

**Identification of novel genetic and prognostic markers in
hereditary and sporadic cancer:
“Two sides of the same coin”**

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Salvatore Piscuoglio
aus Neapel, Italien

Basel 2012

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Primo Schär

Prof. Dr. Gernot Jundt

Prof. Dr. Karl Heinimann

Prof. Dr. Luigi M. Terracciano

Prof. Alfredo Fusco

Basel, den 13-11-2012

Prof. Dr. Jörg Schibler

Dekan

“I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.”

Marie Curie (1867 - 1934)

1. TABLE OF CONTENT

1. TABLE OF CONTENT	iv
2. ABBREVIATIONS.....	vi
3. ABSTRACT	1
4. GENERAL INTRODUCTION.....	4
4.1 Cancer	4
4.2 Cancer Genetics	7
4.2.1 Colorectal Cancer.....	10
4.3 Hereditary cancer syndromes	13
4.3.1 Lynch syndrome (aka HNPCC).....	15
4.3.1.1 Mismatch repair system.....	19
4.3.1.2 Microsatellite instability	21
4.4 Other hereditary colorectal cancer syndromes	23
4.4.1 Familial adenomatous polyposis	23
4.4.1.1 Familial adenomatous polyposis variants	24
4.4.2 Hamartomatous polyposis syndromes	25
4.4.2.1 Juvenile polyposis.....	25
4.4.2.2 Peutz–Jeghers syndrome	26
4.4.2.3 PTEN Hamartoma Tumor Syndrome.....	27
4.5 Diagnostic and prognostic genetic markers in cancer	29
4.6 MicroRNA: biogenesis, functions and targets	34
4.6.1 MicroRNA and cancer	36
4.7 HMG proteins.....	38
4.7.1 HMGA family	38
4.7.2 HMGA expression in normal and neoplastic tissues	40
4.8 Cancer Stem Cell Markers.....	42
4.9 <i>MAGE</i> Gene Family	43
5. AIMS.....	45
6. RESULTS	47
6.1 3'UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancer.....	47

6.2 The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation.	78
6.3 HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma.	146
6.3.1 HMGA1 over-expression represents a poor prognostic index in human breast carcinoma.	155
6.4 Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater.	176
6.5 MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies.	183
7. GENERAL DISCUSSION	196
7.1 3'UTR poly(T/U) tract deletions and altered expression of <i>EWSR1</i> are a hallmark of mismatch repair deficient cancers.	196
7.2 <i>SH2D4A</i> as a novel tumor suppressor gene in CRCs	199
7.3 HMGA proteins as prognostic markers in different tumor entities.	201
7.4 <i>MAGE-A10</i> overexpression in lung, skin and urothelial malignancies.	203
8. APPENDIX	205
8.1 microRNA expression profiling in mismatch repair associated colorectal cancer	205
8.2 Identification of novel recurrent duplication "hot spots" in Lynch syndrome colorectal cancers.	218
9. REFERENCES	233
10. CURRICULUM VITAE	254
11. ACKNOWLEDGMENTS	261

2. ABBREVIATIONS

AC	Amsterdam Criteria
AFAP	Attenuated Familial Adenomatous Polyposis
APC	Adenomatous Polyposis Coli
BRRS	Bannayan-Riley-Ruvalcaba Syndrome
BG	Bethesda Guidelines
CHRPE	Congenital hypertrophy of the retinal pigment epithelium
CS	Cowden Syndrome
CRC	Colorectal Cancer
FAP	Familial Adenomatous Polyposis
HE	Hemotoylin and Eosin Stain
hMLH	Human MutL Homolog
hMSH	Human MutS Homolog
HMPS	Hereditary Mixed Polyposis Syndrome
hPMS	Human Post Meiotic Segregation
HNPCC	Hereditary Non Polyposis Colorectal Cancer
IARC	International Agency for Research on Cancer
IDLs	Insertion or deletion loops
IHC	Immunohistochemistry
<i>KRAS</i>	Kirsten rat sarcoma viral oncogene homolog
LOH	Loss of Heterozygosity
MMR	Mismatch Repair
MSI	Microsatellite Instability
MYH	mutY homolog
PJS	Peutz-Jeghers Syndrome
PCR	Polymerase Chain Reaction
PTEN	Phosphatase and Tensin Homologue
Rb	Retinoblastoma gene
SD	Standard Deviation
TGF β RII	Tumour Growth Factor β Receptor II
TP53	Tumor protein 53
JPS	Juvenile Polyposis Syndrome

3. ABSTRACT

This thesis has focused on the discovery and characterization of novel diagnostic and prognostic markers in various cancer entities, with a special emphasis on colorectal cancer (CRC).

In Switzerland the incidence of colorectal cancer ranks third in males and second in females, with about 4000 new patients diagnosed each year. Incidence trends over the last decades have remained constant in both sexes, whereas mortality rates have been decreasing. Decreasing mortality is thought to be related to improved treatment during the past years as well as generalisation of colorectal cancer screening in the Swiss population. About 80% of colorectal cancers are thought to have occurred by chance (sporadic) with the remainder displaying either familial aggregation (about 15%) or mendelian inheritance (about 5%).

In the first part of this work we identify and characterize a novel target gene locus for microsatellite instability (MSI) consisting of a mononucleotide (T/U)₁₆ tract, EWS16T, located in the 3' UTR of the *Ewing sarcoma break point region 1 (EWSR1)* gene in 319 patients with hereditary and sporadic CRC. We show that the EWS16T locus discriminates MMR proficient from deficient cancers with high diagnostic sensitivity (100%) and specificity (100%). It could thus substantially improve and facilitate MSI analysis in routine daily practice. In addition, biochemical analyses indicate that EWS16T contractions alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in *EWSR1* pre-mRNAs and result in decreased mRNA as well as EWS protein expression. Our findings thus directly implicate the RNA-/DNA-binding Ewing sarcoma protein in MSI-associated colorectal tumorigenesis.

In the second part we characterize a new tumour suppressor gene designated *SH2D4A* located on the short arm of chromosome 8. We demonstrate that *SH2D4A* physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signaling, *SH2D4A* recruits the serine/threonine phosphatase PP1 β to the receptor complex and represses activated STAT3 via dephosphorylation. *SH2D4A* expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1

and Jun B. In addition we show that *SH2D4A* expression is partially lost in human colorectal cancers as a result of chromosomal instability, mutations and epigenetic changes. Finally, diminished *SH2D4A* protein expression was found to correlate with advanced disease stages and was associated with poor prognosis.

In the third part we investigate HMGA1/HMGA2 protein expression in 210 and 1202 patients with pancreatic and breast cancers, respectively. HMGA1 and HMGA2 over-expression was found in a significant number of breast and pancreatic carcinoma samples, and its over-expression positively correlated with grade and stage of the disease. Conversely, no HMGA1 and HMGA2 expression was observed in cancer-free breast and pancreas tissues. Taken together, our findings show that high expression levels of HMGA1 and HMGA2 are related to an unfavorable histological type and a poor prognosis in both, pancreatic and breast cancer. Moreover, these findings further support the notion that these proteins represent appropriate targets for the therapy of human cancer, as suggested by numerous *in vitro* and *in vivo* studies.

In the fourth part of the thesis we evaluate the potential role of the cancer stem cell (CSC) proteins EpCAM, CD44s, CD166 and CD133 in tumors of the ampulla of Vater. CSC expression was determined in 175 carcinoma, 111 adenoma and 152 cancer free-mucosa specimens arranged on a tissue microarray format. The expression of all evaluated marker proteins differed significantly between cancer-free mucosa, adenoma and carcinoma samples. EpCAM expression was significantly correlated with better patient survival. In contrast, increased expression of CD44s, CD166 and CD133 from normal mucosa samples to adenoma and carcinoma was linked to tumor progression but no statistically significant correlation with survival observed. Our findings therefore indicate that in ampullary carcinomas loss of EpCAM expression may be associated with a more aggressive tumor phenotype.

In the fifth part we develop a specific monoclonal antibody for the highly immunogenic member of the MAGE-A family of cancer/testis tumor-associated antigens (C/T TAAs). The antibody was used to stain a multi-tumor tissue microarray comprising more than 2,500 paraffin-embedded specimens of different histological origin. C/T TAA appears to be expressed in a high percentage (>50%) of cancer cells from different tumor types such as lung,

skin, gynecological, stomach and gall bladder cancers. The future characterization of MAGE-A10-specific antibodies might set the stage for the development of targeted active immunotherapy by clarifying potential indications and by allowing the selection of patients eligible for treatment and monitoring of its effectiveness.

4. GENERAL INTRODUCTION

4.1 Cancer

Cancer is a leading cause of death worldwide, according to the International Agency for Research on Cancer (IARC) data, there were 12.7 million new cancer cases in 2008 worldwide, of which 5.6 million occurred in economically developed countries and 7.1 million in economically developing countries. The corresponding estimates for total cancer deaths in 2008 were 7.6 million (about 21,000 cancer deaths a day), 2.8 million in economically developed countries and 4.8 million in economically developing countries [1].

Cancer arises from a loss of normal growth control. In normal healthy tissue, the ratio between cell growth and cell death are in balance. In cancer, this balance is disrupted. This disturbance can result from uncontrolled cell growth or loss of a cell's ability to undergo apoptosis "cell suicide". This is also referred to as programmed cell death and the mechanism by which old or damaged cells can be removed. Cancer being a genetic disease, the genetic material (DNA) of a cell can get damaged, acquiring mutations that affect normal cell growth, division and apoptosis. When this occurs, cells may start to grow uncontrollably and eventually form a mass of tissue called a tumor [2]. Cancer is a disease in which cells display uncontrolled growth, invade and destroy adjacent tissues, and often metastasize to other regions of the body via the lymphatic system or through the bloodstream [2]. These properties discriminate benign from malignant tumors. There are more than 100 different types of cancer. The name of the tumors often comes from the organs or the type of cells affected and can be broadly grouped into the following categories [2]:

- a. **Carcinoma** - a malignant epithelial neoplasm that tends to invade surrounding tissue and to metastasize to distant regions of the body. Carcinomas develop most frequently in the skin, large intestine, lungs, stomach, prostate, cervix, or breast. The tumor is firm, irregular, and nodular, with a well-defined border.

- b. **Sarcoma** - a tumor, often highly malignant, composed of cells derived from connective tissue such as bone, cartilage, muscle, blood vessel, or lymphoid tissue.

Sarcomas usually develop rapidly and metastasize through the lymph channels. Different types are named for the specific tissue they affect: FIBROSARCOMA in fibrous connective tissue; LYMPHOSARCOMA in lymphoid tissues; OSTEOSARCOMA in bone; CHONDROSARCOMA in cartilage; RHABDOMYOSARCOMA in muscle; and LIPOSARCOMA in fat cells.

- c. **Leukemia** - a progressive, malignant neoplasm of the blood-forming organs, marked by diffuse replacement of the bone marrow leukocytes and their precursors in the blood and bone marrow, based on cell type can be divided into myeloid (acute and chronic) and lymphoid (acute and chronic). It is accompanied by a reduced number of erythrocytes and blood platelets, resulting in anemia and increased susceptibility to infection and hemorrhage.
- d. **Myeloma** - an osteolytic neoplasm consisting of a profusion of cells typical of the bone marrow that may develop in many sites and cause extensive destruction of the bone. The tumor occurs most frequently in the ribs, vertebrae, pelvic bones, and flat bones of the skull
- e. **Lymphoma** - Cancer that begins in cells of the immune system that affects lymph cells and tissues, including certain white blood cells (T and B cells), lymph nodes, bone marrow, and the spleen. There are two basic categories of lymphomas. One kind is Hodgkin lymphoma, which is marked by the presence of a type of cell called the Reed-Sternberg cell. The other category is non-Hodgkin lymphomas, which develop from lymphocytes — a type of white blood cell. Non-Hodgkin lymphomas can be further divided into cancers that have an indolent (slow-growing) course and those that have an aggressive (fast-growing) course.
- f. **Central nervous system cancers** - a neoplasm of the brain or spinal cord that characteristically does not spread beyond the cerebrospinal axis, although it may be highly invasive locally and have widespread effects on body functions. Intracranial neoplasms are about four times more common than those arising in the spinal cord. From 20% to 40% of brain tumors are

metastatic lesions from primary cancer elsewhere, such as in the breast, lung, GI tract, kidney, or a site of melanoma [2].

Cancer can be treated by chemotherapy, surgery, radiation therapy, immunotherapy, monoclonal antibody therapy or other methods (experimental cancer treatments are also under development). The choice of therapy depends on the cancer type, the degree and location of the cancer and the patient's general medical condition [2].

4.2 Cancer Genetics

Cancers arise as consequence of accumulating multiple mutations in the genome. Mutations in genes controlling cell growth are responsible for many major human cancers. Two broad classes of genes have been defined: proto-oncogenes (Table 1) and tumor suppressor genes like *Rb*, *VHL*, *APC*, *PTEN*, *TP53*, *CD95*, *ST5*, *YPEL3*, *ST7*, and *ST14* play a key role in cancer initiation and progression. These genes encode many kinds of proteins that help control cell growth and proliferation; mutations in these genes can thus contribute to the development of cancer. Under normal conditions the activities of these two gene classes are optimally balanced. In general, proto-oncogenes allow cell growth whereas tumor suppressor genes inhibit it.

Proto-oncogenes are genes that encode proteins, which can be mutated into a cancer-promoting oncogenes, either by changing the protein-coding segment or by altering its expression. This activation results in a gain of function and drives cell multiplication and thus proliferation.

In contrast tumor suppressor genes encode proteins directly or indirectly inhibiting initiation and/or progression through the cell cycle and other tumor related pathways (e.g. WNT) in which a loss-of-function mutation is oncogenic. A germline mutation in a tumor suppressor gene (e.g. *RB*, *APC*, and *BRCA1*) greatly increases the risk for developing certain types of (hereditary) cancer.

There are also some genetic alterations in DNA that lead to cancer development such as chromosomal translocations and gene promoter (hyper- and hypo-) methylation.

Translocations generate novel chromosomes. In a translocation, a segment from one chromosome is transferred to a nonhomologous chromosome or to a new site on the same chromosome. Translocation events may disrupt gene loci, cause (micro)deletions or result in genes fusion. Some examples are: Ewing's sarcoma, which is the result of a translocation between chromosomes 11 and 22, which fuses the *EWSR1* gene of chromosome 22 to the *FLI1* gene of chromosome 11; chronic myelogenous leukemia (CML) which is the result of a reciprocal translocation between chromosome 9 and 22 (Philadelphia chromosome).

DNA methylation is an important regulator of gene transcription, and its role in

carcinogenesis has been a topic of considerable interest in the last few years. Alterations in DNA methylation are common in a variety of tumors as well as in development. Of all epigenetic modifications, hypermethylation, which represses transcription of the promoter regions of tumor suppressor genes leading to gene silencing, has been most extensively studied [3].

DNA methylation mainly consists of a covalent chemical modification, resulting in the addition of a methyl (CH₃) group at the carbon 5 position of the cytosine ring. Even though most cytosine methylation occurs in the sequence context 5'CG3' (also called the CpG dinucleotide), some involves CpA and CpT dinucleotides [4, 5]. Numerous genes have been found to undergo hypermethylation in cancer like genes involved in cell cycle regulation (*p16^{INK4a}*, *p15^{INK4a}*, *Rb*, *p^{14ARF}*) genes associated with DNA repair (*BRCA1*, *MLH1*, *MGMT*), apoptosis (*DAPK*, *TMS1*), drug resistance, detoxification, differentiation, angiogenesis, and metastasis [5]. Although certain genes such as *RASSF1A* and *p16* are commonly methylated in a variety of cancers, other genes are methylated in specific cancers. One example is the *GSTP1* gene, which is hypermethylated in more than 90% of prostate cancers but is largely unmethylated in acute myeloid leukemia [5]. DNA hypomethylation, on the other hand, is observed in a wide variety of malignancies [6, 7]. It is common in solid tumors such as metastatic hepatocellular cancer and prostate tumors [8, 9].

Table 1: Oncogenes commonly altered in human cancer.

Common Oncogenes Altered in Human Cancers			
Oncogene	Function	Alteration in Cancer	Neoplasm
AKT1	Serine/threonine kinase	Amplification	Gastric carcinoma
AKT2	Serine/threonine kinase	Amplification	Ovarian, breast, pancreas cancer
BRAF	Serine/threonine kinase	Point mutation	Melanoma, lung, colorectal cancer
CTNNB1	Signal transduction	Point mutation	Colon, prostate, melanoma, skin, others
FOS	Transcription factor	Overexpression	Osteosarcoma
ERBB2	Receptor tyrosine kinase	Point mutation, amplification	Breast, ovary, stomach, neuroblastoma
JUN	Transcription factor	Overexpression	Lung cancer
MET	Receptor tyrosine kinase	Point mutation, rearrangement	Osteocarcinoma, kidney, glioma
MYB	Transcription factor	Amplification	AML, CML, colon cancer, melanoma
C-MYC	Transcription factor	Amplification	Breast, colon, gastric, lung
L-MYC	Transcription factor	Amplification	Lung carcinoma, bladder
N-MYC	Transcription factor	Amplification	Neuroblastoma, lung cancer
HRAS	GTPase	Point mutation	Colon, lung, pancreas
KRAS	GTPase	Point mutation	Melanoma, colorectal cancer, AML
NRAS	GTPase	Point mutation	Various carcinomas, melanoma
REL	Transcription factor	Rearrangement, amplification	Lymphoma
WNT1	Growth factor	Point mutation	Retinoblastoma

Compilation of oncogenes commonly altered in human cancer, taken from [2]:
 AML, acute myelogenous leukemia; CML, chronic myelogenous leukemia.

4.2.1 Colorectal Cancer

The colon is part of the digestive system, which is a series of bodily organs beginning at the mouth and ending with the anus, and it is responsible for the final stages of the digestive process (Figure 1).

The colon's function is threefold: to absorb the remaining water and electrolytes from indigestible food matter; to accept and stores food remains that were not digested in the small intestine; and to eliminate solid waste (feces) from the body. The rectum is the segment from end of the colon to the anus. Together, they form a long, muscular tube called the large intestine (also known as the large bowel). Tumors of the colon and rectum are growths arising from the inner wall of the large intestine.

Precursor lesions of the large intestine are called polyps and has been defined as a benign neoplastic growth with variable malignant potential, representing proliferation of epithelial tissue in the lumen of the sigmoid colon, rectum, or stomach, polyps do not invade nearby tissue or spread to other parts of the body. The majority of colorectal cancer (CRC) arises from malignant transformation of an adenomatous polyp (Figure 2).

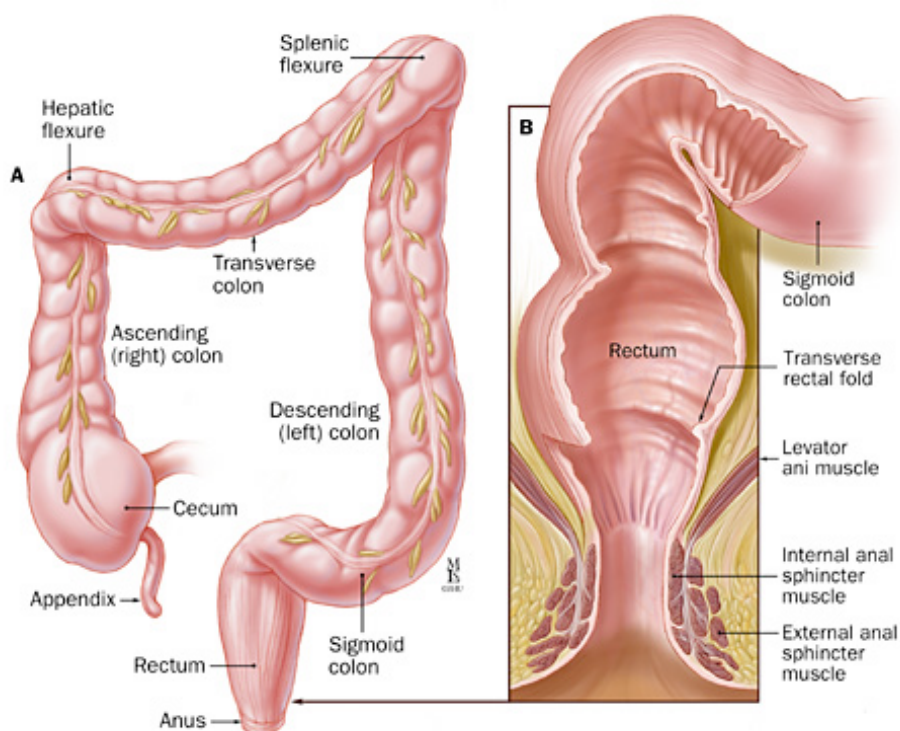
CRC is the fourth most common cancer in men and the third most common cancer in women worldwide [10, 11]. The American Cancer Society's most recent estimates for the number of colorectal cancer cases in the United States for 2011 are: 101,700 new cases of colon cancer and 39,510 of rectal cancers with 49,380 deaths.

Risk factors according to the American Cancer Society for the development of CRC are:

- a. **Age:** peak incidence in the 6Th/7Th decade.
- b. **Nutrition:** a diet that is high in red meat, processed meats and not enough fruits and vegetables consumption is a colon cancer risk.
- c. **A family history of colorectal cancer:** A family history of colorectal cancer or adenomatous polyps increases the risk of colorectal cancer. That risk is even higher if the close relative was diagnosed at a young age or if more than one 1Th degree family member had colon cancer.

- d. **Member of certain racial or ethnic groups:** African Americans get colon cancer more often than other racial groups in the U.S. and are nearly twice as likely to die from it.
- e. **Inherited conditions such as familial adenomatous polyposis,** which causes the development of 100-1000 of polyps in the colon, also raises the risk of colorectal cancer.
- f. **Inflammatory bowel disease:** ulcerative colitis and Crohn's disease are a bowel disease characterized by inflammation with ulcer formation in the lining of the colon (large intestine). People with either of these diseases can develop expansion of immature cells, with a corresponding decrease in the number and location of mature cells, this is called dysplasia, which may turn into cancer [2].

Figure 1: Anatomy of the colon.



Depiction of colorectal anatomy, taken from [12].
A: anatomy of the colon; **B:** anatomy of the rectum.

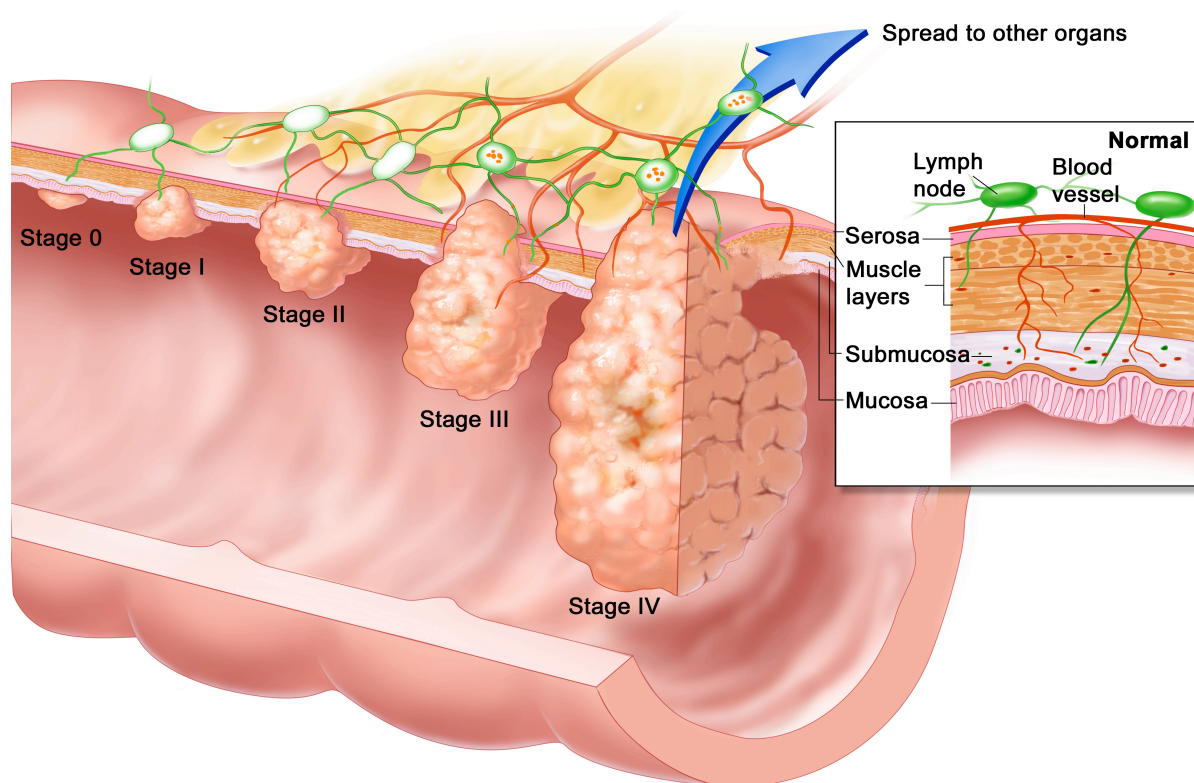
CRC progresses through a series of clinical and histopathological stages ranging from single crypt lesions through small benign tumors (adenomatous

polyps) to malignant cancers (carcinomas) (Figure 2). Stages are usually defined by TNM classification, where **T** describes the size of the tumor and whether it has invaded nearby tissue, **N** describes regional lymph nodes that are involved, **M** describes distant metastasis (spread of cancer from one body part to another). The depth of tumor invasion defines the T stage and increases from T1 (invasion of the submucosa) to T4 (invasion into the serosa or adjacent structures) [13]. Another grading system is Dukes classification that considers the arrangement of the cells rather than the percentage of the differentiated cells. The initial Dukes approach has evolved into the three-grade system. Grade 1 is the most differentiated, with well-formed tubules and the least nuclear polymorphisms and mitoses. Grade 3 is the least differentiated, with only occasional glandular structures, pleomorphic cells and a high incidence of mitoses. Grade 2 is intermediate between Grades 1 and 3 [14].

The development of colorectal cancer is a multistep process that involves an accumulation of mutations in tumor suppressor genes and oncogenes. The model of colorectal tumorigenesis includes several genetic changes that are required for cancer initiation and progression [15]. The earliest trigger genetic event is the inactivation of the *APC* (adenomatous polyposis coli) gene (WNT pathway). Mutations in other tumor suppressor genes like *SMAD4* and *TP53* and oncogenes like *KRAS* and likely several other genes/pathways accompany transitions (in the pathology) of the lesions and drive tumors towards malignancy and metastasis [16]. Alongside with gene mutations, deregulated expression of oncogenes and/or tumor suppressor genes can also occur following epigenetic modifications of their promoters. In chapter 6.2 we describe *SH2D4A* gene as novel tumour suppressor in colorectal cancer, that physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signalling, Shoca-2 recruits the serine/threonine phosphatase PP1 β to the receptor complex and represses activated STAT3 via dephosphorylation. Shoca-2 expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1 and Jun B. Shoca-2 expression may be lost in human colorectal cancers as a result of chromosomal instability, mutations and epigenetic changes and

diminished Shoca-2 protein expression correlates with advanced disease stages and is associated with poor prognosis.

Figure 2: Stages of colon cancer.



The diagram, taken from [2], above illustrates the progression of polyp to a cancer and the cancer's subsequent progression if left untreated.

4.3 Hereditary cancer syndromes

The majority of cancers occur without any predisposition, although genetic factors are believed to have an impact on the susceptibility of an individual to develop cancer [17]. Approximately 5-10% of all cancers are hereditary, which means that germline mutations in specific genes that are known to be related to cancer development can be passed on to offspring. Individuals who inherit one of these gene changes will have a higher likelihood of developing cancer within their lifetime. Inherited cancer syndromes with their genetic causes are listed in Table 2.

Table 2: Examples of hereditary cancer syndromes.

Syndrome	Gene(s)	Location(s)	Tumor types
Breast/Ovarian	BRCA1	17q21	Breast, ovarian
Breast/Ovarian	BRCA2	13q12.3	Breast, ovarian, prostate
Cowden syndrome	PTEN	10q23.3	Pituitary, testicle, thyroid, breast
Familial Adenomatous Polyposis	APC	5q21	GI tract, thyroid
Lynch syndrome	MLH1, MSH2, MSH6, PMS2	2p22-21, 3p21, 2p16, 2q31-33, 7p22	GI tract, ovary, endometrium
Juvenile polyposis coli	SMAD4, BMP1A	18q21.1, 10q23.2	GI tract
Li-Fraumeni	TP53	17p13	Breast, soft tissue, brain, leukaemia
Neurofibromatosis 1	NF1	17q11	Nervous system, skin, muscle, leukaemia
Neurofibromatosis 2	NF2	22q12	Schwann cells, spine skin
Peutz-Jeghers	STK1	19p13.3	Gastrointestinal tract

In hereditary cancer predisposing syndromes, the critical gene is mutated only on one allele but not on the second, “wild type” one. According to Knudson’s widely accepted “two hit theory” which was based on the epidemiological studies of retinoblastoma gene (Rb) [18], both of the alleles of a tumor suppressor gene need to be inactivated before tumorigenesis ensues. Thus, the germline mutation alone is not sufficient to give rise to tumor formation, but the functional wild type allele needs to be inactivated before neoplastic reactions may occur [19].

There are several known and proposed mechanisms, which can cause the inactivation of the wild type allele. It can occur for example by loss of heterozygosity (LOH), somatic deletion [20], or epigenetically by promoter methylation [21]. However, to explain the sometimes even 100% penetrance of hereditary cancer syndromes resulting from a heterozygous mutation in a tumor suppressor gene, it is nowadays believed that also “recessive” mutation have an effect on the tissue phenotype and thus result in a clonal expansion of mutated cells [22]. Dominantly functioning oncogenes can accelerate tumorigenesis also when heterozygous.

4.3.1 Lynch syndrome (aka HNPCC)

Lynch syndrome (LS) represents the most common, autosomal dominantly

inherited cancer predisposition worldwide with incomplete penetrance, accounting for approximately 1-6% of all colorectal malignancies. It is characterized by an increased lifetime risk for colon cancer and cancers of the endometrium, ovary, stomach, small intestine, hepatobiliary tract, urinary tract, brain, and skin [23-25].

It is caused by a germline mutation in a mismatch repair gene (*MLH1*, *MSH2*, *PMS2*, and *MSH6*), *MSH2* on chromosome 2p, and *MLH1* on chromosome 3p account for the majority of genetically defined cases. 50-60% of families with the clinical diagnosis of Lynch syndrome are found to have mutations in these genes; the mutation prevalence depends on features of the family history [26, 27].

The number of polyps is modest without endoscopical differences compared to sporadic polyps but with an early onset (44 years), bigger size and frequency, and with more villous and dysplastic features as well as a more rapid progression to cancer [28, 29]. Without intervention, patients with Lynch syndrome have an 80% lifetime risk of colorectal cancer [30] however the risk of cancer-related mortality remains considerable. Nevertheless, patients with Lynch syndrome-related colorectal cancer appear to have better survival rates when compared to patients with sporadic colorectal cancer [31]. Tumors in LS patients characteristically exhibit microsatellite instability (MSI). The diagnostic settings for LS diagnosis are based on Amsterdam criteria (constantly update) to identify the affected family and the Bethesda guidelines to define the tumors (Table 2) [32-37]. Essentially this guideline specifies that at least three relatives be affected with colorectal cancer, endometrial cancer (female patients), possibly other LS-related malignancies and MSI-H. Genetic testing is improving the approach to patients at risk for LS. The American Gastroenterological Association recommends a testing strategy beginning with MSI testing on tumor tissue from individuals satisfying the revised Bethesda criteria. In persons with MSI-high tumors, germline analysis of hMSH2, hMSH1 and hMSH6 should be assessed. For families in which tumor tissue is not available, germline testing is recommended if any of the first three Bethesda criteria are met (Table 2) [38]. Failure to detect a clearly pathogenic mutation in the proband doesn't exclude Lynch syndrome.

Immunohistochemistry (IHC) testing for the expression of mismatch related

proteins in tumor tissue is a complementary screening method [39], The MMR gene products work in heterodimers: MSH2 with MSH6 or MSH3 protein, and MLH1 with PMS2 or PMS1 protein. A germline mutation in *MSH2* typically results in loss of expression of both proteins MSH2 and MSH6 and a germline mutation in *MLH1* in loss of expression of the proteins MLH1 and PMS2. Germline mutations in *MSH6* and *PMS2* typically however do not result in loss of MSH2 or MLH1 expression because these proteins are still present in other pairings and therefore remain stable expressed [40].

For patients with a definitive or suspected diagnosis of LS, interval surveillance is recommended encompassing: full colonoscopy every one to two years beginning between the ages of 20 and 25 and endometrial and ovarian cancer screening beginning between the of 25 and 35 (for female) [41, 42].

In appendix 8.2 analyzing 12 samples from 6 LS patients (CRC and matched cancer free mucosa) with Affymetrix Whole Genome 2.7M chip and validating the results in a large cohort of LS-related CRCs (50 patients), we describe 3 new hotspot frequently deleted or duplicated in LS related colorectal cancers.

Table 2: Clinical criteria for LS.

Amsterdam	<ol style="list-style-type: none"> 1. Three or more relatives with colorectal cancer, one of whom is a first-degree relative of the other two; 2. Colorectal cancer involving at least two generations; 3. One or more colorectal cancers diagnosed at age <50 years
Amsterdam II	<ol style="list-style-type: none"> 1. Three or more relatives with a HNPCC-associated cancer (colorectal, endometrial, small bowel, ureter, renal pelvis), one of whom is a first-degree relative of the other two; 2. Colorectal cancer involving at least two generations; 3. One or more colorectal cancers diagnosed at age <50 years
Modified Amsterdam	<ol style="list-style-type: none"> 1. Very small families which cannot be further expanded can be considered as HNPCC with only two colorectal cancers in first-degree relatives; colorectal cancer must involve at least two generations, and at least one colorectal cancer must be diagnosed at age <55 years 2. In families with two first-degree relatives affected by colorectal cancer, the presence of a third relative with an unusual early-onset neoplasm or endometrial cancer is sufficient
Young age at onset	Proband diagnosed at age <40 years without a family history fulfilling Amsterdam or Modified Amsterdam criteria
HNPCC-variant	Family history suggestive of HNPCC, but not fulfilling Amsterdam, Modified Amsterdam, or young age at onset criteria
Bethesda	<ol style="list-style-type: none"> 1. Amsterdam criteria 2. Two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers 3. Colorectal cancer and a first-degree relative with colorectal cancer and /or a HNPCC-related extracolonic cancer and/or a colorectal adenoma; one of the cancers diagnosed at age <45 years, and the adenoma diagnosed age <40 years 4. Colorectal cancer or endometrial cancer diagnosed at age <45 years 5. Right-sided colorectal cancer with an undifferentiated pattern on histology diagnosed at age < 45 years 6. Signet-ring cell-type colorectal cancer diagnosed at age <45 years 7. Adenomas diagnosed at age <40 years
Revised Bethesda	<ol style="list-style-type: none"> 1. Colorectal cancer diagnosed at age <50 years 2. Synchronous or metachronous colorectal or other HNPCC-associated tumors regardless of age 3. Colorectal cancer diagnosed at age <60 years with histologic findings of infiltrating lymphocytes, Crohn's-like lymphocytic reaction, mucinous/signet ring differentiation or medullary growth pattern. 4. Colorectal cancer in ≥ 1 first-degree relative(s) with an HNPCC-related tumor, with one of the cancers being diagnosed at age <50 years 5. Colorectal cancer diagnosed in ≥ 2 first- or second-degree relatives with HNPCC-related tumors, regardless of age

Table taken from [43]

4.3.1.1 Mismatch repair system

The DNA Mismatch repair system has an important role in maintaining genomic stability during DNA replication by correcting mismatches and small insertion or deletion loops (IDLs) introduced through errors made by DNA polymerases. Components of the MMR system MutS, MutL, MutH and Uvr were identified in *Escherichia coli* through the genetic studies of mutants [44, 45] and the system has been well conserved during the evolutionary process in all eukaryotic organisms (yeast, mouse and human), [46, 47].

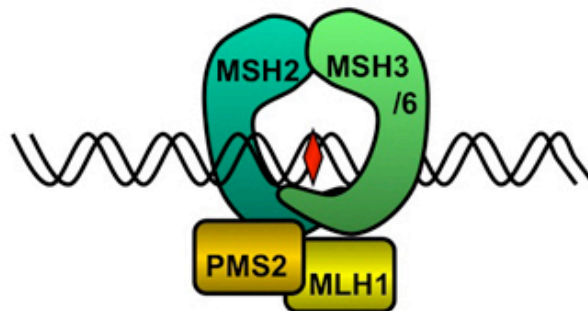
DNA mismatch repair (MMR) can be divided in three phases: initiation, excision and resynthesis (Figure 3). The three MutS homologs, MSH2, MSH3 and MSH6 are involved in the initiation of MMR. The MutS homologs forming heterodimers, recognize DNA damage; the MSH2 and MSH6 dimer (the MutS α complex) recognizes mismatches base-base and single base loops while the MSH2 and MSH3 dimer (hMutS β complex) identifies deletion/insertion loops of more than one base [48-50].

After recognition of the DNA damage the heterodimer MutL α , formed by MLH1 and PMS2, binds MutS α and mediates the interaction between the MutS proteins and enzymes involved in long-patch excision in postreplication mismatch repair [51].

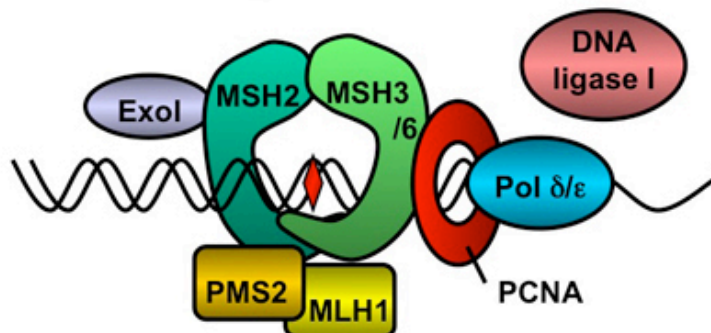
Other proteins are involved in the excision of the damaged strand and resynthesis steps of MMR that are recruited subsequently. Proteins known to be involved are proliferating cell nuclear antigen (PCNA) that is known to work as a processivity factor for replicative polymerases [52], with interactions with MSH2 and MLH1 [53], and with MSH6 [54], as well as exonuclease ExoI, and DNA polymerase δ I [55, 56].

Figure 3: The mismatch repair system (MMR).

A. Initiation



B. Excision & resynthesis



The mismatch repair system (MMR), taken from Helleman *et al.* [57].

A. Initiation of MMR by recognizing the DNA damage by the MutS α or β complex and recruiting the MutL α complex. **B.** Excision of the damaged strand and resynthesis in which exonuclease ExoI, proliferating cell nuclear antigen (PCNA), DNA polymerase δ or ϵ and DNA helicase I are suggested to play a role.

4.3.1.2 Microsatellite instability

The inactivation of the MMR system leads to widespread genomic instability, because the DNA-replication errors induced DNA polymerase slippage during replication of long repetitive DNA sequences. As a consequence, these tumors (MMR deficient) exhibit a specific phenotype called microsatellite instability (MSI), affecting mono-, di-, tri- and tetra nucleotide repetitive sequences [58, 59].

MSI in tumors can be initiated by genetic or epigenetic inactivation of MMR genes. Mouse knockout models have demonstrated that $MLH1^{-/-}$, $PMS2^{-/-}$, $MSH2^{-/-}$ and $MSH3^{-/-}$, exhibit a high frequency of MSI in tumors while $MSH6^{-/-}$ and $PMS1^{-/-}$ mice had a low frequency [60].

Five markers (two mononucleotide repeats: BAT25 and BAT26 and three dinucleotide repeats: D2S123, D5S346, and D17S250) have been recommended by the National Cancer Institute to screen for MSI in LS-related CRCs (often called Bethesda markers) [61]. Tumors with instability at two or more of these markers were defined as MSI-High (MSI-H), while those with instability at one repeat or display no instability were defined as MSI-Low (MSI-L) and those without as microsatellite stable (MSS) tumors, respectively [62]. They occur usually in the Lynch syndrome or sporadically in as many as 10–15% of colorectal, gastric, and endometrial (20-30%) carcinomas [63-65].

Repetitive segments of DNA (dinucleotide, trinucleotide, tetranucleotide, pentanucleotide repeats), scattered throughout the genome in non-coding regions between genes or within genes (introns), often used as markers for linkage analysis because of the naturally occurring high variability in repeat number between individuals. These regions are inherently unstable and susceptible to mutation, Sporadically these repetitive sequences are located in the coding region of genes. The $TGF\beta RII$, BAX , $IGF1R$, $MSH3$, and $MSH6$ genes were the first described genes for instability in MSI-H cancers [66-69]. More recently, additional genes with similar mutations were described encompassing $ACTR11$, $AIM2$, $APAF-1$, $AXIN-2$, $BCL-10$, BLM , $Caspase-5$, $CDX-2$, $CHK-1$, FAS , $GRB-14$, cell cycle protein $hG4-1$, $KIAA0977$, $MBD-4$, $hMLH3$, $NADH$ ubiquinone oxidoreductase, OGT , $PTEN$, $RAD-50$, $RHAMM$, RIZ , $SEC63$, $SLC23AT$, $TCF-4$, and $WISP-3$. MSI-H CRCs [59, 62, 70-84],

whether hereditary or sporadic (through somatic MLH1 promoter hypermethylation), are generally associated with a more favorable prognosis when compared to MSI-L and MSS ones [85]. Recent genome-wide gene expression data on (sporadic) MSS and MSI-H CRCs further demonstrate that tumor development in microsatellite unstable, MMR deficient cancers follows distinct pathogenetic alleys.

In chapter 6.1 we identified a mononucleotide repeat tract (EWS16T) within the 3'UTR of the *EWSR1* gene that is consistently mutated in cancers with microsatellite instability and thus represents a novel target gene locus in mismatch repair-deficient tumors. Our data indicate that contractions at this locus promote SFPQ-mediated distal poly(A) site usage in *EWSR1* pre-mRNAs resulting in decreased EWS expression that is accompanied by altered subcellular localization of the protein.

4.4 Other hereditary colorectal cancer syndromes

4.4.1 Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is an autosomal dominantly inherited disorder with 100% penetrance by age 40 years characterized by the presence of more than hundreds/thousands of benign polyps (adenomas) some of which, inevitably progress to colorectal cancer in the third and fourth decade of life [86, 87]. In addition an extracolonic cancer spectrum may occur, including polyps of the upper gastrointestinal tract, desmoids, osteomas, fibromas, and congenital hypertrophy of the retinal pigment epithelium (CHRPE) [88]. It is thought to be responsible for about 0,5-1% of colorectal cancers diagnosed each year with an estimated prevalence of approximately 1/10,000 [89]. It is caused by germline mutations in the *APC* (*adenomatous polyposis coli*) gene located on the long arm of chromosome 5 (5q21). The majority of mutations (>90%) are nonsense or frameshift mutations that lead to premature stop codons. The resulting protein is truncated and apparently nonfunctional. The bi-allelic inactivation (1st germline and 2nd somatic hit) of the *APC* have a direct role to the disruption of Wnt signalling pathway. Indeed, it is well understood that *APC* acts as a down regulator of β -catenin; and the mutant *APC* protein would therefore lead to the accumulation of cytoplasmic β -catenin. This subsequently leads to an increased level of β -catenin/Tcf complex which alters the expression of genes such as c-myc and cyclin D, and then initiating uncontrollable cell growth.

An *APC* gene mutation can be identified in 80–90% of kindreds with the classical disorder [90]. In addition somatic mutations of the *APC* gene are found in as many as 80% of sporadic tumors, and seems to be involved in initiating tumorigenesis [91].

4.4.1.1 Familial adenomatous polyposis variants

Attenuated Adenomatous Polyposis Coli (AAPC), Gardner syndrome and Turcot syndrome are the three variants of the FAP syndrome. Extraintestinal manifestations distinguish the first two of these variants. AAPC as the name implies, represents a less pronounced FAP. In addition, biallelic mutations in the *MUTYH* gene have been found in 30% of families with multiple adenomas (15–100) who do not exhibit an autosomal dominant inheritance pattern or a germline mutation in the *APC* gene [92]. Patients with attenuated adenomatous polyposis coli develop fewer (10-99) polyps than the typical FAP and at a later age. The polyps tend to be distributed more proximally, and the risk of colon adenocarcinoma, appear to be lower than in a classical FAP[93]. Beside polyposis Gardner syndrome includes the development of benign extracolonic tumors, desmoid tumors, soft-tissue tumors, osteomas and dental abnormalities. Although extremely rare in the general population, desmoids occur in approximately 10% of patients with FAP, most often in the mesentery or abdominal wall following surgical trauma. Despite they do not metastasize, they are locally aggressive and a common cause of death in patients with FAP. The distinction between Gardner syndrome and FAP has become less clear with improved understanding of the genetic basis of these disorders. The same genetic abnormalities are found in patients with either syndrome and both syndromes can be seen within the same family [94]. The additional feature of Turcot syndrome is the growth of central nervous system tumors like medulloblastomas, astrocytomas and ependymomas. In about 70-80% of classical FAP an *APC* gene mutation can be detected. Nevertheless, in few patients with Turcot syndrome, *MLH1* and *MSH2* genes mutation have been described [95].

4.4.2 Hamartomatous polyposis syndromes

This group of disorders has in common the development of hamartomatous polyps in the GI tract, which may also confers an increased risk of cancer. These conditions include Juvenile Polyposis Syndrome (JPS), Peutz-Jeghers Syndrome (PJS), Bannayan-Riley-Ruvalcaba Syndrome (BRRS), Cowden Syndrome (CS), Cronkhite-Canada Syndrome (CCS), and Hereditary Mixed Polyposis Syndrome (HMPS).

4.4.2.1 Juvenile polyposis

JPS is characterized by multiple distinct juvenile polyps in the gastrointestinal tract and an increased risk of colorectal cancer. Clinically it is defined by the presence of five or more juvenile polyps in the colorectum, juvenile polyps throughout the gastrointestinal tract or any number of juvenile polyps, and a positive family history of juvenile polyposis [96, 97].

Juvenile polyposis syndrome (JPS) is a rare (2.8 per 100 000 in children under 10 years of age) autosomal dominantly inherited disorder with germline mutations in the *SMAD4* or *BMPR1A* gene in about 50%-60% of JPS patients [98, 99], and about 10% of JPS patients with a germline mutation in *PTEN* [100].

Patients present with bleeding, anemia, rectal prolapse of a polyp, diarrhea, abdominal pain and/or failure to thrive and in a small subset of patients congenital abnormalities, including cardiac, and bowel rotations have been shown. The most significant cause of mortality in patients with juvenile polyposis is colorectal cancer. In one large series, 18 of 87 patients (21%) developed colorectal cancer. These cancers tended to have early onset (mean age 34 years), were poorly differentiated and had a poor prognosis [96]. Overall, the risk of gastrointestinal malignancy may exceed 50% [101].

4.4.2.2 Peutz–Jeghers syndrome

Peutz–Jeghers syndrome (PJS) is a rare autosomal dominant condition (1 in 150,000 persons), characterized by the development of hamartomatous polyps, benign polyps comprised of several types of typical epithelial cells supported by a thick band of smooth muscle. They are found throughout the gastrointestinal tract (particularly the small bowel) and may cause bleeding, intussusception, or obstruction [102]. Patients usually develop distinctive mucocutaneous pigmented lesions on lips, buccal mucosa, hands, and feet. The lifetime risk of colon cancer is about 39%, and 93% for stomach, small bowel, pancreas, breast, sex cord, uterine, cervical, and melanoma cancers [103-107].

PJS has been associated with germline mutations or deletions in *LKB1* (*STK11*), a serine–threonine kinase that regulates p53-mediated apoptosis and the mammalian target of rapamycin (mTOR) pathway; *STK11* germline alterations can be identified in only 50%-60% of cases of suspected PJS [108-110].

4.4.2.3 PTEN Hamartoma Tumor Syndrome

The *PTEN* hamartoma tumor syndrome (PHTS) includes Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), *PTEN*-related Proteus syndrome (PS), and Proteus-like syndrome are rare polyposis syndromes related to juvenile polyposis.

CS is a multiple hamartoma syndrome with a high risk for benign and malignant tumors of the thyroid, breast, and endometrium [111]. Affected individuals usually have macrocephaly, trichilemmomas, and papillomatous papules and present by the late 20s. The lifetime risk of developing breast cancer is 25%-50%, with an average age of diagnosis between 38 and 46 years. The lifetime risk for thyroid cancer (usually follicular, rarely papillary, but never medullary thyroid cancer) is approximately 10%. The risk for endometrial cancer, although not well defined, may approach 5%-10% [111].

BRRS is a congenital disorder characterized by hamartomatous polyps as well as unusual facies, macrocephaly, developmental delay, and pigmented papules on the penis [43].

PS is a complex, highly variable disorder involving congenital malformations and hamartomatous overgrowth of multiple tissues, as well as connective tissue nevi, epidermal nevi, and hyperostoses [111].

Proteus-like syndrome is undefined but refers to individuals with significant clinical features of PS who do not meet the diagnostic criteria for PS [111].

The diagnosis of PHTS is made only when a *PTEN* mutation is identified. Up to 85% of individuals who meet the diagnostic criteria for CS and 65% of individuals with a clinical diagnosis of BRRS have a detectable *PTEN* mutation. Preliminary data also suggest that up to 50% of individuals with a Proteus-like syndrome and up to 20% of individuals with Proteus syndrome have *PTEN* mutations [111].

PTEN is a tumor suppressor gene. It acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins. Also acts as a lipid phosphatase and this activity is critical for its tumor suppressor function. Antagonizes the PI3K-AKT/PKB signaling pathway by dephosphorylating phosphoinositides and thereby modulating cell cycle progression and cell survival. The unphosphorylated

form cooperates with AIP1 to suppress AKT1 activation. Dephosphorylates tyrosine-phosphorylated focal adhesion kinase and inhibits cell migration and integrin-mediated cell spreading and focal adhesion formation. Plays a role as a key modulator of the AKT-mTOR signaling pathway [112].

4.5 Diagnostic and prognostic genetic markers in cancer

Tumor development starts through the accumulation of genetic and epigenetic alterations that allow cells to escape from the cellular control mechanisms that regulate the homeostatic equilibrium between cell proliferation and cell death [113].

To transform a primary cell into a malignant one *in vitro* requires alterations in the functionality of many processes by which cells control their growth, division and differentiation [114]. Studies on colorectal, breast, lung, pancreatic cancer tissues have demonstrated the view that acquisition of genetic alterations and epigenetic changes could determine the morphological changes that leads to cancer progression [15, 115]. Mutations in cancer cells comprises a large variety of DNA alterations including chromosome copy number, (micro)deletions/(micro)duplications or structural chromosomal alterations like translocations, deletions or amplifications, as well as changes at nucleotide level such as point mutations affecting a single nucleotide at a critical position in a cancer-related gene. It has been shown that these alterations often co-exist in a single tumor [116]. Genetic and epigenetic events represent two complementary mechanisms that are involved in carcinogenesis initiation/progression/metastasis and sometimes coexisting [117].

Biomarkers can be defined as biological molecules found in blood, other body fluids, or tissues, which can indicate a normal or abnormal process, or directly a condition or disease. A biomarker may be further used to see how well the body responds to a treatment for a disease or condition [2, 118]. Biomarkers are an reliable indicators for a disease process and can be divided in five categories: diagnostic, early detection, prognostic and predictive [118]. Each marker should have different characteristics such as specificity, refers to the quantity of control subjects who test negative for the biomarker (patients without disease), and sensitivity refer to quantity of case subject who test positive for the biomarker (patients with confirmed disease).

Some examples are: *TP53* has been found mutated in many tumors and is used as biomarker in epidemiology and for early cancer detection [119-122]; *EGFR* is often mutated in lung adenocarcinomas of never smokers (with

additional mutation in *KRAS*) [123] and is used as biomarker for predicting the response of lung cancer patients to small molecule Tyrosine Kinase Inhibitors (TKIs) such as erlotinib or gefinitib [124-127]; *de novo* methylation of the *p16^{INK4a}* promoter is one of the most frequent epigenetic alterations in human cancer and appears to be the earliest event in some cancer types [128, 129]; *p16^{INKa}*, *p15^{INK4b}*, *RASSF1A*, *MLH1*, *GSTP1*, *CDH1*, *APC*, and *DAPK1* are frequently methylated in circulating DNA and used as biomarkers in the clinic and risk assessment [130]; miR-141 in prostate cancer patients has been shown to be a marker that can distinguish, with significant specificity and sensitivity, patients with cancer from healthy controls [131]. Thus, the discovery of novel gene/pathway alterations could reveal new biomarkers for cancer detection, diagnosis and prognosis.

Tumor markers currently use in clinical practice.

Tumor Marker	Cancer Type	Tissue Analyzed	Usage
ALK gene rearrangements	Non-small cell lung cancer; anaplastic large cell lymphoma	Tumor tissue	To help determine treatment and prognosis
Alpha-fetoprotein (AFP)	Liver cancer; germ cell tumors	Blood	To help diagnose liver cancer and follow response to treatment; to assess stage, prognosis, and response to treatment of germ cell tumors
Beta-2-microglobulin (B2M)	Multiple myeloma; chronic lymphocytic leukemia; some lymphomas	Blood, urine, or cerebrospinal fluid	To determine prognosis and to follow response to treatment
Beta-human chorionic gonadotropin (Beta-hCG)	Choriocarcinoma; testicular cancer	Urine or blood	To assess stage, prognosis, and response to treatment of testicular cancer
BCR-ABL	Chronic myeloid leukemia	Blood and/or bone marrow	To confirm diagnosis and monitor disease status
BRAF mutation V600E	Cutaneous melanoma; colorectal cancer	Tumor tissue	To predict response to targeted therapies
CA15-3/CA27.29	Breast cancer	Blood	To assess whether treatment is working or disease has recurred
CA19-9	Pancreatic cancer; gallbladder cancer; bile duct cancer; gastric cancer	Blood	To assess whether treatment is working
CA-125	Ovarian cancer	Blood	To help in diagnosis, assessment of response to treatment, and evaluation of recurrence
Calcitonin	Medullary thyroid cancer	Blood	To aid in diagnosis, to check whether treatment is working, and to assess recurrence
Carcinoembryonic antigen (CEA)	Colorectal cancer; breast cancer	Blood Blood	To check whether colorectal cancer has spread; to look for breast cancer recurrence and assess response to treatment
CD20	Non-Hodgkin lymphoma	Blood	To determine whether treatment with a targeted therapy is appropriate
Chromogranin A (CgA)	Neuroendocrine tumors	Blood	To help in diagnosis, assessment of treatment response, and evaluation of recurrence

Chromosomes 3, 7, 17, and 9p21	Bladder cancer	Urine	To help in monitoring for tumor recurrence
Cytokeratin fragments 21-1	Lung cancer	Blood	To help in monitoring for recurrence
EGFR mutation analysis	Non-small cell lung cancer	Tumor tissue	To help determine treatment and prognosis
Estrogen receptor (ER)/progesterone receptor (PR)	Breast cancer	Tumor tissue	To determine whether treatment with anti-hormonal therapy (such as tamoxifen) is appropriate
Fibrin/fibrinogen	Bladder cancer	Urine	To monitor progression and response to treatment
HE4	Ovarian cancer	Blood	To assess disease progression and monitor for recurrence
HER2/neu	Breast cancer; gastric cancer; esophageal cancer	Tumor tissue	To determine whether treatment with trastuzumab is appropriate
Immunoglobulins	Multiple myeloma; Waldenstrom macroglobulinemia	Blood and urine	To help diagnose disease, assess response to treatment, and look for recurrence
KIT	Gastrointestinal stromal tumor; mucosal melanoma	Tumor tissue	To help in diagnosis and determining treatment
KRAS mutation analysis	Colorectal cancer; non-small cell lung cancer	Tumor tissue	To determine whether treatment with a particular type of targeted therapy is appropriate
Lactate dehydrogenase	Germ cell tumors	Blood	To assess stage, prognosis, and response to treatment
Nuclear matrix protein 22	Bladder cancer	Urine	To monitor response to treatment
Prostate-specific antigen (PSA)	Prostate cancer	Blood	To help in diagnosis, assess response to treatment, and look for recurrence

Thyroglobulin	Thyroid cancer	Tumor tissue	To evaluate response to treatment and to look for recurrence
Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor (PAI-1)	Breast cancer	Tumor tissue	To determine aggressiveness of cancer (and guide treatment)
70-Gene signature (MammaPrint)	Breast cancer	Tumor tissue	To evaluate risk of recurrence
21-Gene signature (Oncotype DX)	Breast cancer	Tumor tissue	To evaluate risk of recurrence
5-Protein signature (Ova1)	Ovarian cancer	Blood	To pre-operatively assess pelvic mass for suspected ovarian cancer

Table taken from [2]

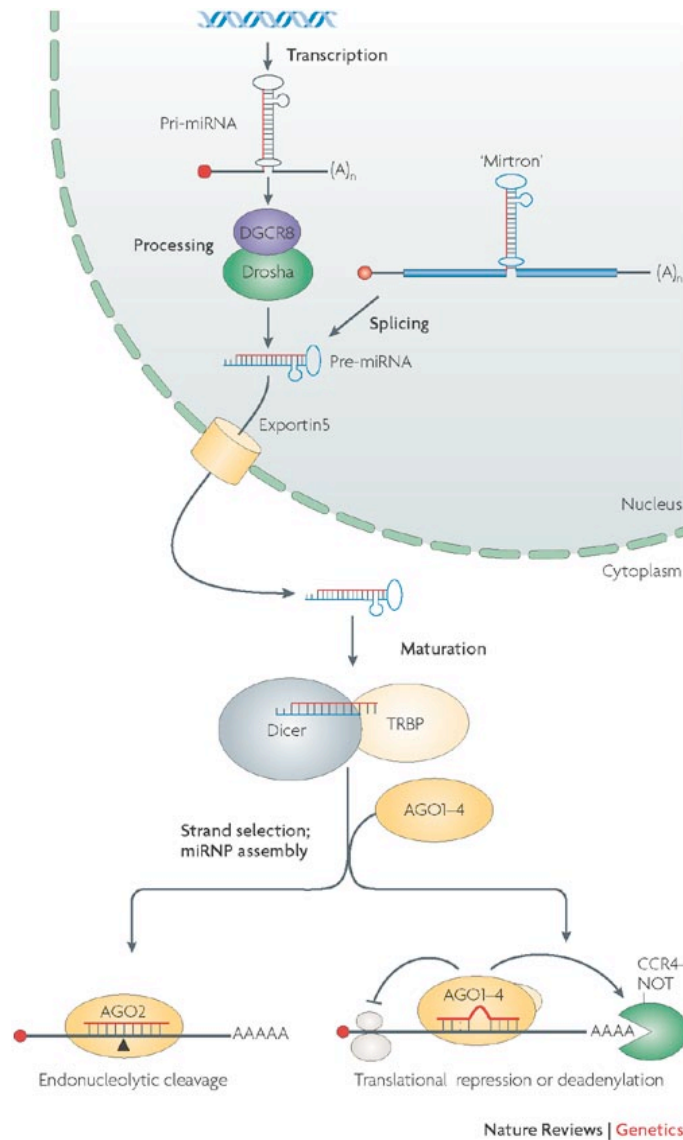
4.6 MicroRNA: biogenesis, functions and targets

MicroRNAs (miRNAs) represent a large family of small non-coding RNAs of about 25 nucleotides in length that serve as effector molecules of sequence-specific gene silencing [132]. It is estimated that the number of miRNAs in the human genome range from about 450 to 1000 and that they control gene expression of about 30% of all protein-coding genes in mammals. The majority of identified miRNAs, are highly evolutionary conserved among many distantly related species, from worms to human, suggesting that miRNAs have roles in essential biological processes, including developmental timing, stem-cell differentiation, signal transduction.

In the nucleus, miRNA genes are generally transcribed by RNA polymerase II or III to form large primary miRNA transcripts (pri-miRNAs). These are further processed by Drosha, an RNase III protein, into 70-nucleotide miRNA precursors (pre-miRNA). After transport into the cytoplasm, pre-miRNAs are further processed by another RNase III enzyme, Dicer, into miRNA duplexes, typically consisting of 19-25 nucleotides in length [133, 134]. Subsequently these duplexes can be loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) and the mature miRNA strand is preferentially retained. Once bound to the 3' untranslated region (UTR) of target mRNAs, the mature miRNA induces cleavage, translational repression or deadenylation, depending on the degree of complementarity [135]. A single miRNA may bind to as many as 200 target genes encoding a broad range of proteins, such as transcription factors, receptors and transporters. In recent years, several approaches have been used to identify miRNA targets [136].

For all known miRNA targets, perfect or near-perfect complementary sites of miRNAs have been conserved. This provides a powerful strategy for the prediction of miRNA targets through computational approaches, and several laboratories have developed different computational methods to achieve this using the available genome database; hundreds of miRNA targets for given miRNAs were predicted [137].

MicroRNA biogenesis



The miRNA processing pathway, taken from Filipowicz *et al.* [132]: canonical maturation includes the production of the primary miRNA transcript (pri-miRNA) by RNA polymerase II or III and cleavage of the pri-miRNA by the microprocessor complex Drosha–DGCR8 (Pasha) in the nucleus. The resulting precursor hairpin, the pre-miRNA, is exported from the nucleus by Exportin-5–Ran-GTP. In the cytoplasm, the RNase Dicer in complex with the double-stranded RNA-binding protein TRBP cleaves the pre-miRNA hairpin to its mature length. The functional strand of the mature miRNA is loaded together with Argonaute (Ago2) proteins into the RNA-induced silencing complex (RISC), where it guides RISC to silence target mRNAs through mRNA cleavage, translational repression or deadenylation, whereas the passenger strand (black) is degraded. In this review we discuss the many branches, crossroads and detours in miRNA processing that lead to the conclusion that many different ways exist to generate a mature miRNA.

4.6.1 MicroRNA and cancer

Early observations suggested a potential role of miRNAs in human carcinogenesis. miRNAs discovered in *C. elegans* and in *Drosophila* were shown to control cell proliferation and apoptosis [138, 139] and their deregulation can contribute to proliferative diseases like cancer. It has also been shown that malignant tumors and tumor cell lines have widespread deregulated miRNA expression compared to normal tissues [140, 141].

One of the questions to be answered is if miRNA expression observed in cancer is cause or rather consequence of malignant transformation [142].

In the past 10 years many studies were performed in this direction, with the first direct evidence shown by Calin *et al.* They identified two miRNA, *mir-15* and *mir-16*, transcribed from genes located in a 30-kb deletion on chromosome 13, which is the most common chromosomal abnormality in chronic lymphocytic leukemia (CLL), consequently they found the two miRNAs absent or downregulated in the majority (68%) of cases compared to normal tissue. This finding suggested that these two miRNAs were causally involved in the pathogenesis of chronic lymphocytic leukemia [143].

Other studies described specific activities of some microRNAs as oncogenes or tumor suppressors. He *et al.* showed the relationship between a miRNA cluster, *mir-17-92 (oncomir-1)* with the Myc oncogenic pathway [144, 145], using a mouse model of human B-cell lymphoma caused by *Myc* oncogene overexpression. They demonstrated that upregulation of the *mir-17-92* cluster accelerated c-Myc-induced tumorigenesis in mice. The oncogenic role of the *mir-17-92* cluster was confirmed by Matsubara *et al.* and Sylvestre *et al.* [146, 147]. On the other hand, O'Donnell *et al.* identified the same cluster of miRNAs, *mir-17-92*, to be regulated by the transcription factor c-Myc which induces expression of a growth factor E2F1. The *mir-17-92* cluster, which is also induced by c-Myc, in contrast, inhibits E2F1 expression, conferring a potential role as a tumor suppressor [144, 145].

MicroRNAs are also involved in the regulation of p53 that regulates part of the stress responses including response to the DNA damage. miR-34, is directly activated by p53 after DNA damage. Overexpression of miR-34 induces cell cycle arrest and promotes apoptosis [148, 149].

It has been shown that expression of many miRNAs is significantly reduced in cancers compared to matched normal tissues. Furthermore, undifferentiated tumors had lower miRNA levels compared with more-differentiated tumors [141, 150], pointing to possible miRNAs functions in terminal differentiation and cell separation. Kumar et al. verified that reduction in miRNA levels promote tumorigenesis by knockdown of Drosha and Dicer in cell lines [151]. These cells with global miRNA downregulation showed enhanced cellular growth *in vitro*. When injected into nude mice, they generated faster growing and more invasive tumors compared to controls. To assess the effect of global miRNA downregulation *in vivo*, the authors deleted Dicer in a mouse model of lung cancer. The enzyme lacking mutant mice developed an increased tumor burden, with a growth in tumor number and tumor size, as well as tumors that were less well differentiated compared to controls. These data suggest that global miRNA downregulation enhances tumorigenesis [151]. Together, these data indicate that altered expression of miRNAs has a pleiotropic effect on tumor suppressors and oncogenes and represents a crucial step in tumorigenesis.

4.7 HMG proteins

High Mobility Group (HMG) proteins, can be divided into three distinguishable families; (1) HMGA, (2) HMGB, and (3) HMGN. All three families share many biochemical features, but each has its own characteristic functional motifs and is expressed in distinctive ways in different cells and tissues [152].

HMG protein families are classified as 'architectural transcription factors' according to their ability to positively and negatively regulate gene transcription by binding to DNA or chromatin in a structure-specific manner. HMG proteins have been implicated in a diverse array of additional nuclear processes like chromatin and nucleosome remodeling events, cell cycle-related chromosomal changes, genetic recombination, DNA replication and repair, apoptosis, and molecular chaperoning [153-157].

4.7.1 HMGA family

The HMGA protein family consists of *HMGA1* and *HMGA2* genes that encode four proteins named HMGA1a, HMGA1b, HMGA1c (spliced isoforms of the HMGA1 gene; locus 6p21) and HMGA2. The HMGA1a, HMGA1b and HMGA2 proteins are composed of 107, 96 and 108 amino acid residues, respectively. Each protein contains three basic domains, named AT-hooks and an acidic C-terminal region, this motif binds the minor groove of a target DNA strand. The HMGA1a protein differs from HMGA1b in that it has an additional insertion of 11 amino acid residues between the first and the second AT-hook domains. The structure of HMGA2, very similar to that of HMGA1b, contained a short peptide of 12 amino acid residues between the third AT-hook and the acidic C-terminal [158].

Until 2004 there were reports of over 50 different eukaryotic and viral genes regulated by HMGA proteins. The vast majority (>35) of these are positively regulated and their inducible expression is controlled by a variety of biological and environmental stimuli. The promoter regions of many of the positively regulated genes contain multiple stretches of AT-rich sequence (AT-hook binding site). Transcriptional activation of these types of promoter often

involves the formation of an “enhanceosome”, a stereo-specific, multi-protein complex that includes HMGA proteins and other transcription factors making specific protein-DNA and protein-protein contacts in intricate, but precise, ways. In the case where HMGA proteins act as negative regulators of gene transcription they often serve as inhibitors of enhanceosome formation, usually by sterically blocking the functional binding of other crucial transcription factors to their recognition sites in gene promoters [159]. Very recently, the down-regulation of the recombination activating gene (RAG2) by HMGA1 proteins has been described [160]. One of the best-studied mechanisms of gene regulation in which HMGA proteins are involved is that of the interferon- β gene (IFN- β). IFN- β expression is due to a multifactorial complex that assembles at the nucleosome-free enhancer region of the gene, formed by the factors NF- κ B, IRF, ATF2/cJun, and the HMGA1a protein. HMGA1a plays a dual function in this context: (i) it induces allosteric changes in the DNA, thus increasing the affinity of the transcription factors for their binding sites and (ii) it establishes protein-protein interactions with the same factors. This new structure, called enhanceosome, is responsible for the modification and the remodeling of a nucleosome that masks the TATA-box; consequently, transcription can start. This remodeling is triggered by the recruitment from the “enhanceosome” of GCN5/PCAF that acetylates the nucleosome and also HMGA1a at K64, the latter modification resulting in the stabilization of the enhanceosome. Later, another acetyltransferase called CBP modifies HMGA1a at K70 destabilizing the enhanceosome and, consequently, repressing transcription [161].

4.7.2 HMGA expression in normal and neoplastic tissues

The HMGA2 gene is not expressed in human tissues except for CD34 positive hematopoietic stem cells. The HMGA1 gene is expressed at low levels in human tissues: a higher expression was observed in testis, skeletal muscle and thymus. Conversely, both genes are widely expressed during embryogenesis. HMGA1 and HMGA2 over-expression was first described in rat thyroid transformed cells and in experimental thyroid tumours. Over-expression of the HMGA proteins was then found to be a common feature of experimental and human malignant neoplasias, including thyroid, prostate, uterus, breast, colorectum, ovary and pancreas carcinomas [158]. Recently, HMGA1 expression has been correlated with the histological grade of human glial tumors [162]. Moreover, the expression level of the HMGA proteins is significantly correlated with parameters of poor prognosis in patients with colorectal cancer. In all of these epithelial/endothelial cell-derived malignant tumors, the over-expressed proteins are full-length non-mutants forms. In contrast to the situation in carcinomas, benign tumors of mesenchymal origin (lipomas, leiomyomas, fibroadenomas, aggressive myxomas, pulmonary hamartomas and endometrial polyps) often contain chromosomal rearrangements that result in the creation of new hybrid genes that code for chimeric proteins in which the AT-hooks of the HMGA proteins are fused to ectopic peptidic sequences [163]. Over-expression of the HMGA proteins appears to be a necessary event in *in vivo* cell transformation. This was demonstrated by experiments in which HMGA expression was blocked by transfecting rat thyroid cells with antisense HMGA constructs. When these cells were infected by the myeloproliferative sarcoma virus and the Kirsten murine sarcoma virus carrying the v-mos and v-ras-Ki oncogenes, respectively, they did not acquire the typical markers of neoplastic transformation (ability to grow in soft agar and induce tumors after injection into athymic mice). Conversely, these markers were shown by the untransfected rat thyroid cells infected with the same murine retroviruses [164]. Over-expression of HMGA1 proteins is also essential in the development of cancer in humans. In fact, an adenovirus carrying the HMGA1 gene in an antisense orientation induces programmed cell death in carcinoma

cell lines derived from human thyroid, lung, colon and breast cancers. Moreover, it has been reported that the overexpression of HMGA1a or HMGA2 leads to neoplastic transformation of both Rat-1a fibroblasts and CB33 cells, whereas the decrease of HMGA1a/b expression abrogates transformation in Burkitt's lymphoma cells [165].

In chapters 6.3 and 6.3.1 we characterized HMGA1, HMGA2, at protein level in different human neoplastic entities, correlating their overexpression with clinic-pathological features.

4.8 Cancer Stem Cell Markers

Stem cells are defined as cells that have the ability to perpetuate themselves through self-renewal and to generate mature cells of a particular tissue through differentiation [166] .

Because normal stem cells and cancer cells share the ability to self-renew, it seems reasonable to propose that newly arising cancer cells appropriate the machinery for self-renewing cell division that is normally expressed in stem cells. Evidence shows that many pathways that are classically associated with cancer may also regulate normal stem cell development [167-169]

In most tissues, stem cells are rare but it has been shown for solid cancers that these cell populations are phenotypically heterogeneous and that only a small proportion of cells are clonogenic. These observations led to the hypothesis that only a few cancer cells are actually tumorigenic and that these tumorigenic cells could be considered as cancer stem cells [166, 170-175].

The cancer stem cell hypothesis postulates that a small subpopulation of cancer cells possessing self-renewal characteristics is responsible for initiating and maintaining cancer growth. According to the CSC model the large populations found in a tumor might represent diverse stages of differentiation. The biological characteristics shared by normal stem cells (NSCs) and CSCs mainly involve self-renewal and differentiation potential, survival ability, niche-specific microenvironment requirements and specific homing to injury sites and may have important implications in terms of new approaches to cancer. The identification of new therapeutic targets based on the CSC model represents a great challenge.

In the chapter 6.4 in accordance to the CSC hypothesis, we evaluated a potential role of CSC proteins in tumors of the ampulla of Vater. Our findings indicate, that in ampullary carcinomas, loss of expression of EpCAM may be linked to a more aggressive tumor phenotype.

4.9 MAGE Gene Family

The melanoma antigen gene (*MAGE*) family is divided into two subfamilies: *MAGE I* and *II*. *MAGE I* consists of a large number of genes located on the X chromosome, including: *MAGE-A* located on Xq28 [176, 177], *MAGE -B* cluster located on Xp21 [178, 179], and *MAGE-C* on Xq26–27 [180]. *MAGE-A*, *-B*, and *-C* is characterized by a large terminal exon encoding the entire protein. Most of them are relevant cancer/testis antigens (CTA) [181] and then silent in normal adult tissues except in male germ cells [182], and highly expressed in various forms of cancer [183]. Studies of *MAGE I* genes, aiming at quantifying mRNA expression in cancer (due to limited specificity of *MAGE* antibodies), have found the highest frequency of *MAGE-I* mRNA expression in different types of cancers like melanoma and lung cancer, particularly the squamous cell type. In contrast, hematopoietic malignancies, including lymphomas and leukemia as well as renal, colon, and pancreatic cancers, displayed notably low *MAGE- I* expression. For instance, *MAGE-A3* mRNA expression has been observed in 85% of non-small-cell lung carcinoma [184], but only in 41% of multiple myeloma [185], 37% of bladder cancer [186] and 10% of the breast cancers [187]. Most investigations showed high expression of MAGEs, except *MAGE-A4* that was often associated with poor outcome [188]. Higher grade and metastatic tumors have also been found to have high MAGEs expression than the primary tumors [189, 190].

The *MAGE-II* family, which includes the *MAGE-D* group, differs from the previously described members since they are almost universally expressed in all tumor free tissues and not related to cancer. They also differ by their genomic structure, the open reading frame of *hMAGE-D2* being split over 11 exons. Importantly, *MAGE-D1* was recently found to interact with the p75 neurotrophin receptor and to facilitate nerve growth factor-dependent apoptosis [191]. *MAGE-D1* was also recently reported to interact with members of the Dlx/Msx homeodomain family and to regulate the transcriptional function of Dlx5 [192].

Encoding tumor-specific antigenic peptides directed against *MAGE-I* expressing cancer cells, *MAGE-I* protein-derived peptides are currently studied as targets for the development of cancer vaccines [193, 194] and

antitumor immunotherapy [195-197]. In addition, expression analysis in various malignancies could be of diagnostic and/or prognostic relevance.

In chapter 6.5 we generated, a MAGE-A10 monoclonal antibodies (mAbs) to staining a multitumor tissue microarray. We described an overexpression of MAGE-A10 in cells from a number of malignancies, including lung, skin and urothelial tumors.

5. AIMS

My thesis focuses on the discovery and characterization of new diagnostic and prognostic markers in various cancer entities, in particular in sporadic and hereditary colorectal cancer. The work can be divided in seven major parts with the subdivision reflecting the structure of the results and the appendix.

The aims of the first part (chapter 6.1) were:

- To elucidate the frequency of contractions/insertions in the 3'UTR of the *EWSR1* gene on MSI detection in MMR deficient and proficient cancers through capillary electrophoresis analysis 319 patients.
- The role of these modifications in mRNA processing and protein expression *in vitro* (siRNA-mediated poly(A) site selection and pull-down assays) and *in vivo* (mRNA and protein expression by RT-PCR and tissue microarray analysis).

The aims of the second part (chapter 6.2) were:

- To understand the type and frequency of genetic as well as epigenetic alterations at the *SH2D4A* locus in colorectal cancer via Sanger sequencing, gene dosage analysis by multiplex ligation-dependent probe amplification (MLPA), quantitative PCR for mRNA and immunohistochemical staining for protein expression.
- To define the *SH2D4A* pathway interactions via immunoprecipitation and cell transfections.
- To test whether *SH2D4A* expression impacts on cell growth via interaction with its partners by knock-down experiments.

The aim of the third part (chapters 6.3 and 6.3.1) was:

- To investigate the prognostic role of HMGA1 and HMGA2 protein in breast and pancreatic cancers applying immunohistochemical techniques on tissue microarrays and gathering of clinic-pathological information and statistical comparison.

The aim of the fourth part (chapter 6.4) was:

- To explore the prognostic usage of cancer stem cell markers in ampulla Vater carcinomas using the tissue microarray technology followed by survival analysis.
-

The aim of the fifth part (chapter 6.5) was:

- To develop a specific *MAGE* monoclonal antibody and investigate its expression in a multitumor array by immunohistochemical techniques.

The aims of the sixth part (chapter 8.1) were:

- To assess the miRNA profile in Lynch syndrome-associated colorectal cancers (CRC) and establish their prognostic and/or therapeutic significance by deep sequencing of MMR deficient and proficient cell lines as well as cancers from Lynch syndrome and sporadic CRC patients.
- Verification and validation of selected, differentially expressed miRNAs using quantitative real-time PCR and cell transfection assays.

The aims of the seventh part (chapter 8.2) were:

- To investigate copy number variation aberrations in CRCs from six Lynch syndrome patients using a high-resolution DNA SNP array.
- To validate the findings in a cohort of 46 Lynch syndrome related and 50 sporadic CRCs.

6. RESULTS

6.1 3'UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancer.

My contribution to this work:

- RNA and DNA extraction from Lynch syndrome and sporadic colorectal cancers
- Molecular assessment of EWS16T contractions/expansions;
- Microsatellite instability analyses of Lynch syndrome and sporadic cancers;
- Quantitative real time PCR experiments;
- Immunohistochemical analysis of CRC whole tissue and microarray sections;
- Data and statistical analysis;
- Manuscript writing;

3'UTR poly(T/U) tract deletions and altered expression of *EWSR1* are a hallmark of mismatch repair deficient cancers

Shivendra Kishore^{1*}, Salvatore Piscuoglio^{2*}, Michal Kovac^{2*}, Annette Gylling³, Friedel Wenzel², Francesca Trapani⁴, Hans Joerg Altermatt⁶, Valentina Mele^{4,5}, Giancarlo Marra⁷, Päivi Peltomäki³, Luigi Terracciano⁴, Mihaela Zavolan^{1§} and Karl Heinimann^{2§}.

¹Biozentrum, University of Basel, Basel, Switzerland.

²Research Group Human Genetics, Department of Biomedicine, University of Basel, and Division of Medical Genetics, University Children's Hospital, Basel, Switzerland.

³Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland.

⁴Institute of Pathology, University Hospital of Basel, Basel, Switzerland.

⁵Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, Basel, Switzerland.

⁶*Pathologie Laenggasse Bern, Bern, Switzerland*

⁷Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland.

** These authors equally contributed to the work*

*§*Corresponding authors:

Mihaela Zavolan

Biozentrum

University of Basel

Klingelbergstrasse 50 / 70

4056 Basel, Switzerland

Phone: +41 61 267 15 77

E-mail: mihaela.zavolan@unibas.ch

and

Karl Heinimann

Research Group Human Genetics

Department of Biomedicine, University of Basel

Mattenstrasse 28

4058 Basel, Switzerland

Phone: +41 61 267 07 73

Fax: +41 61 267 07 78

E-mail: karl.heinimann@unibas.ch

Microsatellite instability (MSI), the genome-wide accumulation of DNA replication errors at repetitive nucleotide sequences, constitutes the hallmark lesion of DNA mismatch repair (MMR) deficient cancers¹. Present in Lynch syndrome-related and about 10-20% of sporadic colorectal (CRC), gastric and endometrial cancers MSI testing is widely used to guide clinical management. The functional significance of MSI at non-coding repeat loci such as the 3' untranslated region (UTR), however, remains largely elusive¹⁻³. Here we describe a mononucleotide (T/U)₁₆ tract, EWS16T, located in the 3' UTR of the *Ewing sarcoma break point region 1 (EWSR1)* gene which discriminates MMR proficient from MMR deficient cancers with 100% sensitivity and specificity. We demonstrate *in vitro* and *in vivo* that contractions at this locus alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in *EWSR1* pre-mRNAs and result in decreased mRNA as well as protein expression. In contrast to their proficient counterparts, MMR deficient CRC display altered subcellular localization of EWS with diffuse cytoplasmic staining. EWS16T thus not only represents a novel monomorphic MSI target locus to accurately identify both, hereditary and sporadic, MMR deficient cancers but contractions therein affect multiple regulatory mechanisms implicating the RNA-/DNA-binding protein EWS in MSI-associated colorectal tumorigenesis.

Our investigation of an extra-osseous Ewing sarcoma (ES) from a *MSH6* mutation carrier (c.3696dupT) previously affected by colon cancer revealed that the tumor displayed the typical hallmarks of LS, i.e. MSI at mononucleotide markers and specific loss of *MSH6* expression (Supplementary Fig. S1), but none of the molecular features commonly associated with ES, i.e. chromosomal translocations involving the *EWSR1* gene on 22q12 (Supplementary Fig. S2). In view of the tumor's MSI-high phenotype we focused on a mononucleotide tract in the 3' UTR of the *EWSR1* gene consisting of 16 thymines (EWS16T; c.*318_*333). Both of the patient's tumors, ES and colon cancer, were found to carry somatic contractions/deletions of 4 and 5 thymines, respectively, prompting us to assess the prevalence of somatic alterations at EWS16T in MMR deficient cancers in general. In contrast to Wheeler et al.⁴ who reported a deletion variant (c.*331_*333delTTT; dbSNP: rs76631619) in James D. Watson's genome, we found the EWS16T locus to be monomorphic in over 300 constitutional DNA samples tested. As a first step, we *analyzed* 85 cancers (78 colorectal (CRCs) and 7 endometrial) and matching leukocyte-derived DNA samples from 78 Swiss LS patients with confirmed MMR germline mutations (58 *MLH1*, 20 *MSH2*). In addition, we investigated 14 sporadic MMR-deficient CRCs with *MLH1* promoter hypermethylation as well as 85 sporadic MMR-proficient CRCs. Assessment of EWS16T tract length by capillary electrophoresis of fluorescently labeled PCR products revealed that all MMR-deficient cancers but none of the MMR-proficient ones displayed novel alleles, i.e. contractions or expansions at the EWS16T tract (Fig. 1). Subsequently, the findings were independently confirmed in a Finnish cohort of 122 patients: all 29 MMR-deficient (12 CRC, 17 gastric cancers) but none of the 93 MMR-proficient (38 CRC and 55 gastric) cancers showed EWS16T tract instability. The majority (72.7%) of somatic alterations consisted of contractions/deletions of 4 or more base pairs (Supplementary Fig. 3). Consistently, we further found that MMR-deficient cell lines (LoVo, HCT15, HCT116) carry only mutated EWS16T alleles (contractions). In contrast, MMR-proficient cell lines (HT29, SW480) as well as 12 MMR-proficient, MSI-low CRCs from Swiss patients were stable at EWS16T (Table 1). The

EWS16T tract thus represents a novel, monomorphic MSI target locus identifying MMR-deficient cancers with 100% sensitivity and specificity. Because somatic EWS16T tract alterations were exclusively present in MMR-deficient cancers, whether hereditary or sporadic, and occurred in all types of cancer investigated (colorectal, gastric and endometrial) we wondered about their possible functional role(s) in MMR-related carcinogenesis. The poly(T/U) tract deletions occur in a region which encodes the 3'UTR of *EWSR1* and do not alter the coding sequence of the *EWSR1* gene. Nonetheless, 3'UTRs contain sequence elements that are important for the post-transcriptional regulation of protein levels. Furthermore, it has been recently demonstrated that changes in 3'UTR length through alternative polyadenylation activates oncogenes⁵. We thus set out to characterize the effect of EWS16T tract deletions on *EWSR1* expression levels. The catalog of 3' end cleavage sites that we recently generated through 3' end sequencing in human embryonic kidney (HEK) 293 cells⁶ shows that *EWSR1* undergoes alternative polyadenylation, generating two transcript forms that differ in the length of their 3'UTRs (Supplementary Fig. 4). The ESW16T tract deletions occur very close to the distal poly(A) signal and may thus result in changes in 3' end processing factor assembly, thereby altering the poly(A) site selection. To investigate this possibility, we cloned the 3'UTR of *EWSR1* downstream of the Renilla luciferase-coding region in a psiCHECK-2 mutant vector in which the synthetic poly(A) site was mutated (psiCHECK-2-SPAm). This construct thus allowed only the usage of the poly(A) signal from the cloned 3'UTR of *EWSR1*. We then generated variant constructs containing poly(T/U) tracts of variable lengths through deletion mutation. With primers that simultaneously detect both the short and the long 3'UTR isoforms in a multiplexed semi-quantitative PCR, we found that deletions in the ESW16T tract promoted the usage of the distal poly(A) site (Figure 2a-b). We further investigated the MMR-proficient and MMR-deficient colon cancer cell lines and found that, consistent with our findings in the heterologous system, MMR deficiency is associated with higher expression of the longer *EWSR1* isoform (Figure 2c). These results indicate that the EWS16T tract deletions alter poly(A) site selection. To determine the factors involved in *EWSR1* poly(A) site selection, we used S1 aptamer-tagged, *in vitro* transcribed wildtype 3'UTR and a 3'UTR variant with 6U deletions in

the EWS16T region to pull down the proteins that associated with these RNAs (Supplementary Fig. 5). In three independent experiments we reproducibly identified a set of A/U-rich element binding proteins that associate with these constructs (Figure 3A and 3B). Interestingly, we found that NF45/90/110, hnRNPC and HuR associate with the wildtype but not with the mutant 3'UTR. NF45 and NF90 have been previously shown to be a part of a heterodimeric complex, nuclear factor of activated T-cells (NFAT), which is required for T-cell expression of interleukin 2, with NF110 being a larger isoform of NF90. NF90 has been shown to regulate mRNA stability and redistribution of nuclear mRNAs in the cytoplasm⁷. The EWS16T mutant preferentially associated with the SFPQ/NONO heterodimer, which is an essential pre-mRNA splicing factor that couples splicing with polyadenylation as a component of a small nuclear ribonucleoprotein (snRNP)-free complex with SNRPA/U1A⁸. To determine which of the RBPs identified above influenced the poly(A) site selection, we knocked down their expression individually with siRNAs (Figure 4a) and assessed the poly(A) site usage in reporter constructs that had either the wildtype or the mutant EWS16T tract (6U deletions) cloned downstream of luciferase. None of the siRNAs influenced the poly(A) site usage in the wildtype constructs in which only the shorter isoform was expressed (Figure 4b). When the 6U deletion construct was used however, the knockdown of SFPQ and hnRNPC strongly influenced poly(A) site selection. Specifically, knockdown of SFPQ promoted proximal site usage, while the knockdown of hnRNPC led to increased expression of the longer 3'UTR isoform (Figure 4c). This result was consistent with the binding pattern of the RBPs. hnRNPC binding was most prominent immediately downstream of the proximal poly(A) site (our unpublished PAR-CLIP data on hnRNPC), while SFPQ most likely bound to the truncated U-tract immediately upstream of the distal poly(A) site, as inferred from the pull down experiments (Figure 3a). Interestingly, knockdown of SFPQ did not affect poly(A) site selection in wildtype constructs, consistent with our earlier findings that SFPQ specifically associated with the construct carrying the 6U deletion. Thus, our results indicate that hnRNPC and SFPQ have antagonistic activity on the processing of mutant *EWSR1* pre-mRNA with hnRNPC promoting the generation of the shorter and SFPQ of the longer 3'UTR isoform.

To finally determine if the choice in polyadenylation site may influence *EWSR1* expression, we performed luciferase assays on the constructs that carried the wildtype or various EWS16T deletion variants. The results shown in Supplementary Fig. 6 indicate a significant downregulation (up to 30%) of protein levels associated with EWS16T tract deletions. We attempted to corroborate our observations *in vivo*, assessing *EWSR1* mRNA expression in a set of 8 LS-related MMR-deficient and 5 sporadic MMR-proficient CRCs relative to matched, tumor-free mucosa by RT-PCR. Four (50%) out of the 8 LS, but only 1 (20%) out of the 5 sporadic CRC showed significantly (>1.2 fold) increased expression of the longer 3'UTR isoform (Figure 5A). The total *EWSR1* mRNA levels, however, were significantly reduced (>1.2 fold) in both LS-related and sporadic CRC tissues (Figure 5B), pointing to additional mechanisms that regulate *EWSR1* expression in sporadic CRCs. To determine the consequences of altered *EWSR1* mRNA expression, we performed immunohistochemical analysis (IHC) of 10 LS-related, MMR-deficient and 9 sporadic, MMR-proficient CRCs. Consistent with the data at the mRNA level, the cancers displayed on average an approx. 30% reduction in EWS expression when compared to matched, tumor-free mucosa. Unexpectedly, however, MMR-deficient and -proficient cancers significantly differed with regard to the subcellular localization of EWS ($P < 0.001$): tumor-free colon mucosa and adenomas from LS and sporadic CRC patients as well as sporadic carcinomas showed exclusively nuclear expression (Fig. 6). In contrast, LS-related CRCs displayed diffuse cytoplasmic EWS expression (Fig. 6f). These results were subsequently confirmed by IHC analysis of a tissue microarray (TMA) containing 64 sporadic and 94 LS-related CRCs: we observed a reduction of approx. 30% in EWS expression in both groups, but only the LS-related cancers showed diffuse cytoplasmic staining for EWS (61% vs 3% of the sporadic cancers). Thus, MMR-deficient CRCs, all carrying somatic EWS16T tract alterations, display a distinct subcellular EWS distribution pattern *in vivo*. Further studies are needed to assess if this is directly related to EWS16T tract alterations or, rather, an indirect consequence of MSI-associated genetic instability affecting e.g. methyl-transferases like PRMT1 known to regulate the localization of EWS⁹ by methylating

Glycine/Arginine-rich motifs located in the arginine-glycine-glycine domains of EWS.

In summary, EWS16T represents a novel, monomorphic MSI target locus which identifies both, hereditary and sporadic, MMR deficient cancers with 100% sensitivity and specificity. The contractions at this locus affect multiple regulatory mechanisms including alternative polyadenylation, mRNA / protein expression and possibly subcellular localization thereby implicating the RNA-/DNA-binding protein EWS, critical for the maintenance of genome integrity¹⁰, in MSI-associated colorectal tumorigenesis.

ACKNOWLEDGEMENTS

We thank the patients and their families for participating in this study and the treating physicians for providing pertinent clinical information and tumor specimens. In particular, we are indebted to Prof. G. Spagnoli, Dept. Biomedicine, University of Basel, for sharing his expertise on cell culturing and harvesting. We also thank Michèle Attenhofer, Nemja Boesch and Sibylle Bertschin for excellent technical assistance.

The study was supported by grants from the European Molecular Biology Organization (long-term fellowship 184-2009 to S.K.), the Academy of Finland (P.P.), the Finnish Cancer Organizations (P.P.), the Sigrid Juselius Foundation (P.P.), Biocentrum Helsinki (P.P.), the European Research Council FP7-ERC-232635 (P.P.), the University of Basel (M.Z.), the Krebsliga beider Basel (K.H.), the Krebsliga Zentralschweiz (K.H.) and Oncosuisse (K.H.).

AUTHOR CONTRIBUTIONS

S.K. performed the in vitro transcription assay, protein pull down assay, poly(A) site selection assay and luciferase assays; S.P. and M.K. performed the MSI analyses on the Swiss patients and quantitative real time PCR experiments; P.P and A.G performed the MSI analysis on the Finnish patients, S.P., F.T. and L.T. analyzed CRC whole tissue and microarray sections; H.A. and G.M. performed immunohistopathological analysis and F.W. FISH analysis on the extraosseous Ewing sarcoma; V.M. cultured and provided the MMR proficient and deficient colorectal cancer cell lines; S.P., S.K., M.K., M.Z. and K.H. analyzed the data and wrote the manuscript; all authors contributed and agreed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1

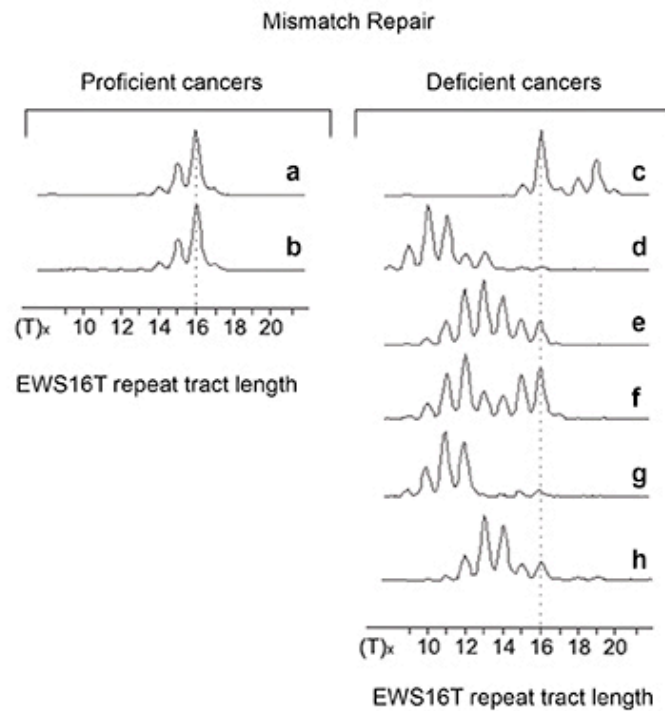


Figure 1: EWS16T tract instability in MMR deficient cancers as determined by capillary electrophoresis of fluorescently labeled PCR products. The dotted line corresponds to the wild-type allele (16 T). MMR-proficient cancers **(a)** colorectal, **(b)** gastric; MMR-deficient cancers with expansion **(c)** or contraction of EWS16T **(d-h)**: **(c)** colorectal, *MSH2* germline mutation carrier, **(d)** colorectal, *MSH2* mutation, **(e)** colorectal, *MLH1* promoter hypermethylation, and cancers with *MLH1* mutation **(f-h)**: **(f)** colorectal, **(g)** gastric, **(h)** endometrial.

Figure 2

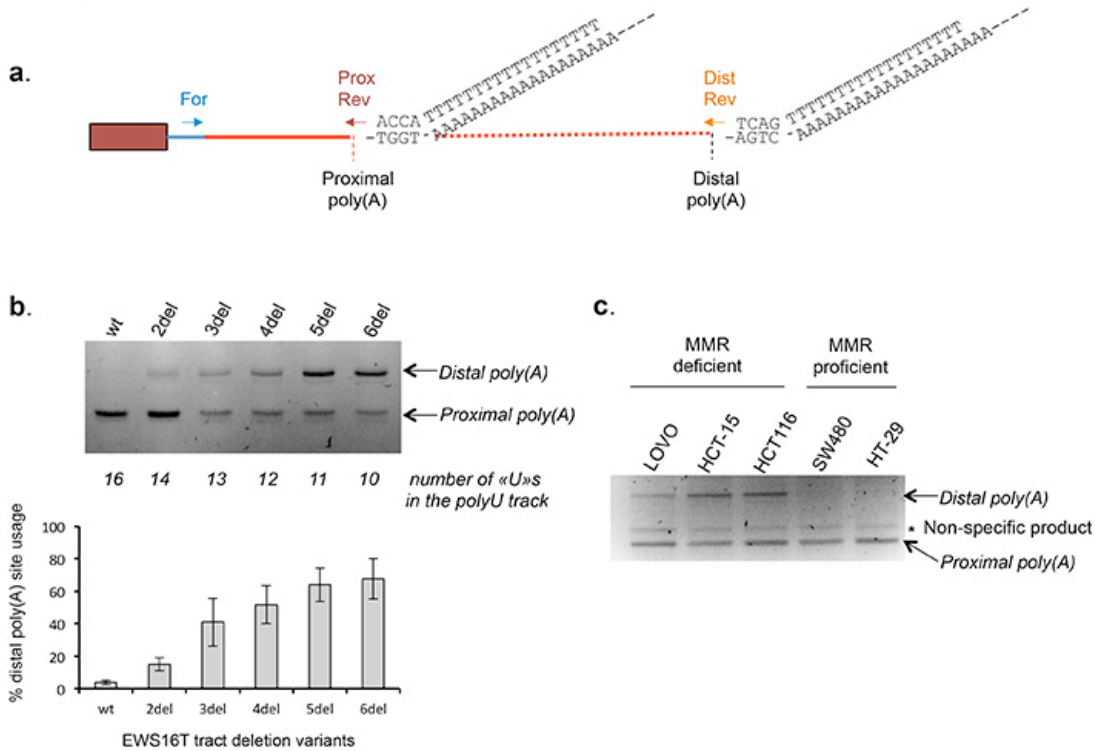


Figure 2: Poly(A) site selection assay through multiplexed PCR. **a)** Schematic diagram of the multiplexed semi-quantitative protocol for poly(A) site selection assay. A single forward primer, which was either vector specific (for analyzing poly(A) usage of reporter constructs) or EWSR1 -3'UTR specific (for analyzing endogenous EWSR1 poly(A) usage), was used along with two terminally anchored reverse primers to yield amplicons representing both long and short variants of EWSR1 3'UTR in a single PCR reaction. **b)** Poly(A) site usage in EWSR1 3'UTR constructs with variable poly(T/U) tracts lengths cloned in psiCHECK-2-SPAm vector along with the quantification of distal poly(A) site usage (percentages). **c)** Poly(A) site usage in endogenous EWSR1 gene across various MMR-deficient and -proficient cell lines.

Figure 3A

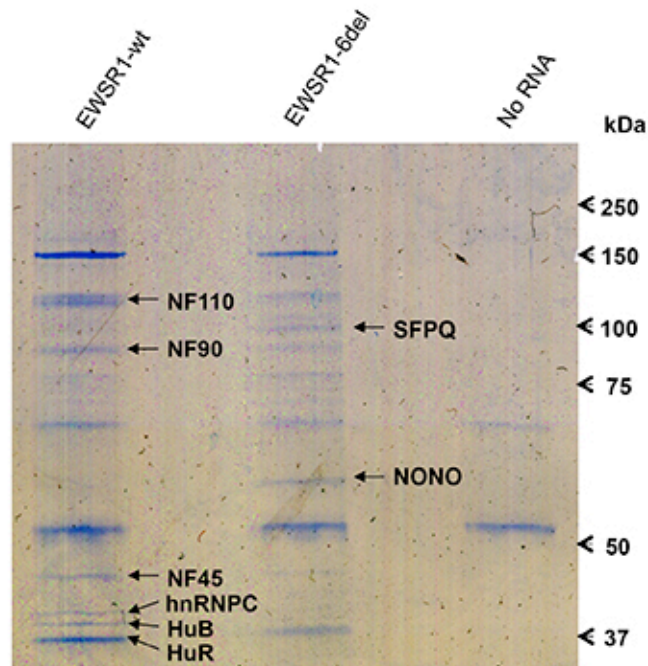


Figure 3: Identification of proteins associated with the 3'UTR of *EWSR1* in a pull down assay. **A)** Colloidal blue stained SDS PAGE gel showing proteins specifically interacting with *in vitro* transcribed, S1 aptamer-tagged *EWSR1*-wt or -6del 3'UTR. "No RNA" refers to the control sample where beads were incubated with cellular lysate without a prior incubation with RNA, allowing detection of any proteins interacting non-specifically with the MyOne Streptavidin beads. Bands specifically present in either wt or 6del lanes were excised and submitted to protein identification by mass spectrometry. Identified proteins have been labeled next to the corresponding bands.

Figure 3B

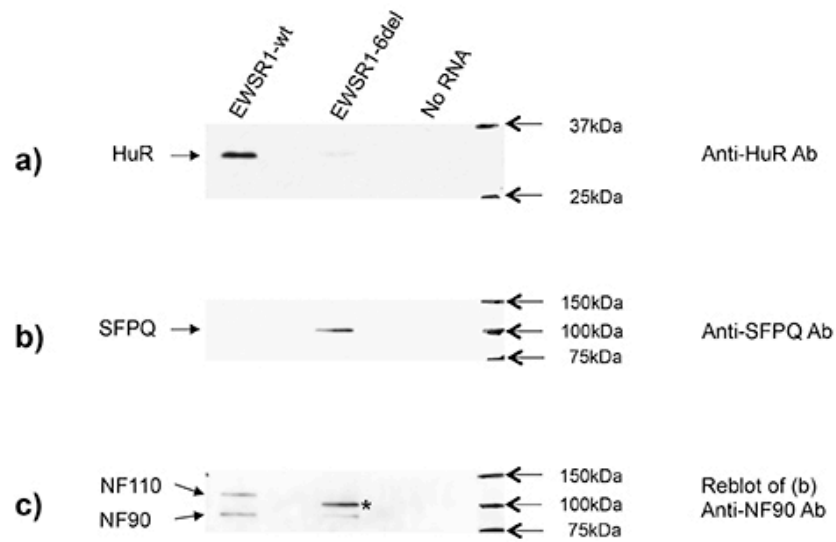


Figure 3: Identification of proteins associated with the 3'UTR of *EWSR1* in a pull down assay. **B)** Western blot confirming the results of mass spectrometry. Western blot with antibody raised against HuR (**a**), SFPQ (**b**) and NF90 (**c**) proteins. The membrane from (**b**) was used to reblot with anti-NF90 in (**c**). Asterix in (**c**) depicts the background signal from SFPQ in (**b**).

Figure 4

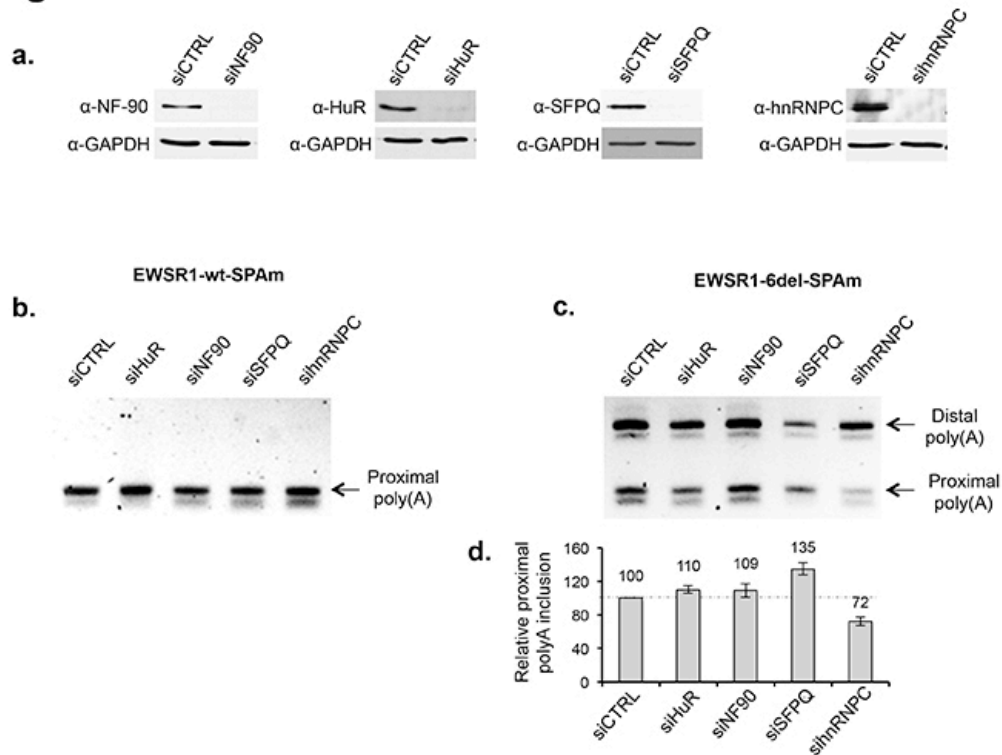


Figure 4: Knockdown of RNA binding proteins identified from pull down assays showing that SFPQ and hnRNPC specifically modulate poly(A) site usage in *EWSR1* 3'UTR reporter constructs. **a)** Western blot showing the knock down of the RNA-binding proteins after 72 hours. Poly(A) site usage in **b)** *EWSR1*-wt-psiCHECK-2-SPAm and **c)** *EWSR1*-6del-psiCHECK-2-SPAm reporter constructs upon knock down of RNA binding proteins. **d)** Quantification of the relative poly(A) site usage in **c)**.

Figure 5A

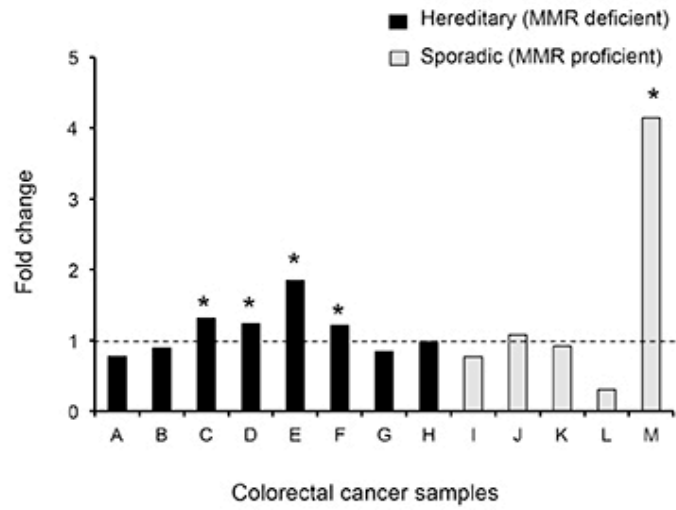


Figure 5: Quantitative Real-Time PCR showing *EWSR1* expression in tumor relative to matched normal tissue. **A)** Difference in polyA site usage in endogenous *EWSR1* mRNA (distal/proximal).

Figure 5B

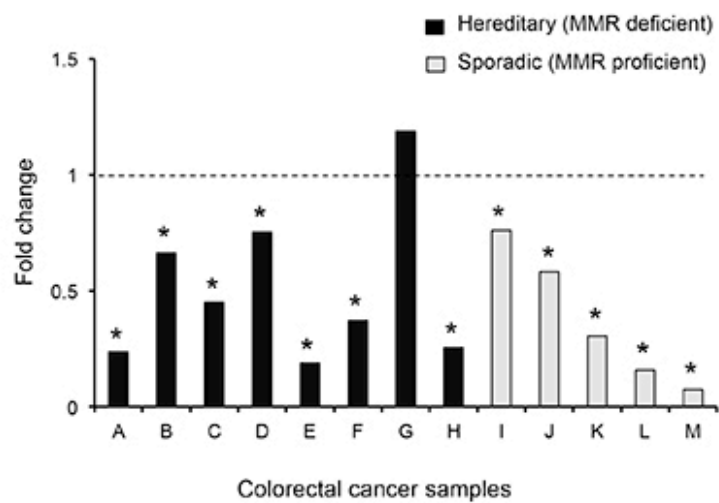


Figure 5: Quantitative Real-Time PCR showing *EWSR1* expression in tumor relative to matched normal tissue. **B)** difference in total *EWSR1* mRNA levels. Samples showing more than 1.2 fold difference are indicated with an asterix.

Figure 6

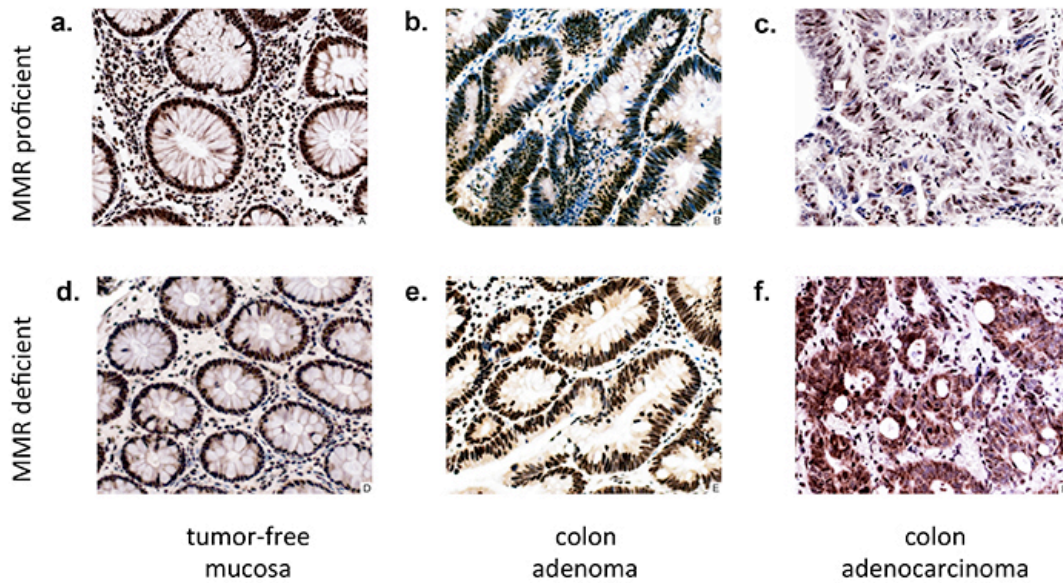


Figure 6: Immunohistochemical staining of EWS. **a)** Normal colonic mucosa from sporadic CRC patient. **b)** Sporadic, MMR proficient colon adenoma with reduced nuclear expression. **c)** Sporadic, MMR proficient colon adenocarcinoma with reduced nuclear expression. **d)** Normal colonic mucosa from Lynch syndrome-related CRC patient. **e)** MMR deficient colon adenoma with reduced nuclear expression. **f)** MMR deficient colon adenocarcinoma with reduced nuclear and diffuse cytoplasmic expression. CRC, colorectal cancer.

Tables
Table 1

Samples analyzed	N	MSI status	EWS16T Status wildtype n (%)	EWS16T Status contraction n (%)	EWS16T Status expansion n (%)
MMR deficient cancers	128	MSI-H	0	126 (98.4)	2 (1.6)
• Colorectal	104	MSI-H	0	102 (98.04)	2 (1.96)
• <i>MLH1</i> deficient	76	MSI-H	0	76 (100)	0
• <i>MSH2</i> deficient	28	MSI-H	0	26 (92.86)	2 (7.14)
• Gastric	17	MSI-H	0	17 (100)	0
• <i>MLH1</i> deficient	16	MSI-H	0	16 (100)	0
• <i>MSH2</i> deficient	1	MSI-H	0	11 (100)	0
• Endometrial	7	MSI-H	0	7 (100)	0
• <i>MLH1</i> deficient	4	MSI-H	0	4 (100)	0
• <i>MSH2</i> deficient	3	MSI-H	0	3 (100)	0
Colon adenomas	37	MSS	37 (100)	0	0
• Lynch syndrome-related	10	MSS	10 (100)	0	0
• <i>MLH1</i> deficient	8	MSS	8 (100)	0	0
• <i>MSH2</i> deficient	2	MSS	2 (100)	0	0
• Sporadic	27	MSS	27 (100)	0	0
MMR proficient cancers	190	MSS/ MSI-L	190 (100)	0	0
• Colorectal	123	MSS	123 (100)	0	0
• Colorectal	12	MSI-L	12 (100)	0	0
• Gastric	55	MSS	55 (100)	0	0

Table 1: EWS16T tract instability in 355 cancer samples. Microsatellite instability was determined according to¹¹. MMR MSI-H: MSI-high, MSI-L: MSI-low, MSS: microsatellite stable.

Methods

DNA and RNA Isolation

To isolate genomic DNA and total RNA from cell lines and fresh/frozen tumor tissue samples Qiagen QIAamp DNA/RNA Mini Kit (Qiagen, Hombrechtikon, Switzerland) and for formalin fixed paraffin embedded (FFPE) tumor samples RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Invitrogen, Carlsbad, CA, USA) were used according to the manufacturers' guidelines.

Analysis of Microsatellite Instability (MSI)

Microsatellite instability was assessed on two independent cohorts of patients encompassing 85 Lynch syndrome related cancers (78 CRCs and 7 endometrial carcinomas), 113 sporadic CRCs, including 14 cases with MLH1 promoter hypermethylation and 12 MMR proficient, MSI-low cancers, and a Finnish cohort including 8 Lynch syndrome gastric related cancers, 50 sporadic CRCs and 64 sporadic gastric cancers. Based on the recommendations of the National Cancer Institute workshop on MSI, a panel of microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250)¹² and two additional microsatellite markers (BAT40, MYCL1) were used to determine MSI status. The 3'UTR poly T(16) tract (EWS16T) of the *EWSR1* gene (Ewing sarcoma breakpoint region 1; RefSeq: NM_005243) was amplified by PCR with the following primers: 5'- AATGTTTCATGGTTGTGATGT-3' (forward FAM-labeled) and 5'-GAAGGATGACTCTTTATAA-3' (reverse). PCR products were analyzed on an ABI 310 Genetic Analyzer with GeneScan Analysis V 3.1 (PE Applied Biosystems, Foster City, CA) and Genotyper 2.0 (PE Applied Biosystems, Foster City, CA) software. Fragment analysis of PCR products allowed determination of novel alleles (expansions or contractions) within the repetitive tract of a given marker. The observed expansions or contractions in EWS16T were verified using a second set of primers covering the locus (Forward primer: 5'-GCATGCTCAGTATCATTGTGG-3'; Reverse primer: 5'-AGGCCGAGAAGGATGACTCT-3') for sequencing analysis of selected samples. Sequencing reactions using the Big Dye terminator chemistry (Applied Biosystems, Foster City, CA) were performed according to the manufacturer's protocol.

Relative expression of *EWSR1* by qPCR

EWSR1 mRNA expression on fresh frozen tissues (8 Lynch syndrome related CRCs and 5 sporadic CRCs both matched with their tumor free mucosa) was assessed using the TaqMan® Probe-Based Gene Expression Analysis (Applied Biosystems, Foster City, CA), and the *EWSR1* probe Hs01580532_g1 (Applied Biosystems, Foster City, CA). The measurements were normalized using the *HPRT1* probes Hs01003270_g1 (Applied Biosystems, Foster City, CA)^{13,14}, and the fold-changes in gene expression were calculated using the standard $\Delta\Delta C_t$ method¹⁵. All retrotranscriptase reactions, including no-template controls, were run on an Applied Biosystem 7900HT thermocycler. Each sample was tested in triplicate unless specified otherwise.

Immunohistochemistry

Several cohorts of patients were studied by immunohistochemical analysis of EWS. Briefly, the tissue samples of the following cohorts of patients were analysed: 37 colon adenoma (9 of which Lynch syndrome related), 19 CRCs (10 of which Lynch syndrome related) and a tissue microarray¹⁶. Patient data including complete follow-up were obtained by retrospective analysis of medical records, regional tumor registries and/or treating physicians. Tissue samples were obtained by surgical or endoscopic excision. Tissues sections of 4 μ m sections of paraffin embedded tissue were immunostained for primary antibody against EWS (Abcam clone 84389 dilution 1:800). Staining was carried out as previously described¹⁷. Immunoreactivity was scored semi-quantitatively by evaluating the number of positive tumor cells over the total number of tumor cells. Nuclear immunoreactivity scores were assigned using 5% intervals and ranged from 0% to 100%. Regarding cytoplasmic expression, the staining intensity was scored as described by Allred *et al.*¹⁸. All samples were examined independently by three different pathologists (S.P., F.T. and L.T.), blinded to clinicopathological and molecular genetic information.

Cell lines

Five colorectal cancer cell lines from the American Type Culture Collection (ATCC, Rockville, MD) were used for this study: four three repair deficient cell lines (HCT116, LoVo, HCT15) and two mismatch repair proficient (SW480 and HT29). HCT116, HCT15, cells (ATCC, Rockville, MD) were cultured in RPMI 1640 (Invitrogen Basel, Switzerland) supplemented with 10% fetal bovine serum FBS, 1% Kanamycin sulphate, 1% GlutaMAX-I, 1% Sodium Pyruvate, 1% non Essential Amino Acids (NEAA), 1% HEPES (all from Invitrogen Basel, Switzerland) and 0.1% 2-mercapto-ethanol (Sigma-Aldrich Basel, Switzerland). HT29 cells were grown in McCoy's 5A Medium (Invitrogen Basel, Switzerland) with 10% fetal bovine serum FBS, Kanamycin sulphate and GlutaMAX-I (all from Invitrogen Basel, Switzerland). SW480 cells were cultured in L-15 Medium (Sigma-Aldrich Basel, Switzerland) with 10% FBS, 1% GlutaMAX-I and 1% Kanamycin sulphate (all from Invitrogen Basel, Switzerland). Cells were maintained at 37°C with 5% CO₂.

HeLa cells at earlier passages were cultured in DMEM with GlutaMAX (Invitrogen) supplemented with 10% fetal bovine serum FBS. Cells were maintained at 37°C with 5% CO₂.

Plasmids and antibodies

Synthetic poly(A) site of Renilla luciferase gene in dual luciferase psiCHECK-2 vector was mutated using primers psi-ck2polyAmutF: 5'-GCGGCCGCTGGCCGCAGCTAAATATCTTTATTTTCA-3' and psi-ck2polyAmutR: 5'-TGAAAATAAAGATATTTAGCTGCGGCCAGCGGCCGC-3' using QuikChange™ Site-Directed Mutagenesis Kit as per manufacturer's instruction to generate psiCHECK-2-Synthetic polyA mutant (psiCHECK-2-SPAm). *EWSR1* 3' UTR was PCR amplified from HEK293 genomic DNA using primers EWSR1_xhoIF: 5'-CCGACTCGAGCGGCCCTACTAGATGCAGAG-3' and EWSR1_notIR: 5'-ATAAGAATGCGGCCGCGAACCAACCGTTTACCTGGA-3'. The PCR amplicons were cloned into pGEM-T Easy and subsequently subcloned into psiCHECK-2-SPAm using *XhoI* and *NotI* restriction site to generate *EWSR1*-wt-3'UTR/psiCHECK-2-SPAm reporter. Mutant constructs with U deletions (*EWSR1*-2/3/4/5/6 del-3'UTR/psiCHECK-2-SPAm) and insertions (*EWSR1*-

2/4/ins-3'UTR/psiCHECK-2-SPAm) were introduced using standard overlap extension PCRs.

Antibodies for Western blots against NF90, SFPQ and HuR were obtained from Santacruz Biotech and Antibody against EWSR1 for Western blots and Immunoprecipitation were obtained from Abcam.

In vitro transcription

In-vitro transcription for pull down assay using S1 aptamer was performed using T7 RiboMAX™ Express Large Scale RNA Production System (Promega) as per manufacturer's instructions. Region flanking the wt- and deleted poly T/U tracks in the 3'UTR of *EWSR1* constructs was amplified using primers T7_ *EWSR1*_IVT_Cf: 5'-GCTTCTAATACGACTCACT ATAGGGAGAAATGGGAACCCCTTGTGAG-3' and *EWSR1*_IVT_Cr: 5'-GAACAGAGGCCGA GAAGGAT-3' to introduce T7 promoter sequences at the 5' end of the amplicons. S1 aptamer sequence was introduced at the 3' end using another round of PCR using T7_ *EWSR1*_IVT_Cf: 5'-GCTTCTAATACGACTCACTATAGGGAGAAATGGGAACCCCTTGT GAG-3' and *EWSR1*_S1apt_IVT_Cr: 5'-CATGCCCCGGCCCGCGACTATCTTACGCACTTG CATGATTC TGGTCGGTCCCATGGATCCGAACAGAGGCCGAGAAGGAT-3'.

Protein Pull-Down Assay

For each sample, 100ul of the MyOne Streptavidin Dynal beads (Invitrogen) was washed twice with one bead volume of solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 0.05 M NaCl) and once with one bead volume of Solution B (DEPC-treated 0.1 M NaCl) and once with RNA binding buffer. The beads were resuspended in one bead volume of RNA binding buffer (100mM NaCl, 50mM Hepes 7.5, 0.5% NP-40 and 10mM MgCl₂) with 100ug of in-vitro transcribed RNA with S1 aptamer sequence and incubated at 10°C for 40 minutes in a thermomixer with intermittent shaking. The beads were washed twice with one volume of RNA wash buffer prior to incubation with the lysate. HEK293 cell pellet from 15cm² dish was lysed in 3ml native lysis buffer (25mM Hepes-KOH pH 7.5, 100mM KCl, 0.5% NP-40, 5mM MgCl₂, 0.5mM DTT , protease inhibitor cocktail, 1mM NaF, 1 mM Na₄VO₄ and 300U of RNasin) for

15 minutes on ice. The lysate was subsequently gently sonicated and centrifuged to remove any cell debris. 200ug of *E. coli* tRNA was additionally added to prevent non-specific binding of proteins to the beads. 1 ml of the lysate was added to the beads coupled to S1 aptamer RNA and also to the beads alone for no RNA control. The mixture was incubated at 4°C on a rotation wheel. After 1 hour of incubation, the beads were washed thrice with Native lysis buffer. The bound proteins were eluted with 100ul Native lysis buffer supplemented with 25mM Biotin for 30-45 minutes at 10°C on a thermomixer with intermittent shaking. 900ul of 100% ethanol was added to the eluate and incubated at -80°C for 2 hours followed by centrifugation to precipitate the eluted proteins. The pellet was air dried and dissolved in 35ul of (SDS loading dye). Prior to loading on the NuVex gradient gels, the samples were heated at 90°C for 5 minutes. After SDS-PAGE electrophoresis, the gel was stained with colloidal blue and bands of interest were excised and sent for mass spectrophotometry.

Poly(A) site selection assay

HeLa cells were transfected with psiCHECK-2-SPAm constructs for 24 hours. Total RNA was isolated from the HeLa cells using TriReagent (Sigma) followed by DNase I (Promega) treatment according to the manufacturer's protocol. Reverse transcription was done with oligo d(T)₁₈ primers for 1 hour. For poly(A) site selection assay, multiplexed polymerase chain (PCR) reaction was set up using a single forward primer specific to the psiCHECK-2 vector (psiCHECK-2-SeqFor: 5'-ATGAAATGGGTAAGTACA-3') and two 4 nucleotide terminally anchored oligo d(T)₁₆ reverse primers specific to the two isoform variants of EWSR1 (EWSpolyAproxR: 5'-TTTTTTTTTTTTTTTTTACCA-3' and EWSpolyAdistR: 5'-TTTTTTTTTTTTTTTTTGACT-3' respectively) to detect only the isoforms generated specifically from the psiCHECK-2-SPAm constructs. PCR was run for 28 cycles and the products were separated on a 2 % Agarose gel. The bands were quantified using the ImageJ software (<http://rsbweb.nih.gov/ij/>).

Luciferase assays

HeLa cells were seeded in a 48 well plate one day prior to transfection. 0.2ug of plasmids (psiCHECK-2-SPAm constructs) were transfected with Lipofectamine 2000 (Invitrogen) for 24 hours. Luciferase assays were done on the transfected cells using The Dual-Luciferase® Reporter Assay System (Promega). Both transfections and luciferase assays were done according to the manufacturer's protocol.

siRNA transfections

Control-siRNAs and siRNAs against hnRNPC, HuR, NF90 and SFPQ and were obtained from Santa Cruz Biotechnology. HeLa cells were reverse transfected with siRNA oligos using RNAiMAX (Invitrogen). After 48 hours the cells were transfected with EWSR1-wt-3'UTR- and *EWSR1-6del-3'UTR*-psiCHECK-2-SPAm-constructs for another 24 hours. The cells were subsequently harvested and split into two aliquots. One aliquot was used to assess the knockdown efficiency of siRNA using Western blot, while the other was used for RNA isolation and subsequent poly(A) site selection assay.

Statistical Analyses

For statistical analysis, the chi-square test (χ^2 test) and Fisher's exact test for nonparametric variables and Student's t-test for parametric variables were used, with all probabilities reported as 2-tailed, considering a $P < 0.05$ to be statistically significant. Calculations were performed using the software program SPSS 17.0 (IBM Corporation, Somers, NY 10589).

Ethical approval

The study is part of the so-called "Basler Studie über familiaere Tumorkrankheiten", Ref.Nr.EK: 258/05 and has been approved by the "Ethikkommission beider Basel". Furthermore, written informed consent was obtained from all Lynch syndrome patients as well as from the sporadic patients.

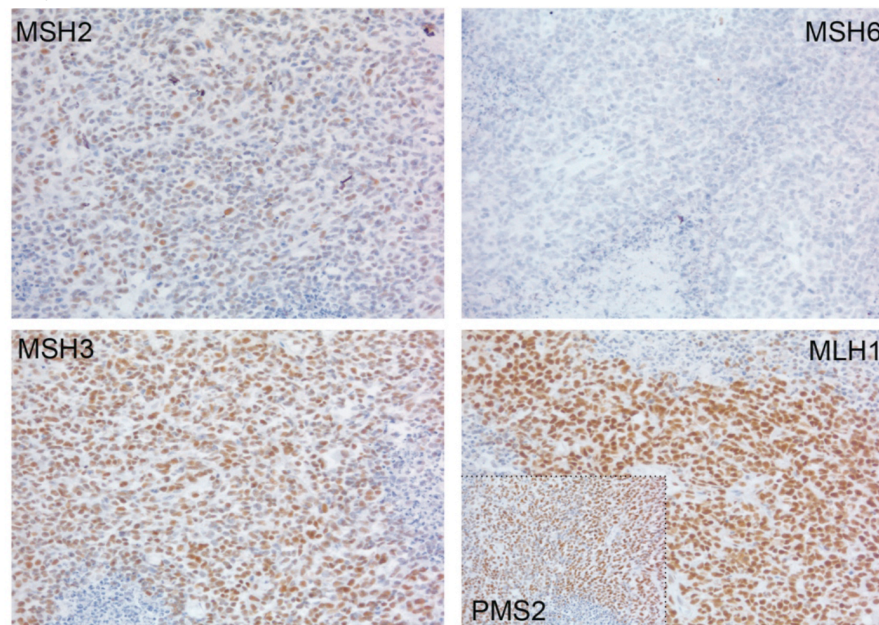
References

1. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087 e2073 (2010).
2. Suraweera, N., *et al.* Conservation of mononucleotide repeats within 3' and 5' untranslated regions and their instability in MSI-H colorectal cancer. *Oncogene* **20**, 7472-7477 (2001).
3. di Pietro, M., *et al.* Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. *Gastroenterology* **129**, 1047-1059 (2005).
4. Wheeler, D.A., *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-876 (2008).
5. Mayr, C. & Bartel, D.P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673-684 (2009).
6. Martin, G. Genome wide analysis of pre-mRNA 3' end processing. *Cell reports* **In press** (2012).
7. Kuwano, Y., *et al.* NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res* **38**, 225-238 (2010).
8. Lutz, C.S., Cooke, C., O'Connor, J.P., Kobayashi, R. & Alwine, J.C. The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. *Rna* **4**, 1493-1499 (1998).
9. Araya, N., *et al.* Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochem Biophys Res Commun* **329**, 653-660 (2005).
10. Paronetto, M.P., Minana, B. & Valcarcel, J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Molecular cell* **43**, 353-368 (2011).
11. Umar, A., *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* **96**, 261-268 (2004).
12. Boland, C.R., *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition:

- development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**, 5248-5257 (1998).
13. de Kok, J.B., *et al.* Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* **85**, 154-159 (2005).
 14. Dydensborg, A.B., Herring, E., Auclair, J., Tremblay, E. & Beaulieu, J.F. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. *Am J Physiol Gastrointest Liver Physiol* **290**, G1067-1074 (2006).
 15. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* **25**, 402-408 (2001).
 16. Zlobec, I., *et al.* Prognostic and predictive value of TOPK stratified by KRAS and BRAF gene alterations in sporadic, hereditary and metastatic colorectal cancer patients. *Br J Cancer* **102**, 151-161 (2010).
 17. Lugli, A., Tzankov, A., Zlobec, I. & Terracciano, L.M. Differential diagnostic and functional role of the multi-marker phenotype CDX2/CK20/CK7 in colorectal cancer stratified by mismatch repair status. *Mod Pathol* **21**, 1403-1412 (2008).
 18. Allred, D.C., Harvey, J.M., Berardo, M. & Clark, G.M. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* **11**, 155-168 (1998).

Supplementary figures and legends
Supplementary Figure 1

Figure S1

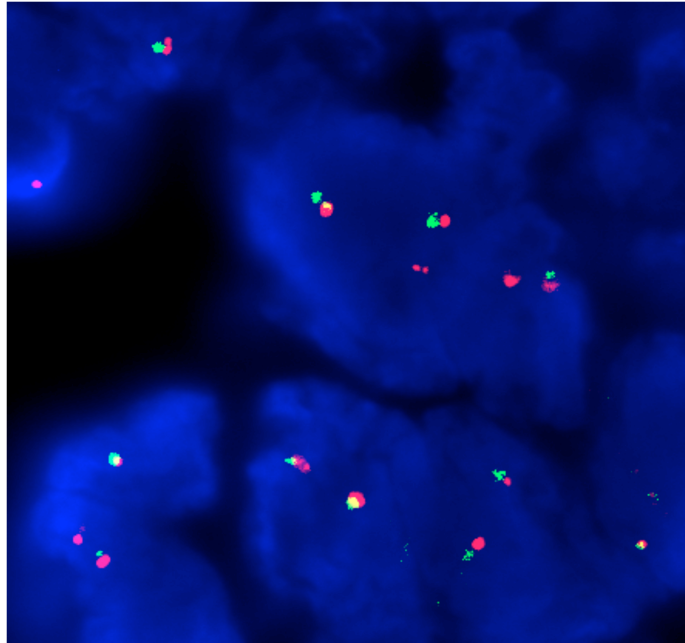


Extra-osseous Ewing sarcoma

Supplementary Figure 1: Immunohistochemical staining of DNA mismatch repair proteins MLH1, MSH2, MSH3, MSH6 and PMS2 in an extra-osseous Ewing sarcoma of a *MSH6* mutation carrier demonstrating selective loss of MSH6 expression.

Supplementary Figure 2

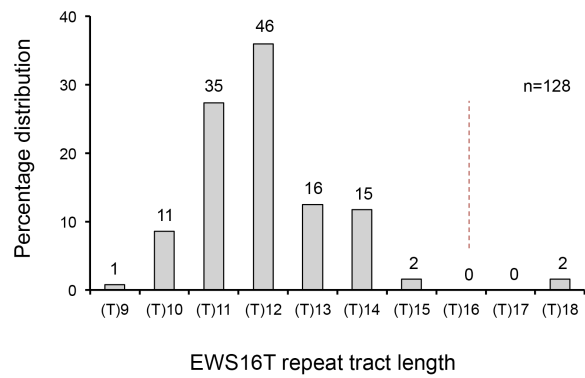
Figure S2



Supplementary Figure 2: Representative extract from fluorescence in situ hybridization (FISH) analysis of the *MHS6*-related extra-osseous Ewing sarcoma specimen using an *EWSR1*-specific break-apart probe (Poseidon™ probe KI-10708 EWS Break). Screening of 200 nuclei gave no evidence for the presence of a rearrangement involving the 22q12 locus (i.e. no evidence for any split red-green signals).

Supplementary Figure 3

Figure S3

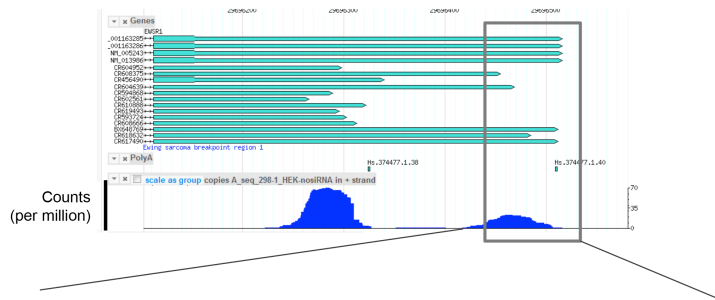


Supplementary Figure 3: Frequency distribution of EWS16T repeat alterations in 128 MMR deficient cancers (104 colorectal, 17 gastric and 7 endometrial cancers). Numbers on top of the bars represent percentages. The dotted line corresponds to the wild-type allele (T)16.

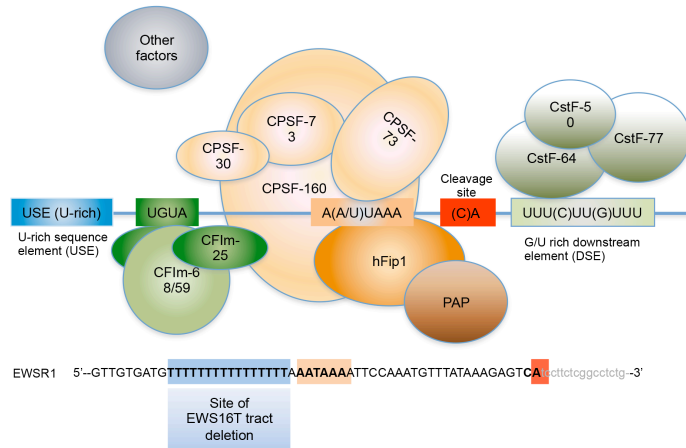
Supplementary Figure 4

Figure S4

a.



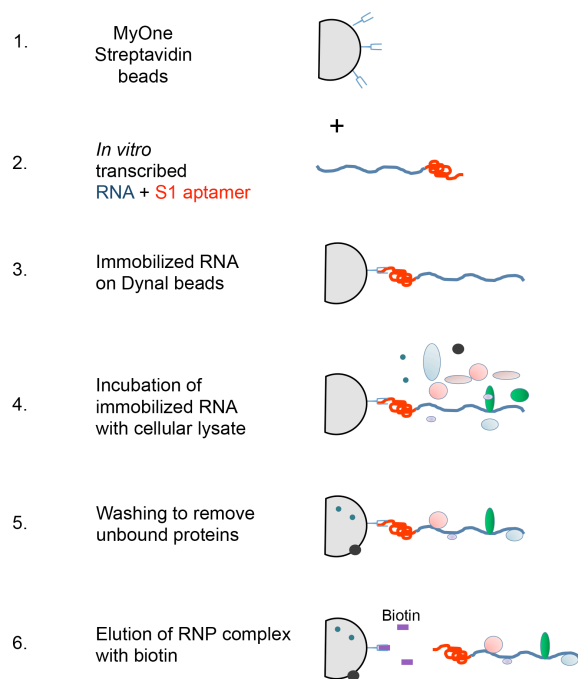
b.



Supplementary Figure 4: The 3'UTR of the *EWSR1* gene harbors two poly(A) signals resulting in the generation of *EWSR1* isoforms with short and long UTRs. **a)** Reads (per million) obtained from poly(A) seq⁷ showing the location of the two poly(A) sites in HEK293 cells, with the proximal poly(A) site being predominantly used over the distal. **b)** Schematic diagram showing the location of the EWS16T tract with respect to the distal poly(A) signal (AAUAAA), and the binding specificity of important 3' end processing factors.

Supplementary Figure 5

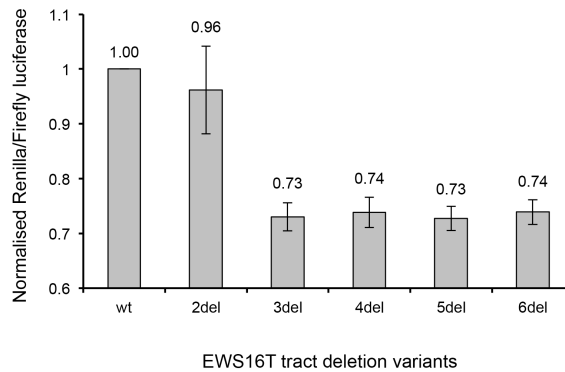
Figure S5



Supplementary Figure 5: Schematic diagram depicting essential steps of the protein pull down assay using *in vitro* transcribed S1 aptamer-tagged RNA. For a more detailed protocol please refer to the Materials and Methods section of the manuscript.

Supplementary Figure 6

Figure S6



Supplementary Figure 6: Luciferase assay showing the changes in Renilla luciferase protein expression from psiCHECK-2-SPAm constructs harboring different EWS16T tract deletion variants in its multiple cloning site. The signals are normalized against internal Firefly luciferase.

6.2 The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation.

My contribution to this work:

- DNA extraction from sporadic colorectal cancers;
- *SH2D4A* mutation analysis by Sanger sequencing of CRC samples;

- Loss of heterozygosity analyses on the CRCs;
- Quantitative real time PCR experiments;
- Immunohistochemical analysis of CRC whole tissue and microarray sections;
- Statistical and Data analysis;
- Manuscript writing

The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation

Sébastien Loeffler^{1,2}, Michal Kovac^{1,2}, Stefan Weis¹, Salvatore Piscuoglio^{1,2}, Saulius Zuklys^{1,2}, Marcel Keller^{1,2}, Katrin Hafen^{1,2}, Daniel Hess³, Kaspar

Truninger¹, Inti Zlobec⁴, Luigi Terracciano⁴, Karl Heinemann^{1,2}, Primo Schär¹
and Georg A. Holländer^{1,2, 5}

Affiliations

1: Department of Biomedicine - Mattenstrasse, University of Basel, 4058 Basel, Switzerland

2: Basel University Children's Hospital, 4058 Basel, Switzerland

3: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

4: Institute of Pathology, University of Basel, 4031 Basel, Switzerland

5: The Department of Paediatrics and The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom

6: Current address: Actelion Pharmaceuticals Ltd, Allschwil, Switzerland

Corresponding author: Georg A. Holländer (georg-a.hollaender@unibas.ch)

Tel: +41 61 695 30 69 / +44 1865234 225

Running title: **SHOCA-2 is a tumor suppressor in colorectal cancer**

Summary

Chromosomal deletions at 8p have been associated with colorectal cancer (CRC), but the tumor suppressor function affected has not been described. Here, we identify *SH2D4A* as a suppressor of CRC mapping to 8p21.3. *SH2D4A* encodes SHOCA-2, a SH2 domain-containing adapter protein that, upon EGF signaling, interacts with the serine/threonine phosphatase PP1b and consequently inhibits STAT3 activity. Conversely, knock-down of SHOCA-2 stimulates c-Myc, Cyclin D1 and Jun B expression and accelerates CRC growth through unopposed STAT3 activity. Loss of SHOCA-2 expression in CRC correlates with advanced disease stages and poor prognosis and is caused by chromosomal rearrangements, point mutations and epigenetic changes. Thus, SHOCA-2 functions as an inhibitor of the EGFR pathway and its absence impacts on CRC development and progression.

Significance

Colorectal cancer (CRC) yearly claims more than 655,000 lives worldwide and represents the third most frequent cause of death in the Western world. Though a 8p21-23 chromosomal deletion has been linked to CRC development and progression, the molecular mechanisms related to this genomic region have not been recognized to date. Encoded on 8p21.3, the SH2 domain-containing SHOCA-2 protein represents a hitherto unrecognized CRC tumor suppressor that limits cell proliferation via a negative feedback loop controlling EGFR signaling. Consequently, a loss of SHOCA-2 expression results in the oncogenic activation of STAT3, a hallmark of many CRC. Since drugs inhibiting EGFR and STAT3 activity offer only limited efficiency, enhancing SHOCA-2 function constitutes a novel insight for the treatment of CRC.

Introduction

Oncogene activation and loss of tumor suppressor activity are responsible for cancer initiation and progression (Vogelstein and Kinzler, 2004). Overexpression of the epithelial cell growth factor receptor (EGFR) is frequently observed in colorectal cancer (CRC), and correlates with a poor clinical outcome (Saif, 2010). EGFR is receptor tyrosine-kinase that relays its activation through the signal transducer and activator of transcription 3 (STAT3) to the nucleus controlling the transcription of genes involved in cell proliferation and survival (Aggarwal et al., 2009). Constitutive activation of the STAT3 signaling pathway as a consequence of genetic mutation is oncogenic, and a hallmark of many CRC (Bromberg et al., 1999; Klampfer, 2008).

Genetic and epigenetic alterations in the expression of tumor suppressor genes are typical features of neoplastic transformation and contribute to both initiation and progression of tumor formation (Weinberg, 1991; Vogelstein and Kinzler, 2004). The loss of heterozygosity (LOH) across variable parts of chromosome 8p has been identified as a frequent abnormality associated with CRC (Emi et al., 1992) and linked to DNA breakage at fragile sites located at 8p12 and 8p22 (Birnbaum et al., 2003). Although predicted from chromosome transfer studies (Tanaka et al., 1996), specific tumor suppressor genes have to date not been identified in this critical chromosomal region. It is thus unknown how genetic alterations in 8p contribute to CRC development and progression.

Here we report on the identification of a new tumor suppressor, termed SHOCA-2. Encoded on 8p21.3, SHOCA-2 interacts with the serine/threonine phosphatase PP1b and consequently inhibits STAT3 activity to control EGFR signaling. Moreover, we characterize the genetic and epigenetic mechanisms that result in a silencing of SHOCA-2 in CRC.

Results

Identification of SHOCA-2 and its loss of expression in colorectal cancer

cDNA Representational Difference Analysis (cDNA-RDA) comparing epithelial cells at two sequential stages of embryonic maturation identified transcripts of a previously uncharacterized gene, designated *SH2D4B* (NCBI gene ID 387694; S.Z. and G.H., unpublished data). Homology searches identified *SH2D4A* as a paralogue (NCBI gene ID 63898; Figure 1A). Both genes are evolutionary conserved (Figure S1A) and ubiquitously expressed, with *SH2D4A* being significantly more abundant than *SH2D4B* (Figure S1B). The proteins encoded by *SH2D4A* and *SH2D4B* display 65% amino acid sequence identity at their C- and N-termini and share several structural motifs including three coiled-coil domains in the middle of the molecule and a Src homology (SH2) domain at the C-terminal end (Figure 1A; Figure S1C). It is for these motifs and the putative function as an adaptor that we named this family of proteins SHOCA, an acronym for SH2-domain containing adaptor protein.

SH2D4A maps to human chromosome 8p21.3 (Figure 1A) and encodes SHOCA-2. Because the chromosomal region surrounding *SH2D4A* (8p21-23) is frequently lost in CRC and other epithelial tumors (Emi et al., 1992), we determined by immunohistochemistry the expression of SHOCA-2 in 400 consecutive, sporadic CRCs (Figure S1D). The significant differences in detectable SHOCA-2 protein correlated with the separate clinical tumor stages (I-IV) (Figures 1B and 1C; Table S1A and S1B); SHOCA-2 expression was reduced or lost of with advanced CRC pathology. Patients with tumors displaying either low or absent SHOCA-2 expression had a poorer outcome (log-rank test $p=0.0118$; Figure 1D; Table S1C) though SHOCA-2 expression, when stratified by disease stage, did not serve on its own as a strong predictor of survival. An identical conclusion was drawn from a Cox regression analysis ($p=0.0134$ with HR (95%CI)=0.56 (0.35-0.89) and a Wilcoxon test ($p=0.0133$) (Table S1C). Gene expression data available in public databases for carcinomas of different origin including bladder, breast, head and neck, esophagus, liver and ovary suggested similar correlations between disease

stage and SHOCA-2 expression (Figure S1E). Finally, further analysis of SHOCA-1 expression (mRNA levels) in CRC samples with absent or reduced SHOCA-2 expression revealed a concurrent loss of SHOCA-1 in approximately half of the samples whereas SHOCA-1 loss was neither observed in healthy mucosa nor in CRC tissues proficient for SHOCA-2 (Figure S1F).

Genetic and epigenetic alterations of the *SH2D4A* locus in CRC

To determine the molecular basis for decreased SHOCA-2 expression, we chose from an unselected cohort of 70 CRC patients 27 subjects with *SH2D4A* alleles that could be distinguished by microsatellite markers and SNPs (Figure S1D). In 17 of these (63%), the primary tumor had lost or diminished SHOCA-2 expression whereas only 3 samples (11%) displayed an *EGFR* gene amplification. A loss of the *SH2D4A* gene and a simultaneous amplification of *EGFR* were noted in 2 out of these 3 individuals. This result is in accordance with the observation that *EGFR* over-expression in CRC is rarely caused by *EGFR* gene amplification (Saif, 2010). Using 3 flanking microsatellite markers and 6 gene-specific SNPs (Figure S2A), we detected LOH at *SH2D4A* in 7 of the 17 tumors (41%). Gene dosage quantification of the short arm of chromosome 8 indicated mono- and biallelic deletions in 6 and 1 of these LOH tumors, respectively (Figure 2A). When normal mucosa was examined, four of the six patients with a monoallelic *SH2D4A* loss in tumor tissue were heterozygous for the intronic SNP rs17128221 (c.342-5T>C). This SNP is located in a mRNA splice acceptor site and causes non-canonical exon 4 skipping, which in turn causes a translational frame shift and premature termination in exon 5 (Figure 2B). Consistent with a pathological effect, the cancer tissue of patients heterozygous for the C allele uniformly demonstrated a selective loss of the T allele (Chi square $p=0.006$) whereas homozygosity for the C allele was never observed in mucosa tissue samples of 83 healthy Caucasian individuals (Figure 2B).

Missense mutations and deletions in the SH2 domain of SHOCA-2 were observed in three of the 10 *SH2D4A* biallelic CRC samples characterized by a low or absent SHOCA-2 expression. These alterations either produced single

amino acid changes (p.Arg324Trp and p.Ser430Phe) with predicted structural effects (Polyphen-2 score of >0.95, SIFT score<0.02), or caused a translational frame shift (p.Ile378fsX15) giving rise to a loss of amino acids 377 to 454 and reduced protein expression (Figure 2C).

To determine causes for low or absent SHOCA-2 expression other than LOH and gene mutations, we next analyzed primary CRC tissue and surrounding healthy mucosa for DNA cytosine methylation in the 5' untranslated region (5'UTR) of *SH2D4A* (Figure 2D). Pyrosequencing of bisulfite-converted genomic DNA (Bettstetter et al., 2007) revealed two specific CpG dinucleotides showing increased methylation in cancer tissue when compared to matched normal mucosa. One of these CpGs is within a canonical Sp1 transcription factor recognition motif and its methylation significantly reduced the Sp1 binding affinity (Figure S2B). Epigenetic changes at the DNA level may therefore represent another cause for reduced *SH2D4A* gene expression in CRC. Taken together, in 15 of 17 CRCs investigated in molecular detail (Figure 2E), distinct genetic and epigenetic alterations explaining the loss or reduction in SHOCA-2 expression could be identified: chromosomal deletions at 8p21.3, genetic mutations in the *SH2D4A* gene, a mRNA splicing defect associated with a specific SNP, and site-specific changes in promoter DNA methylation.

SHOCA-2 negatively regulates EGF-induced STAT3 phosphorylation and transcriptional activity

A proteomic analysis of phospho-tyrosine signaling in non-small-cell lung cancer cell lines driven by EGFR-activating mutations identified a target peptide with sequence homology to SHOCA-2 (Guo et al., 2008). We confirmed that EGFR activation triggered tyrosine phosphorylation of SHOCA-2 in HeLa cells (Figure 3A). To investigate a potential involvement of SHOCA-2 in the activation of the downstream EGFR effectors STAT3, ERK and PI3K-Akt, we stimulated HeLa cells with EGF. The ectopic expression of SHOCA-2 in these cells specifically decreased the STAT3 phosphorylation at Tyr705 (Figure 3B). The phosphorylation of this residue constitutes an important post-translational modification of STAT3 as it triggers its

dimerization, nuclear translocation and DNA binding (Heinrich et al., 2003). By contrast, the relative level of Akt phosphorylation was increased whereas the MAP kinase pathway remained unaffected in EGF simulated HeLa cells (Figure S3A). Because SHOCA-2 had previously been suggested to be associated with the ER alpha/PKC signaling pathway (Li et al., 2009), we were unable to observe in ER α -negative HeLa cells (Arao et al., 2011) overexpressing SHOCA-2 a change in PKC phosphorylation further confirming the regulation of PKC by SHOCA-2 to be dependent on ER α (Figure S3B).

To analyze EGF signaling in cells with reduced SHOCA-2 levels, we generated short hairpin RNA (shRNA) expressing HeLa cell clones in which SHOCA-2 protein was knocked down by 90% (SHOCA-2 KD; Figure 3C). When stimulated with EGF, the extent of STAT3 phosphorylation at tyrosine 705 (Tyr⁷⁰⁵) and serine 727 (Ser⁷²⁷) was inversely correlated with the SHOCA-2 protein level (Figure 3C). In keeping with a significant increase in phosphorylated STAT3, the transcription of the c-Myc, Jun B and Cyclin D1 was enhanced (see below). In contrast, the phosphorylation of AKT was decreased while that of ERK 1/2 remained unchanged in SHOCA-2 KD cells (Figure S3C). These results further corroborated the specific involvement of SHOCA-2 in STAT3-mediated signaling following EGF stimulation.

Using a STAT3-driven luciferase reporter assay (Kreis et al., 2007), we next examined in HeLa cells whether STAT3 transcriptional activity was altered in the presence of decreased or increased SHOCA-2 protein levels. As demonstrated in Figure 3D, EGF treatment of HeLa cells with a significant loss of SHOCA-2 expression (SHOCA-2 KD cells) resulted in a robust transcriptional activity. In contrast, the transcription of the luciferase reporter was significantly reduced in HeLa cells overexpressing SHOCA-2 and exposed to EGF (Figure 3D; Figure S3D). To investigate the nature of SHOCA-2's influence on STAT3, we also tested porcine aortic endothelial cells (PAE) that lack endogenous EGFR expression and mouse embryonic fibroblasts (MEF) rendered deficient for STAT3 expression (Huang et al., 2007; Costa-Pereira et al., 2002). In these cells, SHOCA-2 bound to STAT3

and EGFR and this interaction was always independent of the missing binding partner normally present in wild type cells (Figure 3E, Figure S3E). Moreover, the binding of SHOCA-2 to EGFR and STAT3 was unrelated to its SH2 domain (Figure S3F) and isoform-specific because SHOCA-1 did not physically bind to STAT3 (Figure S3G). Taken together, these results suggest that SHOCA-2 inhibits EGF-induced STAT3 activation through its physical association with both EGFR and STAT3 independently of its SH2 domain.

SHOCA paralogues bind PP1b

Using tandem affinity purification (TAP) and reciprocal immunoprecipitation, we and others (Ewing et al., 2007) identified the serine/threonine Protein Phosphatase-1 beta (PP1b) as an interaction partner of SHOCA-1 and SHOCA-2 (Figure 4A; Figures S4A-C; Table S2). The eukaryotic PP1 protein family is composed of the 3 isoforms a, b and g, whose functionality relies on the association with different regulatory proteins (Ceulemans and Bollen, 2004). Both SHOCA paralogues bound PP1b but neither PP1a or PP1g, demonstrating an isoform-specific interaction (Figure 4B). Confocal immunofluorescence microscopy and immunoprecipitation of PP1b from cytoplasmic and nuclear protein fractions detected the SHOCA-2-PP1b interaction in both subcellular compartments further confirming their physical association (Figure S4D).

The [RK]-X(0,1)-[VI]-{P}-[FW] sequence has previously been identified as a PP1-interacting motif (Hendrickx et al., 2009). We therefore tested whether the KXILF and KX[VI][QH]W sequences located in the N-terminus of SHOCA-1 and -2 may mediate such binding (Figure 4C left panel; Figure S4E). Site-directed mutagenesis was used to alter in the SHOCA isoforms the first motif from KXILF to AXALA (designated PP1_{Mut1}) and the second sequence from KX[VI][QH]W to AXA[QH]A (PP1_{Mut2}; Figure 4C). Overexpression of SHOCA mutant proteins in HEK293 cells demonstrated that only the PP1_{Mut2} proteins failed to interact with PP1b identifying the KX[VI][QH]W sequence as the relevant docking site (Figure 4C, middle and right panels).

STAT3 inhibition by SHOCA-2 requires EGF-induced phosphorylation of Tyr131 and the recruitment of PP1b

We next investigated whether EGFR and/or STAT3 could pre-associate with SHOCA-2 independently of any EGFR stimulation, or, alternatively, whether these interactions required an activation-induced phosphorylation of the individual binding partners. Immunoprecipitation experiments showed that SHOCA-2-flag and endogenous EGFR, STAT3 and PP1b co-precipitated in serum starved, unstimulated HeLa cells independent of prior EGF stimulation (Figure 4D). Following EGF activation, the association of SHOCA-2, STAT3 and PP1b significantly increased in HeLa cells (Figure 4D). Although STAT3 can be indirectly activated through c-SRC (Quesnelle et al., 2007), c-SRC was not associated with the SHOCA-2, STAT3 and PP1b complex demonstrating a SHOCA-2-related, direct inhibition of STAT3 that was independent of a c-SRC engagement (Figure S4F).

Following EGF stimulation, STAT3 is recruited to the phospho-sites Tyr¹⁰⁶⁸ and Tyr¹⁰⁸⁶ within EGFR (Shao et al., 2003). However, in Figure 4D, we demonstrated that EGFR and STAT3 associated prior to EGFR-mediated stimulation. To further characterize this interaction, unstimulated HeLa cells were transfected with mutant forms of either EGFR or STAT3. The EGFR mutant had its Tyr¹⁰⁶⁸ and Tyr¹⁰⁸⁶ replaced by phenylalanine (EGFR Y1068/1086F) and hence lost a motif known for classical STAT3 recruitment (Shao et al., 2003), whereas Tyr⁷⁰⁵ was replaced in the STAT3 mutant to a phenylalanine thus eliminating several essential functions including STAT3 dimerization, translocation to the nucleus and transcriptional activity (Heinrich et al., 2003). Neither the association of mutant EGFR with wild type STAT3 nor the interaction of altered STAT3 with EGFR were affected by these changes (Figure S4G), thus demonstrating that EGFR's Tyr¹⁰⁶⁸/Tyr¹⁰⁸⁶ and STAT3's Tyr⁷⁰⁵ were not critically important for the interaction with each other (Figure S4G). The latter finding is in accordance with a pTyr⁷⁰⁵ independent recruitment of STAT3 to the IL-22 receptor, an interaction that however requires the coiled-coil domain of STAT3 (Dumoutier et al., 2009).

Because EGF-stimulation resulted in Tyr¹³¹ phosphorylation of SHOCA-2 (Figure 3A and (Guo et al., 2008)), we next examined whether this modification is essential for the formation of the SHOCA-2/EGFR/STAT3/PP1b complex and its function. Though changing the tyrosine residue to alanine (Y131A) did not disturb SHOCA-2's ability to complex with EGFR, STAT3 and PP1b (Figure S4H), the point mutation resulted in a constitutive phosphorylation of other SHOCA-2 tyrosine residues and an insensitivity to EGFR signal-mediated inhibition of STAT3 tyrosine phosphorylation (Figure S4I). The regulation of STAT3 activity by EGFR is therefore critically dependent on the phosphorylation of SHOCA-2's Tyr 131.

We next characterized in further detail the conditions under which SHOCA-2 can bind to its partners. Since both wild-type and PP1b interaction-deficient SHOCA-2 variants formed a complex with EGFR/STAT3 (Figure S4H), PP1b binding does not appear to be a prerequisite for the association of SHOCA-2 with STAT3. However, the inactivation of STAT3 was dependent on SHOCA-2's ability to associate with PP1b as the overexpression of SHOCA-2 PP1_{Mut2} failed to alter STAT3 Tyr⁷⁰⁵ phosphorylation following EGF stimulation (Figure 4E). Consequently, STAT3-dependent transcription in a reporter assay was not diminished in HeLa cells overexpressing SHOCA-2 PP1_{Mut2} that had been treated with EGF (Figure 4F). SHOCA-2 thus binds to the EGFR/STAT3 complex independently of PP1b, but requires the association with the phosphatase to modulate STAT3 activity.

SHOCA-2 requires PP1b to inhibit STAT3-dependent tumor cell growth

To address the contribution of SHOCA-2 to cellular growth control, we assessed the expression of STAT3-controlled cell proliferation genes (Aggarwal et al., 2009; Trenerry et al., 2007) in HeLa cells expressing reduced levels of SHOCA-2 (SHOCA-2 KD). The increase of c-Myc, Cyclin D1 and Jun B transcripts in these cells upon EGF stimulation inversely correlated with a robust reduction in SHOCA-2 expression (Figure 5A, left panel). Consequently, the fraction of cells in G2/M of the cell cycle was significantly higher in both unstimulated (8.8% vs. 5% G2/M cells [$p < 0.0002$]) and EGF activated SHOCA-2 KD HeLa cells (21.2% vs. 9.4% G2/M cells, [$p < 0.004$])

when compared to wild type controls (Figure 5A, right panel). A significantly increased cell proliferation (Ki-67 positivity) and higher levels of c-Myc, Jun B, and Cyclin D1 transcripts were also detected in tissue sections of clinically advanced CRC further confirming an inverse correlation between SHOCA-2 expression and STAT3 mediated cellular responses (Figure 5B; Figure S5A). A decrease in SHOCA-2 expression therefore promotes the transcription of cell proliferation factors that in term drive cell cycle progression.

To determine the molecular mechanism by which SHOCA-2 suppresses cell proliferation, we next transfected the colorectal cancer cell line SW480 to over-express STAT3 alone or in combination with either wild-type or mutant SHOCA-2. In a colony formation assay, overexpression of only STAT3 significantly increased the number of colony forming units (Figure 5C), a response that correlated with a higher degree of STAT3 phosphorylation (Figure 5D). Overexpression of STAT3 together with wild-type SHOCA-2 significantly reduced the colony forming potential of SW480 cells as well as STAT3 phosphorylation. This growth suppressive effect was less pronounced when SHOCA-2 mutants deficient in either PP1b interaction (SHOCA-2 PP1_{Mut2}) or Tyr¹³¹ (Y131A) phosphorylation were co-overexpressed instead (Figure 5C). Comparable results using the overexpression of wild type and mutant SHOCA-2 were obtained in the adenocarcinoma cell line H1975 which expresses an EGFR mutation constitutively activating STAT3 (Figure S5B) (Lu et al., 2007). Consistent with a role of SHOCA-2 in suppressing STAT3-mediated cell growth, SW480 cells in which SHOCA-2 expression was knocked down by shRNA diminished their proliferation rate upon exposure to pharmacological inhibition of STAT3 by S31-201, a chemical probe blocking STAT3-STAT3 complex formation and STAT3-DNA binding (Figure S5C). Finally, overexpression of SHOCA-2 in a CRC cell line (SW620) with constitutive STAT3 phosphorylation (Maa et al., 2007) and spontaneously low endogenous SHOCA-2 expression (Figure S5D) suppressed anchorage-independent cell growth (Figure 5E). Taken together, these results demonstrate under different experimental conditions that SHOCA-2 acts as a suppressor of STAT3-driven cancer cell proliferation.

Downregulation of SHOCA-2 alters the cell phenotype and promotes *in vivo* tumor growth

The cobble-stone shape of SW480 cells, which usually express high endogenous SHOCA-2 levels, changed to a spindle-like morphology following the knock-down of SHOCA-2 transcripts (Figure 6A, Figures S5D and S6A). This change in cell shape was paralleled by features of epithelial-mesenchymal transition (EMT) (Christofori, 2006) such as a reduction in E-cadherin and an increase in N-cadherin, vimentin, SNAIL and ZEB1 expression (Figure 6B). These alterations were, however blocked by the inhibition of EGFR using Tyrphostin (AG1478; Figure S6B). Moreover, SHOCA-2 deficient SW480 cells proliferated at a greater rate and displayed a higher sensitivity to EGFR inhibition when compared to mock transfected controls (Figure 6C). Re-establishing mouse Shoca-2 expression in human SHOCA-2 deficient SW480 cells however decreased their proliferation rate (Figure 6D), thus demonstrating again that SHOCA-2 controls STAT3-mediated tumor growth.

To establish an *in vivo* role for SHOCA-2 as a tumor suppressor, SW480 cells that had their SHOCA-2 expression either knocked-down or left unchanged were transplanted into nude mice. Grafts with cells where SHOCA-2 was reduced displayed a significant increase both in tumor incidence and size when compared to control transplants (Figure 6E). Moreover, significant levels of phosphorylated, nuclear STAT3 were detected in tumors that had emerged from grafted cells lacking regular SHOCA-2 expression (Figure 6F, upper left panel). Consistent with this experimental finding, increased STAT3 phosphorylation at residues Tyr⁷⁰⁵ and Ser⁷²⁷ (Figure 6F, upper right and lower panels) were also detected in the majority of CRC biopsies with low SHOCA-2 expression (H-score <100). Taken together, experimental *in vivo* tumor models using CRC cell lines and the detailed analysis of biopsies taken from CRC patients identified the inhibitory role of SHOCA-2 on STAT3 activation.

Discussion

Genomic deletions within the 8p21-23 region constitute a characteristic cytogenetic feature of CRC (Emi et al., 1992) and have been linked to carcinogenesis (Macartney-Coxson et al., 2008). How deletions in this chromosomal region contribute to CRC development has remained unclear since tumor suppressor genes could so far not be allocated to 8p21-23 (Macartney-Coxson et al., 2008; Emi et al., 1992). With the assignment of the SHOCA-2 encoding *SH2D4A* gene to 8p21.3, we have linked genomic instability at this locus with the loss of a specific tumor suppressor activity in CRC. We provide experimental evidence and clinical data that a loss of SHOCA-2 expression results in unopposed STAT3 activation following EGFR stimulation and marks in patients with colorectal cancer accelerated tumor growth as well as poor prognosis. Our findings of a role for SHOCA-2 in the control of tumor growth also extend to malignancies where the mechanism of tumor development and progression are still only poorly understood (Roessler et al., 2011).

EGFR is an upstream activator of multiple pathways involved in cell proliferation, apoptosis and carcinogenesis. Overexpression of EGFR and constitutive activation of its major downstream effector, STAT3, have been linked to cancer progression, a higher risk for metastasis and reduced survival (Quesnelle et al., 2007; Klampfer, 2008; Saif, 2010). Since STAT3 regulates cell growth and tissue homeostasis (Aggarwal et al., 2009), its activation must be under stringent control. We found that SHOCA-2 cooperates with PP1b in modulating the EGFR-induced STAT3 activation, thereby limiting cell proliferation. In view of the specific genetic and epigenetic alterations of *SH2D4A*, that particularly emerge in late stage CRCs, and given the correlation between poor clinical outcome and the loss of SHOCA-2 expression, we conclude that SHOCA-2 is a novel tumor suppressor modulating the activation of the EGFR signaling pathway (Walther et al., 2009). However, in the cohort studied here, the SHOCA-2 status was not a reliable predictor of survival in patients stratified by disease stage. This result is not surprising as it is consistent with the predictive potential of other tumor

suppressors. For example, p53, APC, and DCC also fail to serve as prognostic markers though their loss of expression in colorectal cancer is correlated with increased tumor aggressiveness (Roth, 1999; Walther et al., 2009).

Based on the evidence that SHOCA-2 binds to both EGFR and phosphatase PP1b and is itself phosphorylated upon EGF signaling, we propose that SHOCA-2 controls STAT3 activity through the catalytic activity of PP1b. SHOCA-2 and PP1b form a complex with STAT3 in that SHOCA-2 acts as an adaptor to bring the PP1b phosphatase into physical proximity of STAT3, which in turn effects the dephosphorylation of activated STAT3. Indeed, phosphorylation/dephosphorylation of Ser⁷²⁷ plays an important role in modulating STAT3's transcriptional activity (Levy and Darnell, 2002; Wen et al., 1995; Shen et al., 2004), though the molecular mechanisms underlying Ser⁷²⁷ dephosphorylation have not yet been fully delineated. Though the serine/threonine phosphatases PP1 and PP2A have already been identified to dephosphorylate Ser⁷²⁷ (Lütticken et al., 1995; Woetmann et al., 1999; Togi et al., 2009; Haridas et al., 2009), the regulatory subunit that confers selectivity, specificity, and subcellular localization of PP1 towards STAT3 has so far and contrary to PP2A (Togi et al., 2009) remained unidentified. For the EGF pathway, we now demonstrate that Ser⁷²⁷ phosphorylation of STAT3 decreases on the condition that PP1b associates with activated SHOCA-2.

Dephosphorylation of Tyr⁷⁰⁵ constitutes another critical step in the events leading to STAT3 deactivation (Heinrich et al., 2003). One of the phosphatases involved in this event is the receptor protein tyrosine phosphatase delta (PTPRD), which has been shown to act as a tumor suppressor in both colon cancer and other neoplasms (Zhang et al., 2007; Veeriah et al., 2009). Though PP1b constitutes yet another phosphatase involved in STAT3 inactivation, it remains unknown whether it acts indirectly via a tyrosine-specific phosphatase (e.g. SHP (Neviani et al., 2005)), or directly through target promiscuity (MacKintosh et al., 1996; Shi, 2009). A dual specificity of PP1b's catalytic activity is not unparalleled as a comparable mechanism has been described for the protein tyrosine phosphatase SHP-2

(Wu et al., 2002). Alternatively, the enhanced physical interaction between PP1b, SHOCA-2 and STAT3 following EGF binding to its receptor may be sufficient to inactivate STAT3 secondary to conformational changes independent of any catalytic activity (Lee et al., 2009). The precise molecular mechanism by which Tyr⁷⁰⁵ is de-phosphorylated remains to date unidentified. Irrespective of this lack of knowledge, inappropriate STAT3 regulation is the noticeable consequence of aberrant SHOCA-2 expression and results in an up-regulation of genes involved in cell proliferation.

The reduction of SHOCA-2 expression in the CRC cell line SW480 also produced phenotypic features characteristic of EMT. Such a transformation has been implicated in the conversion of early stage malignancies towards more aggressive tumors showing invasive growth and the formation of metastasis (Thiery et al., 2009). Consistent with a role in EMT, LOH at *SH2D4A* was mostly detected in patients with a disease stage typically associated with metastasis. It is thus conceivable that the loss of SHOCA-2 plays a part in tumor progression not only through its impact on tumor cell growth but possibly also via its contribution to EMT.

EGFR and STAT3 constitute promising drug targets in the treatment of cancer including CRC (Quesnelle et al., 2007). However, therapeutic responses are observed in only 10 to 20% of patients treated with EGFR antagonists (Ciardiello and Tortora, 2008) and none of the available STAT3 inhibitors can to date be considered as viable drug candidates because of their limited efficiency in disrupting STAT3 homodimerization (Yue and Turkson, 2009). Detailed knowledge about the regulatory network that controls STAT3 activity may likely provide provides novel rationale for the design of strategies to interfere with a STAT3-mediated activation in cancer cells.

Experimental Procedures

A detailed description of the methods used can be found in Supplemental Experimental Procedures.

Plasmids, antibodies, cells and reagents.

Antibodies, expression plasmids, cells and reagents were either gifts from investigators or bought commercially, as specified in Extended Experimental Procedures.

Bioinformatic analysis.

The SHOCA-1 (*SH2D4B*) and SHOCA-2 (*SH2D4A*) sequences were analyzed using NCBI, ENSEMBL (Hubbard et al., 2009), the kinBase database (<http://www.kinase.com/kinbase/>), the Toffee software (Notredame et al., 2000), and OncoPrint (Rhodes et al., 2007).

TAP purification and mass spectrometry.

The TAP purification (Chen and Gingras, 2007) and mass spectrometry (Hess et al., 2008) were carried out as previously reported.

Western Blot, immunoprecipitation and cell fractionation.

Cells lysates were either separated on a 8%-12% SDS-PAGE or immunoprecipitated with protein G Plus-Sepharose beads coupled to specified antibodies and analyzed by western blotting. Cell fractionation was achieved using the CellLytic Nuclear extraction kit (Sigma).

Luciferase reporter gene assays.

HeLa cells were transfected with a firefly luciferase reporter SIE (sis-inducible element) plasmid plus a Renilla luciferase plasmid (both Promega) and plasmids that encode either STAT3 or SHOCA-2. Cell extracts were tested using the Dual Luciferase Assay System (Promega) and analyzed using a luminometer. Firefly luciferase activity was normalized to Renilla luciferase activity.

Cell cycle analysis.

Cells were incubated for 15 min with ice-cold hypotonic Propidium Iodide staining solution and subsequently analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases were defined using the mathematical Watson Pragmatic model.

Colony formation and anchorage-independent growth assays.

For colony formation assay, SW480 cells co-transfected with a control vector or plasmids that encode STAT3, SHOCA-2, SHOCA-2 PP1_{Mut2} or SHOCA-2 Y131A were cultured for 15 days under neomycin selection (G418: 1 mg/ml) and then analyzed for colony frequency following fixation and crystal violet staining of the cells. SW620 cells transfected with a control vector or plasmids that encode SHOCA-2, SHOCA-2 PP1_{Mut2} or SHOCA-2 Y131A, cultured under neomycin selection (G418, 1 mg/ml) in a soft agar (CytoSelect, Cell Biolabs) for 15 days and then analyzed for colony frequency.

***In vivo* analyses of tumor formation.**

SW480 cells (5×10^6) were grafted to the flank of 6 week-old female BALB/c nude mice and the size of the tumor was measured weekly. Mice were sacrificed after 5 weeks and tumors were removed for further analysis. The animal experiments were carried out in accordance with the legal requirements of the Swiss veterinary authorities.

Human tissues and immunohistochemistry.

For the genetic and epigenetic study, specimens from 70 CRC patients were obtained from the Department of Surgery of the Kantonsspital Aarau, Switzerland and from of the Department of Gastroenterology of the Inselspital in Berne, Switzerland, with approval of the local medical ethics boards and written consent from patients. Tissues were stored in RNA_{later} reagent at -70°C . Single-punch tumor samples from 501 patients (Institute of Pathology, Stadtspital Triemli, Zürich, Switzerland) were analyzed by tissue microarrays as described previously (Zlobec et al., 2010). SHOCA-2 expression and

STAT3 phosphorylation were quantified using the H score (Hirsch et al., 2003).

Mutational screening, loss of heterozygosity and gene dosage analysis.

The complete coding sequence including flanking intronic regions of the *SH2D4A* gene was analyzed in all cancers lacking SHOCA-2 protein expression and/or that were positive for LOH. LOH for the *SH2D4A* locus was determined using a set of 3 flanking microsatellite markers and 6 gene-specific single nucleotide polymorphisms (SNP) (for details see Extended Experimental Procedures). Gene dosage was determined for each cancer using a chromosome 8-specific multiplex ligation-dependent probe amplification (MLPA) assay (SALSA P014-1A, MRC Holland) and data were processed by the GeneMarker software package (Softgenetics).

Methylation analysis.

Genomic DNA was converted by bisulfite treatment using the EZ DNA Methylation Kit™ according to the manufacturer's instructions (Zymo Research). CpG island regions were amplified by PCR from bisulfite treated DNA and sequenced with the PyroMark Q24 pyrosequencing system (Qiagen) allowing the quantitation of methylated CpG sites (PCR conditions and primer sequences are given in Extended Experimental Procedures). Bisulfite treated genomic DNA, previously *in vitro* methylated with M.SssI methyltransferase (New England BioLabs), served as the positive control.

Statistical analyses.

Student's t test, Likelihood ratio test, Chi-square test, Gehran-Wilcoxon test, Fisher's Exact test, binomial test for equal proportions, Log-rank test, multivariate Cox regression analysis and Kaplan-Meier method were used for statistical analyzes, as indicated.

Acknowledgements

We would like to thank Annick Peter, Dorothea Maass and Ragna Sack for expert technical help; Dr. Giancarlo Marra for microarray data; Dr. Fabrizio Bianchi for statistical analyzes; Dr. Bérengère Fayard and Dr. Laura Trinkle-Mulcahy for valuable discussions; Drs. Nancy Hynes, Thomas Barthlott, Poul Sørensen and Gerhard Christofori for critical reading of the manuscript; Sabrina Harris for secretarial assistance.

This work was supported by grants from the Swiss National Science Foundation (to G.A.H.), the Krebsliga beider Basel (to S.L. and G.A.H.) and the Oxford Biomedical Research Centre (to G.A.H.).

S.L. and G.A.H. designed the experiments, analyzed the data and composed together with P.S. the manuscript; S.L. performed experiments including Western Blot, immunoprecipitation, luciferase assays, mutagenesis, TAP assay, phosphatase assay, bioinformatic analyses, colony formation and anchorage-independent growth assays, and *in vivo* experiments; S.Z., K.H. and M.Ke. participated in the identification and cloning of *SH2D4B* and *SH2D4A* genes; D.H. performed the mass spectrometry analysis; supervised by P.S. and K.H., M.Ko. and S.W. performed the human genetic and epigenetic studies; immunohistochemistry was done by S.P.; K.T. provided human CRC samples; I.Z. performed statistical analyzes; L.T. supplied and analyzed CRC tissue microarray.

The authors declare no competing financial interests.

References

- Aggarwal, B.B., Kunnumakkara, A.B., Harikumar, K.B., Gupta, S.R., Tharakan, S.T., Koca, C., Dey, S., and Sung, B. (2009). Signal transducer and activator of transcription-3, inflammation, and cancer: how intimate is the relationship? *Ann N Y Acad Sci* 1171, 59-76.
- Arao, Y., Hamilton, K.J., Ray, M.K., Scott, G., Mishina, Y., and Korach, K.S. (2011). Estrogen receptor α AF-2 mutation results in antagonist reversal and reveals tissue selective function of estrogen receptor modulators. *Proc Natl Acad Sci U S A* 108, 14986-991.
- Bettstetter, M., Dechant, S., Ruemmele, P., Grabowski, M., Keller, G., Holinski-Feder, E., Hartmann, A., Hofstaedter, F., and Dietmaier, W. (2007). Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin Cancer Res* 13, 3221-28.
- Birnbaum, D., Adélaïde, J., Popovici, C., Charafe-Jauffret, E., Mozziconacci, M.J., and Chaffanet, M. (2003). Chromosome arm 8p and cancer: a fragile hypothesis. *Lancet Oncol* 4, 639-642.
- Bromberg, J.F., Wrzeszczynska, M.H., Devgan, G., Zhao, Y., Pestell, R.G., Albanese, C., and Darnell, J.E. (1999). Stat3 as an oncogene. *Cell* 98, 295-303.
- Ceulemans, H., and Bollen, M. (2004). Functional diversity of protein phosphatase-1, a cellular economizer and reset button. *Physiol Rev* 84, 1-39.
- Chen, G.I., and Gingras, A.C. (2007). Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* 42, 298-305.
- Christofori, G. (2006). New signals from the invasive front. *Nature* 441, 444-450.
- Ciardello, F., and Tortora, G. (2008). EGFR antagonists in cancer treatment. *N Engl J Med* 358, 1160-174.
- Costa-Pereira, A.P., Tininini, S., Strobl, B., Alonzi, T., Schlaak, J.F., Is'harc, H., Gesualdo, I., Newman, S.J., Kerr, I.M., and Poli, V. (2002). Mutational switch of an IL-6 response to an interferon-gamma-like response. *Proc Natl Acad Sci U S A* 99, 8043-47.

Dumoutier, L., de Meester, C., Tavernier, J., and Renauld, J.C. (2009). A new activation modus of STAT3: A tyrosine-less region of the IL-22 receptor recruits STAT3 by interacting with its coiled-coil domain. *J Biol Chem*

Emi, M., Fujiwara, Y., Nakajima, T., Tsuchiya, E., Tsuda, H., Hirohashi, S., Maeda, Y., Tsuruta, K., Miyaki, M., and Nakamura, Y. (1992). Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. *Cancer Res* 52, 5368-372.

Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., et al. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3, 89.

Guo, A., Villén, J., Kornhauser, J., Lee, K.A., Stokes, M.P., Rikova, K., Possemato, A., Nardone, J., Innocenti, G., et al. (2008). Signaling networks assembled by oncogenic EGFR and c-Met. *Proc Natl Acad Sci U S A* 105, 692-97.

Haridas, V., Nishimura, G., Xu, Z.X., Connolly, F., Hanausek, M., Walaszek, Z., Zoltaszek, R., and Gutterman, J.U. (2009). Avicin D: a protein reactive plant isoprenoid dephosphorylates Stat 3 by regulating both kinase and phosphatase activities. *PLoS One* 4, e5578.

Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Müller-Newen, G., and Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J* 374, 1-20.

Hendrickx, A., Beullens, M., Ceulemans, H., Abt, T.D., Van Eynde, A., Nicolaescu, E., Lesage, B., and Bollen, M. (2009). Docking motif-guided mapping of the interactome of protein phosphatase-1. *Chem Biol* 16, 365-371.

Hess, D., Keusch, J.J., Oberstein, S.A., Hennekam, R.C., and Hofsteenge, J. (2008). Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective Omicron-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* 283, 7354-360.

Hirsch, F.R., Varella-Garcia, M., Bunn, P.A., Di Maria, M.V., Veve, R., Bremmes, R.M., Barón, A.E., Zeng, C., and Franklin, W.A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 21, 3798-3807.

Huang, F., Goh, L.K., and Sorkin, A. (2007). EGF receptor ubiquitination is not necessary for its internalization. *Proc Natl Acad Sci U S A* 104, 16904-09.

Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., et al. (2009). Ensembl 2009. *Nucleic Acids Res* 37, D690-97.

Klampfer, L. (2008). The role of signal transducers and activators of transcription in colon cancer. *Front Biosci* 13, 2888-899.

Kreis, S., Munz, G.A., Haan, S., Heinrich, P.C., and Behrmann, I. (2007). Cell density dependent increase of constitutive signal transducers and activators of transcription 3 activity in melanoma cells is mediated by Janus kinases. *Mol Cancer Res* 5, 1331-341.

Levy, D.E., and Darnell, J.E. (2002). Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol* 3, 651-662.

Li, T., Li, W., Lu, J., Liu, H., Li, Y., and Zhao, Y. (2009). SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway. *BMB Rep* 42, 516-522.

Lu, Y., Liang, K., Li, X., and Fan, Z. (2007). Responses of cancer cells with wild-type or tyrosine kinase domain-mutated epidermal growth factor receptor (EGFR) to EGFR-targeted therapy are linked to downregulation of hypoxia-inducible factor-1alpha. *Mol Cancer* 6, 63.

Lütticken, C., Coffey, P., Yuan, J., Schwartz, C., Caldenhoven, E., Schindler, C., Kruijer, W., Heinrich, P.C., and Horn, F. (1995). Interleukin-6-induced serine phosphorylation of transcription factor APRF: evidence for a role in interleukin-6 target gene induction. *FEBS Lett* 360, 137-143.

Maa, M.C., Lee, J.C., Chen, Y.J., Chen, Y.J., Lee, Y.C., Wang, S.T., Huang, C.C., Chow, N.H., and Leu, T.H. (2007). Eps8 facilitates cellular growth and motility of colon cancer cells by increasing the expression and activity of focal adhesion kinase. *J Biol Chem* 282, 19399-9409.

Macartney-Coxson, D.P., Hood, K.A., Shi, H.J., Ward, T., Wiles, A., O'Connor, R., Hall, D.A., Lea, R.A., Royds, J.A., et al. (2008). Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level. *BMC Cancer* 8, 187.

MacKintosh, C., Garton, A.J., McDonnell, A., Barford, D., Cohen, P.T., Tonks, N.K., and Cohen, P. (1996). Further evidence that inhibitor-2 acts like a chaperone to fold PP1 into its native conformation. *FEBS Lett* 397, 235-38.

Neviani, P., Santhanam, R., Trotta, R., Notari, M., Blaser, B.W., Liu, S., Mao, H., Chang, J.S., Galletta, A., et al. (2005). The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. *Cancer Cell* 8, 355-368.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205-217.

Quesnelle, K.M., Boehm, A.L., and Grandis, J.R. (2007). STAT-mediated EGFR signaling in cancer. *J Cell Biochem* 102, 311-19.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincaid-Beal, C., et al. (2007). OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9, 166-180.

Roessler, S., Long, E.L., Budhu, A., Chen, Y., Zhao, X., Ji, J., Walker, R., Jia, H.L., Ye, Q.H., et al. (2011). Integrative Genomic Identification of Genes on 8p Associated with Hepatocellular Carcinoma Progression and Patient Survival. *Gastroenterology*

Roth J. A. (1999). *Clin Cancer Res*. In *p53 prognostication: paradigm or paradox?*. UNITED STATES: [UNKNOWN REFERENCE TYPE]

Saif, M.W. (2010). Colorectal cancer in review: the role of the EGFR pathway. *Expert Opin Investig Drugs* 19, 357-369.

Shao, H., Cheng, H.Y., Cook, R.G., and Tweardy, D.J. (2003). Identification and characterization of signal transducer and activator of transcription 3 recruitment sites within the epidermal growth factor receptor. *Cancer Res* 63, 3923-930.

Shen, Y., Schlessinger, K., Zhu, X., Meffre, E., Quimby, F., Levy, D.E., and Darnell, J.E. (2004). Essential role of STAT3 in postnatal survival and growth revealed by mice lacking STAT3 serine 727 phosphorylation. *Mol Cell Biol* 24, 407-419.

Shi, Y. (2009). Serine/threonine phosphatases: mechanism through structure. *Cell* 139, 468-484.

Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Konishi, M., Oshimura, M., and Miyaki, M. (1996). Suppression of tumorigenicity and invasiveness of colon carcinoma cells by introduction of normal chromosome 8p12-pter. *Oncogene* 12, 405-410.

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

Togi, S., Kamitani, S., Kawakami, S., Ikeda, O., Muromoto, R., Nanbo, A., and Matsuda, T. (2009). HDAC3 influences phosphorylation of STAT3 at serine 727 by interacting with PP2A. *Biochem Biophys Res Commun* 379, 616-620.

Trener, M.K., Carey, K.A., Ward, A.C., and Cameron-Smith, D. (2007). STAT3 signaling is activated in human skeletal muscle following acute resistance exercise. *J Appl Physiol* 102, 1483-89.

Veeriah, S., Brennan, C., Meng, S., Singh, B., Fagin, J.A., Solit, D.B., Paty, P.B., Rohle, D., Vivanco, I., et al. (2009). The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc Natl Acad Sci U S A* 106, 9435-440.

Vogelstein, B., and Kinzler, K.W. (2004). Cancer genes and the pathways they control. *Nat Med* 10, 789-799.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 9, 489-499.

Weinberg, R.A. (1991). Tumor suppressor genes. *Science* 254, 1138-146.

Wen, Z., Zhong, Z., and Darnell, J.E. (1995). Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* 82, 241-250.

Woetmann, A., Nielsen, M., Christensen, S.T., Brockdorff, J., Kalltoft, K., Engel, A.M., Skov, S., Brender, C., Geisler, C., et al. (1999). Inhibition of protein phosphatase 2A induces serine/threonine phosphorylation, subcellular redistribution, and functional inhibition of STAT3. *Proc Natl Acad Sci U S A* 96, 10620-25.

Wu, T.R., Hong, Y.K., Wang, X.D., Ling, M.Y., Dragoi, A.M., Chung, A.S., Campbell, A.G., Han, Z.Y., Feng, G.S., and Chin, Y.E. (2002). SHP-2 is a dual-specificity phosphatase involved in Stat1 dephosphorylation at both tyrosine and serine residues in nuclei. *J Biol Chem* 277, 47572-580.

Yue, P., and Turkson, J. (2009). Targeting STAT3 in cancer: how successful are we? *Expert Opin Investig Drugs* 18, 45-56.

Zhang, X., Guo, A., Yu, J., Possemato, A., Chen, Y., Zheng, W., Polakiewicz, R.D., Kinzler, K.W., Vogelstein, B., et al. (2007). Identification of STAT3 as a substrate of receptor protein tyrosine phosphatase T. *Proc Natl Acad Sci U S A* 104, 4060-64.

Zlobec, I., Molinari, F., Kovac, M., Bihl, M.P., Altermatt, H.J., Diebold, J., Frick, H., Germer, M., Horcic, M., et al. (2010). Prognostic and predictive value of TOPK stratified by KRAS and BRAF gene alterations in sporadic, hereditary and metastatic colorectal cancer patients. *Br J Cancer* 102, 151-161.

(A) Left, location of SHOCA-2 on human chromosome (huChr) 8p. Middle, degree of conservation between the two human SHOCA sequences with higher degrees of conservation specified by darker shades of grey; box, a consensus MyPhoNE sequence typically present in myosin homologues (Hendrickx et al., 2009). Right, predicted domain organization of SHOCA; the amino acid identity between SHOCA-1 and -2 is given for the MyPhoNE and SH2 domains. Of note, SHOCA-2 protein contains only one coiled-coil domain (CC), namely CC2.

(B) Immunohistochemical analysis of a single tissue microarray containing 400 samples from unselected, untreated patients with sporadic CRC. SHOCA-2 expression was quantified using the H-score that determines the percentage of positive tumor cells multiplied by their staining intensity generating individual groups with scores of >200 (n=275), defining a strong expression (green); 100-200 (n=87), identifying a moderate expression (blue); <100 (n=38), characterizing a low or absent expression (red).

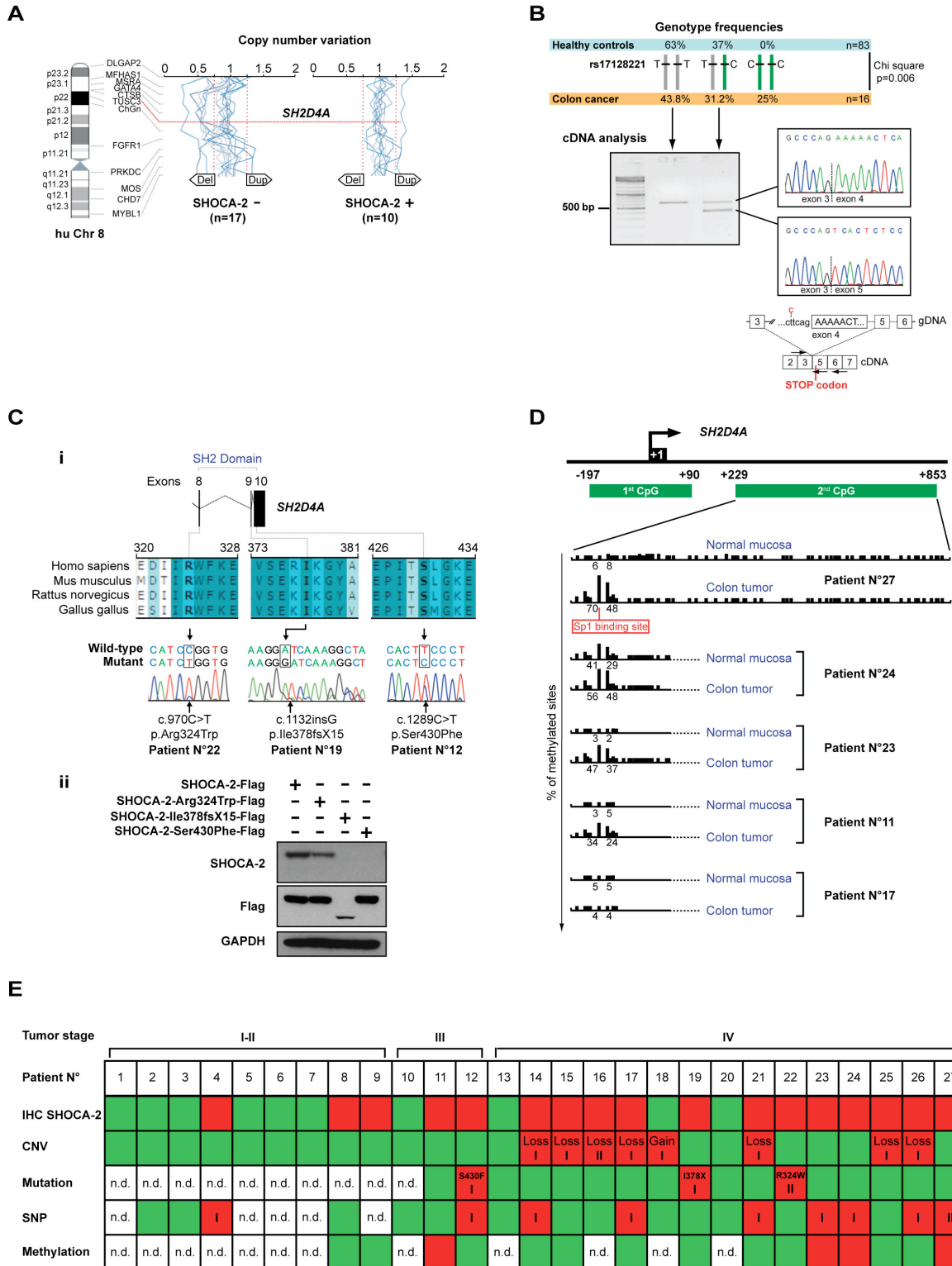
(C) Representative immunohistochemical analysis of tumor tissue of distinct CRC stages depicting either diffuse (>90% cells stained; left), patchy (<30%; middle) or no SHOCA-2 expression (right). T = tumor tissue, N = normal tissue. Scale bar, 200 μ m.

(D) Kaplan-Meier survival analysis of patients with high (green), moderate (blue) and low SHOCA-2 expression (red) for which sufficient clinical information was available (log-rank test $p = 0.0118$ when comparing high to low SHOCA-2 expression).

See also Figure S1 and Table S1

Figure 2. Genetic and epigenetic silencing of *SH2D4A* encoding SHOCA-2.

Figure 2



(A) HuChr8p copy number variation profiles in SHOCA-2-positive and SHOCA-2-negative CRC. Values <0.75 indicate chromosomal deletions (Del) whereas values >1.25 designate chromosomal duplications (Dup). Red dashed lines, threshold.

(B) Top, comparison of the *SH2D4A* c.342-5T>C (rs17128221) SNP frequency between 83 unrelated healthy Caucasian controls and 16 CRC patients (Chi square p value = 0.006). Bottom, SNP rs17128221 RT-PCR (right) and sequencing analysis (left), the diagram depicts exon usage and the arrows indicate the annealing sites for the primers.

(C) i: top, alignment of amino acid sequences for the SH2 domain (AA 320-454) of human SHOCA-2 (AAH82982) with homologues of *mus musculus* (AAI16683), *rattus norvegicus* (NP_001012048), and *gallus gallus* (XP_420452); bottom, sequence analysis of patient tumors. ii: immunoblot analysis of HeLa cells transfected to express Flag-fusion proteins of the indicated human SHOCA-2 mutations.

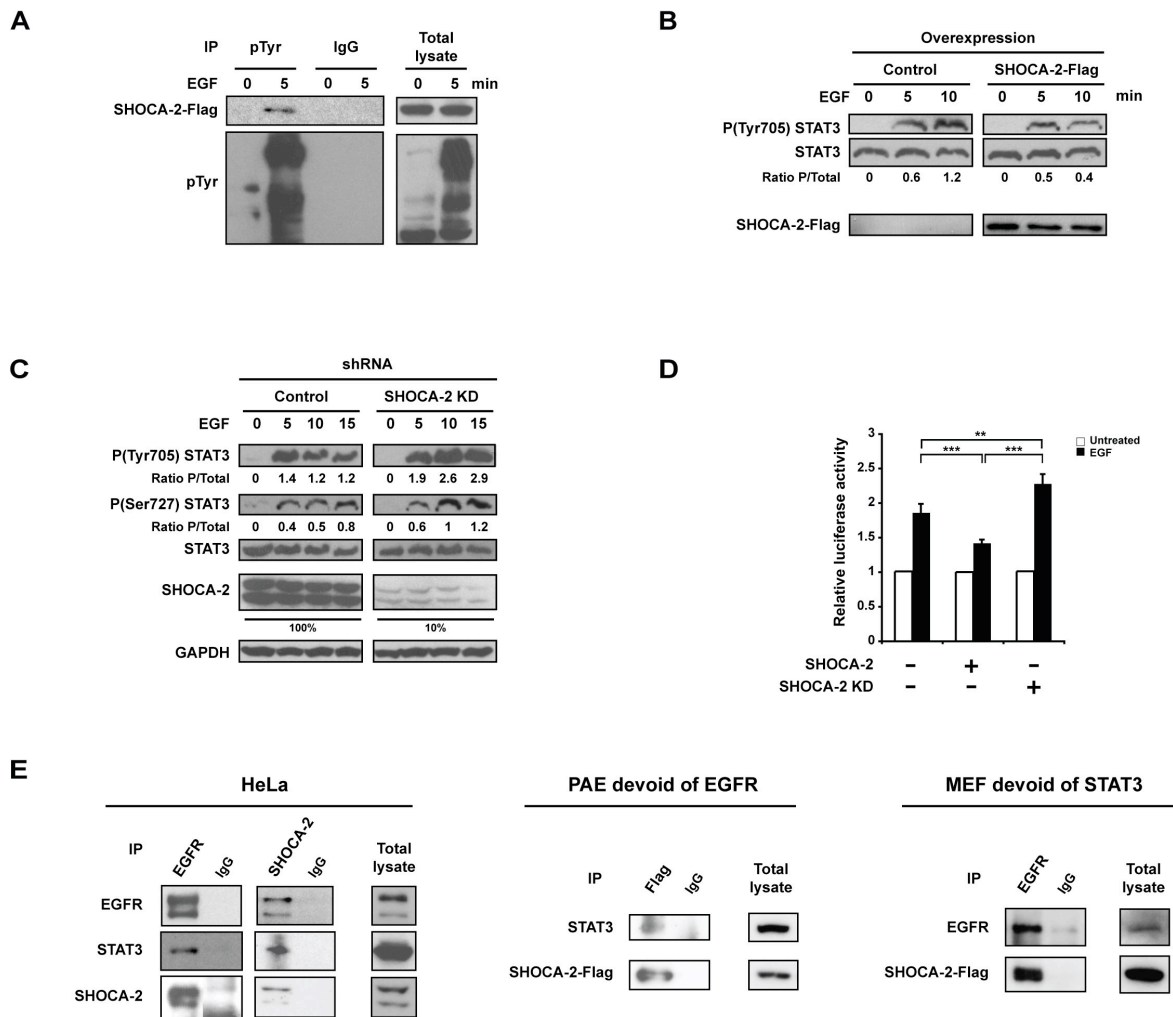
(D) DNA methylation analysis of CpG island located in the 5'UTR of *SH2D4A* comparing normal and tumor tissue.

(E) Summary of all genetic and epigenetic changes detected in a cohort of 27 informative CRC samples; CNV = copy number variation, SNP = single nucleotide polymorphism. Green = normal finding, red = pathological result, n.d. = not done, roman numerals = number of alleles affected.

See also Figure S2

Figure 3. SHOCA-2 is phosphorylated following EGFR activation and inhibits STAT3 activity.

Figure 3



(A) SHOCA-2-Flag expressing, serum-starved HeLa cells were stimulated as indicated with EGF (100 ng/ml). Cell lysates were immunoprecipitated with anti-phosphotyrosine (pTyr) antibody or rabbit IgG and immunoblotted to detect Flag and pTyr.

(B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag and stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry.

(C) Serum-starved HeLa cells that express normal (control, 100%) or 10% of SHOCA-2 (SHOCA-2 KD) were stimulated for 5, 10 and 15 min with EGF (100 ng/ml). The indicated proteins were detected by immunoblotting. Densitometry was used to determine the ratio of phosphoprotein to total protein (designated P/Total).

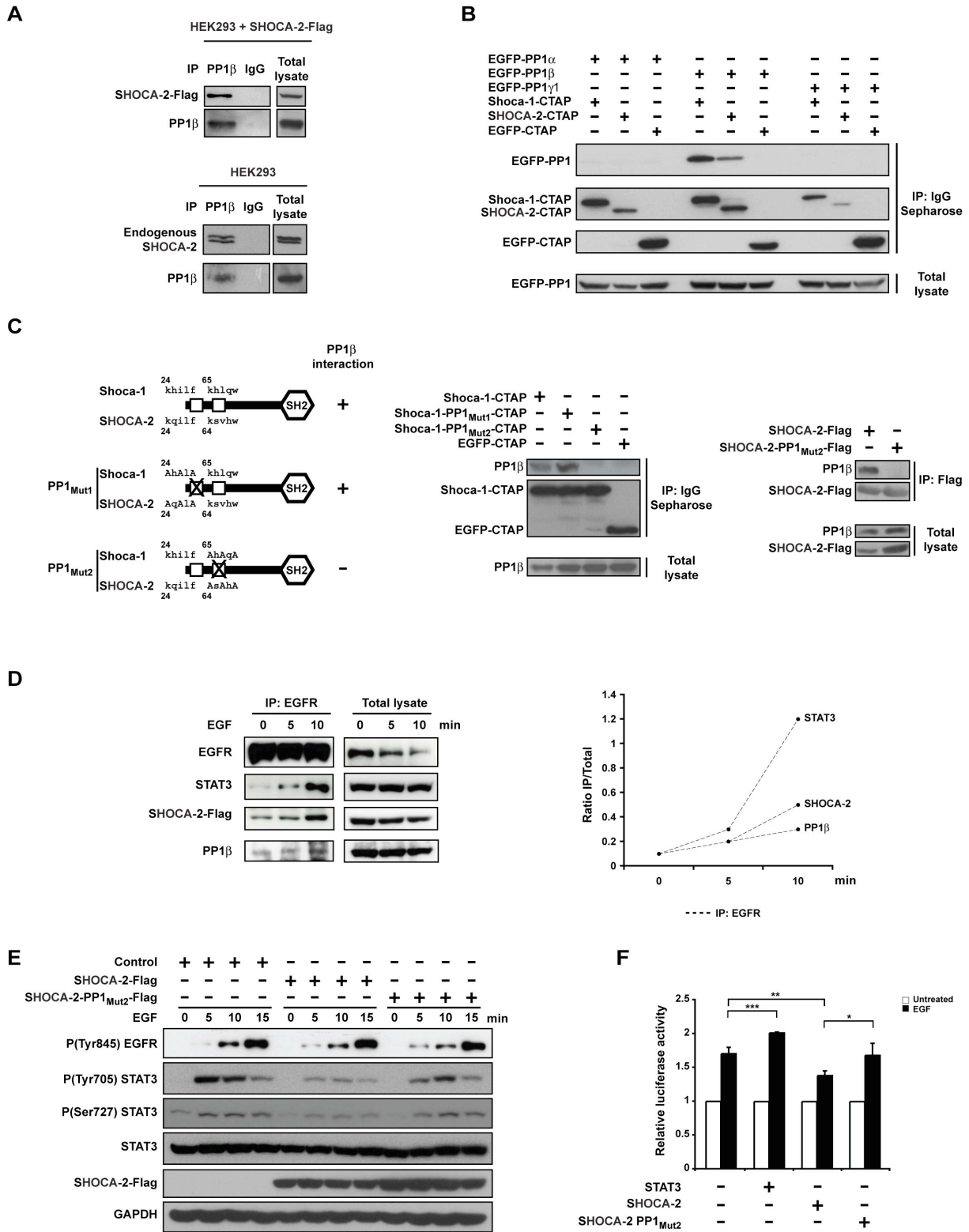
(D) Serum-starved HeLa cells and SHOCA-2 KD HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and a SHOCA-2 expression plasmid were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean \pm SD of three independent, triplicate assays (Student's t test, **p < 0.01; ***p < 0.005).

(E) Immunoblotting of HeLa, PAE and MEF lysates immunoprecipitated with anti-EGFR, anti-SHOCA-2, anti-Flag or mouse IgG antibodies for the detection of EGFR, STAT3, SHOCA-2 and Flag.

See also Figure S3

Figure 4. SHOCA-2 controls STAT3 activity in a PP1b-dependent fashion.

Figure 4



(A) Lysates from HEK293 cells (bottom) and HEK293 cells expressing SHOCA-2-Flag (top) were immunoprecipitated with anti-PP1b antibody or IgG and immunoblotted for PP1b, and Flag (top) or SHOCA-2 (bottom). Human SHOCA-2 is detected as isoforms due to exon 5 skipping.

(B) Lysates from HEK293 cells expressing CTAP fusion proteins and EGFP-tagged PP1 isoforms were immunoprecipitated with IgG sepharose beads (recognizing CTAP fusion proteins) and immunoblotted for the detection of CTAP and EGFP fusion proteins.

(C) Left, diagram representing wild-type and mutant SHOCA altered PP1b - binding sites (boxes) are marked by crosses. Middle, CTAP fusion proteins and PP1b were detected in total and immunoprecipitated lysates from transfected HEK293 cells. Right, SHOCA-2, Flag and PP1b, respectively, were detected in total lysates and anti-Flag immunoprecipitates from HEK293 cells expressing wild type or mutant SHOCA fusion proteins.

(D) Left, serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were analyzed unmanipulated or were immunoprecipitated with anti-EGFR antibody for the detection of EGFR, STAT3, Flag and PP1b by immunoblotting. Right, ratio of immunoprecipitated protein to total protein (designated IP/Total) as determined by densitometry.

(E) Serum-starved HeLa cell co-expressing the indicated proteins were stimulated with EGF (100 ng/ml) for the specified times. Lysates were immunoblotted for the detection of phospho-EGFR (Tyr845), phospho-STAT3 (Tyr705/Ser727), STAT3, Flag, and GAPDH.

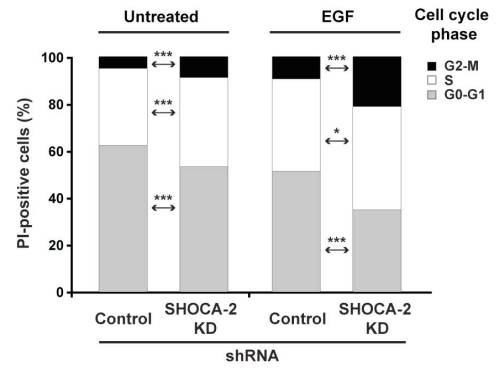
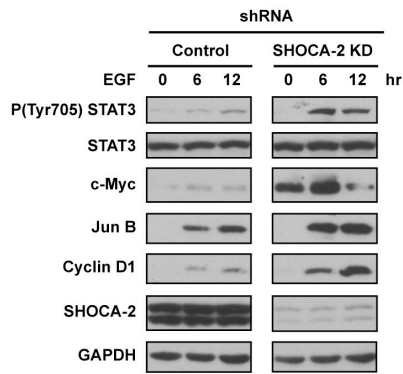
(F) Serum-starved HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and plasmids encoding either STAT3, the human SHOCA-2-Flag fusion protein or a mutant human SHOCA-2-Flag fusion protein unable to associate with PP1b (PP1_{Mut2}) were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean \pm SD of three independent, triplicate assays (Student's t test, *p < 0.05; **p < 0.01; ***p < 0.005).

See also Figure S4 and Table S2

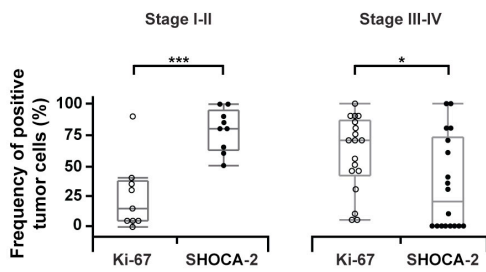
Figure 5. SHOCA-2 and PP1b arrest tumor cell proliferation sustained by STAT3.

Figure 5

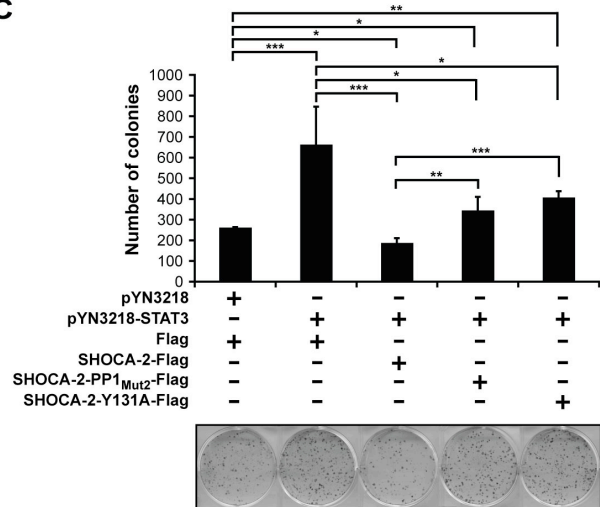
A



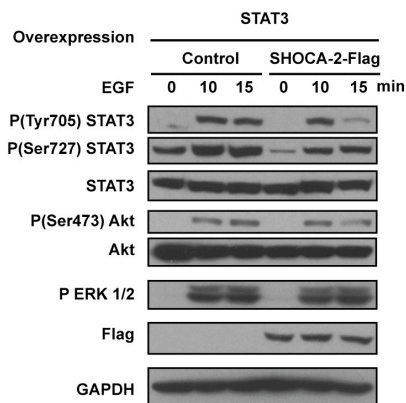
B



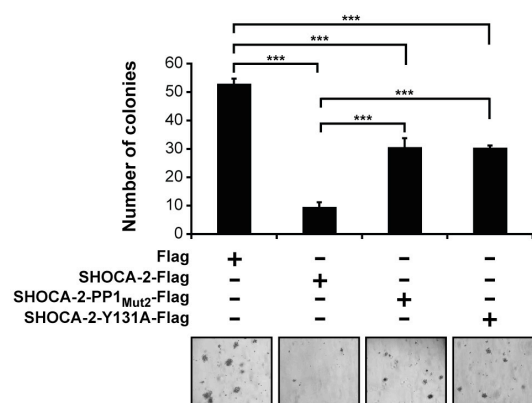
C



D



E



(A) Left panel: Serum-starved HeLa cells that expressed wild type (control) or reduced (SHOCA-2 KD) SHOCA-2 levels were either left untreated or stimulated with EGF (100 ng/ml) for 6 and 12 hrs. Cell lysates were immunoblotted for the detection of phospho-STAT3 (Tyr705), STAT3, c-Myc, Jun B, Cyclin D1, SHOCA-2 and GAPDH. Right panel: After 24 hrs of serum starvation, control and SHOCA-2 KD HeLa cells were stimulated with EGF (100 ng/ml) for 12 hrs. The cell cycle profile was evaluated by flow cytometry using propidium iodide staining (PI). The bar graph shows the percentage of cells in the indicated phases of the cell cycle. The data is representative of 3 separate experiments, each performed in triplicate (* $p < 0.05$; *** $p < 0.005$ using Student's t test).

(B) CRC tissue sections from a cohort of 27 informative CRC patients (see Figure 2E) were analyzed for the proportion of tumor cells that stained for SHOCA-2 and Ki-67 (Wilcoxon test, * $p < 0.05$; *** $p < 0.005$).

(C) SW480 cells transfected to express the indicated empty (pYN3218, Flag) and recombinant vectors were grown for 15 days under neomycin selection and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$).

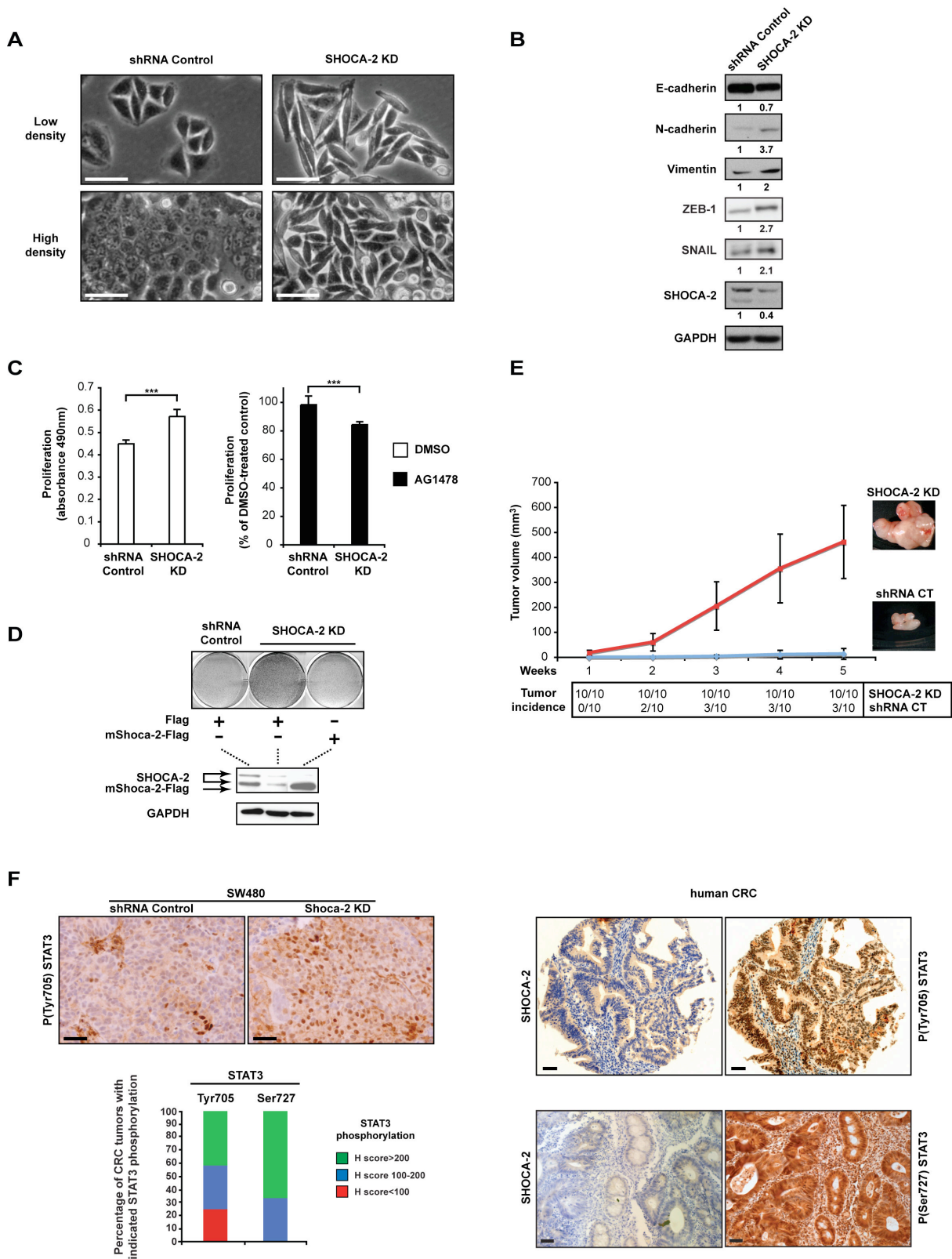
(D) SW480 cells were transiently transfected with a control vector or recombinant vectors as indicated, serum-starved, and then stimulated with EGF (100 ng/ml) for 10 and 15 min. Lysates were immunoblotted to detect phospho-STAT3 (Tyr705 and Ser727), STAT3, phospho-Akt (Ser473), Akt, phospho-ERK1/2 (Thr202/Tyr204), Flag and GAPDH.

(E) SW620 cells expressing wild-type human SHOCA-2 or a SHOCA-2 mutant (PP1_{Mut2}, Y131A) were grown for 15 days under neomycin selection in soft agar and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, *** $p < 0.005$).

See also Figure S5

Figure 6. SHOCA-2 is required for maintaining epithelial morphology and its down-regulation promotes tumorigenesis *in vivo*

Figure 6



(A) Phase-contrast images of control SHOCA-2 expressing SW480 cells and polyclonal SHOCA-2 knockdown (KD) SW480 cells grown at different densities. Scale bar, 100 μ m.

(B) Immunoblot analysis of E-cadherin, N-cadherin, Vimentin, ZEB-1, SNAIL, SHOCA-2 and GAPDH protein expressions in lysates from SW480 stable transfectants (shRNA control, SHOCA-2 KD).

(C) Proliferation of SW480 cells that have been stably transfected with shRNA for the knock-down of SHOCA-2. The cells have been grown in the presence or absence of EGFR inhibitor (AG1478). Data represent the mean \pm SEM of two independent experiments.

(D) SW480 cells with or without a knock-down of SHOCA-2 were transfected to express either a control plasmid (designated Flag) or a murine (m) SHOCA-2-Flag fusion protein that was not targeted by the shRNA used. Transfected cells selected with neomycin and scored for the number of colonies detected (top). Cell lysates were immunoblotted for SHOCA-2 and GAPDH protein expression (bottom). The data are representative of two independent experiments.

(E) Monitoring tumor incidence and size in nude mice (n=5 mice per group) grafted with either SHOCA-2 wild type (shRNA CT) or knock-down SW480 cells (SHOCA-2 KD). Data represent the mean \pm SEM of two independent experiments.

(F) Immunohistochemistry of grafted SW480 tumor cells either wild type or knock-down for SHOCA-2 expression (upper left panel; scale bar, 100 μ m) and of human CRC tissues (right; scale bar, 50 μ m). Left upper panel: Immunohistochemical analysis of in vivo grafted SW480 cells for STAT3 phosphorylated at Tyr⁷⁰⁵; Left lower panel: H-score analysis for Tyr⁷⁰⁵ and Ser⁷²⁷ STAT3 phosphorylation in CRC samples with low to absent SHOCA-2 expression, displayed as a relative percentage of the biopsies investigated. Right: analysis of CRC tissues on consecutive section for the detection of SHOCA-2 and STAT3 phosphorylated either at Tyr⁷⁰⁵ or Ser⁷²⁷.

See also Figure S6

Supplementary Information

The 8p21.3 encoded SHOCA-2 represses STAT3 activation and acts as a tumor suppressor in colorectal cancer

Sébastien Loeffler, Michal Kovac, Stefan Weis, Salvatore Piscuoglio, Saulius Zuklys, Marcel Keller, Katrin Hafen, Daniel Hess, Kaspar Truninger, Inti Zlobec, Luigi Terracciano, Karl Heinimann, Primo Schär and Georg A. Holländer

Inventory of Supplemental Information

Figure S1, related to Figure 1

Table S1, related to Figure 1

Figure S2, related to Figure 2

Figure S3, related to Figure 3

Figure S4, related to Figure 4

Table S2, related to Figure 4

Figure S5, related to Figure 5

Figure S6, related to Figure 6

Supplemental Experimental Procedures

Supplemental References

Supplementary Figure 1

Supplementary Data

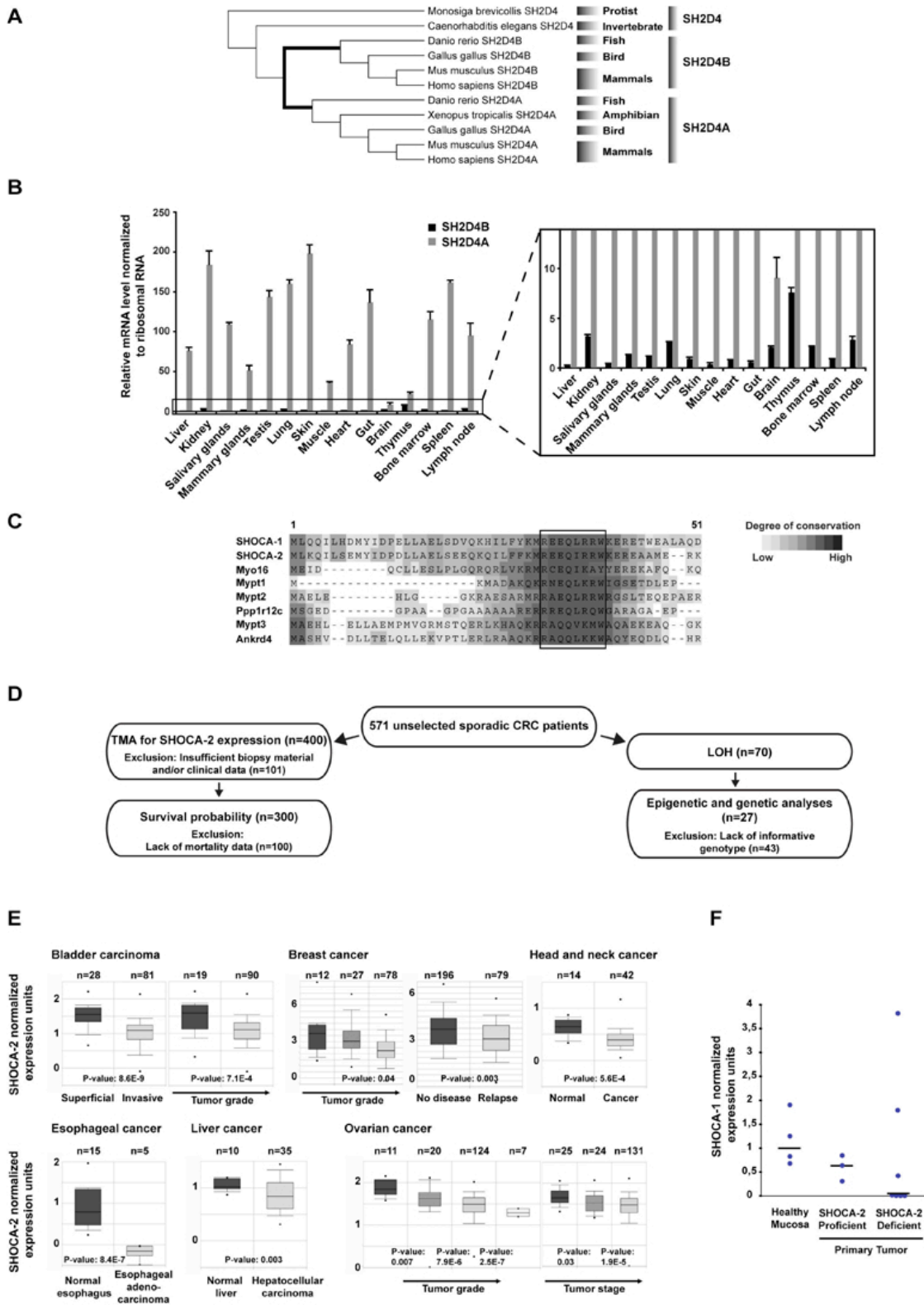


Figure S1. Related to Figure 1. SHOCA sequence analyses and expression profiling in mouse tissue and in tumors. (A) The figure represents an unrooted evolutionary tree analysis of *SH2D4* paralogues using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method. Concurrent with the development of vertebrates (thick line), two distinct *SH2D4* family members developed. (B) Tissues from 10 week-old Balb/c mice were analyzed for *SH2D4B* and *SH2D4A* expression using real time RT-PCR. The results are normalized to 18S rRNA levels. The black bars correspond to *SH2D4B* mRNA levels and the gray bars correspond to *SH2D4A* mRNA levels. (C) The N-terminal amino acid sequences (AA1-51) of human SHOCA-1 and human SHOCA-2 are aligned with human myosin homologues: MYO16 (Myosin-XVI), MYPT1 (Myosin phosphatase-targeting subunit 1, a.k.a. PPP1R12A), MYPT2 (Myosin phosphatase-targeting subunit 2, a.k.a. PPP1R12B), PPP1R12C (Protein phosphatase 1 regulatory subunit 12C), MYPT3 (Myosin phosphatase-targeting subunit 3, a.k.a. PPP1R16A) and ANKRD4 (Ankyrin repeat domain-containing protein 4, a.k.a. PPP1R16B). The box identifies the MyPhone consensus sequence typically present in myosin homologues. (D) From a cohort of 571 unselected patients with sporadic CRC, paraffin-embedded tissue for tissue microarrays (TMA) was available for 501 patients. SHOCA-2 expression analysis was carried out on 400 samples with sufficient tumor tissue (>50%/biopsy) and documented disease stage. Given an informative immunohistochemical analysis and clinical data, survival probability could be calculated for 300 of these patients. LOH was investigated on fresh frozen tissues for 70 of the 571 patients, and epigenetic and genetic analyses were carried out on 27 cases informative for LOH. (E) SHOCA-2 mRNA levels were analyzed by using 9 public gene expression datasets available for meta-analysis (Oncomine)¹. n designates the number of cases included in the analysis. The Y-axis represents expression units that have been normalized as described elsewhere (<http://www.compendiabiotech.com/support/normal.htm>). Shaded boxes represent the interquartile range (i.e. the 25th-75th percentile); whiskers define the 10th-90th percent range; bars indicate the median value; and closed circles identify outliers. The p-value was calculated using Student's t-test. (F) Tissues from human healthy mucosa (n= 4) and tumors (n=10) were analyzed for SHOCA-1 expression using real time RT-PCR. The results are normalized to GAPDH levels. The dots correspond to the relative SHOCA-1 expression of each sample and the bars represent the group's median value. The p-value was calculated using Mann-Whitney test.

Supplementary Table 1

A

H-score		P-value
<100	Stage I vs II	<0.0001
	Stage I vs III	0.0013
	Stage I vs IV	0.0588
	Stage II vs III	0.3692
	Stage II vs IV	0.0143
	Stage III vs IV	0.1083
100-200	Stage I vs II	0.0039
	Stage I vs III	<0.0001
	Stage I vs IV	0.0003
	Stage II vs III	0.1025
	Stage II vs IV	0.3865
	Stage III vs IV	0.4386
>200	Stage I vs II	<0.0001
	Stage I vs III	0.0005
	Stage I vs IV	0.7532
	Stage II vs III	0.4610
	Stage II vs IV	<0.0001
	Stage III vs IV	0.0002

B

H-score		P-value
100-200 vs >200	Stage I vs II	0.2828
	Stage I vs III	0.0189
	Stage I vs IV	0.0009
	Stage II vs III	0.0739
	Stage II vs IV	0.002
	Stage III vs IV	0.1298
<100 vs >200	Stage I vs II	0.0139
	Stage I vs III	0.0368
	Stage I vs IV	0.112
	Stage II vs III	0.5808
	Stage II vs IV	0.5409
	Stage III vs IV	0.8622

C

	Comparison	Log-rank (p-value)	Wilcoxon (p-value)	Cox (p-value)	HR (95% CI)
H-scores	100-200 vs >200	0.2312	0.2164	0.2365	0.81 (0.57-1.15)
Stage	Stage I (no events)	-	-	-	
	Stage II	0.146	0.096	0.159	0.67 (0.38-1.17)
	Stage III	0.582	0.75	0.586	0.82 (0.39-1.7)
	Stage IV	0.843	0.824	0.846	0.94 (0.51-1.74)
H scores	<100 vs >200	0.0118	0.0133	0.0134	0.56 (0.35-0.89)
Stage	Stage I (no events)	-	-	-	
	Stage II	0.355	0.553	0.36	0.62 (0.22-1.73)
	Stage III	0.548	0.551	0.55	0.78 (0.34-1.77)
	Stage IV	0.569	0.609	0.58	0.82 (0.41-1.64)

Table S1. Related to Figure 1. Statistical analyses for association of SHOCA-2 H-score, stage and survival time (A) Differences in the stage distribution among each H-score category (binomial test for equal proportions). (B) Unadjusted p-values for association of H-scores and stage (Chi-Square or Fisher's Exact test). (C) P-values for survival time differences overall and stratified by stage using three different statistical tests.

Supplementary Figure 2

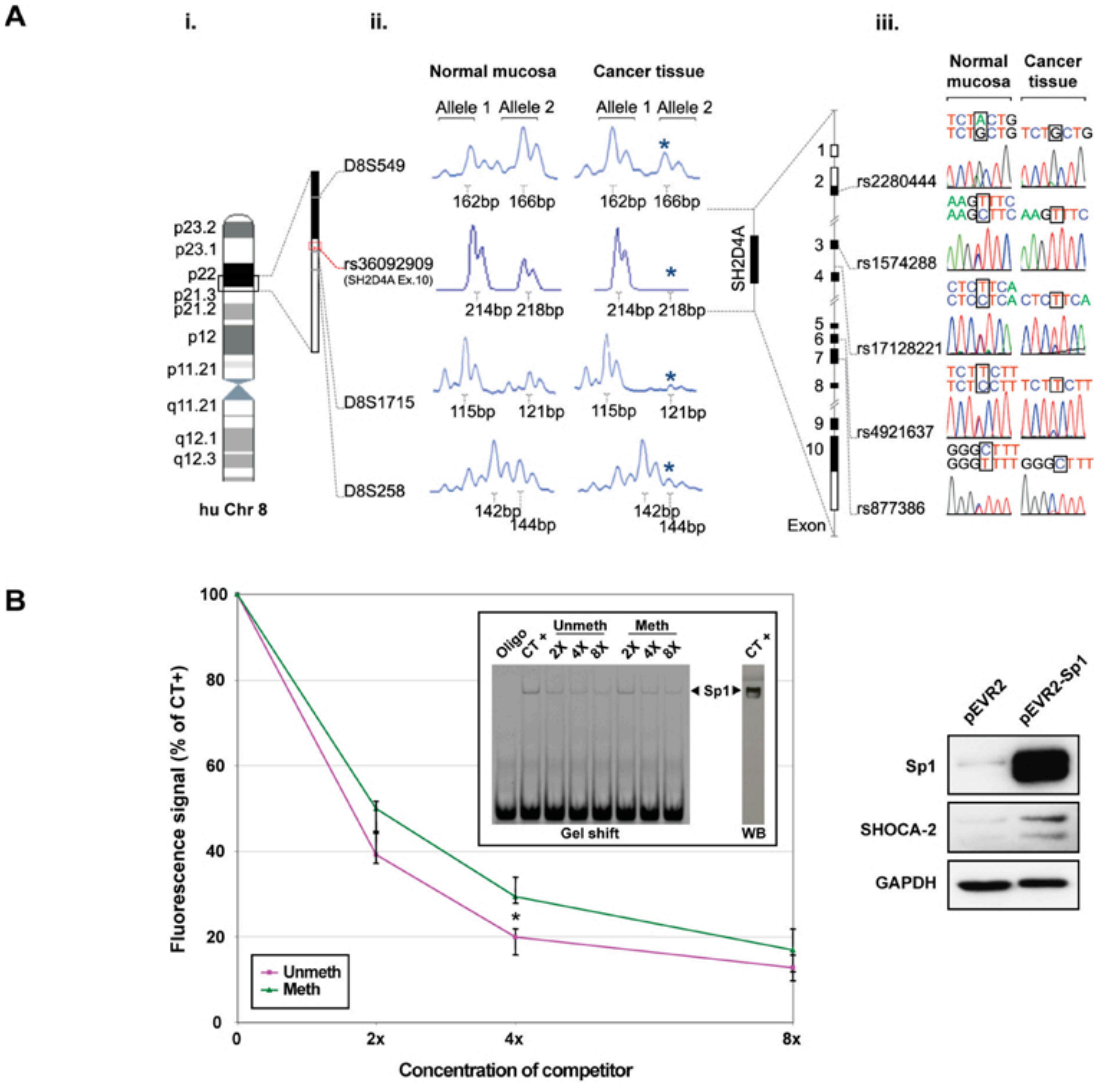
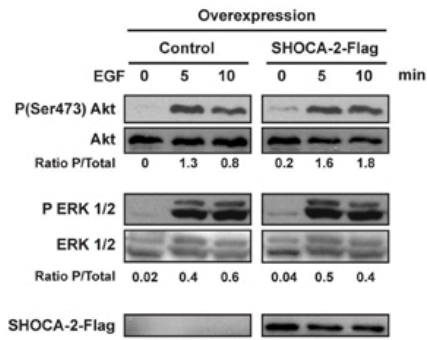


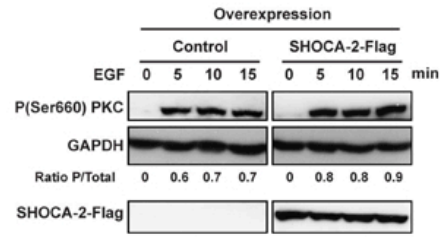
Figure S2. Related to Figure 2. Epigenetic and genetic analyses on *SH2D4A* gene. (A) Loss of Heterozygosity (LOH) analysis for *SH2D4A*. Loss of heterozygosity analysis for the *SH2D4A* locus. Panels **i** and **ii** depict 3 flanking and 1 gene-specific microsatellite marker, respectively, that cover the region (approximately 7Mb) from 8p21.3 to the telomeric boundary of 8p22. Panel **iii** demonstrates for cancer sample number 17 the analysis of 5 gene-specific SNPs and reveals a selective loss of a single *SHOCA-2* allele. (B) Left, Sp1 binding affinity differs between methylated and unmethylated *SH2D4A* DNA oligomers. The Sp1 specific gel-shifts were carried out using a fluorescein-labeled oligomer binding Sp1 (CT+). CT+ and SW48 nuclear extract were incubated in a competitive manner in the absence or presence of increasing concentrations of an unlabeled oligomers. This oligomer sequence, which represents part of the genomic *SH2D4A* sequence, had either an unmethylated (Unmeth) or a methylated (Meth) Sp1 binding site. As a negative control (designated Oligo), CT+ oligomer was incubated in the absence of nuclear extract. Data from four independent experiments are shown and represent the mean and the interquartile range (i.e. the 25th-75th percentile) of the percentage of the fluorescence signal obtained with experimental samples in comparison to control (CT+; *p < 0.05 using Mann-Whitney test). The presence of Sp1 in the shifted bands was confirmed by western blotting (WB). Right, HeLa cells were transiently transfected with either a control vector or a vector encoding Sp1; immunoblotting for the detection of the indicated proteins.

Supplementary Figure 3

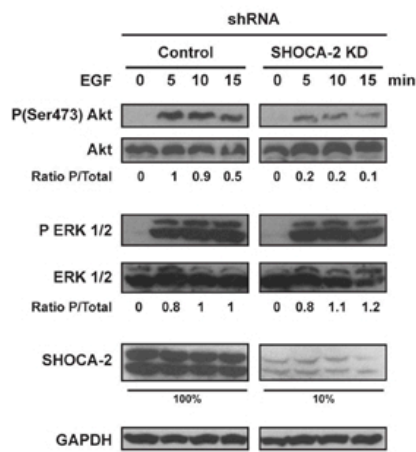
A



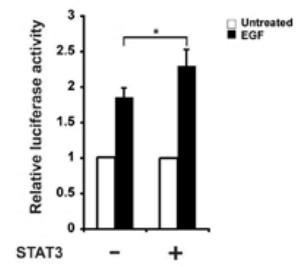
B



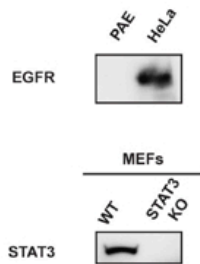
C



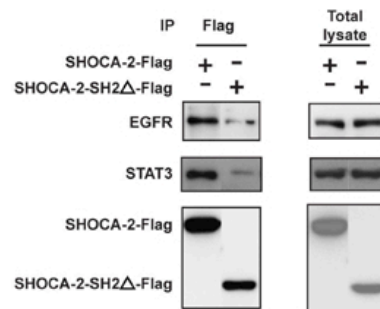
D



E



F



G

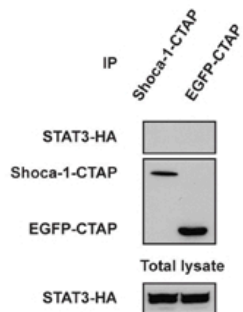


Figure S3. Related to Figure 3. Effect of SHOCA-2 on EGFR-mediated signaling. (A,B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag sequence and were then stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry. (C) HeLa cells were stably transfected either with a plasmid encoding an shRNA specific for SHOCA-2 or with a plasmid devoid of an insert. Following serum starvation for 24 hrs, the cells were stimulated with EGF (100 ng/ml) as indicated. Western blotting for the indicated proteins were performed on aliquots of transfected cell lysates. The efficiency of SHOCA-2 silencing is indicated as a percentage of control transfected cells. Densitometry was used to determine the ratio of phosphoprotein to total protein signal (designated P/Total). (D) A SIE firefly luciferase reporter construct and a renilla luciferase plasmid were transiently transfected together with a STAT3 expression vector into unmanipulated HeLa cells. Transfected cells were serum-starved for 24 hrs and then either stimulated with EGF (100 ng/ml; filled bar) for 6 hrs or left untreated (open bar). The renilla luciferase activity was used to normalize for transfection efficiency. Data are representative of 3 experiments, each performed in triplicate (*p < 0.05; **p < 0.01; ***p < 0.005 using Student's t test). (E) Immunoblotting of lysates from PAE cells, HeLa, and wild type and Stat3 deficient MEFs for the detection of EGFR and Stat3, respectively. (F) HeLa cells were transfected to express a Flag-fusion protein of either the wild-type form of human SHOCA-2 or a mutant form of human SHOCA-2 that lacks the SH2 domain (designated SHOCA-2-SH2 Δ). Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analyzed by immunoblotting using antibodies specific for Flag, STAT3 and EGFR (left panel). Total lysates were used as loading controls (right panel). (G) HEK293 cells were transfected to express STAT3-HA and a C-terminal TAP-tag fusion proteins of SHOCA-1 or EGFP. Cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibodies specific for HA-tag or with Rabbit IgG recognizing the protein A component of the TAP-tag.

Supplementary Figure 4

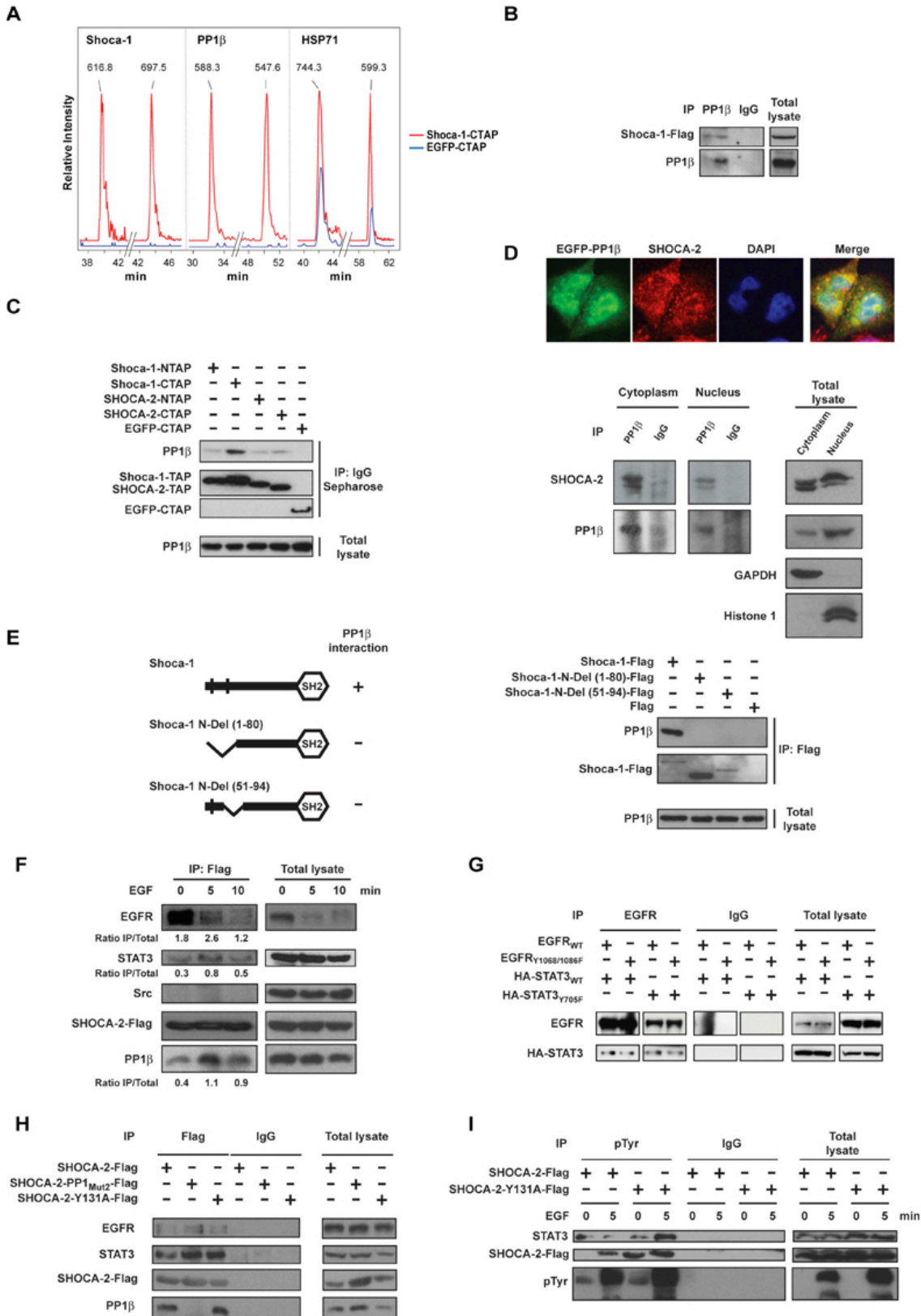


Figure S4. Related to Figure 4. Interaction of SHOCA-2 with the EGFR/STAT3/PP1 β complex. (A) Multiple reaction monitoring (MRM) comparison of two selected peptides from SHOCA-1, PP1 β and HSP71, respectively. Red and blue lines represent peptides retrieved from the SHOCA-1-CTAP and control EGFP-CTAP pull-downs, respectively. The transition signals are labeled with the mass (m/z) of the peptides. Peptides derived from non-specific interactions (e.g. heat shock protein HSP71) were identified in both eluates whereas SHOCA-1 and PP1 β peptides were only found in the SHOCA-1-CTAP immunoprecipitation. (B) Cell lysates from HEK293 cells transfected to express SHOCA-1-Flag fusion protein were immunoprecipitated either with an antibody specific for PP1 β or with goat IgG and then immunoblotted with antibodies specific for Flag and PP1 β . (C) HEK293 cells were transfected to express the indicated TAP-tag fusion proteins. The cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibody specific for PP1 β (loading control) or with Rabbit IgG recognizing the protein A component of the TAP-tag. (D) Top, HeLa cells expressing an EGFP-PP1 β fusion protein were stained with anti-SHOCA-2 antibody and DAPI to detect colocalization by confocal microscopy. Bottom, endogenous SHOCA-2 was detected in cytoplasmic (GAPDH+) and nuclear (Histone H1+) HeLa cell lysates immunoprecipitated with anti-PP1 β . (E) HEK293 cells were transfected to express either a wild-type SHOCA-1-Flag fusion protein or one of two deletion mutants AA1-80 or AA51-94, fused to Flag (left panel; the alleged PP1-binding sites are indicated as vertical bars). Lysates from transfected cells were immunoprecipitated with anti-Flag antibodies and immunoblotted with antibodies directed against Flag or PP1 β (right panel). Interaction of wild type and mutant SHOCA-2 with EGFR and STAT3 (F) Serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with anti-Flag antibody. EGFR, STAT3, c-SRC, Flag, PP1 β were detected by immunoblotting. (G) Lysates of HeLa cells expressing HA-tagged wild-type (WT) STAT3 or the Y705F STAT3 mutant and wild-type (WT) EGFR or the Y1068/1086F EGFR mutant were immunoprecipitated with an EGFR specific antibody. EGFR and HA were detected by immunoblotting. (H) HeLa cells were transfected with an expression plasmid encoding a Flag-fusion either wild-type (WT), PP1_{Mut2} or Y131A mutant protein of human SHOCA-2. Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analysed by immunoblotting using antibodies specific for Flag, STAT3, EGFR and PP1 β (left panel). Total lysates were used as loading controls (right panel). (I) Serum-starved HeLa cells expressing human wild-type SHOCA-2-Flag or the Y131A mutant were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with a phosphotyrosine (pTyr) specific antibody. STAT3, Flag and pTyr were detected by immunoblotting.

Supplementary Table 2

Uniprot	Description	Mass	Shoca1-CTAP		EGFP-CTAP	
			Score	Spectral count	Score	Spectral count
P07437	Tubulin beta chain	49671	1226	56	360	11
B2RSQ7	SH2 domain containing 4B (Shoca1)	51120	926	61	0	0
P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	37187	747	35	0	0
P81947	Tubulin alpha-1B chain	50152	673	51	265	10
P08107	Heat shock 70 kDa protein 1	70052	343	17	88	2
P04517	Genome polyprotein	346164	279	13	182	9
P68104	Elongation factor 1-alpha	50141	228	10	81	3
P08238	Heat shock protein HSP 90-beta	83264	219	5	168	4
P62158	Calmodulin	16838	118	4	266	13
Q9H1R3	Myosin light chain kinase 2	64685	57	6	181	13
P42212	Green fluorescent protein	26886	0	0	357	23

Table S2. Related to Figure 4. LC-MS/MS analysis: comparison of peptides eluted following SHOCA-1- and EGFP-TAP purification. Proteins were identified using 2 or more unique peptides with a MASCOT score above 20. To distinguish between correct and incorrect peptide assignments, protein spectra with less than 4 peptides were manually evaluated. Spectral count: number of spectra identified for each protein.

Supplementary Figure 5

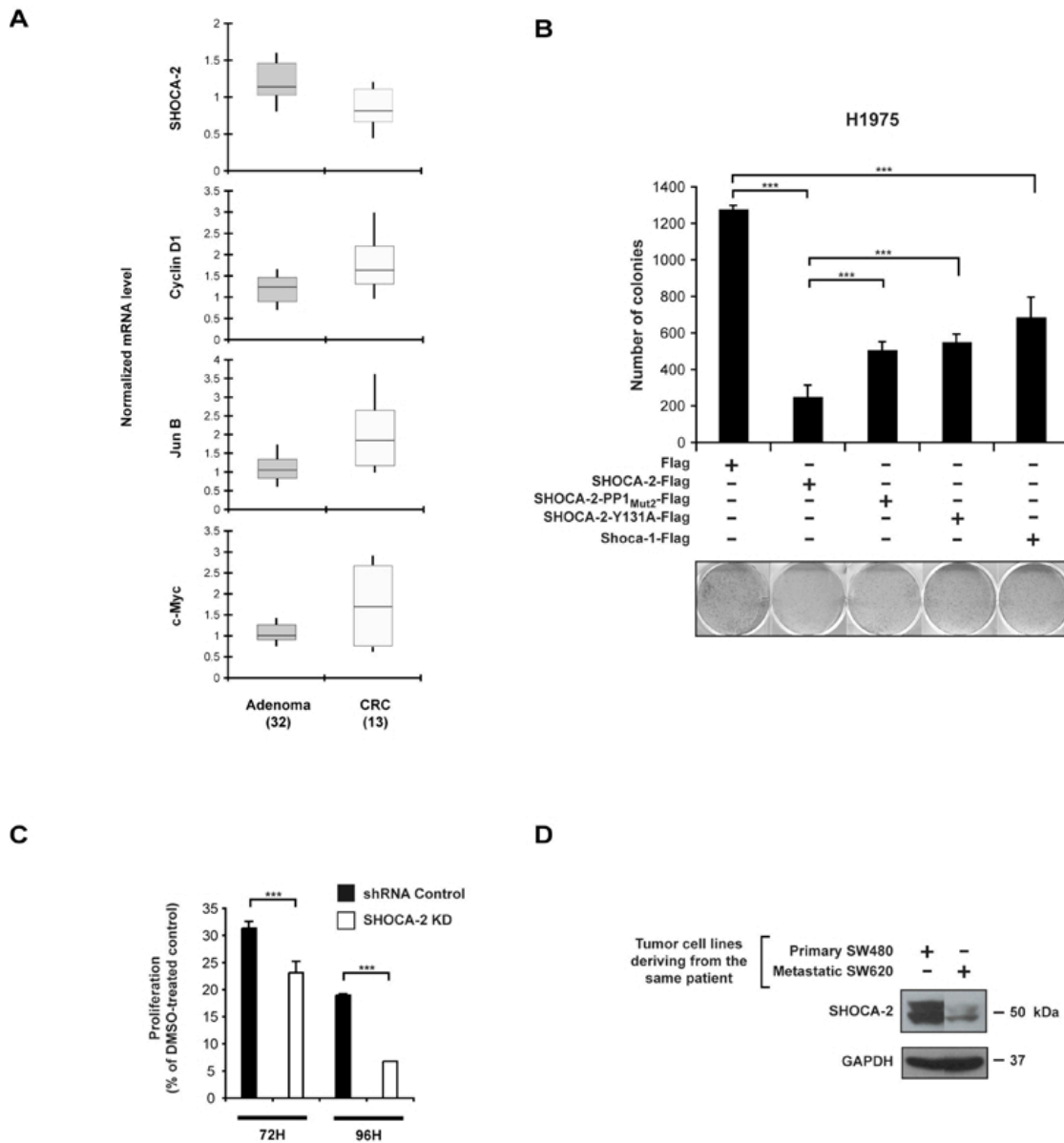
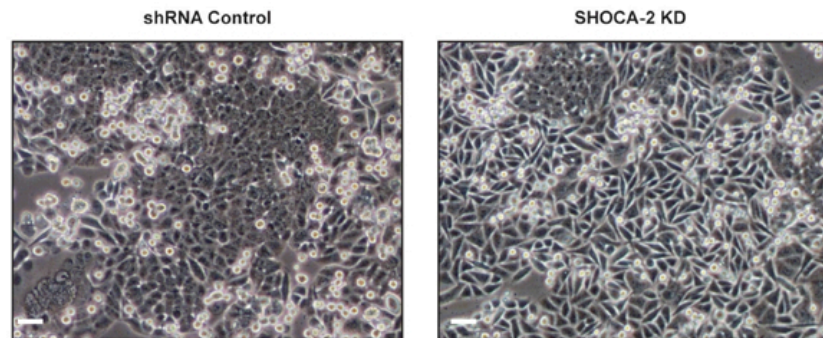


Figure S5. Related to Figure 5. SHOCA-2 influences cell cycle progression and proliferation in STAT3-dependent fashion. (A) mRNA expression of SHOCA-2, Cyclin D1, Jun B, c-Myc, was assessed in 32 adenoma and 13 sporadic colorectal cancers by means of Affymetrix GeneChip U133 expression arrays. Shaded boxes represent interquartile range (i.e. the 25th-75th percentile); whiskers indicate the 10th-90th percentile; bars provide the median value. (B) H1975 cells were transfected to express either a vector without insertion (designated Flag) or a vector encoding Flag-fusion proteins of either wild type SHOCA-2, a SHOCA-2 mutant unable to bind to PP1 β (SHOCA-2-PP1_{Mut2}), or a SHOCA-2 mutant in which the tyrosine at position 131 was exchanged for an alanine (SHOCA-2-Y131A), or wild type Shoca-1. The cells were grown for 15 days under neomycin selection and then scored for the number (bar graphs) and appearance of colonies (microphotographs). The data is representative of two independent experiments performed in triplicates (Student's t test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$). (C) Proliferation of SW480 cells that have been stably transfected with a control (shRNA control) vector or a recombinant vector to knock-down SHOCA-2 expression (SHOCA-2 KD). Successfully transduced cells were grown either in the presence or in the absence of the STAT3 inhibitor S31-201 diluted in DMSO (100 μ M). Data represent the

mean \pm SD of two independent experiments carried out in triplicates. (D) SHOCA-2 protein expression in the CRC cell lines SW480 (primary tumor) and SW620 (metastasis from the same patient) was assessed using Western-blotting.

Supplementary Figure 6

A



B

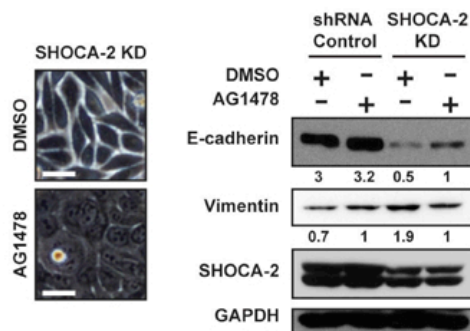


Figure S6. Related to Figure 6. SHOCA-2 knock-down SW480 cells display features of EMT. (A) Low power view of SW480 cells expressing either a control vector (shRNA control, left panel) or a vector to knock-down SHOCA-2 expression (SHOCA-2 KD; right panel). Scale bar, 100 μ m. (B) Left, phase-contrast images of Shoca-2 KD SW480 cells grown in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10 μ M; scale bar: 50 μ m). Right, immunoblot analysis of E-cadherin, Vimentin, Shoca-2 and GAPDH protein expressions in lysates from shRNA control and Shoca-2 KD SW480 cells cultured in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10 μ M).

Supplemental Experimental Procedures

Plasmids, antibodies and reagents

Mouse and human SHOCA-1 and SHOCA-2 sequences were amplified from thymus cDNA libraries by PCR, subcloned into the expression vectors pcDNA-Flag 3.1, CTAP and NTAP (a kind gift from Dr. A.C. Gingras, Samuel Lunenfeld Research Institute at Mount Sinai Hospital, Toronto, Canada) and sequenced. Site-directed mutagenesis (Qiagen) was used to introduce specific mutations. The human EGFP-PP1 α , β , γ 1 constructs were generously provided by Dr. L. Trinkle-Mulcahy (Wellcome Trust Centre for Gene Regulation and Expression, University of Dundee, Scotland, UK), the EGFP-CTAP plasmid was a gift from Dr. S. Meloche (Institute of Research in Immunology and Cancer, Université de Montréal, Canada), the SIE firefly luciferase reporter plasmid was kindly made available by Dr. I. Behrmann (Life Sciences Research Unit, University of Luxembourg, Luxembourg), the EGFR wild-type and Y1068/1086F mutant constructs were generously provided by Dr. S. Sigismund (IFOM, Milano, Italy) and the STAT3 expression vector was obtained courtesy of Dr. J.Y. Yoo (Pohang University of Science and Technology, Pohang, Republic of Korea). Antibodies used were directed against PP2A (Upstate); Flag, ERK 1/2, anti-vimentin (all from Sigma); Histone H1, Jun B, c-Myc, cyclin D1, c-SRC and ZEB-1 (all from Santa-Cruz); GAPDH, SNAIL and PP1 β (all from Abcam); SH2D4A (Abnova); EGF receptor, Tyr845-phosphorylated EGFR, Ser660-phosphorylated PKC, phospho-tyrosine (designated pTyr), Tyr705/Ser727-phosphorylated STAT3, Akt, Ser473-phospho-Akt, Thr202/Tyr204-phospho-ERK 1/2 (all from Cell Signaling); GFP (Roche); anti-N-cadherin (Takara Biomedicals); anti-E-cadherin and STAT3 (both from BD Biosciences). EGF was obtained from Peprotech and used at

a concentration of 100 ng/ml. AG1478 was purchased from Axxora and the STAT3 inhibitor (S3I-201) was obtained from Merck.

Cell culture and transfection

HEK293, HeLa, SW480, SW620 cells (a kind gift from Dr. G. Marra, Institute of Molecular Cancer Research, University of Zurich, Switzerland), Murine embryonic fibroblasts (MEF) either wild type (fl/fl) or deficient for Stat3 (MEF Δ/Δ), a generous gifts of Dr. Valeria Poli, University of Turin, Torino, Italy, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Porcine aortic endothelial (PAE) cells (kindly made available by Dr. P. Berger, Paul Scherrer Institut, Villigen, Switzerland) were grown in F-12 Nutrient Mixture (Ham) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Transient transfections were accomplished using Lipofectamine 2000 (Invitrogen) for HEK293 and Fugene HD (Roche) for HeLa, SW480, SW620 cells according to the manufacturer's recommendations. For stable pShoca-1-CTAP or pEGFP-CTAP transfectants, HEK293 cells were co-transfected with the puromycin resistance plasmid, pBabe-puro (Cell Biolabs). Cells were first plated onto 100-mm plates and allowed to grow to 80% confluence, then rinsed twice with phosphate-buffered saline (PBS) and finally exposed to the plasmids/Lipofectamine-DMEM mixture. After overnight incubation, the cells were washed, split into a 150-mm plate and for selection grown in DMEM containing puromycin (10.8 μ g/ml for the first 48 hrs and then 3.6 μ g/ml). Several puromycin resistant clones were established. Cells expressing comparable amounts of either Shoca-1-CTAP or EGFP-CTAP protein (as analyzed by western-blotting) were selected for further analysis. For the stable knock-down, HeLa and SW480 cells were

transfected using different human SHOCA-2 (*SH2D4A*)-specific shRNA constructs (Origene).

Bioinformatic analysis

Sequences from SHOCA-1 (SH2D4B) and SHOCA-2 (SH2D4A) were retrieved by homology search with ENSEMBL²: *C. elegans* SHOCA, F13B12.6; *Danio rerio* SHOCA-1, ENSDARG00000015144; *Gallus gallus* SHOCA-1, ENSGALG00000002370; *Mus musculus* SHOCA-1, ENSMUSG00000037833; *H. sapiens* SHOCA-1, ENSG00000178217; *Danio rerio* SHOCA-2, ENSDARG00000041378; *Xenopus laevis* SHOCA-2, ENSXETG00000004464; *Gallus gallus* SHOCA-2, ENSGALG00000010122; *Mus musculus* SHOCA-2, ENSMUSG00000053886; *H. sapiens* SHOCA-2, ENSG00000104611. The common SHOCA precursor (Sh2d4) sequence of *Monosiga brevicollis* (30425.AA) was identified in the KinBase database (<http://www.kinase.com/kinbase/>). Sequences were aligned with the Toffee software (<http://tcoffee.vital-it.ch>)³. An unrooted evolutionary tree of SHOCA homologues was generated using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). The gene microarray analysis tool made available by Oncomine¹ (Compendia Bioscience, Ann Arbor, MI, USA: <http://www.oncomine.org>) was used to determine SHOCA-2 mRNA expression in various cancers.

Quantification of SHOCA-2 expression and STAT3 phosphorylation on tissue microarrays & whole-tissue sections

The H score determines the percentage of positive tumor cells multiplied by their staining intensity⁴. Given the size of the cohort of patients analyzed, three ranges of scores were

considered: >200, identifying strong expression or phosphorylation; 100-200, classifying moderate expression or phosphorylation; <100 characterizing a low or absent expression or phosphorylation.

Statistical analyses

The likelihood ratio test and binomial test for differences in proportions were used to compare H-scores between and among disease stages, respectively⁵. Survival time analysis was performed using three different methods: the non-parametric log-rank test, the non-parametric Gehran-Wilcoxon tests and the Cox regression analysis which was carried out after verification of the assumption of proportional hazards; survival time differences were represented using the Kaplan-Meier method⁶. Hazard ratios (HR) and 95% confidence intervals (CI) were obtained to measure effect size with values <1.0 indicating more favorable survival time with higher protein expression. P-values were two-sided and considered statistically significant when $p < 0.05$.

Real-time RT-PCR

Using TRI reagent (Molecular Research Center), total RNA was prepared from various tissues freshly isolated from 10 week-old Balb/c mice. Oligo(dT20) or random-N6-primed cDNA was generated with M-MLV Reverse Transcriptase (Invitrogen-Gibco) employing standard protocols. For quantitative real-time PCR, the SYBR Green-based method was used (Sensimix, Quantace); primers are available upon request. PCR specificity was controlled by analyzing the melting curves and agarose gel electrophoresis of the PCR products. Amounts of specific mRNA were normalized to levels of 18S rRNA transcripts.

Mass spectrometry

TCA precipitated and acetone washed or Chloroform/Methanol treated protein pellets were reduced with TCEP (tris(2-carboxyethyl)phosphine), alkylated with Iodoacetamide and digested with Trypsin. The generated peptides were analyzed by NanoLC-MS/MS on a 4000Q Trap (MDS Sciex) as described⁷. The proteins were identified with Mascot (Matrix Science) searching the Uniprot database⁸. MRM-buddy, a software developed at the Friedrich Miescher Institut (A. di Cara, R. Sack and R. Portmann, unpublished results) was used to extract the MRM-relevant information for the quantification of selected proteins. MRM analysis was done as described⁷.

Western Blot and immunoprecipitation

Total protein (40µg as quantified by the Bradford method; Biorad) was separated on a 8%-12% SDS-PAGE and transferred to nitrocellulose membranes (Biorad). Membranes were blocked with 5% milk (Migros) in TBS-Tween and subsequently incubated overnight with primary antibody using concentrations that have either been previously established or used according to the manufacturer's recommendations. After a one hour incubation with a secondary antibody conjugated to horseradish peroxidase (Southern Biotech), the reaction was visualized by ECL (Pierce). For immunoprecipitation, cell lysates containing equal amounts of total protein were precleared for 1 hr with 40 µl of 50% (wt/vol) protein G Plus-Sepharose beads (Amersham) and incubated for 2-3 hrs at 4°C with primary or control (IgG) antibodies. The beads were added for 1 hr and then washed extensively with lysis buffer; bound proteins were fractionated on a 8%-12% SDS-PAGE, and analyzed by western blotting as detailed above. Immunoprecipitations were performed using RIPA buffer with the notable exception of analyses shown in Figures 3e and 4d

which were carried out using the JS lysis buffer (50 mM HEPES at pH 7.5, 1% glycerol, 50 mM NaCl, 1% Triton X-100, 1.5 mM MgCl₂, 5 mM EGTA).

Confocal microscopy

HeLa cells grown plated on glass coverslips were transiently transfected to express the EGFP-PP1 β fusion protein. Twenty-four hours later, cells were washed in PBS and fixed for 15 min using 4% paraformaldehyde at room temperature (RT). Following three washes with PBS, cells were permeabilized on ice for 15 min using cold 0.1% Triton X-100 (v/v) in PBS. All staining reactions were performed at RT. To prevent unspecific binding, the cells were exposed for 30 min to 5% normal goat serum diluted in PBS. Next, the cells were incubated for 1 hr with the primary antibody, then washed three times in PBS and finally exposed for 30 min to the secondary antibody, an Alexa 555-conjugated goat anti-mouse antibody (Invitrogen). Prior to mounting, nuclei were stained with DAPI (4',6-diamidino-2-phenylindole, diluted 1:10,000 in PBS from a 1 mg/ml stock) and washed twice with PBS. Coverslips were mounted with fluorescent mounting medium (Dako). Images were acquired using an LSM 510 Meta confocal microscope and LSM510 imaging software (Carl Zeiss).

Cell cycle analysis

HeLa cells stably transfected with shRNA plasmid specific for the knock-down of SHOCA-2 or with a control plasmid were serum-starved for 24 hrs prior to a 12 hrs stimulation with EGF (100 ng/ml). After trypsinization and centrifugation, 5×10^5 cells were resuspended in 0.5 ml of cold hypotonic PI (Propidium Iodide) staining solution (50 μ g/ml PI (Sigma), 0.1 mg/ml RNase A, 0.1% Triton X-100, 0.1% sodium citrate).

Following vigorous resuspension, the cells were incubated (15 min) on ice and then analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases of cell cycle were defined using the mathematical Watson Pragmatic model.

Cell proliferation

SW480 cells (5×10^4 cells/well) were cultured in 96-well plates in the absence or presence of AG1478 (10 μ M) or S3I-201 (100 μ M) in 10% serum condition. After 48 hours of treatment, cell proliferation was analyzed using CellTiter 96 aqueous nonradioactive cell proliferation assay kit (Promega, Madison, WI). The cell proliferation index was calculated as a percentage of the absorbance in relation to the untreated control cells.

Human tissues and DNA extraction

The tissues from primary colorectal carcinomas and the macroscopically healthy-appearing neighboring mucosa (7cm proximal of tumor) were resected on the occasion of surgical cancer therapy. All the tissue samples were immediately submerged into RNA $later^{\text{®}}$ RNA stabilization reagent (Qiagen) and kept at room temperature for several hours to guarantee a complete penetration by the reagent. The vials were then stored at -80°C until the tissue was further processed. DNA from approximate 20 mg of tissue was extracted with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's specifications.

LOH and mutational analysis of the *SH2D4A* gene

All canonical exons, the flanking intronic sequences as well as 5' and 3' UTR regions were subjected to PCR amplification (Kapa HiFi™ system ; Kapa Biosystems, USA) and sequenced (BigDye terminator kit v1.1 on ABI 3130xl; both Applied Biosystems, USA) using the set of primers described in the supplementary table. Putative mutations were confirmed by bi-directional sequencing of a second independent PCR product. *In Silico* mutation impact prediction scores (SIFT: <http://sift.jcvi.org> and PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2>) were calculated for all somatic point alterations found. Loss of heterozygosity was assessed using 3 flanking microsatellite markers (D8S549, D8S1715 and D8S258 labeled at 5' end with 6-Carboxyfluorescein (6-FAM) dye) as well as 6 single nucleotide polymorphisms refSNP rs2280444, rs877386, rs1574288, rs17128221, rs4921637, rs36092909) located within the *SH2D4A* gene itself. The relative proportions of alleles 1 and 2 in non-tumor and tumor samples corresponds to the ratio of the heights of the corresponding peaks. An LOH index was calculated by dividing the ratio of allele 2 to allele 1 in the non-tumor DNA by the corresponding ratio in the tumor DNA. LOH positivity was defined as an LOH index of <0.5 (reflecting a substantial loss of allele 1 in the tumor sample) or >1.5 (indicative of substantial loss of allele 2).

Identification and characterization of exon 4 skipping variant of the human *SH2D4A* gene by RT-PCR

First strand cDNA was synthesized from 500ng of total RNA using random-primed reverse transcription (Verso cDNA; Thermo Scientific). PCR amplification was next carried out on a cDNA template, products were separated on a 2% agarose gel and visualized with ethidium bromide. Individual bands were excised from the gel, eluted

(Quiaquick gel extraction kit; Quiagen, Germany), reamplified and sequenced.

Conditions and primers for sequence, LOH and cDNA analyses

For the *SH2D4A* sequence analysis the following PCR conditions were used: initial denaturation of 2min at 94°C followed by 35 cycles of 10sec at 94°C, of 15sec at 56°C and of 15sec at 72°C and a final elongation step of 3min at 72°C. When using the ABI Prism[®] TrueAllele[®] premix, the initial denaturation step was extended to 10 minutes. The following primers were used:

Mutational Analysis	DNA sequence 5'→3'
SH2D4A_x1_For	CGATTGCGCCCCGCCAGTCA
SH2D4A_x1_Rev	CTCCCATGCACCCATGCAAC
SH2D4A_x2_For	GGAACTTTTGCCACAAGTAT
SH2D4A_x2_Rev	CTTGACCTGCTTTGCTTG
SH2D4A_x3_For	ATTGCTACCTACAGATGTTC
SH2D4A_x3_Rev	AAGTGAACTTTTGTGACCCTC
SH2D4A_x4_For	ACGTCTTTACACACCTAGC
SH2D4A_x4_Rev	AAATTCCGCACGTCTACAG
SH2D4A_x5_For	TCACATGCCCGTTGCTGT
SH2D4A_x5_Rev	ATGCCTAGGTATGCAGCCC
SH2D4A_x6_For	AGGAAACTGCTGGATTTGCT
SH2D4A_x6_Rev	ATCAGCTGAGAGCCTGGAAG
SH2D4A_x7_For	ATAAGCAGACAACAAACTG
SH2D4A_x7_Rev	CTCATACAATTCCCCTTTAA
SH2D4A_x8_For	TATATGACTTTTGAGGGC
SH2D4A_x8_Rev	TTCTGGTAGATTAGCACAG
SH2D4A_x9_For	TGCTCTTATTGTGTCTTATTTAGGC
SH2D4A_x9_Rev	CTCAGAAGGCAAATTAAGTT
SH2D4A_x10_For	GGTCACATGTAGATAAGAAA
SH2D4A_x10_Rev	AAAACAAACTCAGCGATGT
LOH Analysis	
D8S1715_For	CAGGTGATGTCCCAGAGG
D8S1715_Rev	CGAACATGAATTAGAAATCCAGTG
D8S258_For	CTGCCAGGAATCAACTGAG
D8S258_Rev	TTGACAGGGACCCACG
D8S549_For	AAATGAATCTCTGATTAGCCAAC
D8S549_Rev	TGAGAGCCAACCTATTTCTACC
rs36092909_For	ACAAATGCCACTGCAACATT

rs36092909_Rev	TACATCATTAAAGGTCATAC
cDNA Analysis	
SH2D4A_cx2.1_For	CAGACCAAAGAAAGAGAATGGCA
SH2D4A_cx2_Rev	ATCTGCCAACATTTGTTGAA
SH2D4A_cx2.1_Rev	CTTTGGATTTTCGCAGAGATG

Multiplex-ligation dependent-probe-amplification (MLPA)

Chromosome 8 copy-number variation was investigated by multiplex ligation-dependent probe amplification assay (kit P014-1A, MRC Holland) according to the manufacturer's protocols. Calculations were performed using GeneMarker software (SoftGenetics). Gene dosage with relative value of 1 is expected for two copies, whereas values of 0.5 and 1.5 indicate loss and gain of one copy, respectively. MLPA results indicative of a somatic deletion were independently confirmed in at least one additional experiment.

Methylation analysis

Genomic DNA from primary tumor samples and adjacent matched healthy appearing mucosa, respectively, was extracted using the QIAamp DNA Mini Kit (Qiagen), and bisulfite converted using the EZ DNA Methylation KitTM (Zymo Research). Promoter associated CpG-island regions of *SH2D4A* were amplified by PCR and biotin labeled either directly using biotin-labelled reverse primers or during a second PCR step using a biotinylated universal primer annealing to a sequence introduced by the reverse primer used in the first PCR. PCR conditions and primer sequences are provided on supplementary table. Labelled PCR products were purified (QIAquick[®] PCR Purification Kit, Qiagen), bound to streptavidin-covered sepharose beads and analyzed by pyrosequencing using the PyroMark Q24 pyrosequencing system (Qiagen). Sequencing-assay design as well as quantification of methylation was performed with the PyroMark

Q24 1.0.10 software. Peripheral blood DNA treated with M.SssI methyltransferase (New England BioLabs) and subsequently bisulfite treated served as positive control.

Conditions and primers for methylation analysis

PCR Amplification of CpG islands	
1 st CpG island forward	5'-GTT TTA TGA ATT AAG AGA GGA GAG GAT AAG TTT ATT T-3'
1 st CpG island reverse	5'-GCC CCC GCC CGC CCC CAA CAC CTA CCA ATA C-3'
Biotinylated primer	5'-Bio-GCC CCC GCC CG-3'
2 nd CpG island forward	5'-AG GTT GTA TGG GTG TAT GGG AG-3'
2 nd CpG island biotinylated reverse	5'-Bio-CAA CTC AAA AAA CCT ACA AAC TAT AAC TCC TTC-3'
PCR conditions:	
1 st CpG island 1 st PCR	Enzyme activation: 95°C for 2min 40 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 25 sec Extension: 72°C for 30 sec 3.5mM MgCl ₂ , 0.5μM primer
1 st CpG island 2 nd PCR	Enzyme activation: 95°C for 2min 20 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 30 sec Extension: 72°C for 30 sec 2.5mM MgCl ₂ , 0.5μM primer
2 nd CpG island	Enzyme activation: 95°C for 4min 35 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 30 sec Extension: 72°C for 55 sec 2.5mM MgCl ₂ , 0.25μM primer
Pyrosequencing primers	

1st CpG island	
Primer 1	5'- GGA GGA GAG GAT AAG T -3'
Primer 2	5'- GTT TTG GAG AGT TTT TAG -3'
Primer 3	5'- GTT ATA AGA GTT GTT TTG -3'
Primer 4	5'- GGT TAA GTG GAT GTG -3'
2nd CpG island	
Primer 1	5'- TAT GGG TGT ATG GGA G -3'
Primer 2	5'- GGG TTT AGG TG -3'
Primer 3	5'- GTT GGT TAG GTG AG -3'
Primer 4	5'- AGG GAG AGG GTT TG -3'
Primer 5	5'- GGG GAG TGT AGG TTT -3'
Primer 6	5'- TAT GGA AAT TTT TTA GGA GG -3'
Primer 7	5'- TGG GGT TTG GGA G -3'
Primer 8	5'- GGG TTG GTT TAG ATT TAG -3'

Gel-Shift, Shift-Western Blot, and Western Blot

The gel-shift method employed was adopted from Harrington et al⁹. A 28 bp long, fluorescein-labeled dsDNA probe containing a Sp1 recognition site was incubated for 15 min at RT with SW48 nuclear extract in the presence of increasing concentrations of an unlabeled 28mer oligomer. The competitor had either a methylated or an unmethylated Sp1 recognition site. The mixture was then run in a 6% native polyacrylamide gel and fluorescence signals were subsequently detected using a biomolecular imager (Typhoon 9400, Amersham Biosciences). Fluorescence intensity of shifted bands was normalized against the sum of signals provided by both unbound oligomers and shifted bands. Experiments were performed in quadruplicates.

Bands from the polyacrylamide gel were transferred onto a nitrocellulose membrane using the wet blotting method and Sp1 was detected via chemiluminescence with anti-

human Sp1 primary antibody (polyclonal rabbit, Millipore) and a HRP-labeled anti-rabbit secondary antibody (GE healthcare).

Sequences of oligomers:

Positive control: Forward: 5'-ACG TTG CAG CCG GGG CGG GGC TTC TGC A-3',

Reverse: 5'-TGC AGA AGC CCC GCC CCG GCT GCA ACG T-3'

Methylated competitor: Forward: 5'-CCT CAC CCC CGC CTC CAC CCC TTC GCG G-3',

Reverse: 5'-F-CCG CGA AGG GCT GGA GGC^m GGG GGT GAG G-3'

Unmethylated competitor: Forward: 5'-CCT CAC CCC CGC CTC CAC CCC TTC GCG

G-3', Reverse: 5'-F-CCG CGA AGG GCT GGA GGC GGG GGT GAG G-3'

Supplemental References

1. Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
2. Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., *et al.* Ensembl 2009. *Nucleic Acids Res* **37**, D690-D697 (2009).
3. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217 (2000).
4. Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Di Maria, M. V., *et al.* Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* **21**, 3798-3807 (2003).
5. Bewick, V., Cheek, L. & Ball, J. Statistics review 8: Qualitative data - tests of association. *Crit Care* **8**, 46-53 (2004).
6. Bewick, V., Cheek, L. & Ball, J. Statistics review 12: survival analysis. *Crit Care* **8**, 389-394 (2004).
7. Hess, D., Keusch, J. J., Oberstein, S. A., Hennekam, R. C. & Hofsteenge, J. Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective Omicron-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* **283**, 7354-7360 (2008).
8. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

9. Harrington, M. A., Jones, P. A., Imagawa, M. & Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc Natl Acad Sci U S A* **85**, 2066-2070 (1988).

6.3 HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma.

My contribution to this work:

- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;
- Manuscript writing;

HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma

Salvatore Piscuoglio, Inti Zlobec, Pierlorenzo Pallante,¹ Romina Sepe,¹ Francesco Esposito,¹ Arthur Zimmermann,² Ioannis Diamantis,³ Luigi Terracciano, Alfredo Fusco¹ & Eva Karamitopoulou²

Institute of Pathology, University of Basel, Basel, Switzerland, ¹Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli 'Federico II', Naples, Italy, ²Institute of Pathology, University of Bern, Bern, and ³Department of Internal Medicine and Gastroenterology, Zofingen Hospital, AG, Switzerland

Date of submission 10 November 2010
Accepted for publication 9 February 2011

Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A & Karamitopoulou E
(2012) *Histopathology* 60, 397–404

HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma

Aims: Pancreatic ductal adenocarcinoma follows a multistep model of progression through precursor lesions called pancreatic intraepithelial neoplasia (PanIN). The high mobility group A1 (HMGA1) and high mobility group A2 (HMGA2) proteins are architectural transcription factors that have been implicated in the pathogenesis and progression of malignant tumours, including pancreatic cancer. The aim of this study was to explore the role of HMGA1 and HMGA2 in pancreatic carcinogenesis.

Methods and results: HMGA1 and HMGA2 expression was examined in 210 ductal pancreatic adenocarcinomas from resection specimens, combined on a tissue microarray also including 40 examples of PanIN and

40 normal controls. The results were correlated with the clinicopathological parameters of the tumours and the outcome of the patients. The percentage of tumour cells showing HMGA1 and HMGA2 nuclear immunoreactivity correlated positively with increasing malignancy grade and lymph node metastasis. Moreover, HMGA1 and HMGA2 expression was significantly higher in invasive carcinomas than in PanINs. No, or very low, expression was found in normal pancreatic tissue.

Conclusions: Our results suggest that HMGA1 and HMGA2 are implicated in pancreatic carcinogenesis and may play a role in tumour progression towards a more malignant phenotype.

Keywords: HMGA1, HMGA2, immunoreactivity, pancreatic ductal adenocarcinoma, pancreatic intraepithelial neoplasia

Abbreviations: CI, confidence interval; HMGA, high mobility group A; NF- κ B, nuclear factor- κ B; PanIN, pancreatic intraepithelial neoplasia; ROC, receiver operating characteristic; SD, standard deviation; TMA, tissue microarray

Address for correspondence: A Fusco, Istituto di Endocrinologia ed Oncologia Sperimentale, del Consiglio Nazionale delle Ricerche, via Pansini 5, 80131 Naples, Italy. e-mail: alfusco@unina.it
L Terracciano, Institute of Pathology, University of Basel, Schönbeinstrasse 40, 4031 Basel. e-mail: lterracciano@uhbs.ch

© 2012 Blackwell Publishing Limited.

Introduction

Pancreatic ductal adenocarcinoma is a common cause of death from cancer, and has a dismal prognosis with currently no effective treatment.¹ Despite important advances in our understanding of the molecular biology of the early stages of neoplastic development, late molecular events that lead to tumour progression are largely unknown. Clinicopathological parameters such as tumour size, lymph node metastases and evidence of blood vessel or lymphatic invasion have been proven to be reliable prognostic determinants in pancreatic cancer.¹ The identification of reliable and reproducible biomarkers would enable better stratification of patients, and eventually provide a guide for individualized therapy. Pancreatic cancer follows a multistep model of progression through non-invasive precursor lesions. Pancreatic intraductal lesions have been classified into four groups of pancreatic intraepithelial neoplasias (PanINs): PanIN-1A, PanIN-1B, PanIN-2, and PanIN-3.² PanIN-3 shows severe epithelial dysplasia, and is most likely to progress to invasive carcinoma.²

The high mobility group A (HMGA) genes encode a family of non-histone chromatin-binding proteins, named for their rapid electrophoretic mobility in polyacrylamide gels.³ HMGA1a and HMGA1b isoforms result from alternative splicing of *HMGA1* mRNA, whereas HMGA2 is encoded by the related gene *HMGA2*.³ HMGA proteins bind the minor groove of AT-rich DNA sequences. Their DNA-binding domain is located in the N-terminal region of the protein, and contains three short basic repeats, the so-called AT-hooks.⁴ Once bound to DNA, the HMGA proteins alter chromatin structure and thereby regulate the transcriptional activity of several genes.⁵ HMGA proteins are normally expressed at high levels during embryonic development, and at very low levels in adult, differentiated tissues.⁶ HMGA proteins participate, as transcriptional regulators, in many cellular functions, including regulation of the cell cycle, cell differentiation, senescence, and neoplastic transformation.⁷ Both *HMGA1* and *HMGA2* have been reported to function as oncogenes and to be overexpressed in almost all human malignancies so far analysed, including ductal pancreatic adenocarcinoma.^{8–20} Moreover, HMGA protein overexpression has been regarded as a poor prognostic feature, as it has often been found to correlate with the presence of metastasis and with reduced survival.²¹

The objective of the present study was to investigate the role of HMGA1 and HMGA2 expression in pancreatic carcinogenesis and to evaluate its prognostic

significance. Using immunohistochemistry we analysed expression in different stages of pancreatic carcinogenesis, including invasive adenocarcinomas, PanINs, and normal pancreatic tissue, in a tissue microarray (TMA) combining 210 ductal adenocarcinomas of the pancreas from resection specimens, 40 cases of PanIN-3, and 40 normal controls.

Materials and methods

PATIENTS AND SPECIMENS

Formalin-fixed and paraffin-embedded tumours and control specimens were retrieved from the archives of the Institute of Pathology, University of Bern. All tumours and controls were reviewed by an experienced pathologist (E.K.). Histological subtypes other than ductal carcinoma were excluded. Tumours were re-staged according to the American Joint Committee on Cancer Staging Manual (seventh edition). Representative tumour areas were selected for the construction of the TMA. The TMA consisted of 210 surgically-resected ductal adenocarcinomas of the pancreas, and included 40 examples of PanIN-3 and 40 normal controls (normal pancreatic tissue and PanINs were selected from areas distant from the carcinomas). The 210 patients comprised 110 males and 100 females, with a mean age of 66.5 years (range: 20–92 years). The study was approved by the ethics committee of the University of Bern.

ASSESSMENT OF BEHAVIOUR

Medical charts were available for 77 of the 210 patients. Of these 77 patients, 60 (78%) died from the disease, and 7 (9%) were alive with recurrent/metastatic disease. The other 10 patients (13%) were alive without disease. The median follow-up time was 16 months. The clinicopathological features of these cases with survival information are given in Table 1.

CONSTRUCTION OF THE TMA

One core tissue biopsy with a diameter of 0.6 mm was taken from a representative region of individual paraffin-embedded pancreatic carcinomas (donor blocks), and placed into a new recipient paraffin block with a semi-automated tissue-arraying device. The presence of tumour tissue on the TMA was verified on a haematoxylin and eosin-stained slide. Two to three tissue cores of each tumour were available for biomarker analysis. Five-micrometre sections were cut with an adhesive-coated slide system (Instrumedics, Hackensack, NJ,

Table 1. Clinicopathological characteristics of cases with survival information (*N* = 77)

Clinicopathological features	Frequency, <i>N</i> (%)
Diagnosis	
Ductal carcinoma	77 (100.0)
Sex	
Female	33 (42.9)
Male	44 (57.1)
Tumour grade	
G1	16 (20.8)
G2	42 (54.6)
G3	19 (24.7)
pT stage	
pT1	3 (4.1)
pT2	12 (16.2)
pT3	52 (70.3)
pT4	7 (9.5)
pN stage	
pN0	27 (38.0)
pN1	44 (62.0)
Metastasis	
Absent	72 (93.5)
Present	5 (6.5)
Tumour diameter (mm), mean ± SD	31.4 ± 14.0
Survival time (months), median (range)	12.0 (0.5–48.0)

SD, Standard deviation.

USA) and examined by immunohistochemistry. The number of samples differed slightly between the individual markers, because of variability in the number of interpretable specimens on TMA sections.

IMMUNOHISTOCHEMISTRY

Freshly cut sections of TMA blocks were used for immunohistochemical staining with anti-HMGA1 and anti-HMGA2 antibodies. Briefly, punches were dewaxed and rehydrated in distilled water. Endogenous peroxidase activity was blocked with 0.5% H₂O₂. The sections were incubated with 10% normal goat serum (Dako Cytomation, Carpinteria, CA, USA) for 20 min,

and then with the primary antibody at room temperature. Optimal staining was achieved after pretreatment in a microwave oven (98°C for 30 min, pH 6, dilution 1:250). Subsequently, sections were incubated with peroxidase-labelled secondary antibody (Dako Cytomation) for 30 min at room temperature. 3,3'-Diaminobenzidine was used as chromogen. Sections were then counterstained with Gill's haematoxylin. As a positive control, a TMA with various normal tissue samples was stained in parallel.

The antibodies used for HMGA1 immunostaining were raised against the synthetic peptide SSSKQQPL-ASKQ, which is specific for HMGA1. They were affinity purified against the synthetic peptide.⁹ For HMGA2 immunohistochemistry, antibodies raised against a synthetic peptide located in the N-terminal region were used.²²

The specificity of immunolabelling was validated by the absence of tumour staining when using antibodies preincubated with the peptide against which the antibodies were raised (data not shown). Similarly, no positivity was observed when tumour samples were incubated with a preimmune serum (data not shown).

IMMUNOHISTOCHEMICAL EVALUATION

Nuclear HMGA1 and HMGA2 staining was scored by two independent observers (S.P. and L.T.) blinded for clinical parameters. Slides were screened semiquantitatively for the percentage of positive cells and the intensity of the signal. At least 100 cells were counted for each punch. The percentage of positive cells per number of cells counted was scored in 10 groups from 0 (0–9%) to 9 (91–100%). The intensity of the signal was graded semiquantitatively into four groups from 0 (no positivity) to 3 (strong positivity). A case was considered to be positive if belonging at least to group 1 for the percentage (i.e. ≥10%), irrespective of intensity. In PanINs and normal controls, the epithelial cells of ductal structures were evaluated.

STATISTICAL METHODS

The selection of clinically important cut-off scores was based on receiver operating characteristic (ROC) curve analysis.^{23,24} At each percentage score, the sensitivity and specificity for each outcome under study were plotted, generating an ROC curve. The score having the closest distance to the point with both maximum sensitivity and specificity, i.e. point (0.0, 1.0) on the curve, was selected as the cut-off score leading to the greatest number of tumours that were correctly classified as having or not having the outcome. In order to

enable the use of ROC curve analysis, the following clinicopathological features were dichotomized: T stage (early, T1 + T2; late, T3 + T4), N stage (N0, no lymph node involvement; N1, any lymph node involvement), tumour grade (low, G1 + G2; high, G3), and survival (death from pancreatic carcinoma or alive).

Chi-square tests were used to study the relationship between HMGA1 and HMGA2 expression and histological subgroups. Differences in HMGA1 and HMGA2 expression between normal tissue, PanIN and carcinoma were investigated with the non-parametric Wilcoxon rank sum test. Univariate survival analysis was carried out with the Kaplan–Meier log-rank test, and multivariate analysis with Cox proportional hazards regression. Hazard ratios and 95% confidence intervals (CIs) were used to determine the effect of each variable on survival time. In addition, logistic regression was performed in univariate and multivariate settings to determine the associations of protein expression and its independent effect on binary outcomes. The odds ratios and 95% CIs were evaluated. A Bonferroni correction for multiple comparisons was performed. A *P*-value ≤ 0.01 (two-sided) was required for the association to be statistically significant. All analyses were carried out with SAS (V9; SAS Institute, Cary, NC, USA).

Results

ANALYSIS OF HMGA1 AND HMGA2 EXPRESSION BY IMMUNOHISTOCHEMISTRY

Table 2 shows the differences in protein expression between normal pancreas, PanIN, and cancer, and Table 3 the correlation of HMGA1 and HMGA2

expression with pT stage, pN stage, and tumour grade. Survival related to protein marker expression is analysed in Table 4. Some representative images of immunohistochemical staining are illustrated in Figures 1 and 2.

PANCREATIC CARCINOMAS VERSUS NORMAL CONTROLS

Pancreatic carcinoma cases generally displayed strong nuclear HMGA1 and HMGA2 immunoreactivity, whereas absent or very low immunoreactivity was observed in normal pancreatic tissue. The mean \pm standard deviation (SD) for the percentage of cells showing HMGA1 and HMGA2 protein expression were found to be 0 ± 0 and 0.2 ± 0.9 , respectively, in normal tissue, as compared with 26.6 ± 30.5 and 16.3 ± 28.4 , respectively, in carcinomas (*P* < 0.001; Table 2).

PANCREATIC CARCINOMA VERSUS PANIN

Percentages of cells showing HMGA1 and HMGA2 protein expression were significantly higher in ductal pancreatic adenocarcinomas (mean \pm SD: 26.6 ± 30.5 and 16.3 ± 28.4 , respectively) than in PanIN cases (11.1 ± 15.0 and 2.7 ± 13.5 , respectively) (*P* < 0.001; Table 2).

PANIN VERSUS NORMAL CONTROLS

Means for the percentages of cells showing HMGA1 and HMGA2 protein expression were significantly higher in PanIN cases (11.1 ± 15.0 and 2.7 ± 13.5 , respectively) than in normal tissue (0 ± 0 and 0.2 ± 0.9 , respectively) (*P* < 0.001; Table 2).

Table 2. Differences in marker expression between normal pancreas, pancreatic intraepithelial neoplasia (PanIN) and cancer

	Normal	PanIN	Cancer	<i>P</i> -value
HMGA1				
<i>n</i>	31	31	183	<0.001
Mean \pm SD	0 ± 0	11.1 ± 15.0	26.6 ± 30.5	
Median (min–max)	0 (0–0)	5.0 (0–50)	10.0 (0–100)	
HMGA2				
<i>n</i>	29	37	191	<0.001
Mean \pm SD	0.2 ± 0.9	2.7 ± 13.5	16.3 ± 28.4	
Median (min–max)	0 (0–5)	0 (0–80)	0 (0–100)	

HMGA, High mobility group A; SD, standard deviation.

Expression is given as percentage of immunoreactive cells. Wilcoxon rank sum test.

Table 3. Protein marker expression related to stage (pT, pN), and tumour grade

	HMGA1	HMGA2
pT stage		
pT1–2	25.7 ± 28.7; 15.0	16.0 ± 28.4; 0.0
pT3–4	27.1 ± 31.3; 10.0	16.9 ± 28.9; 0.0
<i>P</i> -value	0.83	0.952
pN stage		
pN0	18.3 ± 22.8; 5.0	8.4 ± 19.0; 0.0
pN1	32.7 ± 33.6; 20.0	21.5 ± 32.4; 0.0
<i>P</i> -value	0.012	0.039
Tumour grade		
G1–2	22.2 ± 27.3; 10.0	12.0 ± 23.3; 0.0
G3	37.3 ± 35.1; 30.0	26.8 ± 36.0; 5.0
<i>P</i> -value	0.009	0.008

HMGA, High mobility group A. Expression is given as the percentage of immunoreactive tumour cells. Mean ± standard deviation; median values. Wilcoxon rank sum test.

Table 4. Survival analysis related to protein markers using cut-off scores [median values, namely 10% for HMGA1 (nuclear), and 0% for HMGA2 (nuclear)]*

	Total no. of patients	No. of deaths	Median survival time (months) (95% CI)	<i>P</i> -value
HMGA1				
Negative	40	30	14 (10–17)	0.816
Positive	24	20	12.5 (10–22)	
HMGA2				
Negative	37	27	15 (10–24)	0.196
Positive	26	21	14 (12–18)	

CI, Confidence interval; HMGA, high mobility group A. Log-rank test. *Similar results were obtained when expression was analysed as a continuous variable by Cox regression analysis.

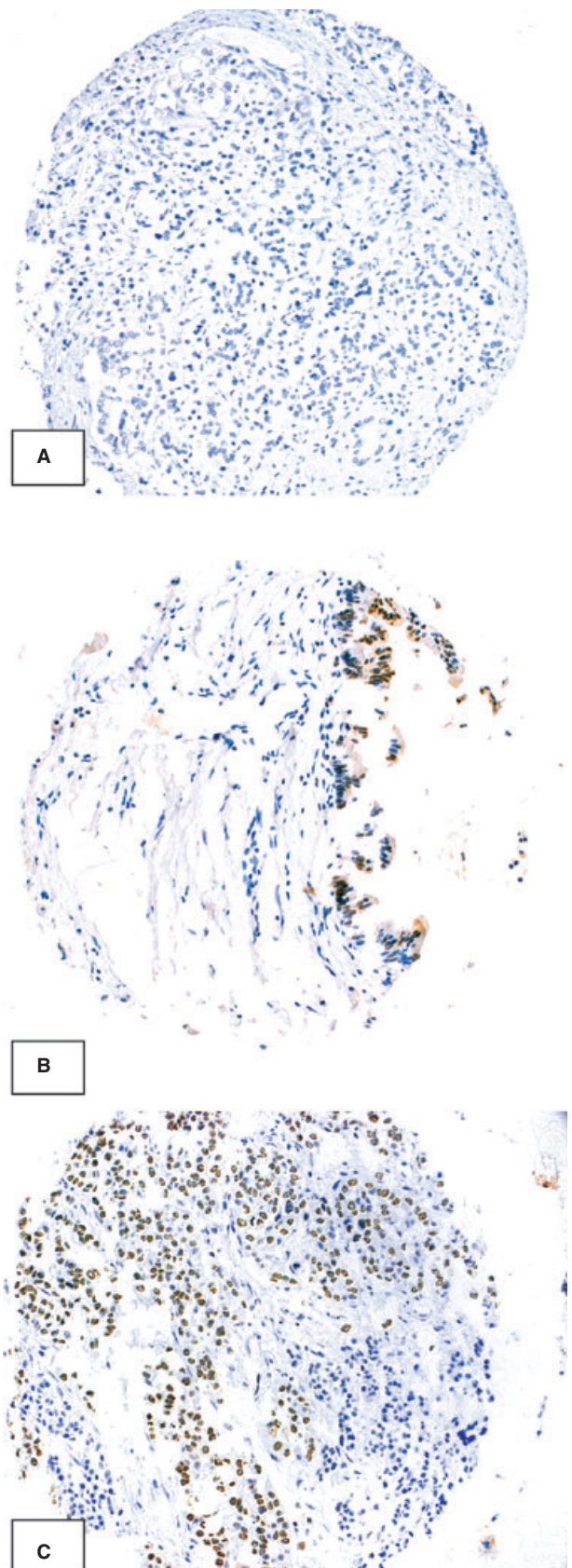


Figure 1. Examples of nuclear immunohistochemical detection of high mobility group A1 (HMGA1). Absent expression in normal pancreatic tissue (A) and moderate expression in pancreatic intra-epithelial neoplasia (PanIN) (B), compared with strong, diffuse expression in pancreatic carcinoma (C).

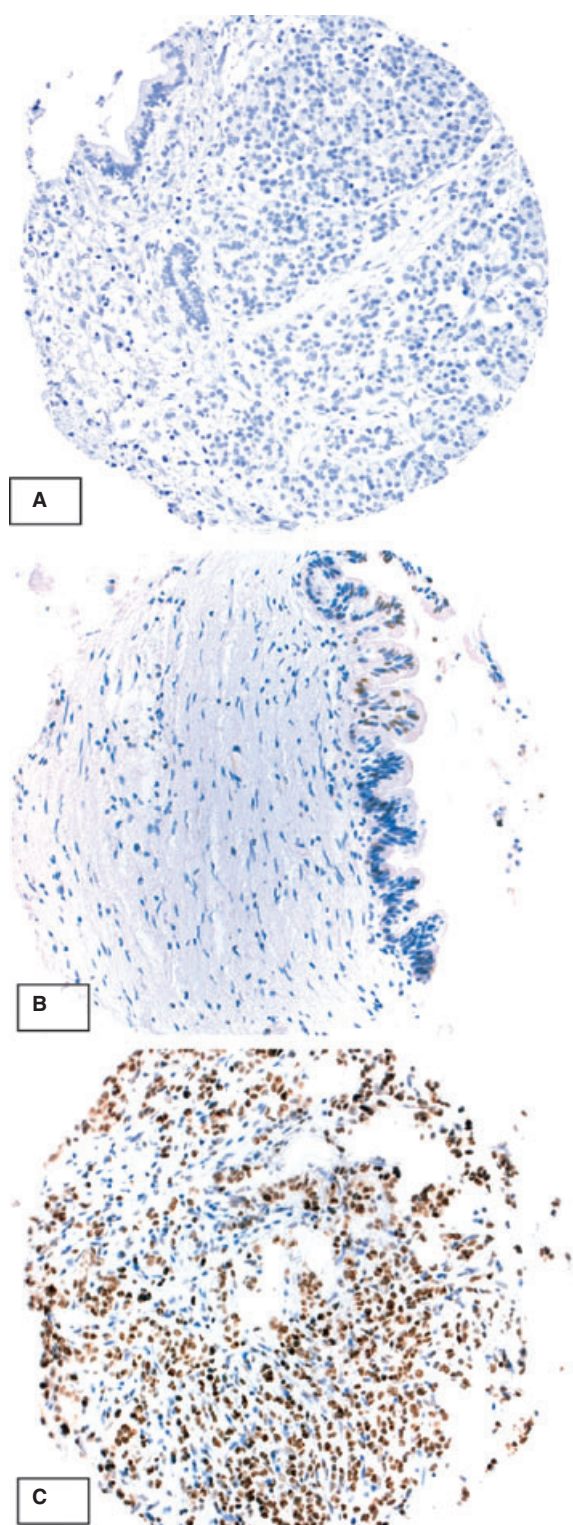


Figure 2. Examples of nuclear immunohistochemical detection of high mobility group A2 (HMGA2). A, Normal pancreatic tissue without HMGA2 expression. B, Pancreatic intraepithelial neoplasia (PanIN) with moderate HMGA2 expression. C, An example of ductal adenocarcinoma with strong nuclear HMGA2 expression.

PROTEIN EXPRESSION AND TUMOUR GRADING

Extent of nuclear HMGA1 and HMGA2 expression showed a positive correlation with higher tumour grade. Cells within poorly differentiated tumours (grade 3) more frequently expressed HMGA1 and HMGA2 (mean percentage expression \pm SD: 37.3 ± 35.1 and 26.8 ± 36.0 , respectively) than those in better-differentiated (grade 1 and 2) tumours (22.2 ± 27.3 and 12.0 ± 23.3 , respectively) ($P = 0.009$ and $P = 0.008$, respectively; Table 3).

PROTEIN EXPRESSION AND TNM CLASSIFICATION OF THE TUMOURS

HMGA1 and HMGA2 expression showed a significant association with the pN stage of the tumours. Mean \pm SD percentage expression levels for HMGA1 and HMGA2 were 18.3 ± 22.8 and 8.4 ± 19.0 , respectively, for nodal negative (pN0) carcinomas, as compared with 32.7 ± 33.6 and 21.5 ± 32.4 , respectively, for nodal positive (pN1) carcinomas ($P = 0.012$ and $P = 0.039$, respectively; Table 3). No association was noted between protein expression and pT stage of the tumours ($P = 0.83$ and $P = 0.952$, respectively; Table 3).

PROGNOSTIC SIGNIFICANCE

Median survival times were 12.5 and 14 months for HMGA1-positive and HMGA2-positive tumours, respectively, as compared with 14 and 15 months for HMGA1-negative and HMGA2-negative tumours, respectively. These differences were not statistically significant ($P = 0.816$ and $P = 0.196$, respectively; Table 4).

Discussion

Although most patients with pancreatic cancer present with advanced disease, the molecular events involved in tumour progression, invasion and metastasis are poorly understood.

In the present study, we investigated the immunohistochemical expression of HMGA1 and HMGA2 in 210 cases of ductal adenocarcinoma of the pancreas combined on a TMA including in addition 40 examples of PanIN-3 and 40 normal controls.

A major finding was the increasing mean protein expression of HMGA1 and HMGA2 between normal pancreatic tissue, PanIN cases, and invasive adenocarcinoma (Table 2). Mean HMGA1 and HMGA2 expression appeared to progressively increase through the

transition from normal tissue to pancreatic cancer. Moreover, HMGA1 and HMGA2 expression showed a positive correlation with malignancy grade and neoplastic progression, becoming higher with the dedifferentiation of the neoplasms and with the presence of lymph node metastasis. Therefore, our data suggest that HMGA1 and HMGA2 expression correlates with a more aggressive phenotype in pancreatic adenocarcinoma. These findings are in keeping with the recent studies of Hristov *et al.*,^{19,20} who also noted an association between HMGA1 and HMGA2 expression and a more malignant phenotype in pancreatic cancer. HMGA1 was found to correlate with advanced tumour grade and decreased survival of their patients, whereas HMGA2 correlated with increasing tumour grade and lymph node metastasis. However, in our study, both HMGA1 and HMGA2 showed a positive correlation with lymph node metastasis. Although we found that patients with HMGA1 and HMGA2 negative tumours tended to survive longer, the association with patient outcome was not statistically significant. This may be because of the short survival time of most patients with pancreatic cancer. In addition, Hristov *et al.*¹⁹ reported that HMGA1 expression correlated with a more advanced PanIN grade, whereas possible differences in protein expression between PanIN and adenocarcinoma were not discussed. In our study, only PanIN-3 lesions were included. Moreover, we demonstrated that mean protein expression was significantly higher in pancreatic ductal adenocarcinomas than in PanIN lesions.

A number of other groups have reported involvement of HMGA genes and proteins in the pathogenesis of pancreatic cancer. In one study, HMGA1 was found to be overexpressed in a small number of pancreatic adenocarcinomas and metastatic lesions, but without association with tumour grade.²⁵ More recently, Liao *et al.*²⁶ reported HMGA1 protein positivity in a high proportion of ductal pancreatic adenocarcinomas, also without correlation with tumour differentiation. However, in this study, in contrast to ours, only staining intensity was analysed, and the number of carcinomas was smaller. HMGA1 expression was also found to be increased in other pancreatic tumours, such as intraductal papillary mucinous neoplasms.²⁶ Studies in pancreatic ductal adenocarcinoma cell lines have shown that *HMGA1* knockdown decreases cellular invasion, anchorage-independent cell growth, and resistance to chemotherapeutic agents.^{26,27} In addition, high-level expression of HMGA1 has been reported in almost all neoplastic tissues, including colon, breast, lung, ovarian, uterine, prostatic, gastric and head and neck carcinomas.^{8–15} *HMGA2* has also

been implicated in the development and progression of human malignancies, including lung adenocarcinomas, breast cancer, and squamous cell carcinomas of the oral cavity.^{16–18} Additionally, *HMGA2* has been reported to play a role in the epithelial–mesenchymal transition that takes place during invasion and metastasis.²⁸ A previous study found increased *HMGA2* expression by reverse transcription polymerase chain reaction in ductal pancreatic adenocarcinomas, and *HMGA2* expression by immunohistochemistry.²⁹ However, associations with grade and outcome were not included in the analysis, and *HMGA2* expression was also found in pancreatic islet cells and, focally, in non-neoplastic ductal epithelial cells. In our study, focal *HMGA2* immunoreactivity was also observed in a very small number of non-malignant ductal epithelial cells.

Regarding the processes underlying the involvement of HMGA genes in neoplastic transformation, it has been hypothesized that this probably occurs through oncofetal transcriptional mechanisms that have not yet been characterized.²¹ It has been suggested that the elevated expression of HMGA1 in tumour cells requires a complex cooperation between SP1 family members and AP1 factors, induced by the activation of Ras GTPase signalling.³⁰ Moreover, the main function of the HMGA proteins, the regulation of gene transcription, is probably based on the ability of HMGA proteins to down-regulate or up-regulate the expression of genes that have a crucial role in the control of cell proliferation and invasion.²¹ In particular, emerging evidence suggests that HMGA1 modulates gene expression, including pathways involved in inflammation, proliferation, transformation, metastatic progression, angiogenesis, and DNA repair. Most transcriptional targets include regulatory elements of nuclear factor- κ B (NF- κ B), a mediator of inflammatory pathways, suggesting that HMGA1 and NF- κ B may cooperate to induce inflammatory signals and drive transformation.³¹

One of the advantages of this study is the use of TMAs, which have provided us with an efficient and cost-effective way of testing a large number of tumour specimens. Concerns could be raised about the TMA technique with regard to the possible limitations in sampling large, heterogeneous tumours. However, previous studies have shown comparable results between whole tissue sections and TMA cores, and have been able to reproduce numerous clinicopathological associations previously found with whole tissue sections.³²

In conclusion, we found increasing mean HMGA1 and HMGA2 expression during neoplastic progression in ductal pancreatic adenocarcinoma, accompanied by

a positive correlation of protein expression with both increasing malignancy grade and the presence of lymph node metastasis. Our results support the idea that HMGA1 and HMGA2 may play a significant role in the late stages of pancreatic carcinogenesis and in the progression towards a more aggressive tumour phenotype.

References

- Hidalgo M. Pancreatic cancer. *N. Engl. J. Med.* 2010; **362**: 1605–1617.
- Hruban RH, Adsay NV, Albores-Saavedra J et al. Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. *Am. J. Surg. Pathol.* 2001; **25**: 579–586.
- Johnson KR, Cook SA, Davison MT. Chromosomal localization of the murine gene and two related sequences encoding high-mobility-group I and Y proteins. *Genomics* 1992; **12**: 503–509.
- Reeves R, Nissen MS. The AT DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.* 1990; **265**: 8573–8582.
- Thanos D, Maniatis T. The high mobility group protein HMG I(Y) is required for NF- κ B-dependent virus induction of the human IFN- β gene. *Cell* 1992; **27**: 777–789.
- Chiappetta G, Avantaggiato V, Visconti R et al. High level expression of the HMGI(Y) gene during embryonic development. *Oncogene* 1996; **13**: 2439–2446.
- Wood LJ, Maher JF, Bunton TE, Resar LM. The oncogenic properties of the HMG-1 gene family. *Cancer Res.* 2000; **60**: 4256–4261.
- Fedele M, Bandiera A, Chiappetta G et al. Human colorectal carcinomas express high levels of high mobility group HMGI(Y) proteins. *Cancer Res.* 1996; **56**: 1896–1901.
- Chiappetta G, Botti G, Monaco M et al. HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. *Clin. Cancer Res.* 2004; **10**: 7637–7644.
- Sarhadi VK, Wikman H, Salmenkivi K et al. Increased expression of high mobility group A proteins in lung cancer. *J. Pathol.* 2006; **209**: 206–212.
- Masciullo V, Baldassare G, Pentimalli F et al. HMGA1 protein over-expression is a frequent feature of epithelial ovarian carcinomas. *Carcinogenesis* 2003; **24**: 1191–1198.
- Bandiera A, Bonifacio D, Manfioletti G et al. Expression of high mobility group I (HMGI) proteins in squamous intraepithelial lesions (SILs) of uterine cervix. *Cancer Res.* 1998; **58**: 426–431.
- Tamimi Y, van der Poel HG, Denyn MM et al. Increased expression of high mobility group protein I (Y) in high grade prostate cancer determined by in situ hybridization. *Cancer Res.* 1993; **53**: 5512–5516.
- Nam ES, Kim DH, Cho SJ et al. Expression of HMGI(Y) associated with malignant phenotype of human gastric tissue. *Histopathology* 2003; **42**: 466–471.
- Rho YS, Lim YC, Park IS et al. High mobility group HMGI(Y) protein expression in head and neck squamous cell carcinoma. *Acta Otolaryngol.* 2007; **127**: 76–81.
- Meyer B, Loeschke S, Schultze A et al. HMGA2 overexpression in non-small cell lung cancer. *Mol. Carcinog.* 2007; **46**: 503–511.
- Rogalla P, Drechsler K, Kazmierczak B, Rippe V, Bonk U, Bullerdiek J. Expression of HMGI-C, a member of the high mobility group protein family, in a subset of breast cancers: relationship to histologic grade. *Mol. Carcinog.* 1997; **19**: 153–156.
- Miyazawa J, Mitoro A, Kawashiri S, Chada KK, Imai K. Expression of mesenchyme-specific gene HMGA2 in squamous cell carcinomas of the oral cavity. *Cancer Res.* 2004; **64**: 2024–2029.
- Hristov AC, Cope L, Delos Reyes M et al. HMGA2 protein expression correlates with lymph node metastasis and increased tumor grade in pancreatic ductal adenocarcinoma. *Mod. Pathol.* 2009; **22**: 43–49.
- Hristov AC, Cope L, Di Cello F et al. HMGA1 correlates with advanced tumor grade and decreased survival in pancreatic ductal adenocarcinoma. *Mod. Pathol.* 2010; **23**: 98–104.
- Fusco A, Fedele M. Roles of HMGA proteins in cancer. *Nat. Rev. Cancer* 2007; **7**: 899–910.
- Fedele M, Visone R, De Martino I et al. HMGA2 induces pituitary tumorigenesis by enhancing E2F1 activity. *Cancer Cell* 2006; **9**: 459–471.
- Hanley J. Receiver operating characteristic (ROC) methodology: the state of the art. *Crit. Rev. Diagn. Imaging* 1989; **29**: 307–337.
- Zlobec I, Steele R, Terracciano L, Jass JR, Lugli A. Selecting immunohistochemical cut-off scores for novel biomarkers of progression and survival in colorectal cancer. *J. Clin. Pathol.* 2007; **60**: 1112–1116.
- Abe N, Watanabe T, Masaki T et al. Pancreatic duct cell carcinomas express high levels of high mobility group I(Y) proteins. *Cancer Res.* 2000; **60**: 3117–3122.
- Liau SS, Rocha F, Matros E, Redston M, Whang E. High mobility group AT-hook 1 (HMGA1) is an independent prognostic factor and novel therapeutic target in pancreatic adenocarcinoma. *Cancer* 2008; **113**: 302–314.
- Liau SS, Wang EE. HMGA1 is a molecular determinant of chemoresistance to gemcitabine in pancreatic adenocarcinoma. *Clin. Cancer Res.* 2008; **14**: 1470–1477.
- Thuault S, Valcourt U, Petersen M, Manfioletti G, Heldin CH, Moustakas A. Transforming growth factor-beta employs HMGA2 to elicit epithelial-mesenchymal transition. *J. Cell Biol.* 2006; **174**: 175–183.
- Abe N, Watanabe T, Suzuki Y et al. An increased high-mobility group A2 expression level is associated with malignant phenotype in pancreatic exocrine tissue. *Br. J. Cancer* 2003; **89**: 2104–2109.
- Cleynen I, Huysmans C, Sasazuki T, Shirasawa S, Van de Ven W, Peeters K. Transcriptional control of the human high mobility group A1 gene: basal and oncogenic Ras-regulated expression. *Cancer Res.* 2007; **15**: 4620–4629.
- Resar LM. The high mobility group A1 gene: transforming inflammatory signals into cancer? *Cancer Res.* 2010; **70**: 436–439.
- Kallionemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum. Mol. Genet.* 2001; **10**: 657–666.

6.3.1 HMGA1 over-expression represents a poor prognostic index in human breast carcinoma.

My contribution to this work:

- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;
- Manuscript writing;

HMGA1 over-expression represents a poor prognostic index in human breast carcinomas

Pierlorenzo Pallante^{1*}, Salvatore Piscuoglio^{2,3*}, Inti Zlobec², Luigi Terracciano² and Alfredo Fusco¹

¹Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Biologia e Patologia Cellulare e Molecolare, Facoltà di Medicina e Chirurgia di Napoli, Università degli Studi di Napoli "Federico II", via Pansini 5, 80131 Naples, Italy.

²Institute of Pathology, Molecular Pathology Division, University of Basel, Schonbeinstrasse 40, 4003 Basel, Switzerland.

³Research Group Human Genetics, Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland

*equally contributed

Running title: HMGA1 in breast carcinomas.

Keywords: HMGA1, breast ductal carcinomas, immunohistochemistry.

Alfredo Fusco,
Istituto di Endocrinologia ed Oncologia Sperimentale del CNR,
via Pansini 5, 80131 Napoli, Italy.
Tel: 39-081-7463602 or 7463749;
Fax: 39-081-2296674.
E-mail: afusco@napoli.com or afusco@unina.it

and

Luigi Terracciano,
Department of Pathology, Molecular Pathology Division,
University of Basel, Schonbeinstrasse 40, 4003 Basel, Switzerland.
Tel: 41-61-26522525 or 2652849;
Fax: 41-61-2653194.
E-mail: lterracciano@uhbs.ch

ABSTRACT

Aim: Breast cancer represents the second leading cause of cancer mortality among women and accounts for more than 40,000 deaths annually. HMGA1 expression has been implicated in the pathogenesis and progression of human malignant tumours, including breast carcinomas. The aim of this study was to evaluate HMGA detection as a possible prognostic index in breast carcinoma by analyzing a large number of breast carcinoma samples.

Methods: HMGA1 expression has been analyzed by immunohistochemistry in a large series of breast carcinoma resections (n=1202) combined on a Tissue Microarray (TMA) mainly including the ductal carcinoma variant. Then, the results were correlated with clinic-pathological parameters and outcome of the patients.

Results: HMGA1 over-expression was found in the large majority of breast carcinoma samples, and its over-expression positively correlates with Her2/neu amplification and progesterone receptor status, while a negative correlation was found with estrogen receptor status. Conversely, no HMGA1 expression was found in normal breast tissues.

Conclusions: The data reported here indicate that the level of HMGA1 expression is related to an unfavourable breast cancer phenotype and poor prognosis, as supported by its strong association with the Her2/neu, PR and ER status that could explain, at least in part, the different behaviour of the human breast carcinoma over-expressing HMGA1.

INTRODUCTION

Breast cancer represents the second leading cause of mortality caused by cancer in women (Ellis et al. 2003). Neoplastic breast diseases comprise benign form, like fibroadenoma, and very aggressive forms, like undifferentiated breast carcinoma. It has been reported that a large series of molecules and patterns (growth factors and their receptors, signal transduction molecules, cell cycle regulators) are altered and deregulated in sporadic breast cancers (Vogelstein and Kinzler 1994).

Nowadays, a series of genetic markers are evaluated to assess the prognosis of breast cancer patients: the BRCA mutational status of patients, the expression of estrogen receptor (ER), progesterone receptor (PR), and the Her2/neu receptor (Her2) (Deroo et al. 2006, Gao et al. 2002, Hynes et al. 1994, Miecznikowski et al. 2010). More recently, TGF-beta has also been considered as a potential prognostic marker for breast cancer patients (Koumoundourou et al. 2007). Moreover, the genetic status has a critical role in assigning the treatment. Indeed, tamoxifen (anti-estrogen agent) and trastuzumab (monoclonal antibody) are the elected chemotherapeutic agents for treating estrogen receptor (ER)-positive breast tumors and human Her2-over-expressing tumors, respectively (Carter et al. 1992, Arteaga et al. 2003). However, these markers are insufficient to predict the prognosis and to indicate the appropriate therapy, and many patients remain over- or under-treated (Cianfrocca and Gradishar, 2009).

HMGA1 protein belongs to the high-mobility group A (HMGA) family that consists of three members: HMGA1a, HMGA1b and HMGA2. Two distinct genes, HMGA1a and HMGA1b generates these three proteins by alternative splicing (Johnson et al. 1989). These proteins are able to bind AT-rich DNA sequences, but do not have transcriptional activity *per se*. However they can alter chromatin structure, therefore modulating the transcriptional activation of genes (Thanos and Maniatis 1992, Grosschedl et al. 1994).

HMGA over-expression is a feature of malignant tumours. Both HMGA genes are widely expressed during embryogenesis and in neoplastic tissues (including pancreas, thyroid, colon, breast, lung, ovary, uterine cervix, prostate, gastric carcinomas, squamous carcinomas of the oral cavity, head and neck tumours), whereas their expression is absent or very low in adult tissues. Their over-

expression represents a poor prognostic index and often correlates with metastases and reduced survival (Fusco and Fedele, 2007). Their oncogenic role has been extensively reported (Wood et al. 2000, Reeves et al. 2001, Berlingieri et al. 2002).

Previous studies of our group performed on a limited number of breast carcinoma samples, of which the clinic-pathological data were available only in a small number of cases, did not show any association between HMGA1 expression and histological grading. Conversely, we found that HMGA1 expression tended to be associated with c-erbB-2 expression (Spearman rho: 0.36; p=0.065), but not with the expression of the receptors for estrogens and progesterone. These results appeared in contrast with previous results, showing that enforced expression of HMGA1 in breast carcinoma cells induced the ability to form primary and metastatic tumours in athymic mice, and that HMGA1 is able to bind to BRCA1 promoter down-regulating its expression (Baldassarre et al. 2003).

The aim of the present work was, therefore, to analyse HMGA1 expression in a very large number of breast carcinoma tissues (n=1024), all of them provided with the most important clinicopathological parameters of patients, such as tumour size, lymph node status, endocrine receptors and Her2 status. Here, we confirm the HMGA1 over-expression in human breast carcinomas, with respect to the normal breast tissues. Moreover, we found that a strong HMGA1 over-expression identifies a subset of breast carcinomas characterized by Her2/neu amplification and progesterone receptor (PR) expression, but lacking estrogen receptor (ER) expression, a signature correlated with a patient poor prognosis.

MATERIALS AND METHODS

Human breast tissue samples

Neoplastic human breast tissue samples and normal controls were collected at the Department of Pathology, University of Basel Switzerland. Specimen obtained from surgical resections were formalin-fixed and paraffin embedded and stored in the archive of the institute. Several expert pathologists (SP, LT and LMT) proceeded to examine and classify tumours according to the American Joint Committee on Cancer (AJCC) staging manual (sixth edition). Tumour areas were accurately selected to be representative of the tumor specimen in the construction of the TMA. The TMA consisted of n=1388 "punches", n=1202 of which (%) were evaluable. Diverse tumoral histotype, as well as clinicopathological characteristic of breast carcinoma samples are precisely indicated in Table 1.

Construction of tissue microarray

Construction of TMA was reported elsewhere (Torhorst et al., 2001). Briefly, representative regions of paraffin-embedded donor blocks corresponding to a single tumor specimen were used to take core biopsies with 0.6 mm diameter. Subsequently, semi-automated apparatus for tissue arraying (Beecher Instruments, Silver Spring, MD) was used to transfer core biopsies into a new recipient. Sections of 5 micrometer thick were cut using an adhesive-coated slide system (Instrumedics Inc., Hackensack, NJ) and analyzed by immunohistochemistry, after verification of TMA by H&E staining.

Immunohistochemistry and immunohistochemical evaluation

TMA sections were used for immunohistochemical analysis of HMGA1 protein expression by using an antibody raised against the N-terminal region of the HMGA1 protein as described elsewhere (Chiappetta et al. 2004). Staining procedures are performed as reported elsewhere (Piscuoglio et al. 2012). Negative controls, to confirm the specificity of the reaction, were performed by omitting the first antibody and by pre-incubating of the first antibody with molar excess of the HMGA1 synthetic peptide. Standard indirect immunoperoxidase procedure (ABC Elite, Vector Laboratories, Burlingame, CA) was used to develop signals. HMGA1 staining was scored by three expert pathologists (SP, LT and LMT) blinded for the clinic-

pathological parameters. A semi-quantitative methodology was used to perform the screening for the percentage of positive cells and for the signal intensity. At least 100 cells were counted in each punch.

Statistical analysis

Statistical correlations between variables were tested using a T student test (paired and unpaired), where appropriate. All tests were two-sided. A p-value <0.05 was considered statistically significant. Analysis was performed using SAS V9.1 (SAS Institute, Cary, NC, USA).

Ethics

All the analyses of this study were performed according to the ethical standards required by the local ethic committee of Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli "Federico II", IEOS-CNR, Napoli, Italy, and Department of Pathology, University of Basel, Basel, Switzerland.

RESULTS

Analysis of HMGA1 expression by immunohistochemistry.

A tissue micro array (TMA) comprising n=1338 cases of human breast carcinomas was analyzed by immunohistochemistry for HMGA1 protein expression using specific antibodies raised against the N-terminal region of the HMGA1 protein as described elsewhere (Chiappetta et al. 2004). TMA consisted mainly of ductal carcinoma tissues (n=963), and also including several normal breast tissue samples as positive controls. The whole set of clinic-pathological features of the patients are reported in Table 1.

A total of n=1202 samples were informative for determination of the HMGA1 expression pattern that resulted always nuclear.

We found that HMGA1 staining was negative in all the normal cases analyzed, while HMGA1 expression was positive in all the breast carcinoma tissues analyzed (Figure 1). As shown in Table 1, no association was found between HMGA1 positivity and the ductal or other histotypes (p=0.883) of breast carcinomas. Indeed, HMGA1 staining, expressed as percentage (mean % \pm sd, median %) of HMGA1-positive cells, was similar among ductal breast carcinomas with different histologic grading (Table 1). No significant association was also found with the pT stage of the tumours (p=0.402, Table 1). As well, no association was observed between HMGA1 protein expression and BRE grade of patients (p=0.2, Table 1).

Conversely, an inverse correlation was found between HMGA1 over-expression and the lymph node status of patients (p=0.293, Table 1), In fact, breast carcinomas not showing lymph-node colonization (pN0) expressed higher HMGA1 levels than their counterpart >pN0. However, this result failed to reach the statistical significance (Table 1), but if we consider only pN0, pN1 and pN2, this relationship is statistically significant (Table 1).

Therefore, these findings indicate a lack of association between HMGA1 expression and the morphologic grading of ductal breast carcinomas.

HMGA1 expression correlates with Her2/neu expression and ER and PR status

Subsequently, we have evaluated the relationship between HMGA1 expression and several indicators of breast carcinoma invasion (Her2/neu amplification, ER and PR status). As reported in Table 1, HMGA1 expression was significantly associated with that of Her2/neu ($p=0.004$). Indeed, HMGA1 and Her2/neu staining are positively correlated: low levels of HMGA1 (26.5 ± 29.7) are associated with a weak Her2 staining (0+1), whereas high levels of HMGA1 (32.2 ± 30.9) are correlated with an intense Her2 staining (2+3).

As far as the correlation of HMGA1 expression with the endocrine status of the breast carcinoma samples is concerned, a significant association has been found with the retention of PR ($p=0.003$) and loss of ER expression ($p=0.007$). Indeed, expression of PR was found positively correlated with high levels of HMGA1 (29.3 ± 31.1) in breast carcinomas (Table 1), while ER expression was found negatively correlated with high levels of HMGA1 (31.1 ± 30.9 , Table 1). Therefore, in human breast carcinomas the over-expression of HMGA1 is positively and negatively correlated with PR and ER status, respectively (Table 1). Representative cases of breast carcinomas expressing HMGA1, PR and ER are reported in Figure 2 (Figure 2A, B, C, D).

DISCUSSION

The aim of our study was to verify whether HMGA1 protein might be an indicator for the diagnosis and prognosis of human breast carcinoma. Therefore, we investigated by immunohistochemistry the expression of the HMGA1 protein in n=1202 breast carcinoma tissues combined on a TMA and a compare it with multiple clinicopathological parameters.

Consistently with our previous studies showing that HMGA1 protein over-expression was found in 60% of ductal carcinomas and in almost all of the lobular carcinomas (Chiappetta et al. 2004), we found that the HMGA1 protein resulted over-expressed in the breast carcinoma specimens compared to the breast normal control tissues and, in addition, a differential HMGA1 protein expression was found between all the carcinoma samples analysed. No association was found between HMGA1 expression and histological grade of ductal carcinomas. Conversely, we have observed an inverse trend between the over-expression of HMGA1 and the pN stage (if we exclude the pN3 value).

A strong association was, instead, observed between HMGA1 over-expression and Her2/neu amplification in breast carcinoma. This result is in line with our previously published data, obtained on a shorter series of human breast carcinomas where HMGA1 over-expression correlated with the amplification of ErbB2 (Chiappetta et al. 2004). Her2 is a transmembrane protein with substantial homology to epidermal growth factor receptor. Amplification of this gene coupled with resultant over-expression of the protein occurs in about 25% of human breast cancers. It has been reported that amplification of c-ErbB2 is associated with a fast proliferation and a poor prognosis of breast cancer (Vogelstein and Kinzler, 1994). Therefore, HMGA1 over-expression could not only represent a predictor of prognosis and outcome, as exerted by amplification of Her2, but it can be also predicted a functional link with the amplification of the Her2 gene. In fact, several studies have demonstrated an accumulation of amplifications of different genomic regions in certain breast cancers considered to exhibit an “amplifier” phenotype (Al-Kuraya et al. 2004, Courjal et al. 1997). Therefore, we could hypothesize that that HMGA1 over-expression may contribute to the accumulation of Her2/neu amplifications, or alternatively that the activation of the ErbB2 transduction pathway may lead to increased HMGA1 protein synthesis.

Likewise, we found a strong relationship between the HMGA1 over-expression and the presence or absence of PR and ER, respectively. It is well known that breast carcinomas not expressing ER are insensitive to estrogen antagonists such as tamoxifen and, as consequence, these carcinomas are insensitive to hormonal treatments and display a negative outcome (Teschendorff et al. 2007). These findings are consistent with HMGA1 over-expression seems associated with a highly malignant phenotype, also representing a poor prognostic index since HMGA1 over-expression often correlates with the presence of metastasis, and with a reduced survival, as it has been extensively reported elsewhere (Abe et al. 2003, Meyer et al. 2007).

On the other hand, the positive correlation between the HMGA1 over-expression and the presence of the PR (Table 1), apparently would link the expression of HMGA1 to a better prognosis, since PR+ breast carcinoma patients respond to the hormonal treatment, even though there are reports indicating that breast cancer patients with ER-/PR+ tumours arose primarily premenopausal and in younger people (Rakha et al. 2007, Rhodes et al. 2009, Cserni et al. 2011). Although these ER-/PR+ patients are generally considered candidates for endocrine therapy, they gain less benefit from adjuvant tamoxifen treatment than ER+/PR+ patients and ER-/PR+ patients <55 years old were found to have significantly worse survival than younger ER+/PR+ patients (Yu et al. 2008). Therefore, considering the outcome, the PR expression needs to be evaluated together with the ER status, whose expression is negatively related to that of HMGA1. Since HMGA1 is a chromatin protein able to up- or down-regulate the expression of important cancer-related genes (Fusco and Fedele, 2007), it could be envisaged that HMGA1 proteins could be able to modulate the expression of both ER and PR expression acting on its promoter. This could have important prognostic significance, since the over-expression of HMGA1 could influence the hormonal responsiveness and, therefore, the outcome of breast cancer patients.

In conclusion, taken together, our findings indicate that the level of HMGA1 expression is related to an unfavourable breast cancer phenotype and poor prognosis, as supported by its strong association with the Her2/neu, PR and ER status that could explain, at least in part, the different behaviour of the human breast carcinoma over-expressing HMGA1. These findings further support the HMGA1 as

an appropriate target for the therapy of human cancer, as suggested by numerous *in vitro* and *in vivo* studies modulating its expression in cancer cells.

ACKNOWLEDGEMENTS

This work was supported by grants from the Associazione Italiana per la Ricerca sul Cancro (AIRC), Progetto Strategico Oncologia, Consiglio Nazionale delle Ricerche, the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MIUR), PNR-CNR Aging Program 2012-2014. We thank the Associazione Partenopea per le Ricerche Oncologiche (APRO) for its support.

CONFLICT OF INTEREST STATEMENT

The authors have declared that no conflict of interest exists.

REFERENCES

Chiappetta G, Botti G, Monaco M, et al. HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression.

Clin Cancer Res 2004;10:7637–44.

Meyer B, Loeschke S, Schultze A, et al. HMGA2 overexpression in non-small cell lung cancer.

Mol Carcinogen 2007;46:503–11.

Abe N, Watanabe T, Izumisato Y, et al. High mobility group A1 is expressed in metastatic adenocarcinoma to the liver and intrahepatic cholangiocarcinoma, but not in hepatocellular carcinoma: its potential use in the diagnosis of liver neoplasms.

J Gastroenterol 2003;38:1144–9.

Ke-da Yu, Gen-hong Di, Jiong Wu, Jin-song Lu, Kun-wei Shen, Guang-yu Liu, Zhenzhou Shen, Zhio-Ming Shao Breast cancer patients with estrogen receptor-negative/progesterone receptor-positive tumors: being younger and getting less benefit from adjuvant tamoxifen treatment

Journal of Cancer Research and Clinical Oncology, December 2008

Johnson KR, Lehn DA, Reeves R. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y.

Mol Cell Biol 1989;9:2114–23.

Grosschedl R, Giese K, Pagel J. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures.

Trends Genet 1994;10:94–100.

Thanos D, Maniatis T. The high mobility group protein HMG I(Y) is required for NF- κ B-dependent virus induction of the human IFN- β gene.

Cell 1992;71:777–89.

Fusco A, Fedele M. Roles of the HMGA proteins in cancer.

Nat Rev Cancer 2007;7:899–910.

Berlingieri MT, Manfioletti G, Santoro M, et al. Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells.

Mol Cell Biol 1995;15:1545–53.

Berlingieri MT, Pierantoni GM, Giancotti V, Santoro M, Fusco A. Thyroid cell transformation requires the expression of the HMG1 proteins.

Oncogene 2002;21:2971–80.

Wood LJ, Maher JF, Bunton TE, Resar LM. The oncogenic properties of the HMG-I gene family.

Cancer Res 2000;60:4256–61.

Reeves R, Edberg DD, Li Y. Architectural transcription factor HMGI (Y) promotes tumor progression and mesenchymal transition of human epithelial cells.

Mol Cell Biol 2001;21:575–94

Vogelstein B, Kinzler KW. Has the breast cancer gene been found?

Cell 1994;79:1–3.

Ram TG, Reeves R, Hosick HL. Elevated high mobility group-I(Y) gene expression is associated with progressive transformation of mouse mammary epithelial cells.

Cancer Res 1993;53(11):2655– 60.

Baldassarre G, Battista S, Belletti B, et al. Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma.

Mol Cell Biol 2003;23:2225–38.

Reeves R, Edberg DD, Li Y. Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells.

Mol Cell Biol 2001;21:575–94.

Chiappetta G, Botti G, Monaco M, Pasquinelli R, Pentimalli F, Di Bonito M, D'Aiuto G, Fedele M, Iuliano R, Palmieri EA, Pierantoni GM, Giancotti V, Fusco A. HMGA1 Protein Overexpression in Human Breast Carcinomas: Correlation with ErbB2 Expression.

Clinical Cancer Research. Vol. 10, 7637–7644, November 15, 2004

Piscuoglio, S., et al., Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater. J Clin Pathol, 2012. **65**(2): p. 140-5.

Ellis IO, et al. (2003) in World Health Organization Classification of Tumours. Pathology and Genetics of Tumors of the Breast and Female Genital Organs, eds Tavassoli FA, Devilee P (IARC Press, Lyon, France), pp 13–59.

Miecznikowski et al. BMC Cancer 2010, 10:573

Deroo B, Korach K: Estrogen receptors and human disease. Journal of Clinical Investigation 2006, 116(3):561-570.

Gao X, Nawaz Z: Progesterone receptors- animal models and cell signaling in breast cancer: Role of steroid receptor coactivators and corepressors of progesterone receptors in breast cancer. Breast Cancer Res 2002, 4(5):182.

Hynes N, Stern D: The biology of erbB-2/neu/HER-2 and its role in cancer. Biochimica et biophysica acta 1994, 1198(2-3):165.

Koumoundourou D, Kassimatis T, Zolota V, Tzorakoeleftherakis E, Ravazoula P, Vassiliou V, Kardamakis D, Varakis J: Prognostic Significance of TGFβ-1 and pSmad2/3 in Breast Cancer Patients with T1-2, N0 Tumours. Anticancer research 2007, 27(4C):2613.

Carter P, et al. (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci USA* 89:4285–4289.

Arteaga CL (2003) Trastuzumab, an appropriate first-line single-agent therapy for HER2-overexpressing metastatic breast cancer. *Breast Cancer Res* 5:96–100.

Banerjee S, et al. (2006) Basal-like breast carcinomas: Clinical outcome and response to chemotherapy. *J Clin Pathol* 59:729–735.

Chia-Chen Liu, Julie Prior, David Piwnica-Worms, and Guojun Bua. LRP6 overexpression defines a class of breast cancer subtype and is a target for therapy. approved January 28, 2010 (received for review October 1, 2009) 5136–5141 | *PNAS* | March 16, 2010 | vol. 107 | no. 11

Vogelstein B, Kinzler KW. Has the breast cancer gene been found? *Cell* 1994;79:1–3.

Al-Kuraya K, Schraml P, Torhorst J, et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res* 2004;64:8534-40.

Courjal F, Theillet C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. *Cancer Res* 1997;57:4368-77.

Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol.* 2007;8(8):R157.

Rakha EA, El-Sayed ME, Green AR et al (2007) Biological and clinical characteristics of breast cancer with single hormone receptor-positive phenotype. *J Clin Oncol* 25:4772–4778

Rhodes A, Jasani B (2009) The oestrogen receptor-negative/ progesterone receptor-positive breast tumour: a biological entity or a technical artefact? *J Clin Pathol* 62:95–96

Cserni G, Francz M, Kálmán E, Kelemen G, Komjáthy DC, Kovács I, Kulka J, Sarkadi L, Udvarhelyi N, Vass L, Vörös A. Estrogen Receptor Negative and Progesterone Receptor Positive Breast Carcinomas-How Frequent are they? *Pathol Oncol Res.* 2011 Jan 26. [Epub ahead of print]

Figures and Tables:

Table 1: Patient characteristics and association of HMGA1 with clinico-pathological features of breast cancer patients. Raw, continuous scores (% positive staining) are used.

Clinico-pathological feature		HMGA1		P-value
		Mean (%) ± SD,	median (%)	
Histological subtype	Ductal carcinoma (n=963)	27.4±30.4	12.5	0.883
	Other (n=375)	26.9±28.4	15	
pT stage	pT1 (n=447)	28.4±30.2	15	0.402
	pT2 (n=656)	27.2±29.9	11.3	
	pT3 (n=84)	26.9±27.9	18.8	
	pT4 (n=134)	24.7±29.7	10	
pN stage	pN0 (n=598)	29.0±30.4	20	0.293
	pN1 (n=452)	26.4±29.5	10	
	pN2 (n=105)	24.7±28.3	10	
	pN3 (n=27)	34.4±34.5	20	
Her2/neu	Staining 0+1 (n=1059)	26.5±29.7	10	0.004
	Staining 2+3 (n=201)	32.2±30.9	20	
ER status	0% (n=322)	31.1±30.9	20	0.007
	>0% (n=978)	26.2±29.5	10	
PR status	0% (n=603)	23.6±28.0	10	0.003
	>0% (n=402)	29.3±31.1	20	

Figure 1

Representative IHC staining of human breast cancers

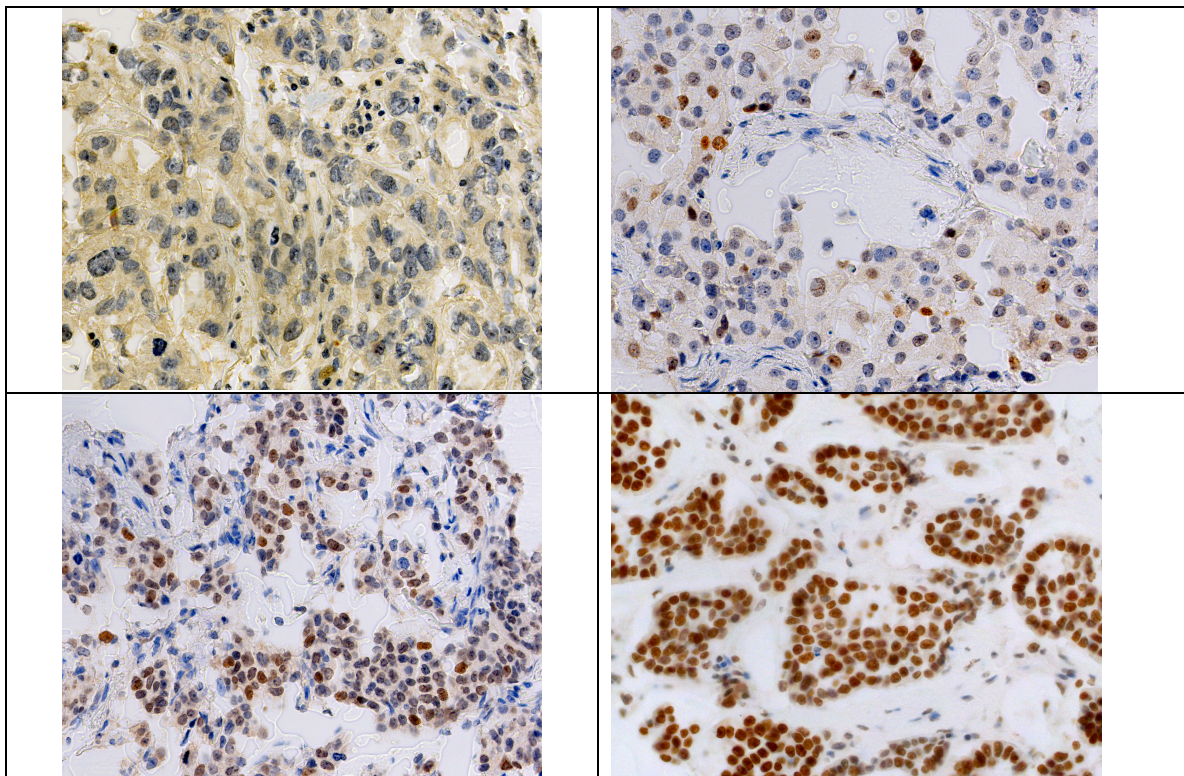


Figure 1. Immunohistochemical analysis of a human breast cancer TMA for HMGA1 protein expression. Examples of HMGA1 immunohistochemistry results in breast carcinomas: negative (A), weak (B), moderate (C), and strong (D) staining.

Figure 2

Representative IHC staining of human ductal breast cancers.

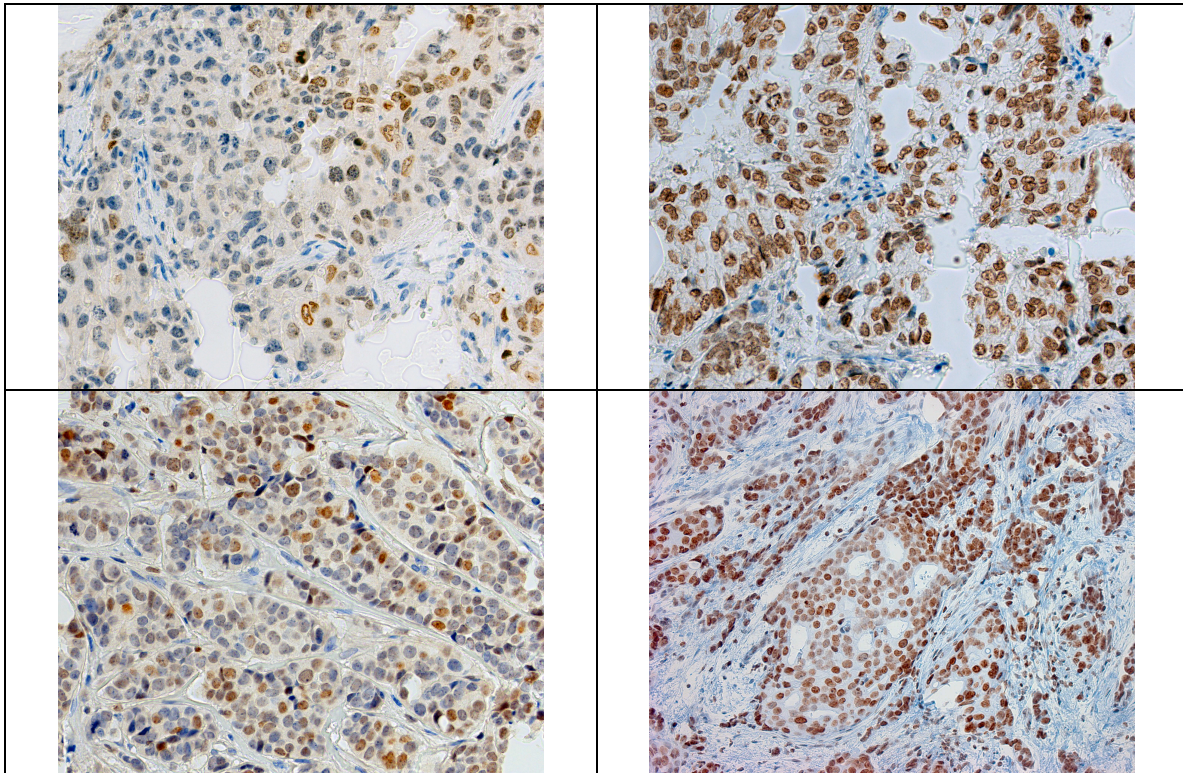


Figure 2. Immunohistochemical analysis of HMGA1 expression in breast ductal carcinomas. ER- breast carcinoma (A). ER+ breast carcinoma (B). PR- breast carcinoma (C). PR+ breast carcinoma (D).

6.4 Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater.

My contribution to this work:

- Immunohistochemical analysis of whole tissue and microarray sections;
- Statistical comparison and survival analysis;
- Data analysis and manuscript writing;

Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater

Salvatore Piscuoglio,¹ Frank S Lehmann,² Inti Zlobec,¹ Luigi Tornillo,¹ Wolfgang Dietmaier,³ Arndt Hartmann,⁴ Peter H Wünsch,⁵ Fausto Sessa,⁶ Petra Rümmele,³ Daniel Baumhoer,¹ Luigi M Terracciano¹

► Additional materials are published online only. To view these files please visit the journal online (<http://jcp.bmj.com/content/65/2.toc>).

¹Institute of Pathology, University Hospital of Basel, Basel, Switzerland
²Department of Gastroenterology and Hepatology, University Hospital of Basel, Basel, Switzerland
³Department of Pathology, University Hospital of Regensburg, Regensburg, Germany
⁴University Hospital of Erlangen, Erlangen, Germany
⁵Municipal Hospital of Nürnberg, Nürnberg, Germany
⁶Anatomic Pathology Unit, University of Insubria, Varese, Italy

Correspondence to
 Prof Luigi M Terracciano,
 Institute of Pathology, University Hospital of Basel,
 Schönbeinstrasse 40, 4003
 Basel, Switzerland;
lterracciano@uhbs.ch

Salvatore Piscuoglio and Frank S Lehmann are equally contributed to this work.

Accepted 13 July 2011
 Published Online First
 30 November 2011

ABSTRACT

Background Carcinomas of the Vaterian system are rare and presumably arise from pre-existing adenomas.

According to the cancer stem cell (CSC) hypothesis, only a small subset of tumor cells has the ability to initiate and develop tumor growth. In colorectal cancer, CD44, CD133, CD166 and EpCAM have been proposed to represent CSC marker proteins and their expression has been shown to correlate with patient survival.

Aims To evaluate a potential role of these CSC proteins in tumors of the ampulla of Vater, we investigated their expression in 175 carcinoma, 111 adenoma and 152 normal mucosa specimens arranged in a Tissue Microarray format.

Materials and methods Membranous immunoreactivity for each protein marker was scored semi-quantitatively by evaluating the number of positive tumor cells over the total number of tumor cells. Median protein expression levels were used as cut-off scores to define protein marker positivity. Clinical data including survival time were obtained by retrospective analysis of medical records, tumor registries or direct contact.

Results The expression of all evaluated marker proteins differed significantly between normal mucosa, adenoma and carcinoma samples. In all markers, we found a tendency towards more constant expression from normal to neoplastic tissue. EpCAM expression was significantly correlated with better patient survival. The increased expression of CD44s, CD166 and CD133 from normal mucosa samples to adenoma and carcinoma was linked to tumor progression. However, there was no statistically significant correlation with survival.

Conclusion Our findings indicate, that in ampullary carcinomas, loss of expression of EpCAM may be linked to a more aggressive tumor phenotype.

INTRODUCTION

The ampulla of Vater combines the terminal and common segment of the bile and pancreatic duct before they enter the duodenum.¹ Carcinomas originating from this complex anatomical unit are uncommon and have an incidence of approximately four to six cases per million population.^{2–3} Carcinomas of the papilla of Vater, defined as junction of the biliary, and pancreatic ducts within the duodenum account for 6%–20% of all peripancreatic tumours⁴ and represent 10%–50% of all cancers resected by pancreaticoduodenectomy.⁵ They can be sited in the ampulloduodenal part of

the papilla of Vater, which is lined by intestinal mucosa. They also can develop in deeper parts of the ampulla, which are lined by pancreaticobiliary duct mucosa. Clinically, tumours of the ampulla of Vater are rapidly detected due to biliary outflow obstruction.^{6–7} Early symptoms as well as differences in tumour biology are held responsible for their favourable clinical outcome (median survival 30–50 months, 5-year survival rate 21%–64%).^{8–9} Histologically, intestinal, pancreaticobiliary, intestinal-mucinous, invasive papillary and poorly differentiated subtypes can be distinguished.¹⁰ The subtypes differ in several clinical and histological aspects including cell type-specific markers, oncogene expression, modes of tumour spread as well as extent and interaction with the extracellular matrix.¹¹ Most authors agree that local spread of the tumour (T stage) is the only significant and independent prognostic factor for this cancer, whereas the predictive value of tumour grade and lymph node metastases is still debated.^{12–13} More recent research data suggest that the prognosis of ampullary cancer may be related to the histological differentiation in intestinal or pancreatobiliary types.¹⁴ In the last years, several molecular markers have been proposed as additional prognostic factors. However, most of these studies have yielded conflicting results and have not been still validated by other reports.^{15–19} Several sources of discrepancy between different reports have been acknowledged mainly due to non-standardised assays often performed on underpowered patient samples that are too small to enable meaningful conclusions to be drawn. Therefore, there is undoubtedly a need for additional prognostic markers for such neoplasia. Recent findings support the concept that cells with the properties of stem cells are integral to the development and perpetuation of several forms of human cancer.^{20–21} Cancer stem cells (CSCs) have low replicative ability, multipotency and resistance to apoptosis and are responsible for tumour development.²² In the different types of digestive tumours, different sets of markers have emerged as the most useful for the identification of CSC. In particular, in intestinal as well as in pancreatic cancer, some markers including CD44, EpCAM, CD166 and CD133 have been indicated as possible CSCs markers. Furthermore, in colorectal cancer, we have shown that their expression inversely correlated with patient survival.²³ However, conflicting results have been

reported about the role of some putative CSC markers in gastrointestinal tract tumours. In particular, contradictory findings have been reported about the association of CD44, in particular of its v6 splicing variant, and tumour progression.^{24–26} Furthermore, while CD133 molecule was initially identified as a reliable CSC marker in human colorectal cancers,^{27 28} a subsequent study has shown that in both mouse and human colorectal cancers, CD133 expression is not restricted to rare cell subsets, but it is detectable in a large majority of tumour cells, irrespective of their tumourigenicity.²⁹ Because of the lacking studies dealing with CSC markers in ampullary tumours, the aim of this study was to elucidate the expression and the prognostic role of CD133, CD166, CD44s, EpCAM expression in ampullary tumours by using a tissue microarray (TMA) including 175 carcinoma, 111 adenoma and 152 normal mucosa specimens of the papilla of Vater.

MATERIAL AND METHODS

Ethics

The study has been approved by the institutional review board of the Department of Pathology, University of Basel, Switzerland. All the analyses were performed according to the ethical standards required by each local ethic committee.

Patients' characteristics and tissue samples

Patients' characteristics have been previously described by our study group.^{30 31} Briefly, the files of the Institute of Pathology, University Hospital Basel (Switzerland), the Institute of Pathology, University of Regensburg (Germany), the Institute of Pathology Nuernberg and the Anatomic Pathology Unit, Department of Human Morphology, University of Insubria, Varese (Italy), were searched for adenomas or carcinomas of the ampulla of Vater over the period from 1985 to 2005. In total, 175 carcinoma, 111 adenoma and 152 normal mucosa samples were retrieved. Sufficient paraffin-embedded tissue for TMA construction was available in all cases. The male-to-female ratio was 3:2; mean age at diagnosis was 63 years (range 15–81 years). To our knowledge, no case was associated with Familial adenomatous polyposis (FAP).

TMA construction

TMAs were constructed from formalin-fixed and paraffin-embedded specimens using a custom-built instrument (Beecher Instruments, Silver Spring, Maryland, USA) as previously described.^{30 31} Briefly, H&E-stained sections were obtained from each selected primary block (donor block) to define representative tissue regions. Core biopsies (0.6 mm cylinders) were taken from the selected tissue regions and then transferred to a paraffin recipient block. The resulting TMA was cut into 4 µm sections, which were used for immunohistochemistry. The number of punches per patient ranged from one to three for both normal tissue and carcinoma and from one to five for adenomas. If more than one punches was obtained, the additional punches were taken from different representative blocks.

Histology and immunophenotyping

All tumours were classified according to the guidelines of the Armed Forces Institute of Pathology using only H&E stains.¹⁰ Mild dysplasia was designated as low-grade dysplasia, whereas moderate and severe dysplasia was considered as high-grade dysplasia. Carcinomas histologically indistinguishable from colorectal carcinomas were classified as intestinal types, whereas carcinomas showing a dense desmoplastic stroma surrounding

small glands or solid nests of tumour cells were referred to as the pancreaticobiliary subtype. Invasive papillary carcinomas typically formed papillary and micropapillary structures in their invasive component and poorly differentiated adenocarcinomas lacked histologic features of glandular or other differentiation. Additionally, an intestinal-mucinous subtype, characterised by any mucinous differentiation, was defined.

For immunohistochemistry, sections were pre-treated with CC1 (Ventana Medical Systems, Tucson, Arizona, USA) and incubated with primary antibodies against CD133, CD44s, CD166 and EpCAM (table 1). Staining procedures were performed on a Benchmark immunohistochemistry staining system (Ventana Medical Systems) using iVIEW-DAB as chromogen.

Membranous immunoreactivity for each protein marker was scored semiquantitatively by evaluating the number of positive tumour cells over the total number of tumour cells. Scores were assigned using 5% intervals and ranged from 0% to 100%. All tissues were scored by an experienced pathologist (LT), blinded to clinicopathological information. To define interobserver agreement, all samples were examined independently by a second pathologist (DB).

Statistical analysis

Statistical correlations between categorical variables were tested using a χ^2 or Fisher exact test, where appropriate. Differences in patient survival were demonstrated using the Kaplan–Meier method and analysed using the log-rank test in univariate analysis. All tests were two sided. p Values <0.05 were considered statistically significant. Cut-off scores were selected by evaluating the receiver operating characteristic curves for each protein marker and the end-point survival. The point on the curve with the shortest distance to the coordinate (0, 1) was selected as the threshold value to classify cases as 'positive/overexpressing' or 'negative/loss'.³² Analysis was performed using SAS V.9.1 (SAS Institute).

RESULTS

Patients' characteristics and tissue samples

Overall, 175 carcinoma, 111 adenoma and 152 normal mucosa samples were retrieved. In patients with carcinomas, we found 19 pT1 (11%), 59 pT2 (34%), 63 pT3 (36%) and 13 pT4 (7%) tumours as well as 17 G1 (10%), 82 G2 (47%) and 55 G3 (31%) cases (no data concerning T stage and grading in 21 samples). Seventy-three (42%) carcinoma patients were node positive and two (1%) had haematogenous metastases at initial diagnosis.

Full clinical data including survival time were available in 133 patients with ampullary carcinoma (76%). Patients were studied up to 164 months after operation. Median follow-up time was 36 months.

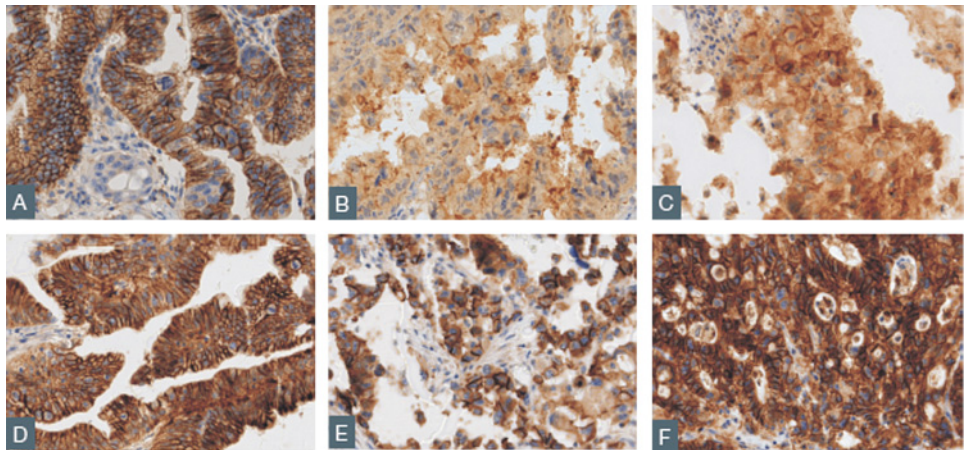
The histological classification of 175 ampullary carcinomas identified 85 intestinal types, 42 pancreaticobiliary types, 23 poorly differentiated adenocarcinomas, 16 intestinal-mucinous types and nine invasive papillary types.

Table 1 Primary antibodies against CD133, CD44, CD166 and EpCAM

Antibody	Dilution/detection	Pretreatment
CD44 (Dako, DF1485)	1:50/BOND	BOND ER2
CD133 (cell signalling, C24B9)	1:100/BOND	Steamer 120°C, pH8
CD166 (Novocastra, MOG/07)	1:200/BOND	BOND ER2
EpCAM (Novocastra, VU-1D9)	1:200/BOND	BOND ER2

Original article

Figure 1 Immunophenotyping in tissue samples from patients with ampullary carcinoma (400×). (A) CD44. (B and C) CD133. (D and E) CD166. (F) EpCAM.



Histological grading was evaluable in 76 of 111 (68%) adenomas and disclosed low-grade dysplasia in 57 of 76 (75%) and high-grade dysplasia in 19 of 76 (25%) cases. All adenomas demonstrated tubular or tubulovillous architecture. Seventy-eight of 111 (70%) adenoma samples were derived from patients with coexisting carcinoma.

Immunophenotyping

Tissue samples of ampullary carcinoma patients expressing CD44, CD133, CD166 and EpCAM are shown in figure 1. Moreover, in table 2, the distribution of the different biomarkers across different diagnostic categories is shown.

The expression of all marker proteins differed significantly between carcinoma, adenoma and normal mucosa samples (table 3).

We have also evaluated the positivity in the two principal histologic types (intestinal type vs pancreatobiliary type). EpCAM was significantly more expressed in intestinal type (table 4).

We have tried also to evaluate if there is some difference between adenomas without coexisting carcinoma and adenomas with coexisting carcinomas. Only CD44 was significantly more expressed in adenomas with coexisting carcinomas ($p=0.043$).

No difference was found between low-grade and high-grade adenomas (data not shown).

Survival

Five-year survival (95% CI) was 45.2 (34 to 56) in EpCAM-positive versus 28.2 (11 to 48) in EpCAM-negative patients ($p<0.05$). EpCAM was not an independent prognostic factor after adjusting for pT and pN stages. Survival curves of both patient groups using the Kaplan–Meier method are demonstrated in figure 2.

DISCUSSION

Tumours of the papilla of Vater are a relatively rare neoplastic entity that came into focus in recent years. Significant overlap exists in phenotypic and molecular characteristics between ampullary and colorectal carcinomas. As in colorectal cancer, the development of ampullary carcinoma from adenomas as precancerous lesions has been well documented, and studies investigating molecular alterations associated with the proposed adenoma–carcinoma sequence have been also performed, including our group.³⁵ However, still missing is a comprehensive analysis of the expression of putative CSC markers in very large groups of patients, amenable to detailed statistical analysis. Moreover, the prognostic significance of the co-expression of multiple CSC markers within the same tumour has not been evaluated so far.

This is the first systematic study assessing the prognostic value of four CSC markers, namely EpCAM, CD44, CD133 and CD166, in a large series of patients with ampullary tumours.

Table 2 Distribution of biomarkers between different diagnostic categories

	Number of cases within each expression category							Sum
	0%	1%–5%	6%–20%	21%–40%	41%–60%	61%–80%	81%–100%	
Normal								
CD133	9	63	30	0	1	1	0	104
CD44	67	20	8	7	3	6	4	115
CD166	81	5	6	3	6	1	13	115
EpCAM	7	1	0	0	2	2	89	101
Adenoma								
CD133	10	61	14	0	0	0	0	85
CD44	25	13	17	5	3	14	12	89
CD166	58	9	5	1	4	3	6	86
EpCAM	0	0	0	1	2	1	89	93
Cancer								
CD133	16	83	45	6	1	0	0	151
CD44	38	15	27	19	9	18	18	144
CD166	84	19	14	9	5	7	10	148
EpCAM	2	0	1	1	3	3	136	146

Table 3 Expression of all evaluated marker proteins in normal mucosa, adenoma and carcinoma samples

	Normal	Adenoma	Carcinoma	p Value
CD44+ (>5%)	29/118 (24.6%)	52/90 (57.8%)	91/145 (62.8%)	<0.001
CD133+ (>5%)	30/104 (28.9%)	11/74 (12.9%)	44/151 (29.1%)	0.012
CD166+ (>0%)	34/120 (28.3%)	28/87 (32.2%)	64/151 (42.4%)	0.044
EpCAM + (100%)	89/104 (85.6%)	88/93 (94.6%)	115/145 (79.3%)	0.005

The overexpression of EpCAM was significantly correlated with better survival time. The increasing expression of CD44, CD166 and CD133 from normal mucosa samples to adenoma and carcinoma was linked to tumour progression. However, there was no statistically significant correlation with survival.

EpCAM is a glycosylated, 30–40 kDa type I membrane protein, which is expressed in a variety of human epithelial tissue cancers, as well as in progenitor and stem cells. It is composed of an extracellular domain with epidermal growth factor and thyroglobulin repeat-like domains, a single transmembrane domain and a short 26 amino acid intracellular domain called EpICD. In normal cells, EpCAM is predominantly located in intercellular spaces, where epithelial cells form very tight junctions. Therefore, on normal epithelia, it is sequestered and may be much less accessible to antibodies than in cancer tissue, where it is homogeneously distributed on the cell surface. Furthermore, EpCAM is part of the signature of cancer-propagating cells in numerous solid tumours as well as in normal progenitor and stem cells.³⁴

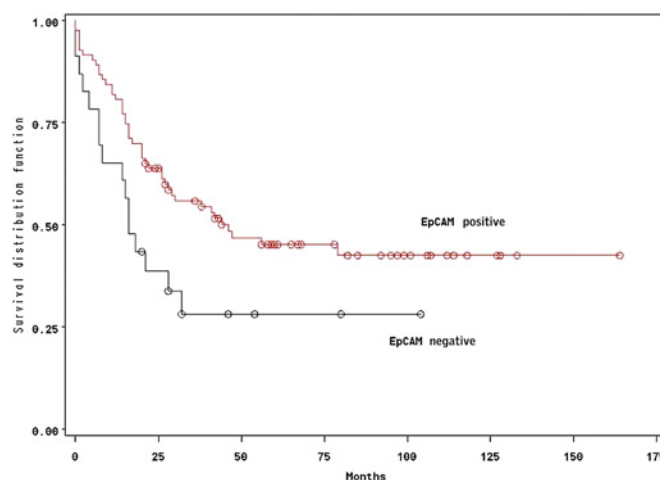
EpCAM was one of the first tumour-associated antigens identified in the late 1970s. Systematic analysis of EpCAM expression for intensity and frequency showed that EpCAM is expressed on essentially all human adenocarcinoma, on certain squamous cell carcinoma, on retinoblastoma and on hepatocellular carcinoma.³⁵

Importantly, EpCAM is part of the signature of cancer-propagating cells in numerous solid tumours and of normal progenitor and stem cells.³⁴ The controversial biological role of EpCAM has recently been discussed by van der Gun *et al.*³⁶ It is of interest that EpCAM overexpression has been associated with both decreased and increased survival time. EpCAM exerts different effects on cell adhesion, either promoting or preventing metastasis.³⁶ The correlation of EpCAM expression and poor survival has been described in several tumour types, including invasive breast cancer,³⁷ urothelial carcinoma of the bladder,³⁸ gallbladder carcinoma^{39 40} and squamous cell carcinoma of the oesophagus.⁴¹ In different tumours, studies on EpCAM-directed immunotherapeutic therapies are currently in clinical development. Therapeutic trials of monoclonal antibodies directed against EpCAM have shown that they may induce antibody-based cellular cytotoxicity by adhering to cytokines such as interleukin 2 or complement-based cytotoxicity by activating T cytotoxic cells.^{42 43}

Table 4 Expression of evaluated markers in the two principal histologic subtypes of ampullary carcinoma

	Intestinal	Pancreatobiliary	p Value
CD44+ (>5%)	82/99 (82.8%)	33/41 (80.5%)	0.805
CD133+ (>5%)	27/88 (30.7%)	12/40 (30.0%)	0.938
CD166+ (>0%)	73/97 (76.8%)	32/44 (72.7%)	0.672
EpCAM+ (100%)	68/75 (90.7%)	25/33 (75.8%)	0.039

The bold value means that EpCAM overexpression is statistically significant.

**Figure 2** Survival of patients with EpCAM-positive and -negative ampullary carcinoma using the Kaplan–Meier method.

Previous studies have yielded conflicting results regarding EpCAM expression and survival in ampullary carcinoma. By Scheunemann *et al.*,⁴⁴ EpCAM expression has been associated with poorer survival in tumours of the papilla of Vater. This study, however, was mainly focused on the frequency and prognostic impact of minimal tumour cell spread in lymph nodes classified as ‘tumour free’ in routine histopathologic evaluation, and EpCAM expression in primary tumours was not analysed.

More recently, Fong *et al.*⁴⁵ were unable to prove a prognostic value of EpCAM overexpression in pancreatic and ampullary carcinoma. In contrast, in our study, we showed that decreased EpCAM membranous expression significantly correlated with biological features of aggressive tumour behaviour. Our study therefore suggests that diminished EpCAM expression is related to tumour invasiveness and progression and it is linked to a more aggressive tumour phenotype. This could also be confirmed by the more frequent expression of EpCAM in intestinal-type than in pancreatobiliary-type tumours. Intestinal-type tumours are in fact associated with a better prognosis.¹⁴

Several reasons for the discrepancies between our results and previous studies can be hypothesised including differences in sample size (power for detecting prognostic differences), methodology (TMA vs whole tissue sections), different clones of antibody and, most importantly, the choice of cut-off scores for the definition of positive staining or staining intensity.

EpCAM is intensely used as a therapeutic target for antibody-based approaches. Future development of EpCAM-directed therapeutics may profit from newly identified functions of EpCAM as mitogenic signal transducer in various ways. An important insight is that EpCAM is apparently needed to maintain distinct cancer cell attributes⁴⁶ and, potentially, the CSC phenotype as well. This function can reduce the risk of immune escape by loss of EpCAM target expression from cancer cells. EpCAM-directed therapies may be selective for those cancer cells with the strongest negative impact on prognosis and for cancer-propagating subsets of malignant cells.

Discrepant results have also been reported regarding the effect of CD44 gene or protein and its splice variants on survival in tumour patients and it is an important receptor that binds hyaluronan (HA). CD44 has previously been considered to be a marker of tumour invasiveness and metastasis. Only recently, it has been described as putative colorectal CSC marker.

Take-home messages

- ▶ CD44, CD133 and CD166 but not EpCAM immunohistochemical expression increased from normal mucosa to adenoma to carcinoma.
- ▶ EpCAM was more expressed in intestinal-type ampullary carcinoma.
- ▶ Loss of EpCAM may be linked to a more aggressive tumour phenotype.
- ▶ Loss of expression of EpCAM, but not of CD44, CD133 or CD166, is linked to poor survival.

However, *CD44* does not seem to belong to the group of genes, such as *OCT4* and *NANOG*, that are central for maintaining stem cell characteristics. Nonetheless, two connections between *CD44* and genes that regulate stem cell characteristics have been described. First, *CD44* is a target of the WNT pathway. Loss of adenomatous polyposis coli (APC) function leads to the constitutive activation of β catenin, a constituent of the WNT signalling pathway. *CD44s* and *CD44v6* expression is restricted to the intestinal crypts in non-transformed tissue, but both *CD44* isoforms are strongly overexpressed in dysplastic crypts and adenomas in humans and mice with mutant APC.⁴⁷ Second, HA-*CD44* binding promotes protein kinase C (PKC) activation and this increases *NANOG* phosphorylation and translocation to the nucleus. Here, it associates with Droscha and an RNA helicase, p68, leading to the transcription of the oncogenic microRNA (miRNA) miR-21 and a reduction in the expression of the tumour suppressor programmed cell death 4. These events initiate the upregulation of the inhibitor of apoptosis (IAP) proteins and multidrug resistant protein 1 (MDR1). *CD44*, in turn, associates with and stabilises MDR1 expression.⁴⁸ This could be one mechanism through which *CD44* contributes to stem cell resistance to chemotherapy, as MDR1 exports several drugs from cells.

In several gastrointestinal tumours, including colorectal cancer,²³ stromal⁴⁹ as well as neuroendocrine tumours,⁵⁰ loss of *CD44* expression has been associated with disease progression and reduced survival. In our study, *CD44* expression was more frequently associated with adenomas with coexisting carcinomas.

Our study is also the first to evaluate the prognostic impact of *CD166* in ampullary carcinoma. We found an increasing expression of *CD166* from normal tissue to carcinoma, suggesting that the increased expression of this marker might be linked to tumour progression. Our data are supported by a recent study in colorectal cancer patients, where a similar increasing expression of *CD166* from normal to neoplastic tissue has been described by Weichert *et al.*⁵¹

We speculate that our findings of decreased rather than increased expression of membranous EpCAM expression and its association with features of tumour progression are mainly a consequence of its cell adhesion function.

In colorectal cancer, Kojima *et al.*⁵² and Horst *et al.*⁵³ reported a significant correlation of increased *CD133* expression and poor clinical outcome. In contrast, in a study on non-small-cell lung cancer, *CD133* expression was not a prognostic factor for survival.⁵⁴ Consistent with the latter study, we found no significant impact of *CD133* on survival in our series of ampullary cancer patients.

In summary, we have provided evidence that in ampullary carcinoma, loss of expression of EpCAM, but not of *CD44*, *CD133* or *CD166*, is linked to poor survival.

Competing interests None.

Patient consent If yes, who signed it?

Ethics approval Ethische Kommission Beider Basel.

Contributors SP, FSL and DB chose the antibodies, collected the data, wrote the introduction and the discussion. LMT and DB performed the immunohistochemical evaluation. SP performed the immunohistochemical staining. IZ performed the statistical evaluation. LT, LMT, PR, FS, PHW and WD revised and reclassified the tumours. LMT and LT revised the manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Allescher HD. Papilla of Vater: structure and function. *Endoscopy* 1989;**21**(Suppl 1):324–9.
2. Neoptolemos JP, Talbot IC, Carr-Locke DL, *et al.* Treatment and outcome in 52 consecutive cases of ampullary carcinoma. *Br J Surg* 1987;**74**:957–61.
3. Goodman MT, Yamamoto J. Descriptive study of gallbladder, extrahepatic bile duct, and ampullary cancers in the United States, 1997–2002. *Cancer Causes Control* 2007;**18**:415–22.
4. Brennan MF. Surgical management of peripancreatic cancer. In: Bland KL, Karakoukis CP, Copeland EM, eds. *Atlas of Surgical Oncology*. Philadelphia, PA: WB Saunders Company, 1995:473.
5. Fisher HP, Zhou H. Pathogenesis of carcinoma of the papilla of Vater. *J Hepatobiliary Pancreat Surg* 2004;**11**:301–9.
6. Benhamiche AM, Jouve JL, Manfredi S, *et al.* Cancer of the ampulla of Vater: results of a 20-year population-based study. *Eur J Gastroenterol Hepatol* 2000;**12**:75–9.
7. Albores-Saavedra J, Schwartz AM, Batich K, *et al.* Cancers of the ampulla of Vater: demographics, morphology, and survival based on 5,625 cases from the SEER program. *J Surg Oncol* 2009;**100**:598–605.
8. Wright NH, Howe JR, Rossini FP, *et al.* Tumours of the small intestine. In: Hamilton SR, Aaltonen LA, eds. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Digestive System*. IARC, 2000:69–92.
9. Di Giorgio A, Alfieri S, Rotondi F, *et al.* Pancreatoduodenectomy for tumors of Vater's ampulla: report on 94 consecutive patients. *World J Surg* 2005;**29**:513–18.
10. Albores-Saavedra J, Henson DE, Klimstra DS. *Tumors of the Gallbladder, Extrahepatic Bile Ducts, and Ampulla of Vater*. Washington, DC: Armed Forces Institute of Pathology, 2000:259–316.
11. Schirmacher P, Büchler MW. Ampullary adenocarcinoma—differentiation matters. *BMC Cancer*. Lyon, 2008;**8**:251.
12. Talamini MA, Moesinger RC, Pitt HA, *et al.* Adenocarcinoma of the ampulla of Vater. A 28-year experience. *Ann Surg* 1997;**225**:590–9.
13. Dorandeu A, Raoul JL, Siriser F, *et al.* Carcinoma of the ampulla of Vater: prognostic factors after curative surgery: a series of 45 cases. *Gut* 1997;**40**:350–5.
14. Westgaard A, Tafjord S, Farstad IN, *et al.* Pancreatobiliary versus intestinal histologic type of differentiation is an independent prognostic factor in resected periampullary adenocarcinoma. *BMC Cancer* 2008;**8**:170.
15. Sessa F, Furlan D, Zampatti C, *et al.* Prognostic factors for ampullary adenocarcinomas: tumor stage, tumor histology, tumor location, immunohistochemistry and microsatellite instability. *Virchows Arch* 2007;**451**:649–57.
16. Park S, Kim SW, Kim SH, *et al.* Lack of microsatellite instability in neoplasms of ampulla of Vater. *Pathol Int* 2003;**53**:667–70.
17. Achille A, Biasi MO, Zamboni G, *et al.* Cancers of the papilla of Vater: mutator phenotype is associated with good prognosis. *Clin Cancer Res* 1997;**3**:1841–7.
18. Imai Y, Tsurutani N, Oda H, *et al.* Genetic instability and mutation of the TGF-beta-receptor-II gene in ampullary carcinomas. *Int J Cancer* 1998;**76**:407–11.
19. Suto T, Habano W, Sugai T, *et al.* Infrequent microsatellite instability in biliary tract cancer. *J Surg Oncol* 2001;**76**:121–6.
20. Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. *Trends Cell Biol* 2005;**15**:494–501.
21. Reya T, Morrison SJ, Clarke MF, *et al.* Stem cells, cancer, and cancer stem cells. *Nature* 2001;**414**:105–11.
22. Alison MR, Islam S. Attributes of adult stem cells. *J Pathol* 2009;**217**:144–60.
23. Lugli A, Iezzi G, Hostettler I, *et al.* Prognostic impact of the expression of putative cancer stem cell markers *CD133*, *CD166*, *CD44s*, *EpCAM*, and *ALDH1* in colorectal cancer. *Br J Cancer* 2010;**103**:382–90.
24. Mulder JW, Kruyt PM, Sewnath M, *et al.* Colorectal cancer prognosis and expression of exon-v6-containing *CD44* proteins. *Lancet* 1994;**344**:1470–2.
25. Herrlich P, Pals S, Ponta H. *CD44* in colon cancer. *Eur J Cancer* 1995;**31**:1110–12.
26. Weg-Remers S, Anders M, von Lampe B, *et al.* Decreased expression of *CD44* splicing variants in advanced colorectal carcinomas. *Eur J Cancer* 1998;**34**:1607–11.
27. O'Brien CA, Pollett A, Gallinger S, *et al.* A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 2007;**445**:106–10.

28. **Ricci-Vitiani L**, Lombardi DG, Pilozi E, *et al*. Identification and expansion of human colon-cancer-initiating cells. *Nature* 2007;**445**:111–15.
29. **Shmelkov SV**, Butler JM, Hooper AT, *et al*. CD133 expression is not restricted to stem cells, and both CD133+ and CD133– metastatic colon cancer cells initiate tumors. *J Clin Invest* 2008;**118**:2111–20.
30. **Baumhoer D**, Zlobec I, Tornillo L, *et al*. Immunophenotyping and oncogene amplifications in tumors of the papilla of Vater. *Virchows Arch* 2008;**453**:579–88.
31. **Ruemmele P**, Dietmaier W, Terracciano L, *et al*. Histopathologic features and microsatellite instability of cancers of the papilla of Vater and their precursor lesions. *Am J Surg Pathol* 2009;**33**:691–704.
32. **Zlobec I**, Steele R, Terracciano L, *et al*. Selecting immunohistochemical cut-off scores for novel biomarkers of progression and survival in colorectal cancer. *J Clin Pathol* 2007;**60**:1112–16.
33. **Scarpa A**, Di Pace C, Talamini G, *et al*. Cancer of the ampulla of Vater: chromosome 17p allelic loss is associated with poor prognosis. *Gut* 2000;**46**:842–8.
34. **Gires O**, Klein CA, Baeuerle PA. On the abundance of EpCAM on cancer stem cells. *Nat Rev Cancer* 2009;**9**:143.
35. **Baeuerle PA**, Gires O. EpCAM (CD326) finding its role in cancer. *Br J Cancer* 2007;**96**:417–23.
36. **van der Gun BT**, Melchers LJ, Ruiters MH, *et al*. EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis* 2010;**31**:1913–21.
37. **Spizzo G**, Obrist P, Ensinger C, *et al*. Prognostic significance of Ep-CAM and Her-2/neu overexpression in invasive breast cancer. *Int J Cancer* 2002;**98**:883–8.
38. **Brunner A**, Prelog M, Verdofer I, *et al*. EpCAM is predominantly expressed in high grade and advanced stage urothelial carcinoma of the bladder. *J Clin Pathol* 2008;**61**:307–10.
39. **Prince S**, Zeidman A, Dekel Y, *et al*. Expression of epithelial cell adhesion molecule in gallbladder carcinoma and its correlation with clinicopathological variables. *Am J Clin Pathol* 2008;**129**:424–9.
40. **Varga M**, Obrist P, Schneeberger S, *et al*. Overexpression of epithelial cell adhesion molecule antigen in gallbladder carcinoma is an independent marker for poor survival. *Clin Cancer Res* 2004;**10**:3131–6.
41. **Stoecklein NH**, Siegmund A, Scheunemann P, *et al*. Ep-CAM expression in squamous cell carcinoma of the esophagus: a potential therapeutic target and prognostic marker. *BMC Cancer* 2006;**6**:165.
42. **Prang N**, Preithner S, Brischwein K, *et al*. Cellular and complement-dependent cytotoxicity of Ep-CAM-specific monoclonal antibody MT201 against breast cancer cell lines. *Br J Cancer* 2005;**92**:342–9.
43. **Mosolits S**, Markovic K, Frödin JE, *et al*. Vaccination with Ep-CAM protein or anti-idiotypic antibody induces Th1-biased response against MHC class I- and II-restricted Ep-CAM epitopes in colorectal carcinoma patients. *Clin Cancer Res* 2004;**10**:5391–402.
44. **Scheunemann P**, Stoecklein NH, Rehders A, *et al*. Frequency and prognostic significance of occult tumor cells in lymph nodes in patients with adenocarcinoma of the papilla of Vater. *HPB (Oxford)* 2007;**9**:135–9.
45. **Fong D**, Steurer M, Obrist P, *et al*. Ep-CAM expression in pancreatic and ampullary carcinomas: frequency and prognostic relevance. *J Clin Pathol* 2008;**61**:31–5.
46. **Maetzel D**, Denzel S, Mack B, *et al*. Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol* 2009;**11**:162–71.
47. **Wielenga VJ**, Smits R, Korinek V, *et al*. Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am J Pathol* 1999;**154**:515–23.
48. **Hao J**, Chen H, Madigan MC, *et al*. Co-expression of CD147 (EMMPRIN), CD44v3–10, MDR1 and monocarboxylate transporters is associated with prostate cancer drug resistance and progression. *Br J Cancer* 2010;**103**:1008–18.
49. **Hsu KH**, Tsai HW, Shan YS, *et al*. Significance of CD44 expression in gastrointestinal stromal tumors in relation to disease progression and survival. *World J Surg* 2007;**31**:1438–44.
50. **Lai CH**, Shan YS, Sy ED, *et al*. The significance of CD44 expression in gastrointestinal neuroendocrine tumors. *Hepatogastroenterology* 2005;**52**:1071–6.
51. **Weichert W**, Knösel T, Bellach J, *et al*. ALCAM / D166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. *J Clin Pathol* 2004;**57**:1160–4.
52. **Kojima M**, Ishii G, Atsumi N, *et al*. Immunohistochemical detection of CD133 expression in colorectal cancer: a clinicopathological study. *Cancer Sci* 2008;**99**:1578–83.
53. **Horst D**, Kriegl L, Engel J, *et al*. CD133 expression is an independent prognostic marker for low survival in colorectal cancer. *Br J Cancer* 2008;**99**:1285–9.
54. **Salnikov AV**, Gladkirch J, Moldenhauer G, *et al*. CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. *Int J Cancer* 2010;**126**:950–8.

6.5 MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies.

My contribution to this work:

- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;

MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies

Elke Schultz-Thater^{1*}, Salvatore Piscuoglio^{2*}, Giandomenica Iezzi¹, Clémentine Le Magnen¹, Paul Zajac¹, Vincenza Carafa², Luigi Terracciano², Luigi Tornillo^{2*} and Giulio C. Spagnoli^{1*}

¹Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, Basel, Switzerland

²Institute of Pathology, University of Basel, Basel, Switzerland

MAGE-A10 is a highly immunogenic member of the MAGE-A family of cancer/testis tumor-associated antigens (C/T TAAs). Studies performed with broadly reactive antibodies have helped to initially characterize this TAA. However, no specific reagents have been developed so far, thus preventing a thorough analysis of its expression in healthy and tumoral tissues. We have produced MAGE-A10 gene product in soluble recombinant form, and we have used it to generate specific monoclonal antibodies (mAbs). One of these reagents, recognizing an epitope located at the COOH terminus of the MAGE-A10 gene product, was used to stain a multitumor tissue microarray comprising more than 2,500 paraffin-embedded specimens including healthy tissues, benign tumors and malignancies of different histological origin. MAGE-A10 protein was identified as an intranuclear protein of an apparent molecular weight of 70 kDa, expressed in normal spermatogonia and spermatocytes but in no other healthy tissue. Most importantly, this C/T TAA appears to be expressed in high (>50%) percentages of cancer cells from a number of malignancies, including lung, skin and urothelial tumors. Unexpectedly, high expression of MAGE-A10 TAA at the protein level was also detectable in gynecological malignancies and stomach and gall bladder cancers. The characterization of MAGE-A10-specific reagents might set the stage for the development of targeted active immunotherapy by clarifying potential indications and by allowing the selection of patients eligible for treatment and the monitoring of its effectiveness.

MAGE-A tumor-associated antigens of the cancer/testis family are expressed in a very limited number of healthy tissues typically including spermatogonia, and, for some of them, thymus and placenta. In contrast, they are expressed in a large variety of malignancies derived from diverse tissues, including, among others, melanoma, lung cancers, prostate cancers, breast cancers, hepatocellular carcinoma and head and neck cancers.^{1,2}

In these tumors, their expression can be used for diagnostic purposes or for the identification of patients potentially benefiting of targeted immunotherapies. Quantitative real-time PCR (qRT-PCR) is currently used for the detection, at

the gene level, of the expression of MAGE-A tumor-associated antigens. On the other hand, only a limited number of reagents are available to detect the corresponding proteins.³

MAGE-A10 is probably the most immunogenic antigen of the MAGE-A family,⁴⁻⁸ and, therefore, it represents a potentially highly attractive target of active specific immunotherapies. This antigen, expressed in the form of a 72-kDa nuclear protein, has been identified thanks to broadly reactive monoclonal antibodies (mAbs) recognizing it together with other members of the MAGE-A family.⁹ However, no exquisitely specific reagents have been developed so far, thus preventing a precise identification of cells expressing it in healthy and cancerous tissues and, in particular, in paraffin-embedded specimens. This information is of critical relevance in the selection of patients potentially eligible for targeted active immunotherapies and for the monitoring of their effectiveness.

Upon expression cloning, we have produced MAGE-A10 in soluble form and we have used it to generate highly specific mAbs. Here, we report that MAGE-A10-specific mAbs are able to identify their target antigen in the nuclei of tumor cells in paraffin-embedded clinical samples. Staining of a multitumor array including more than 2,000 specimens indicates that MAGE-A10 cancer/testis tumor-associated antigen (C/T TAA) is highly expressed in lung, skin and urothelial malignancies.

Key words: cancer/testis tumor-associated antigens, MAGE-A10, immunohistochemistry, tissue microarray, cancer immunotherapy
Grant sponsor: SNF; **Grant numbers:** 320030-120320, 3100A0-122235/1

*E.S.-T., S.P., L.T. and G.C.S. contributed equally to this work

DOI: 10.1002/ijc.25777

History: Received 30 Jun 2010; Accepted 14 Oct 2010; Online 15 Nov 2010

Correspondence to: Giulio C. Spagnoli, Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, 20 Hebelstrasse, 4031 Basel, Switzerland, Tel.: +41-61-265-2378, Fax: +41-61-265-3990, E-mail: gspagnoli@uhbs.ch

Material and Methods

Cell lines

MZ-2 and A375 cell lines are gifts of Dr. Rimoldi (Ludwig Institute, Lausanne, Switzerland). RE cell line is a gift of Dr. Siegrist, formerly at the University of Basel, whereas WM115 cell line was obtained from American Type Culture Collection (Rockville, MD). SK-Mel-37 cell line is a gift of Dr. Jungbluth (Ludwig Institute at Memorial Sloan Kettering Cancer Center, New York, NY). Na8 cell line is a gift of Dr. Jotereau (Nantes, France).¹⁰ All cell lines were cultured in RPMI 1640 medium, supplemented with 10% FCS, nonessential amino acids, glutamine and antibiotics (all from Invitrogen, Basel, Switzerland).

Preparation of MAGE-A10 fusion protein

MAGE-A10 entire gene¹¹ was PCR amplified from cDNA derived from the SK-Mel-37 melanoma cell line by using the following primers, allowing the cloning into a suitable expression vector and including EcoR I and Hind III restriction sites:

forward: 5' CCCGAATTCCCTCGAGCTCCAAAGCG TCAG 3'

reverse: 5' CCCAAGCTTATTCAGGGTAGGAGAA 3'.

A 1110-bp band was excised and inserted into pET-32a vector (Novagen, Madison, WI) allowing inducible expression of inserted genes in the form of fusion proteins containing thioredoxin and a six-histidine tail.¹² The plasmid was used to transform BL21(pLysS) *E. coli* strain. After a 4-hr induction in the presence of IPTG (1 mM final concentration), bacterial cultures were lysed and recombinant proteins were purified under native conditions upon binding to nickel resins (Ni-NTA, Qiagen, Basel, Switzerland) and eluted in the presence of 250 mM imidazole. Production and purification of the recombinant proteins were monitored by SDS-PAGE and Coomassie blue staining (Fig. 1a).¹³ Additional recombinant C/T TAAs was similarly produced in *E. coli*.

Production of monoclonal antibodies

BALB/c mice were repeatedly injected i.p. at 2-week intervals with 100 µg Ni-purified material containing MAGE-A10 gene product, in the presence of Sigma adjuvant system (Sigma, Buchs, Switzerland). Three days after a last injection, animals were sacrificed, and fusions were carried out according to standard methods. Screening of HAT-resistant hybridoma supernatants was performed by ELISA.

Detection of MAGE-A10 gene expression

Total cellular RNA was extracted from the cell lines under investigation (see below), reverse transcribed and tested in qRT-PCR assays in the presence of primers and probes specific for β-actin-positive control gene⁵ and of the following MAGE-A10-specific reagents:

forward: 5' CAGGGAGAGCAAGAGGTCAAGA 3'

reverse: 5' GGGAGTGTGGCAGGACTT 3'

FAM probe: CAGCACTGAAGGAGAAGACCTGCCTGTG.

Specific gene expression was quantitated by using the 2-ΔΔCT method on data normalized by using β-actin as refer-

ence gene.¹⁴ Results were expressed as ratio to SK-Mel-37 reference cell line.

Epitope mapping

The FLITrx random peptide library (Invitrogen, Basel, Switzerland) composed of 1.77×10^8 primary clones of *E. coli* with a dodecamer peptide sequence inserted within the Thioredoxin (TrxA) active-site loop was used to identify epitopes recognized by MAGE-A10-specific mAb, as previously detailed.¹⁵ Briefly, 2 ml of the FLITrx library was induced in the presence of IMC medium containing ampicillin and 100 µg/ml tryptophan at 25°C. Bacteria were then "panned" on 60-mm tissue culture plates (Becton Dickinson, Franklin Lakes, NJ) pretreated with the mAb under investigation (50 µg/ml in sterile water). After the removal of unbound cells, bacteria were amplified o/n in IMC medium at 25°C. After five rounds of panning and culture, *E. coli* were streaked onto RMG medium plates containing 100 µg/ml ampicillin and incubated overnight at 30°C. Individual clones tested positive in Western blot⁷ were selected and amplified o/n at 30°C in RM medium containing ampicillin. DNA was isolated using the NucleoSpin kit (Macherey-Nagel, Oensingen, Switzerland), and plasmids were sequenced by using the FLITrx forward or Rsr reverse sequencing primers.

Tissue microarrays and immunohistochemistry

Tissue microarray (TMA) construction and staining were described in detail elsewhere.¹⁵ Briefly, formalin-fixed and paraffin-embedded tissues were obtained from the archives of the Institute of Pathology at the University of Basel. The multitissue TMA used in our study comprised a total of 2,587 samples including 218 normal tissues, 518 benign tumors and 1,851 malignancies from more than 100 different tumoral tissues. Slides were independently scored by three members of the team, including two experienced pathologists (LTe and LTo). Samples were considered positive if at least 5% of cells showed evidence of staining of moderate or strong intensity.

Results

Production of recombinant MAGE-A10

MAGE-A10 complete open reading frame was PCR amplified from SK-Mel-37 cell line¹⁶ cDNA and cloned into the pET 32a-inducible expression vector. Upon IPTG treatment, the TAA was produced in soluble form within a fusion protein, inclusive of thioredoxin and a polyhistidine tail. After purification on nickel columns, this protein was detectable in Coomassie blue-stained gels with an apparent molecular weight (MW) of 79 kDa (Fig. 1a).⁹ This material was used to immunize mice and screen hybridomas. Similarly produced soluble thioredoxin (TrxA) served as negative control in screening procedures.

Generation of MAGE-A10-specific mAbs recognizing recombinant and native gene products

A number of mAbs (>20) appeared to recognize recombinant MAGE-A10 protein in ELISA assays (data not shown). To verify their capacity to identify the native protein, they were

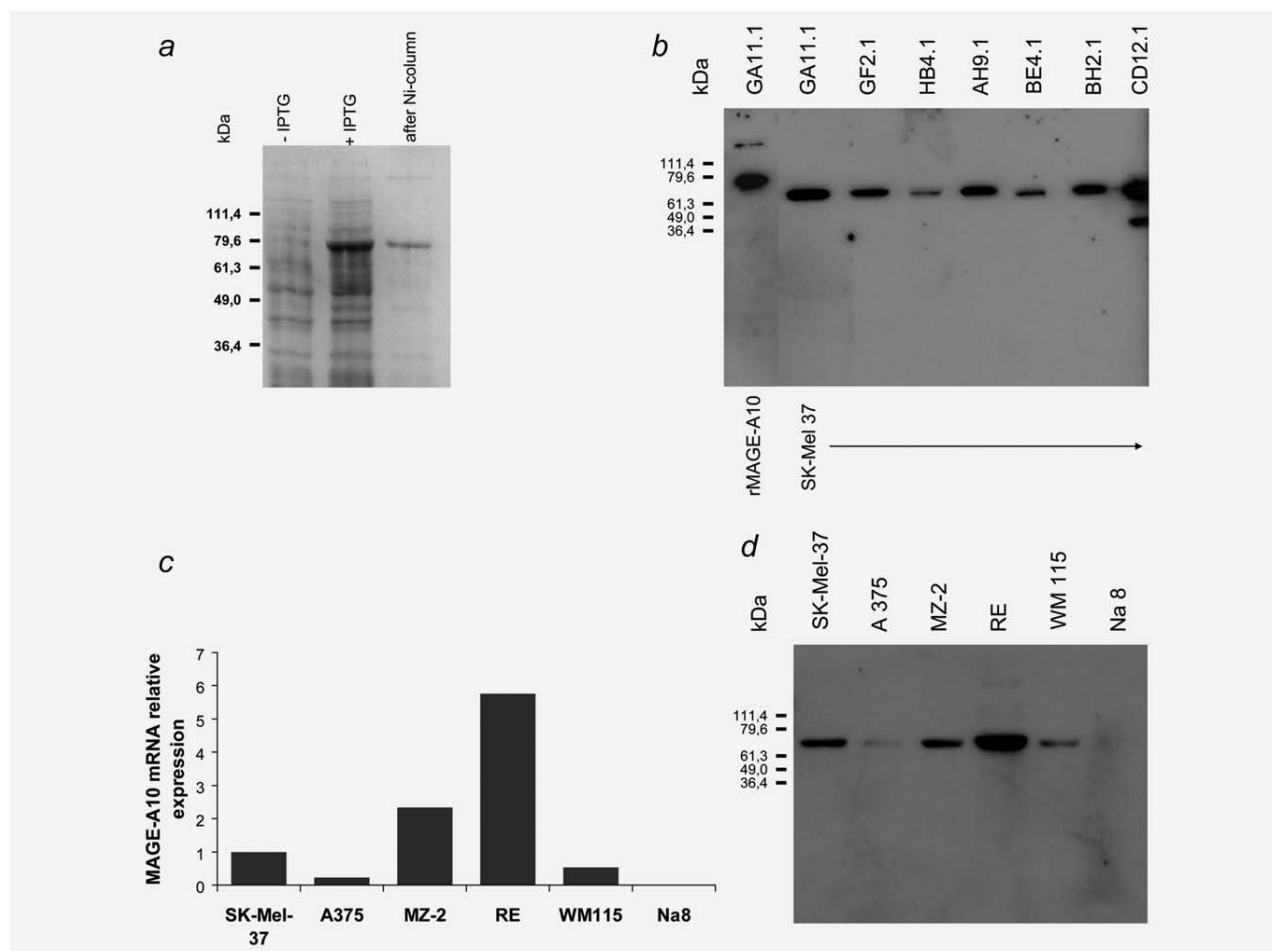


Figure 1. Generation and characterization of MAGE-A10-specific monoclonal antibodies. MAGE-A10 full gene was cloned from SK-Mel-37 cell line and inserted in an IPTG-inducible vector encoding it together with thioredoxin and a polyhistidine tail. IPTG-induced transgene products could be purified on Ni columns (*a*). Supernatants of selected hybridomas were tested in “Western” blot assays on recombinant MAGE-A10 protein preparation (lane 1) and on SK-MEL-37 lysates. Although six supernatants recognize a single 70-kDa protein, CD12.1 hybridoma supernatant also identifies an additional lower MW band (*b*). Total cellular RNA was extracted from the indicated cell lines, reverse transcribed and tested in quantitative real-time PCR assays in the presence of MAGE-A10-specific primers and probes. Data are expressed as fold increase by using SK-MEL-37 cell line as reference (*c*). Panel *d* reports a “Western” blot assay performed by using lysates of the indicated cell lines and GA11.1 hybridoma supernatant.

tested in “Western” blot assays on lysates from SK-Mel-37 cell line, originally utilized for the cloning of the MAGE-A10 gene (see above). Although all reagents recognized the positive control 79-kDa recombinant fusion protein, notably, 15 of 20 mAbs identified a single band of an apparent MW of 70 kDa in SK-Mel-37 lysates. Other mAbs, however, identified extra bands of an apparent 50-kDa MW, suggesting that they might recognize target epitopes shared with additional proteins and, possibly, with other members of the MAGE-A family. Figure 1*b* reports representative examples of the reactivities observed.

To obtain additional evidence of the specificity of the reagents under investigation, SK-Mel-37, RE, WM115, MZ-2, A375 and Na8 melanoma cell lines were tested for MAGE-A10 expression

by qRT-PCR. Expectably, all lines expressed β -actin house keeping gene, whereas MAGE-A10 was expressed to decreasing extents in RE, MZ-2-37, SK-Mel, WM115 and A375 but not in Na8 cell line (Fig. 1*c*). Lysates from all cell lines, equalized in total protein content, were then tested in “Western” blot assays in the presence of the mAbs under investigation. Recognition of the 70-kDa band in cell lysates closely reflected MAGE-A10 gene expression, as detected in the corresponding RNA preparations. A representative blot is reported in Figure 1*d*.

Specificity assessment and epitope mapping

TAA of the MAGE family are characterized by high sequence homology. As a result, mAbs raised by using one TAA as immunogen are frequently characterized by extensive

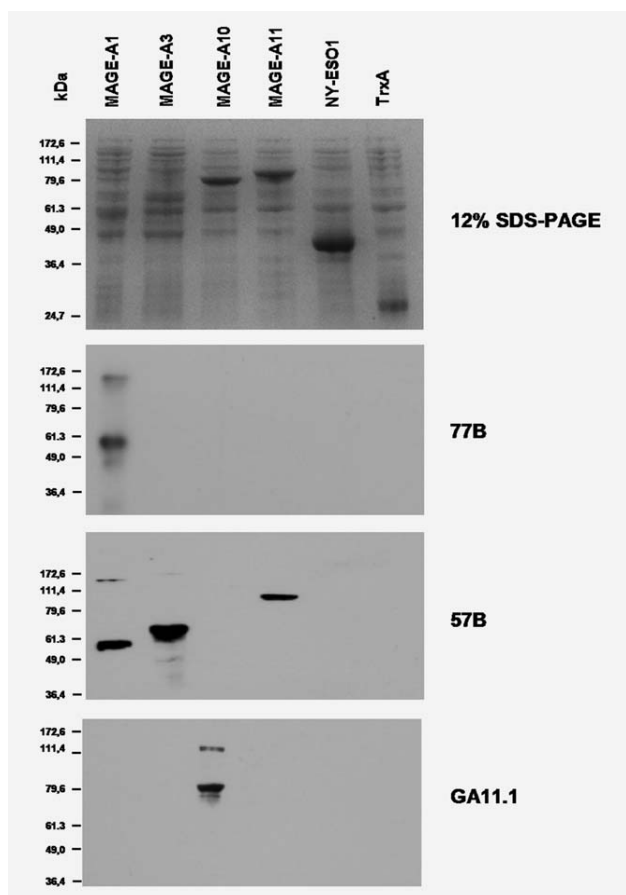


Figure 2. Specificity of anti-MAGE-A monoclonal antibodies. *E. coli* cultures were transformed with expression vectors encoding the indicated proteins, IPTG induced and lysed, as described in "Material and Methods." Equal amounts of proteins were run on a 12% SDS-PAGE gel and blotted. Blots were used in "Western" assays in the presence of 77B, 57B or GA11.1 mAbs generated by immunizing mice with recombinant MAGE-A1, MAGE-A3 and MAGE-A10, respectively.

cross reactivities with other MAGE TAA.^{3,9,17} To unambiguously assess the specificity of the mAbs under investigation, we tested them in "Western" blot assays, by using as target recombinant proteins, lysates of *E. coli* cultures transformed with plasmids encoding different MAGE-A or unrelated proteins, following IPTG induction (Fig. 2). Although 77B mAb only recognized its MAGE-A1 putative target protein, 57B mAb recognized, in addition to its putative MAGE-A3 target, also MAGE-A1 and MAGE-A11.^{13,18} Neither reagent was able to identify MAGE-A10. In contrast, a number of antibodies produced during our study only identified MAGE-A10, as shown, for GA11.1 representative mAb in Figure 2. In no case, evidence of recognition of NY-ESO-1 or thioredoxin (TrxA) could be observed.

Capitalizing on these data, we mapped the specific epitope recognized by GA11.1 IgG1 mAb by using a random peptide

Table 1. Mapping of the epitopes recognized by GA11.1 monoclonal antibody on MAGE-A10 sequence

MAGE-A10	A T G S F S Y P E
GA11-3	G G P H T V S Y P E T
GA11-5	A D N I I S Y P V S G P
GA11-8	S F S Y P S T G A Q T G
GA11-15	L G R S Y P E S E G S Q
GA11-4	F P A G R S Y P N T E A
GA11-12	D A G P S Y P D A T M D

MAGE-A10 sequence:

```
MPRAPKRQRCMPEEDLQSQSETQGLEGAQAPLAVEEDASSSTSTSSSPFPSS
SSSSSSCYPLIPSTPEEVSADETPNPPQSAQIACSSPSVVASLPLDQSDGSS
SQKEESPSTLQVLPDSESLPRSEIDEKVDLVQFLFKYQMKPEITKAEILSVIK
NYEDHFPLLFSEASECMLLVFGIDVKEVDPTGHSFVLVLSGLTYDGMLSDVQSM
KTGILILISIIIEGYCTPEEVIWEALNMMGLYDGMHELIYGEPRKLLTQDWVQENY
EYRQVPGSDPARYEFLWGPRAHAEIRKMSLLKFLAKVNGSDPRSFPLWYEEALKDE
EERAQDRIATDDTTAMASASSSATGSFSYPE.
```

library. Clones from six single colonies expanded following repeated panning on the mAb under investigation were sequenced (Table 1). Notably, all of them expressed the (F)SYPE motif, detectable at the COOH end of the MAGE-A10 protein. Most importantly, this motif is not present in a number of additional members of the MAGE family, including MAGE-B1,-B2,-B6 and MAGE-A1,-2,-3,-4,-5,-6,-9 and -12.

MAGE-A10 is a nuclear protein

Considering its unambiguous specificity, supported by epitope mapping data, we used GA11.1 mAb to identify the intracellular location of MAGE-A10 TAA. Testis sections stained with this mAb showed that the target antigen in healthy testis is exquisitely expressed in spermatogonia and spermatocyte nuclei (Fig. 3a), but, at difference with other C/T TAA, e.g., NY-ESO-1,^{19,20} it is largely undetectable in cell cytoplasm. No other healthy tissue scored positive upon GA11.1 mAb staining (see below).

MAGE-A10 protein expression in cancers

Expression of MAGE-A10 gene has been detected in different types of cancer.^{8,21-29} However, because of the lack of specific reagents, no comprehensive analysis at the protein level on large tumor databases was possible. We used the anti-MAGE-A10 mAbs described above to stain a multitumor TMA, including cancers, benign tumors and corresponding normal tissues for a total of 2,587 samples (Table 2).

Expression in benign tumors was extremely rare ($n = 6/518$, 1.1% of cases). Most interestingly, it was detectable in two cases of moderately dysplastic adenomas of the colon and in benign tumors with remarkable malignant transformation potential, such as paraganglioma ($n = 2$) and a mixed tumor of the salivary glands ($n = 1$).

In contrast, positive staining, limited to tumor cell nuclei, was detectable in 215 of 1,851 cancers (11.6%). For 43 different malignancies, representative numbers of cases ($n = 20$) were available in the TMA under investigation. In 18 of these

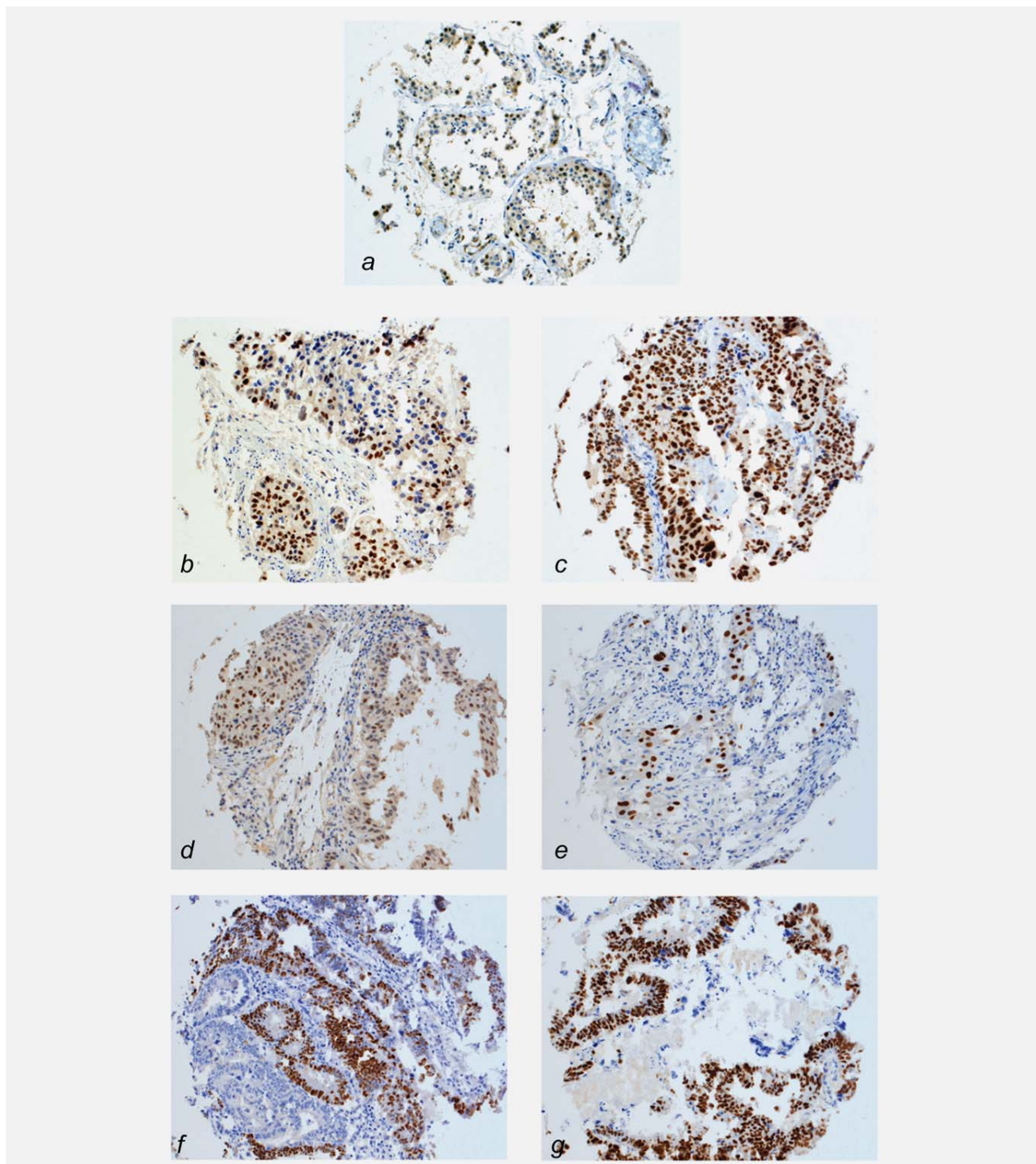


Figure 3. Detection of MAGE-A10 cancer/testis tumor-associated antigens in normal testis and in malignancies of different histological origin. A multitumor tissue microarray from paraffin-embedded sections was stained with GA11.1 hybridoma supernatant according to standard methods. Representative examples of positive stainings of different extents are shown. They include: (a) healthy testis, (b, c) urothelial carcinomas, (d, e) larynx carcinomas, (f, g) colon adenocarcinomas, (h, i) serous carcinomas of the endometrium, (j, k) squamous cell lung carcinoma, (l, m) ovarian serous carcinomas and (n, o) seminomas.

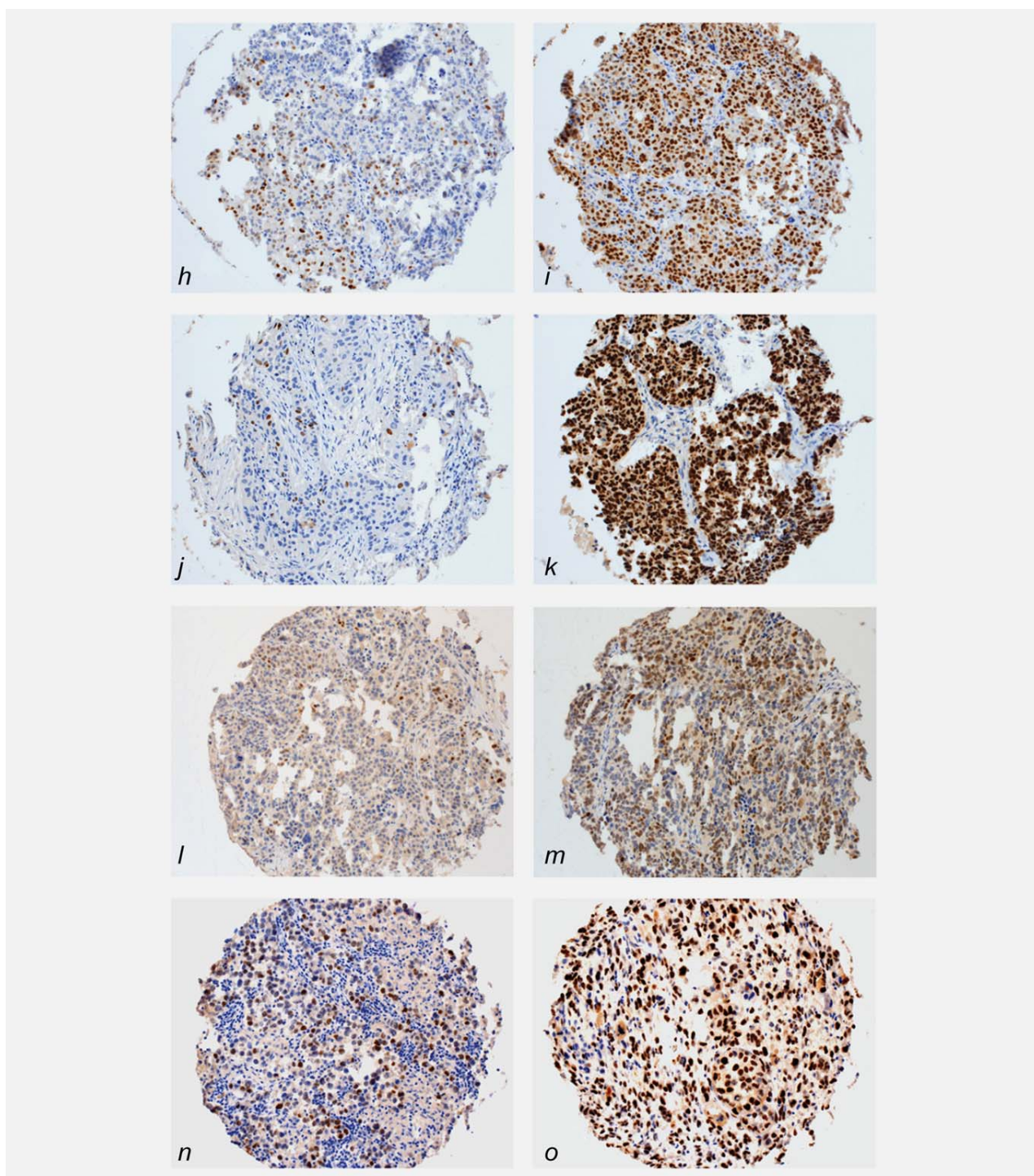


Figure 3. Continued

cancer types, MAGE-A10-specific staining was detectable in >10% of cases, with average percentages of positive malignant cells ranging between 20 (mesothelioma and seminoma) and >60% (lung and bladder cancers, laryngeal squamous cell carcinoma).

In particular, among gynecological cancers,^{30,31} breast tumors of the medullary subtype and endometrioid adenocarcinomas displayed similar levels of positivity (>10% of cases, with >40% positive tumor cells). Ovarian and endometrium serous adenocarcinomas expressed MAGE-A10 more frequently, as

Table 2. Distribution of MAGE-A10-specific staining (GA11.1 mAb) across organs and tissue types (*n* = 2,587)

Organ	Tissue group	Total number of cases	Negative expression (<5%)	Positive expression (>5%)	Average percentage of positive cells
Adrenal gland	Cortical adenoma	12	12 (100)		
	Pheochromocytoma	26	26 (100)		
Brain	Meningioma	35	35 (100)		
	Glioblastoma multiforme	29	29 (100)		
	Oligodendroglioma	13	13 (100)		
	Normal	7	7 (100)		
Breast	Ductal cancer	25	25 (100)		
	Lobular cancer	29	29 (100)		
	Medullary cancer	49	44 (89.8)	5 (10.2)	45
	Tubular cancer	13	13 (100)		
	Mucinous cancer	15	15 (100)		
	Normal	3	3 (100)		
Colon	Adenoma, mild dysplasia	27	27 (100)		
	Adenoma, moderate dysplasia	51	49 (96.1)	2 (3.9)	22.5
	Adenoma, severe dysplasia	23	23 (100)		
	Adenocarcinoma	46	42 (91.3)	4 (8.7)	38.8
	Normal	1	1 (100)		
Endometrium	Endometrioid adenocarcinoma	49	44 (89.8)	5 (10.2)	45
	Serous adenocarcinoma	32	22 (68.8)	10 (31.3)	57
	Normal	13	13 (100)		
Esophagus	Adenocarcinoma	6	6 (100)		
	Squamous cell carcinoma	31	23 (74.2)	8 (25.8)	25
	Small cell carcinoma	1	1 (100)		
	Normal	7	7 (100)		
Gall bladder	Adenocarcinoma	34	29 (85.3)	5 (14.7)	44
	Normal	3	3 (100)		
Kidney	Clear cell renal cell carcinoma	50	50 (100)		
	Papillary renal cell carcinoma	25	25 (100)		
	Chromophobe renal cell carcinoma	10	9 (90.0)	1 (10.0)	85
	Oncocytoma	19	19 (100)		
	Normal	16	16 (100)		
Larynx	Squamous cell carcinoma	24	19 (79.2)	5 (20.8)	61
Liver	Hepatocellular carcinoma	82	78 (95.1)	4 (4.9)	13.8
	Normal	15	15 (100)		
Lung	Squamous cell carcinoma	43	28 (65.1)	15 (34.9)	65.3
	Adenocarcinoma	75	66 (88.0)	9 (12.0)	56.1
	Large cell carcinoma	20	17 (85.0)	3 (15.0)	63.3
	Small cell carcinoma	44	38 (86.4)	6 (13.4)	60
	Bronchioloalveolar adenocarcinoma	6	6 (100)		
	Normal	11	11 (100)		
Lymphatic tissue	Diffuse large B-cell lymphoma	10	10 (100)		
	Non-Hodgkin lymphoma, others	52	52 (100)		
	Mixed cellularity Hodgkin lymphoma	14	14 (100)		

Table 2. Distribution of MAGE-A10-specific staining (GA11.1 mAb) across organs and tissue types (*n* = 2,587) (Continued)

Organ	Tissue group	Total number of cases	Negative expression (<5%)	Positive expression (>5%)	Average percentage of positive cells
	Nodular sclerosis Hodgkin lymphoma	23	22 (95.7)	1 (4.4)	30
	Normal	13	13 (100)		
Myometrium	Leiomyoma	47	47 (100)		
	Normal	9	9 (100)		
Neuroendocrine tissue	Extra-adrenal paraganglioma	8	6 (75.0)	2 (25.0)	15
	Typical carcinoid of the lung	33	33 (100)		
Oral cavity	Squamous cell carcinoma	42	38 (90.5)	4 (9.5)	46.3
	Normal	5	5 (100)		
Ovary	Serous adenocarcinoma	40	32 (80.0)	8 (20.0)	36.3
	Mucinous adenocarcinoma	15	15 (100)		
	Endometrioid adenocarcinoma	42	41 (97.6)	1 (2.4)	25
Pancreas	Adenocarcinoma	48	48 (100)		
	Normal	9	9 (100)		
Parathyroid	Adenoma	31	31 (100)		
	Normal	2	2 (100)		
Peripheral nerves	Neurofibroma	17	17 (100)		
	Schwannoma	23	23 (100)		
Pleura	Mesothelioma	28	23 (82.1)	5 (17.9)	21
Prostate	Adenocarcinoma, hormone-refractory	29	29 (100)		
	Adenocarcinoma, untreated	48	48 (100)		
	Normal	17	17 (100)		
Salivary gland	Warthin tumor	18	18 (100)		
	Pleomorphic adenoma	39	38 (97.4)	1 (2.6)	65
	Adenoid cystic carcinoma	24	24 (100)		
	Normal	13	13 (100)		
Skin	Basalioma	59	40 (67.8)	19 (32.2)	50.5
	Squamous cell carcinoma	33	31 (93.9)	2 (6.1)	60
	Appendageal tumors (benign)	11	11 (100)		
	Malignant melanoma	50	31 (62.0)	19 (38.0)	45.8
	Benign nevus	12	12 (100)		
	Fibrous histiocytoma	12	11 (91.7)	1 (8.3)	10
	Kapillary hemangioma	22	22 (100)		
	Kaposi sarcoma	15	15 (100)		
	Normal	6	6 (100)		
Small intestine	Adenocarcinoma	22	22 (100)		
	Normal	4	4 (100)		
Soft tissue	Liposarcoma	22	22 (100)		
	Malignant fibrous histiocytoma	26	24 (92.3)	2 (7.8)	62.5
	Leiomyosarcoma	45	43 (95.6)	2 (4.4)	52.5
	Tendon sheath, giant cell tumor	14	14 (100)		
	Normal (skeletal muscle)	16	16 (100)		
	Normal (smooth muscle)	5	5 (100)		

Table 2. Distribution of MAGE-A10-specific staining (GA11.1 mAb) across organs and tissue types (*n* = 2,587) (Continued)

Organ	Tissue group	Total number of cases	Negative expression (<5%)	Positive expression (>5%)	Average percentage of positive cells
Stomach	Adenocarcinoma, diffuse type	25	25 (100)		
	Adenocarcinoma, intestinal type	49	38 (77.6)	11 (22.5)	36.8
	Normal	8	8 (100)		
Testis	Seminoma	47	32 (68.1)	15 (31.9)	22.7
	Nonseminomatous germ cell tumors	41	39 (95.1)	2 (4.9)	50
	Normal	17	4 (23.5)	13 (76.5)	24.3
Thymus	Thymoma	37	37 (100)		
	Normal	3	3 (100)		
Thyroid	Follicular adenoma	36	36 (100)		
	Follicular carcinoma	51	51 (100)		
	Papillary carcinoma	19	17 (89.5)	2 (10.5)	17.5
	Normal	12	12 (100)		
Urinary bladder	Noninvasive urothelial carcinoma	36	29 (80.6)	7 (19.4)	64.3
	Infiltrating urothelial carcinoma	67	46 (68.7)	21 (31.3)	68.6
	Normal	3	3 (100)		
Uterus, cervix	Intraepithelial neoplasia, high-grade (CIN III)	17	17 (100)		
Vulva	Squamous cell carcinoma	26	25 (96.2)	1 (3.9)	90

Samples were considered positive when >5% of the indicated cell type showed evidence of positive staining with MAGE-A10-specific GA11.1 mAb. Average percentages of positive cells within individual samples are also reported. Shaded areas refer to types of tissues where a representative (≥ 20) number of cases was available and at least 10% of them showed evidence of MAGE-A10-specific staining.

detected by immunostaining (30 and 20% of cases, with >50 and >30% positive tumor cells, respectively).

MAGE-A10-specific staining was detectable in squamous cell carcinomas of the esophagus²⁴ (8 of 31 cases in 25% of cells on an average) and of the larynx (5 of 24 cases in >60% of tumor cells on an average).

Notably, MAGE-A10 protein expression appears to be particularly frequent in lung cancers. In more than 34% of squamous cell carcinomas and 12, 15 and 13% of adenocarcinomas, large and small cell carcinomas, respectively, more than 50% of tumor cells do express MAGE-A10 protein. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb.

In keeping with previously published gene expression data,²⁹ MAGE-A10 protein expression was also observed in stomach cancers of the "intestinal" type (11 of 49 cases in >35% of tumor cells on an average). Similarly, as expected, MAGE-A10 C/T TAA was also detectable in seminoma specimens.^{19,20}

A sizeable fraction of bladder cancers of noninvasive (7 of 36) or infiltrating (21 of 67) type were found to express

MAGE-10 in >60% of tumor cells. Unexpectedly, high (38–40%) percentages of neoplastic cells were stained by GA11.1 mAb in sizeable percentages of colorectal (8.7%) and gall bladder (14%) cancers.

All in all, the highest percentages of MAGE-A10-positive cells were detectable in lung, skin and urothelial cancers. Figures 3b–3o report representative examples of MAGE-A10-specific staining in different types of tumors from the TMA under investigation.

Discussion

MAGE-A10 ranks among the most antigenic members of the C/T TAA family, together with NY-ESO-1. Specific cytotoxic T lymphocytes (CTLs) have been generated from lymphocytes of untreated patients bearing melanoma,^{8,32} hepatocellular carcinoma,⁴ non-small cell lung cancers⁵ and healthy donors.³³ Furthermore, vaccination has been found to promote specific CTL responsiveness.³⁴ A mAb generated by using recombinant MAGE-1 as immunogen^{9,11} and recognizing MAGE-A10 together with other members of the MAGE-A family has helped to preliminarily characterize a specific gene product and its putative intracellular location. However, the identification of MAGE-A10-positive tumor cells in clinical specimens, a key prerequisite for the development of specific immunotherapy strategies and for their monitoring, has suffered from the absence of mAb of exquisite specificity.

Here, we report the generation and the characterization of MAGE-A10-specific mAbs. One of these reagents, recognizing a discrete MAGE-A10-specific epitope, has been used to stain a multitissue array comprising more than 2,000 specimens and including malignancies and benign tumors of different histological origin together with corresponding healthy tissues.

This reagent has helped to reveal similarities and differences between MAGE-A10 and the other members of the MAGE-A family. First, this TAA is exclusively intranuclear. In this context, its relatively high immunogenicity might be related to the release of nuclear proteins following necrotic cell death, detectable in specific areas, in different types of tumors.³⁵ Second, it identifies relatively large percentages of tumor cells expressing the target proteins in cancers known to express C/T TAA, including squamous cell carcinoma of different histological origin, melanoma, lung and gynecological malignancies and bladder cancers. Third, MAGE-A10 protein, at difference with other C/T TAA, appears to be of limited expression in soft tissue cancers. Fourth, in contrast, it appears to be relatively frequently detectable in gall bladder tumors and, interestingly, in a sizeable subgroup of colorectal cancers.

Considering the high potential immunogenicity of MAGE-A10, the reagents described in our study might prove of critical relevance in the development of targeted active, specific immunotherapies by helping to establish clear indications for treatment and in the monitoring of their impact on tumor progression. Furthermore, as the prognostic relevance of C/T TAA expression in tumors of diverse histological ori-

gin is emerging,^{22,36-39} the availability of these mAbs might set the stage for investigations addressing the role of MAGE-A10 expression in individual tumor entities.

On the other hand, the biological function of the proteins of the MAGE-A family remains largely elusive. MAGE-A gene expression has been shown to suppress p53-dependent apoptosis and to promote the "in vitro" and "in vivo" viability of mast cell lines.^{40,41} Previous studies in a thyroid carcinoma model have suggested that selected MAGE-A genes might control fibronectin-mediated tumor progression.⁴² Notably, MAGE-A11, also characterized by a putatively specific intranuclear location, has been shown to coregulate androgen receptor-mediated transcriptional activity.^{43,44}

Most interestingly, MAGE-A gene expression has been found to be correlated with genome-wide demethylation,⁴⁵ frequently representing an early event during carcinogenesis, associated with hypermethylation of defined tumor suppressor genes.⁴⁶ Strikingly, a number of tumors showing evidence of high MAGE-A10 expression in our study, including lung,⁴⁶ skin,⁴⁷ bladder⁴⁸ and head and neck⁴⁹ cancers, are also known to be characterized by a frequent genome-wide hypomethylation. Also, considering the specific nuclear location of MAGE-A10, further studies are warranted to comparatively address its expression together with the DNA methylation status of repetitive elements, including long interspersed nuclear elements (LINE-1).⁵⁰

Acknowledgements

Our study was partly supported by SNF grants to G.C.S. and L.Te.

References

1. Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol Rev* 2002;188: 22-32.
2. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* 2005;5:615-25.
3. Juretic A, Spagnoli GC, Schultz-Thater E, Sarcevic B. Cancer/testis tumour-associated antigens: immunohistochemical detection with monoclonal antibodies. *Lancet Oncol* 2003;4:104-9.
4. Bricard G, Bouzourene H, Martinet O, Rimoldi D, Halkic N, Gillet M, Chaubert P, Macdonald HR, Romero P, Cerottini JC, Speiser DE. Naturally acquired MAGE-A10- and SSX-2-specific CD8+ T cell responses in patients with hepatocellular carcinoma. *J Immunol* 2005;174:1709-16.
5. Groeper C, Gambazzi F, Zajac P, Bubendorf L, Adamina M, Rosenthal R, Zerkowski HR, Heberer M, Spagnoli GC. Cancer/testis antigen expression and specific cytotoxic T lymphocyte responses in non small cell lung cancer. *Int J Cancer* 2007;120:337-43.
6. Le Gal FA, Widmer VM, Dutoit V, Rubio-Godoy V, Schrenzel J, Walker PR, Romero PJ, Valmori D, Speiser DE, Dietrich PY. Tissue homing and persistence of defined antigen-specific CD8+ tumor-reactive T-cell clones in long-term melanoma survivors. *J Invest Dermatol* 2007;127: 622-9.
7. Sartorius R, Pisu P, D'Apice L, Pizzella L, Romano C, Cortese G, Giorgini A, Santoni A, Velotti F, De BP. The use of filamentous bacteriophage fd to deliver MAGE-A10 or MAGE-A3 HLA-A2-restricted peptides and to induce strong antitumor CTL responses. *J Immunol* 2008; 180:3719-28.
8. Valmori D, Dutoit V, Rubio-Godoy V, Chambaz C, Lienard D, Guillaume P, Romero P, Cerottini JC, Rimoldi D. Frequent cytolytic T-cell responses to peptide MAGE-A10(254-262) in melanoma. *Cancer Res* 2001; 61:509-12.
9. Carrel S, Schreyer M, Spagnoli G, Cerottini JC, Rimoldi D. Monoclonal antibodies against recombinant-MAGE-1 protein identify a cross-reacting 72-kDa antigen which is co-expressed with MAGE-1 protein in melanoma cells. *Int J Cancer* 1996;67:417-22.
10. Gervois N, Guilloux Y, Diez E, Jotereau F. Suboptimal activation of melanoma infiltrating lymphocytes (TIL) due to low avidity of TCR/MHC-tumor peptide interactions. *J Exp Med* 1996;183:2403-7.
11. Rimoldi D, Salvi S, Reed D, Coulie P, Jongeneel VC, De Plaen E, Brasseur F, Rodriguez AM, Boon T, Cerottini JC. cDNA and protein characterization of human MAGE-10. *Int J Cancer* 1999;82:901-7.
12. Schultz-Thater E, Noppen C, Gudat F, Durmuller U, Zajac P, Kocher T, Heberer M, Spagnoli GC. NY-ESO-1 tumour associated antigen is a cytoplasmic protein detectable by specific monoclonal antibodies in cell lines and clinical specimens. *Br J Cancer* 2000;83:204-8.
13. Kocher T, Schultz-Thater E, Gudat F, Schaefer C, Casorati G, Juretic A, Willmann T, Harder F, Heberer M, Spagnoli GC. Identification and intracellular location of MAGE-3 gene product. *Cancer Res* 1995;55:2236-9.
14. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 2008;3:1101-8.

15. Bolli M, Schultz-Thater E, Zajac P, Guller U, Feder C, Sanguedolce F, Carafa V, Terracciano L, Hudolin T, Spagnoli GC, Tornillo L. NY-ESO-1/LAGE-1 coexpression with MAGE-A cancer/testis antigens: a tissue microarray study. *Int J Cancer* 2005;115:960–6.
16. Gure AO, Stockert E, Arden KC, Boyer AD, Viars CS, Scanlan MJ, Old LJ, Chen YT. CT10: a new cancer-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational-difference analysis. *Int J Cancer* 2000;85:726–32.
17. Landry C, Brasseur F, Spagnoli GC, Marbaix E, Boon T, Coulie P, Godelaine D. Monoclonal antibody 57B stains tumor tissues that express gene MAGE-A4. *Int J Cancer* 2000;86:835–41.
18. Schultz-Thater E, Juretic A, Dellabona P, Luscher U, Siegrist W, Harder F, Heberer M, Zuber M, Spagnoli GC. MAGE-1 gene product is a cytoplasmic protein. *Int J Cancer* 1994;59:435–9.
19. Aubry F, Satie AP, Rioux-Leclercq N, Rajpert-De Meyts E, Spagnoli GC, Chomez P, De Backer O, Jegou B, Samson M. MAGE-A4, a germ cell specific marker, is expressed differentially in testicular tumors. *Cancer* 2001;92:2778–85.
20. Satie AP, Rajpert-De Meyts E, Spagnoli GC, Henno S, Olivo L, Jacobsen GK, Rioux-Leclercq N, Jegou B, Samson M. The cancer-testis gene, NY-ESO-1, is expressed in normal fetal and adult testes and in spermatocytic seminomas and testicular carcinoma in situ. *Lab Invest* 2002;82:775–80.
21. Alves PM, Levy N, Bouzourene H, Viatte S, Bricard G, Ayyoub M, Vuilleumier H, Givel JC, Halkic N, Speiser DE, Romero P, Levy F. Molecular and immunological evaluation of the expression of cancer/testis gene products in human colorectal cancer. *Cancer Immunol Immunother* 2007;56:839–47.
22. Gure AO, Chua R, Williamson B, Gonen M, Ferrera CA, Gnjjatic S, Ritter G, Simpson AJ, Chen YT, Old LJ, Altorki NK. Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer. *Clin Cancer Res* 2005;11:8055–62.
23. Jacobs JF, Grauer OM, Brasseur F, Hoogerbrugge PM, Wesseling P, Gidding CE, van de Rakt MW, Figdor CG, Coulie PG, de Vries JJ, Adema GJ. Selective cancer-germline gene expression in pediatric brain tumors. *J Neurooncol* 2008;88:273–80.
24. Lin J, Lin L, Thomas DG, Greenson JK, Giordano TJ, Robinson GS, Barve RA, Weishaar FA, Taylor JM, Orringer MB, Beer DG. Melanoma-associated antigens in esophageal adenocarcinoma: identification of novel MAGE-A10 splice variants. *Clin Cancer Res* 2004;10:5708–16.
25. Muller-Richter UD, Dowejko A, Reuther T, Kleinheinz J, Reichert TE, Driemel O. Analysis of expression profiles of MAGE-A antigens in oral squamous cell carcinoma cell lines. *Head Face Med* 2009;5:10.
26. Peng JR, Chen HS, Mou DC, Cao J, Cong X, Qin LL, Wei L, Leng XS, Wang Y, Chen WF. Expression of cancer/testis (CT) antigens in Chinese hepatocellular carcinoma and its correlation with clinical parameters. *Cancer Lett* 2005;219:223–32.
27. Sharma P, Shen Y, Wen S, Bajorin DF, Reuter VE, Old LJ, Jungbluth AA. Cancer-testis antigens: expression and correlation with survival in human urothelial carcinoma. *Clin Cancer Res* 2006;12:5442–7.
28. Sugita M, Geraci M, Gao B, Powell RL, Hirsch FR, Johnson G, Lapadat R, Gabrielson E, Bremnes R, Bunn PA, Franklin WA. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. *Cancer Res* 2002;62:3971–9.
29. Suzuki S, Sasajima K, Sato Y, Watanabe H, Matsutani T, Iida S, Hosone M, Tsukui T, Maeda S, Shimizu K, Tajiri T. MAGE-A protein and MAGE-A10 gene expressions in liver metastasis in patients with stomach cancer. *Br J Cancer* 2008;99:350–6.
30. Kavalar R, Sarcevic B, Spagnoli GC, Separovic V, Samija M, Terracciano L, Heberer M, Juretic A. Expression of MAGE tumour-associated antigens is inversely correlated with tumour differentiation in invasive ductal breast cancers: an immunohistochemical study. *Virchows Arch* 2001;439:127–31.
31. Yakirevich E, Sabo E, Lavie O, Mazareb S, Spagnoli GC, Resnick MB. Expression of the MAGE-A4 and NY-ESO-1 cancer-testis antigens in serous ovarian neoplasms. *Clin Cancer Res* 2003;9:6453–60.
32. Huang LQ, Brasseur F, Serrano A, De Plaen E, van der Bruggen P, Boon T, Van Pel A. Cytolytic T lymphocytes recognize an antigen encoded by MAGE-A10 on a human melanoma. *J Immunol* 1999;162:6849–54.
33. Bracci L, Schumacher R, Provenzano M, Adamina M, Rosenthal R, Groeper C, Zajac P, Iezzi G, Proietti E, Belardelli F, Spagnoli GC. Efficient stimulation of T cell responses by human IFN- α -induced dendritic cells does not require Toll-like receptor triggering. *J Immunother* 2008;31:466–74.
34. Chianese-Bullock KA, Pressley J, Garbee C, Hibbitts S, Murphy C, Yamshchikov G, Petroni GR, Bissonette EA, Neese PY, Grosh WW, Merrill P, Fink R, et al. MAGE-A1-, MAGE-A10-, and gp100-derived peptides are immunogenic when combined with granulocyte-macrophage colony-stimulating factor and montanide ISA-51 adjuvant and administered as part of a multipptide vaccine for melanoma. *J Immunol* 2005;174:3080–6.
35. Tesniere A, Apetoh L, Ghiringhelli F, Joza N, Panaretakis T, Kepp O, Schlemmer F, Zitvogel L, Kroemer G. Immunogenic cancer cell death: a key-lock paradigm. *Curr Opin Immunol* 2008;20:504–11.
36. Bergeron A, Picard V, LaRue H, Harel F, Hovington H, Lacombe L, Fradet Y. High frequency of MAGE-A4 and MAGE-A9 expression in high-risk bladder cancer. *Int J Cancer* 2009;125:1365–71.
37. Bolli M, Kocher T, Adamina M, Guller U, Dalquen P, Haas P, Mirlacher M, Gambazzi F, Harder F, Heberer M, Sauter G, Spagnoli GC. Tissue microarray evaluation of Melanoma antigen E (MAGE) tumor-associated antigen expression: potential indications for specific immunotherapy and prognostic relevance in squamous cell lung carcinoma. *Ann Surg* 2002;236:785–93.
38. Kocher T, Zheng M, Bolli M, Simon R, Forster T, Schultz-Thater E, Rimmel E, Noppen C, Schmid U, Ackermann D, Mihatsch MJ, Gasser T, et al. Prognostic relevance of MAGE-A4 tumor antigen expression in transitional cell carcinoma of the urinary bladder: a tissue microarray study. *Int J Cancer* 2002;100:702–5.
39. Sielen W, Mecklenburg I, Dango S, Ehrhardt P, Kirschbaum A, Passlick B, Pantel K. Detection of MAGE-A transcripts in bone marrow is an independent prognostic factor in operable non-small-cell lung cancer. *Clin Cancer Res* 2007;13:3840–7.
40. Yang B, O'Herrin SM, Wu J, Reagan-Shaw S, Ma Y, Bhat KM, Gravekamp C, Setaluri V, Peters N, Hoffmann FM, Peng H, Ivanov AV, et al. MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines. *Cancer Res* 2007;67:9954–62.
41. Yang B, O'Herrin S, Wu J, Reagan-Shaw S, Ma Y, Nihal M, Longley BJ. Select cancer testis antigens of the MAGE-A, -B, and -C families are expressed in mast cell lines and promote cell viability in vitro and in vivo. *J Invest Dermatol* 2007;127:267–75.
42. Liu W, Cheng S, Asa SL, Ezzat S. The melanoma-associated antigen A3 mediates fibronectin-controlled cancer progression and metastasis. *Cancer Res* 2008;68:8104–12.
43. Askew EB, Bai S, Hnat AT, Minges JT, Wilson EM. Melanoma antigen gene protein-A11 (MAGE-11) F-box links the androgen receptor NH2-terminal

- transactivation domain to p160 coactivators. *J Biol Chem* 2009;284:34793–808.
44. Karpf AR, Bai S, James SR, Mohler JL, Wilson EM. Increased expression of androgen receptor coregulator MAGE-11 in prostate cancer by DNA hypomethylation and cyclic AMP. *Mol Cancer Res* 2009;7:523–35.
45. De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl Acad Sci USA* 1996;93:7149–53.
46. Liu F, Killian JK, Yang M, Walker RL, Hong JA, Zhang M, Davis S, Zhang Y, Hussain M, Xi S, Rao M, Meltzer PA, et al. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene* 2010;29:3650–64.
47. van Doorn R, Gruis NA, Willemze R, van der Velden PA, Tensen CP. Aberrant DNA methylation in cutaneous malignancies. *Semin Oncol* 2005;32:479–87.
48. Wilhelm CS, Kelsey KT, Butler R, Plaza S, Gagne L, Zens MS, Andrew AS, Morris S, Nelson HH, Schned AR, Karagas MR, Marsit CJ. Implications of LINE1 methylation for bladder cancer risk in women. *Clin Cancer Res* 2010;16:1682–9.
49. Richards KL, Zhang B, Baggerly KA, Colella S, Lang JC, Schuller DE, Krahe R. Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability. *PLoS One* 2009;4:e4941.
50. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, Fuchs CS. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. *Int J Cancer* 2008;122:2767–73.

7. GENERAL DISCUSSION

7.1 3'UTR poly(T/U) tract deletions and altered expression of *EWSR1* are a hallmark of mismatch repair deficient cancers

In this study we describe a mononucleotide (T/U)₁₆ tract, EWS16T, located in the 3' UTR of the *Ewing sarcoma break point region 1 (EWSR1)* gene which discriminates MMR proficient from MMR deficient cancers with 100% sensitivity and 99.5% specificity. We demonstrate *in vitro* and *in vivo* that contractions at this locus alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in *EWSR1* pre-mRNAs and result in decreased mRNA as well as protein expression. In contrast to their proficient counterparts, MMR deficient CRC display altered subcellular localization of EWS with diffuse cytoplasmic staining. EWS16T thus not only represents a novel monomorphic MSI target locus to accurately identify both, hereditary and sporadic, MMR deficient cancers but contractions therein affect multiple regulatory mechanisms implicating the RNA-/DNA-binding protein EWS in MSI-associated colorectal tumorigenesis.

The Ewing sarcoma (EWS) protein is a member of the TET family (TLS/FUS, EWS, and TAF15) of RNA- and DNA-binding proteins, with proposed functions in transcription and RNA processing. The domain composition of TET proteins includes a transcription activation domain at the N terminus and RNA-binding domains, including three RGG boxes and one RRM motif, at the C terminus. Additional domains harbored by TET proteins include an IQ domain, which interacts with calmodulin and is phosphorylated by PKC [198], and one zinc finger motif [199]. EWS interacts with the preinitiation complex TFIID and with subunits of the RNA polymerase II (RNAPII) [200], suggesting its involvement in transcriptional regulation. EWS also interacts with splicing factors, including the U1 snRNP protein U1C, which recognizes 5' splice sites [201], the branchpoint binding protein BBP/SF1 [202], and the spliceosome component YB-1 [203, 204], suggesting a function for EWS in pre-mRNA splicing. Consistent with this potential dual role, EWS has been shown to regulate cyclin D1 transcripts both transcriptionally and at the level of splicing, with

the oncogenic fusion protein EWS-FLI1 promoting the expression of the oncogenic cyclin D1b splice variant in Ewing sarcoma cells [205]. More recently, EWS has been shown to regulate alternative splicing (AS) of the p53 repressor *MDM2* [204, 206]. In addition EWS has been described as component of the microprocessor complex that mediates the genesis of microRNAs [207].

The physiological role of EWS is largely unknown but based on its structural properties this protein is thought to be involved in diverse processes including gene expression, RNA processing / transport and cell signaling. Knockout of EWS in mice results in postnatal lethality, defects in pre-B cell development, meiotic arrest/germ cell apoptosis, premature cellular senescence, and hypersensitivity to ionizing radiation (IR) [208]. These observations suggest additional roles for EWS in homologous recombination, DNA damage response, and maintenance of genome integrity [206].

With respect to tumorigenesis, genetic alterations in *EWSR1* were first observed in Ewing sarcoma, the second most common malignant bone tumor in children [209, 210]. The *EWSR1-FLI1* fusion is the most common, being found in 85% of the cases [211]. The fusion protein retains the N-terminal transcription activation domain but loses the RNA-binding domains, which are replaced with the DNA binding domain of the fusion partner. The fusion proteins are constitutively active and have been shown to alter the transcription of several downstream targets. Ewing Sarcoma is largely thought of as a gain of function phenotype. Loss of the normal *EWSR1* function has been largely overlooked, in spite of the fact that the protein has a very canonical RNA binding domain and has been shown to regulate several RNA processing events in the nucleus [212, 213].

Our results described in details in chapter I suggest that the poly T/U tract in 3'UTR of *EWSR1* gene (EWS16T) represents a novel, quasi-monomorphic MSI target locus which identifies both, hereditary and sporadic, MMR deficient cancers with 100% sensitivity and 99.5% specificity. The contractions at this locus affect multiple regulatory mechanisms including alternative polyadenylation, mRNA / protein expression and possibly subcellular localization thereby implicating the RNA-/DNA-binding protein EWS, in MSI-associated colorectal tumorigenesis. Furthermore we demonstrated the biological effects of MSI-associated 3'UTR contractions on gene expression in vivo and in vitro for the first time.

In summary, due to only few and sketchy data concerning the role of EWS in normal

cell physiology and in CRC-related tumorigenesis, future investigations are needed. In particular to comprehensively characterize the major *EWSR1*-related downstream targets/pathways involved in cell physiology and differentiation the transcriptomes and miRNAomes of human fibroblasts and their progenitor cells, the mesenchymal stem cells should be investigated.

On the other hand, The role(s) of *EWSR1* in colorectal carcinogenesis by the generation of stable *EWSR1* overexpressing/downregulating normal and CRC cell lines, in order to investigate *EWSR1* effect on apoptosis and cell migration *in vitro* as well as its impact in tumorigenesis and its metastatic potential *in vitro* and *in vivo*.

7.2 *SH2D4A* as a novel tumor suppressor gene in CRCs

It is well accepted that the key genetic mutations underlying initiation, progression and transformation of adenomas into CRC include *APC*, *KRAS* and *TP53*, respectively [214]. Mutations of known genes are identified in 60%, 35% and 50% [215] suggesting the existence of more than just one single dominant pathway to CRC development [216].

Heterozygous loss of variable parts of chromosome 8p constitutes a frequent feature of CRC [217] and has been linked to DNA breakage at fragile sites located at 8p12 and 8p22 [218]. Tumour suppressor genes have not been localized to this critical chromosomal region, thus precluding the identification of a likely mechanism for how genetic alterations in 8p contribute to CRC development and progression.

Looking at Oncomine signature of mRNA, we noted that *SH2D4A* protein was deregulated in various cancer types but no information concerning colorectal cancer were available. We started to investigate LOH and copy number variation in an unselected cohort of 70 CRC patients. In 27 subjects the two *SH2D4A* alleles could be separated by microsatellite markers and SNPs. In 14 of these patients (52%), the primary tumour had lost or diminished *SH2D4A* expression. The rate of metastasis was significantly increased among these tumours (12/14; 86%) when compared to *SH2D4A* expressing primary CRC (3/13). Using 3 microsatellite markers and 5 SNPs, *SH2D4A* LOH was detected in 7 of the 14 tumours (50%) marked by a partial or complete lack of *SH2D4A* expression. Gene dosage quantification of the short arm of chromosome 8 revealed a monoallelic deletion in 6 and a biallelic deletion in one of these 7 tumours with LOH. Though 4 of the 6 patients with monoallelic tumours were heterozygous for the intronic SNP rs17128221 (c.342-5T>C) the T allele was selectively lost in their tumour and the C allele provided a splice donor site causing the skipping of exon 4 and a premature termination of translation in exon 5. In addition we demonstrated that *SH2D4A*, physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signalling, *SH2D4A* protein recruits the serine/threonine phosphatase PP1 β to the receptor complex and represses activated STAT3 via dephosphorylation. *SH2D4A* expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1 and Jun B.

Recently in accordance with our data Roessler *et al.* published unsupervised analyses of array comparative genomic hybridization data associated loss of chromosome 8p with poor outcome (reduced survival); somatic copy number alterations correlated with expression of 27.3% of genes analyzed. They associated expression levels of 10 of these genes with patient survival in 2 independent cohorts (comprising 319 cases of hepatocellular carcinoma (HCC) with mixed etiology) and 3 breast cancer cohorts (637 cases). Among the 10-gene signature, a cluster of 6 genes on 8p, (*DLC1*, *CCDC25*, *ELP3*, *PROSC*, *SH2D4A*, and *SORBS3*) were deleted in HCCs from patients with poor outcomes. In vitro and in vivo analyses indicated that the products of *PROSC*, *SH2D4A*, and *SORBS3* have tumor-suppressive activities, along with the known tumor suppressor gene *DLC1* [219].

Thus, *SH2D4A* could represent novel tumor suppressor gene acting in different tumor entities. Due to its interaction with STAT3 in the control of the EGFR signaling pathway, which is involved in the development and progression of several human tumors, including colorectal cancer [220], it may represent a novel promising target for CRC treatment.

In order to understand if germline mutations in *SH2D4A* could contribute in familial colorectal cancer development, whole genome sequencing of 200 index patients from familial colorectal cancer patients, without mutations in any of the known genes, will be done.

7.3 HMGA proteins as prognostic markers in different tumor entities

Our work demonstrates that the percentage of tumour cells showing HMGA1 and HMGA2 nuclear immunoreactivity correlates positively with increasing malignancy of breast and pancreatic tumors. These studies indicate that the overexpression of HMGA1 and HMGA2 is linked to proliferation in the cancer and may have important implications, promoting the growth and spread of tumors. On the molecular basis HMGA1 and HMGA2 interact with several transcription factors, influence gene expression patterns and regulate cell growth, differentiation, apoptosis, and transformation [156]. Their expression has been detected in many kinds of benign and malignant tumors and it is associated with a highly malignant phenotype (poor prognostic index) [221].

HMGA is highly expressed during embryogenesis, but is undetectable or very low in differentiated adult tissues [222, 223], being confined, at least for HMGA2, to the stem cell compartment [224-226]. *In vivo* studies revealed an important role of HMGA proteins in adipogenesis [222, 225], somatic growth [227], cardiac cell growth control [228] and glucose homeostasis [229, 230]. In accordance with this, HMGA mutations have been detected in human diseases such as lipomas [231-233], gigantism [234], dwarfism [235], and diabetes [229]. Moreover, HMGA2 has also been suggested recently by genome-wide SNP studies to influence human height variation [236, 237]. HMGA overexpression is a constant feature of human malignant neoplasms and is frequently associated, with or without gene rearrangements, with human tumors [221, 238]. In addition HMGA proteins also regulate the transcription of genes that are involved in DNA repair. Reeves *et al.* described a number of genes involved in DNA repair that were negatively regulated by HMGA1 in MCF7 human breast-cancer cells, suggesting that HMGA proteins can influence DNA repair by negatively regulating the transcriptional activity of genes involved in various aspects of DNA-damage recognition and removal [239]. Consistently, in another study made by Borrmann *et al.* HMGA2 has been linked to the promoter of the nucleotide excision-repair gene *ERCC1* where it negatively modulates its activity [240]. Moreover, it has been shown that HMGA1 can downregulate BRCA1 expression, which is involved in homologous recombination, by binding directly to its promoter region, and that there

is an inverse correlation between HMGA1 and BRCA1 expression in human breast carcinomas [241]. Consistent with the role of *BRCA1* in DNA double-strand break (DSB) repair, it has been shown that HMGA proteins potentiate genotoxic stress induced by different DNA-damaging agents causing DSBs, such as cisplatin, bleomycin, doxorubicin and X-rays [242, 243]. HMGA proteins can indirectly inhibit DNA repair through cyclin A induction, and it recently has been reported that the cyclin A1–cyclin-dependent kinase 2 complex also regulates DSB repair [221, 244]. Essentially, the data reported in this thesis indicate that HMGA1 and HMG2 are implicated in carcinogenesis and may play a role in the development of a particular phenotype of breast and pancreas carcinomas. Therefore, HMGA1 and HMGA2 expression may represent an indicator of poor prognosis of human tested cancers. Then studies are needed to eluate their role as prognostic factor in tumor progression and maybe as potential target in molecular therapy.

To understand which are the downstream targets directly regulated by HMGA1 and HMGA2, a transcriptome analysis of different tumors developed in *HMGA1* KO mice (Prof. Alfredo Fusco laboratory) is planned; this strategy allow us to define an expression pathway and thus understand the relationship with the most important tumour related pathway and test them *in vivo*.

7.4 *MAGE-A10* overexpression in lung, skin and urothelial malignancies

In this study we produced a *MAGE-A10* protein in soluble recombinant form, and used it to generate specific monoclonal antibodies (mAbs). One of these reagents, recognizing an epitope located at the COOH terminus of the *MAGE-A10* gene product, was used to stain a multitumor tissue microarray comprising more than 2,500 paraffin-embedded specimens including healthy tissues, benign tumors and malignancies of different histological origin. *MAGE-A10* was identified as an intranuclear protein of an apparent molecular weight of 70 kDa, expressed in normal spermatogonia and spermatocytes but in no other healthy tissue. Most importantly, this cancer/testis tumor associated antigen (C/T TAA) appears to be expressed in high (>50%) percentages of cancer cells from a number of malignancies, including lung, skin and urothelial tumors. Unexpectedly, high expression of *MAGE-A10* TAA at the protein level was also detectable in gynecological malignancies and stomach and gall bladder cancers.

MAGE-A proteins are known to be highly expressed in a wide range of cancers like breast, ovary, lung, skin, urothelial and bladder [181, 245-248] and their expression is observed mainly in cancers with malignant phenotypes, [181]. *MAGE* detection also correlates with poor prognosis in cancer patients, underpinning the idea that *MAGE* proteins may contribute actively towards malignancy [247]. The mechanism by which *MAGE-A* overexpression occurs in tumor cells is not totally understood. Usually the expression of *MAGE-A* in somatic tissues is repressed by DNA hypermethylation of CpG dinucleotides in promoters, which acts to prevent access of transcription factors like Ets and SP1 [249, 250]. In tumor cells epigenetic reprogramming can result in promoter hypo-methylation leading to the aberrant expression of one or more of these genes, in fact *MAGE-A* expression can be induced by demethylating agents such as 5-aza-2'-deoxycytidine in non-expressing cells of various origins [249, 250]. Seems to be also that chromatin remodelling events which occur during tumour development, like histone acetylation and methylation, can contribute to changes in *MAGE-A* levels in cancer cells and, at least in some circumstances, can be under control of hormones such as FGF, estrogen, leutenising hormone and directly regulated by microRNAs miR-34a [247, 251-254]. Tumors also show significant differences in signal localization (nuclear vs.

cytoplasmic) [245], suggesting that either different MAGE-A family members display differential localization or that the mechanisms controlling localization are cancer cell-specific.

Several studies have established that MAGE-A proteins can repress p53-mediated transcription, through direct and indirect mechanisms, and inhibit both p53-mediated apoptosis and senescence, two major tumour suppressor mechanisms utilized by p53 [247, 255-259]. These studies also establish the principle that elimination of MAGE-A expression in cultured cells is sufficient to induce p53-mediated apoptosis without the need to use genotoxic agents. This suggests that the development of compounds that block the p53/MAGE-A interaction in cancer cells expressing these proteins could have enormous therapeutic potential that minimizes the requirement for genotoxic approaches and the accompanying side effects. Advances in our knowledge of MAGE structure and interaction with partner proteins are beginning to cover the way towards developing such therapeutic approaches.

8. APPENDIX

8.1 microRNA expression profiling in mismatch repair associated colorectal cancer

My contribution to this work:

- Assessment and optimization of miRNA extraction techniques;
- Total RNA and microRNA extraction from CRC cell lines and human samples;
- miRNA library generation;
- Analysis of deep sequencing data;
- Statistical comparison;

microRNA expression profiling in mismatch repair associated colorectal cancer

Salvatore Piscuoglio¹, Shivendra Kishore², Mihaela Zavolan¹, Luigi Terracciano³ and
Karl Heinimann¹

¹ Research Group Human Genetics, Department of Biomedicine, University of Basel,
and Division of Medical Genetics, University Children's Hospital, Basel, Switzerland.

² Biozentrum, University of Basel, Basel, Switzerland.

³ Institute of Pathology, University Hospital of Basel, Basel, Switzerland

Abstract

Lynch syndrome represents the most common, autosomal dominantly inherited cancer predisposition worldwide and is characterised by early onset colorectal cancer around age 44 years. The syndrome accounts for 3 to 5% of all colorectal cancers in Switzerland and is caused by germ line mutations in DNA mismatch repair (MMR) genes, predominantly MLH1 and MSH2. MMR deficiency leads to genomic instability in the tumor, resulting in the genome-wide accumulation of somatic mutations, in particular at short repetitive sequences giving rise to microsatellite instability (MSI). To date, little is known on the specific molecular genetic alterations, which initiate and promote cancer development in Lynch syndrome patients.

microRNAs (miRNAs) are a family of small non-coding RNAs which are thought to control gene expression of about 30% of all protein-coding genes in humans. They regulate a variety of cellular processes and are likely to have a causal role in carcinogenesis since they are altered in most cancer types, including colon cancer. Their diagnostic, prognostic and therapeutic potential has fuelled miRNA research in recent years. Despite advances on the role of MSI status and miRNA expression in sporadic colorectal cancer, a comprehensive study on miRNA profiles in Lynch syndrome-associated colorectal cancers has not been performed to date.

In a long, painstaking process we evaluated and scrutinized several extraction methods and protocols to obtain high-quality miRNA from snap-frozen MMR proficient and deficient cancers as well as cancer cell lines. Finally, radiolabelling of total RNA with P³² followed by direct excision of the miRNA fraction from a 8-15% polyacrylamide gel and extraction with TRI reagent solution proved to yield the best miRNA quality. We therefore applied the Truseq sequencing on a HiSeq2000 platform and investigated the miRNA profiles by unsupervised analysis using hierarchical clustering and principal component analysis generating a initial list of top/bottom interesting miRNAs.

Currently, selected top most differentially expressed miRNAs will be verified / validated by quantitative real-time PCR on a total of 100 Lynch syndrome and 50 sporadic colorectal cancers. Subsequently, tissue expression of verified miRNAs and potential target proteins will be assessed by in situ hybridization and immunohistochemistry, respectively. Finally, selected miRNAs will be functionally characterised by cell transfection experiments.

Introduction

About 15-20% of all colorectal cancers (CRCs) are thought to arise from an inherited genetic susceptibility to colorectal adenomas and carcinomas [1]. Lynch syndrome (formerly known as hereditary non-polyposis colorectal cancer, HNPCC), is an autosomal dominantly inherited cancer condition with an estimated carrier frequency of 1:2000 (up to 1:200) and thus represents the most common inherited cancer predisposition worldwide. It is estimated to account for about 2-5% of the total colorectal (CRC) cancer burden [2].

It is characterized by the development of colorectal as well as a distinct spectrum of extracolonic cancers (predominantly of the endometrium, ovary, stomach), usually diagnosed before age 50 years. The syndrome is caused by germ line mutations in the mismatch repair (MMR) genes MLH1, MSH2, MSH6, and PMS2, which recognize and correct errors that occur during DNA replication. MSH2 and MLH1 are mutated in about 80% to 90% of patients with typical Lynch syndrome whereas germ line mutations in MSH6 and PMS2 tend to result in a less severe phenotype [3]. Following somatic inactivation of the wild-type allele in the cancer cell MMR deficiency leads to genome-wide accumulation of replication errors predominantly at microsatellite loci, the hallmark lesion of Lynch syndrome, termed microsatellite instability (MSI). In the clinical setting, immunohistochemical assessment of the presence / absence of the MMR proteins in the tumor is used as a preliminary tool, often in conjunction with microsatellite instability testing, to assess MMR proficiency/deficiency and to determine who may benefit from germ line testing [4].

microRNAs (miRNAs) represent a large family of small non-coding RNAs of about 21 nucleotides in length that serve as effector molecules of sequence-specific gene silencing [5]. It is estimated that the number of miRNAs in the human genome range from about 450 to 1000 and that they control gene expression of about 30% of all protein-coding genes in mammals. The majority of identified miRNAs, currently estimated are highly evolutionary conserved among many distantly related species, from worms to human, suggesting that miRNAs have very important roles in essential biological processes, including developmental timing, stem-cell differentiation, signal transduction, disease and cancer. Their causal role in carcinogenesis is further substantiated by the fact that miRNAs are altered in most cancer types, including colon cancer [6].

In the nucleus, miRNA genes are generally transcribed by RNA polymerase II or III to form large primary miRNA transcripts (pri-miRNAs). These are further processed by Drosha, a RNase III protein, into 70-nucleotide miRNA precursors (pre-miRNA). After transport into the cytoplasm, pre-miRNAs are further processed by another RNase III enzyme, Dicer, into miRNA duplexes, typically consisting of 19-25 nucleotides in length [7, 8]. Subsequently these duplexes can be loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) and the mature miRNA strand is preferentially retained. Once bound to the 3'UTR of target mRNAs, the mature miRNA induces cleavage, translational repression or deadenylation, depending on the degree of complementarity [9]. A single miRNA may bind to as many as 200 target genes encoding a broad range of proteins, such as transcription factors, receptors and transporters. In recent years, several approaches have been used to identify miRNA targets [10].

With the advent of massively parallel (“next generation”) sequencing technology genome-wide miRNA profiling has now become possible allowing global assessment of expression-regulating mechanisms in sporadic and hereditary cancer [11, 12].

The results of this study will enhance the scarce knowledge available to date on the role of miRNAs in Lynch syndrome-associated colorectal cancer development. The findings are likely to shed light on intestinal carcinogenesis in general and allow further characterisation of the various pathways involved in MSI-related tumorigenesis. The identification and characterisation of miRNA expression signatures in Lynch syndrome cancer patients may not only improve clinical risk assessment and prognosis, but also offer an opportunity to identify novel therapeutic strategies for cancer patients.

Materials and methods

RNA Isolation

To isolate genomic total RNA from cell lines Tri-Reagent (Ambion) were used according to the manufacturers' guidelines.

Cell lines

Six colorectal cancer cell lines from the American Type Culture Collection (ATCC, Rockville, MD) were used for this study: four repair deficient cell lines (HCT116, LoVo, HCT15, DLD-1) and two mismatch repair proficient (SW480 and HT29). HCT116, HCT15 and DLD-1 cells (ATCC, Rockville, MD) were cultured in RPMI 1640 (Invitrogen Basel, Switzerland) supplemented with 10% fetal bovine serum FBS, 1% Kanamycin sulphate, 1% GlutaMAX-I, 1% Sodium Pyruvate, 1% non Essential Amino Acids (NEAA), 1% HEPES (all from Invitrogen Basel, Switzerland) and 0.1% 2-mercapto-ethanol (Sigma-Aldrich Basel, Switzerland). HT29 cells were grown in McCoy's 5A Medium (Invitrogen Basel, Switzerland) with 10% fetal bovine serum FBS, Kanamycin sulphate and GlutaMAX-I (all from Invitrogen Basel, Switzerland). SW480 cells were cultured in L-15 Medium (Sigma-Aldrich Basel, Switzerland) with 10% FBS, 1% GlutaMAX-I and 1% Kanamycin sulphate (all from Invitrogen Basel, Switzerland). Cells were maintained at 37°C with 5% CO₂.

Preliminary results

To yield the best miRNA quality the total RNA (typically 2 µg) marked radioactively with P³² from each sample was run on denaturing polyacrylamide-urea gels. The approximately 17-25 nucleotide RNAs were excised from the gel, ligated to sequencing adaptors on both ends, and reverse-transcribed according to the manufacturers' guidelines of Illumina. The resulting cDNA library was PCR-amplified for 15 cycles and gel-purified on 6% acrylamide gel. The gel-purified amplicon quality and quantity were analyzed on a 6% acrylamide gel relative to oligonucleotides of known concentration and size. The library (120 µL 1-4pM) was loaded on an Illumina HiSeq 2000 system (Illumina) on a flow cell according to the manufacturer's instructions, where DNA molecules were attached to high-density universal adaptors in the flow cells and amplified. The DNA clusters generated via this process were sequenced with sequencing-by-synthesis technology, where successive high-resolution images of the 4-color fluorescence excitation dependent on the base incorporated during each cycle were captured [13]. Then, the data were pre-processed, including steps of quality check and normalization. miRNA profiles were investigated by unsupervised analysis using hierarchical clustering and principal component analysis (PCA). For computational prediction of miRNA targets and pathway analysis the MirZ web server was used which provides statistical analysis and data mining tools operating on up-to-date databases of sequencing-based miRNA expression profiles and of predicted miRNA target sites [14]

In total, we identified 1240 miRNAs differentially expressed in MMR deficient compared to proficient cancers cell lines. Among the most downregulated miRNAs (Table 1) were hsa-miR-371-5p, involved in Wnt/beta-catenin signaling, a crucial pathway for colorectal carcinogenesis [15], and hsa-miR-200a, which has been shown to inhibit the epithelial-mesenchymal transition [16] and to be deregulated in bladder breast and endometrial cancer (which are, intriguingly, part of the LS-tumor spectrum) [17, 18]. Among the most upregulated miRNAs we observed hsa-miR-141 which is believed to modulate the oxidative stress response [19] and hsa-miR-125b, a negative regulator of p53, also referred to as the “guardian of the genome”[20].

Currently we are analyzing the deep sequencing data of 4 cancer specimens matched with their normal counterpart from 2 LS-related, microsatellite-unstable as well as 2 sporadic microsatellite-stable CRCs.

Brief discussion

MMR-deficient CRCs exhibit mRNA and miRNA expression profiles distinct from their stable, MMR-proficient counterparts. The study by Di Pietro et al. [21], in which our research group also took part, analysed gene expression in proximal colon cancers and was able to divide them into two groups that almost perfectly corresponded with their MMR status. In addition, expression changes in genes involved in apoptosis and the immune response were consistent with the better prognosis of MMR-deficient cancers. Kruhoffer et al. [22] studied mRNA expression in 34 MSI-High (mostly sporadic) and 67 MSS stage II and III colorectal cancers and devised a gene expression signature based on nine genes able to distinguish MMR status.

To determine the degree of gene expression differences which could be explained by CpG island methylation, we assessed the presence of CpG islands in and 5kb up- and down-stream of the top most differentially expressed genes from both studies using the UCSC genome database: only about 50-60% of these genes harbour CpG islands. Thus, about 40-50% of the gene expression changes in MMR-deficient CRCs cannot readily be explained by epigenetic regulation and are likely to be caused by other mechanisms, in particular deregulation of miRNA expression.

The mutator phenotype that results from MMR dysfunction induces the acquisition of additional gene mutations that promote cancer progression [23]. In addition to germline mutations, various pathogenic events, including promoter methylation [24] and reduced histone acetylation [25], result in reduced or absent expression of core MMR proteins, as do microenvironmental factors, such as inflammation and hypoxia [26]. Our preliminary results suggest that some miRNA such as miR-155, could play a role in this multifactorial regulation by causing down-modulation of the core MMR heterodimeric proteins MSH2-MSH6 and MLH1-PMS2. The simultaneous inhibition of these essential MMR components by some miRNA could well explain the observed mutator phenotype. Thus far, only little is known about the role of MSI status and miRNA expression in CRC. A study by Lanza et al. [27] established a sensitive predictive algorithm to correctly distinguish between sporadic MSS and MSI-H CRCs using a 14 miRNA signature in 39 samples. Schepeler et al. [28] used microarrays to profile the expression of 315 miRNAs in normal mucosa samples (n=10) and stage II colon cancers (n=49) differing with regard to microsatellite status

and recurrence of disease. They observed that miR-145 expression was lower in cancer relative to normal tissue and that microsatellite status could be correctly predicted based on miRNA expression profiles. In conclusion, although largely limited to sporadic MMR-deficient CRCs, current data provide evidence that perturbed expression of miRNAs in colon cancer may not only have a functional effect on tumor cell behaviour, but that some miRNAs with prognostic potential could be of clinical importance [29].

The results of this study will enhance the scarce knowledge available to date on the role of miRNAs in Lynch syndrome-associated colorectal cancer development. The findings are likely to shed light on intestinal carcinogenesis in general and allow further characterisation of the various pathways involved in MSI-related tumorigenesis. The identification and characterisation of miRNA expression signatures in Lynch syndrome cancer patients may not only improve clinical risk assessment and prognosis, but also offer an opportunity to identify novel therapeutic strategies for cancer patients.

References

1. Cannon-Albright, L.A., et al., *Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers*. N Engl J Med, 1988. **319**(9): p. 533-7.
2. Lynch, H.T., et al., *Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications*. Eur J Hum Genet, 2006. **14**(4): p. 390-402.
3. Boland, C.R., et al., *The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch syndrome: from bench to bedside*. Fam Cancer, 2008. **7**(1): p. 41-52.
4. Hampel, H., et al., *Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer)*. N Engl J Med, 2005. **352**(18): p. 1851-60.
5. Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg, *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?* Nat Rev Genet, 2008. **9**(2): p. 102-14.
6. Cummins, J.M., et al., *The colorectal microRNAome*. Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3687-92.
7. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex*. Nature, 2004. **432**(7014): p. 231-5.
8. Kim, V.N. and J.W. Nam, *Genomics of microRNA*. Trends Genet, 2006. **22**(3): p. 165-73.
9. Zeng, Y., R. Yi, and B.R. Cullen, *MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms*. Proc Natl Acad Sci U S A, 2003. **100**(17): p. 9779-84.
10. Berezikov, E., E. Cuppen, and R.H. Plasterk, *Approaches to microRNA discovery*. Nat Genet, 2006. **38** **Suppl**: p. S2-7.
11. Sanchez-Navarro, I., et al., *Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen and formalin-fixed, paraffin-embedded breast cancer tissues*. Biotechniques, 2010. **48**(5): p. 389-97.
12. Xi, Y., et al., *Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples*. RNA, 2007. **13**(10): p. 1668-74.

13. Jima, D.D., et al., *Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs*. *Blood*, 2010. **116**(23): p. e118-27.
14. Hausser, J., et al., *MirZ: an integrated microRNA expression atlas and target prediction resource*. *Nucleic Acids Res*, 2009. **37**(Web Server issue): p. W266-72.
15. Zhou, A.D., et al., *beta-Catenin/LEF1 transactivates the microRNA-371-373 cluster that modulates the Wnt/beta-catenin-signaling pathway*. *Oncogene*, 2012. **31**(24): p. 2968-78.
16. Aydogdu, E., et al., *MicroRNA-regulated gene networks during mammary cell differentiation are associated with breast cancer*. *Carcinogenesis*, 2012. **33**(8): p. 1502-11.
17. Wang, G., et al., *Expression of microRNAs in the urine of patients with bladder cancer*. *Clin Genitourin Cancer*, 2012. **10**(2): p. 106-13.
18. Castilla, M.A., et al., *Micro-RNA signature of the epithelial-mesenchymal transition in endometrial carcinosarcoma*. *J Pathol*, 2011. **223**(1): p. 72-80.
19. Mateescu, B., et al., *miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response*. *Nat Med*, 2011. **17**(12): p. 1627-35.
20. Le, M.T., et al., *Conserved regulation of p53 network dosage by microRNA-125b occurs through evolving miRNA-target gene pairs*. *PLoS Genet*, 2011. **7**(9): p. e1002242.
21. di Pietro, M., et al., *Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers*. *Gastroenterology*, 2005. **129**(3): p. 1047-59.
22. Kruhoffer, M., et al., *Gene expression signatures for colorectal cancer microsatellite status and HNPCC*. *Br J Cancer*, 2005. **92**(12): p. 2240-8.
23. Perucho, M., *Tumors with microsatellite instability: many mutations, targets and paradoxes*. *Oncogene*, 2003. **22**(15): p. 2223-5.
24. 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 Res*, 1997. **57**(5): p. 808-11.
25. Edwards, R.A., et al., *Epigenetic repression of DNA mismatch repair by inflammation and hypoxia in inflammatory bowel disease-associated colorectal cancer*. *Cancer Res*, 2009. **69**(16): p. 6423-9.

26. Kondo, A., et al., *Hypoxia-induced enrichment and mutagenesis of cells that have lost DNA mismatch repair*. *Cancer Res*, 2001. **61**(20): p. 7603-7.
27. Lanza, G., et al., *mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer*. *Mol Cancer*, 2007. **6**: p. 54.
28. Schepeler, T., et al., *Diagnostic and prognostic microRNAs in stage II colon cancer*. *Cancer Res*, 2008. **68**(15): p. 6416-24.
29. Schetter, A.J., et al., *MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma*. *JAMA*, 2008. **299**(4): p. 425-36.

Table1: Top 20 miRNAs differentially expressed in MMR deficient compared to MMR proficient colorectal cancers cell lines

miRNA Name	Direction	Pool1 count (MMR-)	Pool1 frequency	Pool2 count (MMR+)	Pool2 frequency	Log(Psame/Pdiff)
hsa-miR-499-5p	↓	1642	0.0010	119032	0.0090	-150763
hsa-miR-371-5p	↓	18184	0.0010	109922	0.0080	-99962
hsa-miR-200a	↓	350694	0.01	342852	0.025	-84781.8
hsa-miR-373	↓	14098	0.0010	81549	0.0060	-73114.1
hsa-miR-200b	↓	704912	0.019	479924	0.035	-56962.2
hsa-miR-372	↓	10283	0.0010	59112	0.0050	-52861.9
hsa-miR-224	↓	41743	0.0020	85516	0.0070	-46316.7
hsa-miR-371-3p	↓	6588	0.0010	44572	0.0040	-41790.8
hsa-miR-429	↓	152804	0.0040	156215	0.012	-41021.6
hsa-miR-141	↑	1762050	0.046	401587	0.029	-38202.7
hsa-miR-10a	↑	1986950	0.052	467442	0.034	-38149.1
hsa-miR-31	↓	602096	0.016	370715	0.027	-32469.1
hsa-miR-100	↑	164399	0.0050	6491	0.0010	-31483.3
hsa-miR-125a-5p	↓	117214	0.0040	119074	0.0090	-30921.8
hsa-miR-24	↓	566033	0.015	347730	0.026	-30194.2
hsa-miR-424	↓	30682	0.0010	56459	0.0050	-28132.9
hsa-miR-148a	↑	303974	0.0080	35600	0.0030	-26735
hsa-miR-125b	↑	125963	0.0040	5836	0.0010	-22502
hsa-miR-499-3p	↓	241	0.0010	16837	0.0020	-21238
hsa-miR-192	↓	627758	0.017	348276	0.026	-20761.9

8.2 Identification of novel recurrent duplication “hot spots” in Lynch syndrome colorectal cancers.

My contribution to this work:

- DNA extraction from Lynch syndrome and sporadic colorectal cancers;
- Analysis of high resolution chip array;
- Selection of candidate genes;
- Validation of selected candidate by qPCR;
- Statistical comparison;
- Manuscript preparation;

Identification of novel recurrent duplication “hot spots” in Lynch syndrome colorectal cancers

Salvatore Piscuoglio¹, Nicole Lüscher¹, Anamaria Dumea¹, Michal Kovac², Luigi Terracciano² and Karl Heinimann^{1§}.

¹Research Group Human Genetics, Department of Biomedicine, University of Basel, and Division of Medical Genetics, University Children's Hospital, Basel, Switzerland.

²Institute of Pathology, University Hospital of Basel, Basel, Switzerland.

Abstract

Background

Lynch syndrome (also Hereditary Non-Polyposis Colon Cancer, HNPCC) represents the most common, autosomal dominantly inherited cancer predisposition worldwide and accounts for 3-5% of the total colorectal cancer (CRC) burden. It is caused by germline mutations in DNA mismatch repair (MMR) genes (mainly *MLH1* and *MSH2*). MMR deficiency results in microsatellite instability (MSI), i.e. genome-wide accumulation of somatic alterations at repetitive DNA sequence motifs, and present in 90% of Lynch syndrome-related cancer. About 80% of MSI tumors have a near-diploid karyotype which stands in clear contrast to the microsatellite stable (MSS) cancers which predominantly are aneuploidy. Thus far, little is known on the type and frequency of, microdeletions and microduplications in LS-related CRCs.

Methods

Here, we applied a high-density CGH microarray-based method using the Affymetrix Whole Genome 2.7 M chip, to study somatic copy number aberrations in CRCs from 12 unrelated Swiss LS patients, and whose cancers displayed microsatellite instability with confirmed germline mutation in *MLH1* or *MSH2*. Next we validated the results in 46 LS-related as well as 50 colorectal cancers by quantitative real-time PCR using locus specific primer pairs.

Results

Copy number assessment by CGH array revealed 2 novel somatic microduplications "hot spot" regions containing *ERCC2* (19q13.32) and *STK40* (1p34.3) gene. The frequency of these copy number aberrations were further validate in our cohort of Lynch syndrome cancers. CNA were present in 24/46 (60.8%) and in 18/46 (39.1%) CRCs for *ERCC2*.and *STK40*, respectively.

Conclusion

Our study identified novel microduplications in Lynch syndrome related CRC located on 1p34.3 and 19q13.32 These genes associated with CN changes in LS-related CRCs warrant further investigation to establish their possible clinical implications. Currently, to confirm that these CNA are an only present in Lynch syndrome CRC 50 sporadic CRCs are under investigation.

Introduction

Hereditary nonpolyposis colorectal cancer (HNPCC), also referred to as the Lynch syndrome (LS), with an estimation between 1:200 and 1:2000 predisposition is the most common form of autosomal dominantly inherited cancer predisposition worldwide. It is characterized by the occurrence of early onset colorectal carcinoma (CRC) as well as a distinct spectrum of extracolonic tumors, such as endometrium, stomach, ovarian, breast and renal pelvis cancers [1, 2] and caused by a germline mutation in mismatch repair (MMR) genes, *MLH1* and *MSH2* (90%) and *MSH6* (10%) [1, 2]. Tumors with MMR deficiency exhibited frequent errors in microsatellite DNA, short segments of DNA containing tandem repeats of mono-, di-, tri- or tetranucleotide [3]. The high-frequency MSI (MSI-H) CRCs have unique clinicopathologic features, such as right-sided, mucinous or poorly differentiated, and stable chromosomal status in the tumors [4].

DNA copy number variation (CNV) and other structural variations in the human genome are increasingly recognized as an alternative source of genetic variation that may influence cancer risk and are a common occurrence in all forms of cancer [5-9]. A typical cancer sample exhibits an average of 17% amplifications and 16% deletions within an entire genome [10]. Somatic copy number alterations have been shown to significantly affect pathways involving tumor suppressor genes such as *TP53*, *APC*, *BRCA1*, *BRCA2*, *PTEN*, and *RB1* and oncogenes, including *HRAS* and *RET* [5, 11]. Detection of these alterations and identification of the specific genes responsible for cancer proliferation can help to subtype cancers at the molecular levels and lead toward more individualized cancer-type specific therapies [12-15].

About 80% of MSI tumors have a near-diploid karyotype and a genetic alterations different from those of microsatellite stable (MSS) cancers [16-20]. Despite the progress of our understanding of CRC genetics, genomic alterations of various subtypes of CRC have not been fully characterized. Copy number variations (CNVs) can contribute to variable levels of gene expression [21], and thus fine-scale copy number (CN) profiling of cancer may further enhance our knowledge about tumorigenesis. This study aim to investigate 12 Lynch syndrome-related CRCs with high-density CGH microarray-based method in search of somatic microduplications and/or microdeletions.

Materials and methods

Patient characteristics

For this study a cohort of 52 Lynch syndrome-related colorectal cancers with identified germline mutation (30 *MLH1*, 16 *MSH2*, 3 *MSH6*, 2 *PMS2*) were analysed for CNV.

In addition, based on the recommendations of the National Cancer Institute workshop on MSI, a panel of microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250) [3] and two additional microsatellite markers (BAT40, MYCL1) were used to determine MSI status.

Patient data including full follow-up were obtained by retrospective analysis of medical records, regional tumor registries and/or treating physicians. Tissue samples were obtained by surgical or endoscopic excision.

Genomic DNA Isolation

To isolate genomic DNA from fresh frozen tumor tissue samples the QIAamp DNA Mini Kit (Qiagen, Hombrechtikon, Switzerland) and for formalin fixed paraffin embedded (FFPE) tumor samples the RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Invitrogen, Carlsbad, CA, USA) were used according to the manufacturers' guidelines.

Cytogenetics Whole-Genome 2.7M Array

A total of 500 ng of genomic DNA of six LS patients (each of them were matched with its normal counterpart), were analyzed using an Affymetrix Cytogenetics Whole-Genome 2.7M Array, a high resolution array containing approximately 400,000 SNP markers and 2.3 million non-polymorphic markers, with high density coverage across cytogenetically significant regions according to the manufacturer's instructions. Data was collected using either GeneChip® Scanner 3000 Dx and CEL files were analyzed using Affymetrix Chromosome Analysis Suite software (ChAS v.1.1). The annotation file used in our analysis can be found on the Affymetrix website, listed as ArrayNA30.2 (hg18). CNVs detected were compared with the Database of Genomic Variants (<http://projects.tcag.ca/variation>) for overlap with known copy number variants using previously described criteria [24].

Determination of selected copy number aberrations by Real-time PCR

Oligonucleotide primers for quantitative PCR (qPCR) were designed for each gene using the AlleleID software (PREMIER Biosoft International, Palo Alto, CA) (*ERCC2* for: 5'-TGGGAAATGAACGGGAAACAG-3'; rev: GGGCAAGACGGACTACGG; *STK40* for: 5'-GTCCTGTTCTCCTGTCTC-3'; rev: GGCTGCGTAATATGATGG-3'; *CFTR* for: 5'- GCATGGGAGGAATAGGTGAA-3'; rev: 5'- CACAATCTACACAATAGGACATGG-3'), to assure maximal efficiency and sensitivity according to the following parameters: avoidance of the formation of self and hetero-dimers, hairpins and self-complementarity, primer length and melting temperature. These properties were further verified using different internet-based interfaces such as Primer-3 [25]. Melting curve analysis was always performed at the end of each PCR assay to control for specificity.

qPCR was performed using standard protocols with 2X iQ SYBR Green supermix (Bio-Rad) on a Bio-Rad iCycler. Briefly, 50 ng DNA was added to 12.5 μ l of SYBR-green PCR master mix (Bio-Rad), with 0.5 μ l (600 nM) of each primer, and water to a final volume of 25 μ l. The reactions were amplified in a single step of 3 min at 95°C and then for 40 cycles of 10 s at 95°C, 1 min at 60°C, with final denaturation step for 10 sec at 95°C. The thermal denaturation protocol was run at the end of the PCR to determine the number of products that were present in the reactions. Experiments were done in triplicate and included non-template controls for each gene. The amount of each gene was normalized to *cystic fibrosis transmembrane conductance regulator (CFTR)* as reference gene. We used a standard analysis to calculate the amplification of the genes by the $2^{-\Delta\Delta C_t}$ method as described previously [26]. Results for each sample were expressed as the N-fold copy number change [27].

Results

Initially we investigated six LS-related CRCs from 4 known *MLH1* and 2 *MSH2* germline mutation carriers matched with their tumor free counterpart for the presence of copy number aberrations (CAN), i.e. micro- deletion/duplication using the Affymetrix Whole-Genome 2.7M CGH array

We observed somatic gains (microduplications) at 1p34.3, 2q24.3, 3p23, 5q23.2, 5q33.1, 6p23, 8q11.1, 10p15.1, 10q25.1, 12p11.23, 17p13.1, 17q24.3, 19q13.32, Xp22.2 and somatic losses (microdeletions) at 19p12, 3p14.2. The overall findings of DNA copy number gains and losses across all samples are shown in Table 1. Interestingly, we identified 2 recurrent novel somatic microduplicated “hot spots” for genomic rearrangements located on 1p34.3 and 19q13.32 containing the *ERCC2* and *STK40* gene, respectively (Figure 1 A and 1 B).

To further assess the frequency at these loci we evaluated 46 LS-related CRCs from 30 *MLH1*, 16 *MSH2*, 3 *MSH6* and 3 *PMS2* mutation carriers by qPCR. The analysis revealed microduplications in *ERCC2* were present in 24/46 (60.8%) CRCs and in *STK40* in 18/46 (39.1%) CRCs (Table 2; Figure 2 A and 2 B).

In order to assess if these novel recurrent “hot spot” region represent somatic alterations typical for LS-related tumorigenesis, a set of 50 sporadic, microsatellite-stable CRCs is currently under investigation.

Brief discussion

The understanding of chromosomal aneuploidies and their role in tumor development is a fundamental problem in cancer biology. Chromosomal aneuploidy, the gain or loss of chromosomes is the most common alteration in cancer. The majority of cancer cells in sporadic MSS CRCs have numerical and structural chromosomal abnormalities with translocations, deletions and other aberrations. In contrast, about 80% of MSI tumors display a near-diploid karyotype and a distinct genetic alteration distinguishable from those observed in MSS cancers [19].

In this study, using the Affymetrix Whole Genome 2.7 M chip array in 6 fresh-frozen Lynch syndrome-associated CRCs we identified 2 novel, recurrent microduplication “hot spot” regions on chromosomes 1p34.3 and 19q13.32 containing the *ERCC2* and *STK40* genes, respectively, and validated the findings by qPCR in a cohort of 46 additional LS-related CRCs. To assess if these alterations are specific to LS related cancers currently 50 sporadic CRCs is currently under investigation

ERCC2 is a key component of the nucleotide excision repair pathway. Further, the protein is an integral member of the basal transcription factor BTF2/TFIIH complex, displays an ATP-dependent DNA helicase activity and belongs to the RAD3/XPD subfamily of helicases [28]. Defects in this gene can result in three different disorders: the cancer-prone syndrome xeroderma pigmentosum complementation group D, trichothiodystrophy, and Cockayne syndrome.

Little is known concerning the *STK40* that may be a negative regulator of NF-kappa-B and p53-mediated gene transcription.

To date, it remains to be clarified if and how these microduplications and amplifications (up to 6n) may affect cancer cell proliferation and progression in Lynch syndrome colorectal cancers.

Among all somatic mutations, non-germline CNVs found in the cancer genomes, also known as copy number alterations/aberrations (CNAs), are frequently observed, e.g., gains of oncogene and losses of tumor suppresser gene loci [22]. Furthermore, the DNA CN states of CRC cases are related to the response of drug treatments, e.g., the degree of CRC-related CNA is associated with response to systemic combination chemotherapy with capecitabine and irinotecan [23].

Figures and Tables

Figure 1: ChAS analysis of LS-related CRCs

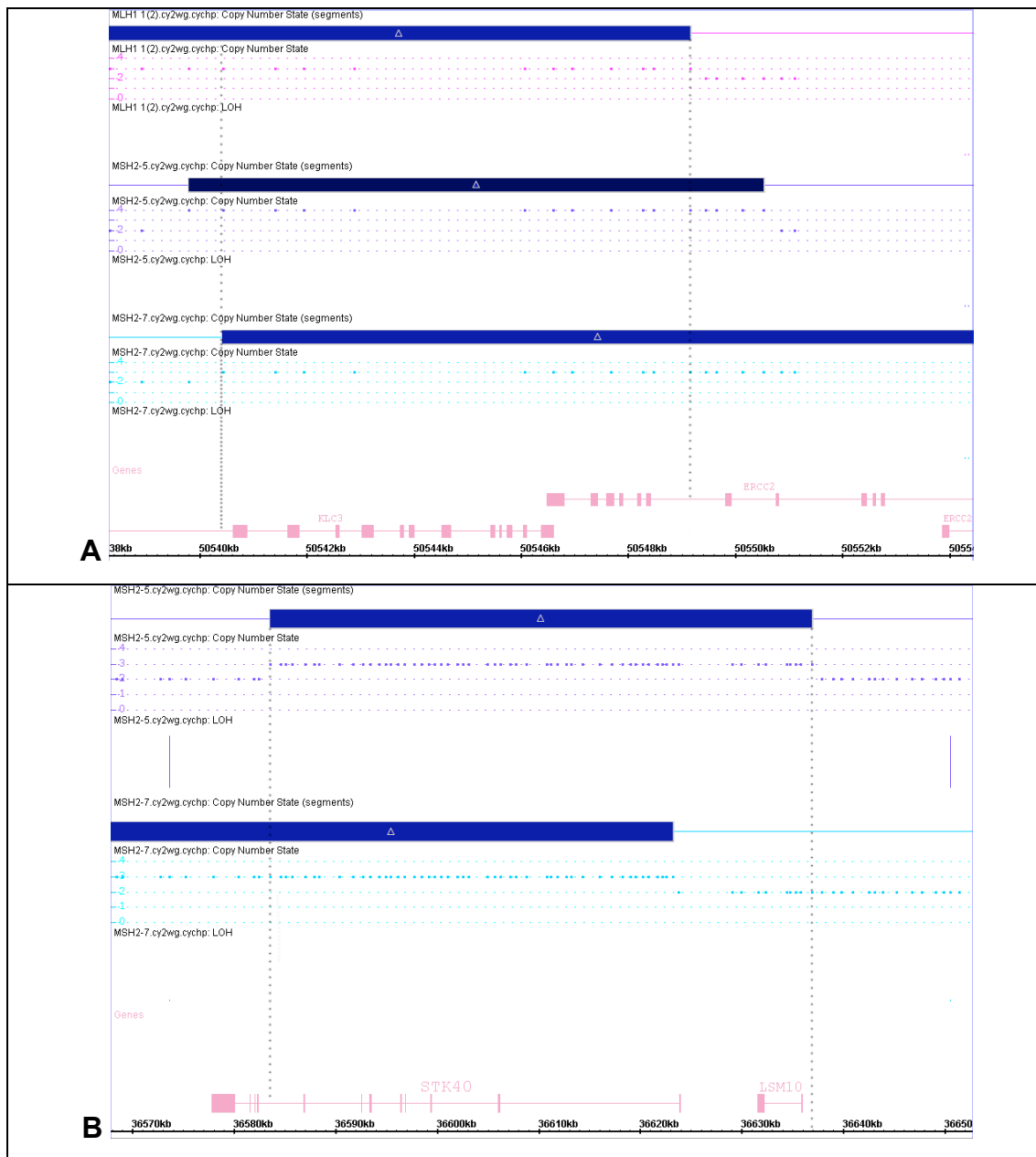


Figure 1: A) Minimal region duplicated on chr19 (19q13.32) including *ERCC2* gene in 3 different patients. B) Minimal region duplicated on chr1 (1p34.3) including *STK40* gene in 2 different patients.

Figure 2: qPCR results in 46 LS-related CRCs: A) CNAs in *ERCC2* gene. B) CNAs in *STK40* gene.

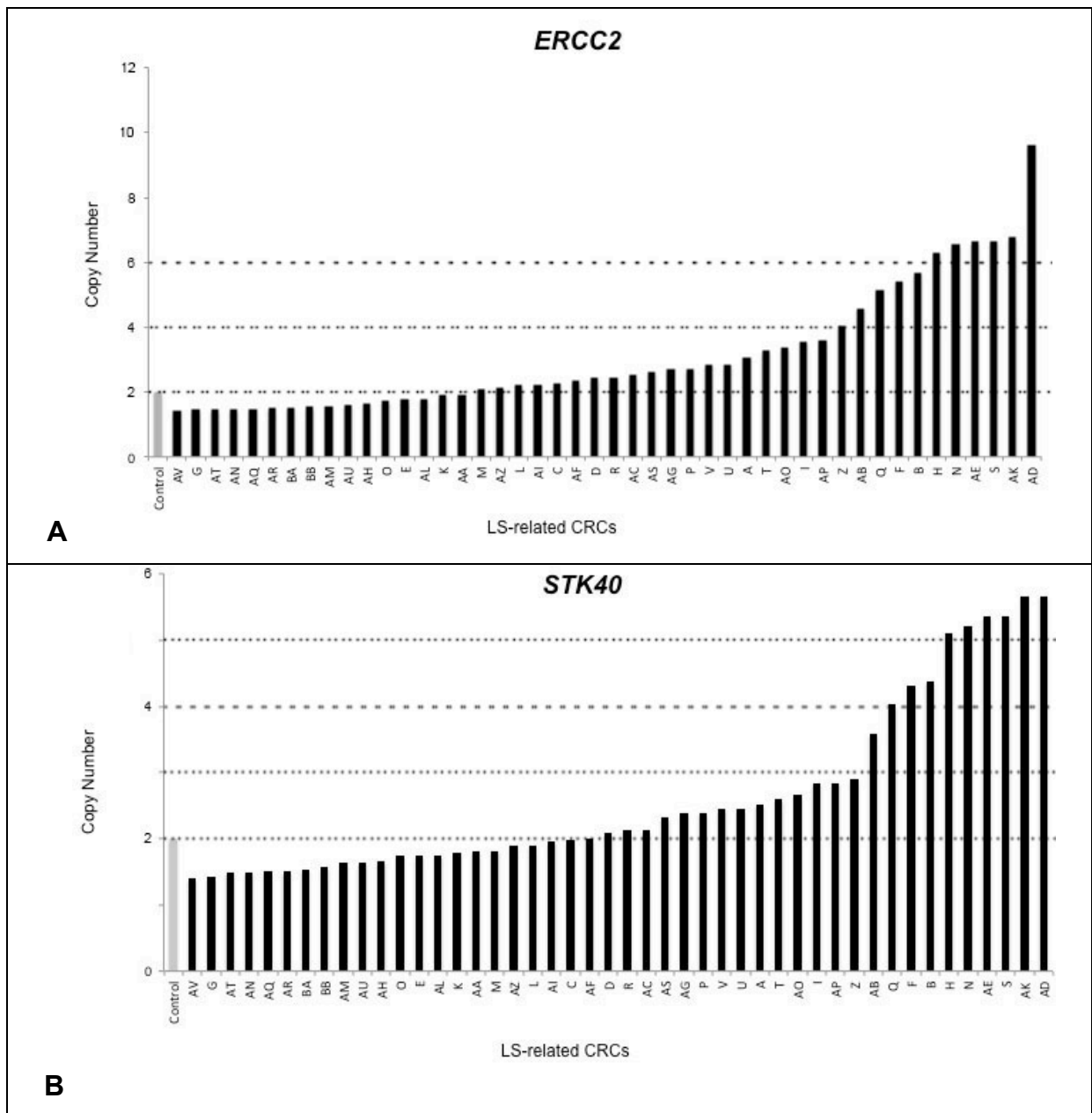


Table 2

Gene name	Copy number aberrations			
	2n (diploid)	3n	4n	>4n
<i>ERCC2</i>	22	11	5	8
<i>STK40</i>	28	8	5	5

Table 2: Degree of copy number aberrations in 46 LS-related CRCs

Table 1

Chromosome	Locus	Alteration	Base Start	Base End	Del Size (Mb)	Patient n.	Genes
1	1p34.3	Duplication	36583571	36636982	53,4	5	STK40, LSM10
			36567450	36623262	55,8	7	STK40
2	2q24.3	Duplication	168993733	169032679	38,9	1	LASS6
			168993733	169036109	42,3	7	
	3p23	Duplication	31535493	31557397	21,9	7	STT3B
			31540540	31557397	16,8	16	
3	3p14.2	Deletion	60421679	60439908	18,2	11	FHIT
			60430695	60450786	20	16	
			60530263	60542710	12,4	11	
			60497433	60554225	56,7	16	
5	5q23.2	Duplication	122869729	122896827	27	11	CSNK1G3
			122871101	122896827	25,7	16	
	5q33.1	Duplication	151109899	151145327	35,4	5	G3BP1
			151131753	151142191	10,4	16	
6	6p23	Duplication	15406555	15428655	22,1	5	JARID2
			15402243	15454282	52	16	
8	8q11.1	Duplication	47134349	47261260	126,9	1	Intragenic region
			47120825	47290072	169,2	7	
10	10p15.1	Duplication	6168193	6192155	23,9	11	RBM17
			6168193	6192694	24,5	16	
	10q25.1	Duplication	111746435	111756922	10,4		ADD3
			111747464	111769074	21,6	1	
12	12p11.23	Duplication	26952817	26993684	40,8	16	FGFR10P2
			26974916	26993684	18,7		
17	17p13.1	Duplication	10499352	10536740	37,3	7	SCO1, MYH3
			10528556	10566728	38,1	11	SCO1
	17q24.3	Duplication	64815529	64839814	24,2	1	MSI2, ABCA5
			64827465	64848873	21,4	16	
19	19q13.32	Duplication	50517505	50549160	31,6	1	ERCC2, KLC3, CKM
			50539800	50550532	10,7	5	ERCC2, KLC3
			50540437	50669798	129,3	5	ERCC2, KLC3, PPP1R13L, CD3EAP, ERCC1
	19p12	Deletion	22987786	23041491	53,7	11	MUC16
			22987786	23041491	53,7	16	
	X	Xp22.2	Duplication	9585012	9610052	25	3
9591311				9623919	32,6	16	

Table 1: CGH analysis results for copy number aberration in 6 LS-related CRCs.

References

1. Zhang, J., et al., *Gene conversion is a frequent mechanism of inactivation of the wild-type allele in cancers from MLH1/MSH2 deletion carriers*. *Cancer Res*, 2006. **66**(2): p. 659-64.
2. Lynch, H.T. and A. de la Chapelle, *Hereditary colorectal cancer*. *N Engl J Med*, 2003. **348**(10): p. 919-32.
3. Boland, C.R., et al., *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer*. *Cancer Res*, 1998. **58**(22): p. 5248-57.
4. Soreide, K., et al., *Microsatellite instability in colorectal cancer*. *Br J Surg*, 2006. **93**(4): p. 395-406.
5. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers*. *Nature*, 2010. **463**(7283): p. 899-905.
6. Degenhardt, Y.Y., et al., *High-content analysis of cancer genome DNA alterations*. *Curr Opin Genet Dev*, 2008. **18**(1): p. 68-72.
7. Hastings, P.J., et al., *Mechanisms of change in gene copy number*. *Nat Rev Genet*, 2009. **10**(8): p. 551-64.
8. Santarius, T., et al., *A census of amplified and overexpressed human cancer genes*. *Nat Rev Cancer*, 2010. **10**(1): p. 59-64.
9. Pinkel, D. and D.G. Albertson, *Array comparative genomic hybridization and its applications in cancer*. *Nat Genet*, 2005. **37 Suppl**: p. S11-7.
10. Wrzeszczynski, K.O., et al., *Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer*. *PLoS One*, 2011. **6**(12): p. e28503.
11. Shlien, A. and D. Malkin, *Copy number variations and cancer*. *Genome Med*, 2009. **1**(6): p. 62.
12. Cerami, E., et al., *Automated network analysis identifies core pathways in glioblastoma*. *PLoS One*, 2010. **5**(2): p. e8918.
13. Chen, S., et al., *Copy number alterations in pancreatic cancer identify recurrent PAK4 amplification*. *Cancer Biol Ther*, 2008. **7**(11): p. 1793-802.
14. Yap, T.A., C.P. Carden, and S.B. Kaye, *Beyond chemotherapy: targeted therapies in ovarian cancer*. *Nat Rev Cancer*, 2009. **9**(3): p. 167-81.

15. Leary, R.J., et al., *Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers*. Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16224-9.
16. Markowitz, S., et al., *Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability*. Science, 1995. **268**(5215): p. 1336-8.
17. Rampino, N., et al., *Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype*. Science, 1997. **275**(5302): p. 967-9.
18. Rowan, A., et al., *Refining molecular analysis in the pathways of colorectal carcinogenesis*. Clin Gastroenterol Hepatol, 2005. **3**(11): p. 1115-23.
19. Souza, R.F., et al., *Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours*. Nat Genet, 1996. **14**(3): p. 255-7.
20. Lin, C.H., et al., *Molecular profile and copy number analysis of sporadic colorectal cancer in Taiwan*. J Biomed Sci, 2011. **18**: p. 36.
21. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes*. Science, 2007. **315**(5813): p. 848-53.
22. Campbell, P.J., et al., *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing*. Nat Genet, 2008. **40**(6): p. 722-9.
23. Postma, C., et al., *DNA copy number profiles of primary tumors as predictors of response to chemotherapy in advanced colorectal cancer*. Ann Oncol, 2009. **20**(6): p. 1048-56.
24. Qiao, Y., et al., *Outcome of array CGH analysis for 255 subjects with intellectual disability and search for candidate genes using bioinformatics*. Hum Genet, 2010. **128**(2): p. 179-94.
25. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers*. Methods Mol Biol, 2000. **132**: p. 365-86.
26. Aarskog, N.K. and C.A. Vedeler, *Real-time quantitative polymerase chain reaction. A new method that detects both the peripheral myelin protein 22 duplication in Charcot-Marie-Tooth type 1A disease and the peripheral myelin protein 22 deletion in hereditary neuropathy with liability to pressure palsies*. Hum Genet, 2000. **107**(5): p. 494-8.

27. Zhang, C., et al., *Deletion and down-regulation of HRH4 gene in gastric carcinomas: a potential correlation with tumor progression*. PLoS One, 2012. 7(2): p. e31207.

9. REFERENCES

1. (IARC), G., *Section of Cancer Information* 2010.
2. NCI. <http://www.cancer.gov>.
3. Das, P.M. and R. Singal, *DNA methylation and cancer*. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2004. **22**(22): p. 4632-42.
4. Dasari, V.K., et al., *DNA methylation regulates the expression of Y chromosome specific genes in prostate cancer*. J Urol, 2002. **167**(1): p. 335-8.
5. Ramsahoye, B.H., et al., *Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(10): p. 5237-42.
6. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts*. Nature, 1983. **301**(5895): p. 89-92.
7. Kim, Y.I., et al., *Global DNA hypomethylation increases progressively in cervical dysplasia and carcinoma*. Cancer, 1994. **74**(3): p. 893-9.
8. Lin, C.H., et al., *Genome-wide hypomethylation in hepatocellular carcinogenesis*. Cancer research, 2001. **61**(10): p. 4238-43.
9. Bedford, M.T. and P.D. van Helden, *Hypomethylation of DNA in pathological conditions of the human prostate*. Cancer research, 1987. **47**(20): p. 5274-6.
10. Parkin, D.M., *International variation*. Oncogene, 2004. **23**(38): p. 6329-40.
11. Parkin, D.M., et al., *Global cancer statistics, 2002*. CA: a cancer journal for clinicians, 2005. **55**(2): p. 74-108.
12. <http://www.medicinenet.com>.
13. Wolpin, B.M. and R.J. Mayer, *Systemic treatment of colorectal cancer*. Gastroenterology, 2008. **134**(5): p. 1296-310.
14. Labianca, R., et al., *Colon cancer*. Critical reviews in oncology/hematology, 2010. **74**(2): p. 106-33.
15. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**(5): p. 759-67.

16. Bellacosa, A., *Genetic hits and mutation rate in colorectal tumorigenesis: versatility of Knudson's theory and implications for cancer prevention*. Genes, chromosomes & cancer, 2003. **38**(4): p. 382-8.
17. Tomlinson, I.P., et al., *A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3*. Nature genetics, 2008. **40**(5): p. 623-30.
18. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma*. Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(4): p. 820-3.
19. Loukola, A., et al., *Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer*. The American journal of pathology, 1999. **155**(6): p. 1849-53.
20. Tomlinson, I.P., R. Roylance, and R.S. Houlston, *Two hits revisited again*. Journal of medical genetics, 2001. **38**(2): p. 81-5.
21. Costello, J.F. and C. Plass, *Methylation matters*. Journal of medical genetics, 2001. **38**(5): p. 285-303.
22. Marsh, D. and R. Zori, *Genetic insights into familial cancers-- update and recent discoveries*. Cancer letters, 2002. **181**(2): p. 125-64.
23. Lynch, H.T., et al., *Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review*. Gastroenterology, 1993. **104**(5): p. 1535-49.
24. Aaltonen, L.A., et al., *Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease*. The New England journal of medicine, 1998. **338**(21): p. 1481-7.
25. Samowitz, W.S., et al., *The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer*. Gastroenterology, 2001. **121**(4): p. 830-8.
26. Syngal, S., et al., *Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing*. JAMA : the journal of the American Medical Association, 1999. **282**(3): p. 247-53.
27. Kovac, M., et al., *Familial colorectal cancer: eleven years of data from a registry program in Switzerland*. Fam Cancer, 2011. **10**(3): p. 605-16.

28. Lynch, H.T., et al., *Natural history of colorectal cancer in hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II)*. Diseases of the colon and rectum, 1988. **31**(6): p. 439-44.
29. Vasen, H.F., et al., *Surveillance in hereditary nonpolyposis colorectal cancer: an international cooperative study of 165 families. The International Collaborative Group on HNPCC*. Diseases of the colon and rectum, 1993. **36**(1): p. 1-4.
30. Aarnio, M., et al., *Cancer risk in mutation carriers of DNA-mismatch-repair genes*. International journal of cancer. Journal international du cancer, 1999. **81**(2): p. 214-8.
31. Bertario, L., et al., *Survival of patients with hereditary colorectal cancer: comparison of HNPCC and colorectal cancer in FAP patients with sporadic colorectal cancer*. International journal of cancer. Journal international du cancer, 1999. **80**(2): p. 183-7.
32. Vasen, H.F., et al., *New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC*. Gastroenterology, 1999. **116**(6): p. 1453-6.
33. Bellacosa, A., et al., *Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects*. American journal of medical genetics, 1996. **62**(4): p. 353-64.
34. Benatti, P., et al., *Tumour spectrum in hereditary non-polyposis colorectal cancer (HNPCC) and in families with "suspected HNPCC". A population-based study in northern Italy. Colorectal Cancer Study Group*. International journal of cancer. Journal international du cancer, 1993. **54**(3): p. 371-7.
35. Vasen, H.F., et al., *The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)*. Diseases of the colon and rectum, 1991. **34**(5): p. 424-5.
36. Rodriguez-Bigas, M.A., et al., *A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines*. Journal of the National Cancer Institute, 1997. **89**(23): p. 1758-62.
37. Umar, A., et al., *Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability*. Journal of the National Cancer Institute, 2004. **96**(4): p. 261-8.

38. Giardiello, F.M., J.D. Brensinger, and G.M. Petersen, *AGA technical review on hereditary colorectal cancer and genetic testing*. *Gastroenterology*, 2001. **121**(1): p. 198-213.
39. Lindor, N.M., et al., *Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors*. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 2002. **20**(4): p. 1043-8.
40. Bellizzi AM, F.W., *Colorectal cancer due to deficiency in DNA mismatch repair function: a review*. *Adv Anat Pathol*, 2009 Nov. **16**(6):
: p. 405-17.
41. Burke, W., et al., *Recommendations for follow-up care of individuals with an inherited predisposition to cancer. I. Hereditary nonpolyposis colon cancer. Cancer Genetics Studies Consortium*. *JAMA : the journal of the American Medical Association*, 1997. **277**(11): p. 915-9.
42. Winawer, S., et al., *Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence*. *Gastroenterology*, 2003. **124**(2): p. 544-60.
43. Strate, L.L. and S. Syngal, *Hereditary colorectal cancer syndromes*. *Cancer causes & control : CCC*, 2005. **16**(3): p. 201-13.
44. Wagner, R., Jr. and M. Meselson, *Repair tracts in mismatched DNA heteroduplexes*. *Proceedings of the National Academy of Sciences of the United States of America*, 1976. **73**(11): p. 4135-9.
45. Cox, E.C., G.E. Degnen, and M.L. Scheppe, *Mutator gene studies in Escherichia coli: the mutS gene*. *Genetics*, 1972. **72**(4): p. 551-67.
46. Modrich, P. and R. Lahue, *Mismatch repair in replication fidelity, genetic recombination, and cancer biology*. *Annual review of biochemistry*, 1996. **65**: p. 101-33.
47. Kolodner, R., *Biochemistry and genetics of eukaryotic mismatch repair*. *Genes & development*, 1996. **10**(12): p. 1433-42.
48. Marsischky, G.T., et al., *Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair*. *Genes & development*, 1996. **10**(4): p. 407-20.
49. Genschel, J., et al., *Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha*. *The Journal of biological chemistry*, 1998. **273**(31): p. 19895-901.

50. Kramer, W., et al., *Cloning and nucleotide sequence of DNA mismatch repair gene PMS1 from Saccharomyces cerevisiae: homology of PMS1 to procaryotic MutL and HexB*. Journal of bacteriology, 1989. **171**(10): p. 5339-46.
51. Boland, C.R. and A. Goel, *Microsatellite instability in colorectal cancer*. Gastroenterology, 2010. **138**(6): p. 2073-2087 e3.
52. Ayyagari, R., et al., *A mutational analysis of the yeast proliferating cell nuclear antigen indicates distinct roles in DNA replication and DNA repair*. Molecular and cellular biology, 1995. **15**(8): p. 4420-9.
53. Umar, A., et al., *Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis*. Cell, 1996. **87**(1): p. 65-73.
54. Flores-Rozas, H., D. Clark, and R.D. Kolodner, *Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mismatch recognition complex*. Nature genetics, 2000. **26**(3): p. 375-8.
55. Bellacosa, A., *Functional interactions and signaling properties of mammalian DNA mismatch repair proteins*. Cell death and differentiation, 2001. **8**(11): p. 1076-92.
56. Peltomaki, P., *Role of DNA mismatch repair defects in the pathogenesis of human cancer*. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2003. **21**(6): p. 1174-9.
57. Helleman, J., et al., *Mismatch repair and treatment resistance in ovarian cancer*. BMC cancer, 2006. **6**: p. 201.
58. Kim, N.G., et al., *Frameshift mutations at coding mononucleotide repeats of the hRAD50 gene in gastrointestinal carcinomas with microsatellite instability*. Cancer research, 2001. **61**(1): p. 36-8.
59. Mori, Y., et al., *Instability typing: comprehensive identification of frameshift mutations caused by coding region microsatellite instability*. Cancer research, 2001. **61**(16): p. 6046-9.
60. Wei, K., R. Kucherlapati, and W. Edelmann, *Mouse models for human DNA mismatch-repair gene defects*. Trends in molecular medicine, 2002. **8**(7): p. 346-53.
61. Boland, C.R., et al., *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of*

- international criteria for the determination of microsatellite instability in colorectal cancer*. Cancer research, 1998. **58**(22): p. 5248-57.
62. Duval, A., et al., *Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers*. Human molecular genetics, 2001. **10**(5): p. 513-8.
 63. Bronner, C.E., et al., *Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer*. Nature, 1994. **368**(6468): p. 258-61.
 64. Papadopoulos, N., et al., *Mutation of a mutL homolog in hereditary colon cancer*. Science, 1994. **263**(5153): p. 1625-9.
 65. Fishel, R., et al., *The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer*. Cell, 1994. **77**(1): p. 1 p following 166.
 66. Markowitz, S., et al., *Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability*. Science, 1995. **268**(5215): p. 1336-8.
 67. Souza, R.F., et al., *Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours*. Nature genetics, 1996. **14**(3): p. 255-7.
 68. Malkhosyan, S., et al., *Frameshift mutator mutations*. Nature, 1996. **382**(6591): p. 499-500.
 69. Rampino, N., et al., *Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype*. Science, 1997. **275**(5302): p. 967-9.
 70. Bertoni, F., et al., *CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers*. Genes, chromosomes & cancer, 1999. **26**(2): p. 176-80.
 71. Calin, G., et al., *The coding region of the Bloom syndrome BLM gene and of the CBL proto-oncogene is mutated in genetically unstable sporadic gastrointestinal tumors*. Cancer research, 1998. **58**(17): p. 3777-81.
 72. Wicking, C., et al., *CDX2, a human homologue of Drosophila caudal, is mutated in both alleles in a replication error positive colorectal cancer*. Oncogene, 1998. **17**(5): p. 657-9.

73. Schwartz, S., Jr., et al., *Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype*. *Cancer research*, 1999. **59**(12): p. 2995-3002.
74. Riccio, A., et al., *The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability*. *Nature genetics*, 1999. **23**(3): p. 266-8.
75. Duval, A., et al., *Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability*. *Cancer research*, 1999. **59**(17): p. 4213-5.
76. Yamamoto, H., et al., *Frameshift mutations in Fas, Apaf-1, and Bcl-10 in gastro-intestinal cancer of the microsatellite mutator phenotype*. *Cell death and differentiation*, 2000. **7**(2): p. 238-9.
77. Liu, W., et al., *Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling*. *Nature genetics*, 2000. **26**(2): p. 146-7.
78. Loukola, A., et al., *Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer*. *The American journal of pathology*, 2000. **157**(2): p. 347-52.
79. Guanti, G., et al., *Involvement of PTEN mutations in the genetic pathways of colorectal cancerogenesis*. *Human molecular genetics*, 2000. **9**(2): p. 283-7.
80. Chadwick, R.B., et al., *Candidate tumor suppressor RIZ is frequently involved in colorectal carcinogenesis*. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(6): p. 2662-7.
81. Ikenoue, T., et al., *Frameshift mutations at mononucleotide repeats in RAD50 recombinational DNA repair gene in colorectal cancers with microsatellite instability*. *Japanese journal of cancer research : Gann*, 2001. **92**(6): p. 587-91.
82. Woerner, S.M., et al., *Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells*. *International journal of cancer. Journal international du cancer*, 2001. **93**(1): p. 12-9.

83. Thorstensen, L., et al., *WNT1 inducible signaling pathway protein 3, WISP-3, a novel target gene in colorectal carcinomas with microsatellite instability*. *Gastroenterology*, 2001. **121**(6): p. 1275-80.
84. Duval, A. and R. Hamelin, *Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability*. *Cancer research*, 2002. **62**(9): p. 2447-54.
85. Popat, S., R. Hubner, and R.S. Houlston, *Systematic review of microsatellite instability and colorectal cancer prognosis*. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 2005. **23**(3): p. 609-18.
86. Lynch, H.T. and A. de la Chapelle, *Hereditary colorectal cancer*. *The New England journal of medicine*, 2003. **348**(10): p. 919-32.
87. Gala, M. and D.C. Chung, *Hereditary colon cancer syndromes*. *Seminars in oncology*, 2011. **38**(4): p. 490-9.
88. Campbell, W.J., R.A. Spence, and T.G. Parks, *Familial adenomatous polyposis*. *The British journal of surgery*, 1994. **81**(12): p. 1722-33.
89. Bulow, S., et al., *The incidence rate of familial adenomatous polyposis. Results from the Danish Polyposis Register*. *International journal of colorectal disease*, 1996. **11**(2): p. 88-91.
90. Powell, S.M., et al., *Molecular diagnosis of familial adenomatous polyposis*. *The New England journal of medicine*, 1993. **329**(27): p. 1982-7.
91. Powell, S.M., et al., *APC mutations occur early during colorectal tumorigenesis*. *Nature*, 1992. **359**(6392): p. 235-7.
92. Sieber, O.M., et al., *Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH*. *The New England journal of medicine*, 2003. **348**(9): p. 791-9.
93. Heinimann, K., et al., *Phenotypic differences in familial adenomatous polyposis based on APC gene mutation status*. *Gut*, 1998. **43**(5): p. 675-9.
94. Paul, P., et al., *Identical APC exon 15 mutations result in a variable phenotype in familial adenomatous polyposis*. *Human molecular genetics*, 1993. **2**(7): p. 925-31.
95. Hamilton, S.R., et al., *The molecular basis of Turcot's syndrome*. *N Engl J Med*, 1995. **332**(13): p. 839-47.

96. Jass, J.R., et al., *Juvenile polyposis--a precancerous condition*. *Histopathology*, 1988. **13**(6): p. 619-30.
97. Giardiello, F.M., et al., *Colorectal neoplasia in juvenile polyposis or juvenile polyps*. *Arch Dis Child*, 1991. **66**(8): p. 971-5.
98. Aretz, S., et al., *High proportion of large genomic deletions and a genotype phenotype update in 80 unrelated families with juvenile polyposis syndrome*. *Journal of medical genetics*, 2007. **44**(11): p. 702-9.
99. Howe, J.R., et al., *Mutations in the SMAD4/DPC4 gene in juvenile polyposis*. *Science*, 1998. **280**(5366): p. 1086-8.
100. Olschwang, S., et al., *PTEN germ-line mutations in juvenile polyposis coli*. *Nature genetics*, 1998. **18**(1): p. 12-4.
101. Howe, J.R., F.A. Mitros, and R.W. Summers, *The risk of gastrointestinal carcinoma in familial juvenile polyposis*. *Ann Surg Oncol*, 1998. **5**(8): p. 751-6.
102. McGarrity, T.J. and C. Amos, *Peutz-Jeghers syndrome: clinicopathology and molecular alterations*. *Cell Mol Life Sci*, 2006. **63**(18): p. 2135-44.
103. Wang, Z.J., et al., *Allele loss and mutation screen at the Peutz-Jeghers (LKB1) locus (19p13.3) in sporadic ovarian tumours*. *British journal of cancer*, 1999. **80**(1-2): p. 70-2.
104. Su, G.H., et al., *Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers*. *The American journal of pathology*, 1999. **154**(6): p. 1835-40.
105. Rowan, A., et al., *Somatic mutations in the Peutz-Jeghers (LKB1/STK11) gene in sporadic malignant melanomas*. *J Invest Dermatol*, 1999. **112**(4): p. 509-11.
106. Trau, H., et al., *Peutz-Jeghers syndrome and bilateral breast carcinoma*. *Cancer*, 1982. **50**(4): p. 788-92.
107. Giardiello, F.M., et al., *Very high risk of cancer in familial Peutz-Jeghers syndrome*. *Gastroenterology*, 2000. **119**(6): p. 1447-53.
108. Jenne, D.E., et al., *Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase*. *Nature genetics*, 1998. **18**(1): p. 38-43.
109. Hemminki, A., et al., *A serine/threonine kinase gene defective in Peutz-Jeghers syndrome*. *Nature*, 1998. **391**(6663): p. 184-7.
110. Zbuk, K.M. and C. Eng, *Hamartomatous polyposis syndromes*. *Nat Clin Pract Gastroenterol Hepatol*, 2007. **4**(9): p. 492-502.
111. Pagon RA, B.T., Dolan CR., ed. *GeneReviews*™ 1993: Seattle (WA).

112. <http://www.genecards.org>.
113. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**(1): p. 57-70.
114. Elenbaas, B., et al., *Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells*. Genes & development, 2001. **15**(1): p. 50-65.
115. Feinberg, A.P., R. Ohlsson, and S. Henikoff, *The epigenetic progenitor origin of human cancer*. Nature reviews. Genetics, 2006. **7**(1): p. 21-33.
116. Sugimura, T., et al., *Multiple genetic alterations in human carcinogenesis*. Environmental health perspectives, 1992. **98**: p. 5-12.
117. Dhillon, V.S., M. Aslam, and S.A. Husain, *The contribution of genetic and epigenetic changes in granulosa cell tumors of ovarian origin*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2004. **10**(16): p. 5537-45.
118. Manne, U., R.G. Srivastava, and S. Srivastava, *Recent advances in biomarkers for cancer diagnosis and treatment*. Drug discovery today, 2005. **10**(14): p. 965-76.
119. Wang, Y.C., et al., *Molecular diagnostic markers for lung cancer in sputum and plasma*. Annals of the New York Academy of Sciences, 2006. **1075**: p. 179-84.
120. Jackson, P.E., et al., *Prospective detection of codon 249 mutations in plasma of hepatocellular carcinoma patients*. Carcinogenesis, 2003. **24**(10): p. 1657-63.
121. Gormally, E., et al., *TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study*. Cancer research, 2006. **66**(13): p. 6871-6.
122. Olivier, M., et al., *The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(4): p. 1157-67.
123. Shigematsu, H. and A.F. Gazdar, *Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers*. International journal of cancer. Journal international du cancer, 2006. **118**(2): p. 257-62.
124. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy*. Science, 2004. **304**(5676): p. 1497-500.

125. Pao, W., et al., *EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(36): p. 13306-11.
126. Pao, W., et al., *Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain*. PLoS medicine, 2005. **2**(3): p. e73.
127. Toyooka, S., K. Kiura, and T. Mitsudomi, *EGFR mutation and response of lung cancer to gefitinib*. The New England journal of medicine, 2005. **352**(20): p. 2136; author reply 2136.
128. Shaw, R.J., et al., *Methylation enrichment pyrosequencing: combining the specificity of MSP with validation by pyrosequencing*. Nucleic acids research, 2006. **34**(11): p. e78.
129. Belinsky, S.A., *Gene-promoter hypermethylation as a biomarker in lung cancer*. Nature reviews. Cancer, 2004. **4**(9): p. 707-17.
130. Laird, P.W., *The power and the promise of DNA methylation markers*. Nature reviews. Cancer, 2003. **3**(4): p. 253-66.
131. Mitchell, P.S., et al., *Circulating microRNAs as stable blood-based markers for cancer detection*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(30): p. 10513-8.
132. Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg, *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?* Nature reviews. Genetics, 2008. **9**(2): p. 102-14.
133. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex*. Nature, 2004. **432**(7014): p. 231-5.
134. Kim, V.N. and J.W. Nam, *Genomics of microRNA*. Trends in genetics : TIG, 2006. **22**(3): p. 165-73.
135. Zeng, Y., R. Yi, and B.R. Cullen, *MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9779-84.
136. Berezikov, E., E. Cuppen, and R.H. Plasterk, *Approaches to microRNA discovery*. Nature genetics, 2006. **38** **Suppl**: p. S2-7.

137. Hausser, J., et al., *MirZ: an integrated microRNA expression atlas and target prediction resource*. Nucleic acids research, 2009. **37**(Web Server issue): p. W266-72.
138. Brennecke, J., et al., *bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila*. Cell, 2003. **113**(1): p. 25-36.
139. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-54.
140. Calin, G.A. and C.M. Croce, *MicroRNA signatures in human cancers*. Nature reviews. Cancer, 2006. **6**(11): p. 857-66.
141. Gaur, A., et al., *Characterization of microRNA expression levels and their biological correlates in human cancer cell lines*. Cancer research, 2007. **67**(6): p. 2456-68.
142. Sassen, S., E.A. Miska, and C. Caldas, *MicroRNA: implications for cancer*. Virchows Archiv : an international journal of pathology, 2008. **452**(1): p. 1-10.
143. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15524-9.
144. He, L., et al., *A microRNA polycistron as a potential human oncogene*. Nature, 2005. **435**(7043): p. 828-33.
145. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression*. Nature, 2005. **435**(7043): p. 839-43.
146. Matsubara, H., et al., *Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92*. Oncogene, 2007. **26**(41): p. 6099-105.
147. Sylvestre, Y., et al., *An E2F/miR-20a autoregulatory feedback loop*. The Journal of biological chemistry, 2007. **282**(4): p. 2135-43.
148. He, L., et al., *A microRNA component of the p53 tumour suppressor network*. Nature, 2007. **447**(7148): p. 1130-4.
149. Chang, T.C., et al., *Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis*. Molecular cell, 2007. **26**(5): p. 745-52.

150. Lu, J., et al., *MicroRNA expression profiles classify human cancers*. Nature, 2005. **435**(7043): p. 834-8.
151. Kumar, M.S., et al., *Impaired microRNA processing enhances cellular transformation and tumorigenesis*. Nature genetics, 2007. **39**(5): p. 673-7.
152. Bustin, M. and R. Reeves, *High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function*. Progress in nucleic acid research and molecular biology, 1996. **54**: p. 35-100.
153. Bianchi, M.E. and M. Beltrame, *Upwardly mobile proteins. Workshop: the role of HMG proteins in chromatin structure, gene expression and neoplasia*. EMBO reports, 2000. **1**(2): p. 109-14.
154. Postnikov, Y.V. and M. Bustin, *Reconstitution of high mobility group 14/17 proteins into nucleosomes and chromatin*. Methods in enzymology, 1999. **304**: p. 133-55.
155. Reeves, R., *Structure and function of the HMGI(Y) family of architectural transcription factors*. Environmental health perspectives, 2000. **108 Suppl 5**: p. 803-9.
156. Reeves, R., *Molecular biology of HMGA proteins: hubs of nuclear function*. Gene, 2001. **277**(1-2): p. 63-81.
157. Thomas, J.O., *HMG1 and 2: architectural DNA-binding proteins*. Biochemical Society transactions, 2001. **29**(Pt 4): p. 395-401.
158. Fedele, M., et al., *Role of the high mobility group A proteins in human lipomas*. Carcinogenesis, 2001. **22**(10): p. 1583-91.
159. Reeves, R. and L. Beckerbauer, *HMGI/Y proteins: flexible regulators of transcription and chromatin structure*. Biochimica et biophysica acta, 2001. **1519**(1-2): p. 13-29.
160. Battista, S., et al., *Loss of Hmga1 gene function affects embryonic stem cell lympho-hematopoietic differentiation*. FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2003. **17**(11): p. 1496-8.
161. Munshi, N., et al., *Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome*. Molecular cell, 1998. **2**(4): p. 457-67.

162. Donato, G., et al., *High mobility group A1 expression correlates with the histological grade of human glial tumors*. *Oncology reports*, 2004. **11**(6): p. 1209-13.
163. Hess, J.L., *Chromosomal translocations in benign tumors: the HMGI proteins*. *American journal of clinical pathology*, 1998. **109**(3): p. 251-61.
164. Berlingieri, M.T., et al., *Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells*. *Molecular and cellular biology*, 1995. **15**(3): p. 1545-53.
165. Wood, L.J., et al., *The oncogenic properties of the HMG-I gene family*. *Cancer research*, 2000. **60**(15): p. 4256-61.
166. Reya, T., et al., *Stem cells, cancer, and cancer stem cells*. *Nature*, 2001. **414**(6859): p. 105-11.
167. Domen, J., K.L. Gandy, and I.L. Weissman, *Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation*. *Blood*, 1998. **91**(7): p. 2272-82.
168. Domen, J. and I.L. Weissman, *Hematopoietic stem cells need two signals to prevent apoptosis; BCL-2 can provide one of these, Kit/c-Kit signaling the other*. *J Exp Med*, 2000. **192**(12): p. 1707-18.
169. Taipale, J. and P.A. Beachy, *The Hedgehog and Wnt signalling pathways in cancer*. *Nature*, 2001. **411**(6835): p. 349-54.
170. Heppner, G.H., *Tumor heterogeneity*. *Cancer research*, 1984. **44**(6): p. 2259-65.
171. Fidler, I.J. and M.L. Kripke, *Metastasis results from preexisting variant cells within a malignant tumor*. *Science*, 1977. **197**(4306): p. 893-5.
172. Fidler, I.J. and I.R. Hart, *Biological diversity in metastatic neoplasms: origins and implications*. *Science*, 1982. **217**(4564): p. 998-1003.
173. Nowell, P.C., *Mechanisms of tumor progression*. *Cancer research*, 1986. **46**(5): p. 2203-7.
174. Southam, C.M., et al., *Hetero- transplantation of human cell lines from Burkitt's tumors and acute leukemia into newborn rats*. *Cancer*, 1969. **23**(2): p. 281-99.
175. Hamburger, A.W. and S.E. Salmon, *Primary bioassay of human tumor stem cells*. *Science*, 1977. **197**(4302): p. 461-3.

176. De Plaen, E., et al., *Structure, chromosomal localization, and expression of 12 genes of the MAGE family*. Immunogenetics, 1994. **40**(5): p. 360-9.
177. Rogner, U.C., et al., *The melanoma antigen gene (MAGE) family is clustered in the chromosomal band Xq28*. Genomics, 1995. **29**(3): p. 725-31.
178. Lurquin, C., et al., *Two members of the human MAGEB gene family located in Xp21.3 are expressed in tumors of various histological origins*. Genomics, 1997. **46**(3): p. 397-408.
179. Muscatelli, F., et al., *Isolation and characterization of a MAGE gene family in the Xp21.3 region*. Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(11): p. 4987-91.
180. Lucas, S., E. De Plaen, and T. Boon, *MAGE-B5, MAGE-B6, MAGE-C2, and MAGE-C3: four new members of the MAGE family with tumor-specific expression*. International journal of cancer. Journal international du cancer, 2000. **87**(1): p. 55-60.
181. Simpson, A.J., et al., *Cancer/testis antigens, gametogenesis and cancer*. Nature reviews. Cancer, 2005. **5**(8): p. 615-25.
182. Chomez, P., et al., *The SMAGE gene family is expressed in post-meiotic spermatids during mouse germ cell differentiation*. Immunogenetics, 1996. **43**(1-2): p. 97-100.
183. Jungbluth, A.A., et al., *Expression of MAGE-antigens in normal tissues and cancer*. International journal of cancer. Journal international du cancer, 2000. **85**(4): p. 460-5.
184. Jang, S.J., et al., *Activation of melanoma antigen tumor antigens occurs early in lung carcinogenesis*. Cancer research, 2001. **61**(21): p. 7959-63.
185. Andrade, V.C., et al., *Prognostic impact of cancer/testis antigen expression in advanced stage multiple myeloma patients*. Cancer Immun, 2008. **8**: p. 2.
186. Picard, V., et al., *MAGE-A9 mRNA and protein expression in bladder cancer*. International journal of cancer. Journal international du cancer, 2007. **120**(10): p. 2170-7.
187. Otte, M., et al., *MAGE-A gene expression pattern in primary breast cancer*. Cancer research, 2001. **61**(18): p. 6682-7.
188. Bandic, D., et al., *Expression and possible prognostic role of MAGE-A4, NY-ESO-1, and HER-2 antigens in women with relapsing invasive ductal breast*

- cancer: retrospective immunohistochemical study.* Croat Med J, 2006. **47**(1): p. 32-41.
189. Bergeron, A., et al., *High frequency of MAGE-A4 and MAGE-A9 expression in high-risk bladder cancer.* International journal of cancer. Journal international du cancer, 2009. **125**(6): p. 1365-71.
190. Brasseur, F., et al., *Expression of MAGE genes in primary and metastatic cutaneous melanoma.* International journal of cancer. Journal international du cancer, 1995. **63**(3): p. 375-80.
191. Salehi, A.H., et al., *NRAGE, a novel MAGE protein, interacts with the p75 neurotrophin receptor and facilitates nerve growth factor-dependent apoptosis.* Neuron, 2000. **27**(2): p. 279-88.
192. Masuda, Y., et al., *Dlxin-1, a novel protein that binds Dlx5 and regulates its transcriptional function.* The Journal of biological chemistry, 2001. **276**(7): p. 5331-8.
193. Castelli, C., et al., *T-cell recognition of melanoma-associated antigens.* J Cell Physiol, 2000. **182**(3): p. 323-31.
194. Atanackovic, D., et al., *Vaccine-induced CD4+ T cell responses to MAGE-3 protein in lung cancer patients.* J Immunol, 2004. **172**(5): p. 3289-96.
195. Nestle, F.O., et al., *Vaccination of melanoma patients with peptide- or tumor lysate-pulsed dendritic cells.* Nat Med, 1998. **4**(3): p. 328-32.
196. Marchand, M., et al., *Tumor regressions observed in patients with metastatic melanoma treated with an antigenic peptide encoded by gene MAGE-3 and presented by HLA-A1.* International journal of cancer. Journal international du cancer, 1999. **80**(2): p. 219-30.
197. Thurner, B., et al., *Vaccination with mage-3A1 peptide-pulsed mature, monocyte-derived dendritic cells expands specific cytotoxic T cells and induces regression of some metastases in advanced stage IV melanoma.* J Exp Med, 1999. **190**(11): p. 1669-78.
198. Deloulme, J.C., et al., *The proto-oncoprotein EWS binds calmodulin and is phosphorylated by protein kinase C through an IQ domain.* The Journal of biological chemistry, 1997. **272**(43): p. 27369-77.
199. Lodomery, M. and G. Dellaire, *Multifunctional zinc finger proteins in development and disease.* Ann Hum Genet, 2002. **66**(Pt 5-6): p. 331-42.

200. Bertolotti, A., et al., *EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes.* Molecular and cellular biology, 1998. **18**(3): p. 1489-97.
201. Knoop, L.L. and S.J. Baker, *The splicing factor U1C represses EWS/FLI-mediated transactivation.* The Journal of biological chemistry, 2000. **275**(32): p. 24865-71.
202. Zhang, D., A.J. Paley, and G. Childs, *The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription.* The Journal of biological chemistry, 1998. **273**(29): p. 18086-91.
203. Chansky, H.A., et al., *Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein.* Cancer research, 2001. **61**(9): p. 3586-90.
204. Dutertre, M., et al., *Cotranscriptional exon skipping in the genotoxic stress response.* Nat Struct Mol Biol, 2010. **17**(11): p. 1358-66.
205. Sanchez, G., et al., *Coupled alteration of transcription and splicing by a single oncogene: boosting the effect on cyclin D1 activity.* Cell Cycle, 2008. **7**(15): p. 2299-305.
206. Paronetto, M.P., B. Minana, and J. Valcarcel, *The Ewing sarcoma protein regulates DNA damage-induced alternative splicing.* Molecular cell, 2011. **43**(3): p. 353-68.
207. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.
208. Li, H., et al., *Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development.* J Clin Invest, 2007. **117**(5): p. 1314-23.
209. Delattre, O., et al., *Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.* Nature, 1992. **359**(6391): p. 162-5.
210. Zucman, J., et al., *Cloning and characterization of the Ewing's sarcoma and peripheral neuroepithelioma t(11;22) translocation breakpoints.* Genes, chromosomes & cancer, 1992. **5**(4): p. 271-7.
211. Sankar, S. and S.L. Lessnick, *Promiscuous partnerships in Ewing's sarcoma.* Cancer Genet, 2011. **204**(7): p. 351-65.

212. Petermann, R., et al., *Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II*. *Oncogene*, 1998. **17**(5): p. 603-10.
213. Zinszner, H., R. Albalat, and D. Ron, *A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP*. *Genes & development*, 1994. **8**(21): p. 2513-26.
214. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development*. *N Engl J Med*, 1988. **319**(9): p. 525-32.
215. Jass, J.R., J. Young, and B.A. Leggett, *Evolution of colorectal cancer: change of pace and change of direction*. *J Gastroenterol Hepatol*, 2002. **17**(1): p. 17-26.
216. *Comprehensive molecular characterization of human colon and rectal cancer*. *Nature*, 2012. **487**(7407): p. 330-7.
217. Emi, M., et al., *Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer*. *Cancer research*, 1992. **52**(19): p. 5368-72.
218. Birnbaum, D., et al., *Chromosome arm 8p and cancer: a fragile hypothesis*. *Lancet Oncol*, 2003. **4**(10): p. 639-42.
219. Roessler, S., et al., *Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival*. *Gastroenterology*, 2012. **142**(4): p. 957-966 e12.
220. Krasinskas, A.M., *EGFR Signaling in Colorectal Carcinoma*. *Patholog Res Int*, 2011. **2011**: p. 932932.
221. Fusco, A. and M. Fedele, *Roles of HMGA proteins in cancer*. *Nature reviews. Cancer*, 2007. **7**(12): p. 899-910.
222. Zhou, X., et al., *Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C*. *Nature*, 1995. **376**(6543): p. 771-4.
223. Chiappetta, G., et al., *High level expression of the HMGI (Y) gene during embryonic development*. *Oncogene*, 1996. **13**(11): p. 2439-46.
224. Rommel, B., et al., *HMGI-C, a member of the high mobility group family of proteins, is expressed in hematopoietic stem cells and in leukemic cells*. *Leuk Lymphoma*, 1997. **26**(5-6): p. 603-7.
225. Anand, A. and K. Chada, *In vivo modulation of Hmgic reduces obesity*. *Nature genetics*, 2000. **24**(4): p. 377-80.

226. Nishino, J., et al., *Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf Expression*. Cell, 2008. **135**(2): p. 227-39.
227. Battista, S., et al., *The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis*. Cancer research, 1999. **59**(19): p. 4793-7.
228. Fedele, M., et al., *Haploinsufficiency of the Hmga1 gene causes cardiac hypertrophy and myelo-lymphoproliferative disorders in mice*. Cancer research, 2006. **66**(5): p. 2536-43.
229. Foti, D., et al., *Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice*. Nat Med, 2005. **11**(7): p. 765-73.
230. Chiefari, E., et al., *The cAMP-HMGA1-RBP4 system: a novel biochemical pathway for modulating glucose homeostasis*. BMC Biol, 2009. **7**: p. 24.
231. Sreekantaiah, C., et al., *Cytogenetic profile of 109 lipomas*. Cancer research, 1991. **51**(1): p. 422-33.
232. Schoenmakers, E.F., et al., *Recurrent rearrangements in the high mobility group protein gene, HMGI-C, in benign mesenchymal tumours*. Nature genetics, 1995. **10**(4): p. 436-44.
233. Ashar, H.R., et al., *Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains*. Cell, 1995. **82**(1): p. 57-65.
234. Ligon, A.H., et al., *Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas*. Am J Hum Genet, 2005. **76**(2): p. 340-8.
235. Buysse, K., et al., *The 12q14 microdeletion syndrome: additional patients and further evidence that HMGA2 is an important genetic determinant for human height*. Eur J Med Genet, 2009. **52**(2-3): p. 101-7.
236. Weedon, M.N., et al., *Genome-wide association analysis identifies 20 loci that influence adult height*. Nature genetics, 2008. **40**(5): p. 575-83.
237. Fedele, M. and A. Fusco, *HMGA and cancer*. Biochimica et biophysica acta, 2010. **1799**(1-2): p. 48-54.

238. Piscuoglio, S., et al., *HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma*. *Histopathology*, 2012. **60**(3): p. 397-404.
239. Reeves, R. and J.E. Adair, *Role of high mobility group (HMG) chromatin proteins in DNA repair*. *DNA Repair (Amst)*, 2005. **4**(8): p. 926-38.
240. Borrmann, L., et al., *High mobility group A2 protein and its derivatives bind a specific region of the promoter of DNA repair gene ERCC1 and modulate its activity*. *Nucleic acids research*, 2003. **31**(23): p. 6841-51.
241. Baldassarre, G., et al., *Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma*. *Molecular and cellular biology*, 2003. **23**(7): p. 2225-38.
242. Baldassarre, G., et al., *HMGA1 protein expression sensitizes cells to cisplatin-induced cell death*. *Oncogene*, 2005. **24**(45): p. 6809-19.
243. Boo, L.M., et al., *High mobility group A2 potentiates genotoxic stress in part through the modulation of basal and DNA damage-dependent phosphatidylinositol 3-kinase-related protein kinase activation*. *Cancer research*, 2005. **65**(15): p. 6622-30.
244. Muller-Tidow, C., et al., *The cyclin A1-CDK2 complex regulates DNA double-strand break repair*. *Molecular and cellular biology*, 2004. **24**(20): p. 8917-28.
245. Sang, M., et al., *Melanoma-associated antigen genes - an update*. *Cancer letters*, 2011. **302**(2): p. 85-90.
246. Scanlan, M.J., et al., *Cancer/testis antigens: an expanding family of targets for cancer immunotherapy*. *Immunol Rev*, 2002. **188**: p. 22-32.
247. Meek, D.W. and L. Marcar, *MAGE-A antigens as targets in tumour therapy*. *Cancer letters*, 2012.
248. Schultz-Thater, E., et al., *MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies*. *International journal of cancer. Journal international du cancer*, 2011. **129**(5): p. 1137-48.
249. De Smet, C., et al., *Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene*. *Immunogenetics*, 1995. **42**(4): p. 282-90.
250. De Smet, C., et al., *DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter*. *Molecular and cellular biology*, 1999. **19**(11): p. 7327-35.

251. Wilson, E.M., *Androgen receptor molecular biology and potential targets in prostate cancer*. Ther Adv Urol, 2010. **2**(3): p. 105-17.
252. Weeraratne, S.D., et al., *miR-34a confers chemosensitivity through modulation of MAGE-A and p53 in medulloblastoma*. Neuro Oncol, 2011. **13**(2): p. 165-75.
253. Wischnewski, F., K. Pantel, and H. Schwarzenbach, *Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells*. Mol Cancer Res, 2006. **4**(5): p. 339-49.
254. Zhu, X., S.L. Asa, and S. Ezzat, *Fibroblast growth factor 2 and estrogen control the balance of histone 3 modifications targeting MAGE-A3 in pituitary neoplasia*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2008. **14**(7): p. 1984-96.
255. Doyle, J.M., et al., *MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases*. Molecular cell, 2010. **39**(6): p. 963-74.
256. Yang, B., et al., *MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines*. Cancer research, 2007. **67**(20): p. 9954-62.
257. Monte, M., et al., *MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(30): p. 11160-5.
258. Marcar, L., et al., *Mage-A cancer/testis antigens inhibit p53 function by blocking its interaction with chromatin*. Cancer research, 2010. **70**(24): p. 10362-70.
259. Nardiello, T., et al., *MAGE-A inhibits apoptosis in proliferating myeloma cells through repression of Bax and maintenance of survivin*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(13): p. 4309-19.

10. CURRICULUM VITAE

PERSONAL INFORMATION

Name Salvatore Piscuoglio
Address Strassburgerallee , 29, 4055 Basel, Switzelan
Telephone +41774303611
E-mail s.piscuoglio@unibas.ch - s.piscuoglio@gmail.com

Nationality Italian

Date of birth 23/02/1982

WORK EXPERIENCE

- Dates (from – to) 01-01-2009 – till now
- Name and address of employer Department of Biomedicine (University of Basel),
Mattenstrasse 28, CH-4058 Basel.
- Type of business or sector Translational research
- Occupation or position held PhD Student (Human Genetics group, PD Karl
Heinimann)
- Main activities and responsibilities MicroRNA, DNA, RNA extraction and sequencing,
library generation and next generation sequencing,
microsatellite analysis, Sanger sequencing, statistical
analysis, immuhistochemistry, tissue microarray
evaluation, long range PCR, real time PCR, MLPA.

WORK EXPERIENCE

- Dates (from – to) 01-04-2008 – 31-12-2008
- Name and address of employer Institute of Pathology (University of Basel), Division of Molecular Pathology (chief Luigi M.Terracciano), Schönbeinstrasse 40. CH-4031
- Type of business or sector Clinical research
- Occupation or position held Research Fellow
- Main activities and responsibilities Tissue micro array building, FISH, RNA and DNA extraction, ISH, microscopy analysis, statistical analysis.

WORK EXPERIENCE

- Dates (from – to) 04/11/2002 - 21/2/2008
- Name and address of employer Department of Biomorphological and Functional Sciences, Section of Pathology, University of Naples "Federico II", Naples, Italy.
Department of Molecular and Cellular Pathology "L.Califano" University of Naples "Federico II").
- Type of business or sector Clinical research
- Occupation or position held Research Fellow
- Main activities and responsibilities Static analysis, DNAploidy, immunohistochemistry reaction, special stain reactions, macroscopic and microscopic analysis, immunofluorescence, PCR, Whole Mount reaction, In situ Hybridization, Rt-Pcr, Fish, Western Blotting, Southern Blot, Mini and Maxi prep.

EDUCATION AND TRAINING

- Dates (from – to) 20/12/2005 - 26/03/2008
- Name and type of organisation providing education and training University of Naples “Federico II” Faculty of Biotechnological Science Medical (curriculum)
- Principal subjects/occupational skills covered Molecular biology, genetics, gene therapy, cellular biology
- Title of qualification awarded Master of Science

Thesis (MSc): **“Chromatin Assembly Factor-1, a marker of prostate cancer progression?”**

- Dates (from – to) 19/09/2001 - 20/12/2005
- Name and type of organisation providing education and training University of Naples “Federico II” Faculty of Biotechnological Science
- Principal subjects/occupational skills covered Biochemistry, molecular and genetic basis of diseases, pathology.
- Title of qualification awarded Bachelor of Science

Thesis (BSc): **“Morphometric evaluation of chromatin texture and CAF-1 expression in oral cancer”**

**PERSONAL SKILLS
AND COMPETENCES**

MOTHER TONGUE

ITALIAN

OTHER LANGUAGES

English

- Reading skills
- Writing skills
- Verbal skills

Fluent
Fluent
Fluent

SOCIAL SKILLS
AND COMPETENCES

I've worked several times in projects where collaboration with other members of the group was essential; furthermore, for 6 years (2000-2006) I got experience as tourist animator and club manager.

ORGANISATIONAL SKILLS
AND COMPETENCES

When I was the superintendent of a staff of 15 people I managed to arrange their jobs and activities, recently I've worked as a responsible for human resources of a magazine.

TECHNICAL SKILLS
AND COMPETENCES

I have good experience in using the main laboratory's tools and I have an excellent familiarity in using computers (Windows, Mac, Ms Office, Macromedia Studio, SPSS for statistic analysis, Photoshop, Statview.)

ARTISTIC SKILLS
AND COMPETENCES

I'm Sound engineer and Dj, and I play guitar

List of Publications:

1. **Piscuoglio S**, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis, Terracciano L, Fusco A and Karamitopoulou E. **HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma**. Histopathology. 2012 Feb;60(3):397-404. doi: 10.1111/j.1365-2559.2011.04121.x.
2. **Piscuoglio S***, Lehmann F*, Zlobec I, Tornillo L, Dietmaier W, Hartmann A, Wünsch PH, Sessa F, MD, Rümmele P, Baumhoer D, Terracciano L. **Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumors of the Ampulla of Vater**. J Clin Pathol. 2012 Feb;65(2):140-5. Epub 2011 Nov 30.
3. Zlobec I, Karamitopoulou E, Terracciano L, **Piscuoglio S**, Iezzi G, Muraro MG, Spagnoli G, Baker K, Tzankov A, Lugli A. **TIA-1 cytotoxic granule-associated RNA binding protein improves the prognostic performance of CD8 in mismatch repair-proficient colorectal cancer**. PLoS One. 2010 Dec 10;5(12):e14282.
4. Schultz-Thater E*, **Piscuoglio S***, Iezzi G, Le Magnen C, Zajac P, Carafa V, Terracciano L, Tornillo L, Spagnoli GC. **MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies**. Int J Cancer. 2010 Nov 15. (equally contributed)
5. Zamparese R, Pannone G, Santoro A, Lo Muzio L, Corsi F, Pedicillo MC, Scillitani EL, Tortorella S, Staibano S, **Piscuoglio S**, Lo Russo L, Bufo P.

Survivin expression in renal cell carcinoma. Cancer Invest. 2008 Nov;26(9):929-35.

Manuscript submitted or in preparation:

1. Loeffler S, Kovac M, **Piscuoglio S**, Zuklys S, Keller M, Hess D, Truninger L, Terracciano L, Heinimann K, Schär P, Holländer G. **The 8p21.3 encoded Shoca-2 represses STAT3 activation and acts as a tumor suppressor in colorectal cancer.** Manuscript in preparation
2. Shivendra Kishore*, **Salvatore Piscuoglio***, Michal Kovac*, Annette Gylling, Friedel Wenzel, Francesca Trapani, Hans Joerg Altermatt, Valentina Mele, Giancarlo Marra, Päivi Peltomäki, Luigi Terracciano, Mihaela Zavolan and Karl Heinimann. **3'UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancers.** Manuscript in preparation. (equally contributed)
3. **Salvatore Piscuoglio**, Nicole Lüscher, Anamaria Dumea and Karl Heinimann. **Identification of new “hot spots” in Lynch syndrome colorectal cancers.** Manuscript in preparation
4. **Salvatore Piscuoglio**, Shivendra Kishore, Mihaela Zavolan and Karl Heinimann. **Distinct microRNA expression profiling in mismatch repair associated colorectal cancer.** Manuscript in preparation
5. Kovac M, Hubert S, Rieder-Knell P, **Piscuoglio S**, Pietrzak M, Marra G, Bartosova Z, Terracciano L, Heinimann K. **Somatic inactivation of DNA mismatch-repair genes in Lynch syndrome cancers.** Manuscript in preparation
6. Kovac M, Kavacova M, **Piscuoglio S**, Roethlisberger B, Heinimann K. **Bioinformatics analysis identifies specific DNA signatures underlying**

different classes of copy number mutation in colorectal cancer gene.

Manuscript in preparation

7. Luca Quagliata^{*}, Mariacarla Andreozzi^{*}, Michal Kovac, Luigi Tornillo, Zuzanna Makowska, Marcus Heim, Karl Heinimann, **Salvatore Piscuoglio[§]** and Luigi Terracciano[§]. **SH2D4A is frequently downregulated in hepatocellular carcinoma.** Manuscript in preparation (equally contributed, corresponding author)
8. **Piscuoglio S^{*}**, Pallante P^{*}, Zlobec I, Fusco A^{*} and Terracciano LM^{*}. **HMGA1 and HMGA2 overexpression in human lung carcinomas.** Manuscript in preparation
9. Pallante P^{*}, **Piscuoglio S^{*}**, Zlobec I, Fusco A^{*} and Terracciano LM^{*}. **HMGA1 overexpression in breast carcinomas: a tissue microarray study.** Manuscript in preparation. (equally contributed)

11. ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my Ph.D. supervisor, Dr. Karl Heinimann. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make research fun for me.

Throughout my PhD period, he provided encouragement, sound advice, good teaching, good company, and lots of good ideas. It was also through his persistence, understanding and kindness that I completed this important step of my life. I would have been lost without him.

I would like to express my profound gratitude and thank to Prof. Luigi Terracciano that during this time has been like a father. He is the one who truly made a difference in my life. He provided me with direction, technical support and became more of a mentor and friend, than a professor. I doubt that I will ever be able to convey my appreciation fully, but I owe him my eternal gratitude.

I would also like to warmly thank my laboratory friends Marianne, Michele, Sybille and Nemya, for their technical advice and assistance, their never ending patience and of course, for the fun times we had together. Every research team should have a people like them!!!

I'd like to thank my PhD father prof. Primo Schär for his thesis supervision and for his advises during this journey; his wisdom, knowledge and commitment to the high standards inspired and motivate me.

I would like to express my heartfelt gratitude to Professor Alfredo Fusco and Dr. Pierlorenzo Pallante for their expertise on HMGA proteins and their sincere friendship.

I would also thank Dr. Shivendra Kinshore and Prof. Mihaela Zavolan for their hard collaborations, their expertise in computational and RNA biology made possible the development of many future projects.

Thanks to Dr. Michal Kovac, that during these years he helped me to solve critical problems, writing, giving wise advice, and helping with various applications.

I am indebted to my many colleagues/friends for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Maricarla, Anamaria, Francesca, Luigi, Luca, Giulio, Markus and Ezia.

I wish to thank my friends Giulia, Mike, Elena, Jenna, Ming, Claus, Daniela, Giancarlo, Meghan, Francesca A. and Luigi for helping me get through the difficult times, and for all the emotional support, entertainment, and caring they provided.

Lastly, and most importantly, I wish to thank my sister Giusy and my parents Carmela and Gaetano. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

