UNIVERSITAT DE BARCELONA

Identification of novel therapeutic targets and tumor suppressor genes in colon cancer using genome-wide high-throughput approaches

Sarah Bazzocco

Identification of novel therapeutic targets and tumor suppressor genes in colon cancer using genome-wide high-throughput approaches

by

Sarah Bazzocco

A thesis submitted in fulfillment of the requirements for

the title of PhD by the University of Barcelona

PhD program in Biomedicine

The thesis was carried out at

Molecular oncology Group

Centre for Molecular Biology and Biochemistry Investigation in Nanomedicine

(CIBBIM-Nanomedicine)

Vall d'Hebron Hospital Research Institute

Director: PhD candidate:

Diego Arango del Corro Sarah Bazzocco

Table of Content

Index of Figures

Index of Tables

Abstract

Colorectal cancer is a disease caused by genetic and epigenetic changes. Inactivation of tumor suppressor genes and activation of oncogenes are key landmarks in tumor progression. However, the list of tumor suppressor genes and oncogenes is far from complete, even in the case of the tumor types that are best characterized, such as colorectal cancer. Colorectal cancer is the second most frequent cause of cancer-related death in the Western world and is a serious health issue for the European Union. Patients having stage III or IV cancer undergo surgery followed by chemotherapy. However, the clinical management of these patients is far from optimal, and only about 30 % of the patients show an objective response to even the best chemotherapeutic agents available. In this study genome-wide high throughput assays were used to better characterize important aspects of the oncogenic progression such as deregulation of proliferation and aberrant expression caused by epigenetic mechanisms.

Because rapid tumor proliferation is associated with poor patient prognosis, here we characterized the transcriptional signature of rapidly proliferating colorectal cancer cells in an attempt to identify genes important to sustain tumor growth and that could be used as novel therapeutic targets. The proliferation rate of 52 colorectal cancer cell lines was determined and genome-wide expression profiling of a subset of these lines was assessed by microarray analysis. The expression of 1,290 genes was significantly correlated with the growth rates of colorectal cancer cells. These included genes involved in cell cycle, RNA processing/splicing and protein transport. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and protoporphyrinogen oxidase (PPOX) were shown to have higher expression in faster growing cancer cells. Importantly, pharmacological and genetic inhibition of GAPDH or PPOX reduced the growth of colon cancer cells *in vitro* and *in vivo*.

To better understand the mechanisms underlying the profound transcriptional reprogramming observed in cancer cells, we investigated the association between the levels of DNA methylation in the promoters of >14,000 genes and the levels of expression of these genes in a panel of 45 colorectal cancer cell lines. A group of cell lines with significantly higher methylation levels was observed, supporting the notion that there is a group of colorectal tumors with a CpG methylator phenotype (CIMP+). A significant negative regulation between methylation and expression levels was observed for 1,409 genes, suggesting that these genes are silenced during the tumorigenic process through this epigenetic mechanism. A significant number of these genes were zinc finger proteins, suggesting an important role of these DNA-binding proteins on the tumorigenic process. Strikingly, approximately one fourth of these genes are not associated with CpG islands, indicating that DNA

methylation outside these CpG rich regions is an important mechanism regulating gene expression and significantly contribute to tumor progression. In addition, we postulate that at least some of those genes have tumor suppressor activity. As a proof-of-concept, we show that restoration of the expression of ZNF238, a gene showing a significant methylation/expression correlation, resulted in reduced growth of colon cancer cells *in vitro* and *in vivo*.

In conclusion, in this study we shed new light on the mechanisms underlying the uncontrolled proliferation of colon cancer cells and the expression reprograming imposed in these cells through CpG methylation. The results of this study may contribute to the identification of novel chemotherapeutic targets for patients with colorectal cancer, and the characterization of novel genes/pathways with tumor suppressor activity, that are epigenetically silenced.

1 Introduction

1.1 The digestive tract

The human digestive tract is a system that comprises organs and glands responsible for processing and digesting foodstuff, which allow the absorption of nutrients. The digestive system consists of a large tube starting with the mouth, passing the esophagus, stomach, small and large intestines, and ending with the anus. Additionally, several glands, like the salivary glands, liver, gall bladder, and pancreas are secreting essential enzymes that support the digestion process (Figure 1).

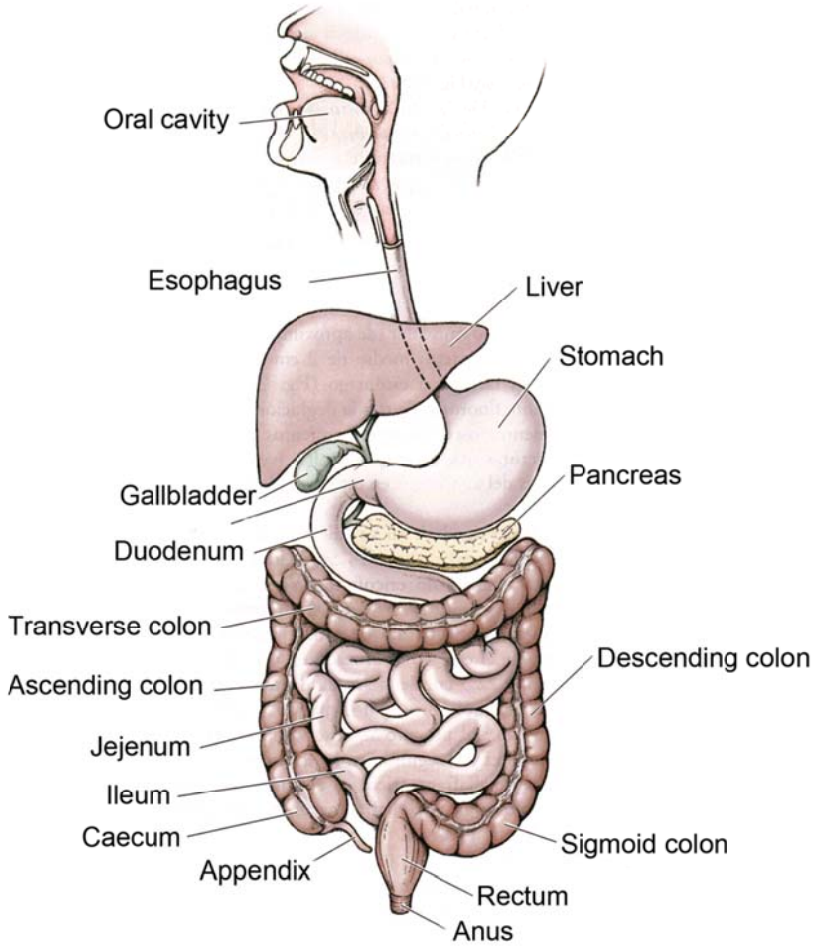

Figure 1: The human digestive tract.
Food gets processed from the uptake in the mouth, passing the esophagus, stomach, small and large intestines and ending in the anus. Figure modified from clinically oriented anatomy 1.

1.2 The human intestine

The human intestine is composed of the small and large intestine. The small intestine is in adults about 5 m long and consists of three distinct functional parts: duodenum, jejunum and ileum (Figure 1). The large intestine, consisting of cecum, ascending, transverse, and descending colon, sigmoid colon, and rectum, is the distal section of the gastrointestinal tract and lies between the ileum to the anus. It is wider than the small intestine, but only 1.5 m long. The large intestine, in its course, describes an arch, which surrounds the small intestine (Figure 1).

1.2.1 Anatomy of the small intestine

The three parts of the small intestine, namely duodenum, jejunum and ileum, are responsible for completion of food digestion and nutrient absorption (Figure 1) [2].

The **duodenum** with its characteristic elongated C shape is the shortest part of the small intestine (25 cm), connects the stomach to the jejunum and is only partially covered by the peritoneum. The duodenum's main function is the digestion of proteins, carbohydrates and fats. The duodenum obtains secretions from the pancreas and the liver thought the pancreatic and common bile ducts, respectively, in order to neutralize the acids from the stomach.

The **jejunum** (about 2.5 m) is the second largest region of the small intestine. It has a thicker intestinal wall than the duodenum and due to the circular and longitudinal smooth muscles it contracts rapidly and vigorously in a wave shaped manner. This movement is called peristalsis and it allows the processed food to move further down the tube. The jejunum's walls are surrounded by extensive arterial blood, which are able to take up all the nutrients.

The **ileum** makes about 60 % of the small intestine in an adult human. It is located in the lower abdominal part. Anatomically, it has thinner intestinal walls that result in slower peristaltic movements and less blood supply that the jejunum. There is no distinct anatomical demarcation between the jejunum and ileum, and both are held in place by the mesentery, a thin, broad-based membrane which is attached to the posterior abdominal wall and allows free movement of the small intestine within the abdominal cavity [2]. The jejunum and ileum present crescent folds of mucosa, known as plicae circulares. The functions of these folds are to retard the passage of the peristalsis and increase the absorptive surface, which results in a more efficient absorption of the nutrients.

The intestinal wall consists of four distinct functional layers: mucosa, submucosa, muscularis propria and serosa (Figure 2) [3].

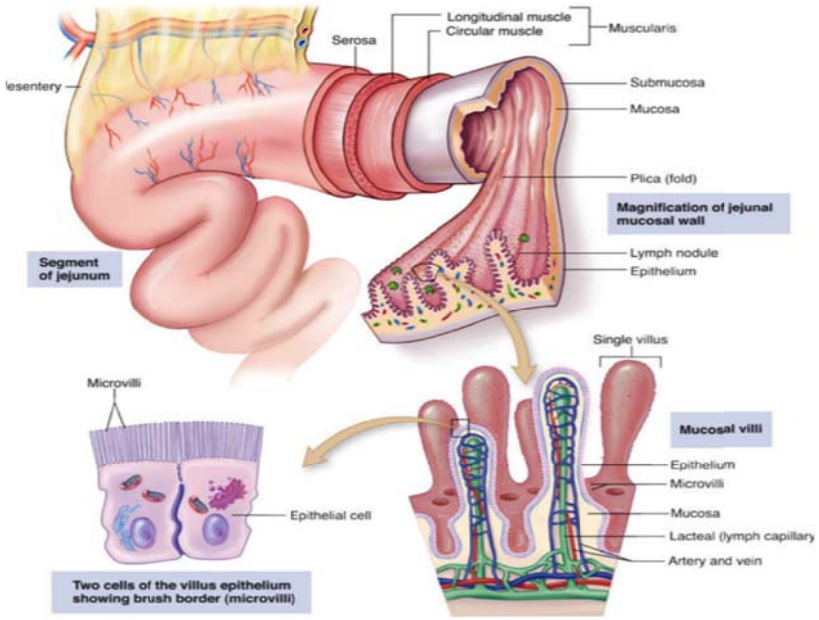

Figure 2: The small intestine
It is a long, narrow, folded or coiled tube extending from the stomach to the large intestine and a region where most digestion and absorption of food takes place. A thin membranous material, the mesentery, supports and somewhat suspends the intestines. Image taken from [4].

The **mucosa** itself comprises three layers: the epithelium, a supporting lamina propria, and the muscularis mucosae. Its surface is further organized in two fundamental structures, in order to enhance the absorption due to the increase of the surface. Those structures are the villi and the crypts or glands of Lieberkühn. The lamina propria is a thin layer of connective tissue lying beneath the epithelium. It functions as a first line of immune defense. In the lamina propria, lymphoid nodules (Payer's patches), lymphocytes and macrophages are present and defend against pathogens like bacteria or viruses. The next layer beneath the lamina propria is the muscularis mucosae. It is a thin smooth muscle layer all along the intestine and separates the mucosa from the submucosa. It keeps the mucosal surface and underlying crypts in a constant state of gentle agitation to expel contents of the glandular crypts and enhance contact between epithelium and the contents of the lumen [2].

The **submucosa** is a layer of dense irregular connective tissue or loose connective tissue. Blood vessels, lymphatic vessels, and nerves run through the submucosa and supply the mucosa.

The **muscularis propria** or muscularis externa consists of two smooth muscle layers. The inner, thicker layer consists of circular fiber, while the external thinner layer consists of longitudinal fibers. This thick muscle layers are combined performing the peristalsis while providing adequate support, elasticity and strength to the small intestine (Figure 2) [2].

Finally, the outer layer of loose supporting tissue is called serosa, which conducts of the major vessels, adipose tissue and nerves (Figure 2) [2].

1.2.2 Histology of the epithelium of the small intestine

The intestinal villi are highly vascularized and project the mucosa into the lumen of the small intestine. The **epithelium**, which covers the villi and the **crypts of Lieberkühn,** present several distinct cell types. In the villi and crypts the most abundant cell type is the **enterocyte** (Figure 3B). These cells present on the luminal surface structures known as microvilli. These microvilli form the brush border of the enterocytes and significantly increase the contact surface with the intestinal contents, improving massively the enterocyte's main functions, namely, the digestion and absorption of nutrients [3]. Scattered among the enterocytes are mucin producing cells known as **goblet cells** (7 % of the epithelial cells) (Table 3B). Mucin, composed by large glycoproteins, is able to form mucus. Mucus is very important for the intestinal function, because it lubricates the intestinal contents and acts as an important defense mechanism against physical and chemical injury caused by ingested food, microbes and the microbial products [5,6] and also as a substrate and niche for the commensal flora that colonizes the intestine [7]. The **enteroendocrine** cells represent 1 % of the cells in the villi and crypts and their function is to secrete multiple peptide hormones which exert control over physiological and homeostatic functions in the digestive tract like sensing the contents of the lumen [8,9]. The **microfold cells** (M cells) are specialized epithelial cells located over the Peyer's patches. These are accumulations of lymphatic tissue that can often be seen macroscopically as large white patches. The M cells transport antigens into the intraepithelial pockets assessed by antigen-presenting cells [10]. The last cell type, which can be found in the epithelium, are the Tuft cells and believed to secrete prostanoids [10]. The **crypts of Lieberkühn** are invaginations between the bases of the villi (Table 2) [2]. Only here **Paneth cells** can be found at the base of the crypts (Table 3B) [2]. They synthesize and secrete large amounts of antimicrobial peptides and proteins. It has been demonstrated that these antimicrobial molecules are key mediators of host-microbe interactions, including homeostatic balance with colonizing microbiota and innate immune protection from enteric pathogens [11]. Paneth cells also secrete factors that help sustain and modulate the epithelial stem and progenitor cells, which cohabitate in the crypts and contribute to the renewal of the epithelium of the small intestine [12].

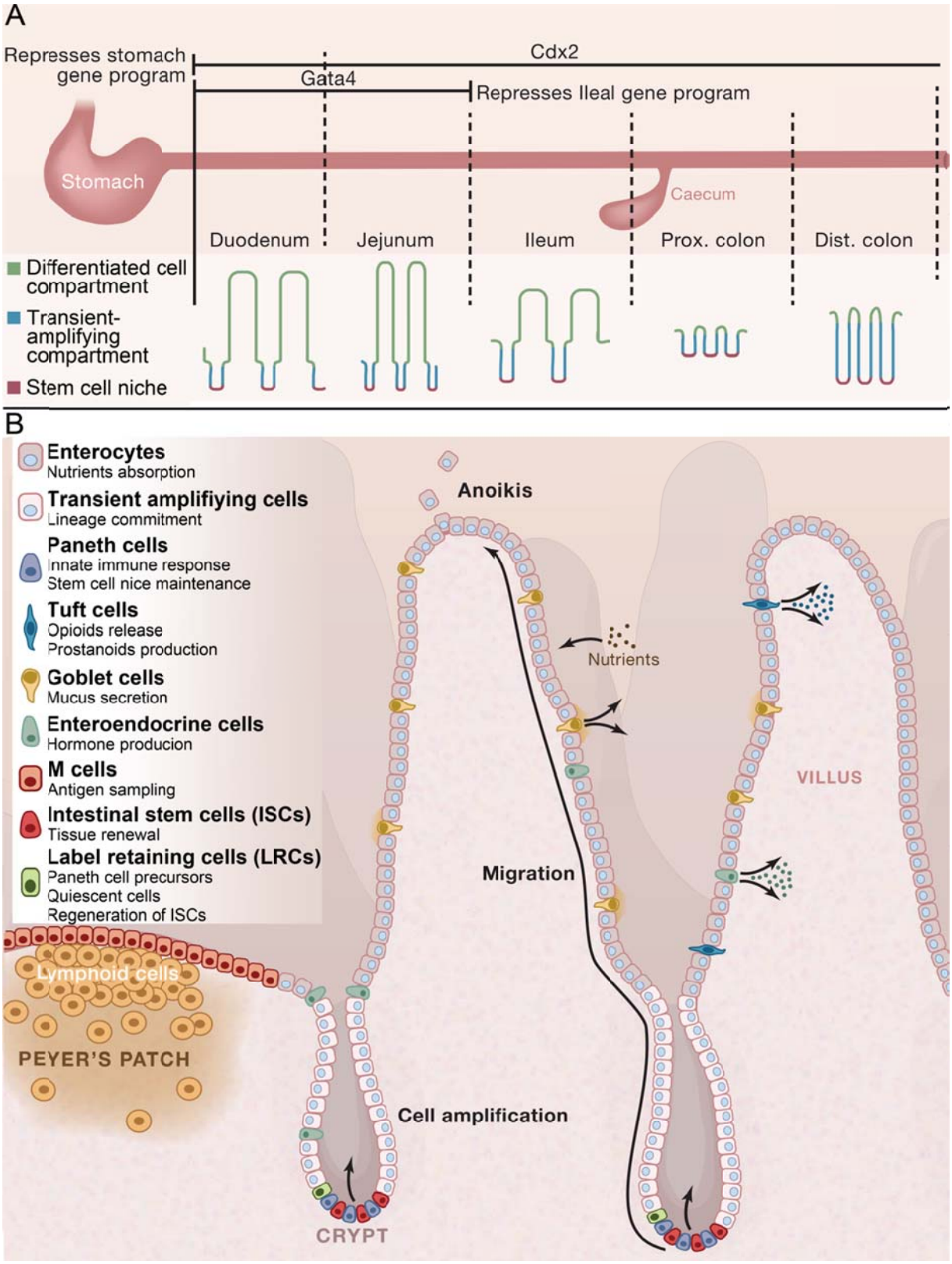

Figure 3: The intestinal crypt
A) Proximodistal organization of the gastrointestinal tract and B) the cell types of the small intestine. Modified from Clevers and Batlle [10].

1.2.3 Anatomy and histology of the large intestine

The first part of the colon, the **cecum**, lies in the right iliac region and dilates downward. The cecum is a pouch like structure 6 to 8 cm in length. It is followed by the colon, which forms the majority of the large intestine. The first part of the colon is called **ascending colon** (12 to 20 cm long)**.** It is called like that because it ascends along the right side of the peritoneal cavity to the hepatic flexure. The colon turns at the hepatic flexure, and as the transverse colon emerges medially and anteriorly into the peritoneal cavity. This longest (40 to 50 cm) and most mobile portion of the colon drapes itself across the anterior abdomen between the hepatic and splenic flexures. The descending colon, is about 30 cm long and travels posteriorly and then inferiorly in the retroperitoneal compartment to the pelvic brim and emerges there into the peritoneal cavity as the sigmoid colon. The rectum with a length of about 10 cm begins at the peritoneal reflection and ends at the anal canal (Figure 1) [2]. The large intestine does not perform digestion. However, it is responsible for water and vitamin uptake and so condense the feces. Moreover the bacteria that colonize the large intestine digest waste products.

The histological pattern of the large intestine differs from the small intestine in the way that villi and circular folds are absent and the crypts are longer, more numerous and closer together. The cell types present in the large intestine are similar to those in the small intestine. However, there is a larger number of goblet cells interspersed amongst the absorptive cells. Stem cells that represent the source of the other epithelial cell types are located at the bases of the intestinal crypts. However, there are no Paneth cells in the crypts of the large intestine. Enteroendocrine cells are located mostly at the base of the crypts and secrete basally into the lamina propria. The structure of the lamina propria, muscularis mucosae and submucosa are similar to the ones of the small intestine. The muscle layer build up by the muscularis propria is more prominent in the large intestine when compared with the small intestine, and consists of distinct inner circular and outer longitudinal layers [2].

1.2.4 Homeostasis of the human intestinal epithelium

The epithelium of the small and large intestine represents the fastest self-renewing tissue of an adult mammal. This self-renewing process is tightly regulated and highly dynamic system where proliferation and differentiation are elegantly balanced. The dividing cells are found in the crypts of Lieberkühn. In the small intestine, the progeny of these dividing cells migrate from the bottom of the

crypts upwards onto the surface of the villi. There no further division occurs and all cells are fully differentiated. Those cell are exposed to the gut contents and are finally shed from the villus tips [13]
As mentioned above, the epithelium of the small intestine consists of six cell types. Enterocytes and goblet cells are most abundant, and more rarely entereoendocrine, Tuft, Panth, and M cells are found. Apart from Paneth cells, that stay their entire live (6-8 weeks) in the bottom of the crypts, all other cell types migrate towards the tip of the villi in about five days [10]. This bottom to top axes is divided in three parts: **the stem cell compartment** at the base of the crypt, **the transient-amplifying compartment** in the middle of the crypts and the **differentiation zone**, which expands from the top third of the crypt, onto the surface of the epithelium and the tip of the villi [10,14]. The renewal of the epithelium needs only a small number of intestinal stem cells (about six intestinal stem cells per crypt) [15]. The intestinal stem cells are located at the bottom of the crypts and mingle with the Paneth cells [10]. While the intestinal stem cells are migrating upwards and enter the transient-amplifying compartment, they undergo a series of very fast divisions (about one every 12 h; Figure 3 A). Further migration to the differentiated cell compartment leads the transient-amplifying cells to become one of the six above mentioned cell types [10].

1.2.4.1 Proliferation and differentiation of intestinal stem cells

The differentiation of the progeny of the intestinal stem cells into one of the six cell types described above is regulated by a number of pathways. This is illustrated in Figure 4 A. First, the precursor cells differentiate into either a secretory or absorptive progenitor cell. This is controlled by the activity of the bHLH transcription factor Math 1. Secretory progenitor cells can further differentiate into the tuft, goblet or endocrine cells. However, *Math1* is repressed in enterocytes by the Notch downstream effector *Hes-1*. In order to become an M cell, the activation of transcription factor *Spi-B* is necessary [10].

It is currently believed that two types of intestinal stem cells can be found in the crypts: a) the Crypt base columnar stem cells are identified by being Leucine-Rich Repeat Containing G Protein-Coupled Receptor 5 positive (Lgr5+). In homeostasis, Lgr5+ intestinal stem cells are generating all types of epithelial cells and b) a minor subset of the intestinal stem cells are consider to be quiescent and build a reserve for intestinal stem Lgr5+ cells. Those quiescent cells are marked by *Bmi1, Lrig1, Tert,* and *Hopx* and are localized around position +4 in the crypt and are called label-retaining cells [11,13,14,16,17]. However, the label-retaining cells are considered not to be intestinal stem cells but rather the precursor cells of Paneth cells, since they are short lived (2-3 weeks), differentiate towards mature Paneth cells, and also express several Paneth-cell-specific genes [18,19].

The intestinal stem cells divide in a stochastic manner and either adopts stem or transient-amplifying fates. Because a niche is of limited size, the growth of those cells is very competitive. Eventually, only surviving clones continue to cover the surface of the crypts whereas existing neighbor clones get extinguished. As a consequence, the crypts drift towards a clonality phenotype within a period of 1-6 month. Therefore, intestinal stem cells last as a population, although, the descents of only one particular intestinal stem cell are present in each crypt [20] (Figure 4 B).

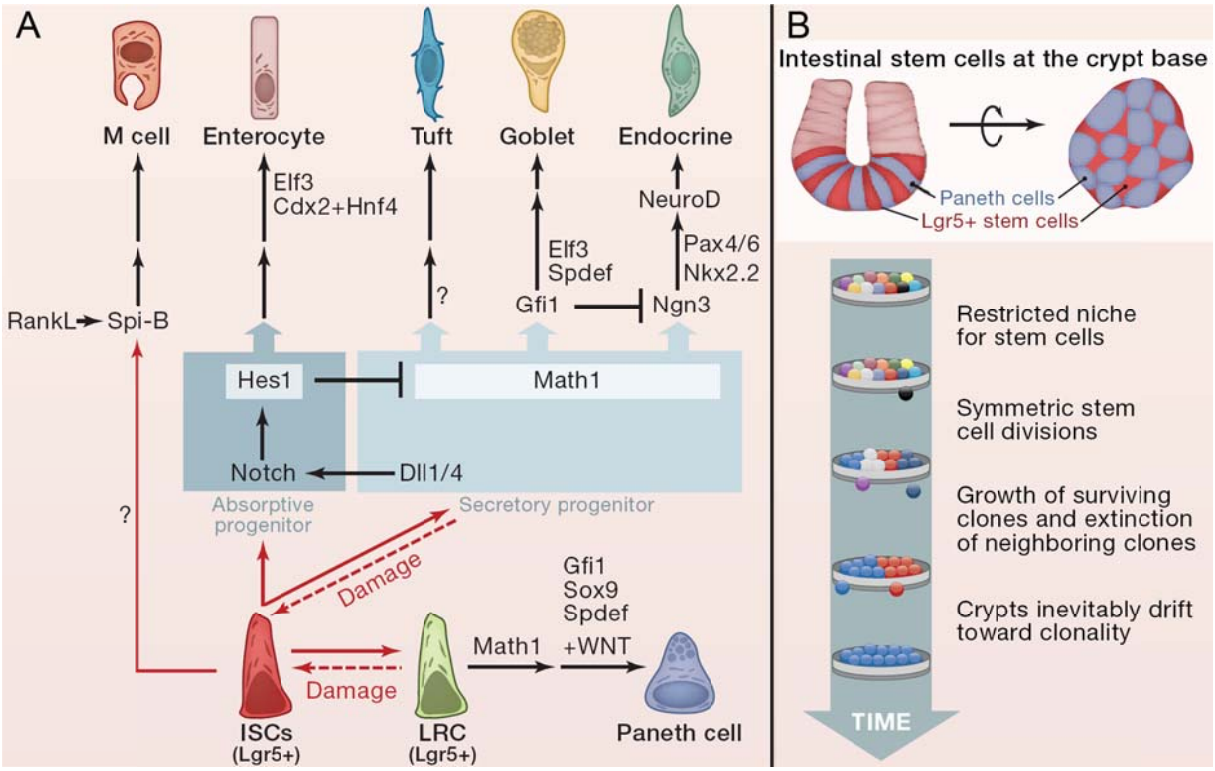

Figure 4 : Proliferation and differentiation in the intestinal epithelium
A) Cell hierarchy and lineage specification, B) Neutral competition in the intestinal stem cells niche. Modified from Clevers and Batlle [10].

1.2.4.2 Signaling in the intestinal stem cells

Wnt signaling pathway maintains proliferation – Intestinal stem cells are known for high Wnt signaling activity. The WNT signaling pathway is activated when WNT ligands (WNT3, WNT6 and WNT9B) bind to the corresponding receptor families (frizzled 5, 6 and 7, and LRP5 and 6). The ligands are highest expressed at the bottom of the crypts, and the receptors are expressed by the epithelial cells of the crypts. Neighboring Paneth cells constitutively secrete the WNT3 ligand, and the surrounding stroma is a Wnt source, too[21] (Figure 5).

The ligand-receptor interaction leads to the activation of Dishevelled family proteins and further downstream changes of the β-catenin levels, which translocate to the nucleus. Nuclear b-catenin

interacts with TCF/LEF family transcription factors to initiate gene expression responsible to maintain the proliferative state of the stem cells. Key target gens of the Wnt signaling pathway are *C-MYC* [22], cyclineD1 [23] and *CD44* [24]. WNT signaling pathway is potentiated when R-sporin binds to LGR4/ LGR5 receptors [25]. WNT signaling pathway is potentiated when R-sporin binds to LGR4/ LGR5 receptors [26].

Notch signaling pathway: first decision between absorptive and secretory cell fate – A common secretory precursor cell is thought to give rise to Goblet, endocrine and Paneth cells , whereas enterocyte precursor cell rise from enterocytes. The transcription factor HES1 regulates the transformation of precursor to enterocyte cells. HES1 is a target of the NOTCH signaling pathway. Notch receptors are activated by ligands expressed on the adjacent cells. The Notch ligands Dll1 and Dll4 are expressed by surrounding cells including Paneth cells [10]. Once activated, the intracellular cytoplasmic domain gets cleaved by □-secretase and translocate to the nucleus. Nuclear Notch binds to the transcription factor CBF1, which leads to the activation and expression genes like HES1. However, secretory precursor cells differentiate into goblet or endocrine cells under the regulation of MATH1 and HFI1 and NGN3 [27–29].

Hedgehog demarcates villus from crypt - The hedgehog pathway demarcates between villus and crypt. It signals from the epithelium to mesenchyme and is required for the formation of villi. When the expression of the two Sonic hedgehog ligands Shh and Ihh is lost in mice, the persistence of a highly proliferative intestinal epithelium due to an increase in Wnt signaling is observed and the formation of villi is completely absent. This process depends on a feedback loop. Mesenchymal cells respond directly to hedgehog signaling from the epithelium, and deliver a signal back to the epithelium by other signaling pathway[30].

BMP pathway inhibits crypt formation - The bone morphogenetic protein (BMP) pathway signals from the mesenchyme to the epithelium [13]. Under the control of the hedgehog signaling, mesenchymal cells express BMP2 and BMP4. Parallely, epithelial cells express the receptor called BMPR1A [30]. The BMP pathway is important for crypt formation. Knocking down BMPR1A in the epithelium causes excessive production of crypt-like structures [31,32] (Figure 5).

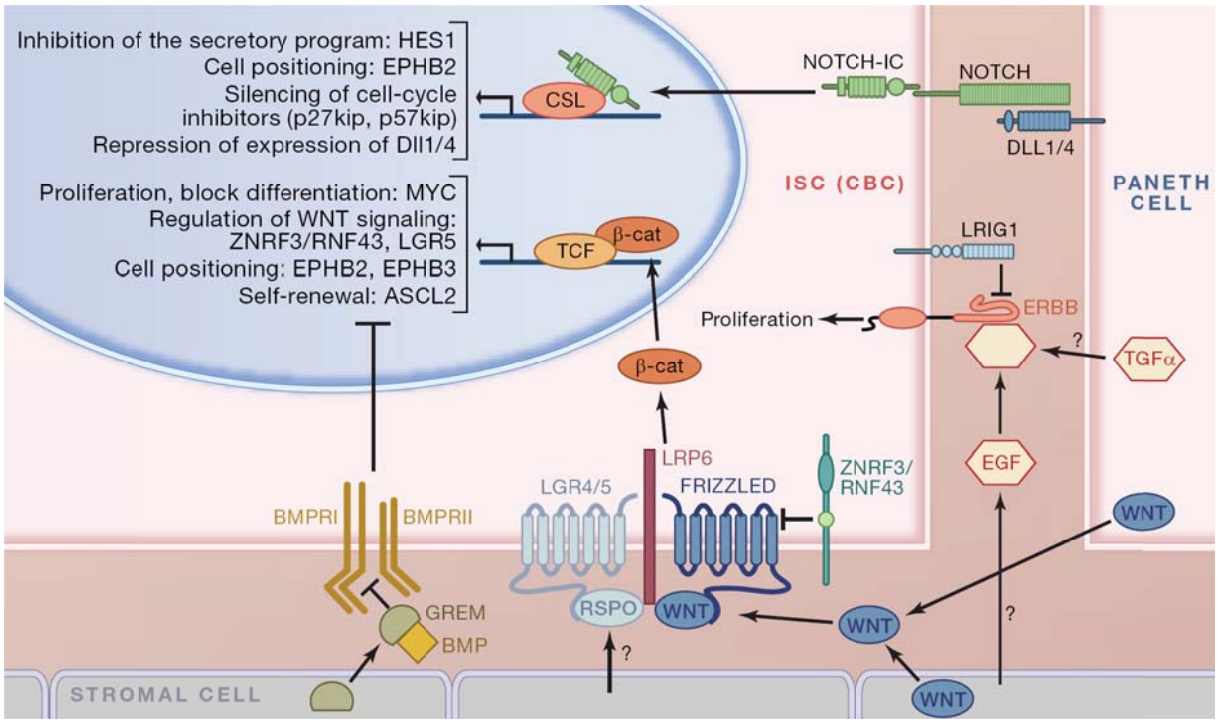

Figure 5: Proliferation and differentiation in the intestinal epithelium.
Signaling in the intestinal stem cells. Modified from Clevers and Batlle [10].

1.3 General characteristics of cancer

Cancer is a multistep process of genetic and epigenetic changes in which gain of function of oncogenes and loss of function of tumor suppressor genes eventually lead to proliferation, maturation, and metastasis.

1.3.1 Types of cancer

According to Cancer Research UK, there are more than 200 types of cancers. They are named according the organ or the cell type they origin from. They are classified into the following main categories:

- Carcinoma – originated from epithelial cells. This group includes the most common cancers including the breast, prostate, lung, pancreas, and colon.
- Sarcoma - originates in connective or supportive tissue, like bone, fat, cartilage, muscle, and blood vessels.
- Germ cell tumor – originated from from pluripotent cells, most often found in the testicle or the ovary.
- Lymphoma, myeloma and leukaemia – derived from cells of the immune system.
- Central nervous system cancers – originated in brain tissues and spinal cord

1.3.2 World and European cancer statistic

Cancers are among the leading causes of morbidity and mortality worldwide, with approximately 14 million new cases and 8.2 million cancer related deaths in 2012. It is expected that annual cancer cases will rise to 22 million within the next two decades. The most common causes of cancer death are cancers of: lung (1.59 million deaths), liver (745 000 deaths), stomach (723 000 deaths), colorectal (694 000 deaths), breast (521 000 deaths), and esophageal cancer (400 000 deaths) [33].

Around one third of cancer deaths are due to five leading behavioral and dietary risks: high body mass index, low fruit and vegetable intake, lack of physical activity, tobacco use, alcohol use. Tobacco use is the most important risk factor for cancer causing around 20 % of global cancer deaths and around 70 % of global lung cancer deaths. Cancer causing viral infections such as HBV/HCV and HPV are responsible for up to 20 % of cancer deaths in low- and middle-income countries [34]. More than 60 % of world's total new annual cases occur in Africa, Asia and Central and South America. These regions account for 70 % of the world's cancer deaths [33]

Ferlay et al. show the estimated numbers of cancer cases and cancer deaths in the 40 European countries for the year 2012 (Figure 6). There were an estimated 3.45 million new cases of cancer (excluding non-melanoma skin cancer, because no attempt was made to estimate incidence and mortality from non-melanoma skin cancer) and 1.75 million deaths from cancer in Europe in 2012 [35]. The most common cancer sites were cancers of the female breast (464,000 cases), followed by colorectal (447,000), prostate (417,000) and lung (410,000). These four cancers represent half of the overall burden of cancer in Europe. Globally, in 2012 the most common causes of death from cancer were cancers of the lung (353,000 deaths), colorectal (215,000), breast (131,000) and stomach (107,000). In the European Union, the estimated numbers of new cases of cancer were approximately 1.4 million in males and 1.2 million in females, and around 707,000 men and 555,000 women died from cancer in 2012 [35].

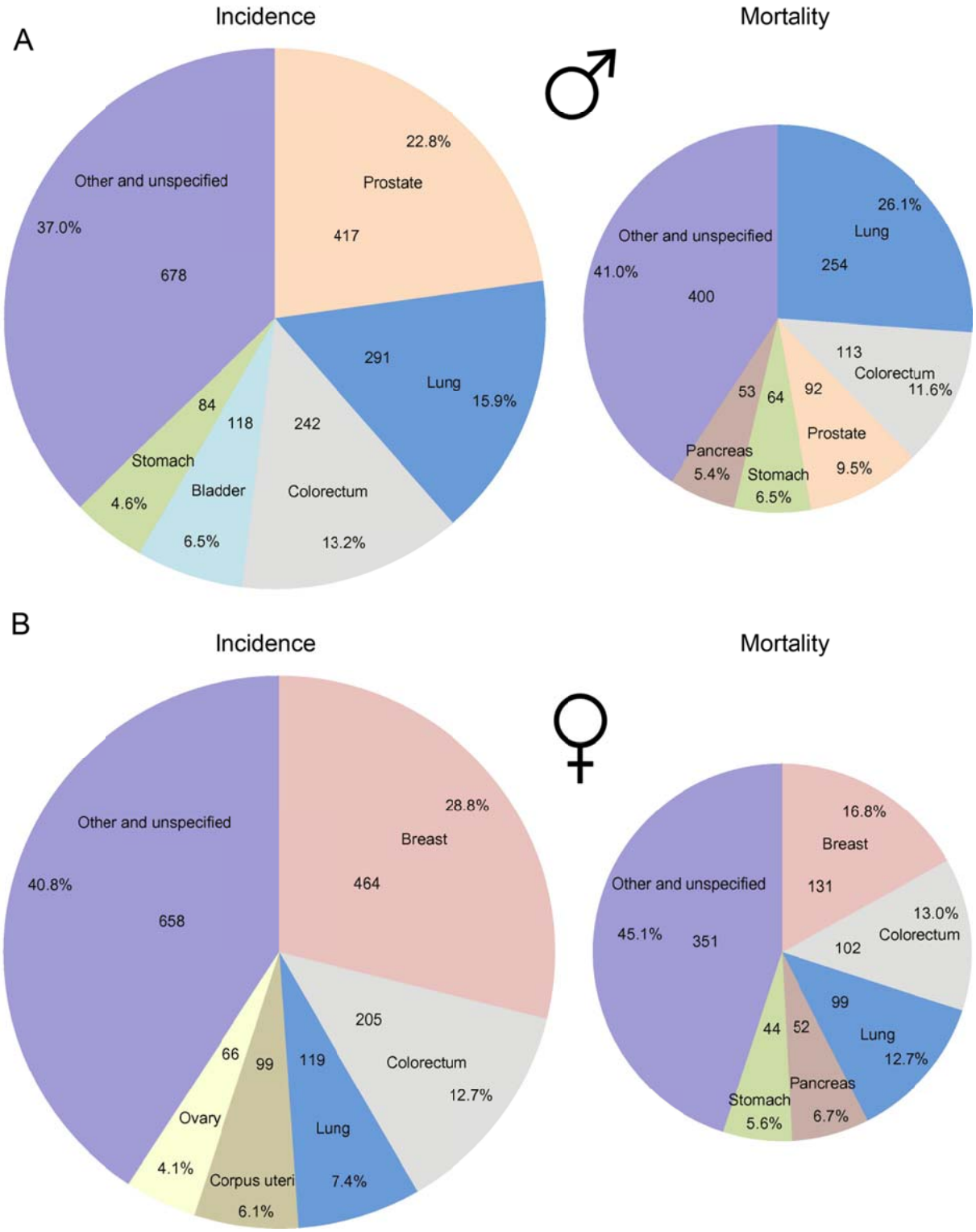

Figure 6: Distribution of the expected cases and deaths for the 5 most common cancers in Europe 2012 in males (A) and females (B).

For each sex, the area of the segment of the pie chart reflects the proportion of the total number of cases or deaths. Graphics taken from Ferlay, et al [35].

1.3.3 Hallmarks of cancer

In 2000, Hanahan and Weinberg published a review article, which focused on the global findings of the past 25 years of cancer research. They suggested that most, if not all cancers have a manifestation of in six essential alterations, which they call “the hallmarks of cancer”, in common. These hallmarks are essential to transform a normal human cell into a malignant cancer cell. Those six hallmarks were listed as followed and represented in Figure 7:

- **Self-sufficiency in growth signals.** Normal tissues carefully regulate the production and release of growth promoting signals. Cancer cells however have the ability to sustain chronic proliferation. These enabling signals are conveyed in large part by growth factors that bind cell-surface receptors, typically containing intracellular tyrosine kinase domains [36,37].

-**Insensitivity to growth inhibitory (anti-growth) signals.** In addition to the capability of sustain proliferation, cancers cells become resistant to signals that negatively regulate cell proliferation. This is largely due to the inactivation of tumor suppressor genes [36,37].

-**Tissue invasion and metastasis.** Most cancer types, eventually invade adjacent tissues, and move from there to distal sites in order to settle and form new colonies. This action is called metastasis and helps the tumor cells to get new space, and nutrients. About 90 % of human cancer deaths are caused due to metastasis [36,37].

-**Limitless replicative potential.** Cancer cells, unlike most normal cells, have an unlimited replicative potential. In most normal cells, the number of cell cycles a cell can undergo before entering senescence or cell death is controlled by multiple tandem repeat hexanucleotides, called telomeres. They are located at the end of the chromosomes and shorten with every cell division. Their function is to protect the DNA from end-to-end-fusions. 85-90 % of the cancers successfully maintain the telomeres by upregulation of the telomerase, which adds hexanuclotide repeats onto the ends of telomeric DNA [36,37].

-**Sustained angiogenesis.** As in normal tissue, tumors require nutrients and oxygen to sustain as well as to draw off metabolic wastes and carbon dioxide. Blood vessel formation (vasculogenesis) and sprouting (angiogenesis) are the solution to this problem. In normal tissues vasculature is formed in the embryonic stage and becomes largely quiescent during the adults stage. Only during wound healing, and female reproductive cycle the angiogenic process is turned on. Cancer cells, however, have the ability to induce an angiogenic switch, in which angiogenesis is almost always activated. This

leads to continuous sprouting and sustains the neoplastic growth. Key factors, which are involved in angiogenesis are the angiogenesis inducing Vascular Endothelial Growth Factor A (VEGF-A) and angiogenesis inhibitor thrombosporin-1 (TSP-1), whose expression are up- or down-regulated, respectively [36,37].

-**Evasion of programmed cell death (apoptosis).** The ability of cancer cells to expand in number is not only due to the rate of cell proliferation but also the rate of attrition. Apoptosis, programmed cell death, represents a major source of attrition. Once the apoptosis machinery is triggered by various physiological signals, it leads to a tightly regulated downstream-activated cascade of signals ultimately leading to cell death. The apoptosis machinery is composed of two entities, the sensors and the effectors. The most commonly lost pro-apoptotic sensor is p53 which is mutated in more than 50 % of the human cancers. This results in the removal of a key DNA damage sensor, which would normally induce the apoptotic cascade [36,37].

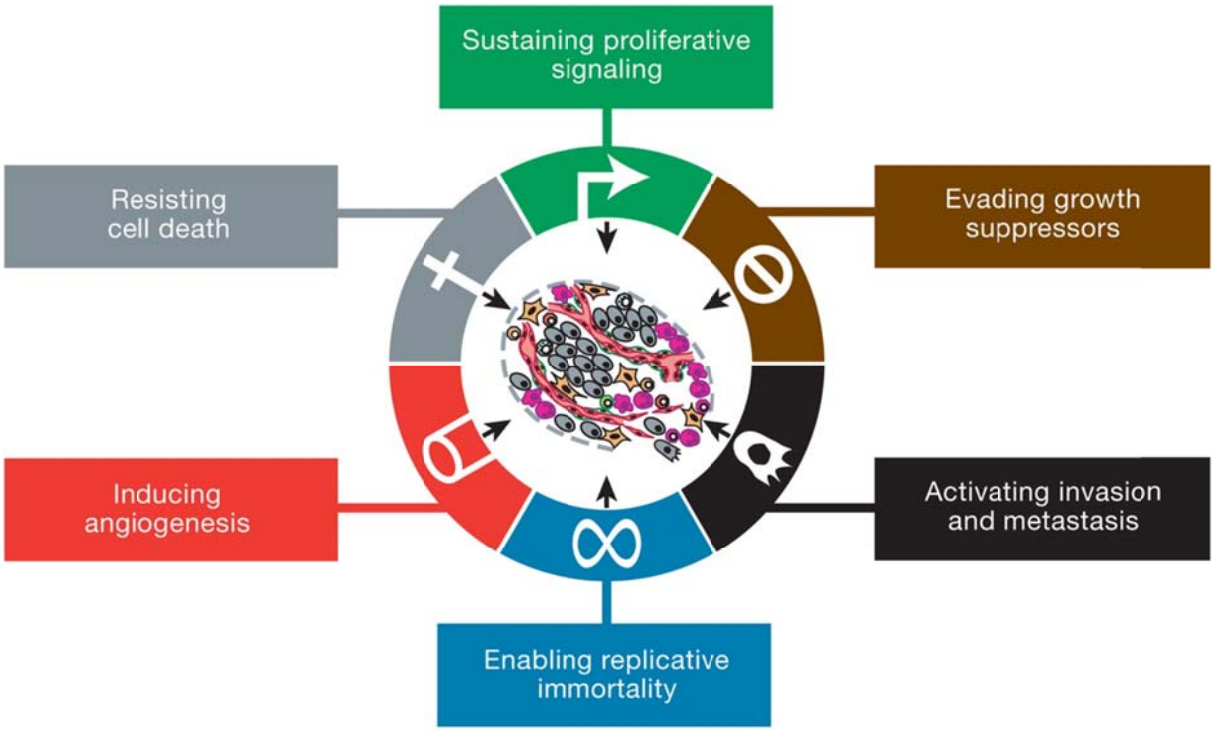

Figure 7: The Hallmarks of Cancer.
The transformation of normal cells into a neoplastic state requires a succession of hallmark capabilities. Image taken from Hanahan and Weinberg [37].

Each of these hallmarks is the cause of successfully overcoming the anticancer defense mechanisms in each cell and tissue type. While it is believed that virtually all cancers must acquire the same six hallmark capabilities, there is a significant difference in the mechanisms involved in performing the hallmarks as well as the order of the onset of each hallmark for each cancer cell type. The acquisition of the six hallmarks is made possible by two enabling characteristics. First, and most prominent, is the development of **genomic instability** in cancer cells, which generate random mutations including chromosome rearrangements; among these are the rare genetic changes that can orchestrate hallmark capabilities. A second, enabling characteristic involves the **inflammatory state** of premalignant and fully malignant lesions that is driven by cells of the immune system, some of which serve to promote tumor progression through various means (Figure 8).

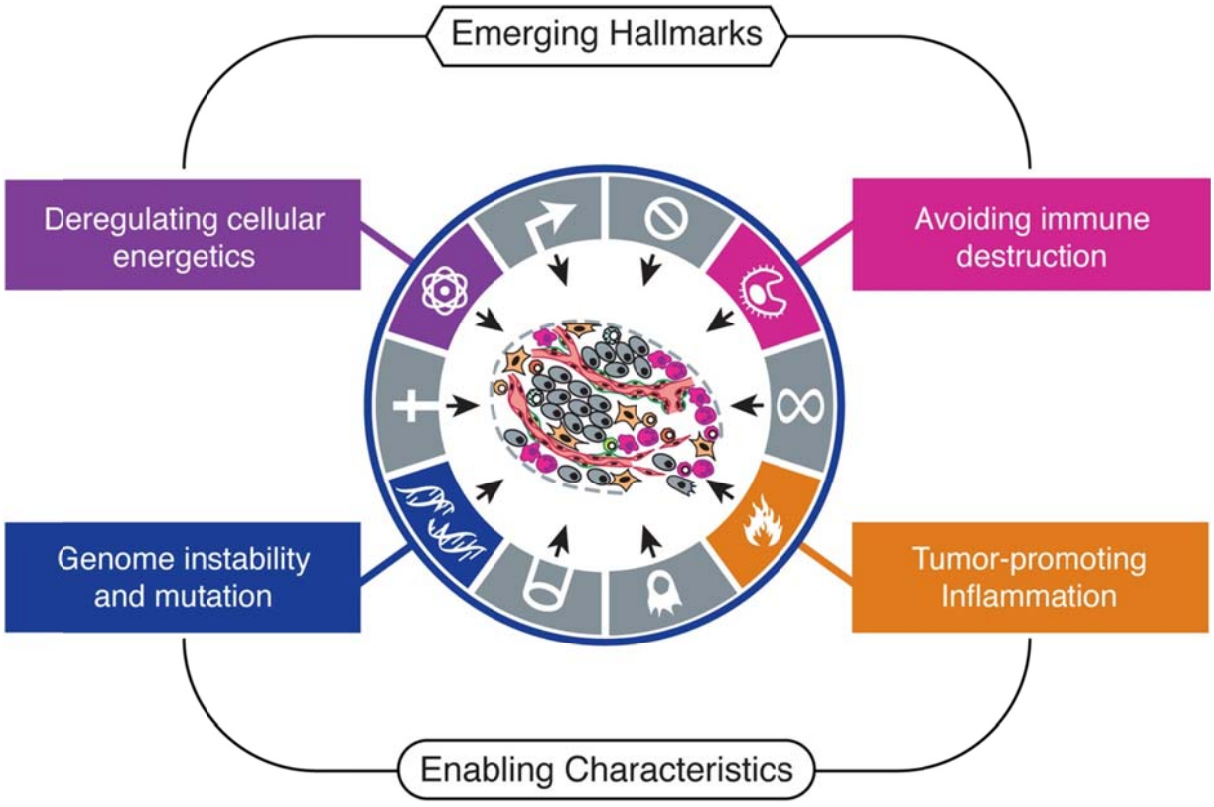

Figure 8: Emerging hallmarks of Cancer.
Increasing research evidence suggests that two additional hallmarks of cancer are involved in the pathogenesis cancer. The first one involves the capability to modify, and reprogram, cellular metabolism in order to sustain neoplastic proliferation. The second allows cancer cells to evade immunological destruction process, in which T and B lymphocytes, macrophages, and natural killer cells are involved. Since their capability is not generalized and fully validated they are marked as emerging hallmarks. Additionally, genomic instability and therefore mutability endow cancer cells with genetic alterations which drive tumor progression. Inflammation by innate immune cells can instead result in their inadvertent support of multiple hallmark capabilities. Thereby inflammation manifests the now widely appreciated tumor-promoting consequences leading to inflammatory responses. Image taken from Hanahan and Weinberg [37].

Recently, other alternative attributes of cancer cells have been proposed to be important for the development of cancer and therefore should be added to the list of core hallmarks. One involves major reprogramming of cellular energy metabolism. Here continuous cell growth and proliferation is guaranteed by replacing the metabolic program that operates in most normal tissues and fuels the physiological operations of the associated cells. Another attribute involves active evasion by cancer cells from attack and elimination by immune cells; this capability shows the dichotomous roles of the immune system which both antagonizes and enhances tumor progression (Figure 8) [37].

1.3.4 Tumor suppressor genes and oncogenes

The previously described hallmarks, driving the conversion of a normal cell into a cancer cell, are the result of alterations in DNA, such as mutations due to replication errors, exposure to carcinogens, or faulty DNA repair. People with inherited germ line mutations are predisposed to cancers, but require additional somatic mutations for cancer to develop. The tumorigenic process is largely driven by mutations in two classes of genes: Tumor suppressor genes and oncogenes. Genes involved in these categories are key players in regulating proliferation, differentiation and cell death.

Tumor suppressor genes generally encode proteins that inhibit cell proliferation or cell growth. In tumor suppressor genes the mutations cause a loss of function. Therefore, most are recessive in nature, because both alleles must be mutated [38]. There are 246 tumor suppressor genes known according to UniProt (searched for keyword “tumor suppressor gene” and *homo sapiens* the 06/10/2015) [39]. Five broad classes of proteins are generally recognized as being encoded by tumor-suppressor genes: a) intracellular proteins, such as the p16 cyclin-kinase inhibitor, that regulate or inhibit progression through a specific stage of the cell cycle, b) receptors for secreted growth factor that function to inhibit cell proliferation, c) checkpoint-control proteins that arrest the cell cycle if DNA is damaged or chromosomes are abnormal, d) proteins that promote apoptosis, and e) enzymes that participate in DNA repair.

Although DNA-repair enzymes do not directly function to inhibit cell proliferation, cells that have lost the ability to repair errors, gaps, or DNA broken ends accumulate mutations in many genes, including those that are critical in controlling cell growth and proliferation. Thus loss-of-function mutations in the genes encoding DNA-repair enzymes promote inactivation of other tumor-suppressor genes as well as activation of oncogenes [40]. Well-described TSGs include genes in pathways such as Wnt/APC (adenomatous polyposis coli gene (*APC*), *AXIN1*, and *CDH1*); apoptosis/cell cycle (cyclin-dependent kinase inhibitor 2A (*CDKN2A*), tumor protein 53 (*TP53*), *RB1*, *TRAF7*, and *CASP8*); chromatin

modification (*ARID1A/B/2*, *ASXL1*, *ATRX*, *CREBBP*, *KDM5C*, *KDM6A*, *MEN1*, *MLL2/3*, *SETD2*, ten-eleven translocation-2 (*TET2*), *WT1*, and *BAP1)*; DNA damage repair (ataxia telangiectasia mutated (*ATM*), ataxia telangiectasia and Rad3 related (*ATR*), *BRCA1/2*, mutL homolog 1 (*MLH1*), and *MSH2/6*); hedgehog (*PTCH1*); Notch (*FBXW7* and *NOTCH1*); phosphoinositide 3- kinase (PI3K)/AKT/mammalian target of rapamycin (mTOR) (*PIK3R1*, phosphatase and tensin homolog (*PTEN*), and *TSC1*); Ras (*CEBPA*, von Hippel-Lindau (*VHL*), and *NF1*); transforming growth factor-b (*SMAD2/4*); and transcriptional regulation (*GATA3* and *RUNX1*) [41].

Oncogenes are proteins, whose product have increased activity and therefore act in a dominant manner and are responsible to promote cancer. A mutation in only one allele is sufficient for an effect [38]. From the 223 known oncogenes according to UniProt (searched for keyword "proto-oncogene" and *homo sapiens* the 06/10/2015) [39], most of them are derived from normal cellular genes. Those genes are called proto-oncogenes. For example, the *RAS* gene is a proto-oncogene that encodes an intracellular signal-transduction protein; the mutant *RAS* gene derived from *RAS* is an oncogene, whose encoded oncoprotein provides an excessive or uncontrolled growth-promoting signal. Because most proto-oncogenes are basic to animal life, they have been highly conserved over eons of evolutionary time [40].

To convert a proto-oncogene into an oncogene generally involves a gain-of-function mutation. Underlying genetic mechanisms associated with oncogene activation include the following: a) Point mutations in a proto-oncogene that result in a constitutively active protein product. b) Point mutations, deletions, or insertions in the promoter region of a proto-oncogene that lead to increased transcription. C) Localized duplication (gene amplification) of a DNA segment that includes a proto-oncogene, leading to overexpression of the encoded protein. d) Chromosomal translocation that dislocates a growth-regulatory gene under the control of a different promoter (chimeric protein) and that causes therefore inappropriate expression of the gene [40,42].

Although a lot of tumor suppressor genes and oncogenes are known, every year new genes of these categories are discovered. Recently, two new genes with tumor suppressor activity in colorectal cancer, namely *Myo1A* [43] and *RhoA* [44] have been discovered by our group. However, many more genes are yet to be discovered, which will help to better understand the underlying mechanisms in cancer.

1.3.5 Epigenetic alterations in cancer

The term epigenetics refers to heritable changes in gene expression, function and/or regulation that do not involve changes in the DNA sequence. In other words, epigenetic changes can affect the phenotype but not the genotype. Cancer is a disease, which include both genetic and epigenetic alterations. The epigenetic changes include three categories: DNA methylation, histone modifications, and chromatin remodelers [45].

In the nucleus, DNA is tightly condensed and wrapped around nuclear proteins called histones. The repeating DNA-histone complex, which consists of 146 base pairs of double-stranded DNA wrapped around eight histone proteins, is called a nucleosome. **Histone posttranslational modification** comprises the largest epigenetic control with over 50 known sites and include modifications like acetylation, methylation, citrullination, phosphorylation, SUMOylation, and ADP-ribosylation [46,47].

DNA methylationoccurs in mammals when a methyl group is covalently transferred from S-adenosylmethionine to the C-5 position of cytosine by a family of cytosine (DNA-5)-methyltransferases (DNMTs). Only cytosine bases that are located 5’ to a guanosine (CpG dinucleotides) are methylated. CpG dinucleotides are underrepresented in the genome. However, short regions rich in CpG content exists and are known CpG islands. Most CpG islands are found in repetitive elements like centromeres, microsatellite sequences, and proximal promoter regions [48,49]. Transcription is mainly by affected by the accessibility of DNA to transcription factors due to the methylation modification made at the CpG regions [50]. Hypomethylation is frequently observed in carcinogenesis [51]. Hypomethylation is frequently observed in carcinogenesis [52–54], resulting in aberrant proliferation of the cell. Hypomethylation can cause higher mutation rates due to mitotic recombination-related defects, like deletions or loss of entire chromosomes [55] In addition, the opposite can happen; promoter hypermethylation can silence genes was they lose their gene function and is a well-characterized epigenetic change in human tumors [56,57]. The presence of CpG island promoter hypermethylation affects genes involved in cell cycle ($p16^{INK4a}$, $p15^{INK4b}$, *Rb*, $p14^{ARF}$), cell-adherence (*CDH1, CDH13*), DNA repair (BRCA1, hMLH1, MGMT), carcinogen-metabolism (*GSTP1*), , apoptosis (*TMS1, DAPK*)[58].

Chromatin has two functions: it serves to condense DNA within the cellular nucleus, and it controls how the DNA is used. In particular, within eukaryotes, specific genes are not expressed unless they are accessible for RNA polymerase and transcription factors. In its default state, the tight coiling of the DNA limits the access of these transcriptiknal regulators to the eukaryotic DNA. The SWI/SNF complexes are a family of multi-subunit complexes that use the energy of adenosine triphosphate

(ATP) hydrolysis to remodel nucleosomes. **Chromatin remodeling** processes mediated by the SWI/SNF complexes are critical for gene expression across a variety of cellular processes, including stemness, differentiation, and proliferation. SWI/SNF mutations, and the subsequent abnormal function of SWI/SNF complexes, are among the most frequent gene alterations in cancer, but the mechanisms by which perturbation of the SWI/SNF complexes promote oncogenesis are not fully elucidated. However, alterations of SWI/SNF genes clearly play a major part in cancer development, progression, and/or resistance to therapy [59,60].

1.4 Colorectal cancer

1.4.1 Origin and clinical classification of colorectal cancer

Adenocaricinomas arise from the epithelium of the colon. Colorectal cancer is the end result of a multistep process, which can take several years to develop. As previously mentioned the intestinal epithelium is tightly regulated for proliferation, differentiation, migration and cell death. Tumorigenesis occurs when those mechanisms fail leading to hyperproliferation. The growth from a carcinoma *in situ* to a metastatic colon cancer passes several stages. Colonic or rectal tumorigenesis is first detected as epithelial hyperplasia, which result in aberrant crypt foci and benign tumors, also known as adenomas or adenomatous polyps and can eventually develop into malignant neoplasia called carcinoma [14]. The origin of the cells from which tumors arise is still not fully understood. There are two controversial, possible scenarios: The top-down hypothesis suggests that adenomas arise from dysplastic cells located on the epithelial surface and expand by lateral migration down into the crypt [61]. The bottom-up mechanism supports the idea of which neoplastically transformed stem cells arise from the crypt base, progress to adenomas and expand by crypt fission to the surface of the epithelium [62]. Current data favors the bottom-up theory since intestinal stem cells arise from the crypts, possess already the machinery to prolonged proliferation and they are the only cells that persist long enough in the tissue to undergo the prolonged sequence of successive mutation and selection demanded by the concept of multistage carcinogenesis [63,64] (Figure 9).

In order to develop neoplasia, adenomas arise as polypoid structures that grow into the colonic lumen. Over time, the villous histology of adenomas becomes more disordered and when invasive cells breach the underlying epithelia basement membrane they are are as cancers [65].

Most colorectal cancer could be prevented if they are detected at a premalignant colon adenoma stage. The same concept is true for early colorectal cancer. Simple surgical reaction of the involved

colonic region would be enough to "cure" a patient. Those considerations led to the recommendations of mass screening for people above 50 years of age that show average risk pattern. However, for individuals, that have a high risk due to family history or other predisposing factors, the onset of screening is earlier [66].

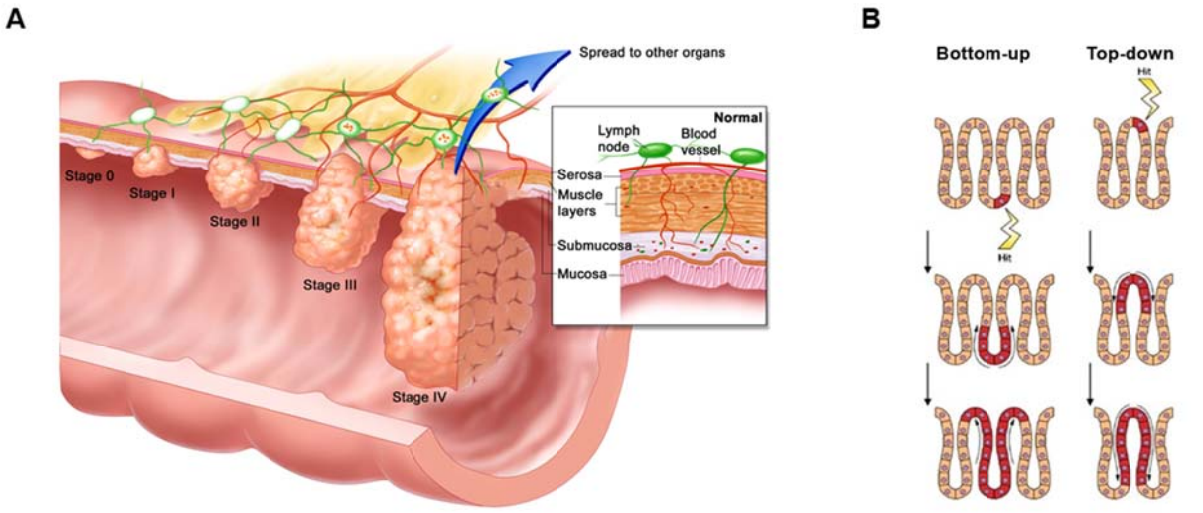

Figure 9: The growth from a polyp to a metastatic tumor.
A) In stage 0, also called carcinoma in situ, abnormal cells are found in the mucosa of the colon wall. These cells may later turn into a carcinoma and metastasize. In stage I, cancer has built in the mucosa of the colon wall and spread to the submucosa and/or the muscle layer of the colon wall. Stage II tumors have spread through the muscle layer of the colon wall to the serosa. In stage III, cancer invaded mucosa, the submucosa and to nearby lymph nodes. In stage IV the cancer has metastasized to preferred organs like lung and liver. Image from the Terese Winslow, US Govt.
B) Bottom-up theory includes the spread of dysplastic cells initiated at the base of the crypt within the stem cell zone, and continues upward which leads to the fact that the whole crypt is dysplastic (cell migration is indicated by the arrows). (C) The opposite is happening in the top-down model, which suggests that a mutation resulting in the appearance of a dysplatic cell can occur at, or near, the top of the crypt. Modified from McDonald et al [67].

colorectal cancer is by far the most common malignancy of the gastrointestinal tract, and it is, without question, a "surgical disease" [68]. An estimated 92 % of colon cancer patients and 84 % of rectal cancer patients undergo surgical resection as the primary modality of treatment [68]. The pathologic assessment of the resection specimen provides data that is essential for patient management, including the estimation of postoperative outcome and the rationale for adjuvant therapy. The essential elements of this assessment include the pathologic determination of TNM stage, tumor type, histological grade, status of resection margins, and vascular invasion [68]. The appropriateness of adjuvant therapy and the prediction of outcome for the patient are, to a large extent, based on the pathologic assessment of the local disease and other tissue-based prognostic factors in the resection specimen [68].

The best estimation of prognosis in colorectal cancer is related to the anatomic extent of the disease determined on pathologic examination of the resection specimen [68]. The staging of colorectal cancer describes how far the disease has spread. The first stage classification was postulated by Dr. Cuthbert Dukes in 1932 and is known as Dukes staging. After revision of the Dukes classification colorectal cancer tumors were staged as follows: Dukes A cancer only affects the innermost lining of the colon or rectal wall or slightly grow into the muscular layer. Dukes B cancer shows a growth through the muscle layer, but are considered of better prognosis than Dukes C (the cancer spreads to at least one lymph node in that area) and Dukes D (the cancer has spread to distant organs like liver or lung).

The Dukes classification has recently been substituted by a more detailed classification called TNM. Between 1943 and 1952, Pierre Denoix came up the new system of classification. This system is based on size and extension of the primary tumor, considers the possibility of affected lymph nodes and includes the presence of distant metastasis.

In the TNM system, the designation "T" refers to the local extent of the primary tumor at the time of diagnosis, before the administration of treatment of any kind [69]. The designation "N" refers to the status of the regional lymph nodes, and "M" refers to distant metastatic disease, including nonregional lymph nodes [68]. The numbers after the letter describe the increasing severity (0 to 4, see Table 1). Pathologic classification typically is based on gross and microscopic examination of the resection specimen of a previously untreated primary tumor [68].

Table 1: Staging of colorectal cancer [69].

	TNM Classification			Dukes Classification	5-year relative survival rate (%)	
Stages	**T[1]**	**N[7]**	**M[11]**	**Stages**	**Colon**	**Rectum**
0	Tis[2]	N0[8]	M0[12]	-	-	-
I	T1[3]	N0	M0	A	92	87
	T2[4]	N0	M0	A		
IIA	T3[5]	N0	M0	B	87	80
IIB	T4[6]	N0	M0	B	63	49
IIIA	T1-T2	N1[9]	M0	C	89	84
IIIB	T3- T4	N1	M0	C	69	71
IIIC	Any T	N2[10]	M0	C	53	58
IV	Any T	Any N	M1[13]	D	11	12

NOTE: [1]T-Primary tumor, [2]Tis-Carcinoma in situ: intraepithelial or invasion of lamina propria, [3]T1-Tumor invades submucosa, [4]T2-Tumor invades muscularis propria, [5]T3-Tumor invades trough the muscualris propria into the subserosa, or into non-peritonealized pericolic or perirectal tissues, [6]T4-Tumor directly invades other organs or structures, and/or perforates visceral peritoneum, [7]N-Reginal lymph nodes, [8]N0-No regional lymph node metastasis, [9]N1-Metastasis in 1 to 3 regional lymph nodes, [10]N2-Metastasis in 4 or more regional lymph nodes, [11]M-Distant metastasis, [12]M0-No distant metastasis, [13]M1- Distant metastasis. Data taken from AJCC and American Cancer society [69,70].

However, additional histopathological variables have been shown to have stage-independent prognostic value. Histological grade has long been known to be of prognostic significance for different types of cancer, including colorectal cancer [68]. Tumor grade is the description of a tumor based on the level of differentiation of the tumor tissue as observed under the microscope and compared to the original surrounding tissue. If the cells of the tumor and the organization of the tumor's tissue are close to those of the normal colonic mucosa, the tumor is called well differentiated, low grade or G1-2, and these tumors tend to grow and spread at a slower rate than tumors that are poorly differentiated. High grade, G3-4 or poorly differentiated tumors grow faster and are more invasive. The factors used to determine tumor grade can vary between different types of cancer. The colorectal grade classification is the next:

G1: Well differentiated (low grade)
G2: Moderately differentiated (intermediate grade)
G3: Poorly differentiated (high grade)
G4: Undifferentiated (high grade)

1.4.2 Incidence and statistics of colorectal cancer

Colorectal cancer incidence rates are rapidly increasing in several areas, even in countries with historically low risk, like Spain, countries within Eastern Asia and Eastern Europe. This trends is based on the fact of a combination of factors that can be divided in non-modifiable risks factors including age (the likelihood of colorectal cancer diagnosis increases progressively from age 40, rising sharply after age 50). More than 90 % of colorectal cancer cases occur in people aged 50 or older. The incidence rate is more than 50 times higher in persons aged 60 to 79 years than in those younger than 40 years. Causes for colorectal cancer involve the following aspects: personal history of adenomatous polyps, inflammatory bowel disease, family history of colorectal cancer or adenomatous polyps, inherited genetic risk, and the environmental risk factors including changes in dietary patterns (excessive red meat and alcohol consumption, low dietary fiber), obesity, and an increased prevalence of smoking among others [71].

Fortunately, due to improved treatments and early detection the mortality rates are decreasing in Western developed countries. Colorectal cancer survival is highly dependent upon stage of disease at diagnosis, and generally spoken, the higher the stage, the lover the survival rates [72]. However in developing countries mortality rates are increasing due to increasing westernized lifestyle, and the limited in economic resources and health infrastructures [73].

The incidence rates of colorectal cancer are slightly higher in men than in women. Elevated rates of incidence were estimated in Central European countries per 100,000 — Slovakia (92), Hungary (87) and Czech Republic (81) in men, and in Norway (54), Denmark (53) and The Netherlands (50) in women. There is up to a fivefold variation in the incidence rates across Europe, with the lowest rates in the Balkan countries of Bosnia Herzegovina (30 in men, 19 in women), Greece (25, 17) and Albania (13, 11). Geographical patterns of mortality partially follow incidence, although mortality is high also in some countries with relatively low incidence rates (Moldova, Russia, Montenegro, Poland and Lithuania) (Figure 10 A and B) [35].

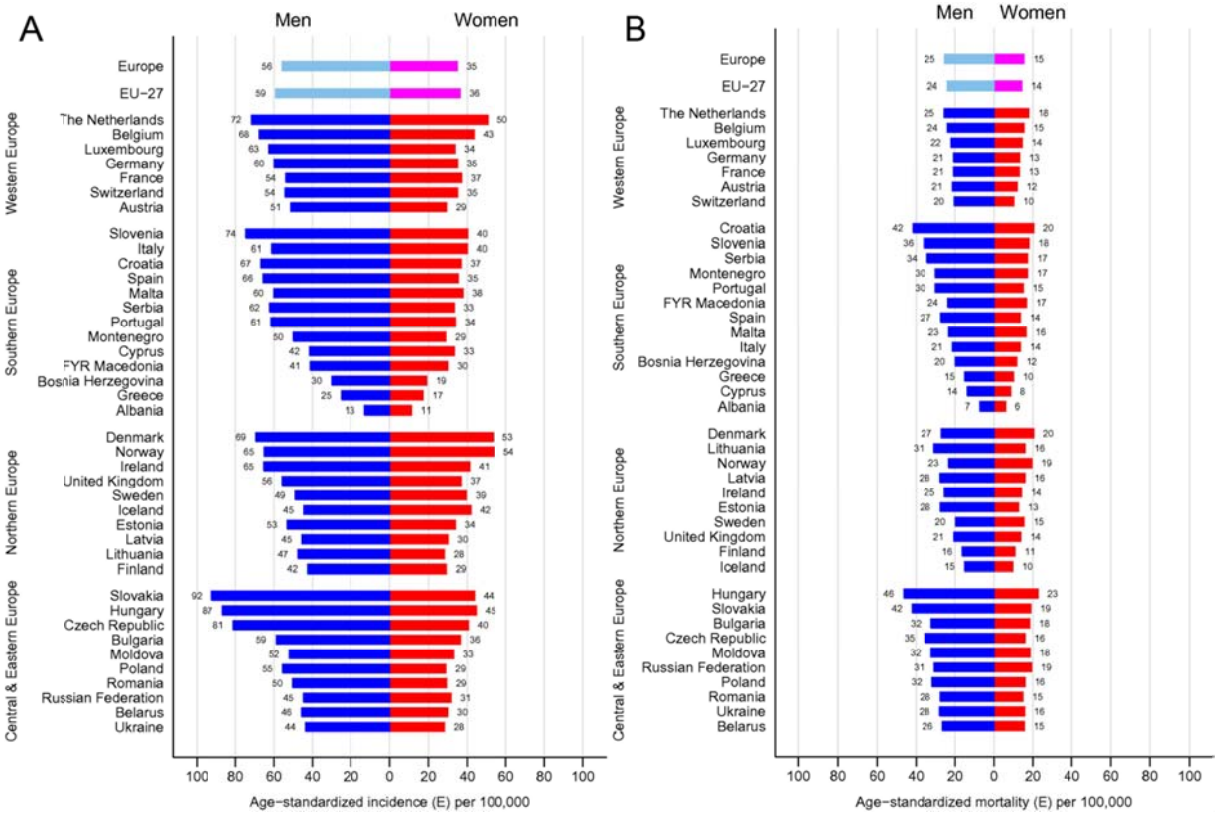

Figure 10: Age-standardized A) incidence rates and B) mortality rates by sex, area and country in Europe in 2012 for colorectal cancer. Modified from Farley et al [35].

1.4.3 Molecular pathways in colorectal cancer

1.4.3.1 The multistep nature from normal epithelium to colorectal cancer

In the early nineties, Fearon and Vogelstein postulated a genetic model, known as Vogelgram, for colorectal tumorigenesis, in which they proposed that colorectal cancer is the result of a series of genetic alterations involving oncogenes and tumor suppressor genes (Figure 11).

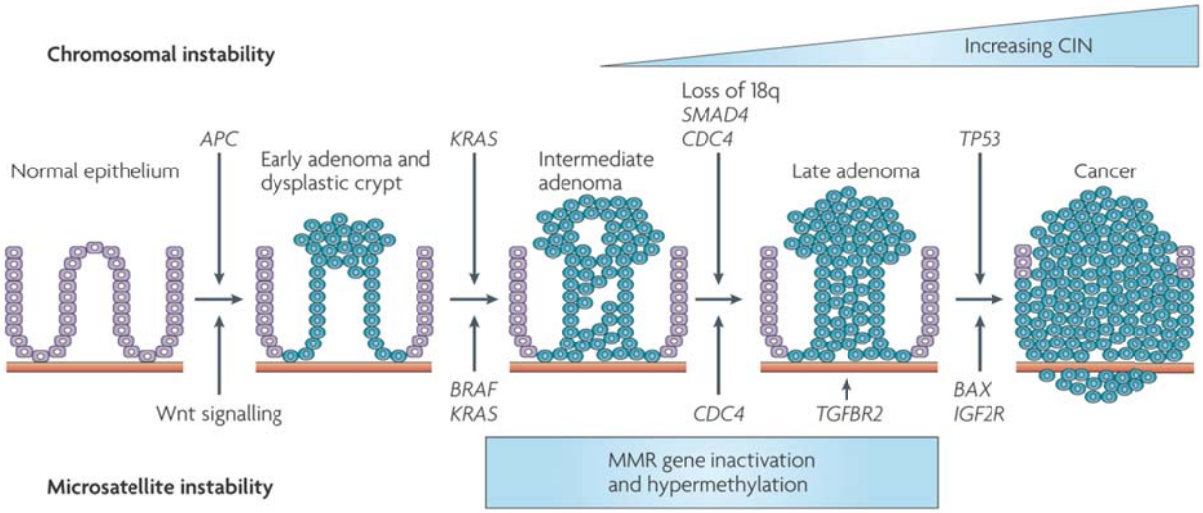

Figure 11: Vogelgram
In order to develop from normal epithelial tissue sporadic colorectal tumors, mutations in APC are required for tumor initiation. Larger adenomas and early carcinomas acquire mutations in KRAS, followed by loss of chromosome 18q with mutation SMAD4, which is downstream of transforming growth factor-β (TGFβ), and mutations in TP53 in carcinoma. These mutations are accompanied with genetic instability, either of microsatellite sequences through loss of mismatch repair (MMR), or of large sections of chromosomes through as yet undefined mechanism called chromosomal instability (CIN). Figure taken from Walther et al [74].

This model proposes that colorectal cancer develops through an adenoma to carcinoma sequence. Most colorectal carcinomas arise from adenomas which gradually progress through increases in size, dysplasia, and the acquisition of villus morphology. Although the process for adenoma progression is contunous, three discrete stages of adenoma formation are depicted in Figure 11. This model proposes that tumorigenesis is a multistep process; initiation, promotion and progression, that occurs throughout decades.

This model suggested that colorectal tumors arise from mutational activation of oncogenes coupled with the mutational inactivation of tumor suppressor genes (the latter changes predominating). Second, least four to five genes need to be mutated for the formation of a malignant tumor. Third, the alterations often occur according to a preferred sequence. The total accumulation of changes

rather than the order of appearance are responsible for determining the tumor's biological properties.

Fourth in some cases, mutant tumor suppressor genes appear to exert a phenotypic effect even when present in the heterozygous state; thus some tumor suppressor genes may not be recessive at the cellular level. The activation of oncogenes might occur by mutation or by amplifications and rearrangements. The inactivation of tumor suppressor genes can occur by allelic losses due to the loss of specific chromosomal regions [75].

In the model proposed by Fearon and Vogelstein points out the adenomatous polyp as the precursor lesion and provides evidence that in the majority of colorectal cancers the first event is the aberrant activation of the APC/β-catenin pathway, followed by RAS/RAF mutations and loss of p53 function at later stages [76]. In recent years this model has become more complex: Instead of a single pathway initiated by biallelic APC loss, and characterized by sequential accumulation of mutations, colon cancer is best characterized by at least three parallel progression mechanisms, each of which has distinct clinicopathologic features.

1.4.3.2 The distinct paths to colorectal cancer

When speaking about colorectal cancer, there are two categories that lead to this disease. Cancer can originate from sporadic mutations or it can originate due to hereditary mutations. It has been estimated that 15–30 % of colorectal cancers may have a major hereditary component, depending on the occurance of first or second-degree relatives. The majority of these cases are **hereditary nonpolyposis colorectal cancer (HNPCC)**, also called Lynch syndrome. In patients with HNPCC, germline defects in mismatch repair genes (primarily *MLH1* and *MSH2*) is the reason for a lifetime risk of colorectal cancer of almost 100 %, with cancers certain by the age of 40 years, on average [77]. Another significant subset of cases is associated with **familial adenomatous polyposis (FAP)**. This syndrome is defined by multiple adenomatous polyps (>100) and carcinomas of the colon and rectum; duodenal polyps and carcinomas; congenital hypertrophy of retinal pigment epithelium [78]. Furthermore, a few other syndromes constitute the remainder of such highly penetrant cases [78].

The remaining 70-85 % of the cases originate from **sporadic mutations** and underlie three distinct pathways: microsatellite instability (MSI), chromosomal instability (CIN), CpG island methylator phenotype (CIMP). Importantly, these three phenotypes are not mutually exclusive and may coexist in the same tumor to some extent (Figure 12) [76].

Genetic instability is important for molecular diversity in neoplasia, which is a prerequisite for the Darwinian selection that characterizes tumor formation and progression. Landmark experiments suggested that virtually all cases had genetic instability, either of microsatellite sequences through

loss of mismatch repair (MSI) or of large sections of chromosomes through as yet largely undefined mechanisms (CIN). This simple model suggests the two paths to develop colon cancer, with all cases having some measurable degree of genetic instability. The model became more complex when it was reported that some colon cancers seemed to fit in neither of the two paths. In parallel, DNA methylation became more prominent and was recognized as a very common epigenetic change in colorectal cancer. This specific pathway of intense DNA hypermethylation was described in colorectal cancer as the CpG island methylator phenotype – CIMP [79]. Furthermore, it is believed that aberrant methylation it plays a key role in colorectal cancer initiation.

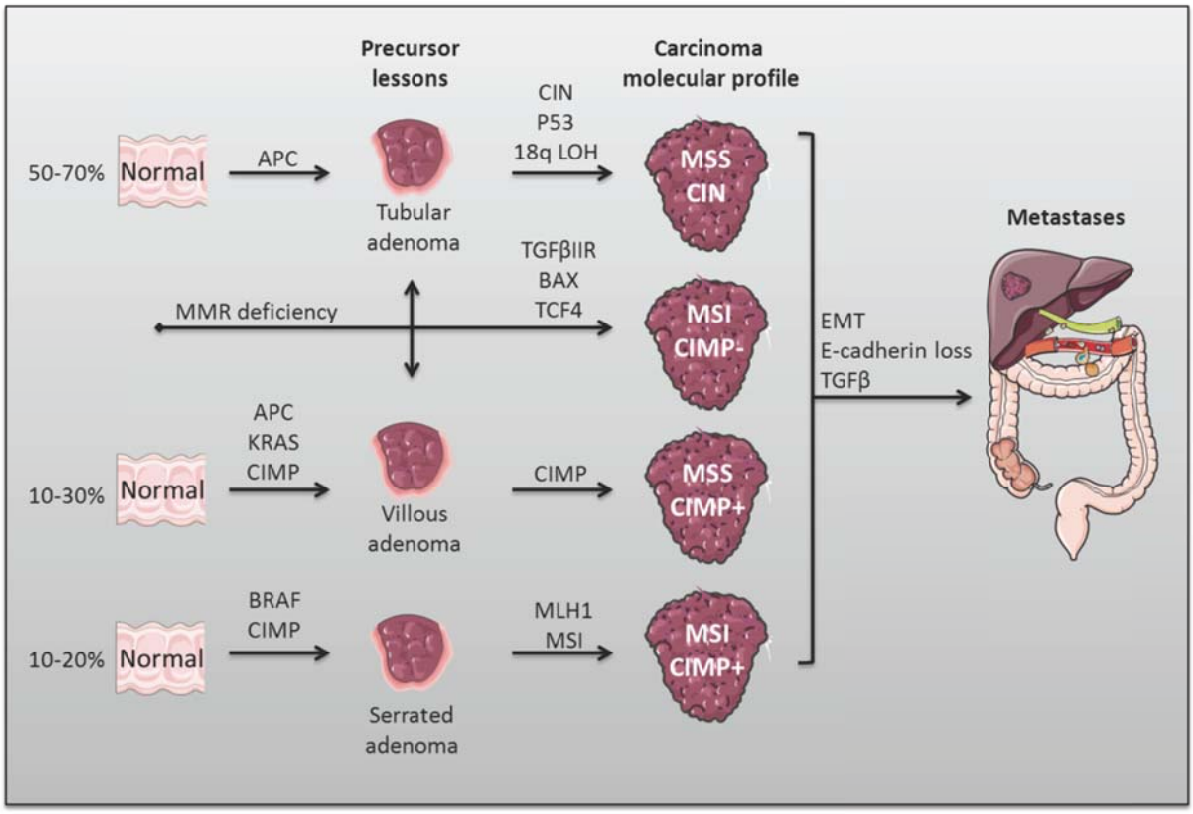

Figure 12: Genomic instability and multiple pathways in colorectal cancer pathogenesis.
Three distinct parallel pathways with the approximate indicated prevalence are implicated in colon cancer pathogenesis: traditional, alternative, and serrated. The traditional and serrated pathways are more homogeneous and clearly distinguishable; the alternative is more heterogeneous. Modified from [76].

1.4.3.3 Microsatellite instability (MSI) - DNA mismatch repair defects in colorectal cancer

DNA damage accumulates in cells over time as a result of exposure to exogenous chemicals and physical agents such as cigarette smoke and ultraviolet light, as well as endogenous reactive metabolites, including reactive oxygen and nitrogen species (ROS and NOS, respectively). Additionally, DNA damage occurs during DNA processing reactions (including DNA replication,

recombination, and repair). Unrepaired DNA damage can generate mutations in somatic or germline cells, which can alter the cellular phenotype causing dysfunction and disease. To prevent the cell's DNA from such deleterious effects and safeguard the integrity of the genome, cells possess multiple mechanisms to repair DNA damage. One such system is the pathway known as DNA mismatch repair (MMR) and is highly conserved among prokaryotes and eukaryotes [80,81].

Briefly, in eukaryotes heterodimeric MutSα (MSH2–MSH6) or MutSβ (MSH2–MSH3) proteins, bind specific base-mispairs or 'looped out' extra nucleotides in DNA. They then bind ATP and recruit MutLα (MLH1–PMS2) heterodimers into the recognition complexes [82]. Accessory proteins are used to create or identify specific strand ends for initiation of excision that proceeds through mismatches. Inactivation of MMR in human cells is associated with hereditary and sporadic intestinal cancers like non-polyposis colorectal cancer (HNPCC) [80].

Microsatellite instability (MSI) is easily detected due to the loss of mismatch-repair function, in which the inability to repair strand slippage within repetitive DNA sequence elements changes the size of the mononucleotide or dinucleotide repeats (microsatellites) that are scattered throughout the genome [83]. The DNA mismatch repair genes are believed to behave like tumor-suppressor genes in which two hits are required to cause a phenotypic effect. Nucleotide insertions or deletions occur randomly in all the microsatellite repeats in coding and non-coding sequences along the genome.

Patients with hereditary HNPCC, also known as the Lynch syndrome, have germ line defects in mismatch repair genes (primarily *MLH1* and *MSH2*) and are therefore predisposed for a lifetime risk of colorectal cancer of about 80 %, with cancers evident by the age of 45 years, on average [84]. On the other hand, in patents with a sporadic instability, representing the great majority, are over 60 years old, and biallelic inactivation of *MLH1* occur in a majority of the cases by MLH1 promoter hypermethylation (about 80 %), whereas the rates of LOH and somatic mutation of MLH1 are about 25 % and about 10 % respectively [85].

TGFβRII gene (transforming growth factor β (TGF-β) receptor type II) from the SMAD pathway is among the most frequently mutated genes in MSI tumors [86,87]. Other genes frequently mutated in MSI tumors are IFRIIR and BAX, EphB2, and Myo1A genes [43,88–90]. Furthermore, mononucleotide repeats in the DNA mismatch repair genes MSH6 and MSH3 have been reported to be mutated [91].

1.4.3.4 – Chromosomal instability (CIN)/microsatellite stability (MSS) - Genetic Instability in colorectal cancer

The majority of colorectal cancers have genetic instability caused by loss or gain of chromosome arms, chromosomal translocations or gene amplifications and are termed chromosomal instable (CIN) tumors or microsatellite stable (MSS) and make about 85 % of the sporadic colorectal cancers [92]. The majority of colorectal cancers have genetic instability caused by loss or gain of chromosome arms, chromosomal translocations or gene amplifications and are termed chromosomal instable (CIN) tumors or microsatellite stable (MSS) and make about 85 % of the sporadic colorectal cancers [78,93].

The model describing the transformation of adenoma to carcinoma by Fearon and Vogelstein emphasizes the somatic mutation events to the pathogenesis of colorectal cancer and helps to understand the molecular mechanisms of the development of CIN tumors (Figure 12) [78,94]. In fact, the accumulation of a characteristic set of mutations in specific tumor suppressor genes and in oncogenes which belong to different signaling pathways. These mutations constitute a second hit, which coupled with the chromosomal defects observed in this type of colorectal cancer.

1.4.3.5 CpG island methylator phenotype (CIMP) - Epigenetic changes in colorectal cancer

Hypermethylation of promoter CpG islands is a mechanism of inactivation of tumor suppressor genes. Aberrant gene promoter methylation has been widely observed to occur in human colorectal cancer. The promoters of approximately 50 % of all genes contain CpG islands. Hypermethylation of these CpG islands seems to be associated with silencing of downstream transcriptional units, which may reflect an epigenetic mechanism that reinforces long-term gene silencing following more transient chromatin modifications. There is a generalized decrease in the total level of DNA methylation (i.e., hypomethylation) in the cancer cells compared with adjacent normal tissues, and DNA hypomethylation is also observed in adenomatous polyps. Although the global trend in colorectal cancer cells is hypomethylation, CpG islands at various promoters show increased methylation often linked to transcriptional silencing of genes that these promoters regulate [78]. This mechanism of transcriptional silencing is especially significant for cancer development/progression when it is responsible for the inactivation of tumor suppressor genes. Recent comprehensive work on human colorectal cancer methylation suggests that the largest number of methylation changes in colorectal cancer occurs not in promoters or in CpG islands but rather more than 2 kb away from the CpG islands at so called CpG island shores [95].

The group of colorectal cancers that show hypermethylation at many different CpG-rich elements fits the CpG island methylator phenotype (CIMP) model (Figure 12). In this context, it has been observed that a subset of these tumors possess an exceptionally high frequency of methylation of some CpG islands, thus suggesting the presence of a distinct trait referred to as CpG island methylator phenotype (CIMP) [96]. The underlying mechanisms of this phenotype are still quiet unknown. An early event in CIMP tumors appears to be a mutation in the *BRAF* proto-oncogene, which promotes proliferation of colonic epithelial cells [97]. In addition, most CIMP colorectal cancer are characterized by promoter CpG island methylation of *MLH1*, resulting in its transcriptional inactivation thus causing a MSI phenotype. Moreover, increased expression of DNA methyltransferases (DNMTs) has been reported in colorectal cancers [98] but there seems to be some discrepancies about this finding [99]. Although it has been more than a decade since CIMP was first identified in colorectal cancer, the path to accepting these tumors, as an etiologically and clinically distinct group of the disease has not been without controversy, and to date, the real cause of CIMP remains unknown.

A lot is known about the mutational status of genes driving colorectal tumorigenesis. However, mutation is not the only event, which inactivates a gene. Epigenetic events, like methylation of the promoter region, can also cause deactivation of genes important for tumor suppression. However, a list, ranking genes involved in colorectal cancer progression/activation inactivated by methylation, is currently lacking.

1.4.3.6 Somatic oncogene and tumor suppressor gene mutations in colorectal cancer

Cancer cell isolated from tumors often display multiple visible defects during mitosis, including abnormal spindle morphology, centrosome amplification, anaphase bridges and lagging chromosomes [100], which lead to a karyotypically unstable and heterogeneous cell population. The consequence of CIN is an imbalance in chromosome number (aneuploidy), subchromosomal genomic amplifications, and a high frequency of loss of heterozygosity (LOH).

Coupled with the typical karyotypic abnormalities observed in CIN tumors is the accumulation of a characteristic set of mutations in specific tumor suppressor genes and oncogenes belonging to different signaling pathways.

APC inactivation is the perhaps the initiation event in 70-80 % of colorectal cancer cases. Due to the high frequency of *APC* gene defects in adenomas and colorectal cancer and the possibility that *APC* inactivation may play a pivotal role in initiating the adenoma-carcinoma pathway, it has been argued that *APC* plays gatekeeper a role in normal colorectal epithelial cells. The term gatekeeper genes

serves to designate genes that are responsible for maintaining a constant cell number in renewing populations and whose mutation leads to a permanent imbalance of cell division over cell death [101].

Recently, it has been proposed that the genomic landscape of colorectal cancer is composed of only a handful of commonly mutated genes which are called "mountains" and a larger number of infrequently mutated genes called "hills" [102]. The large number of hills detected in the colorectal cancer landscape reflects alterations in certain pathways, which are clearly singled out as key factors in tumor formation [102].

The most common mutations occurring in colorectal cancer account for the inactivation of the *APC*, *TP53*, components of the TGFβ pathway tumor suppressor pathway such as *SMAD4* or *TGFBR2* and for the activation of *KRAS*, *BRAF* and *PI3K* oncogenes [78,103,104], which will be discussed in more, detail below.

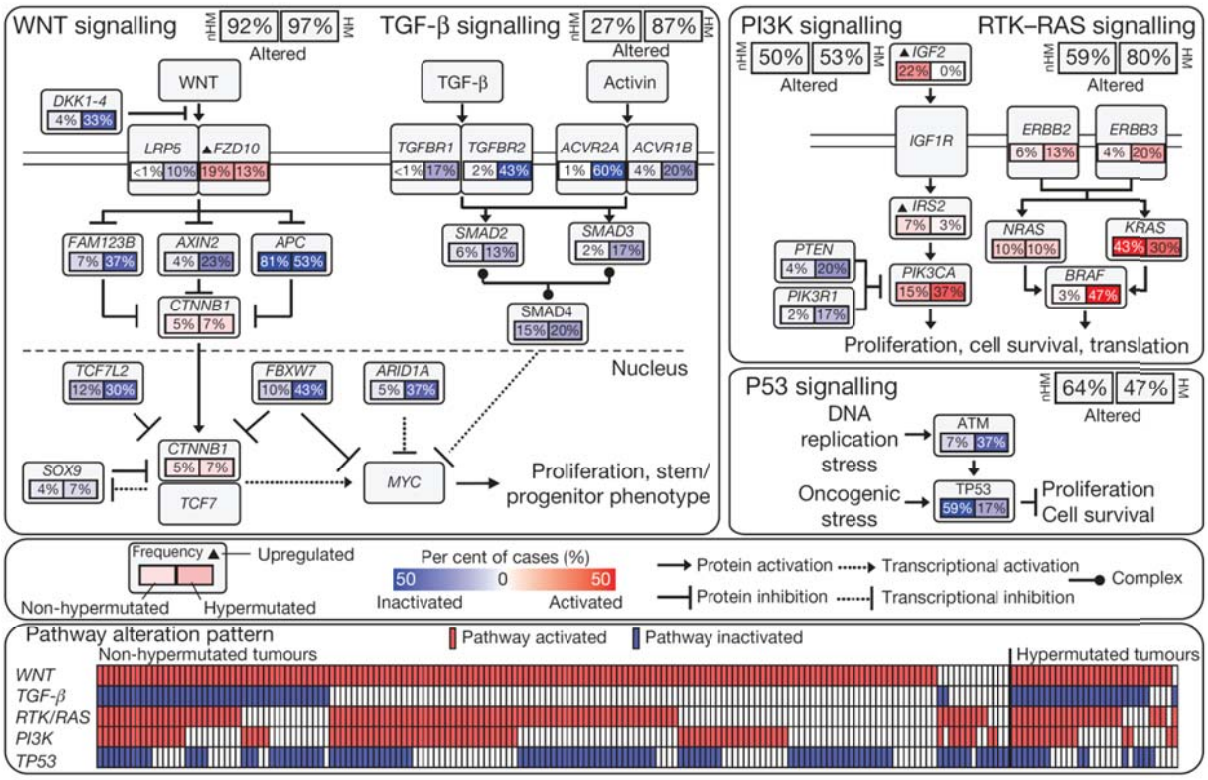

Figure 13: Genetic alterations leading to deregulation of signaling pathways in colorectal cancer.
The most commonly deregulated signaling pathways in colorectal cancer are WNT, TGF-β PI3K, K-Ras and P53 signaling. By using a set of 224 colorectal cancer samples, the Cancer Genome Atlas (TCGA) consortium identified tumors with low mutation rates (84 % of cases) of <8.24x106 bases (median number of non-silent mutations, 58) termed as non-hypermutated and with high mutation rates >12x106 (median number of total mutations, 728) classified as hypermutated. Genetic alterations are defined by somatic mutations, homozygous deletions, high-leve focal amplifications, and in some cases, by significant up- or downregulation of gene expression (IGF2, FZD10, and SMAD4). Alteration frequencies are expressed as a percentage of all cases. Red denotes activated genes and bule denotes inactivated genes. The bottom panel shows for each sample if at least one gene in each of the five pathways described in this figure is altered. Source [104].

APC- Wnt signaling pathway is considered a trigger point event in most colorectal cancer. A frequent mutation in colorectal cancer inactivates the gene that encodes the APC protein. The non-functional APC protein (repressor β-catenin) leads to a deregulation of Wnt signaling and therefore this pathway is constitutively activated. Familial adenomatous polyposis (FAP) an inherited cancer-predisposition syndrome is due to the presence of germ line APC mutations in which more than 100 adenomatous polyps can be developed. It is remarkable to note that the risk of colorectal cancer by the age of 40 years in carriers of the mutant gene is of almost 100 %. The inactivation of both copies of APC by somatic mutations and deletions are present also in most sporadic colorectal adenomas and cancers (70 to 80 % of sporadic colorectal adenomas and carcinomas have somatic mutations that inactivate APC) [64,81,104]. Several results suggest that somatic *APC* mutations are an early and perhaps rate-limiting event in the development of most adenomas [78]. In very small adenomas as well as in advanced adenomas and carcinomas *APC* mutations have almost the same frequency, in contrast to the situation for some other somatically mutated genes in colorectal tumors [75,78,105]. Moreover, somatic APC mutations are present even in the earliest lesions, including microscopic adenomas composed of a few dysplastic glands [78,106] (Figure 13). 50 % of colon tumors with intact APC have been reported to contain mutations leading to a gain-of-function in β-catenin gene (CTNNB1). β-catenin in its primary structure has an N-terminal portion containing the phosphorylation sites for GSK3B important for the degradation. β-catenin mutants lacking these phosphorylation sites are more stable than wild-type protein. Certainly point mutations in or near the four Ser/Thr phosphorylation sites are commonly found in cancer [107]. In the β-catenin degradation complex Axin and conductin interact with β-catenin, GSK3B and APC [108,109]. Axin mutats containing a S614 (phosphorylation site of GSK3B) amino acid substitution caused >4-fold reduction in β-catenin binding. Has been aslo reported in another study that Wnt-induced dephosphorylation of axin leads to reduced β-catenin binding [110].

KRAS and ***BRAF*** - It is known that almost 50 % of colorectal cancers contain *KRAS* somatic mutations. *KRAS* mutations are certainly not required for adenoma initiation however they contribute to colorectal adenoma development. *KRAS* mutations are commonly observed in aberrant crypt foci considered as flat colonic epithelial lesions with altered glandular architecture in which dysplasia is usually absent. *KRAS* mutations are also present in a significant fraction of lesions that are minimally likely to progress to carcinoma like the hyperplastic polyps.

The frequency of *KRAS* mutations in adenomatous polyps clearly depends on the size and degree of dysplasia of the lesion. Adenomas smaller than 1 cm show only 10 % of KRAS mutations, whereas are observed in almost 40 % to 50 % in adenomas greater than 1 cm. Even though the observations that *KRAS* mutations are present in some colorectal lesions with non-malignant potential (e.g., aberrant crypt foci and hyperplastic polyps), when present, mutant KRAS alleles, appear to play an important role in the behavior of advanced colorectal cancer cells [78,81,104].

The gene coding for BRAF is mutated in approximately 5–10 % of colorectal cancers. BRAF is a protein kinase directly activated by RAS proteins that activates the MAPK effectors MEK1 and MEK2. Interestingly, BRAF mutations seem to be deeply associated with the CIMP phenotype (Figure 13) [78,81,111,112].

Phosphatidylinositol 3-kinase (*PI3K*) - Activating somatic mutations in PI3KCA which encodes the catalytic subunit of PI3K appear in one third of colorectal cancers. Less frequents genetic alterations that may include loss of PTEN, an inhibitor of PI3K signaling, likewise amplification of insulin receptor substrate 2 (IRS2), an upstream activator of PI3K signaling, and co-amplification of AKT and PAK4, which are downstream mediators of PI3K signaling (Figure 13) [78,81].

TGF-β - Tumor-Suppressor Pathway – The third step in the progression to colorectal cancer is the mutational inactivation of TGF-β signaling. LOH of chromosome 18q is observed in 70 % of colorectal cancers; 50 % of large and late-stage adenomas; and fewer than 10 % of small and in early-stage adenomas. Two genes located in chromosome 18q are SMAD2 and SMAD4. Both genes encode for proteins that are involved in TGF-β receptor complex. SMAD4 mutations are found in 10–20 % of colorectal cancers, and SMAD2 in 5-15 % of colorectal cancers. 5-43% of tumors has inactivating mutations in the TGF-β type II receptor (TGF□IIR) in colorectal cancers. In addition, in approximately 15 % of MSS-colorectal cancers, somatic mutations have been observed to lead to TGF□IIR inactivation (Figure 13) [78,81,104].

TP53 - The second genetic step in colorectal cancer is the inactivation of the p53 pathway by mutation of TP53. In most tumors, the two TP53 alleles are inactivated, normally due to a combination of a missense mutation that inactivates the transcriptional activity of p53 and a 17p chromosomal deletion that eliminates the second TP53 allele. In many colorectal cancers TP53 remains wild-type, but the pathway might be attenuated by mutations in the pro-apoptotic Bax inducer of apoptosis and other genes. Mutation of the TP53 gene has been shown to occur in the adenoma to carcinoma sequence and most probably occur before metastasis [78,81,104]. The p53 protein

functions as a key transcriptional regulator of genes that encode proteins with functions in cell-cycle checkpoints at the G1/S and G2/M boundaries, in promoting apoptosis, and in restricting angiogenesis [78,113](Figure 13). Selection for *TP53* mutations in colorectal cancer maybe due in part to the loss of *TP53* transcription function in regulating key target genes (e.g., $p21^{WAF/CIP1}$, *PUMA, BAX*, and TP53-inducible gene 3 (*PIG3*)) [113]. However, the missense mutant TP53 proteins present in cancer cells may positively contribute to the cancer phenotype via varied factors and mechanisms [78,113].

In normal cells, MDM2 inhibits p53 function by modulating its transcriptional activity by preventing its interaction with the general transcription machinery. MDM2 forms a heterodimeric complex with MDM4 that promotes the degradation of p53. ARF adds another level of control to the system by inhibiting MDM2 function, the expression of which is in turn also repressed by p53 [114].

We know a great deal about tumor suppressor and oncogenes in colorectal cancer. Although, a full list of those genes is currently not existing, every year new genes with tumor suppressor activity (e.g. *MYO1A* [43] or *RHOA* [44]) or oncogenic activity (e.g. *INPP4B* [115] or *SOCS1* [116]) are discovered. However, much of those genes need yet to be discovered in order to better understand the underlying mechanisms of colorectal cancer.

1.4.4 Treatment of colorectal cancer

1.4.5 Treatment of colon cancer by stage

The treatment and prognosis for patients with colon or rectal cancer depends on the tumor location and stage at the time of diagnosis (Table 1).

Stage 0, which have not grown beyond the inner lining of the colon, are curable with surgery, mainly done by polypectomy (removing the polyp) or local excision through a colonoscope.

Stage I and II - These cancers have grown through several layers of the colon, but they have not invaded to neighboring lymph nodes. Stage I includes cancers that were part of a polyp. If the polyp is removed completely, with clean margins, no further treatment is required. Surgery to remove the cancer and nearby lymph nodes is the most common treatment for early stage (**stage I and II**) colon (94 %) and rectal (74 %) cancer. Colon resection (colectomy) is applied if the tumor is too big to be removed by local excision and is more commonly used for rectal cancer (26 %) than for colon cancer

(7 %), and is often temporary [117]. Occasionally, chemotherapy is given as adjuvant therapy. Adjuvant therapy is an additional cancer treatment given after the primary treatment (mostly surgery) to lower the risk of cancer recurrence.

Stage III - In this stage, the cancer has spread to nearby lymph nodes, but it has not yet spread to other parts of the body. Treatment for these cancers includes first surgery, which is followed by adjuvant chemotherapy. To ensure that, even with clean surgical margins (cut edges 15 cm away from the tumor do not contain tumor mass) possible remaining tumor cells are removed or destroyed, chemotherapy or radiation therapy can be given to the patients. It is of great importance that those cells are targeted and destroyed, in order to avoid the risk of the cancer recurrence. Radiation therapy can be given to the area of the abdomen where the cancer was growing. Chemotherapy is given as adjuvant therapy.

Stage IV - The cancer has spread from the colon to distant organs and tissues. Colorectal cancer most often spreads to the liver, lung, peritoneum (the lining of the abdominal cavity), or distant lymph nodes. Surgery in stage IV cancers is normally not to cure the patient from the cancer but rather to prevent or relieve the patient from the symptoms of the disease. However, if only a few small areas of cancer have metastasized, both primary tumor and metastasis can be removed surgically. When the metastases cannot be surgically removed because they are too large or there are too many of them, palliative chemotherapy is given to the patient before and after surgery. In patients with good response to the palliative treatment showing reduced size/number of metastatic lesions, rescue surgery can be attempted. Chemotherapy in an adjuvant form is usually administered after surgery in an attempt to reduce the probability of recurrence in these patients.

A study showed that in 2010 57.4 % of the stage IV colorectal cancer patients underwent primary tumor resection in the United States. The resulting medium 5-year survival rate is 17.8 % compared to the non-surgically treated patients [118]. However, only 2.2 % achieved a sustained, complete response after chemotherapy, when surgery was not included in the treatment process [119].

1.4.5.1 Chemotherapeutic agents available for colorectal cancer treatment

The drugs currently accepted for routine treatment of colorectal cancer patients are 5-fluorouracil, leucovorin, irinotecan, oxaliplatin, the multi-kinase inhibitor regorafenib and monoclonal antibodies against the vascular endothelial growth factor or the epidermal growth factor receptor (Table 2).

Table 2: Chemotherapeutic agents used for colorectal cancer treatment.

Drugname	Abbreviation	Target/Effect
5-Fluorouracil	5-FU	Irreversible inhibition of thymidylate synthase, which blocks synthesis of the pyrimidine thymidine, that is a nucleoside required for **DNA replication.**
Leucovorin (folinic acid)	-	Given as synergistic combination with 5-FU, due to the enhancement of the inhibition of DNA synthesis.
Oxaliplatin	-	Forms both inter- and intra-strand cross-links in DNA, which prevent **DNA replication and transcription**, causing cell death.
Irinotecan	CPT-11	Activated by hydrolysis to SN-38, it inhibits topoisomerase I by stabilizing the cleavable complex between topoisomerase I and DNA, resulting in DNA breaks, which inhibit **DNA replication** and trigger apoptotic cell death.
Capecitabine	-	Precursor drug to 5-FU, taken orally.
Bevacizumab	-	Monoclonal antibody. Designed to block vascular endothelial growth factor (VEGF), and prevent the growth of new blood vessels, including normal blood vessels and blood vessels that feed tumors. Unlike chemotherapy that attacks the cancer cells, bevacizumab indirectly blocks tumor proliferation through **blockage of the blood supply** that feeds the tumor.
Cetuximab	-	Monoclonal antibody. Inhibits the Epidermal Growth Factor Receptor (EGFR) and so **stops signaling of uncontrolled proliferation.** Only patients whose tumors are *KRAS* wild-type should receive this drug.
Panitumumab	-	Monoclonal antibody. Fully human monoclonal antibody specific to EGFR.
Regorafenib	-	Multi-kinase inhibitor which, targets VEGFR1-3, c-KIT, TIE-2, PDGFR-β, FGFR-1, RET, RAF-1, BRAF and p38 MAP kinase and so **inhibits angiogenesis and proliferation.**

5-FU is an irreversible inhibitor of thymidylate synthase, which was first used in 1957. It was for four decades the only treatment available for locally advanced and metastatic colorectal cancer. 5-FU has been used in a variety of administration routes, dosing patterns, and combination regimens and is most commonly used with leucovorin, a reduced folate, which stabilizes the binding of fluorouracil to thymidylate synthase, thereby enhancing the inhibition of DNA synthesis (5-FU/LV) [120].

Irinotecan is often administrated in combination with the 5-FU and leucovorin, in a regime known as FOLFIRI [121]. Oxaliplatin is a platinum-based drug which causes DNA distortion preventing replication and transcription, which ultimately leads to cellular apoptosis [122]. Oxaliplatin is typically administered in combination with fluorouracil and leucovorin, in a regime known as FOLFOX [123]. In Europe, the first line of treatment is mainly based in FOLFOX (5-FU, leucovorin and oxaliplatin) or FOLFIRI (5-FU, leucovorin and irinotecan) backbones [124]. The administration of capecitabine is recommended as adjuvant chemotherapy in stage III patients. For these patients, capecitabine provides equivalent outcome to intravenous 5-FU and leucovorin, with significantly fewer side effects [125]. Stage IV (metastatic) colorectal cancers are unlikely to be cured with surgery alone and therefore various chemotherapeutic regimes are given to these patients. The different chemotherapeutic agents available are only effective in small subsets of patients (10-30 %) and objective response rates (Complete response plus partial response) are far from optimal (Table 3).

Table 3: Objective response rates of frequent regimes used for metastatic colorectal cancer treatment.

Regimen	Clinical information	Complete response (%)	Partial response (%)	Reference
Leucovorin/5-Fluorouracil	First-line treatment in advanced colorectal cancer	0.5	21.4	[123]
Leucovorin/5-Fluorouracil+Oxaliplatin	First-line treatment in advanced colorectal cancer	1.4	48.6	[123]
FOLFOX	First-Line Treatment in advanced colorectal cancer	10	41.4	[126]
FOLFIRI	Unresectable metastatic colorectal cancer	6	35	[127]
FOLFIRI+Oxaliplatin (FOLFOXIRI)	Unresectable metastatic colorectal cancer	8	58	[127]

However, the remaining 70 % of the patients have no clinical benefit and unnecessarily suffer the significant side effects associated with irinotecan treatment, including diarrhea, vomiting, mucositis, neutropenia, and neutropenic complications, causing treatment discontinuation in >10 % of cases and as many as 1 % of drug-related deaths [128].

Monoclonal antibody-based therapies have been included in the last years in the clinic in combination with conventional chemotherapeutic agents. The European Society of Medical Oncology recommends the use of monoclonal antibodies against the vascular endothelial growth factor (bevacizumab, Avastin) and the epidermal growth factor receptor (cetuximab, Erbitux and panitumumab, Vectibix) in combination with cytotoxic treatments in selected patients with metastatic disease, as these regimens have been associated with improved outcomes compared with chemotherapy alone [129].

All drugs listed in Table 2 except causing severe side effects. Since the organs in the gastrointestinal tract, specifically the small and large intestine, are constantly renewing the epithelial layer, those cells are also targeted by those agents. Therefore, most patients have severe digestive problems, including nausea, vomits, and diarrhea. Treatments with bevacizumab increase the incidence of venous thromboembolism and cetuximab and panitumumab are known to commonly cause a skin rash [130].

1.4.5.2 Strategies to improve the objective response rate in chemotherapy for colorectal cancer patients

In order to overcome the situation of such low objective response rates two strategies can be followed: new and better **biomarkers** should be discovered in order to pick patient suitable for a given chemotherapeutic agents and so improve the objective response rate. In addition, new

therapeutic **targets** should be identified, to open new possibilities for chemotherapeutic treatments using specific inhibitors of those targets.

There are two types of biomarkers: prognostic and predictive markers. Prognostic markers define the natural history of the disease, when no treatment is involved. Predictive markers, however, define the clinical outcome to a specific treatment.

The most powerful prognostic marker is the stage of a tumor, defined with the TNM staging system. Additional features of the tumor, along with clinical features such as obstruction and perforation at the time of diagnosis, are used to better define the poor prognostic subsets of patients (ASCO risk criteria [131]).

The MSI status can also be used as a prognostic marker of colorectal cancer. Several retrospective studies in patients with stage II/III have shown that those with MSI-H tumors have a more favorable stage-adjusted prognosis compared with patients exhibiting MSS or MSI-L [132–134]. For the detection of the MSI status of colorectal tumors, molecular testing relies on the evaluation of microsatellite instability of a panel of microsatellites commonly referred to as the "Bethesda panel", which is widely used in clinical and research testing [135] (Boland et al., 1998). Immunohistochemistry of MLH1, MSH2, MSH6, and PMS2 [136–138] is also used to detect the MSI status of cancer in patients.

Furthermore, the carcinoembryonic antigen (CEA), a high molecular weight glycoprotein, has been used as a biomarker since high serum levels of CEA have been associated with disease progression or recurrence [139].

In the adjuvant setting, there is currently a similar lack of biomarkers capable of predicting response to irinotecan-based treatment. The PETACC3 study, a randomized phase III trial comparing fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of 3,278 stage II and III colon cancer patients [140] did not identify a strong predictive value for several molecular markers investigated, including microsatellite instability, genotyping of *KRAS*, *BRAF*, *UGT1A1*, loss of heterozygosity in chromosome 18q and expression levels of *TERT*, *SMAD4*, *TP53*, *and TYMS* [141–145] (and unpublished findings of the PETACC3 trial). Other large studies, however, have shown that microsatellite instability but not the levels of expression of the cyclin dependent kinase inhibitor (*p27Kip1*) could be useful to predict response to irinotecan adjuvant treatment [146,147].

The most reliable molecular marker commonly used in clinical practice is KRAS. It is an effector molecule responsible for signal transduction from ligand-bound EGFR to the nucleus [148]. Mutations in *KRAS* were found to be associated with resistance to anti-EGFR therapy [149–151]. Therefore it is recommends that the mutational status of *KRAS* is elucidated, both in the primary tumor or metastasis. McLeod and Murray reported a correlation between mutations in *KRAS* and poor prognosis [152 153]. However, the prognostic value of KRAS remains weak and it is of no use for the clinic management of advanced colorectal cancer [142,154] (
Table 4).

Table 4: Prognostic and predictive value of colorectal cancer biomarkers based on stage at presentation and availability as a standard-of-care test in routine clinical practice.

Marker	Prognostic (stage II/III)	Prognostic (stage IV)	Predictive	Standard-of-care test
MSI	Yes	Weak	Yes (adjuvant 5-FU therapy)	Yes
18q LOH	No	No	No	No
KRAS	No	Weak	Yes (anti-EGFR therapy)	Yes
BRAF	Weak	Weak	Yes (anti-EGFR therapy)	Yes
PIK3CA	Weak	No	No	No
PTEN	No	Weak	No	No

NOTE: Information taken from George and Kopetz [155].

On the other hand, new therapeutic targets are needed in order to improve the objective response in patients with advanced colorectal cancer. Apart from the therapeutic targets like topoisomerase I, which is inhibited by irinotecan, or thymidylate synthase, inhibited by 5-fluorouracil, the more modern therapeutic targets are inhibited by monoclonal antibodies. The therapeutic targets of these monoclonal antibodies and the corresponding inhibitors used for chemotherapy of advanced colorectal cancer are discussed in detail below.

Targeting EGFR - EGFR is a tyrosine kinase receptor that belongs to the ErbB family and is abnormally expressed and activated in cancer cells in many tumor types including colorectal cancer. Following stimulation by its natural ligands, EGFR initiates signal transduction cascades that promote cell division, migration and angiogenesis, and inhibit apoptosis [156]. Therefore, it was evident, to develop agents that are able to block this signaling cascade through blockage of EGFR. Cetuximab and panitumumab, both monoclonal antibodies, and direct inhibitor of EGFR are nowadays used in the colorectal cancer therapy. However, clinical trials revealed that patients with a colorectal tumor bearing mutated *KRAS* (in codon 12 or 13) did not benefit from cetuximab, whereas patients with a tumor bearing wild-type *K-ras* did benefit from cetuximab[157,158].

Targeting the angiogenic process – Since the past decade, targeting the neovascularisation process of growing tumors has been widely used in clinical treatment of advanced colorectal cancer patients. Strong evidence that links tumor growth and metastasis with the angiogenesis process in most human tumors, including colorectal cancer, led to the development of specific inhibitors of the angiogenic process. A clear correlation between the microvessel density in the pathology specimens and progression-free and overall survival in patients with colorectal cancer has been demonstrated. VEGF is the most potent and specific angiogenic factor, and its expression in colorectal cancer has been demonstrated to correlate with recurrence and prognosis [156]. In *in vivo* models, the administration of anti-VEGF monoclonal antibodies reduced the size and the number of liver metastasis from human colorectal cancer cell lines [159] and showed a synergic effect when combined with some cytotoxic drugs [160]. Bevacizumab (Avastin) is a recombinant humanized anti-VEGF monoclonal antibody that is used in many tumor types including colorectal cancer [156]. When bevacizumab was given in combination with irinotecan, 5-fluorouracil and locovorin, the progression-free survival improved compared to the chemotherapy alone [161]. However, the effects on objective response rate and overall survival were less consistent in the first and second line advanced colorectal cancer settings. Generally, the benefit varied depending on the choice of chemotherapy backbone given to the patients, but was seemingly greater with irinotecan-based regimens [162].

2 Aims of the study

Colorectal cancer is a disease highly regulated by genetic and epigenetic alterations. Activation of oncogenes and inactivation of tumor suppressor genes are key hallmarks of cancer progression. Methylation of CpG dinucleotides in the promoter region of genes involved in the oncogenic process has been shown to be a key process contributing to tumor initiation and/or progression and significantly contributes to the profound expression reprograming of colorectal tumor cells. However, these molecular events are not fully understood, even in the well-studied colorectal cancer. Furthermore, response rates to the standard chemotherapeutic agents currently used for patients with advanced colorectal cancer are below 30 %. However, understanding better the molecular events involved in colorectal cancer can ultimately improve the colorectal cancer treatments, because genes important for tumor progression or new therapeutic targets as well as tumor suppressor genes can be identified.

Here we hypothesize that genes highly expressed in fast growing tumors are important for tumor progression and that those genes can be used as possible novel therapeutic targets. On the other hand, to better understand the molecular mechanisms driving colorectal cancer, a detailed list of genes whose transcription is frequently regulated by CpG methylation is currently lacking.

Therefore, the aim of the study is to use high-throughput analysis to identify genes, involved in tumor proliferation, which could be novel therapeutic targets, and genes, whose transcription is frequently regulated by CpG methylation in colorectal cancer.

Therefore, the specific aims of this study were:

- Aim 1: to identify molecular mechanisms important for the rapid proliferation observed in aggressive colorectal tumors.
- Aim 2: to identify and preclinically validate new candidate therapeutic targets for colorectal cancer patients.
- Aim 3: to identify new genes regulated by DNA methylation during tumorigenesis in CRC.
- Aim 4: to investigate the tumor suppressor activity of genes silenced epigenetically during colorectal carcinogenesis.

3 Materials and Methods

3.1 Materials used in this study

3.1.1 Human colorectal cancer cell lines

A total of 55 colorectal cancer cell lines were used (see Table 5): Caco2, Colo201, Colo205, Colo320, DLD1, HCT116, HCT15, HCT8, HT29, HUTU80, LOVO, LS1034, LS174T, LS513, RKO, SKCO1, SNUC2B, SW1116, SW403, SW48, SW480, SW620, SW837, SW948, T84, and WIDR were purchased from ATCC (Manassas, VA). HDC108, HDC111, HDC114, HDC133, HDC15, HDC54, HDC75, HDC8, HDC87, and HDC9 were a kind gift from Dr. Johannes Gebert (Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany). HT29-cl16E, HT29-cl19A, HCC2998, KM12, RW7213, and RW2982, were a kind gift from Dr. L.H. Augenlicht (Albert Einstein Cancer Center, Bronx, NY). LIM1215 and LIM2405 were obtained from the Ludwig Institute for Cancer Research in Melbourne. ALA, CO115, FET, Isreco1, Isreco2, Isreco3, V9P, and TC71 were a kind gift from Dr. Hamelin, Paris, France. GP5D, HCA7, and VACO5 were a kind gift from Dr. L.A. Aaltonen (Biomedicum Helsinki, Finland). All lines were maintained in MEM (Life Technologies, Carlsbad, CA) supplemented with 10 % fetal bovine serum, 1x antibiotic/antimycotic (100 units/ml streptomycin, 100 units/ml penicillin, and 0.25 µg/ml amphotericin B), 1x MEM Non-Essential Amino Acids Solution, and 10 mM HEPES buffer solution (all from Life Technologies, Carlsbad, CA) at 37^{o}C and 5 % CO_2. All lines were tested to be negative for mycoplasma contamination (PCR Mycoplasma Detection Set, Takara). Cell lines were cultured until they reached 70 %-80 % confluence and the medium was changed 8h before harvesting the cultures for DNA or RNA extraction. The cell lines used were authenticated, and additionally possible cell line cross-contamination was investigated by clustering analysis of genome-wide mRNA expression and promoter methylation microarray data at the time of these experiments.

Table 5: Cell lines used in this study.

Cell line	Doubling time SRB (h)	MSI/MSS[a]	Histological grade on xenografts	Doublingtime xCELLigence (h)	GI50 5-FU (µM)	GI50 Acifluorfen (µM)	GI50 Na-iodoacetate (µM)	BRAF[b]	KRAS[b]	TP53[b]	APC[b]	PIK3CA[b]	SMAD4[b]	TCF7L2[b]	CTNNB1[b]
HT29-cl16E	43.37	MSS	-	-	-	-	-	-	wt	mut	mut	-	-	-	-
HT29-cl19A	43.4667	MSS	-	-	-	-	-	-	wt	mut	mut	-	-	-	-
ALA	51.5833	MSS	-	-	-	-	-	wt	mut	mut	-	wt	-	-	-
Caco2	51.4071	MSS	2	15.38	3.11	605.10	4.51	wt	wt	mut	mut	wt	mut	wt	mut
Co115	27.66	MSI	-	-	-	-	-	wt	wt	wt	wt	-	-	-	-
Colo201	41.3267	MSS	2	-	2.06	-	-	mut	wt	mut	mut	wt	wt	wt	wt
Colo205	26.7567	MSS	3	-	1.54	-	-	mut	wt	mut	mut	wt	wt	wt	wt
Colo320	24.19	MSS	3	31.31	1.47	-	-	wt	wt	wt	mut	wt	wt	wt	wt
DLD1	25.2667	MSI	3	16.40	1.32	1310.00	10.05	wt	mut	mut	mut	mut	wt	wt	wt
FET	35.4933	MSS	-	-	-	-	-	wt	mut	mut	-	-	-	-	-
GP5D	60.0567	MSI	-	-	-	-	-	wt	mut	wt	mut	mut	wt	mut	wt
HCA7	-	MSI	-	-	-	-	-	wt	wt	mut	mut	wt	wt	wt	wt
HCC2998	44.4	MSS	2	9.69	1.18	320.27	1.47	wt	mut	mut	mut	wt	wt	wt	wt
HCT116	25.0183	MSI	3	9.43	0.69	1004.00	3.29	wt	mut	wt	wt	mut	wt	wt	mut
HCT15	29.5667	MSI	3	11.68	3.11	501.20	4.99	wt	mut	mut	mut	mut	wt	wt	wt
HCT8	27.4867	MSI	3	14.15	2.84	-	-	wt	mut	wt	mut	mut	wt	mut	mut
HDC108	49.7167	MSI	-	-	-	-	-	wt	wt	-	-	-	-	-	-
HDC111	43.98	MSS	-	-	-	-	-	wt	mut	-	-	-	-	-	-
HDC114	50.2833	MSS	-	-	-	-	-	wt	mut	-	-	-	-	-	-
HDC133	99.15	MSS	-	-	-	-	-	-	-	-	-	-	-	-	-
HDC15	59.26	MSS	-	-	-	-	-	wt	wt	-	-	-	-	-	-
HDC54	42.8633	MSS	-	-	-	-	-	wt	wt	mut	mut	wt	wt	wt	mut
HDC75	42.5267	MSS	-	-	-	-	-	wt	mut	-	-	-	-	-	-
HDC8	52.4	MSS	-	-	-	-	-	-	-	-	-	-	-	-	-
HDC87	39.0167	MSS	-	-	-	-	-	wt	mut	wt	mut	wt	wt	wt	wt
HDC9	49.1567	MSI	-	-	-	-	-	wt	wt	-	-	-	-	-	-
HT29	36.9333	MSS	3	28.63	4.20	516.53	5.14	mut	wt	mut	mut	mut	mut	wt	wt
HUTU80	26.9033	MSI	-	-	-	-	-	-	-	-	-	-	-	-	-
Isreco1	20.9567	MSS	-	13.20	-	294.10	7.94	wt	mut	mut	mut	wt	-	-	-
Isreco2	52.524	MSS	-	-	-	-	-	wt	mut	mut	mut	wt	-	-	-
Isreco3	69.3767	MSS	-	77.31	-	-	-	wt	mut	mut	mut	wt	-	-	-
KM12	34.1133	MSI	2	23.61	1.30	-	-	mut	wt	mut	wt	wt	wt	wt	wt
LIM1215	29.5217	MSI	2	22.45	2.97	-	-	wt	wt	wt	wt	wt	wt	wt	mut
LIM2405	30.72	MSI	3	16.42	0.90	-	15.31	mut	wt	wt	mut	wt	wt	wt	wt
LoVo	33.1557	MSI	3	17.94	1.97	-	-	wt	mut	wt	mut	wt	wt	mut	wt
LS1034	40.79	MSS	-	-	-	-	-	wt	wt	mut	mut	wt	-	-	-
LS174T	38.0067	MSI	3	20.99	5.52	341.70	8.19	wt	mut	wt	wt	mut	wt	mut	mut
LS513	33.9267	MSS	-	-	-	-	-	wt	mut	wt	wt	wt	wt	wt	wt
RKO	36.34	MSI	3	-	0.75	614.97	3.45	mut	wt	wt	wt	mut			
RW2982	58.41	MSS	1	41.37	1.62	648.00	-	wt	mut	wt	mut	wt	mut	wt	wt
RW7213	-	MSS	-	-	18.14	-	-	wt	mut	wt	mut	wt	wt	mut	wt
SKCO1	46.56	MSS	3	41.73	2.74	-	-	wt	wt	wt	mut	wt	wt	wt	wt
SNUC2B	49.5067	MSI	-	-	-	-	-	wt	mut	mut	wt	wt	wt	mut	wt
SW1116	68.7029	MSS	1	-	19.50	-	17.10	wt	mut	mut	mut	wt	wt	wt	wt
SW403	38.1886	MSS	1	-	0.68	-	11.86	wt	mut	mut	mut	wt	wt	wt	wt
SW48	42.3267	MSI	3	27.88	4.60	-	-	wt	wt	wt	mut	mut	wt	mut	mut
SW480	30.7425	MSS	3	-	10.69	-	-	wt	mut	mut	mut	wt	wt	wt	wt
SW620	37.0567	MSS	3	27.17	23.09	-	-	wt	mut	mut	mut	wt	wt	wt	wt
SW837	40.6033	MSS	3	54.07	0.80	-	-	wt	mut	mut	mut	wt	wt	wt	wt
SW948	42.4517	MSS	1	11.40	1.35	568.33	2.57	wt	mut	mut	mut	mut	mut	wt	wt
T84	33.902	MSS	2	-	2.60	859.00	-	wt	mut	wt	mut	mut	mut	wt	wt
TC71	28.52	MSI	3	-	1.27	-	-	wt	mut	mut	-	wt	-	-	-
V9P	-	MSS	-	-	-	-	-	-	-	-	-	-	-	-	-
VACO5	28.85	MSI	-	-	-	270.97	1.07	mut	wt	-	-	mut	-	-	-
WiDr	45.6767	MSS	-	27.10	-	-	-	mut	wt	mut	-	wt	-	-	-

NOTE: MSS= Microsatellite stable, MSI= Microsatellite instable, wt= wild-type, mut= mutant.
Information taken from: a) Mazzolini, R. et al. [43]; [b] **Mouradov, D. et al.** [163].

3.1.2 Primary colorectal tumor samples

The data from primary tumor samples used in this study were obtained from The Cancer Genome Atlas (TCGA). mRNA expression levels (Illumina RNAseq and Agilent microarray G4502A) and haematoxylin and eosin-stained high resolution images of formalin-fixed, paraffin-embedded sections of primary tumors were downloaded from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). For light microscopy quantification of mitotic cells in these tumors, three randomly selected fields were selected, and the total number of cells (>500) and mitotic cells was scored blinded from the sample identity.

3.1.3 Antibodies

Table 6: Antibodies used in this study.

Antibody	Source	Reference	Host	Application (dilution)
cleaved PARP (Asp214)	Cell Signaling Technology	9541	rabbit	WB (1:2000)
β-tubulin	Sigma-Aldrich	T4026	mouse	WB (1:1000)
actin	Santa Cruz Biotechnology	sc-10731	rabbit	WB (1:1000)

NOTE: WB=Western blot

3.1.4 Primers

Table 7: Primers used in this study.

Primers name	Application	Location	Sequence 5'-3'	Product size (bp)
TYMS-F	qPCR	E5-6	ACACACTTTGGGAGATGCAC	72
TYMS-R	qPCR	E5-6	GGTTCTCGCTGAAGCTGAAT	72
PPOX-F	qPCR	E5-7	GGCGCTGGAAGGTATCTCTA	72
PPOX-R	qPCR	E5-7	CTGAAGCTGGAATGGCACTA	72
GAPDH-F	qPCR	E8	ACCCACTCCTCCACCTTTGAC	76
GAPDH-R	qPCR	E8	CATACCAGGAAATGAGCTTGACAA	76
SMAD4-F	qPCR	E7-8	AAAACGGCCATCTTCAGCAC	59
SMAD4-R	qPCR	E7-8	AGGCCAGTAATGTCCGGGA	59
CALCOCO2-F	qPCR	E10-11	GAAAGAGAGATTGGAAGGAGAAA	103
CALCOCO2-R	qPCR	E10-11	AGGTACTTGATACGGCAAAGAAT	103
CBX5-F	qPCR	E3-4	ACCCAGGGAGAAGTCAGAAA	101
CBX5-R	qPCR	E3-4	CGATATCATTGCTCTGCTCTCT	101
KLHL3-F	qPCR	E10-12	AGTACTGGCCTAGCATCGGT	149
KLHL3-R	qPCR	E10-12	CGGGAAGCTCCATCATAAC	149
PPP1R14D-F	qPCR	E3-5	AGACTCAGCTGGAGGCCAT	147
PPP1R14D-R	qPCR	E3-5	CAGTGCTGAGGCTGCTAAAG	147
ITGA9-F	qPCR	E27-28	GTTGGTGGGAATCCTCATCT	61
ITGA9-R	qPCR	E27-28	AAAGAAGCCCATCTTCCAGA	61
ZNF238-F	qPCR	E1-2	AGCAGGACTCAGAGGAAAGG	150
ZNF238-R	qPCR	E1-2	CCAGAACAGTGCAGTCACAA	150
18S-F	qPCR	-	AGTCCCTGCCCTTTGTACACA	-
18S- R	qPCR	-	GATCCGAGGGCCTCACTAAAC	-
18S-Probe	qPCR	-	[6FAM]-CGCCCGTCGCTACTACCGATTGG-[TAM]	-
ITGA9-BS-F	BSS	cg13882267	GGGATTTGAGGATTTGTATTTTTT	255
ITGA9-BS-R	BSS	cg13882267	CCTCTACTCCTTCACCCAATTATAA	255
PPP1R14D-BS-F	BSS	cg23382741	AGGTTAGGTTGATAGTAGTTTATATT	225
PPP1R14D-BS-R	BSS	cg23382741	CCTCTATATCCCACCTTCCTAAAAC	225
KLHL3-BS-F	BSS	cg13847070	AAGTTGGAAAGGTGGTAGTGTATTT	282
KLHL3-BS-R	BSS	cg13847070	CCAACAAACCAATAAAAAATCTAATC	282
ZNF238-BST-F	BSS	cg02497700	TTTTTTATTTTATTGGGTAAATGGG	285
ZNF238-BST-R	BSS	cg02497700	CCCAACCCTAATAATAACCACTTC	285
ZNF238-BST-F	BSS	cg23829949	TATATAGATAGGGAGTTAGTGTGT	210
ZNF238-BST-R	BSS	cg23829949	TATACTCAATCTAATCTCTTACTAC	210
BamHI-Kozak-ZNF238-F	Cloning	-	TCGCAGGGATCCgccgccATGGAGTTTCCAGACCATAGTAGAC	-
ZNF238-BamHI-R	Cloning	-	TCACGCGGATCCTTATTTCCAAAGTTCTTGAGAGCTATC	-

NOTE: qPCR= quantitative polymerase chain reaction, BSS=Bisulfite sequencing ,F= forward, R=reverse, E=Exon

3.2 Methods used in this study

3.2.1 Microarray data

3.2.1.1 mRNA expression microarray analysis

The levels of expression of more than 47,000 transcripts and variants, including more than 38,500 well characterized genes and UniGenes, were investigated using GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA) in a subset of 32 colorectal cancer cell lines (CACO2, CO115, COLO201, COLO205, COLO320, DLD1, HCC2998, HCT116, HCT15, HCT8, HT29, IS1, IS2, KM12, LIM1215, LIM2405, LOVO, LS174T, RKO, RW2982, RW7213, SKCO1, SW1116, SW403, SW48, SW480, SW620, SW837, SW948, T84, TC71 and Vaco5).

All cell lines were cultured as described before (Section 3.1.1). Total RNA was extracted with TRIzol Reagent (Life Technologies, Carlsbad, CA) and then labeled and hybridized to Affymetrix HG-U133 Plus 2.0 chips as previously described [164]. The mRNA levels were calculated after RMA (Robust Multichip Average) normalization as described [165]. Clustering analysis was done with dChip software [166]. The mRNA expression microarray data has been deposited at ArrayExpress (E-MTAB-2971).

3.2.1.2 DNA methylation microarray analysis

The quantitative levels of methylation at the single nucleotide resolution level were assessed for a total of 27,578 highly informative CpG sites using HumanMethylation27 Beadchips (Illumina, San Diego, CA). These chips target CpG sites located within the proximal promoter regions of transcription start sites of 14,475 consensus coding sequencing (CCDS) in the NCBI Database (Genome Build 36). On average, two CpG sites were present per CCDS gene and from 3-20 CpG sites for >200 cancer-related and imprinted genes. The levels of DNA methylation were studied in 45 different lines. In addition, one of the lines (SW48) was hybridized twice. Also, an *in vitro* methylated control (CpG Methylated Jurkat Genomic DNA, New England BioLabs) and an unmethylated control (Jurkat DNA amplified *in vitro* with illustra GenomiPhi HY DNA Amplification Kit from GE Healthcare) were included in the experimental design. The DNA from all samples was extracted using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich) and then bisulfite treated and hybridized following manufacturer's recommendations at the Spanish National genotyping Center (CeGen-CRG Genotyping Unit). The levels of methylation were calculated using GenomeStudio software (Illumina). In some analysis, the levels of methylation of primary tumors and normal colonic mucosa samples were determined using HumanMethylation27 beadchips (Illumina) as previously reported in the Gene Expression Omnibus repository (GSE17648) or The Cancer Genome Atlas [104]. Promoter

methylation microarrays data (HumanMethylation27, Illumina) for HCT116 cells where DNA methyltransferase activity has been inactivated either pharmacologically (5-Aza-2'-deoxycytidine treatment) or genetically (*DNMT1* and *DNMT3b* double knockout) can be found in the ArrayExpress (E-MTAB-210) or Gene Expression Omnibus (GSE26990) repositories, respectively [167].

3.2.2 Identification of genes

3.2.2.1 Associations between mRNA expression and the doubling time of cell lines

Given that, for many genes, the relation between expression and growth rate was monotonous but not linear, a Spearman's rank correlation was used to identify genes whose expression was associated with growth rates across a panel of 31 colorectal cancer cell lines. Benjamini-Hochberg procedure was used to correct for multiple hypothesis testing ($p<0.1$).

3.2.2.2 Associations between mRNA expression and promoter methylation levels

Expression data was available for a total of 11,858 (81.92 %) of the 14475 promoters interrogated in the HumanMethylation27 arrays, when using the gene symbol as a common identifier. If there was more than one probe for a given gene/promoter, the average value of expression/methylation was used. The Pearson correlation coefficient was used to identify significant negative correlations ($r<-0.355$; $p<0.05$) between gene expression and promoter methylation in 30 colorectal cancer cell lines. To eliminate from this list genes whose correlation is heavily dependent on a small number of cell lines, the graphs of the expression (Y-axis) and methylation (X-axis) of all 30 cell lines was divided in four quadrants using the median expression/methylation values. Only genes with >5 data points in the upper left and lower right quadrants were further considered. In addition, genes showing no significant Pearson's correlation between methylation and expression, but that showed high expression levels in at least a subset of the cell lines with low methylation levels and low/absent expression in cell lines with high methylation levels (i.e., 'L-shaped' in the scattered plots) were also included.

3.2.3 Functional group enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used to investigate whether there were gene sets with significant enrichment in the number of genes with expression/proliferation correlations or expression/methylation correlations [168]. A Fisher's exact test was used to identify significantly enriched categories of genes associated with cell growth or methylation. The Benjamini-Hochberg procedure was used to correct for multiple hypothesis testing ($p<0.05$).

3.2.4 *In vitro* experiments

3.2.4.1 Determination of the doubling time

To determine the doubling time of each cell line, cells were seeded in seven 96-well plates. Seeding densities varied from $1x10^3$ to $1.5x10^4$ cells/well to ensure control cell densities did not exceed 80 % confluence at the completion of the experiment. The plates were fixed with trichloroacetic acid (final concentration 10 % w/v) at 24h intervals for seven days. Plates were washed with tap water, air dried and stained with 0.4 % (w/v) sulforhodamine B (SRB) for 30 min. Excess SRB was washed out with 1 % acetic acid and the plates were air dried. Cell-bound SRB was solubilized with 10mM Tris buffer pH=10 and absorbance was measured at 590nm using a microplate reader (Sunrise, Tecan, Männedorf, Switzerland). The doubling times were calculated using Prism V.5.01 (GraphPad, San Diego, CA). All experiments were carried out at least 3 times with eight replicates each time.

As an independent approach to assess cell growth, the Roche xCELLigence System was used for real-time monitoring of cell proliferation [169]. Cell lines were seeded in quadruplicate at a density of 5000 cells/well in an E-Plate 96 (Roche Diagnostics, Risch-Rotkreuz, Switzerland). The Real-Time Cell Analyzer MP instrument (Roche Diagnostics, Risch-Rotkreuz, Switzerland) together with the E-Plate 96, was placed in a cell culture incubator maintained at 37°C with 5 % CO_2, and continuous electrical impedance measurements were taken hourly for eight days. Doubling times were calculated using Cell Index data from the exponential growth phase for each cell line, with RTCA software version 1.2.1.

3.2.4.2 Proliferation assay

Two experimental approaches were used to determine the proliferation rate of cell lines: indirectly using sulforhodamine B (SRB) staining, or directly counting the number of cells. Briefly, to determine the cell growth by SRB, cells were seeded in seven 96-well plates, $1x10^3$ cells per well. The plates were fixed and stained with SRB. Cell-bound SRB was solubilized and absorbance was measured at 590 nm as described above. All experiments were carried out at least 3 times with eight replicates each time.

To determine the growth by cell counting, cells were seeded into 6-well plates at a density of $3x10^4$ cells per well and allowed to adhere overnight. Cells were trypsinized and stained with trypan blue; viable cells were counted using a hemocytometer at times 0, 24, 48, 72, and 96 h. Growth curves presented are the average of three independent experiments carried out in triplicates.

3.2.4.3 Growth inhibition assay

The dose resulting in 50 % growth inhibition (GI_{50}) in the presence of 5-FU, acifluorfen, sodium iodoacetate, oxadiazon (all from Sigma-Aldrich, St. Louis, MO) or CGP 3466B maleate (Tocris, UK), compared to the corresponding control, was determined as described [170,171]. Briefly, $5x10^3$ cells per well were seeded in 96-well plates. Twenty-four hours after seeding, cells were treated with 5-FU (0, 0.01, 0.1, 0.5, 1, 2.5, 5, 10, 25, 50, 100, and 500 µM in PBS*)*, acifluorfen (0, 5, 25, 100, 200, 300, 400, 500, 750, 1000, 2000, and 3000 µM in DMSO), Na iodoacetate (0, 0.01, 0.1, 1, 2.5, 5, 7.5, 10, 20, 30, 60, and 120 µM in PBS), oxadiazon (0, 25, 50, 100, 200, 300, 400, 500, 600, 750, 1000, and 1250 µM in DMSO) or CGP 3466B maleate (0, 5, 10, 25, 50, 75, 100, 125, 250, 500 and 750 µM in DMSO) for 72 h. Cells were fixed with trichloroacetic acid and stained with sulforhodamine B, as described above. One plate of each cell line was fixed to assess cell number at the time when drug treatment started (=T_0). After absorbance was measured at 590nm using a microplate reader (Sunrise, Tecan, Männedorf, Switzerland), the blank (medium incubated for 24 h or 72 h) was subtracted. In order to calculate the GI_{50}, three measurements are necessary: Ti= absorbance of cells after treatment (for each drug concentration), T_0= absorbance at the beginning of the treatment, and C= absorbance of cells without treatment (incubated with complete growth medium for 24 h). Using these measurements, cellular responses was calculated for growth inhibition. The formula used is: If $Ti \geq T_0$, the calculation is $100 \times [(Ti - T_0)/(C - T_0)]$. If $Ti \leq T_0$, cell killing has occurred and can be calculated from $100 \times [(Ti - T_0)/T_0]$. Thus, for each drug-cell line combination, a dose-response curve was generated and growth inhibition of 50 % (GI_{50}) was calculated from $100 \times [(Ti - T_0)/(C - T_0)] = 50$, which is the drug concentration causing a 50 % reduction in the net protein increase in control cells during the drug incubation [172,173]. These experiments were carried out at least three times in quadruplicates.

3.2.4.4 Apoptosis and cell cycle analysis

Two hundred thousand cells were seeded in triplicate in 6-well plates. Control wells reached a confluence of approximately 80 % at the completion of the experiment. 24 h after seeding, cells were treated with 0, 10, 20,or 30 µM sodium iodoacetate or 0, 400, 800, 1200 µM acifluorfen (both Sigma-Aldrich, St. Louis, MO) for 72 h. Both, floating and adherent cells, were harvested, washed with cold PBS, and resuspended in 50 µg/ml propidium iodide, 0.1 % sodium citrate, and 0.1 % Triton X-100. Cells were stained for 2 h at 4 °C, and 10,000 cells were analyzed for DNA content using a FacsCalibur Flow Cytometer (Becton Dickinson, Franklin Lakes, NJ). The percentage of cells with a subdiploid DNA content was quantified using WinList 2.0 (Verity Software House, Topsahm, NE). The cell cycle profile was established using the ModFit 2.0 (Verity Software House, Topsahm, NE).

3.2.4.5 Clonogenic assay

Five hundred HCT116 or DLD1 cells were seeded in triplicate in 6-well plates. 24h after seeding, cells were treated with 0 or 15 μM sodium iodoacetate or 0 or 1200 μM acifluorfen for 9 h. The medium containing the drug was washed off and replaced with fresh medium without drug. Colony formation was monitored over the following 2–3 weeks. Cultures were stained with 1 % crystal violet for 30 min, washed with distilled water, air dried, and the number of colonies was determined blinded from the sample identity. Each cell line was assayed three times in triplicates.

3.2.4.6 RNAi knockdown of PPOX and GAPDH

HCT116 cells ($2x10^5$) were seeded in 6-well plates and 24h later they were transfected with control On-TARGET plus non-targeting siRNA, or siRNA pools of 4 siRNAs against GAPDH or PPOX (D-001810-10-05, D-001830-10-05 or L-008383-00-0005, respectively; Dharmacon, Lafayette, CO) using Lipofectamine 2000 (Life Technologies, Carlsbad, CA). Expression levels and cell numbers were assessed 72h after transfection as described below.

3.2.4.7 Protein extraction and Western blot

Seven hundred and fifty thousand cells were seeded in 6-well plates. 24h after seeding, cells were treated with 0, 10 or 20 μM sodium iodoacetate or 0, 400 or 800 μM acifluorfen for 24 h. Cells were harvested, washed with cold PBS, and cell pellets resuspended in 0.1 ml lysis buffer (25 mM Hepes, pH=7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 % NP-40, 1 mM DTT, 10 % glycerol and protease inhibitors (cOmplete ULTRA Tablets, Mini, EDTA-free, Roche, Mannheim, Germany)). Aliquots of the cleared supernatant containing total protein (25 μg) were loaded on a 15 % acrylamide gel. After gel electrophoresis, proteins were transferred to a PVDF membrane and probed with anti-cleaved PARP (Asp214), anti-β-tubulin, or anti-actin antibody (Table 6).

3.2.4.8 RNA extraction and quantitative RT-PCR

Cell cultures were harvested at 70-80 % confluence and total RNA was extracted using TRIzol Reagent (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Total RNA (2 μg) was reverse transcribed using the High Capacity cDNA BS-R Transcription kit (Life Technologies, Carlsbad, CA), and relative mRNA levels of PPOX, GAPDH, TYMS, CALCOCO2, CBX5, SMAD4, ITGA9, KLHL3, PP1R14D, and ZNF238 were assessed by Real-Time PCR using SYBR Green Master Mix (Life Technologies, Carlsbad, CA). 18S rRNA (Taqman Master Mix, Life Technologies, Carlsbad, CA) was used as a standardization control for the $2^{-\Delta\Delta Ct}$ method as described before [164]. The primers used are listed in Table 7.

3.2.4.9 DNA extraction and bisulfite sequencing

Genomic DNA was isolated from each line using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO) and bisulfite treated with EZ DNA Methylation-Gold Kit (Zymo Research) both following the manufacturer's instructions and then PCR-amplified with the following conditions: 95°C for 5 min, then 40 cycles of 94 °C for 30 seconds, 59.1 °C (ITGA9 and PPP1R14D), 60.4 °C (ZNF238) or 64.1 °C (KLHL3) for 45 seconds and 72°C for 30 seconds, and then the last elongation at 72 °C for 10 min. Primers for DNA amplification and sequencing were designed using MethPrimer software [174] and are listed in Table 7 .The amplified regions were sequenced using the Macrogen sequencing facility (Macrogen Inc, Korea).

3.2.5 ZNF238 overexpression in colon cancer cells

3.2.5.1 ZNF238 cloning

The plasmid pENTR233-ZNF238 (Clone ID HsCD00376023) was used to obtain the ZNF238 sequence. In order to amplify the insert using PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA), a forward primer containing BamHI and Kozak sequence, and a reverse primer containing also a BamHI sequence were designed (Table 7). The PCR product and the pIRES2-EGFP vector (Chlontech, Mountain View, CA) were digested with BamHI and ligated. StBl3 *E.coli* barcterias were transformed by electroporation and amplified. Colonies were picked and the orientation of the insert was verified by digestion and sequencing of the plasmid. The pZNF238-IRES2-EGFP is referred to as ZNF238, and the pIRES2-EGFP as EV.

3.2.5.2 ZNF238 overexpression in colon cancer cells

HCT116 cells were stably transfected with pZNF238-IRES2-EGFP and are referred to as HCT116-ZNF238, and the pIRES2-EGFP as HCT116-EV using the Lipofectamine 2000 transfection agent (Thermo Fisher Scientific, Waltham, MA) according to manufacturer's instruction. After neomycin selection, the population of EGPF positive cells was enriched by FACS Aria (BD, Franklin Lakes, NJ) to obtain cultures with >90 % EGFP positive cells. ZNF238 overexpression was confirmed by quantitative real-time RT-PCR.

3.2.6 *In vivo* animal experiments

3.2.6.1 Drug effects *in vivo* using a xenograft model

Six to seven weeks old female and male NOD/SCID mice were purchased from Charles River (Wilmington, MA). The mice were maintained under sterile conditions and the experiments were carried out under observance of the protocol approved by the ethical committee for animal experimentation from the University Hospital Vall d'Hebron, Barcelona. The tumors were established by subcutaneous injection of $2x10^6$ DLD1, Isreco1 or HCC2998 cells, $2.5x10^6$ HCT116 cells, $1x10^6$ HT29 cells, $3x10^6$ RKO cells, and $5x10^6$ T84 cells, all resuspended in 100 µl sterile PBS. When the tumors reached a volume of about $80mm^3$, the animals were randomized to groups treated with vehicle (PBS), 5-fluorouracil, acifluorfen or sodium iodoacetate (50 mg/kg, 168 mg/kg and 18.4 mg/kg, respectively) three times per week intraperitoneally. The long (L) and short (S) axis of the tumor were measured with a caliper five times a week. The tumor volume was calculated using the formula: $V=LxS^2x0.52$.

3.2.6.2 Determination of the grade of differentiation of cell lines in a xenograft model

Six to seven weeks old male NOD/SCID mice were purchased from Charles River (Wilmington, MA), and experiments carried out under observance of a protocol approved by the Institute's oversight committee for animal experimentation. Tumors were established by subcutaneous injection of $5x10^6$ cells in 200 µl of a 1:1 PBS:matrigel solution into the right flank. When the tumors were >1000 mm^3, they were formalin-fixed, paraffin embedded and Hematoxylin and Eosin stained sections were used to score tumor grade by an experienced pathologist blinded from the sample identity.

3.2.6.3 The effect of overexpressing ZNF238 in cells *in vivo* using a xenografts model

Six to seven weeks old female and male NOD/SCID mice were purchased from Charles River (Wilmington, MA). The mice were maintained under sterile conditions and the experiments were carried out under observance of the protocol approved by the ethical committee for animal experimentation from the University Hospital Vall d'Hebron, Barcelona. The tumors were established by subcutaneous injection HCT116 cells either overexpressing ZNF238 or the EV control, respectively. Thus, $2.5x10^6$ cells were resuspended in 100 µl of sterile PBS and were injected s.c. The long (L) and short (S) axis of the tumor were measured with a caliper three times per week. The tumor volume was calculated using the formula: $V=LxS^2x0.52$. After euthanization of the animals the tumors were resected and the tumor weight was determined.

4 Results

4.1 Highly expressed genes in rapidly proliferating tumor cells as new targets for colorectal cancer treatment

Colorectal cancer is a disease highly regulated by genetic and epigenetic alterations and activation of oncogenes and downregulation of tumor suppressor genes are key hallmarks of cancer proliferation. Patients with advanced colorectal cancer show an objective survival rate of about only 30 % to chemotherapeutic treatments. Therefore, it is of great importance, to better understand the molecular events involved in colorectal cancer, including genes important for tumor progression, tumor suppressor genes and new therapeutic targets. These achievements can ultimately improve the colorectal cancer treatments. Here we hypothesize that genes highly expressed in fast growing tumors are important for tumor progression and that those genes can be used as possible novel therapeutic targets.

4.1.1 Proliferation of colorectal cancer cell lines

Significant variability has been observed in the growth rates of colorectal tumors [175–178]. Here, we thoroughly characterized the growth rates of a large panel of human colorectal cancer cell lines derived from colorectal tumors. The doubling time of these 52 cell lines was initially determined using an indirect sulforhodamine B (SRB) assay to quantify the total protein content in cell line cultures at 24h intervals over one week. Cell line growth demonstrated the expected lag phase before reaching an exponential growth phase followed by a growth plateau (Figure 14B, Appendix 1)[179]. Significant variability was observed in the doubling time during the exponential growth phase of this panel of cell lines (Figure 14A and Table 5). For a subset of 22 lines, we validated these results using an independent technique based on electrical impedance as the readout for real-time noninvasive cell growth monitoring (xCELLigence, Roche Diagnostics, Mannheim, Germany), and we found good correlation between the doubling time calculated through both approaches (Pearson's r=0,66, p=0.0007; Figure 14C).

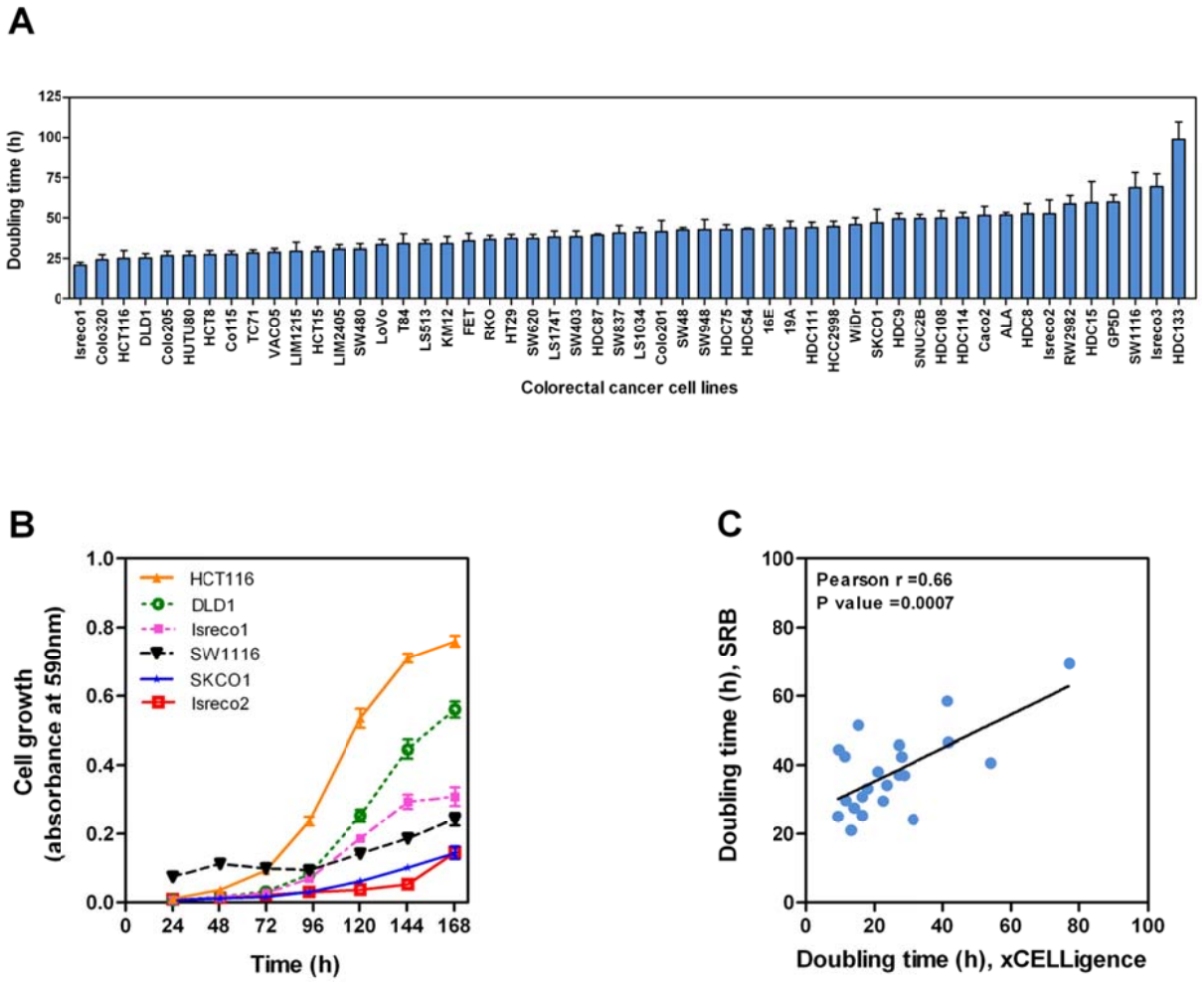

Figure 14: Growth of colorectal cancer cell lines.
A) Histogram showing the doubling time of all 52 cell lines used in this study determined with the SRB assay (average of three experiments ± SEM). B) For a panel of 52 colorectal cancer cell lines, cells were seeded in seven 96 well plates, harvested daily for 7 consecutive days and cell number assessed indirectly using sulforhodamine B (SRB) staining and colorimetric quantification. Representative cell lines with rapid and slow growth are shown. C) The growth of a subset of 22 of these cell lines was assessed using electrical impedance as the readout for the number of cells (xCELLigence). The doubling time was calculated with both techniques and the correlation of the results obtained with both methods is shown.

Inactivation of mismatch repair genes results in the accumulation of mutations throughout the genome which manifests as microsatellite instability (MSI) in approximately 15 % of colorectal tumors [180]. However, the majority of colorectal tumors shows no microsatellite instability and instead displays chromosomal instability with large chromosomal abnormalities, and is referred to as microsatellite stable (MSS) or chromosomal instable (CIN) tumors. We found here that cell lines with microsatellite instability grew significantly faster than microsatellite stable lines (Figure 15A). A subset of 27 of these cell lines were grown as subcutaneous xenografts in immunodeficient mice, and the histological grade of the tumors formed was determined. Microsatellite instable tumors have been shown to be associated with high tumor grade [180]. In good agreement, higher tumor grade was found to be associated with an MSI phenotype in these cell lines (χ^2, $p<0.05$), and faster growth was observed in cell lines that generated high grade tumors when grown as xenografts, compared to lines generating low/moderate grade tumors (Figure 15B). No associations were found between cell line

doubling time and the mutational status of the genes most frequently mutated in colorectal tumors, such as *APC*, *TP53*, *KRAS*, *BRAF*, *PIK3CA*, *SMAD4*, *TCF7L2* and *CTNNB1* (Figure 15C-J).

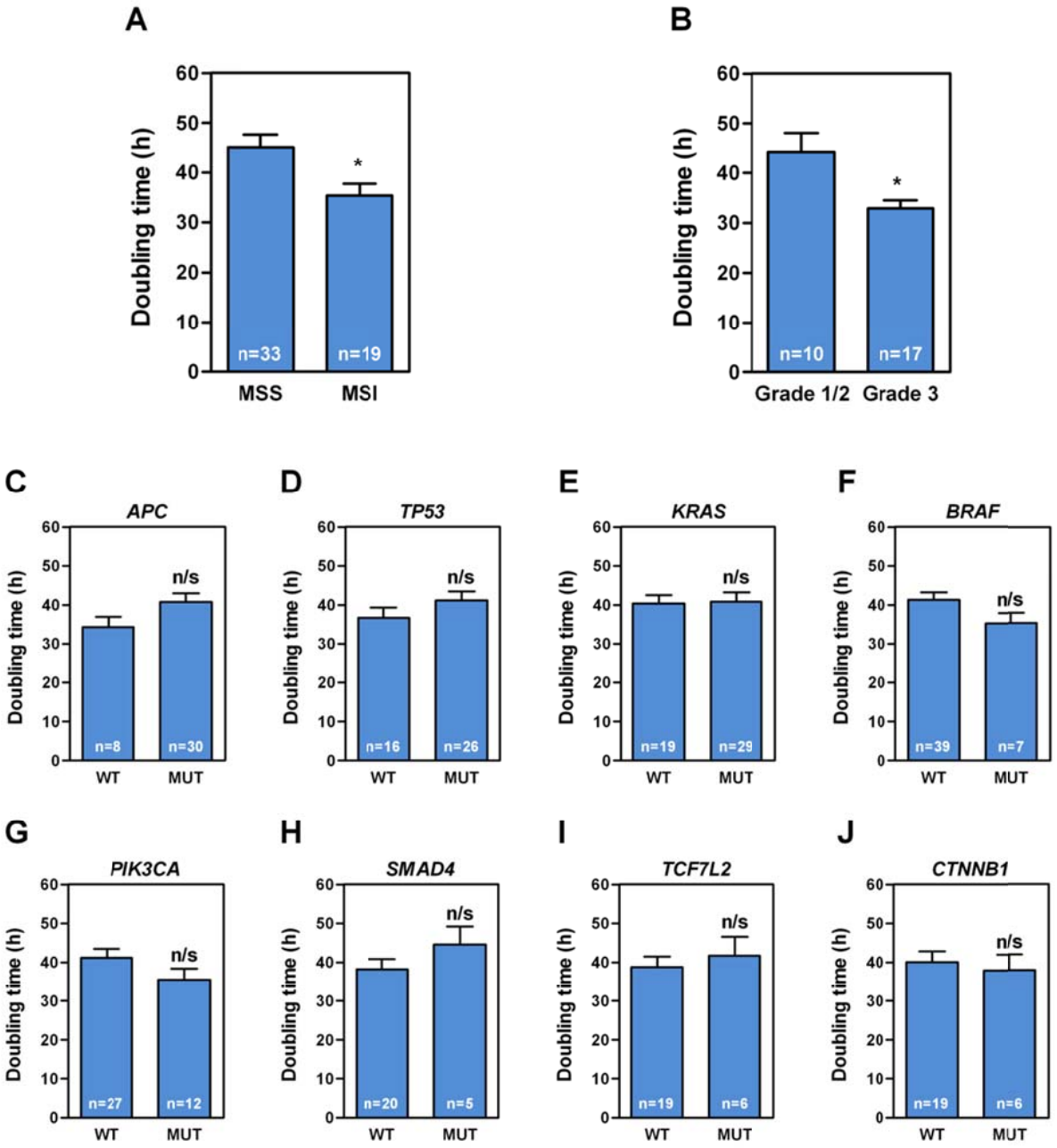

Figure 15: The doubling time of colorectal cancer cell lines is associated with microsatellite instability and tumor grade but not with the mutational status of the most frequently mutated genes in colorectal tumors.
A) Colorectal cancer cell lines with an MSI phenotype showed significantly faster growth (lower doubling time) compared to lines without MSI phenotype. B) Cell lines growing as poorly differentiated (grade 3) tumors in subcutaneous xenografts in immunodeficient mice had faster growth than cell lines displaying moderately/highly differentiated histology (grade 1 and 2). The average doubling time of cell lines that are either wild type or mutant for C) *APC*, D) *TP53*, E) *KRAS*, F) *BRAF*, G) *PIK3CA*, H) *SMAD4*, I) *TCF7L2*, and J) *CTNNB1* was calculated to assess the possible effects on tumor growth of the most frequent mutations observed in colorectal tumors. n: number of cell lines. The mean ± SEM is shown. Asterisks indicate Student's t-test $p<0.05$. n/s: Student's T-test $p>0.05$.

4.1.2 Expression profiling of colorectal cancer cell lines with different growth rates

High proliferation rates in colorectal tumors have been previously associated with poor patient prognosis [175–178], and although the molecular mechanisms regulating the progression of tumor cells through the different phases of the cell cycle are well characterized, the key rate-limiting steps are not fully understood. Here we used microarray analysis to perform global gene expression profiling on a subset of these colorectal cancer cell lines (n=31) to investigate the molecular mechanisms underlying the differences in growth rates.

For this analysis, we considered genes with expression levels significantly above background in 23 of the 31 cell lines investigated (>75 %). Of the 11,512 genes investigated, the expression of 1,290 (11.2 %) was significantly correlated with the doubling time of these cell lines (966 negatively and 324 positively correlated) Spearman's correlation, BH FDR<0.1 for at least one probe (Table 8, and Figure 16; for complete list of genes see Appendix 1 as referred to Supplementary Table S2).

Table 8: Top 20 probes with highest correlation coefficient (positive and negative) between gene expression and doubling time in a panel of 31 colorectal cancer cell lines; BH FDR<0.1.

ProbeSets ID	Gene Symbol	Gene Name	Spearman ρ	Spearman *P*	BH (FDR) adjusted *P*
227257_s_at	CACUL1	CDK2-associated, cullin domain 1	-0.74	1.85E-06	1.25E-02
230069_at	SFXN1	Sideroflexin 1	-0.74	2.21E-06	1.25E-02
222983_s_at	PAIP2	Poly(A) binding protein interacting protein 2	-0.74	2.34E-06	1.25E-02
201968_s_at	PGM1	Phosphoglucomutase 1	-0.73	2.89E-06	1.25E-02
223443_s_at	AMZ2P1	Archaelysin family metallopeptidase 2 pseudogene 1	-0.73	3.18E-06	1.25E-02
201051_at	ANP32A	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	-0.72	4.59E-06	1.25E-02
207124_s_at	GNB5	Guanine nucleotide binding protein (G protein), beta 5	-0.72	4.67E-06	1.25E-02
219978_s_at	NUSAP1	Nucleolar and spindle associated protein 1	-0.72	5.12E-06	1.25E-02
1554740_a_at	IPP	Intracisternal A particle-promoted polypeptide	-0.72	6.11E-06	1.25E-02
221677_s_at	DONSON	Downstream neighbor of SON	-0.71	7.67E-06	1.25E-02
220465_at	CEBPA-AS1	CEBPA antisense RNA 1 (head to head)	0.65	8.56E-05	2.69E-02
229690_at	FAM109A	Family with sequence similarity 109, member A	0.65	6.78E-05	2.61E-02
200070_at	CNPPD1	Cyclin Pas1/PHO80 domain containing 1	0.65	6.59E-05	2.61E-02
203201_at	PMM2	Phosphomannomutase 2	0.65	6.50E-05	2.61E-02
201368_at	ZFP36L2	ZFP36 ring finger protein-like 2	0.66	6.15E-05	2.61E-02
1569679_at	CDH22	Cadherin 22, type 2	0.66	5.90E-05	2.61E-02
209509_s_at	DPAGT1	Dolichyl-phosphate (UDP-N-acetylglucosamine) N-acetylglucosaminephosphotransferase 1 (GlcNAc-1-P transferase)	0.66	4.90E-05	2.51E-02
239588_s_at	---	---	0.68	3.06E-05	2.12E-02
208987_s_at	KDM2A	Lysine (K)-specific demethylase 2A	0.69	2.02E-05	1.70E-02
214316_x_at	CTC-425F1.4	---	0.72	4.76E-06	1.25E-02
1554696_s_at	TYMS	Thymidylate synthetase	-0.68	2.76E-05	1.97E-02
238117_at	PPOX	Protoporphyrinogen oxidase	-0.6	3.77E-04	3.74E-02
AFFX-HUMGAPDH/M33197_5_at	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	-0.55	1.30E-03	5.74E-02

NOTE: TYMS, GAPDH, and PPOX are also shown. =Spearman correlation coefficient. BH (FDR)= Benjamini & Hochberg, false discovery rate.

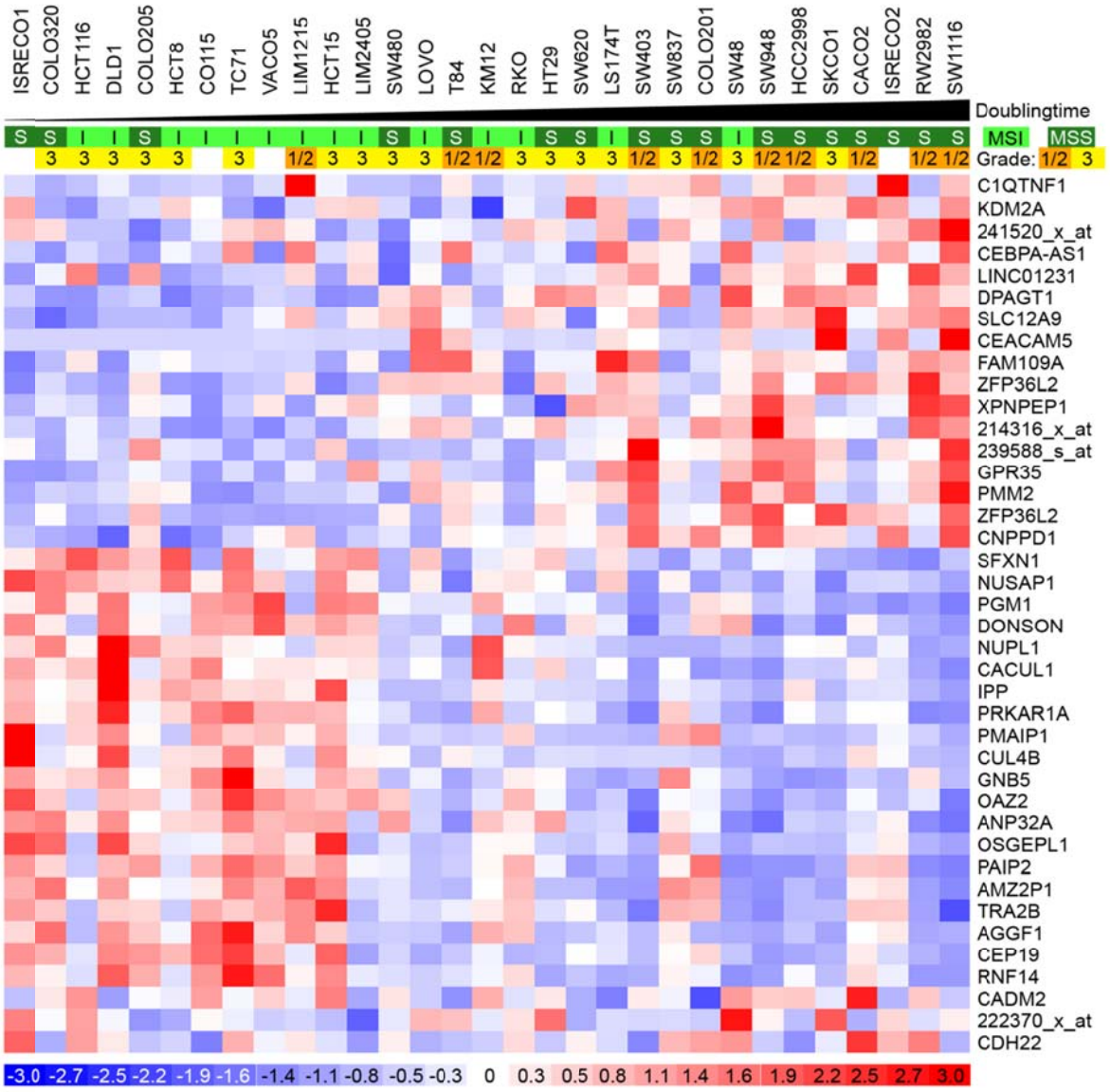

Figure 16: Associations between gene expression and growth of colorectal cancer cells.
Clustering analysis of the 40 genes (rows) whose expression is best correlated with the doubling time of a panel of 31 colorectal cancer cell lines. Cell lines (columns) are ordered by increasing doubling times. Genes with relative expression levels above or below the mean are shown in red and blue, respectively (color scale is shown at the bottom).

In order to validate the mRNA expression microarray data, six of these genes expression levels were independently assessed using quantitative real-time RT-PCR. A significant correlation was observed with mRNA levels quantified by microarray analysis (Figure 17).

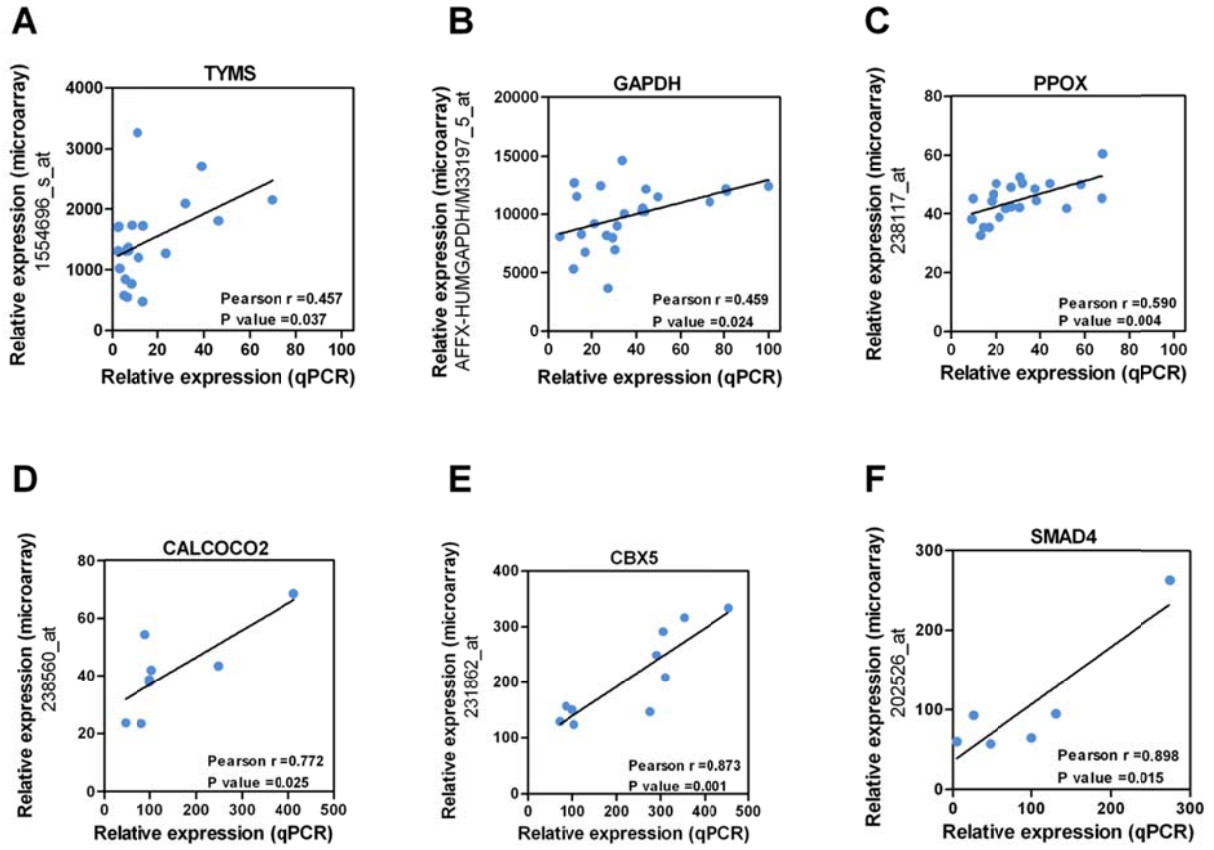

Figure 17: Independent validation of mRNA expression levels.
To independently validate the quantification of mRNA expression of the microarray experiments, the expression of A) TYMS, B) GAPDH, C) PPOX, D) CALCOCO2, E) CBX5, and F) SMAD4 was assessed by real-time RT-PCR. The Pearson's correlation coefficient (r) and p value are shown.

Among the genes whose expression was found to be significantly correlated with the doubling time of the cell lines were multiple genes known to be key cell cycle regulators, including multiple cyclins (A2, B1, B2, E2, F, I, and T2), cyclin-dependent kinases (CDKs; 1, 2, 9 and 13), the CDK inhibitor 2D (p19) and the cell division cycle (CDC) proteins 5L, 6, 14B, 25C, 27, and 37 (

Figure 18 and for complete list of genes see Supplementary Table S2 as referred in Appendix 1).

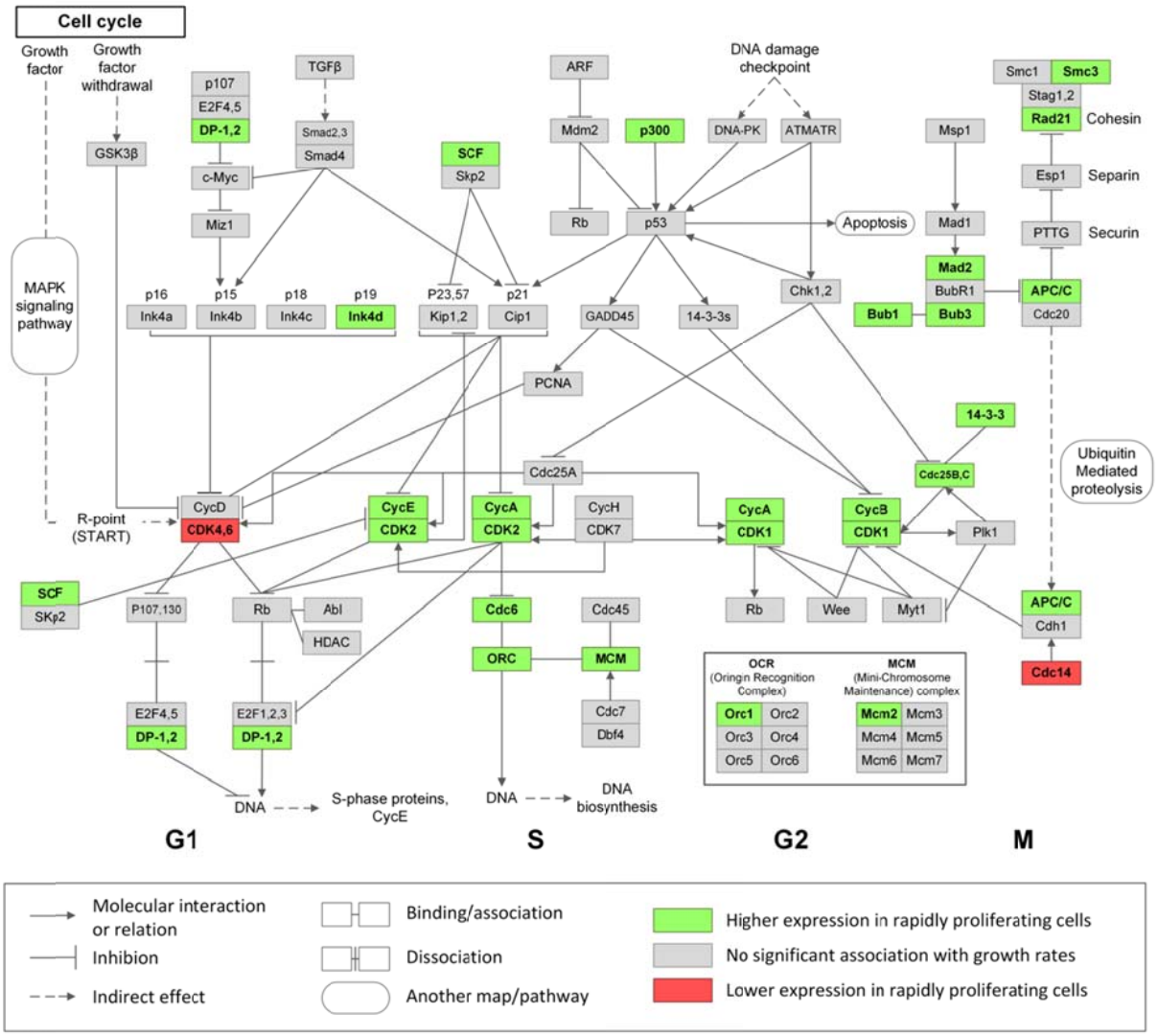

Figure 18: Cell cycle KEGG pathway showing genes with expression levels significantly correlated with the growth of a panel of 31 colorectal cancer cell lines.
Genes are represented by rectangular boxes. Green: higher relative levels in rapidly proliferating cells (Spearman's correlation, FDR<0.1); red: lower levels in rapidly proliferating cells (Spearman's correlation, FDR<0.1); and grey: present on the chip, but not significantly correlated.

Consistently, functional group enrichment analysis also identified groups of genes involved in biological processes, cellular functions or molecular compartments. Biological processes, that have long been known to participate in cell cycle regulation both in normal and tumor cells, such as Gene Ontology biological process categories involved in cell cycle, mitosis, RNA processing and DNA metabolic process (Table 9, Appendix 2). In addition, other groups of functionally related genes whose expression levels are associated with growth rates included RNA splicing, protein transport and ubiquitin-dependent protein catabolic process (Table 9, Appendix 2). The majority of cellular compartments are intracellular organelle or membrane-bounded organelle (Table 10, Appendix 2). The main molecular functions involve binding to various molecules like proteins or nucleic acids (Table 11, Appendix 2).

Table 9: Functional group enrichment analysis using Gene Ontology categories.
The 20 most significantly enriched biological processes are shown (Benjamini-Hochberg-corrected p<0.05).

Term [1]	Count [2]	% [3]	p Value [4]	Pop Hits [5]	Fold Enrichment	Benjamini adjusted *P*
GO:0044260 cellular macromolecule metabolic process	525	40.70	2.00E-29	5214	1.46	6.09E-26
GO:0044237 cellular metabolic process	622	48.22	1.59E-28	6636	1.36	2.41E-25
GO:0043170 macromolecule metabolic process	543	42.09	5.14E-24	5710	1.38	3.90E-21
GO:0009987 cellular process	849	65.81	4.22E-24	10541	1.17	4.27E-21
GO:0008152 metabolic process	660	51.16	7.54E-20	7647	1.25	4.58E-17
GO:0044238 primary metabolic process	610	47.29	2.23E-19	6923	1.28	1.13E-16
GO:0044267 cellular protein metabolic process	259	20.08	4.44E-16	2355	1.60	1.93E-13
GO:0007049 cell cycle	116	8.99	1.24E-15	776	2.17	4.64E-13
GO:0000278 mitotic cell cycle	70	5.43	1.57E-14	370	2.75	5.33E-12
GO:0016070 RNA metabolic process	129	10.00	1.78E-14	938	2.00	5.40E-12
GO:0008380 RNA splicing	59	4.57	3.66E-14	284	3.02	1.01E-11
GO:0016071 mRNA metabolic process	69	5.35	5.17E-14	370	2.71	1.31E-11
GO:0006397 mRNA processing	63	4.88	6.91E-14	321	2.85	1.62E-11
GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	330	25.58	8.74E-13	3409	1.41	1.90E-10
GO:0006996 organelle organization	160	12.40	1.06E-12	1332	1.74	2.14E-10
GO:0000087 M phase of mitotic cell cycle	48	3.72	3.72E-12	224	3.11	7.07E-10
GO:0016043 cellular component organization	255	19.77	5.64E-12	2498	1.48	1.01E-09
GO:0022402 cell cycle process	85	6.59	1.02E-11	565	2.18	1.55E-09
GO:0000377 RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	38	2.95	9.40E-12	153	3.61	1.59E-09
GO:0000398 nuclear mRNA splicing, via spliceosome	38	2.95	9.40E-12	153	3.61	1.59E-09

NOTE: Total number of genes in the query list mapped to any gene set in this ontology is 972. Total number of genes on the background list mapped to any gene set in this ontology is 14116.
[1]Term-GO= Gene Ontology term; [2]Count: number of genes associated with this gene set; [3]Percentag: gene associated with this gene set/total number of query genes; [4]p value: modified Fisher Exact p-value; [5]Pop Hits: number of genes annotated to this gene set on the background list.

Table 10: Functional group enrichment analysis using Gene Ontology categories.
The 20 most significantly enriched cellular compartments are shown (Benjamini-Hochberg-corrected p<0.05).

Term [1]	Count [2]	% [3]	p Value [4]	Pop Hits [5]	Fold Enrichment	Benjamini adjusted *P*
GO:0005622 intracellular	939	72.79	5.73E-63	10995	1.31	3.22E-60
GO:0044424 intracellular part	921	71.40	1.16E-62	10624	1.32	3.26E-60
GO:0043229 intracellular organelle	790	61.24	1.09E-41	8977	1.34	2.03E-39
GO:0043226 organelle	790	61.24	2.10E-41	8989	1.34	2.95E-39
GO:0043227 membrane-bounded organelle	715	55.43	6.68E-36	7989	1.37	7.50E-34
GO:0043231 intracellular membrane-bounded organelle	714	55.35	1.06E-35	7982	1.37	9.94E-34
GO:0044446 intracellular organelle part	454	35.19	6.92E-35	4225	1.64	4.86E-33
GO:0044422 organelle part	456	35.35	6.70E-35	4251	1.64	5.38E-33
GO:0044428 nuclear part	251	19.46	1.18E-32	1822	2.11	7.34E-31
GO:0005634 nucleus	504	39.07	1.32E-30	5077	1.52	7.40E-29
GO:0005737 cytoplasm	651	50.47	8.87E-29	7319	1.36	4.53E-27
GO:0070013 intracellular organelle lumen	231	17.91	4.54E-26	1779	1.98	2.13E-24
GO:0031981 nuclear lumen	200	15.50	1.38E-25	1450	2.11	5.97E-24
GO:0031974 membrane-enclosed lumen	235	18.22	5.26E-25	1856	1.93	2.11E-23
GO:0043233 organelle lumen	231	17.91	1.09E-24	1820	1.94	4.10E-23
GO:0032991 macromolecular complex	332	25.74	8.17E-22	3155	1.61	2.87E-20
GO:0043228 non-membrane-bounded organelle	285	22.09	5.91E-21	2596	1.68	1.95E-19
GO:0043232 intracellular non-membrane-bounded organelle	285	22.09	5.91E-21	2596	1.68	1.95E-19
GO:0005654 nucleoplasm	132	10.23	1.01E-19	882	2.29	3.16E-18
GO:0005829 cytosol	164	12.71	7.53E-16	1330	1.88	2.30E-14

NOTE: Total number of genes in the query list mapped to any gene set in this ontology is 1041. Total number of genes on the background list mapped to any gene set in this ontology is 15908.
[1]Term-GO= Gene Ontology term; [2]Count: number of genes associated with this gene set; [3]Percentag: gene associated with this gene set/total number of query genes; [4]p value: modified Fisher Exact p-value; [5]Pop Hits: number of genes annotated to this gene set on the background list.

Table 11: Functional group enrichment analysis using Gene Ontology categories.
The 20 most significantly enriched molecular funcions are shown (Benjamini-Hochberg-corrected p<0.05).

Term [1]	Count [2]	% [3]	p Value [4]	Pop Hits [5]	Fold Enrichment	Benjamini adjusted *P*
GO:0005515 protein binding	703	54.50	1.09E-19	8154	1.24	1.07E-16
GO:0000166 nucleotide binding	238	18.45	2.17E-12	2245	1.53	1.07E-09
GO:0003723 RNA binding	101	7.83	8.28E-12	718	2.03	2.72E-09
GO:0005488 binding	939	72.79	1.54E-10	12531	1.08	3.78E-08
GO:0003676 nucleic acid binding	302	23.41	8.60E-09	3264	1.34	1.69E-06
GO:0017076 purine nucleotide binding	190	14.73	1.47E-07	1918	1.43	2.42E-05
GO:0032553 ribonucleotide binding	183	14.19	1.77E-07	1836	1.44	2.49E-05
GO:0032555 purine ribonucleotide binding	183	14.19	1.77E-07	1836	1.44	2.49E-05
GO:0003824 catalytic activity	434	33.64	6.11E-07	5198	1.21	7.51E-05
GO:0001883 purine nucleoside binding	159	12.33	1.80E-06	1601	1.43	1.77E-04
GO:0017111 nucleoside-triphosphatase activity	85	6.59	2.07E-06	728	1.69	1.85E-04
GO:0001882 nucleoside binding	160	12.40	1.71E-06	1612	1.43	1.87E-04
GO:0005524 ATP binding	148	11.47	2.74E-06	1477	1.45	2.25E-04
GO:0032559 adenyl ribonucleotide binding	149	11.55	3.58E-06	1497	1.44	2.71E-04
GO:0030554 adenyl nucleotide binding	155	12.02	4.53E-06	1577	1.42	3.18E-04
GO:0016462 pyrophosphatase activity	86	6.67	5.28E-06	757	1.64	3.46E-04
GO:0016818 hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	86	6.67	6.26E-06	760	1.63	3.85E-04
GO:0016817 hydrolase activity, acting on acid anhydrides	86	6.67	7.68E-06	764	1.62	4.44E-04
GO:0003712 transcription cofactor activity	49	3.80	1.08E-05	363	1.95	5.89E-04
GO:0008094 DNA-dependent ATPase activity	15	1.16	2.45E-05	57	3.80	1.27E-03

NOTE: Total number of genes in the query list mapped to any gene set in this ontology is 1049. Total number of genes on the background list mapped to any gene set in this ontology is 14143.
[1]Term-GO= Gene Ontology term; [2]Count: number of genes associated with this gene set; [3]Percentag: gene associated with this gene set/total number of query genes; [4]p value: modified Fisher Exact p-value; [5]Pop Hits: number of genes annotated to this gene set on the background list.

4.1.3 Identification of PPOX and GAPDH as new candidate therapeutic targets

High rates of proliferation are associated with poor patient prognosis and at least some of the genes with higher relative expression in the tumors with faster growth are likely to be necessary to sustain rapid proliferation. We therefore hypothesized that targeting these genes could impair tumor growth. Genome-wide microarray analysis of the panel of 31 colorectal cancer cell lines investigated identified 966 genes with significantly higher expression in rapidly proliferating tumor cells (genes with negative Spearman's r in list of genes see Appendix 1 as referred to Supplementary Table S2). Importantly, thymidylate synthase (TYMS), the direct target of the well-established chemotherapeutic agent 5-fluorouracil (5-FU), the gold standard agent for the treatment of colorectal cancer patients for over five decades, was among the top 50 genes with highest negative correlation between doubling time and gene expression (Figure 19A). Because of the availability of chemical inhibitors, we selected two additional genes, protoporphyrinogen oxidase (PPOX) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), that showed significant negative correlations between gene expression and the doubling time of colorectal cancer cell lines (Figure 19B-C).

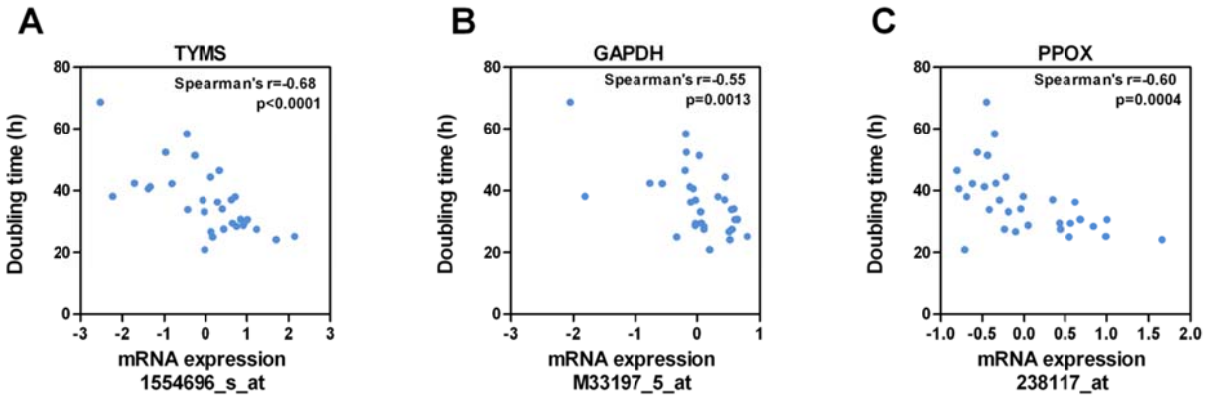

Figure 19: Correlations between expression levels of the novel candidate therapeutic targets and the doubling time in colorectal cancer cell lines.
The expression of thymidylate synthase (TYMS; A), the direct target of the chemotherapeutic agent 5-FU, as well as glyceraldehyde 3-phosphate (GAPDH; B) and protoporphyrinogen oxidase (PPOX; C) were negatively correlated with the doubling time of a panel of 31 colorectal cancer cell lines.

Importantly, the levels of expression of PPOX and GAPDH were significantly correlated with the rates of proliferation (percentage of mitotic cells) in a cohort of 36 primary colorectal tumors (Figure 20). No associations were observed between PPOX/GAPDH mRNA levels and tumor size, site, pathological TNM staging, venous invasion, patient age, gender or overall survival (Cox regression p>0.56) in an extended cohort of 433 colorectal primary tumors (Table 12). A modest reduction in PPOX levels was observed in late stage tumors (Table 12).

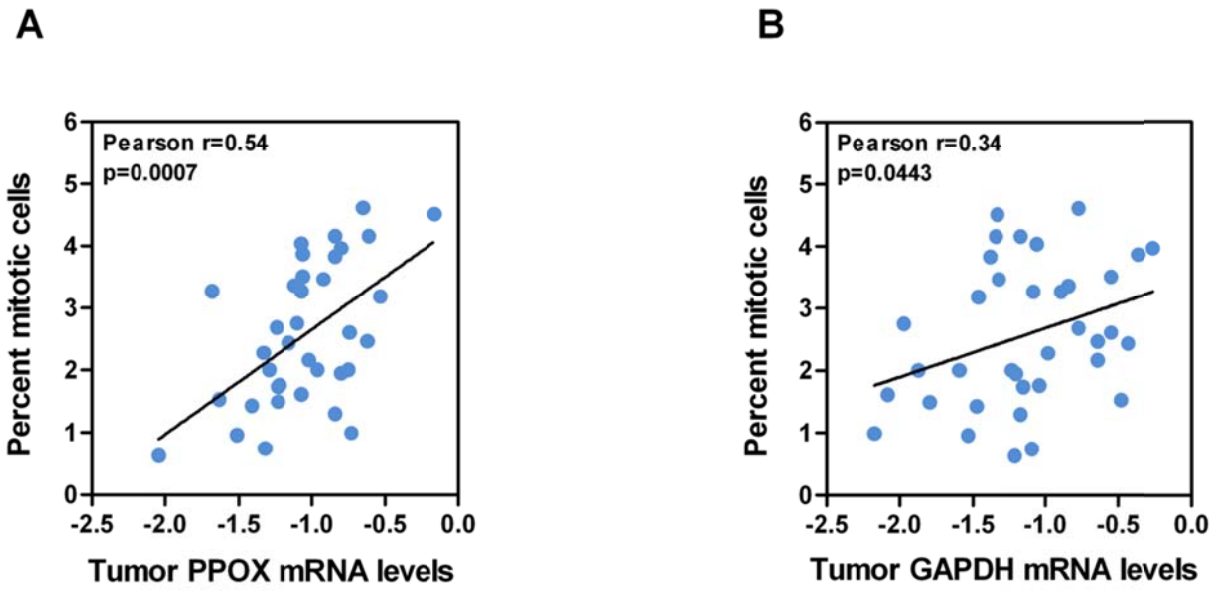

Figure 20: Correlation between PPOX or GAPDH expression and rates of proliferation in primary tumors.
The correlation between the mRNA levels of A) PPOX and B) GAPDH expression in a cohort of 36 primary colorectal tumors (TCGA) and their proliferation rates (percentage of mitotic cells) is shown. The Pearson's correlation coefficient (r) and p value are shown.

Table 12: Associations between PPOX or GAPDH levels and clinicopathological features of a cohort of 433 colorectal tumors from the TCGA.

	TYMS		PPOX		GAPDH	
Gender Mean±SD (n)		p value		p value		p value
Male	9.58±0.06 (196)	0.4120	7.54±0.03 (196)	0.77[1]	16.05±0.04 (196)	0.41[1]
Female	9.58±0.05 (228)		7.55±0.03 (228)		16.10±0.04 (228)	
Site Mean±SD (n)						
Colon	9.67±0.04 (320)	0.0142	7.51±0.03 (320)	0.64[1]	16.13±0.03 (320)	0.22[1]
Rectum	9.30±0.07 (101)		7.48±0.05 (101)		15.93±0.05 (101)	
Venous invasion Mean±SD (n)						
No	9.60±0.05 (284)	0.1927	16.09±0.03 (284)	0.46[1]	7.49±0.03 (284)	0.69[1]
Yes	9.45±0.09 (81)		15.99±0.06 (81)		7.51±0.05 (81)	
Pathologic T Mean±SD (n)						
T1	9.66±0.27 (12)	0.1751	16.09±0.16 (12)	0.68[2]	7.65±0.11 (12)	0.24[2]
T2	9.72±0.09 (66)		16.04±0.07 (66)		7.42±0.07 (66)	
T3	9.59±0.47 (66)		16.06±0.03 (66)		7.50±0.28 (66)	
T4	9.38±0.12 (53)		16.12±0.07 (53)		7.58±0.07(53)	
Pathologic N Mean±SD (n)						
N0	9.73±0.05 (237)	0.0003	16.13±0.04 (237)	0.08[2]	7.47±0.03 (237)	0.26[2]
N1	9.37±0.07 (106)		16.05±0.05 (106)		7.54±0.05 (106)	
N2	9.45±0.08 (76)		15.97±0.06 (76)		7.56±0.05 (76)	
Pathologic M Mean±SD (n)						
M0	9.64±0.05 (285)	0.1083	16.12±0.03 (285)	0.38[1]	7.50±0.03 (285)	0.86[1]
M1	9.23±0.09 (58)		15.90±0.07 (58)		7.54±0.06 (58)	
Tumor stage						
I	9.79±0.09 (65)	0.0003	16.10±0.07 (65)	0.042	7.44±0.06 (65)	0.48[2]
II	9.71±0.07 (157)		16.16±0.05 (157)		7.48±0.04 (157)	
III	9.47±0.07 (125)		16.02±0.05 (125)		7.54±0.04 (125)	
IV	9.28±0.07 (62)		15.92±0.07 (62)		7.53±0.06 (62)	
Tumor size pearson's r (n)	0.00 (424)	0.9800	0.00 (424)	0.93[3]	0.04 (424)	0.35[3]
Age pearson's r (n)	-0.07 (293)	0.1963	0.02 (293)	0.67[3]	-0.01 (293)	0.82[3]

[1]Student's T-test; [2]ANOVA; [3]Pearson's correlation

4.1.4 Inhibition of PPOX and GAPDH reduces the growth of colorectal cancer cells *in vitro*

We then used 5-FU, acifluorfen and Na iodoacetate, specific chemical inhibitors of TYMS, PPOX and GAPDH, respectively, to investigate whether their activity is necessary for the growth of colon cancer cells. As expected, treatment with 5-FU, a chemotherapeutic agent clinically used for the treatment of colorectal cancer, efficiently inhibited the growth of colon cancer cells (Figure 21A). Similarly, acifluorfen and Na iodoacetate treatment resulted in a dose-dependent inhibition of the growth of colon cancer cells (Figure 21B-C), that was not dependent on the growth rates of the cell lines (Figure 21D-F).

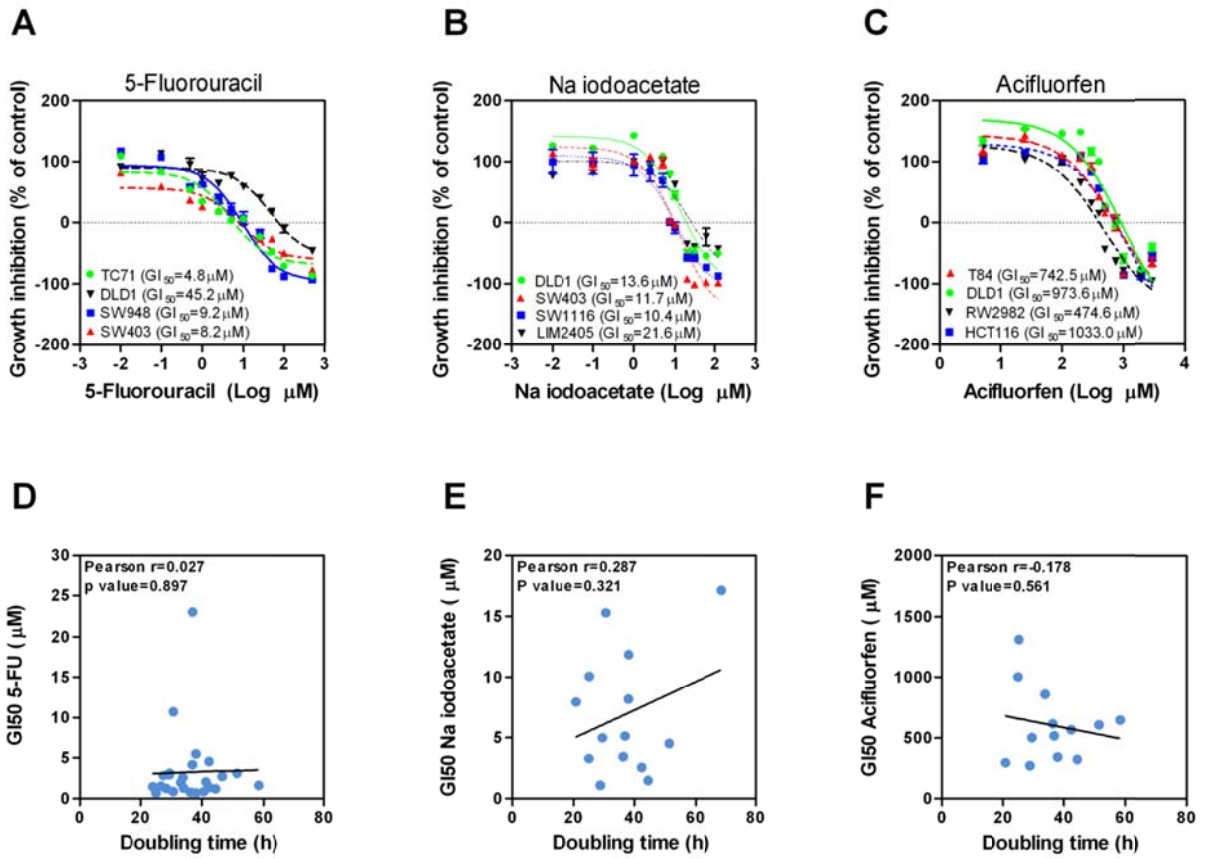

Figure 21: Direct targeting of the novel candidate therapeutic targets inhibits the growth in colorectal cancer cell lines but is not dependent of the growth rate of the cell lines.
Inhibition of A)TYMS with 5-FU, B) GAPDH with Na iodoacetate and C) PPOX with acifluorfen resulted in a dose dependent inhibition of the growth of different colorectal cancer cell lines. The GI50 of the cell lines tested is shown for A) 5-fluorouracil (TC71, DLD1, SW948, SW403), for B) Na iodoacetate (DLD1, SW403, SW116, LIM2405), and for C) acifluorfen (T84, DLD1, RW2982, HCT116). The doubling time of colon cancer cell lines is plotted against the sensitivity (GI50) of these cell lines to the D) TYMS inhibitor 5-FU, E) the GAPDH inhibitor Na iodoacetate or F) the PPOX inhibitor acifluorfen. The Pearson's correlation coefficient (r) and p- value are shown.

Treatment of colorectal cancer cell lines with acifulorfen or Na iodoacetate resulted in significant changes in the cell cycle. Treatment with three different concentrations of acifluorfen (Figure 22) or Na iodoacetate (Figure 23) showed an increased percentage of cells with a subdiploid content of DNA, characteristic of apoptotic cells, compared to control untreated cells.

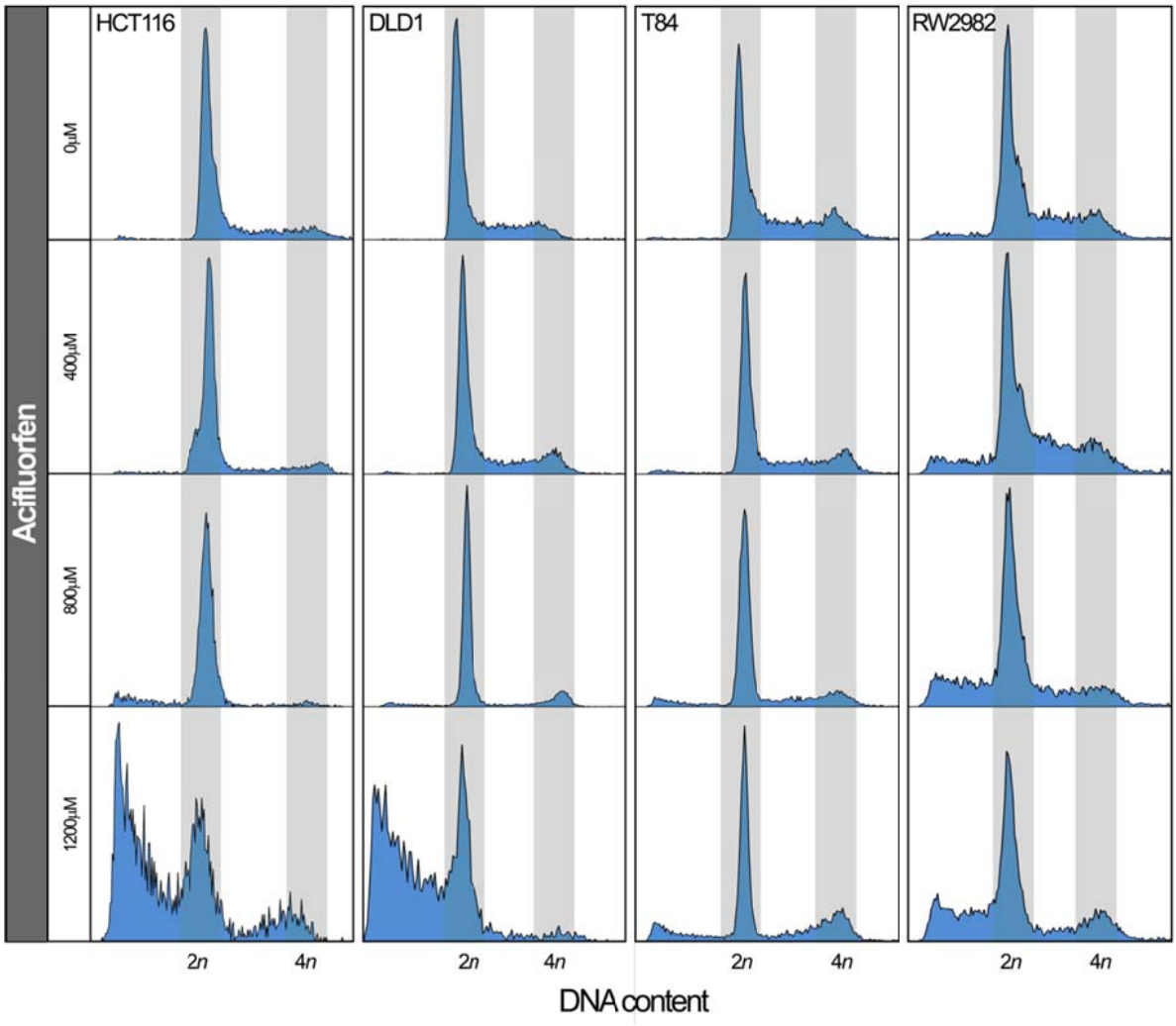

Figure 22: Effects of acifluorfen treatment on the cell cycle of colon cancer cells.
The effects of acifluorfen on the distribution of cells in the different phases of the cell cycle were assessed by propidium iodide and flow cytometry analysis. The results of a representative experiment are shown. Histograms represent the G0/G1, G2/M and S-phase of the cell cycle (DNA content *2n*, *4n* and 2<n<4, respectively), and the sub-diploid fraction is indicated as <*2n*.

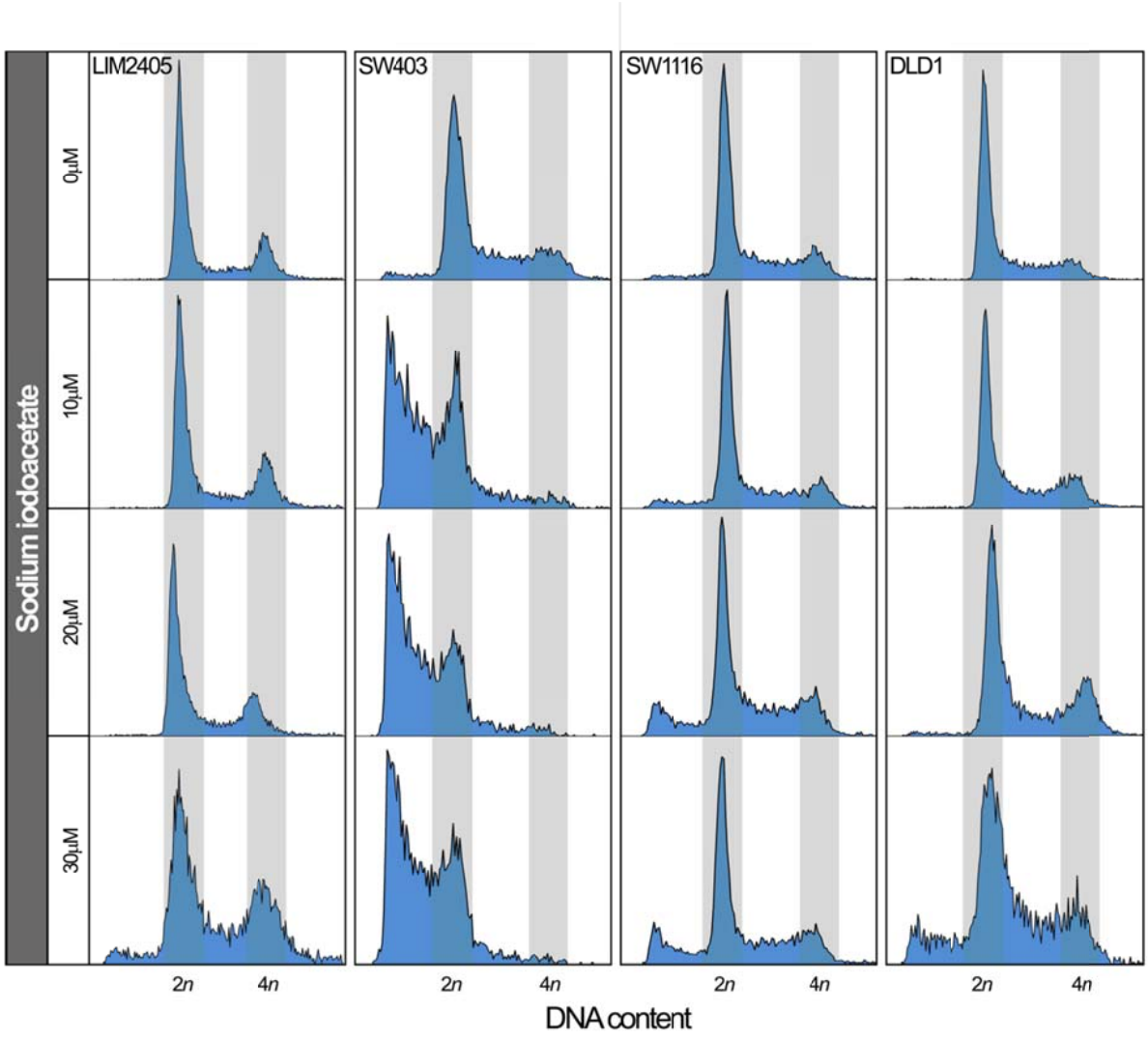

Figure 23: Effects of Na iodoacetate treatment on the cell cycle of colon cancer cells.
The effects of Na iodoacetate on the distribution of cells in the different phases of the cell cycle were assessed by propidium iodide and flow cytometry analysis. The results of a representative experiment are shown. Histograms represent the G0/G1, G2/M and S-phase of the cell cycle (DNA content *2n*, *4n* and 2<n<4, respectively), and the sub-diploid fraction is indicated as <*2n*.

The quantification of three individual experiments as shown in Figure 22 and Figure 23, revealed acifluorfen treatment was associated with an arrest of the cell cycle in the G0/G1 phase (Figure 24A-D). Cells treated with Na iodoacetate did not consistently show the same effect (Figure 24E-H).

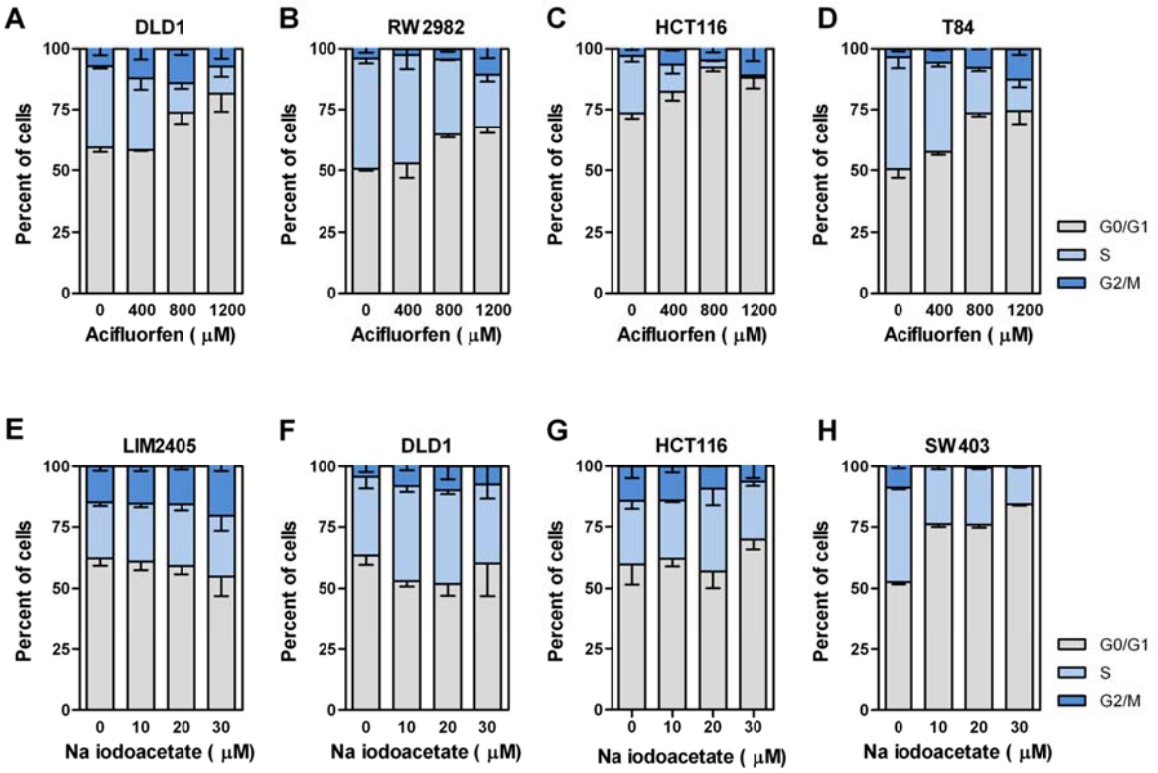

Figure 24: Effects of acifluorfen or Na iodoacetate treatment on the cell cycle of colon cancer cells.
The number of cells in the G0/G1, G2/M and S-phase of the cell cycle (DNA content *2n*, *4n* and 2<n<4, respectively) after acifluorfen (A-D) or Na iodoacetate (E-H) in the indicated cell lines was quantified and the average (±SEM) of three independent experiments, each carried out in triplicate, is shown.

Consistently, flow cytometry analysis of propidium iodide-stained cells after acifluorfen or Na iodoacetate treatment revealed induction of apoptosis, presented by the sub-diploid fraction of the cells (Figure 25A and B). Additionally, PARP cleavage, an indicator of apoptosis is shown in (Figure 25C and D). Moreover, both acifluorfen and Na iodoacetate significantly reduced the long term (>2 weeks) clonogenic capacity of colon cancer cells after short term (9h) treatment, suggesting that these agents could cause cell death in addition to growth inhibition (Figure 25E and F).

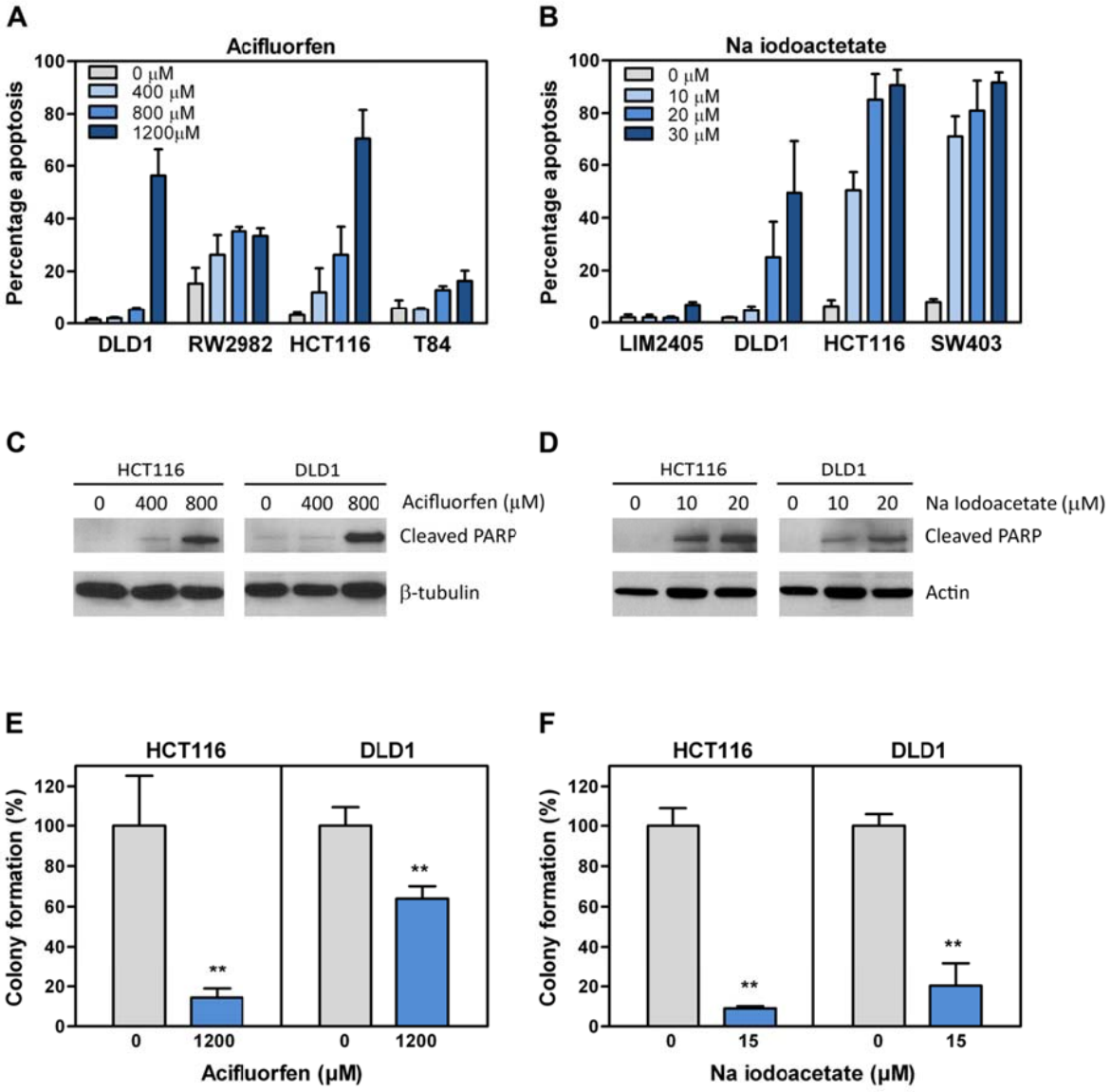

Figure 25: Effects of acifluorfen and Na iodoacetate on the clonogenic potential of colon cancer cells and induction of apoptosis.
The percentage of apoptotic cells after treatment of different colon cancer cell lines with the indicated concentrations of A) acifluorfen or B) Na iodoacetate was assessed by quantification of the number of cells with a sub-diploid content of DNA as determined by propidium iodide staining and flow cytometry. The average (±SEM) of three independent experiments each in triplicate is shown. The indicated colon cancer cell lines were treated with C) acifluorfen or D) Na iodoacetate for 24h, and the presence of cleaved poly ADP ribose polymerase (PARP) was assessed by Western blotting. The number of macroscopically visible colonies after treatment of HCT116 and DLD1 cells with E) acifluorfen and F) Na iodoacetate was assessed 2-3 weeks after 9h treatment with the indicated concentrations. Three independent experiments were carried out in triplicate and the average percentage (±SEM) relative to untreated controls is shown.

Treatment of colon cancer cells with two additional chemically unrelated inhibitors of PPOX and GAPDH (oxadiazon and CGP 3466B maleate, respectively) also resulted in a dose-dependent growth inhibition of colon cancer cells (

Figure 26A-B). Moreover, RNAi-based knock down of PPOX and GAPDH also interfered with the growth of colon cancer cells (

Figure 26C-F), further indicating that PPOX and GAPDH are necessary for proliferation of colon cancer cells.

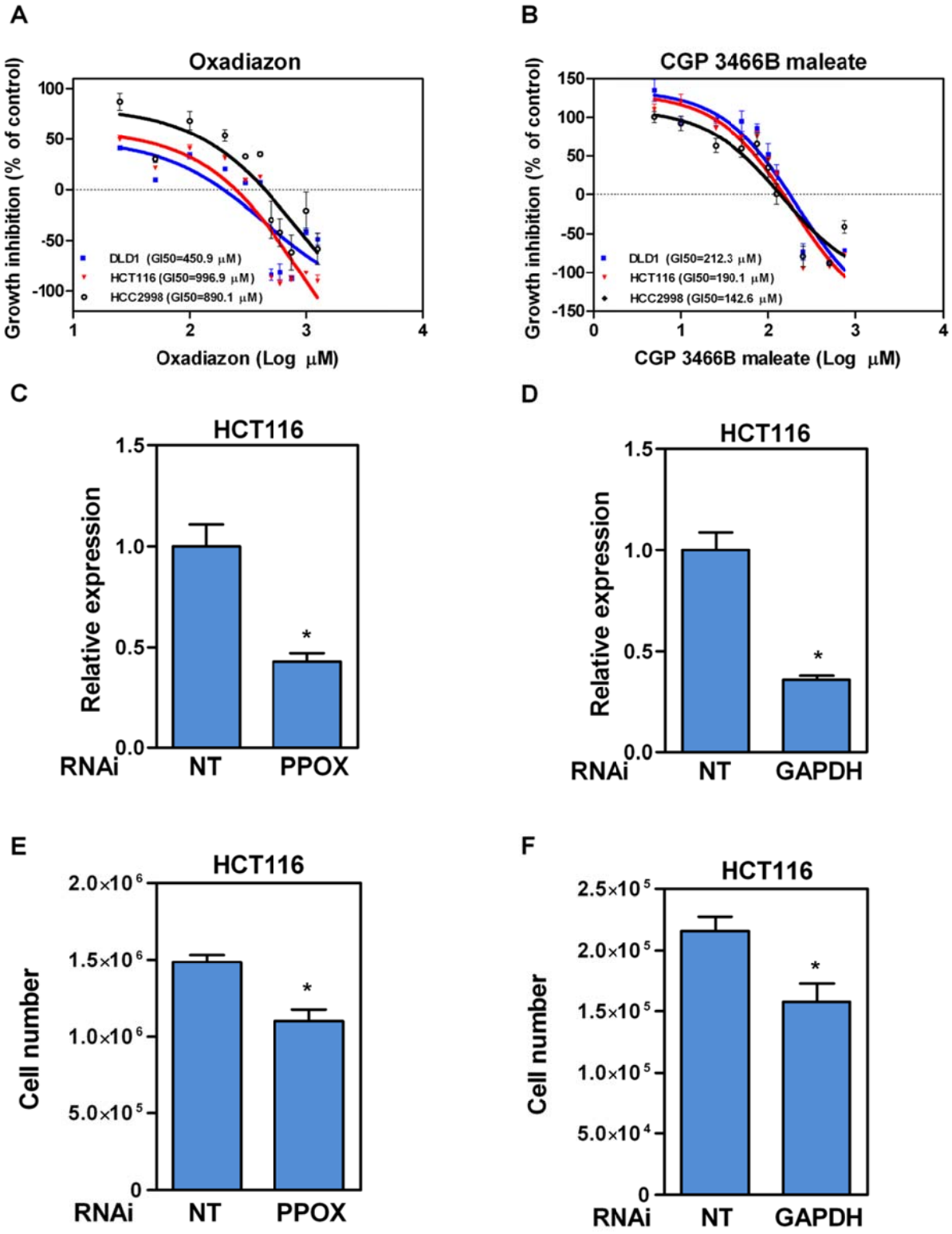

Figure 26: Effects of PPOX and GAPDH inhibition on the growth of colon cancer cells.

Treatment with the PPOX inhibitor A) oxadiazon or the B) GAPDH inhibitor CGP 3466B maleate caused a dose-dependent reduction in the growth of colon cancer cells, as determined by sulforhodamine B staining. Transfection of RNAi oligos targeting C) PPOX or D) GAPDH resulted in a reduction in the levels of mRNA expression of these genes compared to untreated controls or cells transfected with a non-target (NT) RNAi (all 10nM) as determined by qPCR. The number of cells 72h after control non-target (NT), E) PPOX or F) GAPDH RNAi transfection was directly counted. The average (±SEM) of three independent experiments each in triplicate is shown. *Student's T-test $p<0.05$.

4.1.5 PPOX inhibition reduces the growth of colon cancer cells in a xenograft model

The *in vitro* experiments suggested that PPOX and GAPDH could constitute novel therapeutic targets for colorectal cancer. To further investigate this possibility, we used a xenograft model in NOD/SCID immunodeficient mice. DLD1 and HCT116 cells were subcutaneously injected into the flanks of 24 animals and when the tumors reached a volume of 80mm^3 the animals were randomized to a control group, or groups treated with acifluorfen, Na iodoacetate or 5-FU (Figure 27A–B). As expected, 5-FU treatment reduced the growth of these colon cancer cell lines (Figure 27A). Although treatment with the GAPDH inhibitor Na iodoacetate did not have any effect on the growth of these cell lines, systemic administration of the PPOX inhibitor acifluorfen resulted in a significant inhibition of the growth of DLD1 cells (Figure 27A). To further investigate the sensitivity of colon cancer cell lines to acifluorfen and Na iodoacetate, additional cell lines were subcutaneously injected into immunodeficient NOD/SCID mice that were treated with these agents. While Na iodoacetate did not significantly affect the growth of these additional cell lines, the growth of T84 and Isreco1 cells was significantly reduced in animals treated with acifluorfen (Figure 27 C-H). We did not include the 5-FU treatment, since we were interested in a growth inhibition compared to the untreated control (PBS group). The loss of weight during the treatments was within the accepted rage imposed by the ethical committee for animal experimentation from the University Hospital Vall d'Hebron, Barcelona. (Appendix 3). Collectively, these results indicate that PPOX could constitute a novel therapeutic target for the treatment of colon cancer.

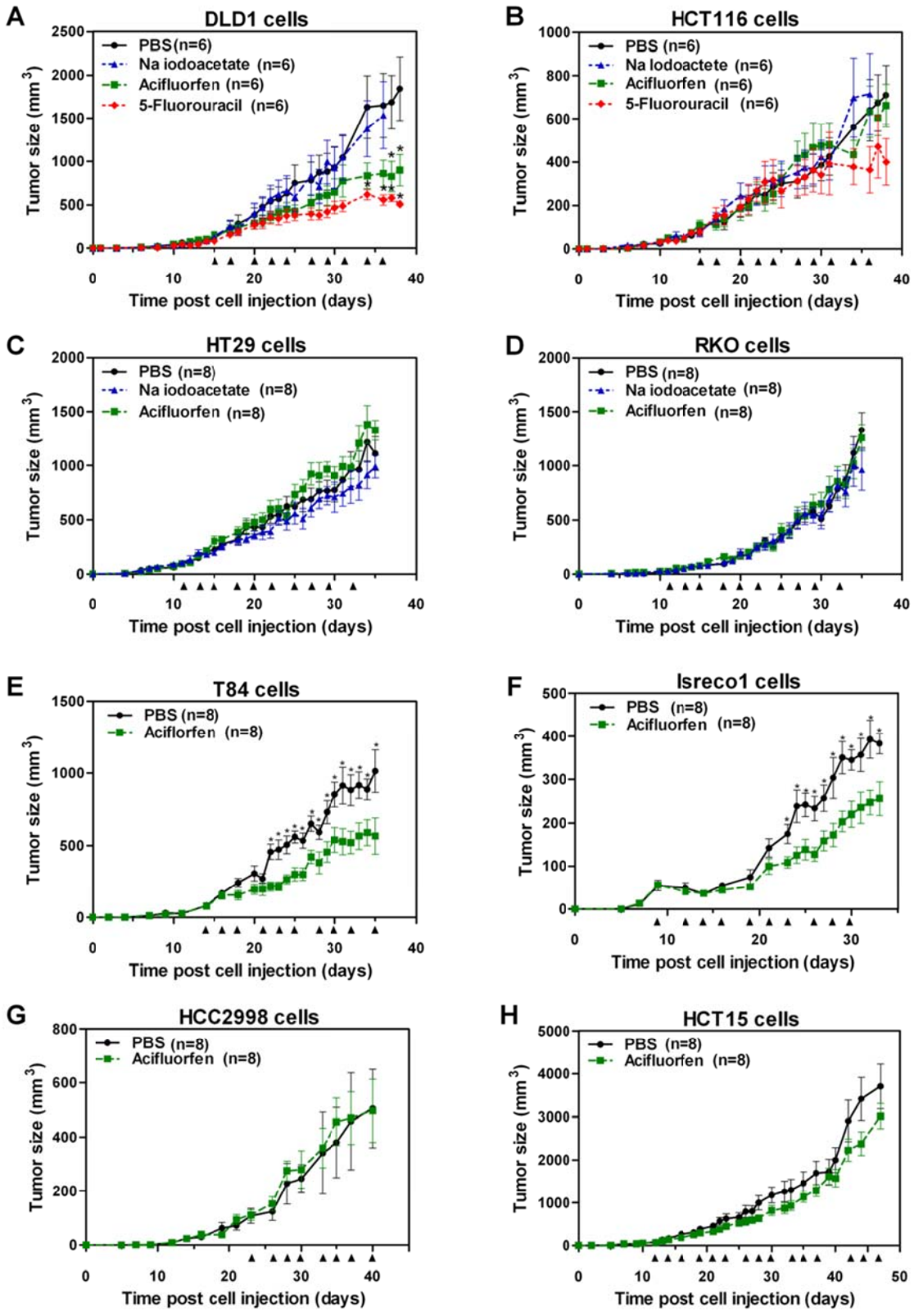

Figure 27: Effects of GAPDH and PPOX inhibition on tumor growth using a xenograft model.
Groups of NOD/SCID immunodeficient mice with A) DLD1, B) HCT116, C) HT29, D) RKO, E) T84, F) Isreco1, G) HCC2998, and H) HCT15 cells as subcutaneous xenografts were treated (i.p.) three times per week with acifluorfen (168mg/kg), Na iodoacetate (18.4mg/kg), 5-fluorouracil (50mg/kg) or vehicle PBS as indicated, starting when the tumors reached approximately 80mm3. Arrowheads in the X-axis indicate treatment times. Tumor size was monitored over time. Asterisk indicates statistically significant differences (Student's T-test, $p<0.05$) in the mean tumor size in the control (PBS) group and treatment groups (5-FU or acifluorfen). The mean ± SEM is shown.

4.2 Expression regulation by epigenetic silencing

4.2.1 Genome-wide analysis of CpG methylation in colorectal cancer cell lines

To investigate the levels of CpG methylation throughout the genome of colorectal cancer cells we used a microarray approach with single CpG dinucleotide resolution to quantify the levels of methylation of the promoters of 14,475 consensus coding sequences (CCDS) in a panel of 45 different colorectal cancer cell lines. The study design included an unmethylated control (mean methylation 5 %), an *in vitro* methylated control (mean methylation 86 %; Figure 28A) and a biological replicate showing excellent correlation between independent methylation measurements (Figure 28B; Pearson $r=0.998$, $p<0.0001$). The average methylation per sample for all the >27,000 individual CpGs interrogated ranged from 22 % to 43 % with an average of 31±5 % (Figure 28A; mean±SD). The average methylation per CpG for all the 45 lines in the study showed a bimodal distribution. Most (>50 %) of the >27,000 CpGs had levels of methylation below 20 %. However, a significant number of CpGs (>30 %) showed methylation levels above 50 % (Figure 28C).

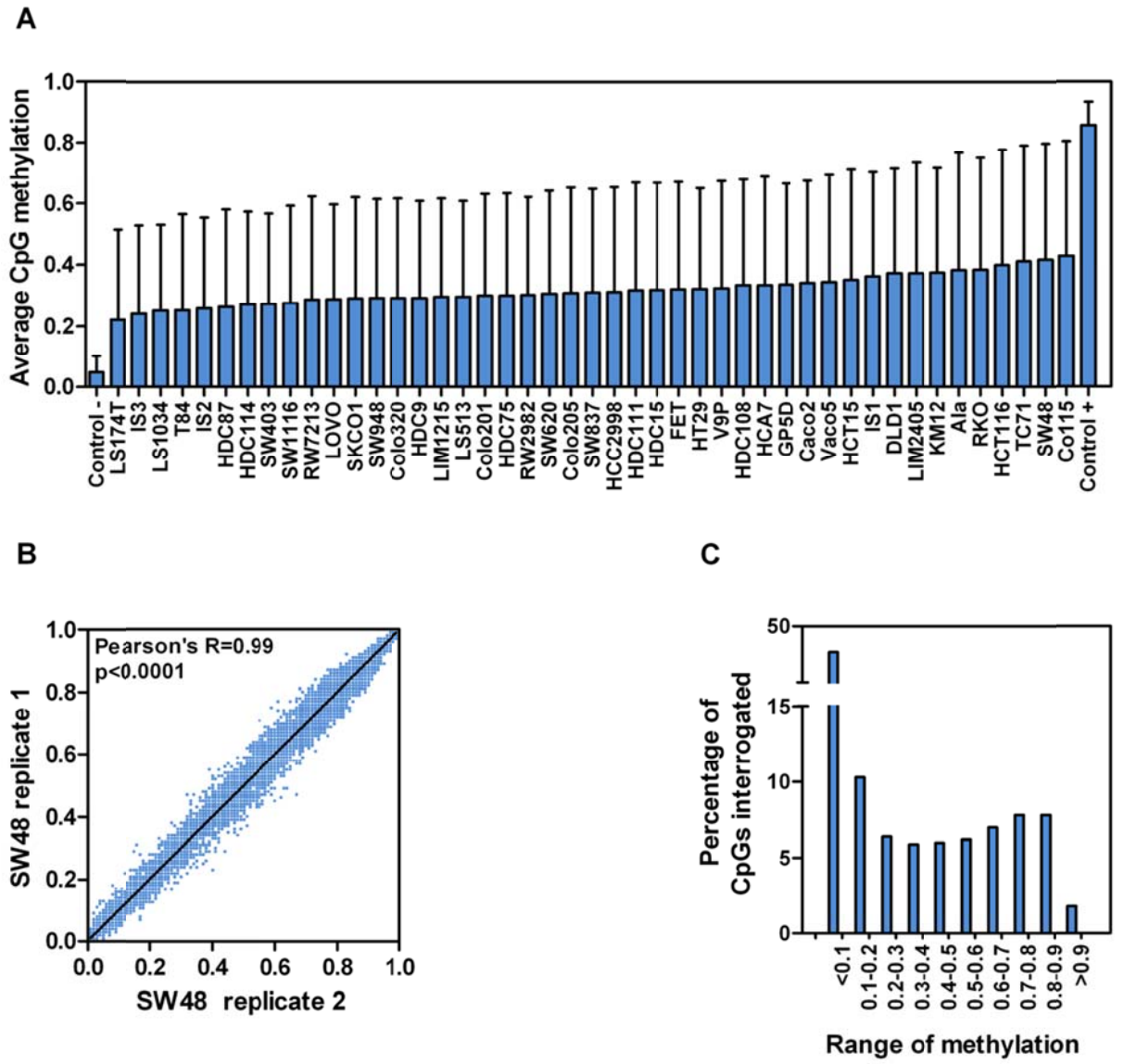

Figure 28: CpG methylation in colorectal cancer cell lines.
A) The average methylation of all the >27000 CpGs interrogated, representing >14,400 genes was variable and ranged from 0.22 to 0.43 with an overall average in all CpGs in all 45 cell lines of 0.31±0.05 (mean±SD). The unmethylated control (NoMet: whole genome amplification with GenomiPhi, GE Healthcare) and the methylated control (MET: in vitro methylated with SssI methylase; New England Biolabs) controls have an average methylation of 0.05 and 0.86, respectively. Colorectal cancer cell lines range from 0.22 to 0.43. B) Correlation between the levels of methylation of >27,000 CpGs in replicate determinations for the cell line SW48 (Pearson's r=0.99; p<0.0001). C) The majority of the CpGs investigated (51.2 %) show average methylation below 20 %, while 30.6 % of the CpGs have average methylation levels greater than 50 % in these 45 colorectal cancer cell lines.

In agreement with the reported association between MSI and CIMP+, the overall levels of methylation were significantly higher in MSI lines compared to MSS lines (Student's T-test p=0.001; Figure 29A). In addition, the average levels of methylation across the >14,000 promoters investigated was significantly lower in the cell lines with *APC* mutations (Student's T-test p=0.043; Figure 29B). Moreover, there was a significant association between higher methylation levels and faster growth in this panel of 45 colorectal cancer cell lines (Figure 29C; Pearson's r=-0.39; p=0.010), suggesting that higher levels of promoter CpG methylation may contribute to the uncontrolled proliferation

characteristic of tumor cells by inhibiting tumor suppressor genes. No associations were found between the overall methylation levels of these 45 cell lines and mutations in *KRAS*, *BRAF, TP53*, or *PIK3CA* (Student's T-test p>0.25; Figure 29D-G).

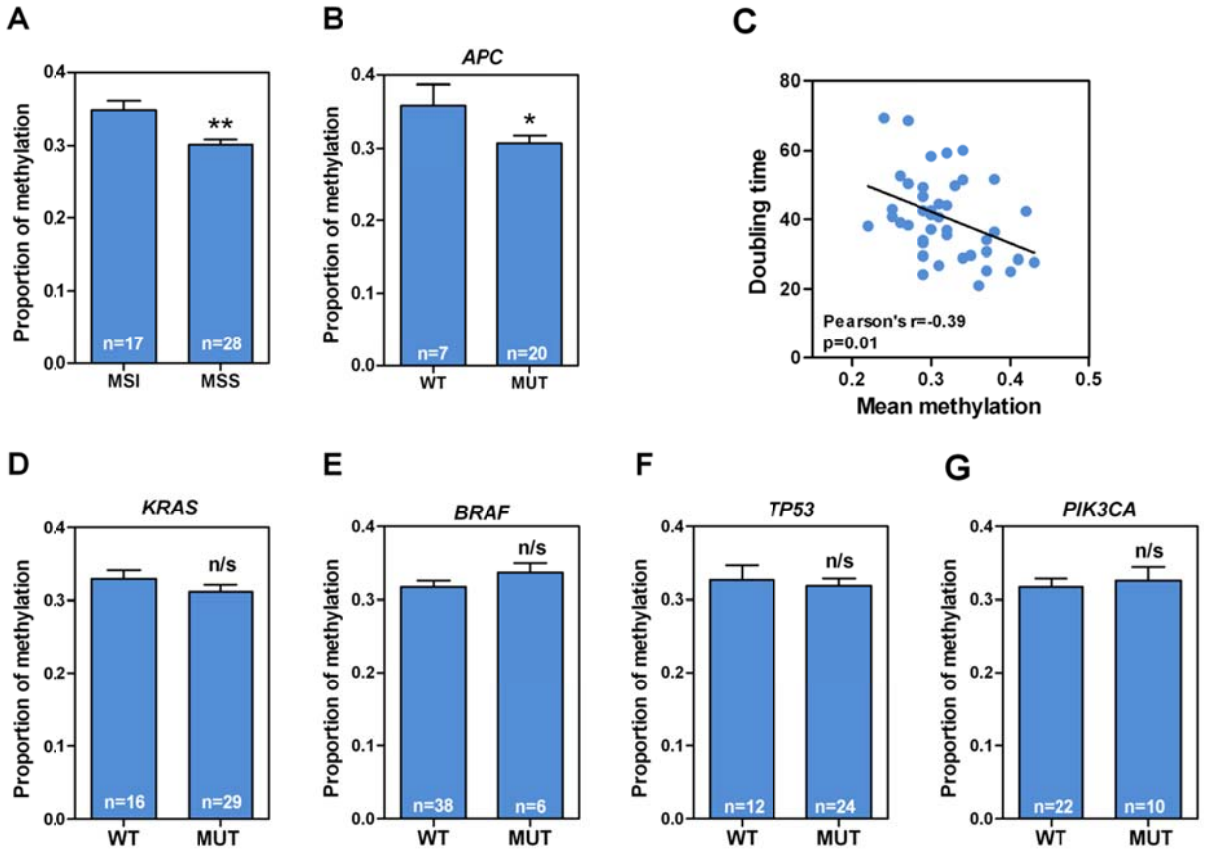

Figure 29: Association of the overall levels of methylation with frequent mutations in colorectal tumors.
A) The overall levels of methylation were significantly (Student's t-test p=0.0037) higher in MSI lines (n=17; 0.35±0.05) than in MSS lines (n=29; 0.3±0.04). B) Lower average methylation levels associated significantly with *APC* mutations (Student's t-test p=0.0428). C) There was a significant negative correlation (Pearson's R=-0.39; p=0.01) between the average levels of methylation and the doubling time of 45 colon cancer cell lines. Student's T-Test: *p<0.05; **p<0.001. No associations were observed between the levels of overall methylation in >27,000 CpG sites and the mutations status of D) *KRAS*, E) *BRAF*, F) *TP53*, and G) *PIC3CA*.

We used bisulfite sequencing as an independent technique to validate the results of the methylation microarrays. We randomly selected four genes showing wide differences in methylation levels in this set of cell lines, and after bisulfite treatment, a genomic region containing the CpG interrogated by the methylation arrays was directly PCR amplified and sequenced (Figure 30E). Bisulfite sequencing results were in perfect agreement with the methylation levels observed with the methylation microarrays (Figure 30A-D).

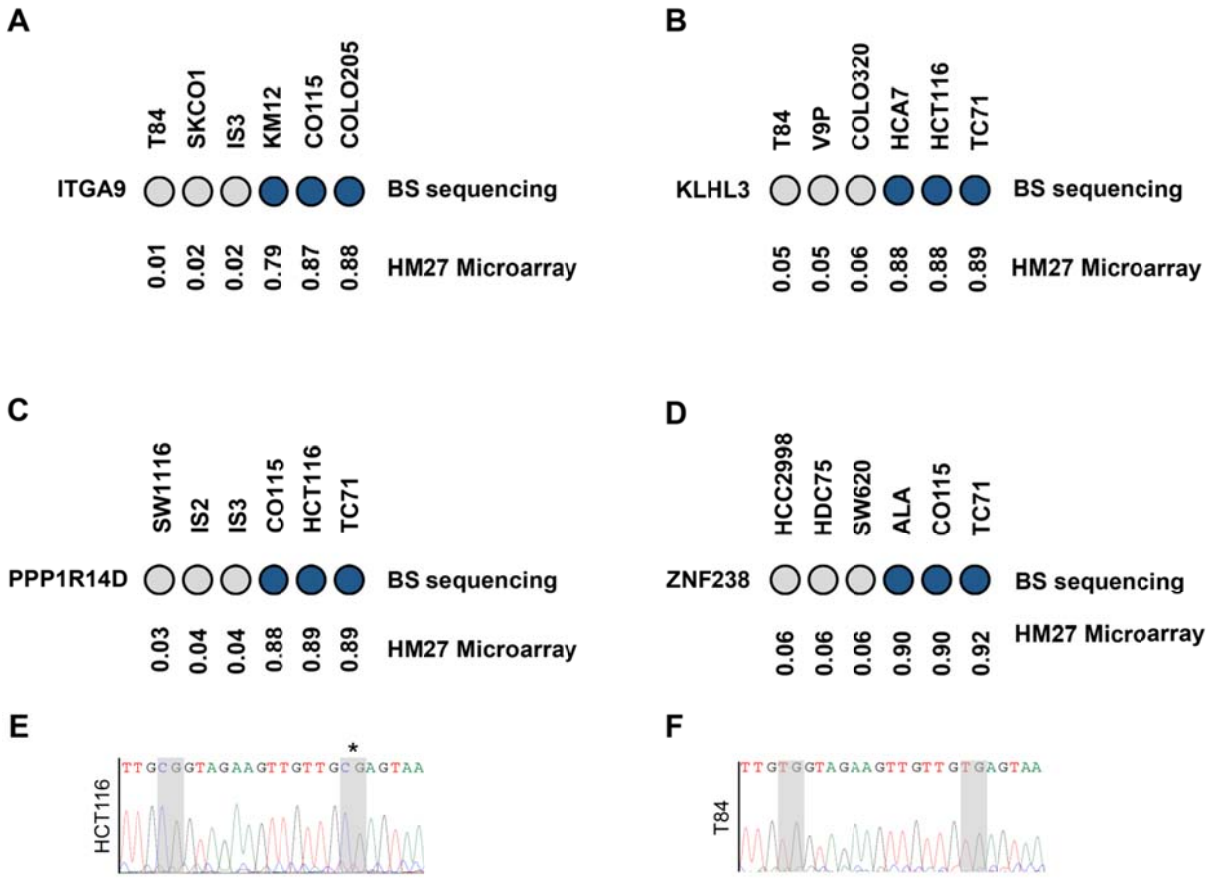

Figure 30: Independent validation of the of the methylation microarrays.
The levels of promoter methylation observed in the genes A) *ITGA9*, B) *KLH3*,C) *PPP1R14D*, and D) *ZNF238* by bisulfite sequencing (gray circle: unmethylated; blue circles: methylated) and by Illumina's HumanMethylation27 microarrays (numbers under the bisulfite sequencing data) are shown for the colorectal cancer cell lines indicated. E) Representative example of bisulfite sequencing assessment of methylation, showing the CpG interrogated by the arrays (asterisk) and a second CpG dinucleotide in fully methylated HCT116 (keeping CpG after bisulfite treatment) and F) unmethylated T84 cells (change of cytosine to thymine after bisulfite treatment). A total of 8, 5, 3 or 11 CpG sites were analyzed for the promoter of *ITGA9*, *KLH3*, *PPP1R14D*, and *ZNF238* respectively.

4.2.2 CpG island methylator phenotype

It has been proposed that widespread CpG island hypermethylation defines a subset of colorectal tumors, known as 'CpG island methylator phenotype positive' (CIMP+) [96,181]. However, the existence of CIMP+ tumors has been challenged in some studies because colorectal tumors could not be clearly dichotomized as CIMP+ and CIMP- since a continues gradient was observed in the overall levels of promoter methylation [182,183]. In good agreement, here we found that if the methylation levels of all the >27,000 CpGs interrogated or a subset 1,516 of them with the highest methylation variability among samples (1.0<SD/Mean<1000) is considered, the distribution of the number of cell lines with different average methylation levels did not show any evidence of bimodality. The original definition of CpG island methylator phenotype used a set of promoters that do not show high methylation levels in normal colonic samples-'type C' promoters [96]. Here, we found that the average methylation of the of a selection of 643 'Type C' promoters in this panel of 45 colon cancer cell lines, showed a

bimodal distribution and could be used to classify a group of 18 cell lines as CIMP+ (Figure 31A). Moreover, clustering analysis of the levels of methylation of these 643 'Type C' promoters demonstrated that CIMP+ cell lines form a homogeneous branch of the dendrogram generated (Figure 31B). Therefore, a threshold methylation value of 0.6 could be used in this series to classify tumor samples as CIMP+ or CIMP-.

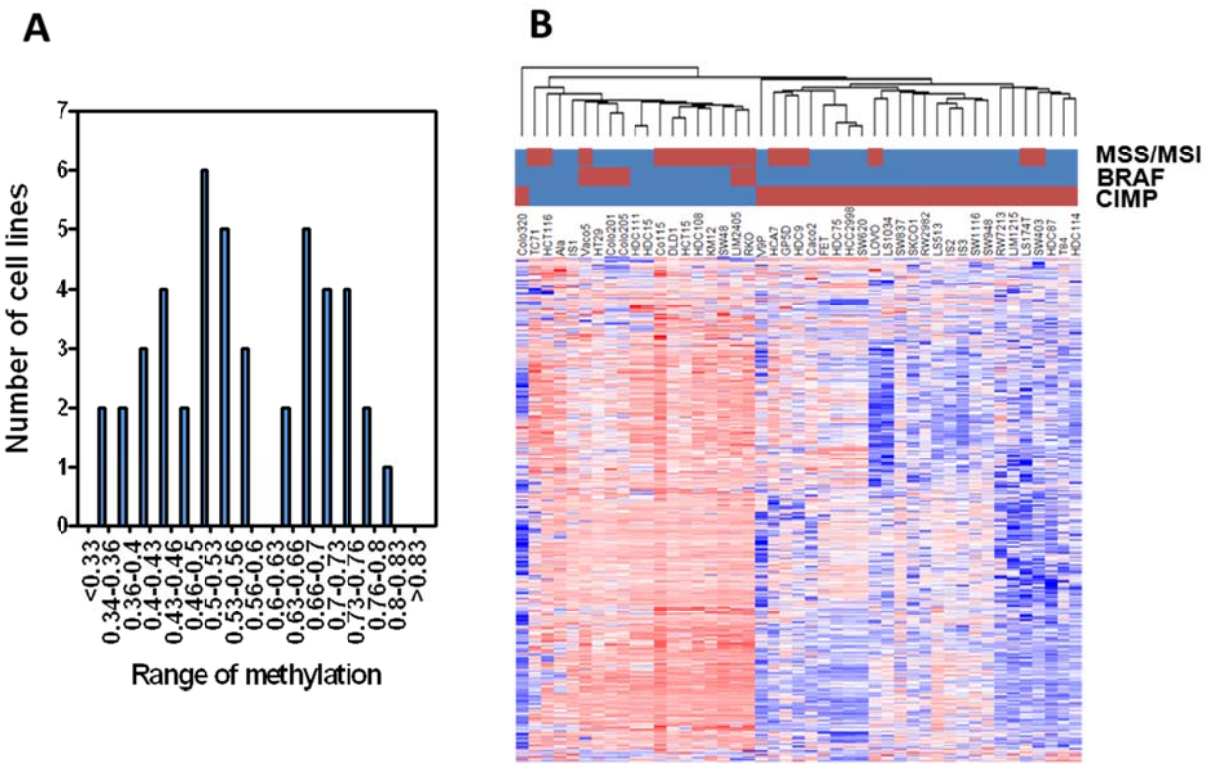

Figure 31: CpG methylator phenotype in colorectal cancer cell lines.
(A) Using 643 Type C promoters a bimodal distribution of the levels of methylation of these 45 the cell lines was observed. (B) Clustering analysis of the levels of methylation of these 643 'Type C' promoters.

4.2.3 Genes transcriptionally regulated by promoter methylation in colorectal cancer

Although it is widely accepted that aberrant CpG methylation is a frequent event in colorectal cancer and other cancer types and that it is important in the transcriptional regulation of many genes, a detailed list of genes whose transcription is frequently regulated by CpG methylation is currently lacking. Here we used microarray analysis (Affymetrix U133 Plus 2.0) to assess the levels of expression of >47,000 transcripts. Expression data was available for a total of 11,858 (81.92 %) of the 14,475 promoters interrogated in the HumanMethylation27 arrays. To validate the expression data

obtained with the mRNA microarrays with an independent technique, we used Real-Time RT-PCR quantification of the levels of expression of four randomly selected genes. The quantitative RT-PCR results obtained are in perfect agreement with the mRNA microarray results (Figure 32A-D).

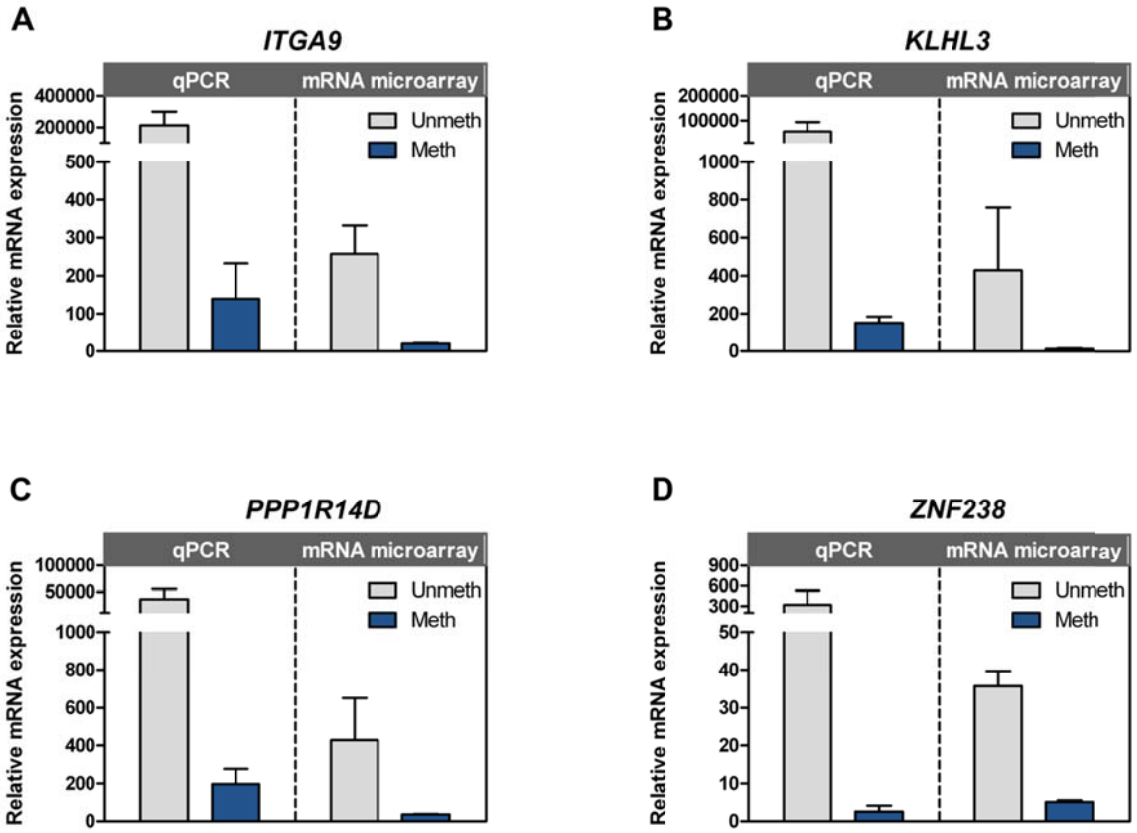

Figure 32: Independent validation of the of the mRNA microarrays.
The comparison of mRNA expression levels observed by quantitative Real-Time RT-PCR and by microarray is shown for A) *ITGA9*, B) *KLH3*, C) *PPP1R14D*, and D) *ZNF238*.

We found a significant inverse correlation (Pearson r<-0.355) between the levels of mRNA expression and the levels of promoter methylation of 1,409 genes (9.7 %). To eliminate genes whose correlation between expression and methylation levels was driven by a small number of samples, we used the criteria described in the 'Materials and Methods' (section 3.2.2.2) and identified a subset of 643 candidate genes whose expression is regulated by promoter methylation (Figure 33A-D).

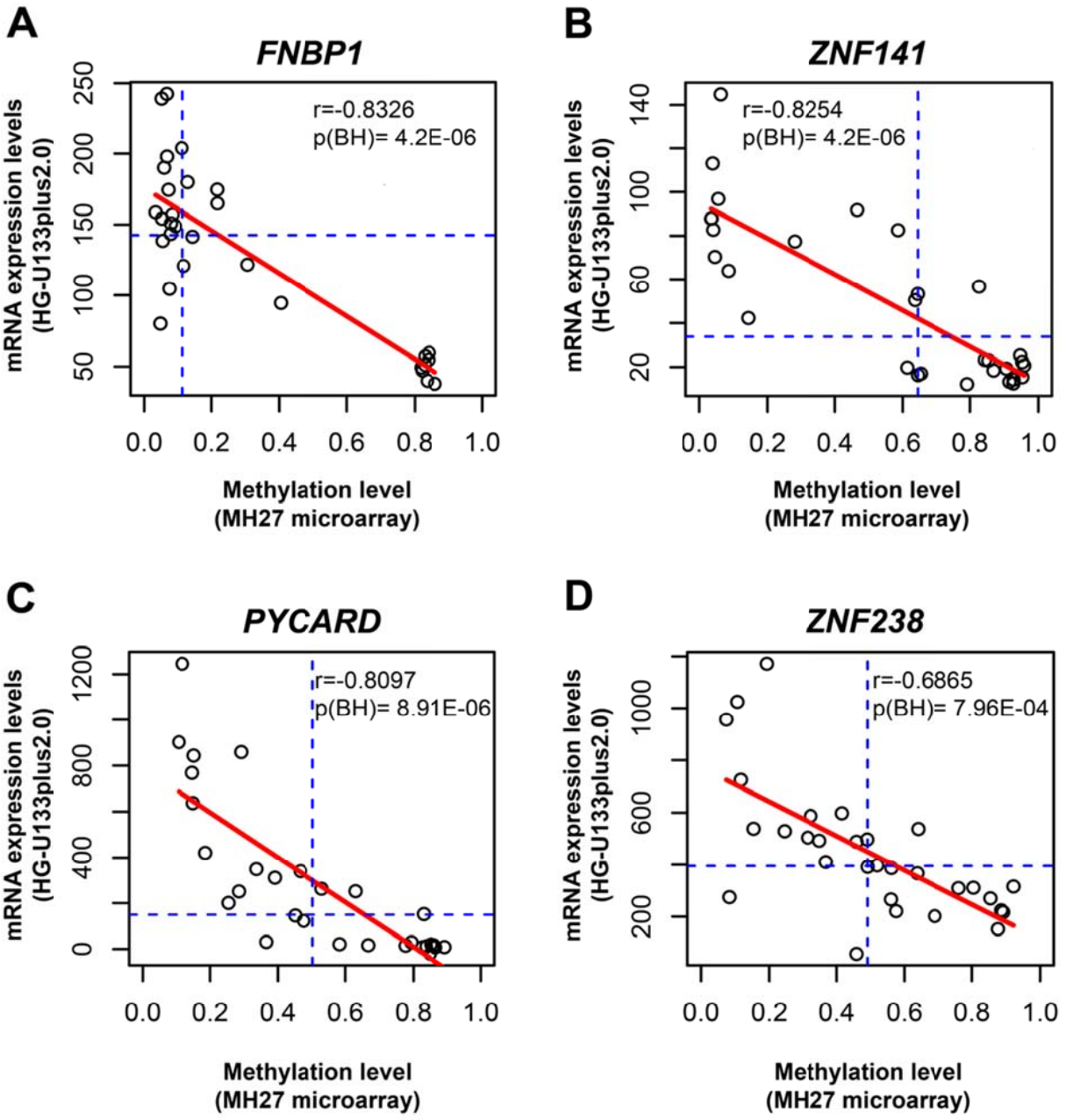

Figure 33: Representative examples of genes showing significant correlations between mRNA and methylation levels. (A-D) shows examples of genes with significant negative correlations between mRNA expression and promoter methylation levels. A significant association was observed between high promoter methylation and low levels of expression.

Moreover, we identified an additional set of 24 genes whose expression was not significantly correlated with the levels of methylation (Pearson's r>-0.355), but are likely to be regulated by promoter methylation (Figure 34A-D; see full list Appendix 4, Table 18).

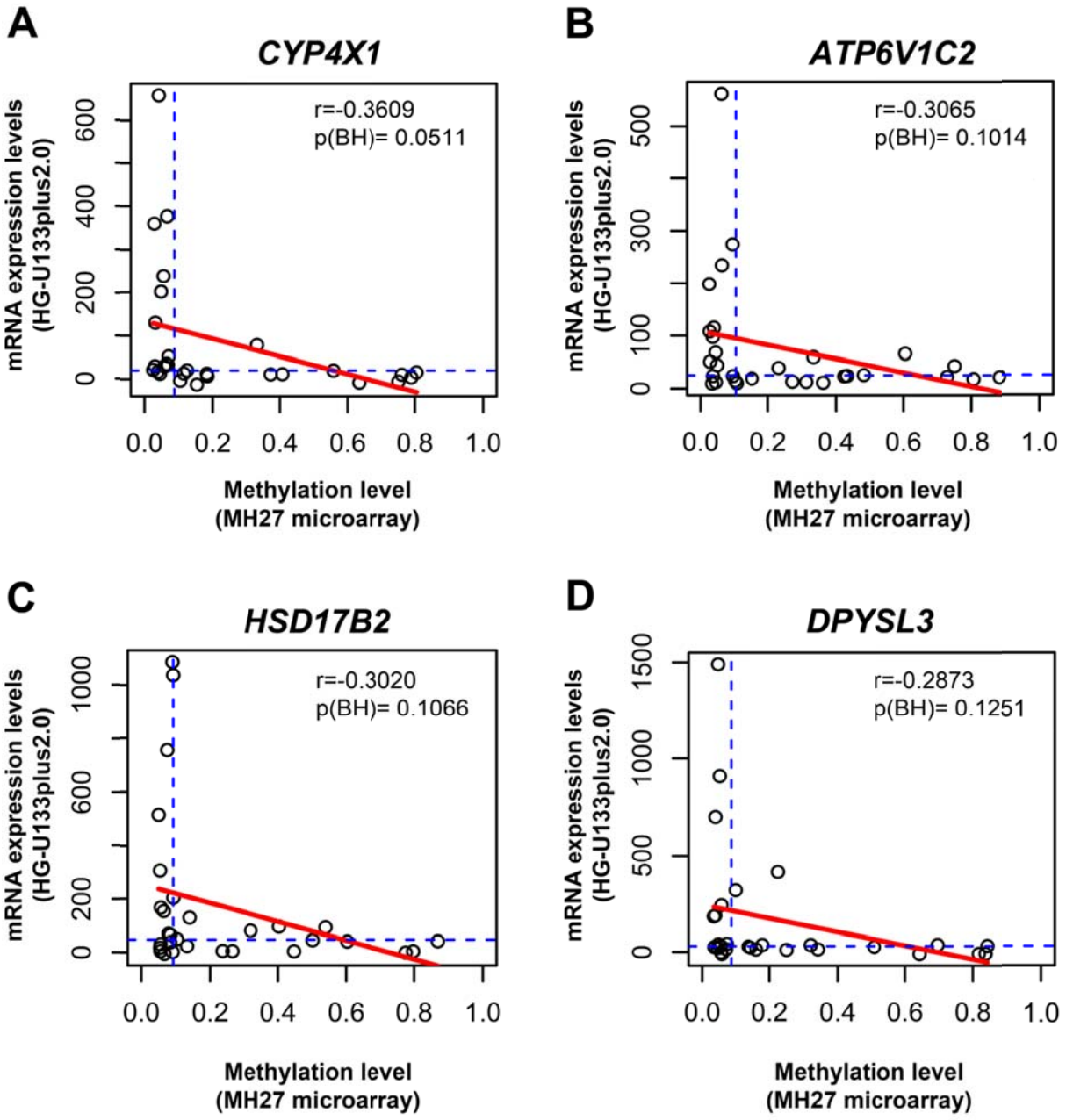

Figure 34: Representative examples of genes showing the "L-shape" between mRNA and methylation levels.
(A-D) are representative examples of genes with no significant correlations between mRNA and promoter methylation levels, but that show the expected 'L-shaped' profile characteristic of genes whose expression is regulated by promoter methylation in this type of plots.

We therefore identified a total of 667 of the 11,858 genes investigated (5.6 %) as candidate genes whose expression is regulated by promoter methylation (Table 13; top 100 genes in Appendix 5, Table 19). Because high levels of promoter methylation are expected to result in low levels of gene expression (i.e., negative correlations), to assess the robustness of the analysis, we searched for genes whose expression was positively correlated with the levels of methylation and found 325 genes. As expected, we observed a strong bias towards negatively correlated genes (1409 of all 1734;

81.25 % of all significantly correlated genes; Binomial test p<10E-30), convincingly showing that there are underlying biological reasons for the observed associations (namely, promoter methylation negatively regulates gene expression) and providing a good estimate of the proportion of false positive genes in our analysis.

To further assess whether these genes were regulated by CpG methylation of their promoter regions, we used mRNA expression microarray analysis of HCT116 colon cancer cells, HCT116 cells treated with the demethylating agent 5-Aza-2'-deoxycytidine, and HCT116 DKO that are double knockout for the DNA methyltransferases DNMT1 and DNMT3B. When looking at the genes that had methylation levels >50 % in parental HCT116 cells (n=260 genes), we found that the majority of the genes (69.6 % and 65.2 %) had higher levels of expression in the HCT116+AZA or HCT116 DKO cells compared to the parental HCT116 cells (binomial p= $6.3x10^{-11}$ and $3.5x10^{-7}$, respectively). Therefore, the combined analysis of mRNA expression and promoter methylation levels identified a core set of 667 genes that are frequently regulated at the transcriptional level by aberrant CpG promoter methylation in colorectal tumors (Table 13).

Table 13: Top ten genes with significant correlations and top ten genes which are L-shaped.

Gene symbol	Type[a]	Name	Pearson r	BH(FDR) adjusted p
FNBP1	Correlated	Formin binding protein 1	-0.83262	4.2E-06
ZNF141	Correlated	Zinc finger protein 141	-0.82544	4.2E-06
PYCARD	Correlated	PYD and CARD domain containing	-0.80969	8.91E-06
MRPS21	Correlated	Mitochondrial ribosomal protein S21	-0.80535	9.02E-06
KLHL3	Correlated	Kelch-like 3 (Drosophila)	-0.77506	4.83E-05
GIPC2	Correlated	GIPC PDZ domain containing family, member 2	-0.77011	5.35E-05
BST2	Correlated	Bone marrow stromal cell antigen 2	-0.76228	7.07E-05
GPR56	Correlated	G protein-coupled receptor 56	-0.75568	8.79E-05
TFCP2	Correlated	Transcription factor CP2	-0.73731	0.000197
LGALS4	Correlated	Lectin, galactoside-binding, soluble, 4	-0.73337	0.000214
CYP4X1	L-Type	Cytochrome P450, family 4, subfamily X, polypeptide 1	-0.36093	0.051151
RNF186	L-Type	Ring finger protein 186	-0.34999	0.059221
KCNK15	L-Type	Potassium channel, subfamily K, member 15	-0.34759	0.061044
C14orf50	L-Type	Chromosome 14 open reading frame 50	-0.34738	0.061117
IGF2BP1	L-Type	Insulin-like growth factor 2 mRNA binding protein 1	-0.34241	0.065142
ATP6V1C2	L-Type	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C2	-0.30653	0.101408
ADAMTS15	L-Type	ADAM metallopeptidase with thrombospondin type 1 motif, 15	-0.30215	0.106561
HSD17B2	L-Type	Hydroxysteroid (17-beta) dehydrogenase 2	-0.30199	0.106606
ST6GALNAC3	L-Type	In multiple clusters	-0.29999	0.108933
PLAU	L-Type	Plasminogen activator, urokinase	-0.29673	0.112905

[a] Type: correlated= significant negative correlation, L-shaped= not significant negative correlation, but likely to be regulated by promoter methylation.

The analysis of genes regulated by promoter methylation in primary colorectal tumors is complicated by the contamination of tumor samples with normal tissue. Tumor samples contain a significant proportion of normal cells from infiltrating lymphocytes, stromal cells, lymph/blood vessels, etc that can significantly interfere with the quantification of the levels of methylation and expression observed in tumor samples. However, to further investigate the regulatory effects of promoter methylation on gene expression we used a series of 222 primary colorectal tumors. Methylation and mRNA expression information was available for 12,719 genes in this series of primary tumors. As observed with the cell lines, there was a strong bias for negative correlations, as 3,494/12,719 (27.47 %) showed a significant negative correlation (p (BH)= <0.05) between methylation and mRNA levels and only 1,012/12,719 genes showed significant positive correlations (p (BH)= <0.05; 7.96 %; Binomial p<10E-30). Of the 667 genes identified as regulated by promoter methylation in the tumor

lines, expression and methylation data in the primary tumors was available from 659 genes. A negative correlation (p(BH) <0.05) was observed between expression and methylation for 424/659 genes (64.33 %; Appendix 5, Table 1), while only 9 of these 659 genes (1.37 %) had a significant positive correlation between mRNA and methylation levels. Therefore, despite the interference of the contamination with DNA derived from the normal tissue in primary tumor samples, evidence of epigenetic silencing could be observed for the majority (>64 %) of the genes initially identified in colorectal cancer cell lines.

4.2.4 Functional group enrichment analysis of genes regulated by CpG promoter methylation

To gain a better understanding of the biological function of the genes regulated by CpG methylation, we used functional group enrichment analysis. Of all the categories analyzed (see methods), we identified 17 functional groups significantly enriched (average fold enrichment >2) in the number of genes regulated by CpG promoter methylation (Fisher's Exact test, Benjamini correction, p<0.025; Table 14). Strikingly, all these overlapping categories are related to zinc finger proteins. The largest of these categories (Interpro IPR013087: Zinc finger, C2H2-type/integrase, DNA-binding) contains a total of 38 zinc finger domain proteins, including a large proportion of transcription factors (Table 14).

Table 14: Functional group enrichment analysis run with DAVID.

Category[1]	Term[2]	Count[3]	%[4]	PValue[5]	List Total[6]	Pop Hits[7]	Pop Total[8]	Fold Enrichment	Benjamini Adjusted *P*
UP_SEQ_FEATURE	zinc finger region:C2H2-type 10	28	4.27	8.70E-08	639	147	10929	3.26	7.70E-05
UP_SEQ_FEATURE	zinc finger region:C2H2-type 11	26	3.96	5.43E-08	639	126	10929	3.53	9.61E-05
UP_SEQ_FEATURE	zinc finger region:C2H2-type 9	29	4.42	4.72E-07	639	169	10929	2.93	2.78E-04
UP_SEQ_FEATURE	zinc finger region:C2H2-type 8	31	4.73	8.50E-07	639	194	10929	2.73	3.76E-04
UP_SEQ_FEATURE	zinc finger region:C2H2-type 4	38	5.79	2.05E-06	639	278	10929	2.34	6.06E-04
UP_SEQ_FEATURE	domain:KRAB	25	3.81	1.89E-06	639	141	10929	3.03	6.71E-04
UP_SEQ_FEATURE	zinc finger region:C2H2-type 6	33	5.03	3.23E-06	639	228	10929	2.48	8.17E-04
SMART	SM00349:KRAB	25	3.81	4.64E-06	332	151	5723	2.85	9.74E-04
UP_SEQ_FEATURE	zinc finger region:C2H2-type 7	31	4.73	5.02E-06	639	211	10929	2.51	1.11E-03
UP_SEQ_FEATURE	zinc finger region:C2H2-type 12	21	3.20	9.70E-06	639	115	10929	3.12	1.91E-03
UP_SEQ_FEATURE	zinc finger region:C2H2-type 5	34	5.18	1.35E-05	639	255	10929	2.28	2.38E-03
INTERPRO	IPR001909:Krueppel-associated box	25	3.81	4.48E-06	578	151	10095	2.89	4.63E-03
UP_SEQ_FEATURE	zinc finger region:C2H2-type 13	17	2.59	3.67E-05	639	87	10929	3.34	5.88E-03
UP_SEQ_FEATURE	zinc finger region:C2H2-type 2	37	5.64	5.10E-05	639	308	10929	2.05	7.49E-03
INTERPRO	IPR013087:Zinc finger, C2H2-type/integrase, DNA-binding	38	5.79	1.56E-05	578	310	10095	2.14	8.03E-03
UP_SEQ_FEATURE	zinc finger region:C2H2-type 3	37	5.64	8.73E-05	639	316	10929	2.00	1.18E-02
UP_SEQ_FEATURE	zinc finger region:C2H2-type 1	33	5.03	2.02E-04	639	280	10929	2.02	2.52E-02

[1]Category: Gene Ontology categories; [2]Term: Gene set name; [3]Count: number of genes associated with this gene set; [4]Percentag: gene associated with this gene set/total number of query genes, [5]p value: modified Fisher Exact p-value; [6]List Total: number of genes in the query list mapped to any gene set in this ontology; [7]Pop Hits: number of genes annotated to this gene set on the background list; [8]Pop Total: number of genes on the background list mapped to any gene set in this ontology.

4.2.5 CpG methylation outside of CpG islands can regulate gene expression

CpG islands are regions of increased density of CpG dinucleotides [184,185]. CpG methylation within CpG islands has long been known to regulate transcriptional activity. Surprisingly, it has recently been reported that CpG methylation in the 'shore' of these dense CpG island regions may be even more relevant for the transcriptional silencing of the associated promoter regions. However, the role of CpG methylation in regions not associated with CpG islands in transcriptional regulation remains unclear. Here we used genome-wide methylation and mRNA expression data on 30 cell lines to gain further insight into this question. Of the 667 candidate genes whose expression was regulated by CpG methylation, 489 (73.3 %) were associated with a CpG island as defined by classical Gardiner-Garden and Frommer [185] (UCSC Genome Browser, GRCh37/hq19; Appendix 5). However, the remaining 178 genes (26.7 %) with significant associations between mRNA and methylation levels, were not associated with a conventionally defined CpG island. Moreover, using less stringent criteria for defining CpG islands throughout the human genome using a Hidden Markov model-based approach [186], 123 genes (18.16 %) showing good correlations between expression and methylation levels in colorectal cancer cells, that were not associated with any CpG islands. Messenger RNA expression data before and after treatment with 5-Aza-2'-deoxycytidine in HCT116 colon cancer cells was available for 667 of the 667 genes identified as being epigenetically regulated. Of these, 260 genes showed >50 % methylation in the associated CpG dinucleotides investigated and 92 of these (35.38 %) had no CpG islands associated. Importantly, 71 of these 92 genes (77.17 %; Binomial test 5.8×10^{-8}) showed elevated expression levels after 5-Aza-2'-deoxycytidine treatment. Collectively these data indicate that CpG methylation outside CpG islands can regulate the transcriptional activity of a significant number of genes in colorectal cancer.

4.2.6 ZNF238 has tumor suppressor activity in colorectal cancer cells

Genetic and epigenetic events can silence gene expression. Tumor suppressor genes are known to be inactivated in tumors e.g. through mutation. We found 667 genes, whose expression is regulated by the methylation of the promoter. Furthermore, doing a functional group enrichment analysis of those genes, we found the largest of the categories contained 38 zinc finger domain proteins regulated by CpG promoter methylation. Here we hypothesize genes with high levels of methylation and resulting low expression are possible tumor suppressor genes. In order to test our hypothesis, we picked zinc finger protein 238 (ZNF238, also known as RP58 or ZBTB18) one of the zinc finger proteins from the above mentioned enrichment analysis and performed *in vitro* and *in vivo* studies.

We stably overexpressed the ZNF238-IRES-EGFP (from now on referred to as ZNF238) and its corresponding control IRES-EGFP (Empty vector =EV) in HCT116 cells. Parental HCT116 cells showed low levels of ZNF238 expression in mRNA microarray analysis, but high levels of methylation in MH27 microarray. The expression levels were checked by FACS for EGFP (Figure 35A) as well as qPCR using specific primers for ZNF238 (Figure 35B). Parental cells, as well as the HCT116-EV cells showed low endogenous ZNF238levels whereas the expression of ZNF238 was 90 times higher in HCT116-ZNF238 cells (Figure 35B).

Next, we used these generated cell lines to conduct some *in vitro* proliferation studies (Figure 35C-D). Reintroducing (overexpressing) HCT116-ZNF238 cells significantly reduced their growth compared to the EV (IRES-EGFP) when assessed the cell growth with and indirect method (Figure 35C) Furthermore, to validate these results using an independent technique, cell growth was measured directly by cell counting(Figure 35D).

To test our hypothesis *in vivo*, we used NOD/SCID mice and injected subcutaneous HCT116-EV cells in the left flank and HCT116-ZNF238 in the right flank. In good agreement with the *in vitro* findings, tumors with high levels of ZNF238 expression grew significantly slower than the corresponding EV (Figure 35E). Furthermore, HCT116-ZNF238 showed a lower tumor mass at the end of the experiment, compared to the HCT116-EV cells (Figure 35F). Collectively, these results demonstrate that ZNF238 has tumor suppressor activity in colorectal cancer.

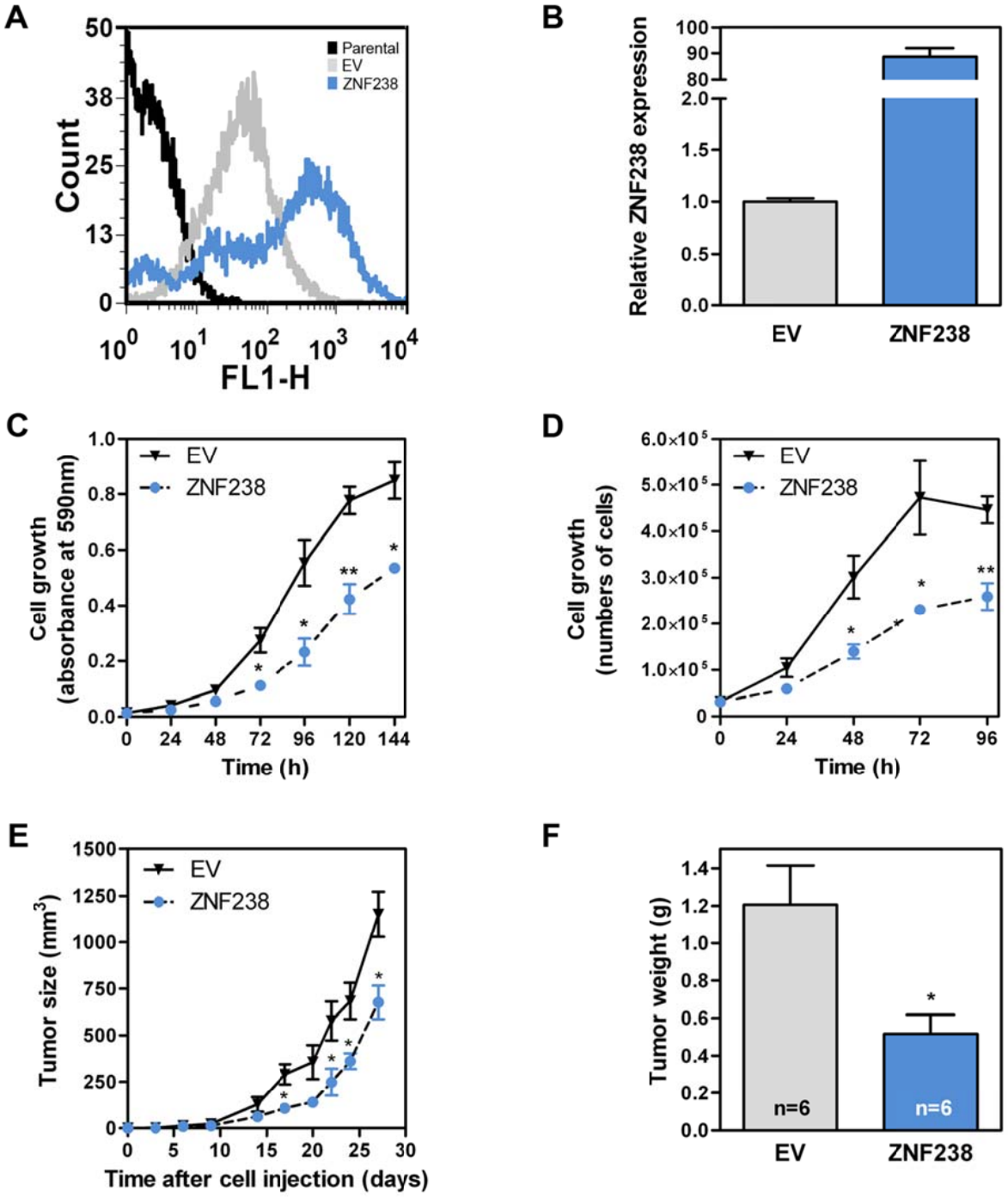

Figure 35: Role of ZNF238 on the growth of colorectal cancer cells.
A) Using flow cytometry, the overexpression of EGFP was detected in HCT116-EV and HCT116-ZNF238 cells compared to the parental cells. B) The expression was analyzed using qPCR with specific primers for ZNF238. C)To assess possible changes in cell growth, cells were seeded in seven 96-well plates and every 24h one plate was fixed and cells were then stained with SRB and the absorbance measured. D) cells were seeded in 24-well plates and every 24 h the cells were trypsinized and counted. Experiments were conducted three times in quadruplicates. E) EV and ZNF238 cells were implanted subcutaneously in immunodeficient NOD/SCID mice and the growth of the tumors was monitored over time.F) After euthanization of the animals tumors were resected and weighted. Animals n=6 per group. Student's T-test $p<0.05$.

5 Discussion

Colorectal cancer is a disease that is caused by genetic and epigenetic changes and patients with advanced colorectal cancer show an objective response rate of about only 30 % to current chemotherapeutic treatments. Inactivation of tumor suppressor genes and activation of oncogenes are key landmarks in tumor progression. The molecular events driving colorectal cancer are among the best characterized and it is now realized that tumor suppressor genes can be inactivated either genetically or epigenetically, thus contributing to tumor progression. However, new tumor suppressor genes and oncogenes are discovered every year, and the complete list of genes involved in colorectal cancer initiation and progression is far from complete. For example, we have recently described the important role of *MYO1A* [43] and *RHOA* [44] as genes with tumor suppressor activity in colon cancer. Finding new tumor suppressor genes will facilitate the understanding of the underlying molecular mechanisms in colorectal cancer. On the other hand, newly identified genes responsible for faster proliferation of a tumor, can be used as new therapeutic targets. This would widen the range of chemotherapeutic agents available for patients and potentially contribute to personalize the treatment for colorectal cancer.

While the genetic events that drive the tumorigenic process are relatively well characterized for colorectal cancer, the epigenetic events and their impact on the transcriptional reprogramming observed in colorectal tumors have not been as extensively investigated. Cytosine methylation in CpG dinucleotides is important for normal development and cell differentiation in higher organisms. These methylation marks lead to chromatin condensation and gene silencing and CpG methylation is inherited by the daughter cells after cell division. Aberrant methylation of CpG dinucleotides in the 5' regulatory regions of genes that are involved in the oncogenic process has emerged as an important mechanism leading to the initiation and/or progression of different tumor types, including tumors of the colon and rectum. However, the detailed transcriptional reprogramming resulting from aberrant methylation in colorectal tumors remains to be thoroughly characterized. The advent of high throughput technologies to investigate genome-wide levels of both mRNA expression and CpG methylation allows the systematic analysis of the genes regulated by CpG methylation throughout the genome of colorectal tumors. Although recent genome-wide studies have analyzed the genomic distribution of hypermethylated CpGs in a small number of colorectal tumors [95], a detailed analysis of the subset of these events that are important for gene expression regulation is currently lacking.

5.1 New lists of genes involved in the oncogenic process of colorectal cancer

5.1.1 Genes important for proliferation inhibition

Activation of oncogenes and inactivation of tumor suppressor genes eventually promotes proliferation and metastasis of cancer cells. Therefore, the comprehensive identification of oncogenes and tumor suppressor genes is important to understand the underlying molecular mechanisms involved in cancer.

Growth rates of colorectal tumors vary largely and faster proliferation is associated with poor patient prognosis [175–178]. In order to obtain a list of genes that are important for tumor proliferation we considered genes that are highly expressed in rapidly proliferating tumor cells. Additionally, because those genes are highly expressed in fast growing tumors, we considered those genes as good therapeutic targets. We identified 1290 genes, whose growth rates were significantly correlated with the expression levels in colorectal cancer cell lines. From those 1290 genes the majority, 966 genes, showed significant negative correlations. In other words, those genes showed higher expression levels in cell lines with faster growth (= low doubling time). The remaining 324 genes showed a significantly positive correlation, with lower expression in faster proliferating cells. In other words, those genes were highly expressed in slow growing colorectal cancer cells and therefore could be important for tumor growth inhibition.

Considering the fact that tumor suppressor genes are inactivated in tumors by e.g. mutation or methylation in the promoter region of a gene, we asked ourselves, which genes have high promoter methylation and show low expression in colorectal cancer cell lines. 667 candidate genes were identified whose expression was associated with the promoter methylation. Among those genes, the majority was negatively correlated, which means in genes with high levels of promoter methylation the genes were expressed at low levels. Therefore, those genes might be important for tumor growth inhibition.

Importantly, we found in the two above mentioned lists of genes, that might inhibit proliferation (with 324 and 667 genes, respectively), well studied tumor suppressor genes like *SMAD4* (Transforming growth factor-β), *MLH1* (DNA damage repair pathway), *RUNX1* (transcriptional regulation), and *CDH1* (Wnt/APC pathway) [41]. This strongly suggesting that new tumor suppressor genes could be identified with the methods described here.

This is further supported by the observation that a significant (Chi square p-value 0.02) overlap exists between the 667 genes with high methylation associated with low expression levels, and the 324 genes highly expressed in cancer cells with slower proliferation rates.

5.1.2 Genes important for proliferation

Genes important to sustain rapid proliferation of a tumor can be used as new therapeutic targets. There are currently a limited number of chemotherapeutic agents approved for their routine use in the fight against colorectal cancer, namely the antimetabolite 5-fluorouracil (5-FU), the platinum compound oxaliplatin and the topoisomerase I inhibitor irinotecan, in addition to the targeted agents cetuximab, panitumumab, bevacizumab and regorafenib. However, the response rate to each of these drugs used as single agents is below 30 % and the identification of novel therapeutic targets and the subsequent development of new chemotherapeutic agents would likely improve the survival of these patients. Here we hypothesized that inhibition of genes highly expressed in rapidly proliferating colorectal cancer cells can interfere with tumor growth, and these genes are therefore good candidate chemotherapeutic targets. In support of this hypothesis, we provide here a list of 966 genes that have significantly higher expression in the tumor cell lines with higher proliferation rates. Importantly, we found that the direct target of 5-FU, the gold standard agent for the treatment of colorectal cancer patients for over five decades [187], thymidylate synthetase, was among the genes showing a highly significant correlation between its expression level and the rate of tumor cell growth.

As a proof of concept, we selected two additional genes with high expression levels in rapidly proliferating colorectal cancer cells: Protoporphyrinogen oxidase (PPOX), which catalyzes the 6-electron oxidation of protoporphyrinogen IX to form protoporphyrin IX, the penultimate reaction of heme biosynthesis and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), which catalyzes the sixth step of glycolysis, the conversion of glyceraldehyde 3-phosphate to D-glycerate 1,3-bisphosphate. Those two enzymes have known specific, chemical inhibitors, which are commercially available and were therefore selected for further study. Importantly, the two genes were not the ones with the most significant correlations, because for the genes from the top of the list, no chemical inhibitors were commercially available. However, genetic inactivation of selected genes with significant associations between expression and tumor growth could be used to identify the most promising therapeutic targets, for which novel specific inhibitors could then be developed. Na iodoacetate and oxadiazon inhibit GAPDH and acifluorfen and CGP 3466B maleate inhibit PPOX. We found that the clinically used TYMS inhibitor 5-FU as well as GAPDH and PPOX inhibitors, significantly reduced the growth of colon cancer cells at micromolar concentrations. Moreover, using a preclinical subcutaneous xenograft model, we could demonstrate that at least the PPOX inhibitor acifluorfen

was able to inhibit the growth of colon cancer cell lines (3 of 7; 42.3 %). The response rate observed for acifluorfen is comparable with the response observed clinically with established chemotherapeutic agents. Patients respond to 5-FU mono-treatment with about 15 % [188,189] and in combination with folinic acid about 30 % [190] .

Heme plays critical roles in multiple processes involving oxygen metabolism. This includes proteins that transport or store oxygen such as hemoglobin and myoglobin, but is also important in mitochondrial respiratory chain complexes, in cytochrome P450s and in other enzymes that use or detoxify oxygen such as peroxidases and catalases [191]. Our findings are consistent with the observation that inhibition of heme synthesis significantly reduced proliferation in lung cancer cells [192].

It has long been known that most cancer cells predominantly produce energy by a high rate of glycolysis and lactate production, an observation known as the Warburg effect [193]. Therefore, it has been suggested before that differences in the metabolisms of tumor cells could offer a therapeutic window [194,195]. Recently, GAPDH has been shown to translocate to the nucleus and to be implicated in several non-metabolic processes, including transcriptional regulation and apoptosis [196,197]. We show here that inhibition of GAPDH efficiently reduces the growth of colon cancer cells *in vitro*. Interestingly, although the GAPDH inhibitor Na iodoacetate has been shown to reduce the growth of Ehrlich ascites carcinoma (EAC) cells and xenografts of colon cancer cells at doses similar or lower than the one used *in vivo* in this study [198,199], no significant effects were observed here on the growth of subcutaneous xenografts of four different colon cancer cell lines. However, no toxicity was observed at the doses used, and based on the *in vitro* effects observed, it remains possible that Na iodoacetate treatment at higher doses and/or in other tumor cell lines, may interfere with tumor growth.

5.2 Cell lines as a tool to investigate genes involved in proliferation and epigenetic silencing in colorectal cancer

In this study we have used colorectal cancer cell lines to identify possible genes important for rapid proliferation and inhibition colorectal cancer growth. The use of cell lines has a number of advantages when compared to the use of primary tumor material as the starting point for these high throughput analyses. Cell lines grown *in vitro* are not contaminated with non-tumor cells such as blood/lymph vessels, infiltrating immune cells or stromal cells, which are normally present in primary

tumor samples at varying proportions that could account for as much as 50 % of the cells and significantly complicate the interpretations of the results [200,201]. In addition, cell lines provide an unlimited supply of material that is widely available, they are easy to propagate, and are therefore amenable for high-throughput assays. Moreover, confirmation of the findings can be easily achieved by manipulation of the same tumor cells used in the screening analysis.

Regarding the use of cell lines to investigate the molecular mechanisms regulating proliferation of colorectal tumors, we showed clearly that the proliferation rate in cell lines *in vitro* correlated with their growth *in vivo* when injected subcuaniously in immunodeficient NOD/SCID mice. Furthermore, primary colorectal cancer tumors also showed a significant correlation between the mitotic rate and the expression of *PPOX* or *GAPDH*, further indicating that the proliferation of cancer cells *in vitro* is a good surrogate for their proliferation *in vivo*.

In primary colorectal tumors, an association has been reported between high grade (poorly differentiated) tumors [202,203] or microsatellite instability [204] and faster proliferation rates. Consistently, here we show that cell lines that form high-grade tumors when grown as xenografts or have microsatellite instability, proliferate significantly faster than cell lines forming low-grade (differentiated) tumors or microsatellite stable lines. These results indicate that the proliferative profile of the cell line panel used here closely recapitulates the characteristics of primary colorectal tumors [163], and collectively suggests that cell lines are a suitable model for the investigation of this disease and a good tool to identify genes important to sustain rapid proliferation of colorectal cancer cells and possible new therapeutic targets.

Work on ovarian and colorectal cancer cell lines showed that it is important to use the correct cell lines and the adequate number of cell lines for an *in vitro* study [200]. In ovarian cancer e.g. the two most frequently used cell lines for *in vitro* studies are not very representative cell lines. They have wild-type TP53 and uncharacteristic mutations [200,201]. Therefore, it is of great importance to use a large panel of single tissue type cell lines, which covers and represents the heterogeneity of primary tumors [200,201]. Large, annotated cell line collections may help to discover anticancer agents, biomarker, and can be used in functional and mechanistic testing, and high-throughput screening for drug discovery among other applications [202,205]. In addition, the study of Mouradov et al using 70 colorectal cancer cell lines and The Cancer Genome Atlas (TCGA) tumor samples, concluded that cell lines represent the three main subtypes of colorectal cancer tumors, namely non-hypermutated, hypermutated with MSI, and hypermutated without MSI [163]. Moreover, Ahmed et al thoroughly examined 24 colorectal cancer cell lines for their genetic features (*BRAF*, *KRAS*, *PIK3CA* and *PTEN*

mutation status) as well as their methylation status and concluded that this selection of cell lines can be helpful for further *in vitro* models used for descriptive and functional research [206].
Here a panel of 55 colorectal cancer cell lines was used, of which we evaluated growth rates in 52 of them and obtained expression patterns in 32 of them and promoter methylation data in 45 of them. We used this large number of cell lines to better capture the heterogeneity observed in primary colorectal tumors. This should accurately capture the broad range of gene expression, methylation and proliferation rates that tumors also present.

On the other hand, cell lines have some disadvantages compared to primary tumors: it is a known problem for several decades, that cell lines are misidentified and/or cross-contaminated [207–209]. This problem among cell lines bears the risk of drawn false conclusions in the experiments. This problem can be solved nowadays by using standardized protocols for authentication of human cell lines [207]. Furthermore, most cell lines have grown over decades in monolayers as opposed to three-dimensional cultures, high oxygen tension (21 %, whereas physiological conditions are 2-5 %) and controlled growth media (e.g. no secreted growth factors from neighboring cell types). These reasons could cause that cell lines not always reflect the primary tumor entity [208].

When comparing cell lines and primary tumors, globally similar genetic alterations, including genome-wide mutations, DNA copy number, and driver gene profiles are detected [163]. Furthermore, gene expression profiles of colorectal cancer cell lines showed a broad representation of the situation of primary tumors [210]. Other studies show however the opposite, namely promoter methylation is increased in tumor-derived cell lines when compared to primary tumors or xenografts in various types of cancer [211–213] and in hepatocellular carcinoma, microarray profiling showed distinct differences between primary tumors and commonly used preclinical models like cell lines [214].
It also has been shown that when comparing cancer cell lines (from labs owing a huge panel of colorectal cancer lines like the Bodmer lab) with primary tumors (from TCGA or COSMIC), the mutation frequency of those most frequently mutated genes in colorectal cancer is higher in cell lines compared to the primary tumors[200]. The reason for this discrepancy between the mutation landscape in cell lines and primary tumors depends on the source of the data, due to overrepresentation of mismatch repair deficiency (and thus the mutations highly associated with this hypermutated subtype) in the cell line panels compared with primary cancers [200]. Another possibility for this is that at the time most of the cell lines currently in use were isolated, the chance of obtaining a cell line from a tumor was influenced by its genetic and epigenetic make-up[200].

Whether primary tumors or cell lines derived from tumors are the better tool, remains unclear, and depends probably mostly on the type of study, including the tumor type and molecular event investigated.

5.3 Frequently mutated genes and associations with the overall methylation and growth rates colorectal cell lines

Cancer is a multistep process that arises over the course of several years, resulting from the accumulations of mutations and clonal selection that lead to gain of function of oncogenes and top the loss of function of tumor suppressor genes. However, in our study we found that the global mutational status of the genes most frequently mutated in colorectal tumors (*APC*, *TP53*, *KRAS*, *BRAF*, *PIK3CA*, *SMAD4*, *TCF7L2* and *CTNNB1*) were not associated with the growth rates of colon cancer cell lines. This suggests that these common genetic changes, when considered individually, do not have a significant effect on the proliferation rates of colon cancer cells.

However, the overall levels of methylation in cell lines were significantly higher in MSI lines compared to MSS lines, and the average levels of methylation were significantly lower in the cell lines with *APC* mutations.

It is known that *BRAF* mutation are tightly associated with the 'CpG island methylator phenotype' (CIMP) characterized by aberrant hypermethylation of many genes, including the mismatch repair gene *MLH1*[181,215]. Sporadic colorectal cancer with MSI characteristically includes the absence of significant familial clustering, biallelic methylation of the MLH1 promoter; absence of MLH1 and PMS2 proteins; and frequent mutation (usually V600E) in *BRAF* [180]. *BRAF* mutations, detected in sporadic but not familial colorectal cancers with MSI, are associated with reduced mortality[180]. However, in our study *BRAF* mutational status is not associated with methylation levels. One reason for this discrepancy could be that we have used cell lines and not primary tumors, since the cell lines do not always represent the situation of a primary tumor. Moreover, as previously reported, CIMP+ cases were found to include all the *BRAF* mutant lines as well as the majority (9/17; 53 %) of MSI lines [96].

Moreover, there was a significant association between higher methylation levels and faster growth suggesting that higher levels of promoter CpG methylation may contribute to the uncontrolled proliferation characteristic of tumor cells by inhibiting tumor suppressor genes.

5.4 Growth rates are not exclusively dependent on genes regulating the cell cycle

Pioneer studies performed genome-wide transcriptional profiling of synchronized proliferating human fibroblasts or HeLa cells [216,217]. The findings of these studies identified genes that are expressed in a periodical manner during the cell cycle. These studies identified a group of genes involved in cell motility and remodeling of the extracellular matrix was mainly expressed during M phase, indicating a mechanism for tight regulation of proliferation and invasive cellular behavior [216]. Other genes are involved in well know processes during cell cycle like DNA replication, chromosome segregation, and cell adhesion [217]. It is important to highlight, however, that this periodical waves of gene expression observed as cells progress through the cell cycle take place in all dividing cells, regardless of the proliferation rate they exhibit and limited progress has been made in the identification of genes with differential expression patterns in tumors cells with high and low proliferation rates [218,219].

In this regard, previous studies using the NCI60 set, that is composed of 60 cell lines representing nine different cancer types, including colon represented by seven cell lines, found that the growth rate of cell lines is closely related with genes involved in the cell cycle, and RNA and protein synthesis [218–220]. Our present study, carried out on a larger set of colorectal cancer cell lines (n= 31), confirmed these earlier findings. Furthermore, using the cell lines of the NCI data set, it has been shown that genes involved in cholesterol metabolism, iron metabolism and fatty acid metabolism were negatively correlated with the growth rates of colon cell lines [218]. However, we could not confirm these finding in our study. Instead, we found additional groups of functionally related genes significantly correlated with the growth rates of colorectal cancer cells including several categories related with protein metabolism, such as translation, protein transport and cellular protein catabolic process. The reason for this discrepancy might lay in the number of cell lines used for the study. As previously mentioned, it is of great importance to choose a set of cell lines, large enough to give statistical power and probably even more important, fully capture the heterogeneity of primary tumors.

5.5 CIMP phenotype of colorectal cancer cell lines

Recently, there has been some controversy regarding the existence of a new tumorigenic mechanism resulting from the widespread hypermethylation of CpG islands (CIMP+ tumors) [182,183]. Although the

important role of CpG methylation in tumor progression is well documented, the existence of the CpG methylator phenotype (CIMP) critically depends on the existence of a bimodal distribution in the levels of methylation, rather than a continuous gradient in the levels of methylation in colorectal tumors. The quantitative analysis of >27,000 CpGs in a panel of 45 cell lines derived from colorectal tumors allows the genome-wide analysis of the levels of CpG methylation in large sample sets. Moreover, unlike studies with primary tumors, the interpretation of the results of methylation analysis of cell lines is not complicated by the presence of a variable proportion of 'contaminating' normal cells (stromal, infiltrating lymphocytes, blood/lymph vessels, etc). Although using the methylation status of all >27,000 CpGs interrogated by the HumanMethylation27 chips, unsupervised cluster analysis of the 45 cell lines in this study identified a group of cell lines with higher methylation levels, the distribution of cell lines with high/low methylation was not bimodal. However, the initial definition of the CpG methylator phenotype was based in the methylation status of cancer-related promoters that are not methylated in normal colonic mucosa samples (Type C promoters) [96]. Using 643 Type C promoters that do not show methylation in a series of 22 normal colonic samples (GSE17648), or the 5 CIMP markers that have recently been used to define the CIMP phenotype, not only clustering analysis identified an extended set of samples showing significantly higher methylation levels, but also the distribution of cell lines with high/low methylation was bimodal. Although the molecular mechanisms underlying this tumorigenic pathway remain to be determined, these results support the existence of a type of colorectal tumors driven by the higher incidence of promoter methylation. In addition, cell lines with higher methylation levels showed significantly faster growth than cell lines with lower levels of CpG methylation. Although from this analysis it is not possible to establish a causal effect, this is consistent with the observation that several genes that inhibit cell cycle progression (such as, *RPRM*, *PLAGL1* and *p16*) are frequently methylated and their expression is regulated by these epigenetic defects.

5.6 Gene expression regulation through methylation in colorectal cancer

5.6.1 Genes regulated by promoter methylation

Although CpG promoter hypermethylation was found throughout the genome of colorectal cancer cell lines, only the expression of a subset of the genes showing frequent promoter methylation were shown to be regulated through aberrant methylation. This study describes details of the genome-wide transcriptional reprogramming resulting from aberrant methylation in colorectal tumors at the level of individual genes. We found that the expression of 667 of the 11,858 genes investigated (5,6

%) was silenced by promoter methylation. These genes included important tumor suppressor genes known to be frequently methylated in colorectal tumors such as E-cadherin (CDH1) [221–223], CA9 (Carbonic anhydrase IX) [224–226], the serine/threonine protein kinase DAPK2 [227] and mismatch repair gene MLH1 [228].

In addition, other important tumor suppressor genes and oncogenes that have not previously been reported to be methylated in colorectal tumors were also identified. These included *F11R* (junctional adhesion molecule A) and *CLD1* (Claudin 1), two important regulators of tight junction assembly in epithelial sheets [229–231], the proapoptotic Bcl2-family member *BOK* (BCL2-related ovarian killer) [232,233] and the Ephrin ligand EFNB2 (*Ephrin B2*) [234,235].

5.6.2 CpG methylation outside CpG islands can regulate gene expression

The role of cytosine methylation in the context of CpG islands in the regulation of gene expression has been known for a long time. Although 45% of all human gene promoters do lie within a CpG island, little is known about their regulation and potential transcriptional control mechanism [236]. A recent study has revealed that a subset of the human methylome is highly stable across various cell types, including cancer cells [237]. The majority of these ultrastable regions are always unmethylated and flanked by regions with high methylation. Those regions, called ravines, are suggested to encompass CpG islands and flanking regions. This methylation pattern of these ravines are associated with a particularly high expression levels of genes contained I this region [237]. Studies showed that genes having no CpG island in the promoter region are regulated by methylation: DNA methylation at CpG poor *RUNX3* promoter 1 and *LAMB3* promoter can lead to silencing of those genes [236] . A study in chronic myelomonocytic leukemia showed that leukemia-associated TET2 mutations have a minor effect on DNA methylation throughout the human genome, and preferentially result in hypermethylation at selected non-CpG island sites that are enriched at transcription factor-binding sites and enhancers [236].

Here, we found that 24 of the 667 genes that are regulated by CpG methylation are not associated with a CpG island. These included multiple genes known to be important during colorectal cancer progression such as *MYO1A*, a brush border myosin that we have recently shown to have important tumor suppressor effects in this organ [43,238]. Other genes known to be important in the oncogenic process are Cadherin 17 [239,240], selenium binding protein 1 [241], *TNF*, BCL2-like 14 (apoptosis facilitator).

5.7 Zinc finger proteins are regulated by methylation in colorectal cancer tumors

In our study a significant enrichment of zinc finger proteins was observed in the number of gene associated with the methylation in colorectal tumors. 38 zinc finger domain proteins were identified, of which a larger number of them were transcription factors and likely to be important in amplifying the transcriptional reprogramming imposed by CpG methylation in colorectal cancer cells. We hypothesized that at least some of the genes having high promoter methylation and low expression might exhibit tumor suppressor activity. To investigate this possibility we selected *ZNF238* (also known as *RP58* or *ZBTB18*) one of the genes whose methylation was negatively correlated with the levels of expression.

Zinc finger proteins are a large family of metalloproteins that use zinc for as a structural stabilizer of the domain [242]. Zinc finger proteins are structurally diverse and involved in various cellular processes, such as replication and repair, transcription and translation regulation, metabolism and signaling, cell proliferation and apoptosis. They typically bind to a wide variety of compounds, such as nucleic acids, proteins and small molecules [243].

Zinc finger proteins are deregulated in several types of cancer. ZPO2 is involved in mammary gland homeostasis and deregulation of ZPO2 may promote breast cancer development [244]. Gα12 overexpressed in hepatocellular carcinoma causes zinc finger E-box binding homeobox 1 (ZEB1) induction by deregulating TP53-responsive miRNAs, which facilitates epithelial-mesenchymal transition and growth in liver tumors [245]. LYAR, a cell growth-regulating zinc finger protein, is associated with cytoplasmic ribosomes in male germ and cancer cells [246]. Human melanomagenesis as loss of zinc finger E-box binding homeobox 2 (ZEB2) expression is associated with reduced patient survival [247]. Furthermore, it has been shown that some zinc finger proteins have tumor suppressor activity, including *ZNF382* in nasopharyngeal, esophageal, colon, gastric, and breast cancer [248], *ZNF331* gastric carcinogenesis [249] and, *ZNF545* breast cancer [250].

ZNF238 is a C2H2-type zinc finger protein, that acts at the molecular level as a transcriptional repressor and is involved in chromatin assembly [251,252]. Structurally, at the N-terminus ZNF238 presents a BTB/POZ-ZF [Broad complex, Tramtrack, Bric à brac (BTB) or poxvirus and zing finger (POZ)-zinc finger] domain, which is responsible for DNA binding. This domain is highly conserved between human, mice, and zebrafish. At the COOH-terminal part of the protein four zinc fingers are located [253]. In order to avoid any possible steric hindrance of the folding and DNA-binding process we decided not do directly add a GFP moiety in frame with ZNF238. Instead, we created a construct

using and IRES-EGFP after the ZNF238 coding sequence. The IRES (internal ribosome entry site) serves as a translation initiator after the ZNF238 and so ZNF238 and EGFP are expressed as two individual proteins under the regulation of the same promoter.

ZNF238 has been shown to play a role in various developmental processes such as cerebral cortex development [254] and myogenesis [255]. DNA methylation is known to play an important role during myogeneic differenciation. Using a human myoblast differentiation model, global DNA methylation level increased and the methylation patterns looked more different from those of mesenchymal stem cells. *De novo* methylation of non-CpG island promoters of muscle-related genes like ID4 and ZNF238 was more often associated with transcriptional down-regulation than that of CpG island promoters [256]. However the role of ZNF238 in colorectal cancer has not been investigted.

In order to study the effects of overexpressing ZNF238 in a colorectal cancer cell line, we used HCT116 colon cancer cells, which have low endogenous ZNF238 mRNA levels, and high levels of methylation of its promoter. We found that restoration of ZNF238 expression significantly reduced the growth of HCT116 both *in vitro* and when implanted subcutaneously in immunodeficient NOD/SCID mice. This is consistent with a tumor suppressor activity of ZNF238 in colorectal cancer. Consistently, similar results have been demonstrated in brain tumor, where reinstating ZNF238 in medulloblastoma and glioblastoma multiforme cells decreased cell proliferation and promoted cell death which lead to the conclusion that ZNF238 is a novel brain tumor suppressor [253].

Taken together, these results strongly suggest that the list of 667 genes is enriched in genes with tumor suppressor activity and that the picture of the epigenetically regulated genes which contributes to the oncogenic process is incomplete.

6 Conclusions

Colorectal cancer is a disease caused by genetic and epigenetic alterations. Inactivation of tumor suppressor genes and activation of oncogenes are key landmarks in tumor progression. However, the list of tumor suppressor genes and oncogenes is far from complete, even in the case of the tumor types that are best characterized, such as colorectal cancer. Because patients with advanced colorectal cancer are treated with chemotherapy, and only about 30 % of the patients show an objective response, it is of importance to improve the clinical management of these patients. In this study genome-wide high throughput assays were used to better characterize important aspects of the oncogenic progression such as deregulation of proliferation and aberrant expression caused by epigenetic mechanisms.

Therefore, the main conclusions of this study are:

- Groups of functionally related genes were found, whose growth rates were significantly correlated with the expression levels in colorectal cancer cells, including protein metabolism, such as translation, protein transport and cellular protein catabolic process.
- PPOX and GAPDH were identified as candidate therapeutic targets and as a proof of concept, acifluorfen was demonstrated to inhibit the growth of colorectal cancer cells *in vitro* and *in vivo*, which identified PPOX as a novel candidate chemotherapeutic target for the treatment of colorectal cancer.
- 667 genes showing significant associations between DNA methylation and expression levels during tumorigenesis in colorectal cancer and an enrichment of zinc finger proteins was identified among those genes.
- ZNF238, a zinc finger protein whose methylation is negatively correlated with its levels of expression, was shown to inhibit tumor growth *in vitro* and *in vivo*.

These results contribute to the identification of novel chemotherapeutic targets for patients with colorectal cancer and a better understanding of the characterization colorectal cancer.

7 Bibliography

1. Moore, K. & Dalley, A. *Clinically oriented anatomy*. (Lippincott-Williams&Wilkins, 2006).
2. Standring, S. *et al. Gray's anatomy: The Anatomical Basis of Clinical Practice*. (Churchill Livingstone, 2009).
3. Young, B., Lowe, J. ., Stevens, A. & Health, J. . *Wheather's Functional Histology*. (Elsevier, New York, 2006).
4. Lee. Digestion.
5. Deplancke, B. & Gaskins, H. R. Microbial modulation of innate defense: Goblet cells and the intestinal mucus layer. *Am. J. Clin. Nutr.* **73,** (2001).
6. Kim, Y. S. & Ho, S. B. Intestinal goblet cells and mucins in health and disease: Recent insights and progress. *Curr. Gastroenterol. Rep.* **12,** 319–330 (2010).
7. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).
8. Sternini, C., Anselmi, L. & Rozengurt, E. Enteroendocrine cells: a site of 'taste' in gastrointestinal chemosensing. *Curr. Opin. Endocrinol. Diabetes. Obes.* **15,** 73–78 (2008).
9. Moran, G. W., Leslie, F. C., Levison, S. E., Worthington, J. & McLaughlin, J. T. Enteroendocrine cells: neglected players in gastrointestinal disorders? *Therap. Adv. Gastroenterol.* **1,** 51–60 (2008).
10. Clevers, H. & Batlle, E. SnapShot: The intestinal crypt. *Cell* **152,** 1198–1198.e2 (2013).
11. Clevers, H. C. & Bevins, C. L. Paneth cells: maestros of the small intestinal crypts. *Annu. Rev. Physiol.* **75,** 289–311 (2013).
12. Bastide, P. *et al.* Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *J. Cell Biol.* **178,** 635–648 (2007).
13. Crosnier, C., Stamataki, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat. Rev. Genet.* **7,** 349–359 (2006).
14. Pinto, D. & Clevers, H. Wnt, stem cells and cancer in the intestine. *Biol. Cell* **97,** 185–196 (2005).
15. Baker, A. M. *et al.* Quantification of Crypt and Stem Cell Evolution in the Normal and Neoplastic Human Colon. *Cell Rep.* **8,** 940–947 (2014).
16. Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nat. Genet.* **40,** 915–920 (2008).
17. Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* **31,** 3079–3091 (2012).
18. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449,** 1003–1007 (2007).

19. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459,** 262–265 (2009).

20. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143,** 134–144 (2010).

21. Pellegrinet, L. *et al.* Dll1- and Dll4-mediated Notch signaling is required for homeostasis of intestinal stem cells. *Gastroenterology* **140,** 1230–1240 (2012).

22. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281,** 1509–1512 (1998).

23. Tetsu, O. & McCormick, F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398,** 422–426 (1999).

24. Wielenga, V. J. *et al.* Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am. J. Pathol.* **154,** 515–523 (1999).

25. De Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476,** 293–297 (2011).

26. Koo, B.-K. *et al.* Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* **488,** 665–669 (2012).

27. Jensen, J. *et al.* Control of endodermal endocrine development by Hes-1. *Nat. Genet.* **24,** 36–44 (2000).

28. Yang, Q., Bermingham, N. a, Finegold, M. J. & Zoghbi, H. Y. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* **294,** 2155–2158 (2001).

29. Fortini, M. E. Notch Signaling: The Core Pathway and Its Posttranslational Regulation. *Dev. Cell* **16,** 633–647 (2009).

30. Madison, B. B. *et al.* Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* **132,** 279–289 (2005).

31. Haramis, A.-P. G. *et al.* De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* **303,** 1684–1686 (2004).

32. He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat. Genet.* **36,** 1117–21 (2004).

33. Stewart, B. W., Wild, C. P. *World Cancer Report 2014*. (IARC Nonserial Publication, 2014).

34. De Martel, C. *et al.* Global burden of cancers attributable to infections in 2008: A review and synthetic analysis. *Lancet Oncol.* **13,** 607–615 (2012).

35. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *Eur. J. Cancer* **49,** 1374–1403 (2013).

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

37. Hanahan, D. & Weinberg, R. a. Hallmarks of cancer: The next generation. *Cell* **144,** 646–674 (2011).

38. Pecorino, L. *Molecular Biology of Cancer: Mechanisms, Targets, and Therapeutics*. (Oxford University Press Inc., 2006).

39. UniProt. UniProtKB results. (2015). at <http://www.uniprot.org/uniprot/>

40. Lodish, H. *et al.* in *Mol. Cell Biol.* (2000).

41. Morris, L. G. T. & Chan, T. a. Therapeutic targeting of tumor suppressor genes. *Cancer* **121,** 1357–1368 (2015).

42. Chial, H. Proto-oncogenes to oncogenes to cancer. *Nat. Educ.* **1,** 33 (2008).

43. Mazzolini, R. *et al.* Brush border Myosin Ia has tumor suppressor activity in the intestine. *Proc Natl Acad Sci U S A* **109,** 1530–1535 (2012).

44. Rodrigues, P. *et al.* RHOA inactivation enhances Wnt signalling and promotes colorectal cancer. *Nat. Commun.* **5,** 5458 (2014).

45. Hammoud, S. S., Cairns, B. R. & Jones, D. a. Epigenetic regulation of colon cancer and intestinal stem cells. *Curr. Opin. Cell Biol.* **25,** 177–183 (2013).

46. Jenuwein, T. The epigenetic magic of histone lysine methylation. *FEBS J.* **273,** 3121–3135 (2006).

47. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).

48. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3,** 415–428 (2002).

49. Klose, R. J. & Bird, A. P. Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31,** 89–97 (2006).

50. Jones, P. a & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10,** 805–811 (2009).

51. Gama-sosa, M. a. *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11,** 6883–6894 (1983).

52. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301,** 89–92 (1983).

53. Sharrard, R. M., Royds, J. a, Rogers, S. & Shorthouse, a J. Patterns of methylation of the c-myc gene in human colorectal cancer progression. *Br. J. Cancer* **65,** 667–672 (1992).

54. Cheah, M. S., Wallace, C. D. & Hoffman, R. M. Hypomethylation of DNA in human cancer cells: a site-specific change in the c-myc oncogene. *J. Natl. Cancer Inst.* **73,** 1057–65 (1984).

55. Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. *Nature* **395,** 89–93 (1998).

56. Ting, A. H., McGarvey, K. M. & Baylin, S. B. The cancer epigenome--components and functional correlates. *Genes Dev.* **20,** 3215–3231 (2006).

57. Esteller, M. Epigenetic gene silencing in cancer: The DNA hypermethylome. *Hum. Mol. Genet.* **16,** 50–59 (2007).

58. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21,** 5427–5440 (2002).

59. Masliah-Planchon, J., Bièche, I., Guinebretière, J.-M., Bourdeaut, F. & Delattre, O. SWI/SNF Chromatin Remodeling and Human Malignancies. *Annu. Rev. Pathol.* 145–171 (2014). doi:10.1146/annurev-pathol-012414-040445

60. Wolffe, a P. Chromatin remodeling: why it is important in cancer. *Oncogene* **20,** 2988–2990 (2001).

61. Shih, I. M. *et al.* Top-down morphogenesis of colorectal tumors. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2640–2645 (2001).

62. Preston, S. L. *et al.* Bottom-up histogenesis of colorectal adenomas: Origin in the monocryptal adenoma and initial expansion by crypt fission. *Cancer Res.* **63,** 3819–3825 (2003).

63. Medema, J. P. & Vermeulen, L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. *Nature* **474,** 318–326 (2011).

64. Humphries, A. & Wright, N. a. Colonic crypt organization and tumorigenesis. *Nat. Rev. Cancer* **8,** 415–424 (2008).

65. Markowitz, S. D., Dawson, D. M., Willis, J. & Willson, J. K. Focus on colon cancer. *Cancer Cell* **1,** 233–236 (2002).

66. Schoen, R. E. The case for population-based screening for colorectal cancer. *Nat Rev Cancer* **2,** 65–70 (2002).

67. McDonald, S., Preston, S., Lovell, M., Wright, N. & Jankowski, J. Mechanisms of Disease: from stem cells to colorectal cancer. *Nat Clin Pr. Gastroenterol Hepatol* **5,** 267–275 (2006).

68. Compton, C. C. Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod. Pathol.* **16,** 376–88 (2003).

69. Egner, J. R. AJCC Cancer Staging Manual. *JAMA J. Am. Med. Assoc.* **304,** 1726 (2010).

70. American Cancer Society. What are the survival rates for colorectal cancer by stage? at <http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-survival-rates>

71. Haggar, F. a, Boushey, R. P. & Ph, D. Colorectal Cancer Epidemiology : Incidence , Mortality , Survival , and Risk Factors. *Clin. Colon Rectal Surg.* **6,** 191–197 (2009).

72. Jemal, A. *et al.* Annual report to the nation on the status of cancer, 1975-2001, with a special feature regarding survival. *Cancer* **101,** 3–27 (2004).

73. Jemal, A., Bray, F. & Ferlay, J. Global Cancer Statistics. *CA Cancer J Clin* **61,** 69–90 (2011).

74. Walther, A. *et al.* Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer* **9,** 489–499 (2009).

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

76. Pancione, M., Remo, A. & Colantuoni, V. Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. *Patholog. Res. Int.* **2012,** (2012).

77. Jacob, S. & Praz, F. DNA mismatch repair defects: Role in colorectal carcinogenesis. *Biochimie* **84,** 27–47 (2002).

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

79. Issa, J. P. Colon cancer: It's CINor CIMP. *Clin. Cancer Res.* **14,** 5939–5940 (2008).

80. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18,** 85–98 (2008).

81. Markowitz, S. D. & Bertagnolli, M. M. Molecular basis of colorectal cancer. *N. Engl. J. Med.* **361,** 2449–2460 (2009).

82. Wang, H. & Hays, J. B. Human DNA mismatch repair: Coupling of mismatch recognition to strand-specific excision. *Nucleic Acids Res.* **35,** 6727–6739 (2007).

83. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363,** 558–561 (1993).

84. Aaltonen, L. A. *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science (80-.).* **260,** 812–816 (1993).

85. Kuismanen, S. a, Holmberg, M. T., Salovaara, R., de la Chapelle, a & Peltomäki, P. Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. *Am. J. Pathol.* **156,** 1773–1779 (2000).

86. Vilkki, S. *et al.* Screening for microsatellite instability target genes in colorectal cancers. *J. Med. Genet.* **39,** 785–789 (2002).

87. Parsons, R. *et al.* Microsatellite Instability and Mutations of the Transforming Growth Factor beta Type II Receptor Gene in Colorectal Cancer. *Cancer Res.* 5548–5550 (1995).

88. Souza, R. F. *et al.* Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nat. Genet.* **14,** 255–7 (1996).

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

90. Alazzouzi, H. *et al.* Mechanisms of inactivation of the receptor tyrosine kinase EPHB2 in colorectal tumors. *Cancer Res* **65,** 10170–10173 (2005).

91. Malkhosyan, S., Rampino, N., Yamamoto, H. & Perucho, M. Frameshift mutator mutations. *Nature* **8,** 499–500 (1996).

92. Trautmann, K. *et al.* Chromosomal instability in microsatellite-unstable and stable colon cancer. *Clin. Cancer Res.* **12,** 6379–6385 (2006).

93. Ogino, S. *et al.* Prognostic significance and molecular associations of 18q loss of heterozygosity: A cohort study of microsatellite stable colorectal cancers. *J. Clin. Oncol.* **27,** 4591–4598 (2009).

94. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer—the stable evidence. **7,** 153–162 (2010).

95. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41,** 178–186 (2009).

96. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96,** 8681–8686 (1999).

97. Snover, D. C. Update on the serrated pathway to colorectal carcinoma. *Hum. Pathol.* **42,** 1–10 (2011).

98. Etoh, T. *et al.* Increased DNA methyltransferase 1 (DNMT1) protein expression correlates significantly with poorer tumor differentiation and frequent DNA hypermethylation of multiple CpG islands in gastric cancers. *Am. J. Pathol.* **164,** 689–699 (2004).

99. Eads, C. a. CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression. *Cancer Res.* **59,** 2302–2306 (1999).

100. Ganem, N. J., Godinho, S. a. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460,** 278–282 (2009).

101. Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386,** 761, 763 (1997).

102. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science (80-.).* **318,** 1108–1113 (2007).

103. Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* **361,** 2449–2460 (2009).

104. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).

105. Polakis, P. The many ways of Wnt in cancer. 45–51 (2007). doi:10.1016/j.gde.2006.12.007

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

107. Polakis, P. Wnt signaling and cancer Wnt signaling and cancer. 1837–1851 (2000). doi:10.1101/gad.14.15.1837

108. Behrens, J. Functional interaction of an axin homolog, conductin, with [bgr]-catenin, APC, and GSK3[bgr]. *Science (80-.).* **280,** 596–599 (1998).

109. Yamamoto, H. *et al.* Axil, a member of the Axin family, interacts with both glycogen synthase kinase 3beta and beta-catenin and inhibits axis formation of Xenopus embryos. *Mol. Cell. Biol.* **18,** 2867–75 (1998).

110. Willert, K., Shibamoto, S. & Nusse, R. Wnt-induced dephosphorylation of Axin releases ??-catenin from the Axin complex. *Genes Dev.* **13,** 1768–1773 (1999).

111. Rustgi, A. K. BRAF: A Driver of the Serrated Pathway in Colon Cancer. *Cancer Cell* **24,** 1–2 (2013).

112. Rad, R. *et al.* A Genetic Progression Model of BrafV600E-Induced Intestinal Tumorigenesis Reveals Targets for Therapeutic Intervention. *Cancer Cell* **24,** 15–29 (2013).

113. Vousden, K. H. & Prives, C. Blinded by the Light: The Growing Complexity of p53. *Cell* **137,** 413–431 (2009).

114. Brown, C. J., Lain, S., Verma, C. S., Fersht, A. R. & Lane, D. P. Awakening guardian angels: drugging the p53 pathway. *Nat. Rev. Cancer* **9,** 862–873 (2009).

115. Guo, S. T. *et al.* INPP4B is an oncogenic regulator in human colon cancer. *Oncogene* 1–13 (2015). doi:10.1038/onc.2015.361

116. Tobelaim, W. S. *et al.* Tumour-promoting role of SOCS1 in colorectal cancer cells. *Sci. Rep.* **5,** 14301 (2015).

117. Siegel, R. *et al.* Cancer Treatment and Survivorship Statistics , 2012. *CA. Cancer J. Clin.* **62,** 220–241 (2012).

118. Hu, C.-Y. *et al.* Time trend analysis of primary tumor resection for stage IV colorectal cancer: less surgery, improved survival. *JAMA Surg.* **150,** 245–251 (2015).

119. Ferrarotto, R. *et al.* Durable complete responses in metastatic colorectal cancer treated with chemotherapy alone. *Clin. Colorectal Cancer* **10,** 178–182 (2011).

120. Meyerhardt, J. a & Mayer, R. J. Systemic therapy for colorectal cancer. *N. Engl. J. Med.* **352,** 476–487 (2005).

121. Chee, C. E. & Sinicrope, F. A. Targeted Therapeutic Agents for Colorectal Cancer. *Gastroenterol Clin North Am* **39,** 601–613 (2010).

122. Kelland, L. The resurgence of platinum-based cancer chemotherapy. *Nat. Rev. Cancer* **7,** 573–584 (2007).

123. De Gramont, A. *et al.* Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol* **18,** 2938–2947 (2000).

124. Zhao, Z. *et al.* Patterns of treatment with chemotherapy and monoclonal antibodies for metastatic colorectal cancer in Western Europe. *Curr. Med. Res. Opin.* **28,** 221–229 (2012).

125. Twelves, C. *et al.* Capecitabine as adjuvant treatment for stage III colon cancer. *N. Engl. J. Med.* **352,** 2696–2704 (2005).

126. Cassinello, J. *et al.* Activity and safety of oxaliplatin with weekly 5-fluorouracil bolus and low-dose leucovorin as first-line treatment for advanced colorectal cancer. *Clin. Colorectal Cancer* **3,** 108–12 (2003).

127. Falcone, a. *et al.* Phase III Trial of Infusional Fluorouracil, Leucovorin, Oxaliplatin, and Irinotecan (FOLFOXIRI) Compared With Infusional Fluorouracil, Leucovorin, and Irinotecan (FOLFIRI) As First-Line Treatment for Metastatic Colorectal Cancer: The Gruppo Oncologico Nor. *J. Clin. Oncol.* **25,** 1670–1676 (2007).

128. Saltz, L. B. *et al.* Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med* **343,** 905–914 (2000).

129. Van Cutsem, E., Nordlinger, B. & Cervantes, a. Advanced colorectal cancer: ESMO clinical practice guidelines for treatment. *Ann. Oncol.* **21,** 93–97 (2010).

130. Hansel, T. T., Kropshofer, H., Singer, T., Mitchell, J. A. & George, A. J. T. The safety and side effects of monoclonal antibodies. *Nat. Rev. Drug Discov.* **9,** 325–338 (2010).

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

132. Ribic, C. M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N. Engl. J. Med.* **349,** 247–257 (2003).

133. French, A. J. *et al.* Prognostic significance of defective mismatch repair and BRAF V600E in patients with colon cancer. *Clin. Cancer Res.* **14,** 3408–3415 (2008).

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

135. 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).

136. Lindor, N. *et al.* Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *J Clin Oncol* **20,** 1043–1048 (2002).

137. Marcus, V. *et al.* Immunohistochemistry for hMLH1 and hMSH2: a practical test for DNA mismatch repair-deficient tumors. *Am J Surg Pathol* **23,** 1248–1255 (1999).

138. Thibodeau, S. N. *et al.* Altered Expression of hMSH2 and hMLH1 in Tumors with Microsatellite Instability and Genetic Alterations in Mismatch Repair Genes Advances in Brief Altered Expression of hMSH2 and hMLH1 in Tumors with Microsatellite Instability and Genetic Alterations in M. 4836–4840 (1996).

139. Fakih, M. G. & Padmanabhan, A. CEA monitoring in colorectal cancer. What you should know. *Oncology (Williston Park).* **20,** 579–587; discussion 588, 594, 596 passim (2006).

140. Van Cutsem, E. *et al.* Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* **27,** 3117–3125 (2009).

141. Tejpar, S. *et al.* Microsatellite instability (MSI) in stage II and III colon cancer treated with 5FU-LV or 5FU-LV and irinotecan (PETACC 3-EORTC 40993-SAKK 60/00 trial). *J Clin Oncol* **27,** (suppl; abstr 4001) (2009).

142. Roth, A. D. *et al.* Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J Clin Oncol* **28,** 466–474 (2010).

143. Bosman, F. T. *et al.* Tissue biomarker development in a multicentre trial context: a feasibility study on the PETACC3 stage II and III colon cancer adjuvant treatment trial. *Clin Cancer Res* **15,** 5528–5533 (2009).

144. Roth, A. D. *et al.* Stage-specific prognostic value of molecular markers in colon cancer: Results of the translational study on the PETACC 3-EORTC 40993-SAKK 60-00 trial. *J Clin Oncol* **27,** (suppl; abstr 4002) (2009).

145. Ogino, S. *et al.* KRAS mutation in stage III colon cancer and clinical outcome following intergroup trial CALGB 89803. *Clin Cancer Res* **15,** 7322–7329 (2009).

146. Bertagnolli, M. M. *et al.* Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *J Clin Oncol* **27,** 1814–1821 (2009).

147. Bertagnolli, M. M. *et al.* p27Kip1 in stage III colon cancer: implications for outcome following adjuvant chemotherapy in cancer and leukemia group B protocol 89803. *Clin Cancer Res* **15,** 2116–2122 (2009).

148. Markman, B., Javier Ramos, F., Capdevila, J. & Tabernero, J. EGFR and KRAS in colorectal cancer. *Adv. Clin. Chem.* **51,** 71–119 (2010).

149. Di Fiore, F. *et al.* Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. *Br. J. Cancer* **96,** 1166–1169 (2007).

150. Amado, R. G. *et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* **26,** 1626–1634 (2008).

151. Lièvre, A. *et al.* KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J. Clin. Oncol.* **26,** 374–379 (2008).

152. Esteller, M. *et al.* K-ras and p16 aberrations confer poor prognosis in human colorectal cancer. *J Clin Oncol* **19,** 299–304 (2001).

153. McLeod, H. L. & Murray, G. I. Tumour markers of prognosis in colorectal cancer. *Br. J. Cancer* **79,** 191–203 (1999).

154. Andreyev, H. J., Norman, a R., Cunningham, D., Oates, J. R. & Clarke, P. a. Kirsten ras mutations in patients with colorectal cancer: the multicenter 'RASCAL' study. *J. Natl. Cancer Inst.* **90,** 675–684 (1998).

155. George, B. & Kopetz, S. Predictive and prognostic markers in colorectal cancer. *Curr. Oncol. Rep.* **13,** 206–215 (2011).

156. Tabernero, J. *et al.* Targeted therapy in advanced colon cancer: the role of new therapies. *Ann. Oncol.* **15 Suppl 4,** iv55–62 (2004).

157. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359,** 1757–1765 (2008).

158. Allegra, C. J. *et al.* American Society of Clinical Oncology Provisional Clinical Opinion: Testing for KRAS Gene Mutations in Patients With Metastatic Colorectal Carcinoma to Predict Response to Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy. *J Clin Oncol* (2009). doi:JCO.2009.21.9170 [pii] 10.1200/JCO.2009.21.9170

159. Warren, R. S., Yuan, H., Matli, M. R., Gillett, N. A. & Ferrara, N. Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. *J. Clin. Invest.* **95,** 1789–1797 (1995).

160. Borgstrom, P., Gold, D. P., Hillan, K. J. & Ferrara, N. Importance of VEGF for breast cancer angiogenesis in vivo: implications from intravital microscopy of combination treatments with an anti-VEGF neutralizing monoclonal antibody and doxorubicin. *Anticancer Res.* **19,** 4203–4214 (1999).

161. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* **350,** 2335–2342 (2004).

162. Heinemann, V., Douillard, J. Y., Ducreux, M. & Peeters, M. Targeted therapy in metastatic colorectal cancer – An example of personalised medicine in action. *Cancer Treat. Rev.* **39,** 592–601 (2013).

163. Mouradov, D. *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res.* 3238–3247 (2014). doi:10.1158/0008-5472.CAN-14-0013

164. Arango, D. *et al.* Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* **129,** 874–884 (2005).

165. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).

166. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98,** 31–36 (2001).

167. Sproul, D. *et al.* Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A* **108,** 4364–4369 (2011).

168. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4,** 44–57 (2009).

169. Ke, N., Wang, X., Xu, X. & Abassi, Y. A. The xCELLigence system for real-time and label-free monitoring of cell viability. *Methods Mol. Biol.* **740,** 33–43 (2011).

170. Mariadason, J. M. *et al.* Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res* **63,** 8791–8812 (2003).

171. Dopeso, H. *et al.* Aprataxin tumor levels predict response of colorectal cancer patients to irinotecan-based treatment. *Clin. Cancer Res.* **16,** 2375–82 (2010).

172. Monks, A. *et al.* Feasibility of a High-Flux Anticancer Drug Screen Using a Diverse Panel of Cultured Human Tumor Cell Lines. *JNCI J. Natl. Cancer Inst.* **83,** 757–766 (1991).

173. Skehan, P. *et al.* New colorimetric cytotoxicity assay for anticancer-drug screening. *J. Natl. Cancer Inst.* **82,** 1107–12 (1990).

174. Li, L.-C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18,** 1427–1431 (2002).

175. Sinicrope, F. A. *et al.* Apoptotic and mitotic indices predict survival rates in lymph node-negative colon carcinomas. *Clin. Cancer Res.* **5,** 1793–1804 (1999).

176. al-Sheneber, I. F., Shibata, H. R., Sampalis, J. & Jothy, S. Prognostic significance of proliferating cell nuclear antigen expression in colorectal cancer. *Cancer* **71,** 1954–9 (1993).

177. Mayer, A. *et al.* The prognostic significance of proliferating cell nuclear antigen, epidermal growth factor receptor, and mdr gene expression in colorectal cancer. *Cancer* **71,** 2454–60 (1993).

178. Kovac, D. *et al.* Proliferating cell nuclear antigen (PCNA) as a prognostic factor for colorectal cancer. *Anticancer Res.* **15,** 2301–2

179. Bazzocco, S. *et al.* HIGHLY EXPRESSED GENES IN RAPIDLY PROLIFERATING TUMOR CELLS AS NEW TARGETS FOR COLORECTAL CANCER TREATMENT. *Clin. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-2457

180. Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138,** 2073–2087.e3 (2010).

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

182. Anacleto, C. *et al.* Colorectal cancer 'methylator phenotype': fact or artifact? *Neoplasia* **7,** 331–335 (2005).

183. Yamashita, K., Dai, T., Dai, Y., Yamamoto, F. & Perucho, M. Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* **4,** 121–131 (2003).

184. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99,** 3740–3745 (2002).

185. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196,** 261–282 (1987).

186. Irizarry, R. A., Wu, H. & Feinberg, A. P. A species-generalized probabilistic model-based definition of CpG islands. *Mamm Genome* **20,** 674–680 (2009).

187. Heidelberger, C. *et al.* Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **179,** 663–6 (1957).

188. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. *J. Clin. Oncol.* **10,** 896–903 (1992).

189. Folprecht, G. *et al.* Efficacy of 5-fluorouracil-based chemotherapy in elderly patients with metastatic colorectal cancer: a pooled analysis of clinical trials. *Ann. Oncol.* **15,** 1330–1338 (2004).

190. Labianca, R., Pancera, G. & Luporini, G. Factors influencing response rates for advanced colorectal cancer chemotherapy. *Ann. Oncol.* **7,** 901–906 (1996).

191. Padmanaban, G., Venkateswar, V. & Rangarajan, P. N. Haem as a multifunctional regulator. *Trends Biochem. Sci.* **14,** 492–6 (1989).

192. Hooda, J. *et al.* Enhanced heme function and mitochondrial respiration promote the progression of lung cancer cells. *PLoS One* **8,** e63402 (2013).

193. Warburg, O. Injuring of Respiration the Origin of Cancer Cells. *Science (80-.).* **123,** (1956).

194. Tennant, D. A., Durán, R. V & Gottlieb, E. Targeting metabolic transformation for cancer therapy. *Nat. Rev. Cancer* **10,** 267–77 (2010).

195. Kroemer, G. & Pouyssegur, J. Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell* **13,** 472–82 (2008).

196. Zheng, L., Roeder, R. G. & Luo, Y. S Phase Activation of the Histone H2B Promoter by OCA-S, a Coactivator Complex that Contains GAPDH as a Key Component. *Cell* **114,** 255–266 (2003).

197. Hara, M. R. *et al.* S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nat. Cell Biol.* **7,** 665–74 (2005).

198. Sánchez-Aragó, M. & Cuezva, J. M. The bioenergetic signature of isogenic colon cancer cells predicts the cell death response to treatment with 3-bromopyruvate, iodoacetate or 5-fluorouracil. *J. Transl. Med.* **9,** 19 (2011).

199. Fahim, F. a, Esmat, A. Y., Mady, E. a & Ibrahim, E. K. Antitumor activities of iodoacetate and dimethylsulphoxide against solid Ehrlich carcinoma growth in mice. *Biol. Res.* **36,** 253–62 (2003).

200. Wilding, J. L. & Bodmer, W. F. Cancer cell lines for drug discovery and development. *Cancer Res.* **74,** 2377–2384 (2014).

201. Domcke, S., Sinha, R., Levine, D. a, Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4,** 2126 (2013).

202. Saleh, H., Jackson, H., Khatib, G. & Banerjee, M. Correlation of bcl-2 oncoprotein immunohistochemical expression with proliferation index and histopathologic parameters in colorectal neoplasia. *Pathol. Oncol. Res.* **5,** 273–9 (1999).

203. Georgescu, C. V., Săftoiu, A., Georgescu, C. C., Ciurea, R. & Ciurea, T. Correlations of proliferation markers, p53 expression and histological findings in colorectal carcinoma. *J. Gastrointestin. Liver Dis.* **16,** 133–9 (2007).

204. Michael-Robinson, J. M. *et al.* Proliferation, Apoptosis, and Survival in High-Level Microsatellite Instability Sporadic Colorectal Cancer. *Clin Cancer Res* **7,** 2347–2356 (2001).

205. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–307 (2012).

206. Ahmed, D. *et al.* Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis* **2,** e71 (2013).

207. Yu, M. *et al.* A resource for cell line authentication, annotation and quality control. *Nature* **520,** 307–311 (2015).

208. Gillet, J.-P., Varma, S. & Gottesman, M. M. The Clinical Relevance of Cancer Cell Lines. *JNCI J. Natl. Cancer Inst.* **105,** 452–458 (2013).

209. Torsvik, a, Rosland, G. V & Bjerkvig, R. Spontaneous transformation of stem cells in vitro and the issue of cross-contamination. *Int J Biol Sci* **8,** 1051–1054 (2012).

210. Schlicker, A. *et al.* Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med. Genomics* **5,** 66 (2012).

211. Smiraglia, D. J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10,** 1413–1419 (2001).

212. Paz, M. F. *et al.* A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res.* **63,** 1114–21 (2003).

213. Hennessey, P. T. *et al.* Promoter methylation in head and neck squamous cell carcinoma cell lines is significantly different than methylation in primary tumors and xenografts. *PLoS One* **6,** e20584 (2011).

214. Wang, W. *et al.* Microarray profiling shows distinct differences between primary tumors and commonly used preclinical models in hepatocellular carcinoma. *BMC Cancer* **15,** 828 (2015).

215. Fang, M., Ou, J., Hutchinson, L. & Green, M. R. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Mol. Cell* **55,** 904–15 (2014).

216. Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nat. Genet.* **27,** 48–54 (2001).

217. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13,** 1977–2000 (2002).

218. Waldman, Y. Y., Geiger, T. & Ruppin, E. A genome-wide systematic analysis reveals different and predictive proliferation expression signatures of cancerous vs. non-cancerous cells. *PLoS Genet.* **9,** e1003806 (2013).

219. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24,** 227–35 (2000).

220. Feizi, A. & Bordel, S. Metabolic and protein interaction sub-networks controlling the proliferation rate of cancer cells and their impact on patient survival. *Sci. Rep.* **3,** 3041 (2013).

221. Lind, G. E. *et al.* A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines. *Mol. Cancer* **3,** 28 (2004).

222. Garinis, G. a *et al.* Hypermethylation-associated transcriptional silencing of E-cadherin in primary sporadic colorectal carcinomas. *J Pathol* **198,** 442–449 (2002).

223. Lee, S., Hwang, K. S., Lee, H. J., Kim, J.-S. & Kang, G. H. Aberrant CpG island hypermethylation of multiple genes in colorectal neoplasia. *Lab. Investig.* **84,** 884–893 (2004).

224. Jakubickova, L., Biesova, Z., Pastorekova, S., Kettmann, R. & Pastorek, J. Methylation of the CA9 promoter can modulate expression of the tumor-associated carbonic anhydrase IX in dense carcinoma cell lines. *Int J Oncol* **26,** 1121–1127 (2005).

225. Nakamura, J. *et al.* Expression of hypoxic marker CA IX is regulated by site-specific DNA methylation and is associated with the histology of gastric cancer. *Am J Pathol* **178,** 515–524 (2011).

226. Zatovicova, M. *et al.* Carbonic anhydrase IX as an anticancer therapy target: preclinical evaluation of internalizing monoclonal antibody directed to catalytic domain. *Curr Pharm Des* **16,** 3255–3263 (2010).

227. Mittag, F. *et al.* DAPK promotor methylation is an early event in colorectal carcinogenesis. *Cancer Lett* **240,** 69–75 (2006).

228. Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6870–5 (1998).

229. Nakagawa, S. *et al.* Expression of CLDN1 in colorectal cancer: a novel marker for prognosis. *Int J Oncol* **39,** 791–796 (2011).

230. Konopka, G., Tekiela, J., Iverson, M., Wells, C. & Duncan, S. A. Junctional adhesion molecule-A is critical for the formation of pseudocanaliculi and modulates E-cadherin expression in hepatic cells. *J Biol Chem* **282,** 28137–28148 (2007).

231. Singh, A. B. *et al.* Claudin-1 Upregulates the Repressor ZEB-1 to Inhibit E-Cadherin Expression in Colon Cancer Cells. *Gastroenterology* (2011). doi:S0016-5085(11)01221-2 [pii] 10.1053/j.gastro.2011.08.038

232. Zeilstra, J. *et al.* WNT signaling controls expression of pro-apoptotic BOK and BAX in intestinal cancer. *Biochem Biophys Res Commun* **406,** 1–6 (2011).

233. Zeilstra, J. *et al.* Deletion of the WNT target and cancer stem cell marker CD44 in Apc(Min/+) mice attenuates intestinal tumorigenesis. *Cancer Res* **68,** 3655–3661 (2008).

234. Batlle, E. *et al.* EphB receptor activity suppresses colorectal cancer progression. *Nature* **435,** 1126–1130 (2005).

235. Liu, W. *et al.* Effects of overexpression of ephrin-B2 on tumour growth in human colorectal cancer. *Br J Cancer* **90,** 1620–1626 (2004).

236. Han, H. *et al.* DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* **20,** 4299–4310 (2011).

237. Edgar, R., Tan, P. P. C., Portales-Casamar, E. & Pavlidis, P. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin* **7,** 28 (2014).

238. Mazzolini, R. *et al.* Brush border myosin Ia inactivation in gastric but not endometrial tumors. *Int J Cancer* (2012). doi:10.1002/ijc.27856

239. Lee, N. P., Poon, R. T., Shek, F. H., Ng, I. O. & Luk, J. M. Role of cadherin-17 in oncogenesis and potential therapeutic implications in hepatocellular carcinoma. *Biochim Biophys Acta* **1806,** 138–145 (2010).

240. Su, M.-C., Yuan, R.-H., Lin, C.-Y. & Jeng, Y.-M. Cadherin-17 is a useful diagnostic marker for adenocarcinomas of the digestive system. *Mod Pathol* **21,** 1379–1386 (2008).

241. Li, T. *et al.* Expression of selenium-binding protein 1 characterizes intestinal cell maturation and predicts survival for patients with colorectal cancer. *Mol Nutr Food Res* (2008). doi:10.1002/mnfr.200700331

242. Lee, S. J. & Michel, S. L. J. Structural Metal Sites in Nonclassical Zinc Finger Proteins Involved in Transcriptional and Translational Regulation. (2014).

243. Krishna, S. S. Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Res.* **31,** 532–550 (2003).

244. Shahi, P. *et al.* The Transcriptional Repressor ZNF503/Zeppo2 Promotes Mammary Epithelial Cell Proliferation and Enhances Cell Invasion. *J. Biol. Chem.* **290,** 3803–3813 (2015).

245. Yang, Y. M. *et al.* Galpha12 gep oncogene deregulation of p53-responsive microRNAs promotes epithelial-mesenchymal transition of hepatocellular carcinoma. *Oncogene* **34,** 2910–2921 (2015).

246. Yonezawa, K., Sugihara, Y., Oshima, K., Matsuda, T. & Nadano, D. Lyar, a cell growth-regulating zinc finger protein, was identified to be associated with cytoplasmic ribosomes in male germ and cancer cells. *Mol. Cell. Biochem.* **395,** 221–229 (2014).

247. Denecker, G. *et al.* Identification of a ZEB2-MITF-ZEB1 transcriptional network that controls melanogenesis and melanoma progression. *Cell Death Differ.* **21,** 1250–1261 (2014).

248. Cheng, Y. *et al.* KRAB zinc finger protein ZNF382 is a proapoptotic tumor suppressor that represses multiple oncogenes and is commonly silenced in multiple carcinomas. *Cancer Res.* **70,** 6516–6526 (2010).

249. Yu, J. *et al.* Zinc-finger protein 331, a novel putative tumor suppressor, suppresses growth and invasiveness of gastric cancer. *Oncogene* **32,** 307–317 (2013).

250. Xiao, Y. *et al.* Zinc-Finger Protein 545 Inhibits Cell Proliferation as a Tumor Suppressor through Inducing Apoptosis and is Disrupted by Promoter Methylation in Breast Cancer. *PLoS One* **9,** e110990 (2014).

251. Aoki, K. *et al.* RP58 associates with condensed chromatin and mediates a sequence-specific transcriptional repression. *J. Biol. Chem.* **273,** 26698–704 (1998).

252. Meng, G. *et al.* Structural analysis of the gene encoding RP58 , a sequence-specific transrepressor associated with heterochromatin. **242,** 59–64 (2000).

253. Tatard, V. M., Xiang, C., Biegel, J. a & Dahmane, N. ZNF238 is expressed in postmitotic brain cells and inhibits brain tumor growth. *Cancer Res.* **70,** 1236–46 (2010).

254. Heng, J. I.-T. *et al.* The Zinc Finger Transcription Factor RP58 Negatively Regulates Rnd2 for the Control of Neuronal Migration During Cerebral Cortical Development. *Cereb. Cortex* 1–11 (2013). doi:10.1093/cercor/bht277

255. Yokoyama, S. *et al.* A Systems Approach Reveals that the Myogenesis Genome Network Is Regulated by the Transcriptional Repressor RP58. *Dev. Cell* **17,** 836–848 (2009).

256. Miyata, K. *et al.* DNA methylation analysis of human myoblasts during in vitro myogenic differentiation: de novo methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation. *Hum. Mol. Genet.* **24,** 410–423 (2015).

8 Appendix

Appendix 1

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

Highly Expressed Genes in Rapidly Proliferating Tumor Cells as New Targets for Colorectal Cancer Treatment

Sarah Bazzocco, Higinio Dopeso, Fernando Carton-Garcia, et al.

Clin Cancer Res Published OnlineFirst May 5, 2015.

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

Cancer Therapy: Preclinical

Clinical Cancer Research

Highly Expressed Genes in Rapidly Proliferating Tumor Cells as New Targets for Colorectal Cancer Treatment

Sarah Bazzocco[1,2], Higinio Dopeso[1,2], Fernando Carton-Garcia[1,2], Irati Macaya[1,2], Elena Andretta[1,2], Fiona Chionh[3], Paulo Rodrigues[1,2], Miriam Garrido[1], Hafid Alazzouzi[1,2], Rocio Nieto[1,2], Alex Sanchez[4,5], Simo Schwartz Jr[2,6], Josipa Bilic[1,2], John M. Mariadason[3], and Diego Arango[1,2]

Abstract

Purpose: The clinical management of colorectal cancer patients has significantly improved because of the identification of novel therapeutic targets such as EGFR and VEGF. Because rapid tumor proliferation is associated with poor patient prognosis, here we characterized the transcriptional signature of rapidly proliferating colorectal cancer cells in an attempt to identify novel candidate therapeutic targets.

Experimental Design: The doubling time of 52 colorectal cancer cell lines was determined and genome-wide expression profiling of a subset of these lines was assessed by microarray analysis. We then investigated the potential of genes highly expressed in cancer cells with faster growth as new therapeutic targets.

Results: Faster proliferation rates were associated with microsatellite instability and poorly differentiated histology. The expression of 1,290 genes was significantly correlated with the growth rates of colorectal cancer cells. These included genes involved in cell cycle, RNA processing/splicing, and protein transport. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and protoporphyrinogen oxidase (PPOX) were shown to have higher expression in faster growing cell lines and primary tumors. Pharmacologic or siRNA-based inhibition of GAPDH or PPOX reduced the growth of colon cancer cells *in vitro*. Moreover, using a mouse xenograft model, we show that treatment with the specific PPOX inhibitor acifluorfen significantly reduced the growth of three of the seven (42.8%) colon cancer lines investigated.

Conclusion: We have characterized at the transcriptomic level the differences between colorectal cancer cells that vary in their growth rates, and identified novel candidate chemotherapeutic targets for the treatment of colorectal cancer. *Clin Cancer Res; 1–10.* ©2015 AACR.

Introduction

Colorectal cancer is one of the three most prevalent types of cancer in the western world and accounts for over 1.2 million new cases and 600,000 deaths every year worldwide (1). The genetic and epigenetic defects and the sequence of how these events accumulate during tumor progression are well characterized. Our current knowledge of the molecular mechanisms underlying the development of colorectal cancer is the result of extensive investigation in previous decades and the new light shed by the more recent genome-wide efforts such as The Cancer Genome Atlas—TCGA project (2) or the Encyclopedia Of DNA Elements—ENCODE (3).

Since the late 1990s, when expression microarray analysis became popular in the field, it was apparent that many genes were involved in the regulation of the cell cycle (4, 5). Because a major hallmark of cancer is uncontrolled rapid proliferation, it was not surprising to find that many of the genes that control cell-cycle progression were deregulated in the different tumor types investigated, compared with the corresponding normal tissue (4, 6). However, despite some early studies (7, 8), the genes with higher expression in rapidly proliferating tumor cells compared with slowly cycling tumors are not as well characterized. This is of considerable clinical relevance because it has been repeatedly observed that rapid tumor proliferation is associated with poor patient prognosis (9–12). Moreover, some of the most widely used chemotherapeutic agents for various types of cancer are inhibitors of proteins that are involved in cell proliferation, such as hydroxyurea, methotrexate, and doxorubicin/etoposide, which target ribonucleotide reductase, dihydrofolate reductase, and topoisomerase II, respectively. Notably, 5-fluorouracil (5-FU),

[1]Group of Molecular Oncology, CIBBIM-Nanomedicine, Vall d'Hebron University Hospital Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain. [3]Ludwig Institute for Cancer Research Melbourne-Austin Branch and Olivia Newton-John Cancer Research Institute, Melbourne, Victoria, Australia. [4]Unitat d'Estadística i Bioinformàtica, Vall d'Hebron University Hospital Research Institute (VHIR), Barcelona, Spain. [5]Departament d'Estadística, Universitat de Barcelona, Barcelona, Spain. [6]Group of Drug Delivery and Targeting, CIBBIM-Nanomedicine, Vall d'Hebron University Hospital Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Diego Arango, Vall d'Hebron Research Institute, Passeig Vall d'Hebron, 119-129, Barcelona 08035, Spain. Phone: 34-93-274-6739; Fax: 34-93-489-3893; E-mail: diego.arango@vhir.org

doi: 10.1158/1078-0432.CCR-14-2457

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

Translational Relevance

The identification of novel therapeutic targets would significantly improve the clinical management of colorectal cancer patients. Because rapidly proliferating tumors are associated with poor patient prognosis, here we assessed the growth rates of a panel of 52 colorectal cancer cell lines and used microarray analysis to identify a subset of 966 genes with high expression levels in colorectal tumor cells with faster growth. As a proof of concept, we then demonstrated that similar to the thymidylate synthase inhibitor 5-fluorouracil (5-FU), a well-established therapeutic agent, pharmacologic inhibition of protoporphyrinogen oxidase (PPOX), a gene with significantly higher expression levels in rapidly proliferating tumor cells, resulted in reduced growth *in vitro* and in a preclinical xenograft model. Importantly, PPOX and other genes highly expressed in rapidly growing tumor cells might constitute novel therapeutic targets for colorectal cancer patients.

the gold standard agent for the treatment of colorectal cancer patients for over five decades (13), targets thymidylate synthetase, an important gene required for cell proliferation. In addition to 5-FU, the therapeutic options currently approved for the treatment of colorectal cancer are limited, and include irinotecan, oxaliplatin, and the targeted agents cetuximab/panitumumab, bevacizumab, and regorafenib. When used as single agents, the response rates for these drugs is below 30% and there is a clear need for the improvement of the clinical management of these patients that the identification of new therapeutic targets and novel agents would bring about.

In this study, we used a panel of 52 colorectal cancer cell lines to investigate different features associated with the growth rates of these cells. We found that higher proliferation rates in colorectal cancer cells were associated with a microsatellite instable (MSI) phenotype and poor differentiation. In addition, we used microarray analysis of a subset of 31 of these cell lines to determine the expression signature of rapidly proliferating tumor cells. Moreover, we identified protoporphyrinogen oxidase (PPOX) as a novel chemotherapeutic target candidate, and using chemical inhibitors or siRNA-based knockdown we confirmed that targeting of PPOX *in vitro* and *in vivo* significantly interferes with tumor growth.

Materials and Methods

Cell culture and primary tumor samples

A total of 52 colorectal cancer cell lines were used: Caco2, Colo201, Colo205, Colo320, DLD1, HCT116, HCT15, HCT8, HT29, HUTU80, LoVo, LS1034, LS174T, LS513, RKO, SKCO1, SNUC2B, SW1116, SW403, SW48, SW480, SW620, SW837, SW948, T84, and WiDr were purchased from the ATCC. HDC108, HDC111, HDC114, HDC133, HDC15, HDC54, HDC75, HDC8, HDC87, and HDC9 were a kind gift from Dr. Johannes Gebert (Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany). HT29-cl16E, HT29-cl19A, HCC2998, KM12, and RW2982, were a kind gift from Dr. L.H. Augenlicht (Albert Einstein Cancer Center, Bronx, NY). LIM1215 and LIM2405 were obtained from the Ludwig Institute for Cancer Research (Melbourne, Australia). ALA, Co115, FET, Isreco1, Isreco2, Isreco3, and TC71 were a kind gift from Dr. Richard Hamelin (INSERM U434 CEPH, Paris, France). GP5D and VACO5 were a kind gift from Dr. L.A. Aaltonen (Biomedicum Helsinki, Helsinki, Finland). All lines were obtained more than 6 months before the beginning of the experiments in this study and maintained in MEM (Life Technologies) supplemented with 10% fetal bovine serum, 1× antibiotic/antimycotic (100 U/mL streptomycin, 100 U/mL penicillin, and 0.25 μg/mL amphotericin B), 1× MEM nonessential amino acids solution, and 10 mmol/L HEPES buffer solution (all from Life Technologies). All lines were tested to be negative for *Mycoplasma* contamination (PCR Mycoplasma Detection Set; Takara). Cell lines were cultured until they reached 70% to 80% confluence and the medium was changed 8 hours before harvesting the cultures for RNA extraction. The cell lines used were not authenticated, but possible cell line cross-contamination was investigated by clustering analysis of genome-wide mRNA expression microarray data at the time of these experiments.

The data from primary tumor samples used in this study were obtained from TCGA. mRNA expression levels (Illumina RNAseq and Agilent microarray G4502A) and hematoxylin and eosin–stained high-resolution images of formalin-fixed, paraffin-embedded sections of primary tumors were downloaded from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). For light microscopy quantification of mitotic cells in these tumors, three random fields were selected, and the total number of cells (>500) and mitotic cells was scored blinded from the sample identity.

Doubling time

To determine the doubling time of each cell line, cells were seeded in seven 96-well plates. Seeding densities varied from 1×10^3 to 1.5×10^4 cells per well to ensure control cell densities did not exceed 80% confluence at the completion of the experiment. The plates were fixed with trichloroacetic acid (final concentration 10% *w*/*v*) at 24-hour intervals for 7 days. Plates were washed with tap water, air-dried and stained with 0.4% (*w*/*v*) sulforhodamine B (SRB) for 30 minutes. Excess SRB was washed out with 1% acetic acid and the plates were air dried. Cell-bound SRB was solubilized with 10 mmol/L Tris buffer pH 10 and absorbance was measured at 590 nm using a microplate reader (Sunrise, Tecan). The doubling times were calculated using Prism V.5.01 (GraphPad). All experiments were carried out at least three times with eight replicates each time.

As an independent approach to assess cell growth, the Roche xCELLigence System was used for real-time monitoring of cell proliferation (14). Cell lines were seeded in quadruplicate at a density of 5,000 cells per well in an E-Plate 96 (Roche Diagnostics, GmbH). The Real-Time Cell Analyzer MP instrument (Roche Diagnostics, GmbH), together with the E-Plate 96, was placed in a cell culture incubator maintained at 37°C with 5% CO_2, and continuous electrical impedance measurements were taken hourly for 8 days. Doubling times were calculated using Cell Index data from the exponential growth phase for each cell line, with RTCA software version 1.2.1.

Growth inhibition assay

The dose resulting in 50% growth inhibition (GI_{50}) in the presence of 5-FU, acifluorfen, sodium iodoacetate, oxadiazon (all from Sigma-Aldrich), or CGP 3466B maleate (Tocris), compared with the corresponding control, was determined as described previously (15, 16). Briefly, 5×10^3 cells per well were seeded in

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

96-well plates. Twenty-four hours after seeding, cells were treated with 5-FU (0, 0.01, 0.1, 0.5, 1, 2.5, 5, 10, 25, 50, 100, and 500 μmol/L), acifluorfen (0, 5, 25, 100, 200, 300, 400, 500, 750, 1,000, 2,000, and 3,000 μmol/L), Na iodoacetate (0, 0.01, 0.1, 1, 2.5, 5, 7.5, 10, 20, 30, 60, and 120 μmol/L), oxadiazon (0, 25, 50, 100, 200, 300, 400, 500, 600, 750, 1,000, and 1,250 μmol/L), or CGP 3466B maleate (0, 5, 10, 25, 50, 75, 100, 125, 250, 500, and 750 μmol/L) for 72 hours. Cells were fixed with trichloroacetic acid and stained with SRB, as described above. One plate of each cell line was fixed to assess cell number at the time when drug treatment started. The GI_{50} was calculated as described previously (17, 18). These experiments were carried out at least three times in quadruplicates.

Apoptosis and cell-cycle analysis

Two hundred thousand cells were seeded in triplicate in 6-well plates. Control wells reached a confluence of approximately 80% at the completion of the experiment. Twenty-four hours after seeding, cells were treated with 0, 10, 20, or 30 μmol/L sodium iodoacetate or 0, 400, 800, or 1,200 μmol/L acifluorfen (both Sigma-Aldrich) for 72 hours. Both, floating and adherent cells, were harvested, washed with cold PBS, and resuspended in 50 μg/mL propidium iodide, 0.1% sodium citrate, and 0.1% Triton X-100. Cells were stained for 2 hours at 4°C, and 10,000 cells were analyzed for DNA content using a FacsCalibur Flow Cytometer (Becton Dickinson). The percentage of cells with a subdiploid DNA content was quantified using WinList 2.0 (Verity Software House). The cell-cycle profile was established using the ModFit 2.0 (Verity Software House).

Protein extraction and Western blot analysis

Seven hundred and fifty thousand cells were seeded in 6-well plates. Twenty-four hours after seeding, cells were treated with 0, 10, or 20 μmol/L sodium iodoacetate or 0, 400, or 800 μmol/L acifluorfen for 24 hours. Cells were harvested, washed with cold PBS, and cell pellets were resuspended in 0.1 mL of lysis buffer (25 mmol/L Hepes, pH 7.5, 150 mmol/L NaCl, 5 mmol/L $MgCl_2$, 1% NP-40, 1 mmol/L DTT, 10% glycerol and protease inhibitors). Aliquots of the cleared supernatant containing total protein (25 μg) were loaded on a 15% acrylamide gel. After gel electrophoresis, proteins were transferred to a polyvinylidene difluoride (PVDF) membrane and probed with rabbit polyclonal anti-cleaved PARP (Asp214) antibody (#9541; Cell Signaling Technology; 1:2,000), mouse monoclonal anti-β-tubulin antibody (T4026; Sigma-Aldrich; 1:1,000), or rabbit polyclonal anti-actin antibody (Santa Cruz Biotechnology, H-300; 1:1,000).

Clonogenic assay

Five hundred HCT116 or DLD1 cells were seeded in triplicate in 6-well plates. Twenty-four hours after seeding, cells were treated with 0 or 15 μmol/L sodium iodoacetate or 0 or 1,200 μmol/L acifluorfen for 9 hours. The medium containing the drug was washed off and replaced with fresh medium without drug. Colony formation was monitored over the following 2 to 3 weeks. Cultures were stained with 1% crystal violet for 30 minutes, washed with distilled water, air dried, and the number of colonies was determined blinded from the sample identity. Each cell line was assayed three times, each time in triplicate.

RNA extraction and quantitative RT-PCR

Cell cultures were harvested at 70% to 80% confluence and total RNA was extracted using TRizol Reagent (Life Technologies) according to the manufacturer's instructions. Total RNA (2 μg) was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Life Technologies), and relative mRNA levels of PPOX, GAPDH, TYMS, CALCOCO2, CBX5, and SMAD4 were assessed by Real-Time PCR using SYBR Green Master Mix (Life Technologies). 18S rRNA (TaqMan Master Mix, Life Technologies) was used as a standardization control for the $2^{-\Delta\Delta Ct}$ method as described before (19). The primers used were TYMS-qPCR-forward 5′-ACA CAC TTT GGG AGA TGC AC-3′, TYMS-qPCR-reverse 5′-GGT TCT CGC TGA AGC TGA AT-3′, PPOX-qPCR-forward 5′-GGC GCT GGA AGG TAT CTC TA-3′, PPOX-qPCR-reverse 5′-CTG AAG CTG GAA TGG CAC TA-3′, GAPDH-qPCR-forward 5′-ACC CAC TCC TCC ACC TTT GAC-3′, GAPDH-qPCR-reverse 5′-CAT ACC AGG AAA TGA GCT TGA CAA-3′, SMAD4-qPCR-forward 5-AAA ACG GCC ATC TTC AGC AC-3′, SMAD4-qPCR-reverse 5′-AGG CCA GTA ATG TCC GGG A-3′, CALCOCO2-qPCR-forward 5′-GAA AGA GAG ATT GGA AGG AGA AA-3′, CALCOCO2-qPCR-reverse 5′-AGG TAC TTG ATA CGG CAA AGA AT-3′, CBX5-qPCR-forward 5′-ACC CAG GGA GAA GTC AGA AA-3′, CBX5-qPCR-reverse 5′-CGA TAT CAT TGC TCT GCT CTC T-3′, 18S-qPCR-forward 5′-AGT CCC TGC CCT TTG TAC ACA-3′, 18S-qPCR-reverse 5′-GAT CCG AGG GCC TCA CTA AAC-3′, and 18S-Probe 5′-[6FAM]-CGC CCG TCG CTA CTA CCG ATT GG-[TAM]-3′.

Microarray mRNA expression analysis

All cell lines were cultured as described above. Total RNA was extracted with TRizol Reagent (Life Technologies) and then labeled and hybridized to Affymetrix HG-U133 Plus 2.0 chips as previously described (19). The mRNA levels were calculated after RMA (Robust Multichip Average) normalization as described previously (20). Clustering analysis was done with dChip software (21). Microarray data have been deposited at ArrayExpress (E-MTAB-2971).

Given that, for many genes, the relation between expression and growth rate was monotonic but not linear, a Spearman rank correlation was used to identify genes whose expression was associated with growth rates across a panel of 31 colorectal cancer cell lines. The Benjamini–Hochberg procedure was used to correct for multiple hypothesis testing ($P < 0.1$). To investigate whether there were gene sets with significant enrichment in the number of genes with expression/proliferation correlations, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (22). A Fisher exact test was used to identify significantly enriched categories of genes associated with cell growth. The Benjamini–Hochberg procedure was used to correct for multiple hypothesis testing ($P < 0.05$).

RNAi knockdown of PPOX and GAPDH

HCT116 cells (2×10^5) were seeded in 6-well plates and 24 hours later they were transfected with control On-TARGET plus nontargeting siRNA, or siRNA pools against GAPDH or PPOX (D-001810-10-05, D-001830-10-05, or L-008383-00-0005, respectively; Dharmacon) using Lipofectamine 2000 (Life Technologies). Expression levels and cell numbers were assessed 72 hours after transfection as described above.

Drug effects *in vivo* using a xenograft model

Six- to 7-week-old female and male NOD/SCID mice were purchased from Charles River Laboratories. The mice were maintained under sterile conditions and the experiments were carried

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

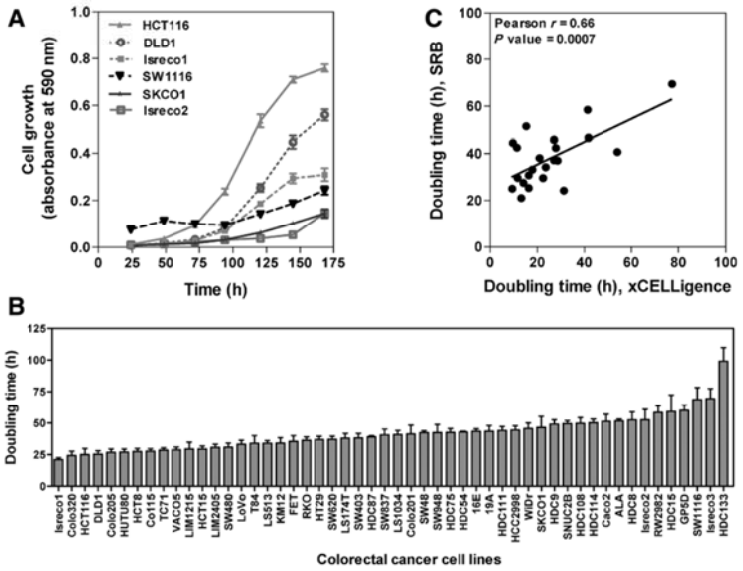

Figure 1.
Growth of colorectal cancer cell lines. A, for a panel of 52 colorectal cancer cell lines, cells were seeded in seven 96-well plates and harvested daily for 7 consecutive days and cell number was assessed indirectly using SRB staining and colorimetric quantification. Representative cell lines with rapid and slow growth are shown. B, histogram showing the doubling time of all 52 cell lines used in this study determined with the SRB assay (average of three experiments ± SEM). C, the growth of a subset of 22 of these cell lines was assessed using electrical impedance as the readout for the number of cells (xCELLigence). The doubling time was calculated with both techniques, and the correlation of the results obtained with both methods is shown.

out under observance of the protocol approved by the ethical committee for animal experimentation of the University Hospital Vall d'Hebron (Barcelona, Spain). The tumors were established by subcutaneous injection of 2×10^6 DLD1, Isreco1 or HCC2998 cells, 2.5×10^6 HCT116 cells, 1×10^6 HT29 cells, 3×10^6 RKO cells, and 5×10^6 T84 cells, all resuspended in 100-μL sterile PBS. When the tumors reached a volume of about 80 mm^3, the animals were randomized to groups treated with vehicle (PBS), 5-FU, acifluorfen, or sodium iodoacetate (50, 168, and 18.4 mg/kg, respectively) three times per week intraperitoneally. The long (L) and short (S) axis of the tumor were measured with a caliper five times a week. The tumor volume was calculated using the formula: $V = L \times S^2 \times 0.52$.

Determination of the grade of differentiation of cell lines in a xenograft model

Six- to 7-week-old male NOD/SCID mice were purchased from Charles River Laboratories, and experiments carried out under observance of a protocol approved by the Institute's oversight committee for animal experimentation. Tumors were established by subcutaneous injection of 5×10^6 cells in 200 μL of a 1:1 PBS: Matrigel solution into the right flank. When the tumors were >1,000 mm^3, they were formalin-fixed, paraffin-embedded and hematoxylin and eosin–stained sections were used to score tumor grade by an experienced pathologist blinded from the sample identity.

Results

Proliferation of colorectal cancer cell lines

Significant variability has been observed in the growth rates of colorectal tumors (9–12). Here, we thoroughly characterized the growth rates of a large panel of human colorectal cancer cell lines derived from colorectal tumors. The doubling time of these 52 cell lines was initially determined using an indirect SRB assay to quantify the total protein content in cell line cultures at 24-hour intervals over 1 week. Cell line growth demonstrated the expected lag phase before reaching an exponential growth phase followed by a growth plateau (Fig. 1A). Significant variability was observed in the doubling time during the exponential growth phase of this panel of cell lines (Fig. 1B and Supplementary Table S1). For a subset of 22 lines, we validated these results using an independent technique based on electrical impedance as the readout for real-time noninvasive cell growth monitoring (xCELLigence; Roche Diagnostics), and we found good correlation between the doubling time calculated through both approaches (Pearson $r = 0.66$; $P = 0.0007$; Fig. 1C).

Inactivation of mismatch repair genes results in the accumulation of mutations throughout the genome that manifests as microsatellite instability (MSI) in approximately 15% of colorectal tumors (23). However, the majority of colorectal tumors show no MSI and instead display chromosomal instability with large chromosomal abnormalities, and are referred to as microsatellite stable (MSS) or chromosomal instable (CIN) tumors. We found here that MSI cell lines grew significantly faster than MSS lines (Fig. 2A). A subset of 27 of these cell lines was grown as subcutaneous xenografts in immunodeficient mice, and the histologic grade of the tumors formed was determined. MSI tumors have been shown to be associated with high tumor grade (23). In good agreement, higher tumor grade was found to be associated with an MSI phenotype in these cell lines (χ^2, $P < 0.05$), and faster growth was observed in cell lines that generated high-grade tumors when grown as xenografts, compared with lines generating low/moderate-grade tumors (Fig. 2B). No associations were found between cell line doubling time and the mutational status of the genes most frequently mutated in colorectal tumors, such as *BRAF*, *KRAS*,

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

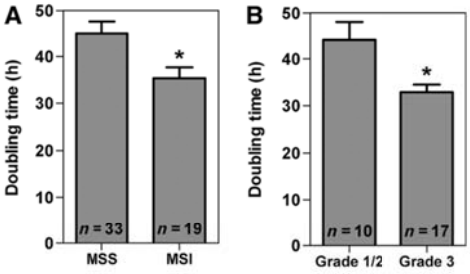

Figure 2.
The doubling time of colorectal cancer cell lines is associated with MSI and tumor grade. A, colorectal cancer cell lines with an MSI phenotype showed significantly faster growth (lower doubling time) compared with lines without MSI phenotype. B, cell lines growing as poorly differentiated (grade 3) tumors in subcutaneous xenografts in immunodeficient mice had faster growth than cell lines displaying moderately/highly differentiated histology (grade 1 and 2). N: number of cell lines. The mean ± SEM is shown. Asterisks indicate the Student *t* test $P < 0.05$.

TP53, *APC*, *PIK3CA*, *SMAD4*, *TCF7L2*, and CTNNB1 (Supplementary Fig. S1).

Expression profiling of colorectal cancer cell lines with different growth rates

High proliferation rates in colorectal tumors have been previously associated with poor patient prognosis (9–12), and although the molecular mechanisms regulating the progression of tumor cells through the different phases of the cell cycle are well characterized, the key rate-limiting steps are not fully understood. Here, we used microarray analysis to perform global gene expression profiling on a subset of these colorectal cancer cell lines ($n = 31$) to investigate the molecular mechanisms underlying the differences in growth rates.

For this analysis, we considered genes with expression levels significantly above background in 23 of the 31 cell lines investigated (>75%). Of the 11,512 genes investigated, the expression of 1,290 (11.2%) was significantly correlated with the doubling time of these cell lines (966 negatively and 324 positively correlated; Spearman correlation, BH FDR < 0.1 for at least one probe; Fig. 3A; Table 1; and Supplementary Table S2). The expression levels of six of these genes were independently assessed using quantitative real-time RT-PCR and a significant correlation was observed with mRNA levels quantified by microarray analysis (Supplementary Fig. S2).

Among the genes whose expression was found to be significantly correlated with the doubling time of the cell lines were multiple genes known to be key cell-cycle regulators, including multiple cyclins (A2, B1, B2, E2, F, I, and T2), cyclin-dependent kinases (CDKs; 1, 2, 9, and 13), the CDK inhibitor 2D (p19), and the cell division cycle (CDC) proteins 5L, 6, 14B, 25C, 27, and 37 (Fig. 3B and Supplementary Table S2). Consistently, functional group enrichment analysis also identified groups of genes that have long been known to participate in cell-cycle regulation both in normal and tumor cells, such as Gene Ontology biologic process categories involved in cell cycle, mitosis, RNA processing, and DNA metabolic process (Supplementary Table S3). In addition, other groups of functionally related genes whose expression levels are associated with growth rates included RNA splicing, protein transport, and ubiquitin-dependent protein catabolic process (Supplementary Table S3).

Identification of new candidate therapeutic targets

High rates of proliferation are associated with poor patient prognosis and at least some of the genes with higher relative expression in the tumors with faster growth are likely to be necessary to sustain rapid proliferation. We therefore hypothesized that targeting these genes could impair tumor growth. Genome-wide microarray analysis of the panel of 31 colorectal cancer cell lines investigated identified 966 genes with significantly higher expression in rapidly proliferating tumor cells (genes with negative Spearman *r* in Supplementary Table S2). Importantly, thymidylate synthase (TYMS), the direct target of the well-established chemotherapeutic agent 5-FU, was among the top 50 genes with highest negative correlation between doubling time and gene expression (Fig. 4A). Because of the availability of chemical inhibitors, we selected two additional genes, PPOX and glyceraldehydes-3-phosphate dehydrogenase (*GAPDH*), which showed significant negative correlations between gene expression and the doubling time of colorectal cancer cell lines (Fig. 4B and C). Importantly, the levels of expression of PPOX and GAPDH were significantly correlated with the rates of proliferation (percentage of mitotic cells) in a cohort of 36 primary colorectal tumors (Supplementary Fig. S3). No associations were observed between PPOX/GAPDH mRNA levels and tumor size, site, pathologic T/N/M, venous invasion, patient age, gender, or overall survival (Cox regression $P > 0.56$) in an extended cohort of 433 colorectal primary tumors (Supplementary Table S4). A modest reduction in PPOX levels was observed in late-stage tumors (Supplementary Table S4).

Inhibition of PPOX and GAPDH reduces the growth of colorectal cancer cells *in vitro*

We then used 5-FU, acifluorfen, and Na iodoacetate, specific chemical inhibitors of TYMS, PPOX, and GAPDH, respectively, to investigate whether their activity is necessary for the growth of colon cancer cells. As expected, treatment with 5-FU, a chemotherapeutic agent clinically used for the treatment of colorectal cancer, efficiently inhibited the growth of colon cancer cells (Fig. 4D). Similarly, acifluorfen and Na iodoacetate treatment resulted in a dose-dependent inhibition of the growth of colon cancer cells (Fig. 4E and F), which was not dependent on the growth rates of the cell lines (Supplementary Fig. S4). Moreover, both acifluorfen and Na iodoacetate significantly reduced the long term (>2 weeks) clonogenic capacity of colon cancer cells after short-term (9 hours) treatment, suggesting that these agents could cause cell death in addition to growth inhibition (Supplementary Fig. S5A and S5B). Consistently, flow cytometry analysis of propidium iodide–stained cells after acifluorfen or Na iodoacetate treatment revealed the presence of a significant proportion of cells with a subdiploid amount of DNA (Fig. 4G–I and Supplementary Fig. S5C and S5D), and PARP cleavage (Supplementary Fig. S5E and S5F), indicating that these agents induced apoptotic death in colon cancer cells. In addition, acifluorfen treatment was also associated with an arrest of the cell cycle in the G_0–G_1 phase (Fig. 4G–I and Supplementary Fig. S6).

Treatment of colon cancer cells with two additional chemically unrelated inhibitors of PPOX and GAPDH (oxadiazon

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

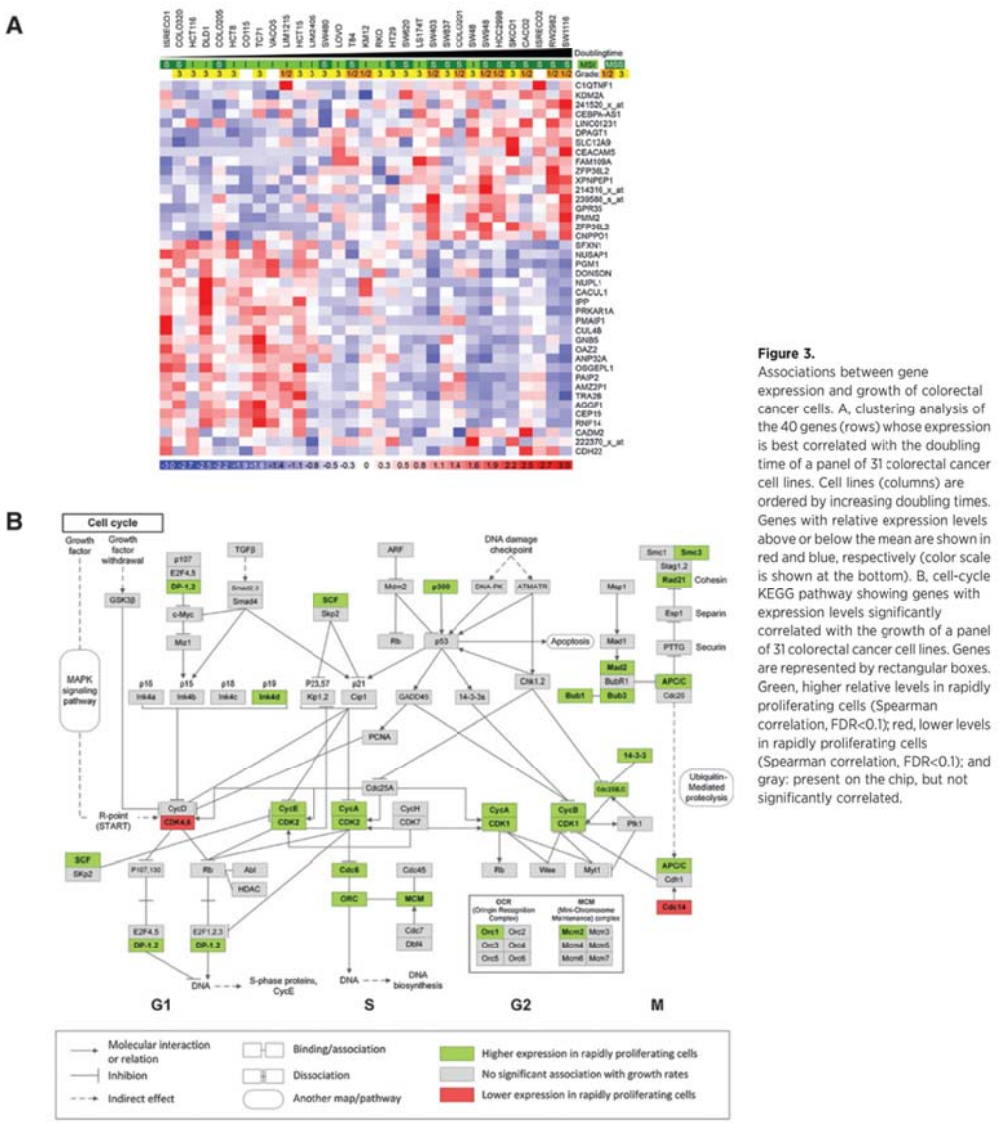

Figure 3.
Associations between gene expression and growth of colorectal cancer cells. A, clustering analysis of the 40 genes (rows) whose expression is best correlated with the doubling time of a panel of 31 colorectal cancer cell lines. Cell lines (columns) are ordered by increasing doubling times. Genes with relative expression levels above or below the mean are shown in red and blue, respectively (color scale is shown at the bottom). B, cell-cycle KEGG pathway showing genes with expression levels significantly correlated with the growth of a panel of 31 colorectal cancer cell lines. Genes are represented by rectangular boxes. Green, higher relative levels in rapidly proliferating cells (Spearman correlation, FDR<0.1); red, lower levels in rapidly proliferating cells (Spearman correlation, FDR<0.1); and gray: present on the chip, but not significantly correlated.

and CGP 3466B maleate, respectively) also resulted in a dose-dependent growth inhibition of colon cancer cells (Supplementary Fig. S7A and S7B). Moreover, RNAi-based knockdown of PPOX and GAPDH also interfered with the growth of colon cancer cells (Supplementary Fig. S7C–S7F), further indicating that PPOX and GAPDH are necessary for proliferation of colon cancer cells.

PPOX inhibition reduces the growth of colon cancer cells in a xenograft model

The *in vitro* experiments above suggested that PPOX and GAPDH could constitute novel therapeutic targets for colorectal cancer. To further investigate this possibility, we used a xenograft model in NOD/SCID immunodeficient mice. DLD1 and HCT116 cells were subcutaneously injected into the flanks

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

Table 1. Top 20 probes with highest correlation coefficient (positive and negative) between gene expression and doubling time in a panel of 31 colorectal cancer cell lines

ProbeSets ID	Gene symbol	Gene name	Spearman ρ	Spearman *P*	BH (FDR) adjusted *P*
227257_s_at	CACUL1	CDK2-associated, cullin domain 1	−0.74	1.85E−06	1.25E−02
230069_at	SFXN1	Sideroflexin 1	−0.74	2.21E−06	1.25E−02
222983_s_at	PAIP2	Poly(A) binding protein interacting protein 2	−0.74	2.34E−06	1.25E−02
201968_s_at	PGM1	Phosphoglucomutase 1	−0.73	2.89E−06	1.25E−02
223443_s_at	AMZ2P1	Archaelysin family metallopeptidase 2 pseudogene 1	−0.73	3.18E−06	1.25E−02
201051_at	ANP32A	Acidic (fleucine-rich) nuclear phosphoprotein 32 family, member A	−0.72	4.59E−06	1.25E−02
207124_s_at	GNB5	Guanine nucleotide binding protein (G protein), beta 5	−0.72	4.67E−06	1.25E−02
219978_s_at	NUSAP1	Nucleolar and spindle associated protein 1	−0.72	5.12E−06	1.25E−02
1554740_a_at	IPP	Intracisternal A particle-promoted polypeptide	−0.72	6.11E−06	1.25E−02
221677_s_at	DONSON	Downstream neighbor of SON	−0.71	7.67E−06	1.25E−02
220465_at	CEBPA-AS1	CEBPA antisense RNA 1 (head to head)	0.65	8.56E−05	2.69E−02
229690_at	FAM109A	Family with sequence similarity 109, member A	0.65	6.78E−05	2.61E−02
200070_at	CNPPD1	Cyclin Pas1/PHO80 domain containing 1	0.65	6.59E−05	2.61E−02
203201_at	PMM2	Phosphomannomutase 2	0.65	6.50E−05	2.61E−02
201368_at	ZFP36L2	ZFP36 ring finger protein-like 2	0.66	6.15E−05	2.61E−02
1569679_at	CDH22	Cadherin 22, type 2	0.66	5.90E−05	2.61E−02
209509_s_at	DPAGT1	Dolichyl-phosphate (UDP-N-acetylglucosamine) N-acetylglucosaminephosphotransferase 1 (GlcNAc-1-P transferase)	0.66	4.90E−05	2.51E−02
239588_s_at	—	—	0.68	3.06E−05	2.12E−02
208987_s_at	KDM2A	Lysine (K)-specific demethylase 2A	0.69	2.02E−05	1.70E−02
214316_x_at	CTC-425F1.4	—	0.72	4.76E−06	1.25E−02
1554696_s_at	TYMS	Thymidylate synthetase	−0.68	2.76E−05	1.97E−02
238117_at	PPOX	Protoporphyrinogen oxidase	−0.60	3.77E−04	3.74E−02
AFFX-HUMGAPDH/M33197_5_at	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	−0.55	1.30E−03	5.74E−02

NOTE: TYMS, GAPDH, and PPOX are also shown.

of 24 animals, and when the tumors reached a volume of 80 mm^3 the animals were randomized to a control group, or groups treated with acifluorfen, Na iodoacetate, or 5-FU. As expected, 5-FU treatment reduced the growth of these colon cancer cell lines (Fig. 5A and B). Although treatment with the GAPDH inhibitor Na iodoacetate did not have any effect on the growth of these cell lines, systemic administration of the PPOX inhibitor acifluorfen resulted in a significant inhibition of the growth of DLD1 cells (Fig. 5A and B). The average weight changes of control mice and 5-FU, Na iodoacetate or acifluorfen-treated animals are shown in Fig. 5C. To further investigate the sensitivity of colon cancer cell lines to acifluorfen and Na iodoacetate, additional cell lines were subcutaneously injected into immunodeficient NOD/SICD mice that were treated with these agents. Although Na iodoacetate did not significantly affect the growth of these additional cell lines, the growth of T84 and Isreco1 cells was significantly reduced in animals treated with acifluorfen (Supplementary Fig. S8). Collectively, these results indicate that PPOX could constitute a novel therapeutic target for the treatment of colon cancer.

Discussion

Significant variability has been reported in the rates of proliferation of colorectal cancer tumors, and faster proliferation is associated with poor patient prognosis (9–12). In primary colorectal tumors, an association has been reported between high-grade (poorly differentiated; refs. 24, 25) or MSI (26) and faster proliferation rates. Here, we show that cell lines that form high-grade tumors when grown as xenografts or have microsatellite instability proliferate significantly faster than cell lines forming low-grade (differentiated) tumors or MSS lines. These results indicate that the proliferative profile of the cell line panel used here closely recapitulates the characteristics of primary colorectal tumors. This is consistent with our recent findings demonstrating that the mutational landscape of colorectal cancer cell lines closely resembles that of primary colorectal cancers (27), and collectively establish cell lines as suitable models for the investigation of this disease. Interestingly, the mutational status of the genes most frequently mutated in colorectal tumors did not correlate with the growth rates of colon cancer cell lines, suggesting that these common genetic changes, when considered individually, do not have a consistent effect on the proliferation rates of colon cancer cells. However, the expression of 11.2% of the genes investigated was associated with the growth of these cell lines, indicating that changes in proliferation are fine-tuned at the transcriptional level in colon cancer cells.

Although several pioneer studies identified genes periodically expressed during different phases of the cell cycle (4, 5), limited progress has been made regarding the identification of genes with differential expression patterns in tumors with high and low proliferation rates (7, 8). Consistent with previous studies using a small number of cancer lines from nine different tumor types (NCI60 set containing 7 colon cancer cell lines; refs. 7, 8), we found that genes involved in cell cycle, RNA, and protein synthesis are closely correlated with the growth rates of colon cancer cells.

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

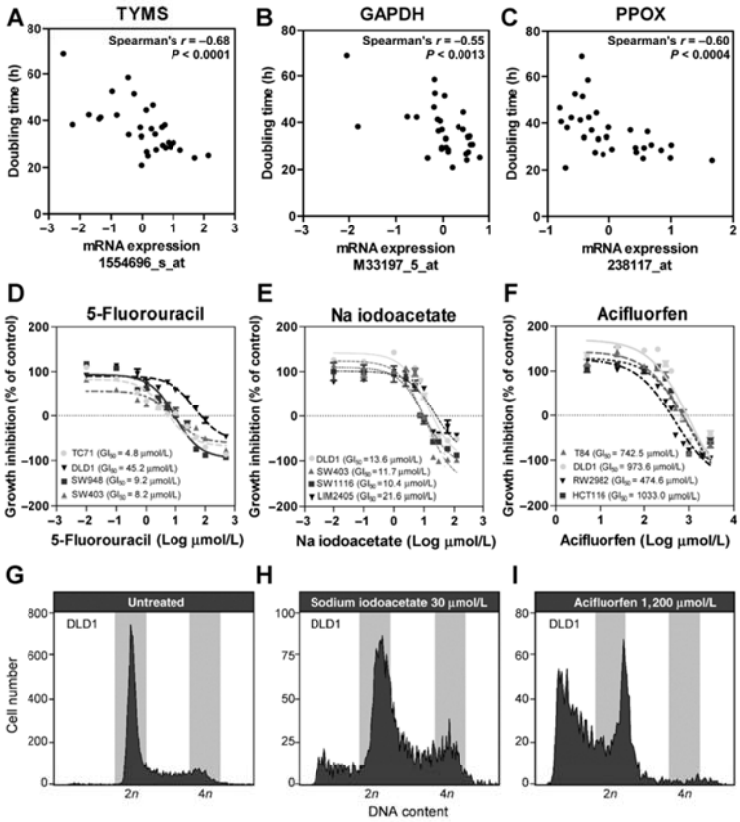

Figure 4.
Genes highly expressed in rapidly proliferating colon cancer cells as novel candidate therapeutic targets. The expression of TYMS (A), the direct target of the chemotherapeutic agent 5-FU, as well as GAPDH (B) and PPOX (C) were negatively correlated with the doubling time of a panel of 31 colorectal cancer cell lines. Inhibition of TYMS with 5-FU (D), GAPDH with Na iodoacetate (E), and PPOX with acifluorfen (F) resulted in a dose-dependent inhibition of the growth of different colorectal cancer cell lines. The GI50 of the cell lines tested is shown. G–I, the effects of acifluorfen and sodium iodoacetate treatment on the cell cycle of colon cancer cells. Cells were stained with propidium iodide and analyzed by flow cytometry.

Here, we found that analysis of a larger set of colon cancer lines ($n = 31$) that widely vary in their growth rates did not confirm some of the findings made on the NCI colon cancer cells, such as cholesterol metabolism, iron metabolism, and fatty acid metabolism. However, we found additional groups of functionally related genes significantly correlated with the growth rates of colorectal cancer cells including several categories related with protein metabolism, such as translation, protein transport, and cellular protein catabolic process (Supplementary Table S3).

There are currently a limited number of chemotherapeutic agents approved for their routine use in the fight against colorectal cancer, namely the antimetabolite 5-FU, the platinum compound oxaliplatin and the topoisomerase I inhibitor irinotecan, in addition to the targeted agents cetuximab, panitumumab, bevacizumab, and regorafenib. However, the response rate to each of these drugs used as single agents is below 30% and the identification of novel therapeutic targets and the subsequent development of new chemotherapeutic agents would likely improve the survival of these patients. Here, we hypothesized that inhibition of genes highly expressed in rapidly proliferating colorectal cancer cells can interfere with tumor growth, and these genes are therefore good candidate chemotherapeutic targets. In support of this hypothesis, we found that the direct target of 5-FU, thymidylate synthetase, was among the genes showing a highly significant correlation between its expression level and the rate of tumor cell growth. As a proof of concept, we selected two additional genes with high expression levels in rapidly proliferating colorectal cancer cells (GAPDH and PPOX) with known specific inhibitors for the encoded proteins (Na iodoacetate and oxadiazon or acifluorfen and CGP 3466B maleate, respectively) and found that, as the TYMS inhibitor 5-FU, GAPDH, and PPOX inhibitors significantly reduced the growth of colon cancer cells at micromolar concentrations. Moreover, using a preclinical subcutaneous xenograft model, we could demonstrate that at least the PPOX inhibitor acifluorfen was able to inhibit the growth of colon cancer cell lines (3 of 7; 42.3%). PPOX catalyzes the 6-electron oxidation of protoporphyrinogen IX to form protoporphyrin IX, the penultimate reaction of heme biosynthesis. Heme plays critical roles in multiple processes involving oxygen metabolism. This includes proteins that transport or store oxygen such as hemoglobin and myoglobin, but is also important in mitochondrial respiratory chain complexes, in cytochrome P450s, and in other enzymes that use or detoxify oxygen such as peroxidases and catalases (28). Our findings are consistent with the observation that inhibition of

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

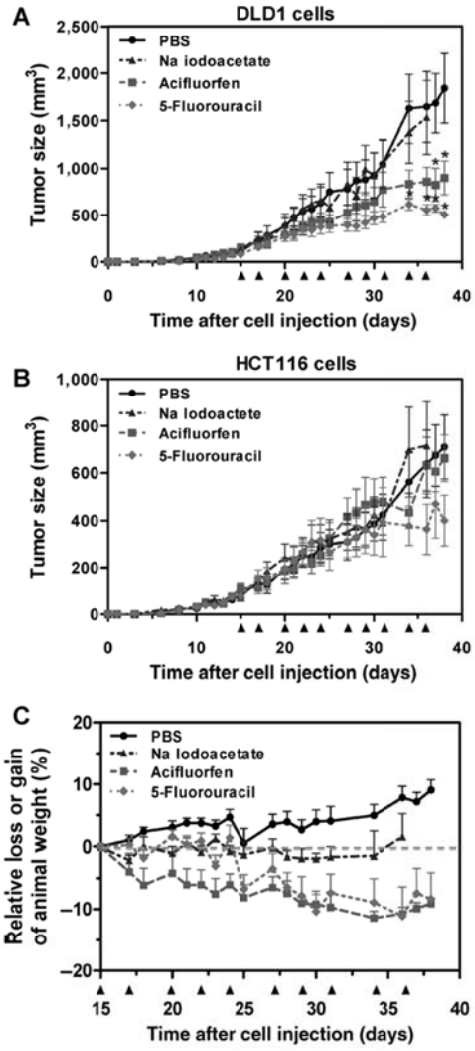

Figure 5.
Effects of GAPDH and PPOX inhibition on tumor growth using a xenograft model. Groups of NOD/SCID immunodeficient mice ($n = 6$ per group) with DLD1 and HCT116 cells as subcutaneous xenografts were treated (i.p.) three times per week with acifluorfen (168 mg/kg), Na iodoacetate (18.4 mg/kg), 5-FU (50 mg/kg), or vehicle PBS, starting when the tumors reached approximately 80 mm^3. Arrowheads in the *X*-axis indicate treatment times. Tumor size was monitored over time for DLD1 (A) and HCT116 (B) cells. Percentage animal weight gain/loss after drug treatment is shown in panel (C). Asterisks indicate statistically significant differences (Student *t* test, $P < 0.05$) in the mean tumor size in the control (PBS) group and treatment groups (5-FU or acifluorfen). The mean ± SEM is shown.

heme synthesis significantly reduced proliferation in lung cancer cells (29).

It has long been known that most cancer cells predominantly produce energy by a high rate of glycolysis and lactate production, an observation known as the Warburg effect (30). Therefore, it has been suggested before that differences in the metabolisms of tumor cells could offer a therapeutic window (31, 32). GAPDH catalyzes the sixth step of glycolysis, the conversion of glyceraldehyde 3-phosphate to D-glycerate 1,3-bisphosphate. Recently, GAPDH has been shown to translocate to the nucleus and to be implicated in several nonmetabolic processes, including transcriptional regulation and apoptosis (33, 34). We show here that inhibition of GAPDH efficiently reduces the growth of colon cancer cells *in vitro*. Interestingly, although the GAPDH inhibitor Na iodoacetate has been shown to reduce the growth of Ehrlich ascites carcinoma (EAC) cells and xenografts of colon cancer cells at doses similar or lower than the one used *in vivo* in this study (35, 36), no significant effects were observed here on the growth of subcutaneous xenografts of four different colon cancer cell lines. However, no toxicity was observed at the doses used, and based on the *in vitro* effects observed, it remains possible that Na iodoacetate treatment at higher doses and/or in other tumor cell lines, may interfere with tumor growth.

Importantly, we provide here a list of 966 genes that have significantly higher expression in the tumor cell lines with higher proliferation rates. The two genes that were further investigated in this study were selected because of the availability of specific chemical inhibitors. However, genetic inactivation of selected genes with significant associations between expression and tumor growth could be used to identify the most promising therapeutic targets, for which novel specific inhibitors could then be developed.

In summary, we found that the proliferation of colorectal cancer cells is significantly associated with higher tumor grade and an MSI phenotype. In addition, microarray transcriptomic analysis of a panel of 31 colorectal cancer cell lines shed new light on the molecular mechanisms regulating the uncontrolled proliferation of colorectal cancer cells. Moreover, we demonstrate that genes with high expression in rapidly proliferating tumor cells are good candidates for therapeutic targeting. As a proof of concept, we demonstrate that acifluorfen inhibits the growth of colorectal cancer cells *in vitro* and *in vivo*, identifying PPOX as a novel candidate chemotherapeutic target for the treatment of colorectal cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S. Bazzocco, H. Dopeso, D. Arango
Development of methodology: S. Bazzocco, H. Dopeso, F. Carton-Garcia, E. Andretta, M. Garrido, H. Alazzouzi, R. Nieto
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Bazzocco, H. Dopeso, F. Carton-Garcia, I. Macaya, P. Rodrigues, M. Garrido, J. Bilic, J.M. Mariadason
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Bazzocco, H. Dopeso, I. Macaya, E. Andretta, F. Chionh, M. Garrido, H. Alazzouzi, A. Sanchez, J. Bilic, D. Arango
Writing, review, and/or revision of the manuscript: S. Bazzocco, H. Dopeso, F. Chionh, S. Schwartz Jr, J.M. Mariadason, D. Arango
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Bazzocco, H. Dopeso, M. Garrido, H. Alazzouzi
Study supervision: S. Bazzocco, H. Dopeso, S. Schwartz Jr, D. Arango

Published OnlineFirst May 5, 2015; DOI: 10.1158/1078-0432.CCR-14-2457

Grant Support

This study was partially funded by grants of the Spanish Ministry for Economy and Competitiveness (CP05/00256, TRA2009-0093, SAF2008-00789, PI12/03103, and PI12/01095), the Association for International Cancer Research (AICR13-0245), and Agència de Gestió d'Ajuts Universitaris i de Recerca (SGR 157) to D. Arango. S. Bazzocco is funded by a predoctoral fellowship from the ISCiii (FI10/00519).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 29, 2014; revised April 8, 2015; accepted April 27, 2015; published OnlineFirst May 5, 2015.

References

1. Jemal A, Bray F, Ferlay J. Global Cancer Statistics. CA Cancer J Clin 2011;61:69–90.
2. Network TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.
3. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
4. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 2002;13:1977–2000.
5. Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, Sapinoso L, et al. Transcriptional regulation and function during the human cell cycle. Nat Genet 2001;27:48–54.
6. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci USA 2004;101:9309–14.
7. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, et al. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–35.
8. Waldman YY, Geiger T, Ruppin E. A genome-wide systematic analysis reveals different and predictive proliferation expression signatures of cancerous vs. non-cancerous cells. PLoS Genet 2013;9:e1003806.
9. Sinicrope FA, Hart J, Hsu HA, Lemoine M, Michelassi F, Stephens LC. Apoptotic and mitotic indices predict survival rates in lymph node-negative colon carcinomas. Clin Cancer Res 1999;5:1793–804.
10. al-Sheneber IF, Shibata HR, Sampalis J, Jothy S. Prognostic significance of proliferating cell nuclear antigen expression in colorectal cancer. Cancer 1993;71:1954–9.
11. Mayer A, Takimoto M, Fritz E, Schellander G, Kofler K, Ludwig H. The prognostic significance of proliferating cell nuclear antigen, epidermal growth factor receptor, and mdr gene expression in colorectal cancer. Cancer 1993;71:2454–60.
12. Kovac D, Rubinic M, Krasevic M, Krizanac S, Petrovecki M, Stimac D, et al. Proliferating cell nuclear antigen (PCNA) as a prognostic factor for colorectal cancer. Anticancer Res 1995;15:2301–2.
13. Heidelberger C, Chaudhuri NK, Danneberg P, Mooren D, Griesbach L, Duschinsky R, et al. Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. Nature 1957;179:663–6.
14. Ke N, Wang X, Xu X, Abassi YA. The xCELLigence system for real-time and label-free monitoring of cell viability. Methods Mol Biol 2011;740:33–43.
15. Mariadason JM, Arango D, Shi Q, Wilson AJ, Corner GA, Nicholas C, et al. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. Cancer Res 2003;63:8791–812.
16. Dopeso H, Mateo-Lozano S, Elez E, Landolfi S, Ramos Pascual FJ, Hernández-Losa J, et al. Aprataxin tumor levels predict response of colorectal cancer patients to irinotecan-based treatment. Clin Cancer Res 2010;16:2375–82.
17. Monks A, Scudiero D, Skehan P, Shoemaker R, Paull K, Vistica D, et al. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. J Natl Cancer Inst 1991;83:757–66.
18. Skehan P, Storeng R, Scudiero D, Monks A, McMahon J, Vistica D, et al. New colorimetric cytotoxicity assay for anticancer-drug screening. J Natl Cancer Inst 1990;82:1107–12.
19. Arango D, Laiho P, Kokko A, Alhopuro P, Sammalkorpi H, Salovaara R, et al. Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 2005;129:874–84.
20. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249–64.
21. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 2001;98:31–6.
22. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
23. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology 2010;138:2073–87.e3.
24. Saleh H, Jackson H, Khatib G, Banerjee M. Correlation of bcl-2 oncoprotein immunohistochemical expression with proliferation index and histopathologic parameters in colorectal neoplasia. Pathol Oncol Res 1999;5:273–9.
25. Georgescu CV, Săftoiu A, Georgescu CC, Ciurea R, Ciurea T. Correlations of proliferation markers, p53 expression and histological findings in colorectal carcinoma. J Gastrointestin Liver Dis 2007;16:133–9.
26. Michael-Robinson JM, Reid LE, Purdie DM, Walsh MD, Pandeya N, Simms LA, et al. Proliferation, apoptosis, and survival in high-level microsatellite instability sporadic colorectal cancer. Clin Cancer Res 2001;7:2347–56.
27. Mouradov D, Sloggett C, Jorissen RN, Love CG, Li S, Burgess AW, et al. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. Cancer Res 2014;74:3238–47.
28. Padmanaban G, Venkateswar V, Rangarajan PN. Haem as a multifunctional regulator. Trends Biochem Sci 1989;14:492–6.
29. Hooda J, Cadinu D, Alam MM, Shah A, Cao TM, Sullivan L a, et al. Enhanced heme function and mitochondrial respiration promote the progression of lung cancer cells. PLoS ONE 2013;8:e63402.
30. Warburg O. Injuring of respiration the origin of cancer cells. Science 1956;123:309–14.
31. Tennant DA, Durán RV, Gottlieb E. Targeting metabolic transformation for cancer therapy. Nat Rev Cancer 2010;10:267–77.
32. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell 2008;13:472–82.
33. Zheng L, Roeder RG, Luo Y. S phase activation of the histone H2B Promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. Cell Elsevier 2003;114:255–66.
34. Hara MR, Agrawal N, Kim SF, Cascio MB, Fujimuro M, Ozeki Y, et al. S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. Nat Cell Biol 2005;7:665–74.
35. Sánchez-Aragó M, Cuezva JM. The bioenergetic signature of isogenic colon cancer cells predicts the cell death response to treatment with 3-bromopyruvate, iodoacetate or 5-fluorouracil. J Transl Med 2011;9:19.
36. Fahim FA, Esmat AY, Mady EA, Ibrahim EK. Antitumor activities of iodoacetate and dimethylsulphoxide against solid Ehrlich carcinoma growth in mice. Biol Res 2003;36:253–62.

Appendix 2

Table 15: Functional group enrichment analysis showing all enriched biological process Gene Ontology categories.

Term	Count	%	P Value	Pop Hits	Fold Enrichment	Benjamini adjusted *P*
GO:0044260 cellular macromolecule metabolic process	525	40.70	2.00E-29	5214	1.46	6.09E-26
GO:0044237 cellular metabolic process	622	48.22	1.59E-28	6636	1.36	2.41E-25
GO:0043170 macromolecule metabolic process	543	42.09	5.14E-24	5710	1.38	3.90E-21
GO:0009987 cellular process	849	65.81	4.22E-24	10541	1.17	4.27E-21
GO:0008152 metabolic process	660	51.16	7.54E-20	7647	1.25	4.58E-17
GO:0044238 primary metabolic process	610	47.29	2.23E-19	6923	1.28	1.13E-16
GO:0044267 cellular protein metabolic process	259	20.08	4.44E-16	2355	1.60	1.93E-13
GO:0007049 cell cycle	116	8.99	1.24E-15	776	2.17	4.64E-13
GO:0000278 mitotic cell cycle	70	5.43	1.57E-14	370	2.75	5.33E-12
GO:0016070 RNA metabolic process	129	10.00	1.78E-14	938	2.00	5.40E-12
GO:0008380 RNA splicing	59	4.57	3.66E-14	284	3.02	1.01E-11
GO:0016071 mRNA metabolic process	69	5.35	5.17E-14	370	2.71	1.31E-11
GO:0006397 mRNA processing	63	4.88	6.91E-14	321	2.85	1.62E-11
GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	330	25.58	8.74E-13	3409	1.41	1.90E-10
GO:0006996 organelle organization	160	12.40	1.06E-12	1332	1.74	2.14E-10
GO:0000087 M phase of mitotic cell cycle	48	3.72	3.72E-12	224	3.11	7.07E-10
GO:0016043 cellular component organization	255	19.77	5.64E-12	2498	1.48	1.01E-09
GO:0022402 cell cycle process	85	6.59	1.02E-11	565	2.18	1.55E-09
GO:0000377 RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	38	2.95	9.40E-12	153	3.61	1.59E-09
GO:0000398 nuclear mRNA splicing, via spliceosome	38	2.95	9.40E-12	153	3.61	1.59E-09
GO:0000375 RNA splicing, via transesterification reactions	38	2.95	9.40E-12	153	3.61	1.59E-09
GO:0034641 cellular nitrogen compound metabolic process	345	26.74	1.00E-11	3670	1.37	1.60E-09
GO:0006396 RNA processing	83	6.43	1.17E-11	547	2.20	1.70E-09
GO:0044265 cellular macromolecule catabolic process	100	7.75	2.18E-11	725	2.00	3.02E-09
GO:0019538 protein metabolic process	276	21.40	4.49E-11	2812	1.43	5.94E-09
GO:0019941 modification-dependent protein catabolic process	84	6.51	5.82E-11	574	2.13	7.38E-09
GO:0043632 modification-dependent macromolecule catabolic process	84	6.51	5.82E-11	574	2.13	7.38E-09
GO:0022403 cell cycle phase	67	5.19	9.28E-11	414	2.35	1.13E-08
GO:0000280 nuclear division	45	3.49	9.67E-11	220	2.97	1.13E-08
GO:0007067 mitosis	45	3.49	9.67E-11	220	2.97	1.13E-08
GO:0030163 protein catabolic process	88	6.82	1.14E-10	622	2.05	1.28E-08
GO:0044257 cellular protein catabolic process	86	6.67	1.27E-10	603	2.07	1.38E-08
GO:0006807 nitrogen compound metabolic process	347	26.90	1.95E-10	3778	1.33	2.05E-08
GO:0000279 M phase	57	4.42	2.06E-10	329	2.52	2.08E-08
GO:0051603 proteolysis involved in cellular protein catabolic process	85	6.59	2.32E-10	600	2.06	2.28E-08

GO:0009057 macromolecule catabolic process	102	7.91	3.15E-10	781	1.90	2.99E-08
GO:0048285 organelle fission	45	3.49	3.76E-10	229	2.85	3.46E-08
GO:0006511 ubiquitin-dependent protein catabolic process	46	3.57	7.17E-10	242	2.76	6.42E-08
GO:0051301 cell division	52	4.03	7.71E-10	295	2.56	6.70E-08
GO:0010467 gene expression	284	22.02	1.08E-09	2999	1.38	9.11E-08
GO:0006259 DNA metabolic process	73	5.66	2.39E-09	506	2.10	1.97E-07
GO:0043933 macromolecular complex subunit organization	91	7.05	8.45E-09	710	1.86	6.76E-07
GO:0044248 cellular catabolic process	118	9.15	2.07E-08	1024	1.67	1.62E-06
GO:0044085 cellular component biogenesis	115	8.91	3.80E-08	1001	1.67	2.89E-06
GO:0065003 macromolecular complex assembly	84	6.51	6.70E-08	665	1.83	4.97E-06
GO:0022607 cellular component assembly	103	7.98	1.33E-07	887	1.69	9.62E-06
GO:0033554 cellular response to stress	73	5.66	2.47E-07	566	1.87	1.75E-05
GO:0009056 catabolic process	133	10.31	3.03E-07	1253	1.54	2.09E-05
GO:0006974 response to DNA damage stimulus	54	4.19	3.33E-07	373	2.10	2.25E-05
GO:0051276 chromosome organization	64	4.96	6.91E-07	485	1.92	4.57E-05
GO:0034645 cellular macromolecule biosynthetic process	253	19.61	1.34E-06	2812	1.31	8.65E-05
GO:0009059 macromolecule biosynthetic process	253	19.61	2.35E-06	2832	1.30	1.49E-04
GO:0000075 cell cycle checkpoint	21	1.63	2.79E-06	91	3.35	1.73E-04
GO:0007346 regulation of mitotic cell cycle	28	2.17	4.52E-06	152	2.68	2.75E-04
GO:0046907 intracellular transport	77	5.97	4.62E-06	657	1.70	2.76E-04
GO:0034621 cellular macromolecular complex subunit organization	49	3.80	5.79E-06	357	1.99	3.32E-04
GO:0006260 DNA replication	32	2.48	5.75E-06	190	2.45	3.36E-04
GO:0044249 cellular biosynthetic process	296	22.95	6.10E-06	3442	1.25	3.44E-04
GO:0051340 regulation of ligase activity	19	1.47	7.39E-06	81	3.41	4.08E-04
GO:0006412 translation	46	3.57	8.45E-06	331	2.02	4.59E-04
GO:0060255 regulation of macromolecule metabolic process	281	21.78	1.01E-05	3259	1.25	5.39E-04
GO:0006281 DNA repair	41	3.18	1.15E-05	284	2.10	6.01E-04
GO:0051716 cellular response to stimulus	89	6.90	1.65E-05	820	1.58	8.52E-04
GO:0043161 proteasomal ubiquitin-dependent protein catabolic process	21	1.63	1.73E-05	102	2.99	8.74E-04
GO:0010498 proteasomal protein catabolic process	21	1.63	1.73E-05	102	2.99	8.74E-04
GO:0051726 regulation of cell cycle	45	3.49	1.88E-05	331	1.97	9.35E-04
GO:0080090 regulation of primary metabolic process	279	21.63	4.16E-05	3291	1.23	2.01E-03
GO:0070271 protein complex biogenesis	60	4.65	4.12E-05	505	1.73	2.02E-03
GO:0006461 protein complex assembly	60	4.65	4.12E-05	505	1.73	2.02E-03
GO:0009058 biosynthetic process	297	23.02	4.75E-05	3542	1.22	2.25E-03
GO:0022618 ribonucleoprotein complex assembly	16	1.24	5.52E-05	69	3.37	2.58E-03
GO:0051438 regulation of ubiquitin-protein ligase activity	17	1.32	6.66E-05	78	3.17	3.06E-03
GO:0051439 regulation of ubiquitin-protein ligase activity during mitotic cell cycle	16	1.24	7.83E-05	71	3.27	3.55E-03
GO:0043412 biopolymer modification	143	11.09	8.27E-05	1526	1.36	3.69E-03

GO:0015031 protein transport	81	6.28	8.61E-05	762	1.54	3.79E-03
GO:0006310 DNA recombination	20	1.55	8.80E-05	105	2.77	3.81E-03
GO:0022613 ribonucleoprotein complex biogenesis	28	2.17	1.02E-04	180	2.26	4.34E-03
GO:0051246 regulation of protein metabolic process	62	4.81	1.12E-04	546	1.65	4.58E-03
GO:0043687 post-translational protein modification	115	8.91	1.11E-04	1182	1.41	4.59E-03
GO:0051351 positive regulation of ligase activity	16	1.24	1.09E-04	73	3.18	4.61E-03
GO:0045184 establishment of protein localization	81	6.28	1.16E-04	769	1.53	4.63E-03
GO:0007093 mitotic cell cycle checkpoint	12	0.93	1.15E-04	43	4.05	4.65E-03
GO:0007017 microtubule-based process	35	2.71	1.31E-04	253	2.01	5.17E-03
GO:0006464 protein modification process	136	10.54	1.33E-04	1453	1.36	5.18E-03
GO:0019222 regulation of metabolic process	299	23.18	1.42E-04	3621	1.20	5.43E-03
GO:0034622 cellular macromolecular complex assembly	41	3.18	1.50E-04	318	1.87	5.67E-03
GO:0051649 establishment of localization in cell	87	6.74	1.78E-04	852	1.48	6.66E-03
GO:0033043 regulation of organelle organization	31	2.40	1.94E-04	217	2.07	7.16E-03
GO:0010605 negative regulation of macromolecule metabolic process	77	5.97	1.99E-04	734	1.52	7.28E-03
GO:0006325 chromatin organization	46	3.57	2.21E-04	378	1.77	7.98E-03
GO:0016568 chromatin modification	36	2.79	2.85E-04	274	1.91	1.01E-02
GO:0032268 regulation of cellular protein metabolic process	54	4.19	3.07E-04	474	1.65	1.08E-02
GO:0051641 cellular localization	92	7.13	3.20E-04	928	1.44	1.11E-02
GO:0006508 proteolysis	102	7.91	3.50E-04	1054	1.41	1.20E-02
GO:0031323 regulation of cellular metabolic process	284	22.02	3.90E-04	3464	1.19	1.32E-02
GO:0031145 anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	14	1.09	4.09E-04	65	3.13	1.37E-02
GO:0000226 microtubule cytoskeleton organization	23	1.78	4.46E-04	147	2.27	1.48E-02
GO:0009892 negative regulation of metabolic process	79	6.12	4.81E-04	780	1.47	1.58E-02
GO:0006261 DNA-dependent DNA replication	13	1.01	4.89E-04	58	3.26	1.59E-02
GO:0006414 translational elongation	18	1.40	4.95E-04	101	2.59	1.59E-02
GO:0009411 response to UV	13	1.01	5.76E-04	59	3.20	1.83E-02
GO:0051329 interphase of mitotic cell cycle	18	1.40	6.24E-04	103	2.54	1.96E-02
GO:0051437 positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle	14	1.09	6.45E-04	68	2.99	2.00E-02
GO:0031570 DNA integrity checkpoint	12	0.93	6.83E-04	52	3.35	2.10E-02
GO:0051443 positive regulation of ubiquitin-protein ligase activity	14	1.09	8.60E-04	70	2.90	2.61E-02
GO:0051325 interphase	18	1.40	8.74E-04	106	2.47	2.62E-02
GO:0031396 regulation of protein ubiquitination	17	1.32	1.26E-03	100	2.47	3.71E-02
GO:0010608 posttranscriptional regulation of gene expression	28	2.17	1.32E-03	211	1.93	3.82E-02
GO:0007088 regulation of mitosis	12	0.93	1.31E-03	56	3.11	3.83E-02
GO:0051783 regulation of nuclear division	12	0.93	1.31E-03	56	3.11	3.83E-02
GO:0000077 DNA damage checkpoint	11	0.85	1.34E-03	48	3.33	3.85E-02
GO:0016044 membrane organization	43	3.33	1.63E-03	381	1.64	4.61E-02

GO:0080135 regulation of cellular response to stress	17	1.32	1.73E-03	103	2.40	4.84E-02

NOTE: Explanation of the individual sections: **Category**: Gene Ontology categories: biological process (BP), cellular compartment (CC) and molecular function (MF); **Term**: Gene set name; **Count**: number of genes associated with this gene set; **Percentag**: gene associated with this gene set/total number of query genes; **P-value**: modified Fisher Exact P-value; **List Total**: number of genes in the query list mapped to any gene set in this ontology; **Pop Hits**: number of genes annotated to this gene set on the background list; **Pop Total**: number of genes on the background list mapped to any gene set in this ontology; **Benjamini**: Benjamini-Hochberg-corrected p-value.

The significantly enriched biological processes are shown (Benjamini-Hochberg-corrected $p<0.05$.Total number of genes in the query list mapped to any gene set in this ontology is 972. Total number of genes on the background list mapped to any gene set in this ontology is 14116.

Table 16: Functional group enrichment analysis showing all enriched cellular compartment Gene Ontology categories.

Term	Count	%	P Value	Pop Hits	Fold Enrichment	Benjamini adjusted *P*
GO:0005622 intracellular	939	72.79	5.73E-63	10995	1.31	3.22E-60
GO:0044424 intracellular part	921	71.40	1.16E-62	10624	1.32	3.26E-60
GO:0043229 intracellular organelle	790	61.24	1.09E-41	8977	1.34	2.03E-39
GO:0043226 organelle	790	61.24	2.10E-41	8989	1.34	2.95E-39
GO:0043227 membrane-bounded organelle	715	55.43	6.68E-36	7989	1.37	7.50E-34
GO:0043231 intracellular membrane-bounded organelle	714	55.35	1.06E-35	7982	1.37	9.94E-34
GO:0044446 intracellular organelle part	454	35.19	6.92E-35	4225	1.64	4.86E-33
GO:0044422 organelle part	456	35.35	6.70E-35	4251	1.64	5.38E-33
GO:0044428 nuclear part	251	19.46	1.18E-32	1822	2.11	7.34E-31
GO:0005634 nucleus	504	39.07	1.32E-30	5077	1.52	7.40E-29
GO:0005737 cytoplasm	651	50.47	8.87E-29	7319	1.36	4.53E-27
GO:0070013 intracellular organelle lumen	231	17.91	4.54E-26	1779	1.98	2.13E-24
GO:0031981 nuclear lumen	200	15.50	1.38E-25	1450	2.11	5.97E-24
GO:0031974 membrane-enclosed lumen	235	18.22	5.26E-25	1856	1.93	2.11E-23
GO:0043233 organelle lumen	231	17.91	1.09E-24	1820	1.94	4.10E-23
GO:0032991 macromolecular complex	332	25.74	8.17E-22	3155	1.61	2.87E-20
GO:0043228 non-membrane-bounded organelle	285	22.09	5.91E-21	2596	1.68	1.95E-19
GO:0043232 intracellular non-membrane-bounded organelle	285	22.09	5.91E-21	2596	1.68	1.95E-19
GO:0005654 nucleoplasm	132	10.23	1.01E-19	882	2.29	3.16E-18
GO:0005829 cytosol	164	12.71	7.53E-16	1330	1.88	2.30E-14
GO:0005623 cell	1024	79.38	1.04E-15	14827	1.06	2.68E-14
GO:0044464 cell part	1024	79.38	9.99E-16	14826	1.06	2.81E-14
GO:0030529 ribonucleoprotein complex	83	6.43	3.41E-14	515	2.46	8.71E-13
GO:0044444 cytoplasmic part	431	33.41	5.52E-14	4895	1.35	1.35E-12
GO:0043234 protein complex	252	19.53	8.87E-12	2588	1.49	2.08E-10
GO:0044451 nucleoplasm part	78	6.05	2.23E-10	555	2.15	5.01E-09
GO:0005694 chromosome	66	5.12	2.85E-09	460	2.19	6.16E-08
GO:0005681 spliceosome	30	2.33	5.69E-09	132	3.47	1.19E-07
GO:0005730 nucleolus	86	6.67	1.59E-08	698	1.88	3.18E-07
GO:0044427 chromosomal part	56	4.34	3.61E-08	386	2.22	6.99E-07
GO:0015630 microtubule cytoskeleton	71	5.50	5.12E-08	549	1.98	9.60E-07
GO:0016607 nuclear speck	23	1.78	6.50E-07	103	3.41	1.18E-05
GO:0005819 spindle	28	2.17	8.97E-07	147	2.91	1.58E-05
GO:0016604 nuclear body	30	2.33	1.36E-06	168	2.73	2.32E-05
GO:0031967 organelle envelope	72	5.58	2.49E-06	620	1.77	4.12E-05
GO:0031975 envelope	72	5.58	2.80E-06	622	1.77	4.50E-05
GO:0005739 mitochondrion	110	8.53	3.75E-06	1087	1.55	5.86E-05
GO:0005813	33	2.56	2.44E-05	224	2.25	3.70E-04

centrosome						
GO:0048471 perinuclear region of cytoplasm	39	3.02	2.77E-05	288	2.07	4.09E-04
GO:0005643 nuclear pore	17	1.32	4.29E-05	79	3.29	6.18E-04
GO:0005815 microtubule organizing center	35	2.71	4.96E-05	253	2.11	6.96E-04
GO:0005856 cytoskeleton	126	9.77	9.28E-05	1381	1.39	1.27E-03
GO:0005840 ribosome	30	2.33	1.62E-04	215	2.13	2.17E-03
GO:0044429 mitochondrial part	62	4.81	3.07E-04	595	1.59	4.00E-03
GO:0046930 pore complex	17	1.32	4.10E-04	95	2.73	5.23E-03
GO:0019866 organelle inner membrane	39	3.02	4.54E-04	329	1.81	5.53E-03
GO:0005743 mitochondrial inner membrane	37	2.87	4.48E-04	306	1.85	5.58E-03
GO:0030530 heterogeneous nuclear ribonucleoprotein complex	7	0.54	5.09E-04	17	6.29	5.95E-03

NOTE: Explanation of the individual sections: **Category**: Gene Ontology categories: biological process (BP), cellular compartment (CC) and molecular function (MF); **Term**: Gene set name; **Count**: number of genes associated with this gene set; **Percentag**: gene associated with this gene set/total number of query genes; **P-value**: modified Fisher Exact P-value; **List Total**: number of genes in the query list mapped to any gene set in this ontology; **Pop Hits**: number of genes annotated to this gene set on the background list; **Pop Total**: number of genes on the background list mapped to any gene set in this ontology; **Benjamini**: Benjamini-Hochberg-corrected p-value.

The significantly enriched cellular compartments are shown (Benjamini-Hochberg-corrected $p<0.05$.Total number of genes in the query list mapped to any gene set in this ontology is 1041. Total number of genes on the background list mapped to any gene set in this ontology is 15908.

Table 17: Functional group enrichment analysis showing all enriched molecular function Gene Ontology categories.

Term	Count	%	P Value	Pop Hits	Fold Enrichment	Benjamini Adjusted *P*
GO:0005515 protein binding	703	54.50	1.09E-19	8154	1.24	1.07E-16
GO:0000166 nucleotide binding	238	18.45	2.17E-12	2245	1.53	1.07E-09
GO:0003723 RNA binding	101	7.83	8.28E-12	718	2.03	2.72E-09
GO:0005488 binding	939	72.79	1.54E-10	12531	1.08	3.78E-08
GO:0003676 nucleic acid binding	302	23.41	8.60E-09	3264	1.34	1.69E-06
GO:0017076 purine nucleotide binding	190	14.73	1.47E-07	1918	1.43	2.42E-05
GO:0032553 ribonucleotide binding	183	14.19	1.77E-07	1836	1.44	2.49E-05
GO:0032555 purine ribonucleotide binding	183	14.19	1.77E-07	1836	1.44	2.49E-05
GO:0003824 catalytic activity	434	33.64	6.11E-07	5198	1.21	7.51E-05
GO:0001883 purine nucleoside binding	159	12.33	1.80E-06	1601	1.43	1.77E-04
GO:0017111 nucleoside-triphosphatase activity	85	6.59	2.07E-06	728	1.69	1.85E-04
GO:0001882 nucleoside binding	160	12.40	1.71E-06	1612	1.43	1.87E-04
GO:0005524 ATP binding	148	11.47	2.74E-06	1477	1.45	2.25E-04
GO:0032559 adenyl ribonucleotide binding	149	11.55	3.58E-06	1497	1.44	2.71E-04
GO:0030554 adenyl nucleotide binding	155	12.02	4.53E-06	1577	1.42	3.18E-04
GO:0016462 pyrophosphatase activity	86	6.67	5.28E-06	757	1.64	3.46E-04
GO:0016818 hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	86	6.67	6.26E-06	760	1.63	3.85E-04
GO:0016817 hydrolase activity, acting on acid anhydrides	86	6.67	7.68E-06	764	1.62	4.44E-04
GO:0003712 transcription cofactor activity	49	3.80	1.08E-05	363	1.95	5.89E-04
GO:0008094 DNA-dependent ATPase activity	15	1.16	2.45E-05	57	3.80	1.27E-03
GO:0051082 unfolded protein binding	22	1.71	3.64E-05	115	2.76	1.79E-03
GO:0008134 transcription factor binding	61	4.73	4.11E-05	513	1.72	1.92E-03
GO:0016879 ligase activity, forming carbon-nitrogen bonds	34	2.64	5.62E-05	231	2.12	2.51E-03
GO:0016881 acid-amino acid ligase activity	30	2.33	1.30E-04	201	2.15	5.55E-03
GO:0004386 helicase activity	23	1.78	2.42E-04	140	2.37	9.88E-03
GO:0016874 ligase activity	46	3.57	4.91E-04	390	1.70	1.84E-02
GO:0016887 ATPase activity	41	3.18	4.73E-04	334	1.77	1.85E-02
GO:0003735 structural constituent of ribosome	25	1.94	5.54E-04	168	2.15	2.00E-02
GO:0008234 cysteine-type peptidase activity	22	1.71	6.91E-04	141	2.25	2.40E-02
GO:0031202 RNA splicing factor activity, transesterification mechanism	8	0.62	9.07E-04	24	4.81	3.03E-02
GO:0019787 small conjugating protein ligase activity	24	1.86	1.10E-03	166	2.09	3.53E-02
GO:0042623 ATPase activity, coupled	34	2.64	1.14E-03	272	1.80	3.56E-02
GO:0004523 ribonuclease H activity	5	0.39	1.27E-03	8	9.02	3.84E-02

NOTE: Explanation of the individual sections: **Category**: Gene Ontology categories: biological process (BP), cellular compartment (CC) and molecular function (MF); **Term**: Gene set name; **Count**: number of genes associated with this gene set; **Percentag**: gene associated with this gene set/total number of query genes; **P-value**: modified Fisher Exact P-value; **List Total**: number of genes in the query list mapped to any gene set in this ontology; **Pop Hits**: number of genes annotated to this gene set on the background list; **Pop Total**: number of genes on the background list mapped to any gene set in this ontology; **Benjamini**: Benjamini-Hochberg-corrected p-value.

The significantly enriched molecular finctions are shown (Benjamini-Hochberg-corrected p<0.05.Total number of genes in the query list mapped to any gene set in this ontology is 1049. Total number of genes on the background list mapped to any gene set in this ontology is 15143.

Appendix 3

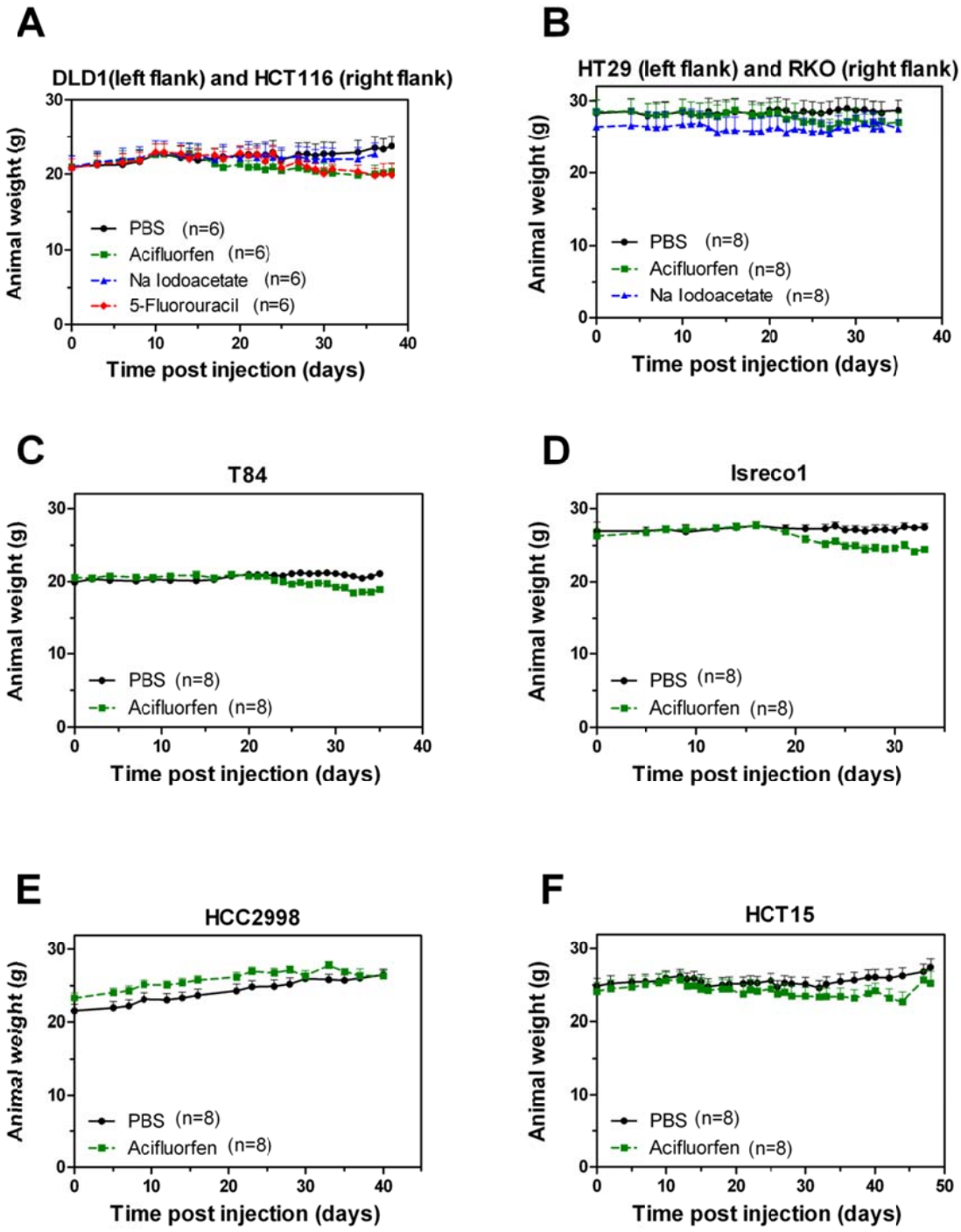

Figure 36: Animals weight of figures treated with PBS, 5-FU, Na Iodoacetate or acifluorfen.
Mice were injected with A) DLD1 (left flank) and HCT116 cells (right flank), B) HT29 (left flank) and RKO cells (right flank), C) T84 cells, D) Isreco1 cells, E) HCC2998 cells, F) HCT15 cells.

Appendix 4

Table 18: All L-shaped correlated genes in cell lines and TCGA tumor samples.

		Cell lines			TCGA tumors		
Gene symbol	Gene name	Pearson r	Pearson p	BH, FDR	Pearson r	Pearson p	BH, FDR
CYP4X1	Cytochrome P450, family 4, subfamily X, polypeptide 1	-0.361	4.94E-02	5.12E-02	-0.191	4.37E-03	1.19E-02
RNF186	Ring finger protein 186	-0.350	5.73E-02	5.92E-02	-0.056	4.06E-01	4.93E-01
KCNK15	Potassium channel, subfamily K, member 15	-0.348	5.92E-02	6.10E-02	-0.012	8.57E-01	8.94E-01
C14orf50	Chromosome 14 open reading frame 50	-0.347	5.93E-02	6.11E-02	-0.065	3.36E-01	4.48E-01
IGF2BP1	Insulin-like growth factor 2 mRNA binding protein 1	-0.342	6.33E-02	6.51E-02	-0.190	4.46E-03	1.19E-02
ATP6V1C2	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C2	-0.307	9.87E-02	1.01E-01	-0.348	1.00E-07	1.21E-06
ADAMTS15	ADAM metallopeptidase with thrombospondin type 1 motif, 15	-0.302	1.04E-01	1.07E-01	-0.055	4.11E-01	4.93E-01
HSD17B2	Hydroxysteroid (17-beta) dehydrogenase 2	-0.302	1.04E-01	1.07E-01	-0.077	2.51E-01	4.25E-01
ST6GALNAC3	In multiple clusters	-0.300	1.07E-01	1.09E-01	-0.222	8.49E-04	5.09E-03
PLAU	Plasminogen activator, urokinase	-0.297	1.11E-01	1.13E-01	0.000	1.00E+00	1.00E+00
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	-0.292	1.17E-01	1.19E-01	-0.082	2.26E-01	4.18E-01
DPYSL3	Dihydropyrimidinase-like 3	-0.287	1.23E-01	1.25E-01	-0.065	3.35E-01	4.48E-01
SNAI2	Snail homolog 2 (Drosophila)	-0.284	1.28E-01	1.30E-01	0.047	4.86E-01	5.30E-01
DLL1	Delta-like 1 (Drosophila)	-0.272	1.45E-01	1.47E-01	-0.051	4.46E-01	5.09E-01
EPHB6	EPH receptor B6	-0.268	1.52E-01	1.54E-01	0.075	2.66E-01	4.25E-01
GPR87	G protein-coupled receptor 87	-0.249	1.84E-01	1.86E-01	0.069	3.06E-01	4.48E-01
DACH1	Dachshund homolog 1 (Drosophila)	-0.246	1.90E-01	1.92E-01	-0.210	1.62E-03	7.66E-03
GALNT5	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)	-0.245	1.91E-01	1.93E-01	-0.265	6.47E-05	5.17E-04
IL7	Interleukin 7	-0.244	1.93E-01	1.95E-01	-0.130	5.24E-02	1.26E-01
FHOD3	Formin homology 2 domain containing 3	-0.238	2.04E-01	2.06E-01	-0.423	4.61E-11	1.11E-09
PDZRN3	PDZ domain containing ring finger 3	-0.226	2.28E-01	2.29E-01	-0.207	1.92E-03	7.66E-03
MX1	Myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	-0.208	2.69E-01	2.70E-01	-0.101	1.33E-01	2.66E-01
FRMD6	FERM domain containing 6	-0.163	3.88E-01	3.88E-01	-0.107	1.12E-01	2.45E-01
SULT2A1	Sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1	-0.143	4.50E-01	4.50E-01	-0.196	3.28E-03	1.12E-02

Appendix 5

Table 19: 100 most significant correlation genes (correlation between methylation and expression levels) in cell lines and TCGA tumor samples.

		Cell lines			TCGA tumors		
Gene symbol	Gene name	Pearson r	Pearson p	BH, FDR	Pearson r	Pearson p	BH, FDR
FNBP1	Formin binding protein 1	-0.8326	7.04E-09	4.20E-06	-0.2138	1.35E-03	2.81E-03
ZNF141	Zinc finger protein 141	-0.8254	1.24E-08	4.20E-06	-0.6084	6.62E-24	2.86E-22
PYCARD	PYD and CARD domain containing	-0.8097	3.95E-08	8.91E-06	-0.2469	2.03E-04	5.57E-04
MRPS21	Mitochondrial ribosomal protein S21	-0.8053	5.33E-08	9.02E-06	-0.6718	1.43E-30	1.15E-28
KLHL3	Kelch-like 3 (Drosophila)	-0.7751	3.57E-07	4.83E-05	-0.7149	3.99E-36	5.29E-34
GIPC2	GIPC PDZ domain containing family, member 2	-0.7701	4.74E-07	5.35E-05	-0.4667	2.03E-13	2.65E-12
BST2	Bone marrow stromal cell antigen 2	-0.7623	7.31E-07	7.07E-05	-0.7999	7.44E-51	8.38E-48
GPR56	G protein-coupled receptor 56	-0.7557	1.04E-06	8.79E-05	-0.3154	1.61E-06	7.57E-06
TFCP2	Transcription factor CP2	-0.7373	2.62E-06	1.97E-04	-0.3591	3.66E-08	2.40E-07
PPP1R14D	Protein phosphatase 1, regulatory (inhibitor) subunit 14D	-0.7222	5.29E-06	3.25E-04	-0.5799	2.24E-21	7.31E-20
C1orf59	Chromosome 1 open reading frame 59	-0.7151	7.24E-06	3.35E-04	-0.8105	3.44E-53	6.37E-50
S100P	S100 calcium binding protein P	-0.7161	6.95E-06	3.35E-04	-0.5770	3.92E-21	1.23E-19
SPAG16	Sperm associated antigen 16	-0.7146	7.43E-06	3.35E-04	-0.6944	2.34E-33	2.29E-31
S100A6	S100 calcium binding protein A6	-0.7122	8.22E-06	3.48E-04	-0.1565	1.96E-02	2.46E-02
ZNF571	Zinc finger protein 571	-0.7099	9.07E-06	3.61E-04	-0.3236	8.28E-07	4.11E-06
STK33	Serine/threonine kinase 33	-0.7041	1.16E-05	4.13E-04	-0.4020	4.90E-10	4.11E-09
ZNF238	Zinc finger protein 238	-0.6865	2.35E-05	7.96E-04	-0.4011	5.37E-10	4.48E-09
IRX2	Iroquois homeobox 2	-0.6848	2.52E-05	8.12E-04	-0.7113	1.27E-35	1.59E-33
C1orf172	Chromosome 1 open reading frame 172	-0.6757	3.56E-05	9.22E-04	-0.2545	1.26E-04	3.70E-04
NNT	Nicotinamide nucleotide transhydrogenase	-0.6788	3.16E-05	9.22E-04	-0.1525	2.30E-02	2.80E-02
RAB25	RAB25, member RAS oncogene family	-0.6749	3.66E-05	9.22E-04	-0.2916	1.00E-05	3.93E-05
RIPK3	Receptor-interacting serine-threonine kinase 3	-0.6783	3.22E-05	9.22E-04	-0.2028	2.39E-03	4.56E-03
RPP25	Ribonuclease P/MRP 25kDa subunit	-0.6747	3.69E-05	9.22E-04	-0.1932	3.85E-03	6.77E-03
TACSTD2	Tumor-associated calcium signal transducer 2	-0.6776	3.31E-05	9.22E-04	-0.2844	1.68E-05	6.25E-05
ZNF420	Zinc finger protein 420	-0.6738	3.81E-05	9.22E-04	-0.4009	5.50E-10	4.57E-09
CXorf26	Chromosome X open reading frame 26	-0.6721	4.07E-05	9.50E-04	-0.3009	5.00E-06	2.10E-05
FADS1	Fatty acid desaturase 1	-0.6705	4.31E-05	9.72E-04	-0.2305	5.35E-04	1.29E-03
LY75	Lymphocyte antigen 75	-0.6682	4.68E-05	1.02E-03	-0.4804	3.11E-14	4.57E-13
S100A4	S100 calcium binding protein A4	-0.6638	5.48E-05	1.10E-03	-0.6009	3.20E-23	1.32E-21
ZNF512	Zinc finger protein 512	-0.6628	5.69E-05	1.10E-03	-0.5910	2.47E-22	9.14E-21
PRTFDC1	Phosphoribosyl transferase domain containing 1	-0.6617	5.91E-05	1.11E-03	-0.4438	3.86E-12	4.28E-11
LYZ	Lysozyme	-0.6560	7.21E-05	1.32E-03	-0.2949	7.82E-06	3.14E-05
MTERF	Mitochondrial transcription termination factor	-0.6547	7.54E-05	1.34E-03	-0.7933	1.86E-49	1.67E-46
GSTP1	Glutathione S-transferase pi 1	-0.6519	8.29E-05	1.40E-03	-0.6411	3.79E-27	2.25E-25
REPIN1	Replication initiator 1	-0.6520	8.26E-05	1.40E-03	-0.3394	2.16E-07	1.23E-06

C17orf73	Chromosome 17 open reading frame 73	-0.6476	9.60E-05	1.57E-03	-0.1345	4.53E-02	4.66E-02
ZNF71	Zinc finger protein 71	-0.6472	9.72E-05	1.57E-03	-0.6426	2.64E-27	1.61E-25
LXN	Latexin	-0.6418	1.16E-04	1.83E-03	-0.3816	4.10E-09	3.08E-08
KRTCAP3	Keratinocyte associated protein 3	-0.6405	1.21E-04	1.87E-03	-0.1729	9.86E-03	1.44E-02
KRT20	Keratin 20	-0.6378	1.32E-04	1.95E-03	-0.3519	7.12E-08	4.52E-07
PRAF2	PRA1 domain family, member 2	-0.6382	1.31E-04	1.95E-03	-0.3588	3.75E-08	2.46E-07
KCNE3	Potassium voltage-gated channel, Isk-related family, member 3	-0.6362	1.39E-04	1.99E-03	-0.5674	2.40E-20	6.71E-19
TNS4	Tensin 4	-0.6358	1.41E-04	1.99E-03	-0.3245	7.68E-07	3.84E-06
HNF4A	Hepatocyte nuclear factor 4, alpha	-0.6307	1.66E-04	2.04E-03	-0.4721	9.79E-14	1.32E-12
HOXB6	Homeobox B6	-0.6313	1.63E-04	2.04E-03	-0.4745	7.00E-14	9.86E-13
MT2A	Metallothionein 2A	-0.6334	1.53E-04	2.04E-03	-0.1901	4.47E-03	7.64E-03
SGK2	Serum/glucocorticoid regulated kinase 2	-0.6303	1.69E-04	2.04E-03	-0.6435	2.14E-27	1.34E-25
TMEM125	Transmembrane protein 125	-0.6337	1.51E-04	2.04E-03	-0.1366	4.20E-02	4.41E-02
ZNF502	Zinc finger protein 502	-0.6332	1.54E-04	2.04E-03	-0.7183	1.31E-36	1.97E-34
ZNF215	Zinc finger protein 215	-0.6294	1.73E-04	2.06E-03	-0.5047	9.04E-16	1.56E-14
SERPINB1	Serpin peptidase inhibitor, clade B (ovalbumin), member 1	-0.6257	1.95E-04	2.23E-03	-0.3235	8.33E-07	4.13E-06
RPH3AL	Rabphilin 3A-like (without C2 domains)	-0.6248	2.00E-04	2.25E-03	-0.4411	5.38E-12	5.84E-11
ZNF350	Zinc finger protein 350	-0.6232	2.10E-04	2.33E-03	-0.7164	2.44E-36	3.33E-34
PON3	Paraoxonase 3	-0.6221	2.17E-04	2.33E-03	-0.4004	5.81E-10	4.80E-09
ZNF570	Zinc finger protein 570	-0.6221	2.17E-04	2.33E-03	-0.6068	9.34E-24	3.90E-22
PAX6	Paired box 6	-0.6209	2.25E-04	2.37E-03	-0.1783	7.75E-03	1.19E-02
ZNF566	Zinc finger protein 566	-0.6205	2.28E-04	2.37E-03	-0.4094	2.17E-10	1.92E-09
HEPH	Hephaestin	-0.6169	2.54E-04	2.51E-03	-0.3062	3.31E-06	1.46E-05
MAPK13	Mitogen-activated protein kinase 13	-0.6173	2.51E-04	2.51E-03	-0.3250	7.36E-07	3.70E-06
SLC26A11	Solute carrier family 26, member 11	-0.6167	2.56E-04	2.51E-03	-0.1521	2.34E-02	2.84E-02
RUSC2	RUN and SH3 domain containing 2	-0.6152	2.68E-04	2.59E-03	-0.4642	2.81E-13	3.59E-12
HOXB3	Homeobox B3	-0.6136	2.80E-04	2.67E-03	-0.4633	3.16E-13	4.00E-12
ADAM8	ADAM metallopeptidase domain 8	-0.6101	3.11E-04	2.93E-03	0.1916	4.16E-03	7.22E-03
BOLA1	BolA homolog 1 (E. coli)	-0.6064	3.46E-04	2.94E-03	-0.2707	4.36E-05	1.46E-04
GUCY2C	Guanylate cyclase 2C (heat stable enterotoxin receptor)	-0.6064	3.47E-04	2.94E-03	-0.3482	9.91E-08	6.09E-07
NES	Nestin	-0.6064	3.47E-04	2.94E-03	-0.2705	4.43E-05	1.48E-04
PDZK1	PDZ domain containing 1	-0.6087	3.24E-04	2.94E-03	-0.4461	2.89E-12	3.28E-11
STAP2	Signal transducing adaptor family member 2	-0.6067	3.43E-04	2.94E-03	-0.3386	2.31E-07	1.31E-06
TGFBI	Transforming growth factor, beta-induced, 68kDa	-0.6088	3.23E-04	2.94E-03	-0.4901	7.85E-15	1.22E-13
ZNF550	Zinc finger protein 550	-0.6053	3.57E-04	2.99E-03	-0.7039	1.28E-34	1.41E-32
FAM111A	Family with sequence similarity 111, member A	-0.6018	3.95E-04	3.18E-03	-0.2741	3.45E-05	1.18E-04
HOXB5	Homeobox B5	-0.6024	3.89E-04	3.18E-03	-0.3675	1.65E-08	1.14E-07
ZNF14	Zinc finger protein 14	-0.6021	3.92E-04	3.18E-03	-0.6781	2.59E-31	2.20E-29
FABP1	Fatty acid binding protein 1, liver	-0.6001	4.15E-04	3.31E-03	-0.4839	1.90E-14	2.85E-13
IRF6	Interferon regulatory factor 6	-0.5985	4.34E-04	3.34E-03	-0.1646	1.41E-02	1.90E-02
NETO2	Neuropilin (NRP) and tolloid (TLL)-like 2	-0.5987	4.32E-04	3.34E-03	-0.3292	5.18E-07	2.71E-06
PRDM5	In multiple clusters	-0.5986	4.33E-04	3.34E-03	-0.4327	1.49E-11	1.51E-10
FBP1	Fructose-1,6-bisphosphatase 1	-0.5955	4.72E-04	3.59E-03	-0.3163	1.50E-06	7.11E-06
PLEK2	Pleckstrin 2	-0.5924	5.15E-04	3.79E-03	-0.1353	4.41E-02	4.56E-02

ZNF655	Zinc finger protein 655	-0.5927	5.10E-04	3.79E-03	-0.7227	3.07E-37	5.54E-35
DBN1	Drebrin 1	-0.5911	5.33E-04	3.88E-03	-0.3242	7.87E-07	3.93E-06
ANXA3	Annexin A3	-0.5878	5.84E-04	3.99E-03	-0.3759	7.27E-09	5.26E-08
C6orf150	Chromosome 6 open reading frame 150	-0.5894	5.59E-04	3.99E-03	-0.4361	9.88E-12	1.04E-10
GALNT11	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)	-0.5886	5.72E-04	3.99E-03	-0.4341	1.26E-11	1.29E-10
SHC3	SHC (Src homology 2 domain containing) transforming protein 3	-0.5884	5.75E-04	3.99E-03	-0.4628	3.38E-13	4.25E-12
VAMP5	Vesicle-associated membrane protein 5 (myobrevin)	-0.5881	5.80E-04	3.99E-03	-0.1740	9.39E-03	1.38E-02
C10orf58	In multiple clusters	-0.5858	6.16E-04	4.13E-03	-0.2770	2.83E-05	9.94E-05
ARHGEF10	Rho guanine nucleotide exchange factor (GEF) 10	-0.5841	6.45E-04	4.17E-03	-0.5437	1.67E-18	3.86E-17
HTR4	5-hydroxytryptamine (serotonin) receptor 4	-0.5841	6.46E-04	4.17E-03	-0.2344	4.29E-04	1.07E-03
ITGA9	Integrin, alpha 9	-0.5840	6.47E-04	4.17E-03	-0.3945	1.09E-09	8.77E-09
LMCD1	LIM and cysteine-rich domains 1	-0.5849	6.31E-04	4.17E-03	-0.3281	5.69E-07	2.95E-06
MYO1A	Myosin IA	-0.5822	6.80E-04	4.23E-03	-0.4100	2.05E-10	1.83E-09
SNX19	Sorting nexin 19	-0.5823	6.77E-04	4.23E-03	-0.2944	8.15E-06	3.25E-05
TNFRSF1B	Tumor necrosis factor receptor superfamily, member 1B	-0.5806	7.08E-04	4.36E-03	-0.3444	1.39E-07	8.19E-07
EPM2AIP1	EPM2A (laforin) interacting protein 1	-0.5785	7.50E-04	4.57E-03	-0.7740	1.10E-45	7.11E-43
CKB	Creatine kinase, brain	-0.5774	7.71E-04	4.66E-03	-0.2184	1.05E-03	2.29E-03
C10orf116	Chromosome 10 open reading frame 116	-0.5760	8.01E-04	4.76E-03	-0.4048	3.62E-10	3.09E-09
CST7	Cystatin F (leukocystatin)	-0.5752	8.16E-04	4.76E-03	-0.3104	2.39E-06	1.08E-05
MST1R	Macrophage stimulating 1 receptor (c-met-related tyrosine kinase)	-0.5750	8.22E-04	4.76E-03	-0.2660	5.95E-05	1.92E-04