [149]. However, in about 25-30 % of early stage colon cancer cases, the disease still recurs and progresses after surgical management, and may ultimately be fatal [108].

Therefore, high expression of RBP7 may be particularly useful to identify patients with colon cancer that are at high risk for disease progression, and thus may be candidates for adjuvant chemotherapy and increased clinical attention, despite low clinical stage. Additionally, we found that also on the mRNA level, *RBP7* was significantly linked with poor outcome in an independent collection of 457 colon cancers from TCGA that included all tumor stages, and also was an independent prognostic biomarker in this case collection. This further validated and broadened the results from our own tissue collection. In addition, in this data set RBP7 levels increased with T-category and thus with the depth of bowel wall infiltration, suggesting a link of RBP7 and tumor invasion. Collectively, these findings demonstrated significant biomarker potential of RBP7 on protein and mRNA levels that may be useful for risk stratification in patients with colon cancer.

Disease progression of colon cancer requires invasive growth of tumor cells into surrounding tissues, blood vessels, and lymphatics [150]. Invasion often is tied to a loss of epithelial characteristics during epithelial-mesenchymal transition (EMT) [151]. Looking at thousands of different curated and hallmark gene sets with GSEA, we found that *RBP7* was most strongly linked to a multicancer invasiveness signature, as well as to a hallmark EMT gene signature [141-143].

This is of particular interest because markers that indicate invasion and EMT in colon cancer are scarce, and detection of typical EMT biomarkers such as ZEB1, SNAIL1, or Vimentin in cancer tissues can be challenging, as reflected by yet few convincing in situ studies [152,153]. On the contrary, we demonstrate that RBP7 can be robustly detected by immunohistochemistry in primary colon cancer, and quantified by digital image analysis. We therefore suggest that RBP7 may serve as a surrogate marker that indicates the overall degree of tumor invasion and EMT within colon cancer, which also may explain its association with poor prognosis. However, before implementation in a

clinical setting, i.e. to complement tumor staging, further independent confirmation of these findings will be mandatory.

Ectopic expression of RBP7 increased migration and invasion, which demonstrates a direct functional contribution of RBP7 to the malignant traits of colon cancer cells. This may explain the association of RBP7 with invasion, EMT, and poor prognosis that we observed in colon cancer case collections. Previous studies demonstrated that another member, RBP4 and its receptor, are potent oncogenes in human breast and colon cancer cells that drive malignant transformation [154]. Furthermore, RBP4 expression in colon cancer has been associated with poor prognosis, promoted growth in xenograft models, and increased the expression of putative cancer stem cell antigens [147]. Considering these findings, our data provide a new link of retinol metabolism, invasion, and EMT in colon cancer through RBP7. However, the exact mechanism by which RBP7 drives these malignant traits and affects the transcriptome of colon cancer cells remains to be determined. Further study also is required to elaborate whether therapeutic interference with retinol metabolism and RBP7 may be a strategy to target invasion, EMT, and colon cancer progression.

6.2 An unbiased screen for prognostic marker identifies Annexin A9 for independent risk stratification in colon cancer

Disease prognosis and survival probability of colorectal cancer yet can most robustly be assessed by tumor staging, describing the extent of the disease [95]. However, since tumors of identical stage can behave significantly different in aggressiveness followed by different outcomes, there is need for additional robust and validated molecular predictors of biological behavior and disease progression, so that treatment protocols may be personalized based on individual risk. Although a multitude of molecular prognostic markers for colon cancer have been published, many markers lack robust independent validation, or are commercial and proprietary in interpretation [136,155].

In this work, using a large open access data set of 457 colon cancer cases from the TCGA, we identified genes whose expression is significantly linked to patient survival and tumor metastasis. For identification of these genes, we applied large scale ROC curve analyses, a method for establishing medical diagnostic tests [156]. The full list of genes provides information on prognostic power and independence of staging and other key clinical variables. Moreover, we added information on functional properties of proteins encoded by these genes. We suggest this may be a useful resource to develop robust open access molecular signatures to predict more precisely colon cancer outcome, and to identify individual factors driving tumor progression or indicating response to personalized cancer therapy and subsequently to avoid unnecessary treatment, toxicity and the financial costs associated with therapy.

Among the top predictors of poor survival and metastasis, ANXA9 most significantly was independent of other clinical variables, and thisfinding validates previous studies on ANXA9 and prognosis in colon cancer [157,158]. Moreover, this provided a rationale to select this factor for validation on the protein level by immunostaining in two independent case collections, stratified for survival and metastasis, respectively. Using

a collection of UICC stage II colon cancer cases, high ANXA9 protein expression indicated poor cancer specific survival, independent of other clinical variables. Patients with Stage II colon cancer generally have a good outcome. However, a high risk group of these patients may have a worse outcome. Therefore, identification of patients with stage II colon cancer who are at high risk of poor oncologic outcome is very important to select for cases that may benefit from additional chemotherapy. Based on these observations, we suggest that assessing ANXA9 expression may identify patients with more aggressive stage II colon cancer that may benefit from adjuvant chemotherapy despite low tumor stage. Due to robust detection of ANXA9 in tumor cells by immunostaining, and absence of possibly confounding ANXA9 expression in surrounding stromal cells, assessing this marker may well be integrated into routine pathology workup of colorectal cancer specimens and guide the decision for adjuvant therapy for patients with stage II colon cancer. Furthermore, a strong correlation of ANXA9 expression and liver metastasis was demonstrated in a second independent case control collection, and this finding further strengthens the validity of this marker in predicting colon cancer outcome.

This work explains an approach for unbiased extraction of prognostically useful markers for colon cancer, and determines ANXA9 as the most robust candidate in predicting colon cancer outcome and progression on the mRNA and protein levels. ANXA9 is a member of the annexin family of calcium and phospholipid binding proteins and only little is known about its cellular regulation and function [159-161]. Although functional characterization of this protein in colon cancer is beyond the scope of this work, its localization within tumor cells of colon cancer is notable and might guide further investigations. We found ANXA9 expression primarily in colon cancer cells at the leading tumor edge, including tumor cells that apparently dissociate from the tumor mass. These cells have also been termed budding colon cancer cells and are attributed certain characteristics, such as loss of epithelial markers and putative cancer stem cell

traits. It might therefore be speculated that ANXA9 marks colon cancer cells undergoing epithelial-mesenchymal transition or putative colon cancer stem cells, a hypothesis that might explain its strong link to poor survival and metastasis. Collectively, this work confirms that tumor cells at the leading tumor edge indeed might be drivers for malignant progression and potential therapeutic targets. Additionally, we propose Annexin A9 as an independent prognostic predictor of poor outcomes in colon cancer by means of immunohistochemical analysis.

SUMMARY

For patients with colorectal cancer, the risk for disease recurrence and death mainly depends on disease stage. Yet, patients with early stage colon cancer may still succumb to the disease. Therefore, to improve the management of patients with colorectal cancer, new biomarkers for risk stratification are needed that are independent of tumor stage.

Here, we demonstrate that RBP7 is a strong prognostic biomarker in colon cancer that functionally contributes to the malignant phenotype of colon cancer cells. We quantified RBP7 expression in colon cancer tissue by digital image analysis, and high levels of RBP7 protein and mRNA expressions were associated with poor cancer specific survival. Additionally, GSEA analysis and cell migration and invasion assays demonstrated that RBP7 is functionally linked to invasion and epithelial-mesenchymal transition in colon cancer.

Furthermore, we illustrate here an unbiased approach using publically available TCGA data to identify new biomarkers that may aid in colorectal cancer risk stratification beyond clinical staging. By this approach Annexin A9 was identified and validated as an independent prognostic predictor of poor outcomes and that was associated with distant metastasis in independent colon cancer case collections on the protein level.

Collectively, these findings provide a rationale for considering RBP7 and Annexin A9 as promising independent predictors for prognosis. These may be useful for risk stratification in patients with colorectal cancer and aid in improving patient management.

ZUSAMMENFASSUNG

Bei Patienten mit kolorektalen Karzinomen ist das Risiko für Tumorrezidive nach Therapie und für eine tödlich verlaufende Krebserkrankung eng mit dem Tumorstadium assoziiert. Allerdings können auch Patienten mit frühen Tumorstadien letztlich an ihrer Krebserkrankung versterben. Um die Behandlung von Patienten mit kolorektalen Karzinomen zu verbessern sind daher neue Biomarker erforderlich, die das Risiko für einen Krankheitsprogress unabhängig vom Tumorstadium vorhersagen können.

Hier zeigen wir, dass RBP7 ein starker prognostischer Biomarker bei Kolonkarzinomen ist, der funktionell zu den malignen Eigenschaften von Kolonkarzinomzellen beiträgt. Die RBP7-Expression in Kolonkarzinomengewebe konnte durch digitale Bildanalyse quantifiziert werden, und hohe Spiegel von RBP7 Protein als auch von *RBP7* mRNA waren mit einem schlechten krebsspezifischen Überleben verbunden. Zusätzlich zeigten GSEA-Analysen sowie Zellmigrations- und Invasions-Assays, dass RBP7 funktionell zu den invasiven Eigenschaften von Kolonkarzinomzellen beiträgt und mit epithelial-mesenchymaler Transition (EMT) assoziiert ist.

Darüber hinaus zeigen wir hier einen unvoreingenommenen Ansatz unter Verwendung von öffentlich verfügbaren TCGA-Daten, um neue Biomarker zu identifizieren, die sich für die Risikostratifizierung beim Kolonkarzinom über das klinische Stadium hinaus eignen. Über diesen Ansatz konnten wir Annexin A9 als unabhängigen prognostischen Prädiktor für schlechtes Überleben und für Metastasierung in zwei unabhängigen Kolonkarzinomkollektiven identifizieren und validieren.

Zusammengenommen zeigen diese Ergebnisse, dass sich RBP7 und Annexin A9 als neue und vielversprechende unabhängige Biomarker eignen könnten, um die Prognose bei Patienten mit Kolonkarzinomen vorherzusagen und deren Behandlung zu verbessern.

ABBREVIATIONS

A	AJCC	American Joint Committee on Cancer
	ANXA9	Annexin A9
	APC	Adenomatous polyposis coli
	AUC	Area under the curve
B	BAX	BCL2 associated X
	BM	Basement membrane
	BSA	Bovine serum albumin
	BRAF	B-Raf proto-oncogene, serine/threonine kinase
C	CCT6B	Chaperonin containing TCP1 subunit 6B
	CDH1	Cadherin-1
	CDKN2A	Cycline Dependent Kinase Inhibitor 2A
	CEA	Carcinoembryonic Antigen
	CIN	Chromosomal Instability
	CIMP	CpG island methylator phenotype
	CMS	Consensus molecular subtype
	CNAs	Copy number alterations
	CRBPs	Cytoplasmic retinol binding proteins
	CRC	Colorectal cancer
	CTCs	Circulating tumor cells
D	DAB	3,3'-Diaminobenzidine
	DMEM	Dulbecco's Modified Eagle Medium
	DMSO	Dimethylsulfoxide

	DNA	Deoxyribonucleic acid
E	ECM	Extracellular Matrix
	E. coli	Escherichia coli
	EDTA	Ethylene Diamine Triacetic Acid
	ERK	Extracellular signal-regulated kinases
	EGFR	Epidermal growth factor receptor
	EMT	Epithelial-mesenchymal transition
	EMT-TFs	Epithelial-mesenchymal transition - inducing transcription factors
	ERK	Extracellular signal-regulated kinase
	ES	Enrichment score
F	FAP	Familial adenomatous polyposis
	FBS	Fetal bovine serum
	FDA	Food and Drug Administration
	FFPE	Formalin-fixed, paraffin-embedded
G	G-protein	guanine nucleotide-binding proteins
	GTPase	Guanosine triphosphatase
	GSEA	Gene set enrichment analysis
H	H&E	Hematoxylin and Eosin
	HOTAIR	HOX transcript antisense RNA
	HR	Hazard ratio
	HRP	Horseradish peroxidase
I	IDT	Integrated DNA technologies
	IHC	Immunohistochemistry

K	KRAS	KRAS proto-oncogene, GTPase
L	LOH	Loss of heterozygosity
	LVI	Lymphovascular invasion
M	MAPK	Mitogen-activated protein kinase
	MCU	Mitochondrial calcium uniporter
	MEK	Mitogen-activated protein kinase kinase
	MET	Mesenchymal-epithelial transition
	miRNAs	microRNA
	MLH1	Mut L homologue 1
	MMR	Mismatch Repair
	MSI	Microsatellite instability
	MSI-H	Microsatellite instability-High
	MSI-L	Microsatellite instability-Low
	MSS	Microsatellite Stable
N	NaCl	Sodium chloride
	NOS	Not otherwise specified
	NRAS	NRAS proto-oncogene, GTPase
P	PD-1	programmed-cell-death protein 1
	PI3K	Phosphatidylinositol-3-Kinase
	PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
	PNI	Perineural invasion
	PTEN	Phosphatase and tensin homologue deleted on chromosome 10

	PVDF	polyvinylidene difluoride
Q	QuPath	Quantitative Pathology
R	RAF	Raf Proto-Oncogen, Serine/Threonine Kinase
	RBP7	Retinol Binding Protein 7
	RIPA	Radioimmunoprecipitation assay
	ROC	Receiver operating characteristic
S	SDS	Sodium dodecyl sulfate
	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
	SMAD	homolog of the Drosophila protein, mothers against decapentaplegic
	SLUG	SNAI2; snail homolog 2
	SNAI1	snail homolog 1, a Zn finger protein
	SSLs	Sessile serrated lesions
T	TCGA	The Cancer Genome Atlas
	TCHH	Trichohyalin
	TFs	Transcription factors
	TGF-β	Transforming growth factor-β
	TGFβRII	Transforming growth factor-β receptor II
	TMAs	Tissue microarrays
	TNM	Tumor, Nodes, Metastasis
	TP53	Tumor Protein 53
	TSAs	Traditional serrated adenomas
	TWIST	Twist-related protein 1

U	UICC	Union for International Cancer Control
W	WHO	World Health Organization
	Wnt	Wingless-type /integration site family member
Z	ZEB	Zinc-finger E-box-binding homeobox
	z^{surv}	z-scores for survival
	z^{met}	z-scores for metastasis

REFERENCES

1. Elmasry, M.*, et al.* RBP7 is a clinically prognostic biomarker and linked to tumor invasion and EMT in colon cancer. *J Cancer* **10**, 4883-4891 (2019).

2. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians* **69**, 7-34 (2019).

3. Ferlay, J.*, et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer* **103**, 356-387 (2018).

4. Bray, F.*, et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**, 394-424 (2018).

5. Araghi, M.*, et al.* Global trends in colorectal cancer mortality: projections to the year 2035. *International Journal of Cancer* **144**, 2992-3000 (2019).

6. Chaffer, C.L. & Weinberg, R.A. A perspective on cancer cell metastasis. *Science* **331**, 1559-1564 (2011).

7. Van Cutsem, E. & Oliveira, J. Advanced colorectal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology* **20 Suppl 4**, 61-63 (2009).

8. Chan, A.T. & Giovannucci, E.L. Primary prevention of colorectal cancer. *Gastroenterology* **138**, 2029-2043.e2010 (2010).

9. Song, M., Garrett, W.S. & Chan, A.T. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology* **148**, 1244-1260.e1216 (2015).

10. Vilar, E. & Gruber, S.B. Microsatellite instability in colorectal cancer—the stable evidence. *Nature Reviews Clinical Oncology* **7**, 153 (2010).

11. Brenner, H., Kloor, M. & Pox, C.P. Colorectal cancer. *Lancet* **383**, 1490-1502 (2014).

12. Jasperson, K.W., Tuohy, T.M., Neklason, D.W. & Burt, R.W. Hereditary and familial colon cancer. *Gastroenterology* **138**, 2044-2058 (2010).

13. Lynch, H.T. & de la Chapelle, A. Hereditary colorectal cancer. *N Engl J Med* **348**, 919-932 (2003).

14. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).

15. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-649 (1998).

16. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

17. Fearon, E.R. Molecular genetics of colorectal cancer. *Annu Rev Pathol* **6**, 479-507 (2011).

18. Vogelstein, B.*, et al.* Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).

19. Tomasetti, C., Vogelstein, B. & Parmigiani, G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1999-2004 (2013).

20. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719-724 (2009).

21. Tomasetti, C., Marchionni, L., Nowak, M.A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 118-123 (2015).

22. Grady, W.M. & Markowitz, S.D. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. *Dig Dis Sci* **60**, 762-772 (2015).

23. Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170 (1996).

24. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337 (2012).

25. Vogelstein, B.*, et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-532 (1988).

26. Santini, D.*, et al.* High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice. *Oncologist* **13**, 1270-1275 (2008).

27. Takaku, K.*, et al.* Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. *Cell* **92**, 645-656 (1998).

28. Miyaki, M.*, et al.* Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* **18**, 3098-3103 (1999).

29. Nassif, N.T.*, et al.* PTEN mutations are common in sporadic microsatellite stable colorectal cancer. *Oncogene* **23**, 617-628 (2004).

30. Velho, S.*, et al.* The prevalence of PIK3CA mutations in gastric and colon cancer. *Eur J Cancer* **41**, 1649-1654 (2005).

31. Chalhoub, N. & Baker, S.J. PTEN and the PI3-kinase pathway in cancer. *Annu Rev Pathol* **4**, 127-150 (2009).

32. Lewitowicz, P.*, et al.* Conventional colon adenomas harbor various disturbances in microsatellite stability and contain micro-serrated foci with microsatellite instability. *PLoS One* **12**, e0172381 (2017).

33. Gonzalez, R.S., Washington, K. & Shi, C. Current applications of molecular pathology in colorectal carcinoma. *Applied Cancer Research* **37**, 13 (2017).

34. Jass, J.R. & Smith, M. Sialic acid and epithelial differentiation in colorectal polyps and cancer--a morphological, mucin and lectin histochemical study. *Pathology* **24**, 233-242 (1992).

35. Bettington, M.*, et al.* The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* **62**, 367-386 (2013).

36. Pai, R.K., Bettington, M., Srivastava, A. & Rosty, C. An update on the morphology and molecular pathology of serrated colorectal polyps and associated carcinomas. *Modern Pathology* **32**, 1390-1415 (2019).

37. Davies, H.*, et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).

38. Dhomen, N. & Marais, R. New insight into BRAF mutations in cancer. *Curr Opin Genet Dev* **17**, 31-39 (2007).

39. Weisenberger, D.J.*, et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793 (2006).

40. Noffsinger, A.E. Serrated polyps and colorectal cancer: new pathway to malignancy. *Annu Rev Pathol* **4**, 343-364 (2009).

41. Rex, D.K.*, et al.* Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol* **107**, 1315-1329; quiz 1314, 1330 (2012).

42. The Cancer Genome Atlas, N.*, et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330 (2012).

43. Tsang, A.H.*, et al.* Current and future molecular diagnostics in colorectal cancer and colorectal adenoma. *World J Gastroenterol* **20**, 3847-3857 (2014).

44. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623-627 (1997).

45. Geigl, J.B., Obenauf, A.C., Schwarzbraun, T. & Speicher, M.R. Defining 'chromosomal instability'. *Trends Genet* **24**, 64-69 (2008).

46. Thompson, S.L., Bakhoum, S.F. & Compton, D.A. Mechanisms of chromosomal instability. *Curr Biol* **20**, R285-295 (2010).

47. Pino, M.S. & Chung, D.C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059-2072 (2010).

48. Holland, A.J. & Cleveland, D.W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* **10**, 478-487 (2009).

49. Michor, F., Iwasa, Y., Vogelstein, B., Lengauer, C. & Nowak, M.A. Can chromosomal instability initiate tumorigenesis? *Seminars in cancer biology* **15**, 43-49 (2005).

50. Levine, M.S. & Holland, A.J. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev* **32**, 620-638 (2018).

51. Kane, M.F.*, et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer research* **57**, 808-811 (1997).

52. Cunningham, J.M.*, et al.* Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer research* **58**, 3455-3460 (1998).

53. Herman, J.G.*, et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6870-6875 (1998).

54. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087.e2073 (2010).

55. Hegde, M., Ferber, M., Mao, R., Samowitz, W. & Ganguly, A. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genet Med* **16**, 101-116 (2014).

56. Rampino, N.*, et al.* Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**, 967-969 (1997).

57. Markowitz, S.*, et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336-1338 (1995).

58. Deng, G.*, et al.* BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res* **10**, 191-195 (2004).

59. Toyota, M.*, et al.* CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 8681-8686 (1999).

60. Merlo, A.*, et al.* 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* **1**, 686-692 (1995).

61. Lao, V.V. & Grady, W.M. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 686-700 (2011).

62. Shen, L.*, et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18654-18659 (2007).

63. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

64. Talmadge, J.E. & Fidler, I.J. AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer research* **70**, 5649-5669 (2010).

65. Lambert, A.W., Pattabiraman, D.R. & Weinberg, R.A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670-691 (2017).

66. Anderson, R.L.*, et al.* A framework for the development of effective anti-metastatic agents. *Nature Reviews Clinical Oncology* **16**, 185-204 (2019).

67. Martin, T.A. & Jiang, W.G. Loss of tight junction barrier function and its role in cancer metastasis. *Biochim Biophys Acta* **1788**, 872-891 (2009).

68. Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**, 992-1009 (2011).

69. Lu, P., Takai, K., Weaver, V.M. & Werb, Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb Perspect Biol* **3**(2011).

70. Cristofanilli, M.*, et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351**, 781-791 (2004).

71. Peinado, H.*, et al.* Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* **17**, 302-317 (2017).

72. Valastyan, S. & Weinberg, R.A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275-292 (2011).

73. Riihimaki, M., Hemminki, A., Sundquist, J. & Hemminki, K. Patterns of metastasis in colon and rectal cancer. *Scientific reports* **6**, 29765 (2016).

74. Sheth, K.R. & Clary, B.M. Management of hepatic metastases from colorectal cancer. *Clin Colon Rectal Surg* **18**, 215-223 (2005).

75. Massague, J. & Obenauf, A.C. Metastatic colonization by circulating tumour cells. *Nature* **529**, 298-306 (2016).

76. Lee, J.M., Dedhar, S., Kalluri, R. & Thompson, E.W. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* **172**, 973-981 (2006).

77. Shook, D. & Keller, R. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev* **120**, 1351-1383 (2003).

78. Kalluri, R. & Weinberg, R.A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420-1428 (2009).

79. Nieto, M.A., Huang, Ruby Y.-J., Jackson, Rebecca A. & Thiery, Jean P. EMT: 2016. *Cell* **166**, 21-45 (2016).

80. Yang, J. & Weinberg, R.A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**, 818-829 (2008).

81. Kalluri, R. & Weinberg, R.A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420-1428 (2009).

82. Chaffer, C.L., San Juan, B.P., Lim, E. & Weinberg, R.A. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev* **35**, 645-654 (2016).

83. Mayor, R. & Etienne-Manneville, S. The front and rear of collective cell migration. *Nat Rev Mol Cell Biol* **17**, 97-109 (2016).

84. Brabletz, T. To differentiate or not--routes towards metastasis. *Nat Rev Cancer* **12**, 425-436 (2012).

85. Tsai, J.H. & Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev* **27**, 2192-2206 (2013).

86. Vu, T. & Datta, P.K. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. *Cancers (Basel)* **9**(2017).

87. De Craene, B. & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* **13**, 97-110 (2013).

88. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* **16**, 488-494 (2014).

89. Lu, W. & Kang, Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Dev Cell* **49**, 361-374 (2019).

90. Thiery, J.P., Acloque, H., Huang, R.Y. & Nieto, M.A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-890 (2009).

91. Prieto-Garcia, E., Diaz-Garcia, C.V., Garcia-Ruiz, I. & Agullo-Ortuno, M.T. Epithelial-to-mesenchymal transition in tumor progression. *Med Oncol* **34**, 122 (2017).

92. Canel, M., Serrels, A., Frame, M.C. & Brunton, V.G. E-cadherin-integrin crosstalk in cancer invasion and metastasis. *J Cell Sci* **126**, 393-401 (2013).

93. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196 (2014).

94. Peixoto, P.*, et al.* EMT is associated with an epigenetic signature of ECM remodeling genes. *Cell Death & Disease* **10**, 205 (2019).

95. Gunderson, L.L., Jessup, J.M., Sargent, D.J., Greene, F.L. & Stewart, A.K. Revised TN categorization for colon cancer based on national survival outcomes data. *J Clin Oncol* **28**, 264-271 (2010).

96. Jessup, J.M.*, et al.* Colon and rectum. in *AJCC cancer staging manual* (ed. Amin, M.B.) 251-394 (Springer, Chicago, 2017).

97. Brierley, J.D., Gospodarowicz, M. K. & Wittekind C. (ed.) *Union For International Cancer Control TNM Classification of Malignant Tumors* 8th edn, (Wiley-Blackwell, Hoboken, NJ, 2017).

98. Labianca, R.*, et al.* Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology* **24 Suppl 6**, vi64-72 (2013).

99. Pita-Fernandez, S.*, et al.* Intensive follow-up strategies improve outcomes in nonmetastatic colorectal cancer patients after curative surgery: a systematic review and meta-analysis. *Annals of oncology : official journal of the European Society for Medical Oncology* **26**, 644-656 (2015).

100. Wolmark, N.*, et al.* The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. *J Clin Oncol* **11**, 1879-1887 (1993).

101. Group, Q.C. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *The Lancet* **370**, 2020-2029 (2007).

102. Schippinger, W.*, et al.* A prospective randomised phase III trial of adjuvant chemotherapy with 5-fluorouracil and leucovorin in patients with stage II colon cancer. *British journal of cancer* **97**, 1021-1027 (2007).

103. Compton, C.C.*, et al.* Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* **124**, 979-994 (2000).

104. Bertagnolli, M.M.*, et al.* Microsatellite instability and loss of heterozygosity at chromosomal location 18q: prospective evaluation of biomarkers for stages II and III colon cancer--a study of CALGB 9581 and 89803. *J Clin Oncol* **29**, 3153-3162 (2011).

105. Roth, A.D.*, et al.* Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. *J Natl Cancer Inst* **104**, 1635-1646 (2012).

106. Grinnell, R.S. The grading and prognosis of carcinoma of the colon and rectum. *Ann Surg* **109**, 500-533 (1939).

107. Nagtegaal, I.D., Arends, M.J., D., O.R. & Lam, A.K. Tumors of the colon and rectum. in *WHO classification of tumors of the digestive System* (ed. WHO classification of tumours editorial board) 157-192 (International Agency for Research on Cancer, Lyon, 2019).

108. Dotan, E. & Cohen, S.J. Challenges in the management of stage II colon cancer. *Semin Oncol* **38**, 511-520 (2011).

109. Betge, J.*, et al.* Intramural and extramural vascular invasion in colorectal cancer: prognostic significance and quality of pathology reporting. *Cancer* **118**, 628-638 (2012).

110. Liebig, C.*, et al.* Perineural invasion is an independent predictor of outcome in colorectal cancer. *J Clin Oncol* **27**, 5131-5137 (2009).

111. Benson, A.B., 3rd*, et al.* American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *J Clin Oncol* **22**, 3408-3419 (2004).

112. Schmoll, H.J.*, et al.* ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Annals of oncology : official journal of the European Society for Medical Oncology* **23**, 2479-2516 (2012).

113. Lim, S.B.*, et al.* Prognostic significance of lymphovascular invasion in sporadic colorectal cancer. *Dis Colon Rectum* **53**, 377-384 (2010).

114. Alotaibi, A.M., *et al.* Prognostic and Oncologic Significance of Perineural Invasion in Sporadic Colorectal Cancer. *Ann Surg Oncol* **24**, 1626-1634 (2017).

115. Pawlik, T.M., *et al.* Effect of surgical margin status on survival and site of recurrence after hepatic resection for colorectal metastases. *Ann Surg* **241**, 715-724 (2005).

116. Koelzer, V.H. & Lugli, A. The tumor border configuration of colorectal cancer as a histomorphological prognostic indicator. *Front Oncol* **4**, 29-29 (2014).

117. Rozek, L.S., *et al.* Tumor-Infiltrating Lymphocytes, Crohn's-Like Lymphoid Reaction, and Survival From Colorectal Cancer. *J Natl Cancer Inst* **108**(2016).

118. Sis, B., *et al.* Desmoplasia measured by computer assisted image analysis: an independent prognostic marker in colorectal carcinoma. *Journal of clinical pathology* **58**, 32-38 (2005).

119. Loree, J.M., *et al.* Classifying Colorectal Cancer by Tumor Location Rather than Sidedness Highlights a Continuum in Mutation Profiles and Consensus Molecular Subtypes. *Clin Cancer Res* **24**, 1062-1072 (2018).

120. Thirunavukarasu, P., *et al.* C-stage in colon cancer: implications of carcinoembryonic antigen biomarker in staging, prognosis, and management. *J Natl Cancer Inst* **103**, 689-697 (2011).

121. Gryfe, R., *et al.* Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med* **342**, 69-77 (2000).

122. Popat, S., Hubner, R. & Houlston, R.S. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* **23**, 609-618 (2005).

123. Lanza, G., *et al.* Immunohistochemical test for MLH1 and MSH2 expression predicts clinical outcome in stage II and III colorectal cancer patients. *J Clin Oncol* **24**, 2359-2367 (2006).

124. Merok, M.A., *et al.* Microsatellite instability has a positive prognostic impact on stage II colorectal cancer after complete resection: results from a large, consecutive Norwegian series. *Annals of oncology : official journal of the European Society for Medical Oncology* **24**, 1274-1282 (2013).

125. Hutchins, G.*, et al.* Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol* **29**, 1261-1270 (2011).

126. Yi, M.*, et al.* Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors. *Mol Cancer* **17**, 129-129 (2018).

127. Yokota, T. Are KRAS/BRAF mutations potent prognostic and/or predictive biomarkers in colorectal cancers? *Anticancer Agents Med Chem* **12**, 163-171 (2012).

128. Van Cutsem, E.*, et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* **27**, 1386-1422 (2016).

129. Benson, A.B., 3rd*, et al.* Colon cancer, version 3.2014. *J Natl Compr Canc Netw* **12**, 1028-1059 (2014).

130. Sepulveda, A.R.*, et al.* Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of Clinical Oncology. *J Clin Oncol* **35**, 1453-1486 (2017).

131. Taieb, J.*, et al.* Prognostic Value of BRAF and KRAS Mutations in MSI and MSS Stage III Colon Cancer. *J Natl Cancer Inst* **109**(2017).

132. Horst, D., Kriegl, L., Engel, J., Kirchner, T. & Jung, A. Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. *Cancer Invest* **27**, 844-850 (2009).

133. Graham, R.P.*, et al.* Tumor Budding in Colorectal Carcinoma: Confirmation of Prognostic Significance and Histologic Cutoff in a Population-based Cohort. *Am J Surg Pathol* **39**, 1340-1346 (2015).

134. Guinney, J.*, et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356 (2015).

135. Dienstmann, R.*, et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer* **17**, 79-92 (2017).

136. Phipps, A.I.*, et al.* Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology* **148**, 77-87.e72 (2015).

137. Song, N.*, et al.* Clinical Outcome From Oxaliplatin Treatment in Stage II/III Colon Cancer According to Intrinsic Subtypes: Secondary Analysis of NSABP C-07/NRG Oncology Randomized Clinical Trial. *JAMA Oncol* **2**, 1162-1169 (2016).

138. Bankhead, P.*, et al.* QuPath: Open source software for digital pathology image analysis. *Scientific reports* **7**, 16878 (2017).

139. Ruifrok, A.C. & Johnston, D.A. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* **23**, 291-299 (2001).

140. Liberzon, A. A description of the Molecular Signatures Database (MSigDB) Web site. *Methods Mol Biol* **1150**, 153-160 (2014).

141. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).

142. Anastassiou, D.*, et al.* Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC cancer* **11**, 529 (2011).

143. Liberzon, A.*, et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).

144. Napoli, J.L. Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: Effects on retinoid metabolism, function and related diseases. *Pharmacol Ther* **173**, 19-33 (2017).

145. Das, B.C.*, et al.* Retinoic acid signaling pathways in development and diseases. *Bioorg Med Chem* **22**, 673-683 (2014).

146. Berry, D.C., Levi, L. & Noy, N. Holo-retinol-binding protein and its receptor STRA6 drive oncogenic transformation. *Cancer Res* **74**, 6341-6351 (2014).

147. Karunanithi, S.*, et al.* RBP4-STRA6 Pathway Drives Cancer Stem Cell Maintenance and Mediates High-Fat Diet-Induced Colon Carcinogenesis. *Stem Cell Reports* **9**, 438-450 (2017).

148. Kudryavtseva, A.V.*, et al.* Deregulation of RBP4 and RBP7 Genes in Clear Cell Renal Cell Carcinoma. *Asian J Pharm* **11**, 183-186 (2017).

149. Stintzing, S. Management of colorectal cancer. *F1000Prime Rep* **6**, 108-108 (2014).

150. Kirchner, T. & Brabletz, T. Patterning and nuclear beta-catenin expression in the colonic adenoma-carcinoma sequence. Analogies with embryonic gastrulation. *Am J Pathol* **157**, 1113-1121 (2000).

151. Blaj, C.*, et al.* Oncogenic Effects of High MAPK Activity in Colorectal Cancer Mark Progenitor Cells and Persist Irrespective of RAS Mutations. *Cancer research* **77**, 1763-1774 (2017).

152. Spaderna, S.*, et al.* A transient, EMT-linked loss of basement membranes indicates metastasis and poor survival in colorectal cancer. *Gastroenterology* **131**, 830-840 (2006).

153. Toiyama, Y.*, et al.* Increased expression of Slug and Vimentin as novel predictive biomarkers for lymph node metastasis and poor prognosis in colorectal cancer. *Carcinogenesis* **34**, 2548-2557 (2013).

154. Berry, D.C., Levi, L. & Noy, N. Holo-retinol-binding protein and its receptor STRA6 drive oncogenic transformation. *Cancer research* **74**, 6341-6351 (2014).

155. Erstad, D.J., Tumusiime, G. & Cusack, J.C., Jr. Prognostic and Predictive Biomarkers in Colorectal Cancer: Implications for the Clinical Surgeon. *Ann Surg Oncol* **22**, 3433-3450 (2015).

156. Hajian-Tilaki, K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med* **4**, 627-635 (2013).

157. Miyoshi, N.*, et al.* ANXA9 gene expression in colorectal cancer: A novel marker for prognosis. *Oncol Lett* **8**, 2313-2317 (2014).

158. Yu, S., Bian, H., Gao, X. & Gui, L. Annexin A9 promotes invasion and metastasis of colorectal cancer and predicts poor prognosis. *Int J Mol Med* **41**, 2185-2192 (2018).

159. Boczonadi, V. & Maatta, A. Annexin A9 is a periplakin interacting partner in membrane-targeted cytoskeletal linker protein complexes. *FEBS Lett* **586**, 3090-3096 (2012).

160. Mussunoor, S. & Murray, G.I. The role of annexins in tumour development and progression. *The Journal of pathology* **216**, 131-140 (2008).

161. Chlystun, M., Markoff, A. & Gerke, V. Structural and functional characterisation of the mouse annexin A9 promoter. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1742**, 141-149 (2004).

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. David Horst, for providing me the opportunity to carry out my doctoral research in his lab. Prof. Horst has given me a fantastic support in all stages of my PhD studies, guiding me from the initial laboratory training to the discussion of the final results. In fact, this work could not have been done without his knowledge, research suggestion and direction.

I would also like to express my great thanks to my supervisor Prof. Dr. Andreas Jung for his scientific guidance, valuable advices, and accepting me to complete my doctoral research in his lab after moving of Prof. Horst to Berlin. Furthermore, I would like to thank Prof. Dr. Thomas Kirchner for supporting this work.

Besides, I would like to thank all the former and present members in Prof. Horst and Prof. Jung research groups, for their support and valuable discussions.

Additionally, I am especially thankful to the German Academic Exchange Service (DAAD) and the Egyptian Ministry of Higher Education and Scientific Research (MHESR) for giving me the opportunity to pursue my doctoral degree in Germany. This work was supported by a grant from the German-Egyptian Research Long Term Scholarship Program (GERLS).

Furthermore, I am deeply grateful to my parents, brother and sister for their spiritual supporting, encouragement and endless love.

Last but not least, I would like to express my sincere appreciation to my husband, Dr. El-Amir Sheleel, and our wonderful sons, Eyad and Mohammed. Words cannot express how grateful I am to them. Thank you so much for your love, understanding, patience and support which make the completion of this thesis possible.