University of Insubria
Department of Medicine and Surgery
PhD in Experimental and Translational Medicine
XXX cycle
Coordinator: Prof. Daniela Negrini

DNA HYPOMETHYLATION AND HYPERMETHYLATION IN COLORECTAL CANCER INITIATION

Tutor: Prof. Daniela Furlan

PhD candidate:
Dr. Davide Trapani

Academic year: 2017-2018

INDEX

INTRODUCTION

1. COLORECTAL CANCER (CRC)

1.1 EPIDEMIOLOGY AND NATURAL HISTORY

Colorectal cancer (CRC) is the third most common tumour in men and the second in women, accounting for 10% of all tumour types worldwide. Incidence is higher in males (ratio: 1.4) and for both genders there is a 10-fold difference in incidence between several regions. With 608000 deaths estimated each year (about 8% of all cancer deaths), CRC is the fourth most common cancer-related cause of death in the world. However, mortality has declined progressively in many Western countries: this can be attributed to cancer screening programs, removal of adenomas, early detection of cancerous lesions and availability of more effective therapies, chiefly for early stage disease (1).

CRC results from the accumulation of genetic and epigenetic alterations, as well as biochemical changes in the macro- and microenvironment, which lead to the transformation of normal colonic epithelium into benign adenomas and eventually adenocarcinoma, known as the adenoma-carcinoma sequence (2-3).
The earliest identifiable lesion of the colonic mucosa is the aberrant crypt focus (ACF), characterized by histological features including darker staining, raised appearance, and crypt size at least three times larger than adjacent normal mucosa (4). Individual cells in ACF are morphologically normal, however the thickened layer of cells within crypts leads to crowding and mucosal folding. A subset of these microscopic mucosal abnormalities are believed to be the precursors of CRC. Some of these lesions show mutations in *APC* or *KRAS* or *BRAF* genes and have increased expression of proliferative markers (5-8).

Polyps are benign gland-forming mucosal projections, which can be categorized as adenomatous or serrated. Adenomatous polyps, or adenomas, may be tubular, tubulovillous, or villous adenomas (9-10).
The serrated pathway is distinct from the conventional adenoma-carcinoma pathway, characterized by serrated architecture of the epithelial compartment (9,11).
Hyperplastic polyps are the most common type of serrated lesion and are generally less than 5 mm in size, found in the distal colon, and rarely progress to CRC (12).

1.2 RISK FACTORS

It is well known that the risk of CRC is associated with personal predisposing features or environmental factors (13). The most significant contributory agents include:

- genetic factors, indeed, the greatest risk factor for CRC is family history (14). Several hereditary CRC and polyposis syndromes have been characterized. The hereditary CRC syndromes are classified based upon the clinical presence or absence of colonic polyps as a major disease manifestation and the presence of known causative genetic mutation (15). However, only 5-10% of CRCs can be attributed to germline mutations in highly penetrant genes associated with hereditary CRC syndromes, leaving a proportion of CRC risk that may be associated to familial predisposition due to germline variants in currently unidentified genes. Epidemiological studies demonstrated that an individual's risk of CRC doubles if one first-degree relative also has CRC, and quadruples if two first-degree relatives are affected by CRC (16-17). Currently, it is widely accepted that the accumulation of a number of low-risk, low-penetrance alleles and a combination of genetic and environmental factors may contribute to significant CRC risk (18-21);
- age is considered an important risk factor for sporadic colon cancer: nearly 70% of patients with colon cancer are over 65 years of age. CRC is rare before 40 years even if data from SEER (Surveillance, Epidemiology, and End Results) and Western registries show an increasing incidence in the 40–44 years group (22);
- chronic disease history: chronic inflammation diseases, such as Crohn's bowel syndrome or ulcerative colitis can induce colorectal carcinogenesis and the risk increases further with early age at diagnosis of inflammatory bowel disease, duration of symptoms and severity of the inflammation. In this context, the gut microbiota and dysbiosis situations may have a crucial role (23-25);
- presence of adenomas: there is a high risk of CRC in individuals whose adenomas are not removed (26). These lesions are typically asymptomatic and are often found incidentally during colonoscopy performed for unrelated symptoms or for CRC screening. Although not all colonic polyps are adenomas and more than 90% of adenomas do not progress to cancer, a differential diagnosis that takes into account the various types of colorectal polyps and the accurate identification of those that will progress to cancer remain challenging. High-risk features of colonic polyps are currently based on size, number, and pathologic characteristics. More specifically,

multiple adenomas, lesions that are larger than 1 cm in size, villous or tubulovillous histology and high grade dysplasia are considered important risk factors for CRC development (10). Some studies only recently suggested that adenomas with somatic *KRAS* or *BRAF* mutations, might provide information about the risk of developing metachronous advanced neoplasia during follow-up for patients diagnosed with polyps (27);

- lifestyle: smoking, environmental factors such as a diet rich in unsaturated fats and red meat, excessive alcohol consumption, and reduced physical activity predispose to CRC (28-31). Unfortunately, specifically dietary and environmental factors that contribute to CRC in Western countries are not particularly well defined, and the majority of CRCs arise in people with poorly defined risk profiles.

1.3 SCREENING, STAGING AND PROGNOSIS

Several screening strategies exist for asymptomatic persons at average risk for developing CRC. These strategies have allowed for earlier detection at more curable stages and have resulted in reduced mortality rates (32-35). The guaiac fecal occult blood test (gFOBT) is a non-invasive procedure able to detect small amounts of blood in stool.

Clinical trials have demonstrated that individuals with positive occult blood tests have three to four times higher risk for developing CRC compared to those with negative tests (36) and that gFOBT reduced CRC mortality by 15-33% (37-40). Actually, gFOBT has low sensitivity for CRC (25-38%) and advanced adenomas (16-31%) (41). For this reason, the gold standard of examination is colonoscopy, which allows for full colonic examination combined with polypectomy in a single session. Studies have shown that colonoscopy reduces CRC incidence by 67-77% and CRC mortality by 31-65% (42-45). This method has the highest level of sensitivity and specificity of any screening method and is the final assessment step of any current screening program (46-47).

The stage at which CRC is diagnosed determines the treatment options and is currently the strongest predictor of survival. The tumor-node-metastasis (TNM) is the most widely used cancer staging system: T refers to the local extent of untreated primary tumour at the time of diagnosis and initial workup; N refers to the status of the regional lymph nodes; M refers to distant metastatic disease. Pathological classification is based on gross/microscopic examination of the resected specimen of an untreated

primary tumour and clinical classification is based on a variety of techniques including physical examination, radiologic imaging, endoscopy, biopsy, and surgical exploration (48). Using the TNM staging system, the five-year disease-specific survival rate of CRC is around 90% for stage I CRCs, 85% for stage II, 70% for stage III, and 10-15% for stage IV disease (49-50).

2. MOLECULAR PATHWAYS IN CRC

CRC has long been considered a single disease process with shared causality, clinical characteristics, and prognosis. Recently, applications of molecular technologies coupled with extensive analysis of precursor lesions and hereditary forms of the disease, have shown that CRC is a heterogeneous and complex disorder that develops as a consequence of accumulation of both genetic and epigenetic alterations. In CRC, there are at least three major pathways contributing to instability: chromosomal instability (CIN), microsatellite instability (MSI) and epigenetic instability, known as the CpG island methylator phenotype (CIMP). The genetic and/or epigenetic alterations common to the progression of these three pathways are shown in Figure 1 (51). These three pathways are not always mutually exclusive. About 25% of MSI CRCs exhibit chromosomal instability and 12% of CIN tumors have high level of MSI (52). CIMP is most often found in MSI positive CRCs, but up to 35% of CIMP positive tumors exhibit high levels of chromosomal abnormalities (53-55). The biological relevance of these overlapping pathways and their role in prognosis are not yet fully understood.

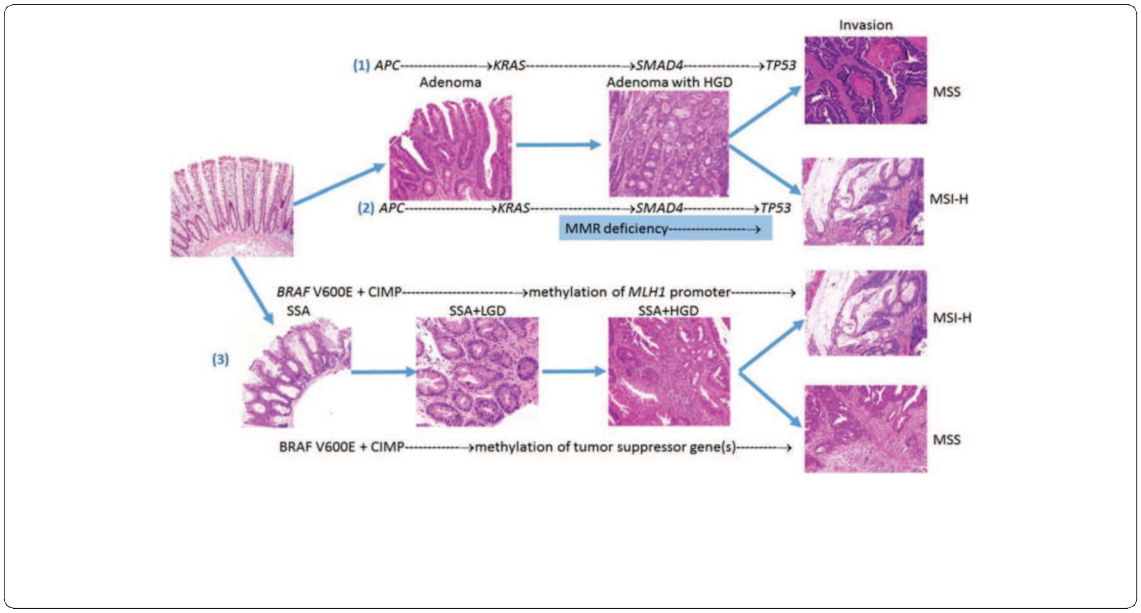


Figure 1. Three major pathways leading to colorectal cancers. (1) Conventional adenoma-carcinoma sequence with oncogene (e.g. *KRAS*) activation and tumor suppressor (e.g. *APC, SMAD4 and TP53*) inactivation, resulting in microsatellite stable (MSS) cancers; (2) Microsatellite instability (MSI) pathway with mismatch repair (MMR) protein, resulting in MSI-high (MSI-H) cancers; (3) Serrated pathway with CpG island methylation phenotype, resulting in either MSI-H cancers if methylation occurs in *MLH1* promoter or MSS cancers if methylation occurs in tumor suppressor genes. HGD: high-grade dysplasia; LGD: low-grade dysplasia; SSA: sessile serrated adenoma (51).

2.1 CIN

The CIN pathway is also known as the adenoma-carcinoma sequence and it follows a predictable progression of genetic and corresponding histologic changes. The inactivation of the *APC* gene is the earliest event associated with adenoma formation, through the activated Wnt-signaling (56-57). The physiological role of APC is to abolish the Wnt signaling cascade when appropriate by binding to a complex of proteins including β-catenin, glycogen synthase kinase-3β (GSK3β), and casein kinase 1 (CK1) on an axis inhibition protein 1 (AXIN1) scaffold. This is known as the β-catenin destruction complex, which facilitates the phosphorylation of β-catenin by GSK3β and CK1 leading to its proteasomal degradation, aided by β-transducin-repeat-containing protein (βTRCP). Mutations of *APC*, occurring in >75% of CRC tumours, increase transcriptional activity of β-catenin (57-58).
Alternatively, mutations can occur in β-catenin (*CTNNB1*) or other Wnt components to bring about the same activation (59-60).

Subsequent *KRAS* mutations occur in about 10% of adenomas but are observed in about 50% of adenomas with high grade dysplasia, whereas deletions (or other alterations) of genes in chromosome 18q affect adenoma growth and progression. The biallelic loss or inactivation of *TP53* and other genetic abnormalities like mutations in *TGFβR* and *PIK3CA* drive the activation of the adenoma-carcinoma transition (61-62). Although according to this multistep model at least 7 distinct mutations are required for tumorigenesis, genome-wide sequencing of CRCs have calculated about 80 mutated genes per tumor, with less than 15 mutations considered to be true drivers (62). Moreover, whether CIN creates the appropriate environment for the accumulation of these mutations or *vice versa* remains unknown.

2.1.1 THE ROLE OF DNA HYPOMETHYLATION IN CRC

The first-described epigenetic change in human cancer was the global DNA hypomethylation in 1983 (63). However, for several decades, it was almost ignored, with attention mainly focused on the hypermethylation of genes that are silenced in cancer.

Different investigations sustain a causal link between DNA hypomethylation and CIN, reporting an association between this epigenetic defect and aneuploidy in human colorectal cancer cell lines (64), changes in DNA methyltransferase activity and loss of imprinting. Moreover, hypomethylation of satellite DNA at the juxtacentromeric heterochromatin and chromatin decondensation was observed in cancer cells (65-66).

This epigenetic event is frequently linked with altereted chromatin structure, changes in DNA methyltransferase activity and loss of imprinting. The resultant aberrant transcription and chromosomal instability is believed to contribute to disease onset or progression and increased tumor frequency and malignancy (Figure 2).

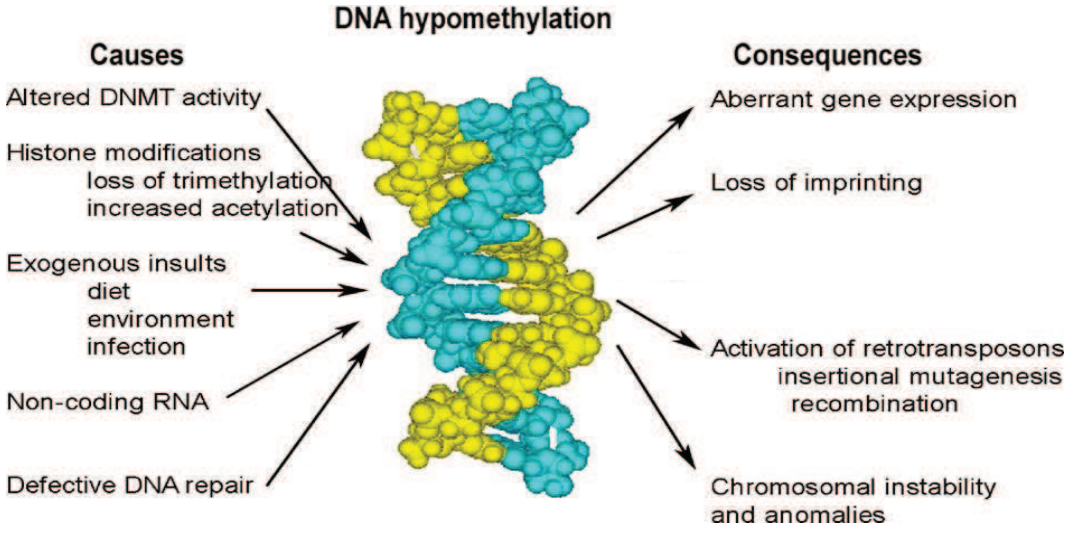


Figure 2. The figure shows potential causes and consequences of DNA hypomethylation (65).

Global DNA hypomethylation in cancer was mostly associated with repeated DNA elements such as Long Interspersed Nucleotide Elements (LINE), and for a long time this epigenetic alteration continued to receive rather little attention. Indeed, due to the lack of any obvious function, LINE have long been regarded as parasitic components of genomes. This view has now been changed for two main reasons: a structural and a functional one. First, sequencing of higher eucaryotic DNAs indicated that large portions of the genomes are formed by retro-elements (45% in humans),

whereas protein-encoding genes do not exceed 2% in the human genome (The-ENCODE-Consortium Nature 2012). Second, LINEs play functional roles during various (physiological) and pathological processes (67). In somatic cells, LINE-1 or L1 (the only active retro-trasposon in humans covering approximately 17% of the entire genome) is actively suppressed by several mechanisms that have been recently described, namely: i) silencing of L1 activity by DNA methylation (68); ii) binding of PIWI RNAs that target L1 transcripts leading their degradation (68); iii) control of L1 activity by SIRT-6, a chromatin-associated protein that is required for normal base excision repair (BER) of DNA damage in mammalian cells (69).

In several type of cancers L1 are strongly hypomethylated and, due to the abundance of these elements in human genome, this is taken as a valid surrogate of genome-wide methylation. A large body of evidence supports that global hypomethylation involving L1 elements is an important cause of their reactivation in cancer (68). In CRC, L1 hypomethylation significantly correlates with various clinico-pahological and molecular variables. L1 hypomethylation can be found early in carcinogenesis since the colonic mucosa from CRC patients has been reported to show lower global methylation levels compared with control individuals (70-74). Karyn L et al. recently have shown that deletion of *Dnmt1* in the adult intestinal epithelium of *Apc* $^{Min/+}$ mice causes accelerated formation of adenomas and results in acute hypomethylation and genomic instability (75). These results confirm the important role of DNA methylation in preserving genomic integrity during intestinal tumorigenesis.
In addition, L1 hypomethylation is inversely correlated with MSI and the CIMP phenotype in CRC (76). Moreover it has been associated with more aggressive progression of CRC and with CRC familial clustering (77). Taking these findings together, LINE-1 methylation level may serve as a potential tumor biomarker for prognostication as well as familial cancer risk assessment (78).

The tumorigenic function of L1 elements occur by both retrotransposition-dependent and independent mechanisms. The former may result in chromosomal rearrangements and target gene inactivation. The latter can exhert epigenetic regulation by generating endo-siRNAs, forming chimeric L1-transcripts (LCT) or changing the expression of adjacent genes by providing alternative promoters or splicing sites (79-80) (Figure 3). In particular, demethylation of the L1 promoter leads to the activation of both sense (SP) and anti-sense (ASP) promoter regions. The SP controls canonical transcription of L1 ORFs, while the ASP drives illegitimate transcription of

neighbouring sequences (68). The role and features of these chimeric transcripts are not fully understood, but they represent an emerging mechanism for the regulation of gene expression.

The high number of L1 elements in the human genome could activate many illegitimate transcripts. *In silico* analyses have revealed up to 911 new putative antisense chimeric transcripts across the genome, but only a few of these L1-containing genes are implicated in cancer (81).

The *MET* oncogene is one of the well-known example of these genes, since the hypomethylation of L1 sequence in its second intron can drive the transcription of a chimeric isoform of the *MET* gene, i.e. L1-*MET* (82). Interestingly, L1-*MET* was reported to be a negative regulator of canonical MET expression *in vitro* (83). Conversely, there are recent reports describing that increased MET protein levels in CRC and in hepatocellular carcinomas are associated with a higher transcription levels of L1-*MET* mRNA (84-85).

The biological roles for L1-*MET* mRNA expression seem controversial, however, these studies suggest that the biological role of L1-*MET* is cellular context dependent. Finally, methylation of the *MET* gene plays an indispensable role modulating its expression in some cancers. It has been shown that the *MET* gene is hypomethylated and thus overexpressed in pancreatic ductal adenocarcinomas (PDAC). Importantly, the hypomethylation status of the *MET* gene is correlated with low overall survival and disease-free survival (86-87).

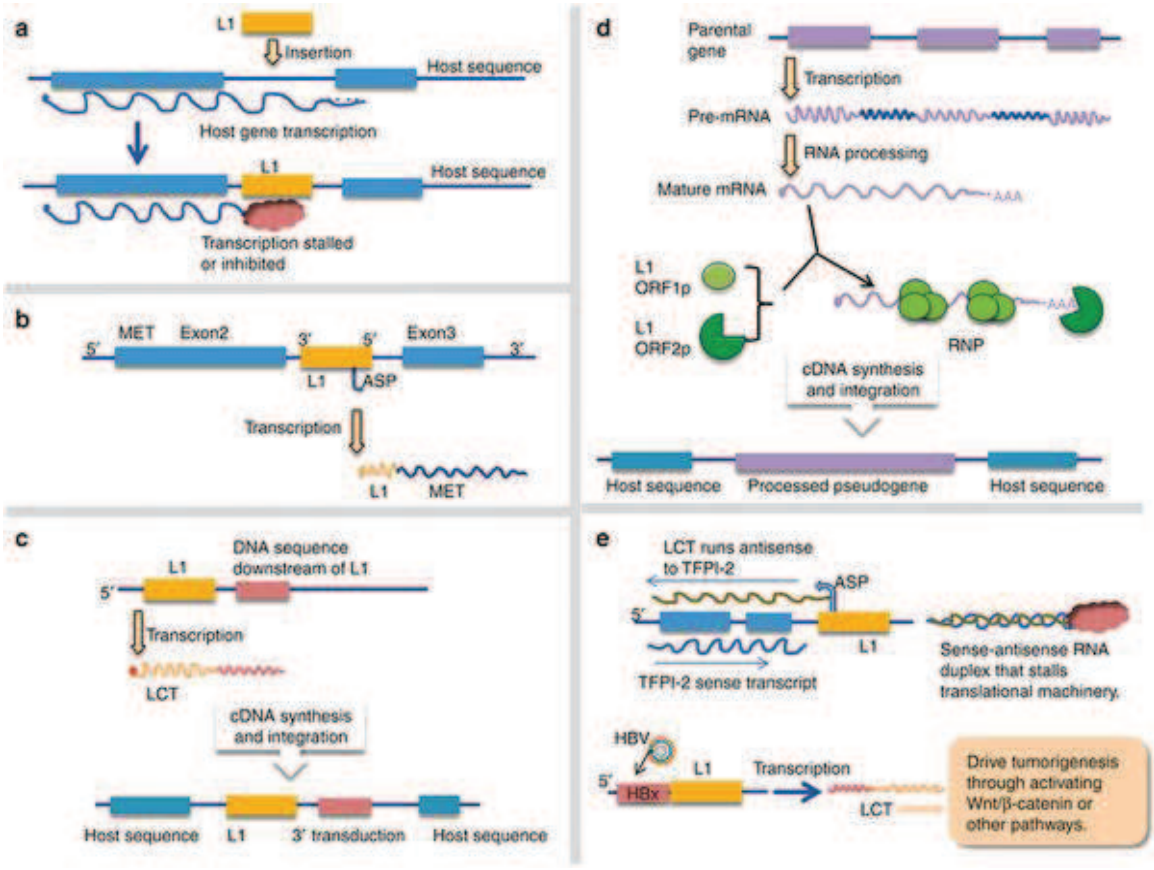


Figure 3. Representation of potential pathogenic functions of L1 in cancer.

(a) L1 insertion mediated inhibition of host gene transcription: L1 can potentially act to slow RNA pol II elongation, dissociate it from the template, or induce premature termination of transcription. (b) L1 insertion–mediated oncogene activation: the ASP within L1 inserted antisense to gene *MET* serves as a transcription start site to drive MET expression. (c) 3'transduction: downstream sequence of L1 3' end is transcribed together with L1 and the resultant LCT is reverse-transcribed and integrated into a new locus by L1 retrotransposition machinery. (d) L1-mediated formation of processed pseudogenes: mature mRNA (lacking introns) is reverse-transcribed and integrated into a new locus by the L1 retrotransposition machinery to generate processed pseudogenes that lack introns and are punctuated by a 3' poly-A tail. (e) Functions of LCTs in cancer: ASP within L1 drives the transcription of LCT that runs antisense to upstream TFPI-2 gene. The expression of TFPI-2 is inhibited by this LCT (upper). HBx from the HBV genome drives LCT that is transcribed partially from HBx and partially from L1 sequence. This LCT functions as an oncogenic long noncoding RNA that can activate Wnt/Functions of LCTs in cancer: ASP within L1 drives the tranare shown for gene *MET* and *TFPI-2*. ASP, antisense promoter; HBV, hepatitis B virus; L1, long interspersed nuclear element-1; LCT, L1 chimeric transcript; ORF1p, protein encoded by L1 open reading frame 1; ORF2p, protein encoded by L1 open reading frame 2; RNP, ribonucleoprotein (68).

2.1.2 DNA HYPOMETHYLATION AND OXIDATIVE STRESS

It has been considered that elevated levels of reactive oxygen species (ROS), down regulation of ROS cleaners, and antioxidant enzymes are related to various tumors (88). Besides, high levels of ROS contribute to tumor development through both genetic and epigenetic mechanism.

Published studies have reported that DNA global demethylation is associated with oxidative DNA damage and several mechanisms have been proposed (89-90). ROS can directly trigger the oxidation of macromolecules within the cell and DNA oxidation involves in a variety of damaged sites.

The most common studied oxdidative DNA lesion was the 8-Oxo-2'-deoxyguanosine (8-OHdG) that nowadays is widely used as a biomarker for oxidative stress.

8-OHdG can also exert an influence on the DNA methylation of nearby cytosine and its presence negatively affect adjacent sites methylation (91-92).

In this context, Turk et al. made an additional discovery describing how the oxidative damage on the nascent strand could suppress DNA methylation through the target cytosine one or two base pairs away from the damaged guanine (93).

Moreover, in an oxidative stress condition with decreased availability of *S*-adenosylmethionine, a depletion of the methyl pool has been shown to cause DNA hypomethylation in folate-deficient models (94).

On the other hand, DNA methylation may also influence the mutagenic effect of ROS. Notably, 8-OHdG may function as a premutagenic base instructing DNA polymerases to incorporate dAMP, a mismatch pairing inducing G→T transversions (95).

Thus, the methylation of Cyt 5′ to 8-OHdG may interfere with efficient repair of this lesion, which partially explains why not only C, but also G in CpG dinucleotides are a hot spot for mutations (96).

8-OHdG is repaired by 8-oxoguanine-DNA glycosylase (OGG1) and methylation of the adjacent cytosine abolished stimulation of OGG1 by repair endonuclease APEX1 (97). Moreover, the exonuclease activity of APEX1 may remove mCyt 5′ to 8-OHdG, providing an intriguing possibility for DNA demethylation coupled with oxidative damage repair.

Finally, other types of oxidative damage could also contribute to DNA demethylation. 5-hydroxymethylcytosine (5hmC) is the result of the hydroxylation of methylcytosine through attack by ROS. Valinluck et al. demonstrated that this

hydroxylation involved in more than 90% reduction in cytosine methylation and these results suggested that DNA methylation could reduce in subsequent rounds of cell division (98). Thus oxidative damage may trigger perturbation.
Additionally, through repair pathways, which considered being responsible for active demethylation in undividing cells, the incorrectly hydroxylated 5mC could be returned to unmethylated cytosine (99).

2.2 MSI

About 15% of sporadic CRCs arise from MSI pathway that is associated with inactivation of DNA mismatch repair (MMR) proteins, leading to mutation accumulation. The MMR machinery consists of a family of enzymes (MLH1, MLH3, MSH2, MSH3, MSH6 and PMS2) with the capability of recognizing and repairing base-base mismatches or insertion-deletion loops (IDL) generated by DNA replication errors acquired during the S-phase of the cell cycle (100). MSI in sporadic CRC is mainly caused by DNA methylation of the cytosine and guanine (CpG)-rich promoter sequences of *MLH1* that causes transcriptional repression and loss of gene function (101-102).

There is a substantial body of literature about the use of *MLH1* methylation and/or MSI as a biomarker in the classification of CRC. However, the biological mechanisms underlying the methylation have not been investigated in detail until recently. Fang et al. have demonstrated in cancer cell lines that *BRAF* oncogenic mutations mediate the CpG island methylation phenotype (CIMP) resulting in hypermethylation at *MLH1* and other CIMP marker genes, via the transcriptional repressor MAFG (103).

Typical for MSI tumours are frameshift mutations in specific genes such as b-catenin, *TGFbRII*, *BAX* (104-105). In MSI tumorigenesis, serrated adenomas are most common and may transform into carcinoma without a component of adenomatous dysplasia (106). However, traditional adenomas may also be observed in MSI pathway. It is well-known that MSI sporadic CRCs display distinct clinico-pathological features such as proximal colonic site, mucinous or signet ring cell type, poor differentiation, presence of infiltrating lymphocytes, frequent *BRAF* mutations and fewer *KRAS* and *TP53* mutations (107).

MSI tumours are also associated with larger tumour size and more favourable outcome at Stage II/III whereas stage IV MSI tumours have a poor prognosis. MSI

CRCs respond poorly to commonly used chemotherapy, such as 5-fluorouracil but are targetable by immune checkpoint inhibitors due to their high number of neoantigens (108-110).

2.3 CIMP

A third pathway through which CRC progresses is the CIMP (111-113). It consists of the aberrant hypermethylation of CpG dinucleotide sequences localized in the promoter regions of tumor suppressor genes, causing their loss of expression. The mechanisms that underlay aberrant DNA methylation in cancer cells have been well studied in the last years and several data suggest that DNA methylation may be an early event in the development of CRC (114). Most CpG islands lack methylation in normal colon mucosa but during CRC initiation and progression, specific hypermethylation events are frequent and affect several signaling pathways, including Wnt, Receptor Tyrosine Kinases, NOTCH, TP53, PI3K, Retinoic Acid, and IGF as well as other pathways involved in cell cycle regulation, transcription regulation, DNA repair/stability, apoptosis, adhesion angiogenesis, invasion and metastasis, axon guidance, transmembrane glycoproteins, peptide hormones, and chromatin organization (115).

CIMP is found in approximately 20%–30% of CRC and it was reported that clinical features of CIMP CRCs are often similar to those associated with MSI (116). Interestingly, the precursor lesions of most of CIMP CRCs are believed to be sessile serrated adenomas (SSA), exhibiting both *BRAF* mutation and high level of CpG island methylation (117). Most of these lesions, particularly those in the proximal colon, have been so far under-recognized and missed during colonoscopy, qualifying these lesions as the main cause of interval cancers. It is estimated that 10%-20% of CRCs evolve through this alternative, serrated pathway, with a distinct genetic and epigenetic profile.

3. HEREDITARY COLORECTAL CANCER SYNDROMES

As reported above, about 5 to 10% of CRCs can be attributed to inherited susceptibility syndromes (118–119). These hereditary syndromes are broadly classified upon the clinical presence or absence of multiple colorectal polyps into two categories. Here, we focused on the three most frequent forms of hereditary CRC currently managed in the diagnostic routine of Genetic Counseling Service: MUTYH-associated polyposys (MAP) and familial adenomatous adenomatous polyposis (FAP) *versus* Lynch syndrome (LS).

3.1 MUTYH-ASSOCIATED POLYPOSIS (MAP)

In 2002, Al-Tassan et al. described for the first time MAP syndrome, a recessive form of polyposis associated with biallelic mutations of *MUTYH* gene (120) (OMIM 604933). This gene is located at chromosome locus 1p34.1, is 11.2 kb long and has 16 exons.

The MUTYH protein is a base excision repair (BER) glycosylase involved in the repair of one of the most frequent and stable forms of oxidative damage, oxidation of a guanine leading to 8-OHdG. When an oxoG:A mismatch is present in the DNA-template, in the next round of replication a G:C to T:A transversion will occur. MUTYH recognizes an oxoG:A mismatch and excises the undamaged adenine base using a base–flipping mechanism. DNA polymerases can subsequently restore an oxoG:C pair that can be acted upon by another BER-glycosylase, OGG1, to replace the oxidized guanine with a guanine (121-123) (Figure 4).

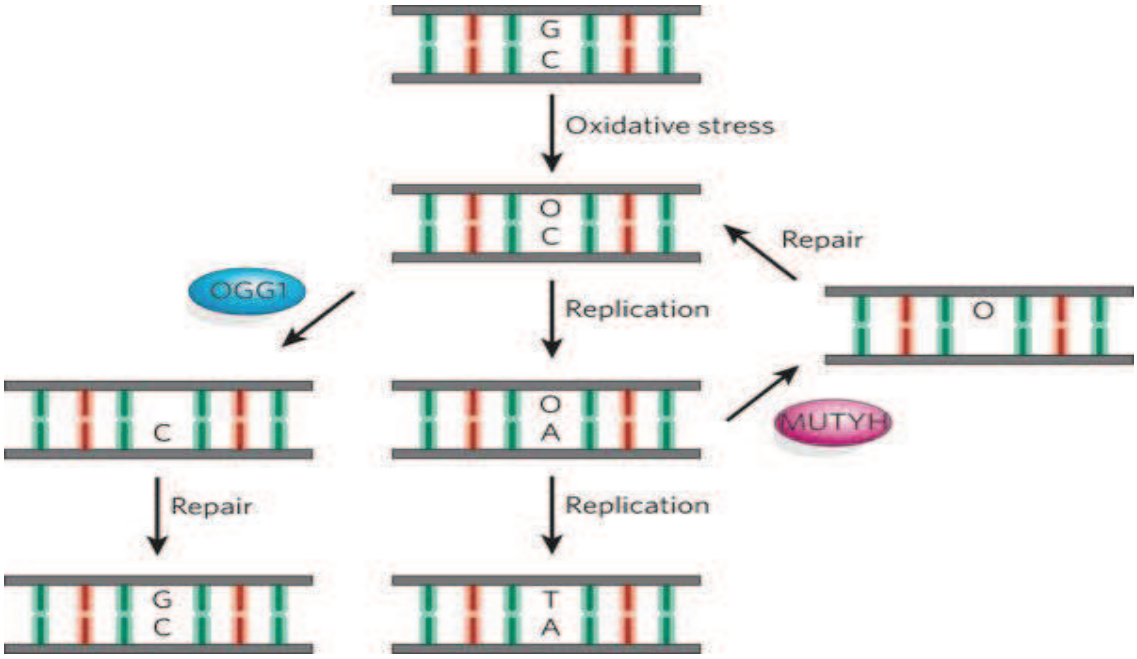


Figure 4. Pathways for the removal of 8-OHdG in human cells.

To date, 308 unique sequence variants for the MUTYH gene have been reported in the Leiden Open Variation Database (LOVD) of the *InSiGHT* (International Society for Gastrointestinal Hereditary Tumours). These are predominantly missense mutations, but also small deletions and duplications may be observed.

In Caucasian populations, a biallelic status for the hot spot mutations p.Y179C and/or p.G396D is reported in up to 70% of MAP patients. Furthermore, 90% of the western MAP population carries at least one of these mutations (124).

The MAP phenotype resembles that of *APC*-linked attenuated familial adenomatous polyposis (AFAP) with the appearance of a limited number of adenomas (generally 30–100) in the fourth to fifth decade of life. However, unlike AFAP, approximately 60% of MAP patients show colorectal cancer at presentation (125). MAP carcinogenesis displays peculiar molecular features that characterize disease progression; CIN is detectable during the early stages in MAP tumours (126) and the somatic molecular fingerprint of this syndrome is an excess of *KRAS* c.34G>T transversions, due to the failure of the impaired MUTYH to repair the mismatches induced by the 8-oxoG variant base (127).

Since oxidative stress occurs in various cell types, *MUTYH* inactivation can be expected to predispose not only to intestinal, but also to extraintestinal lesions. Key extracolonic manifestations include predisposition to duodenal adenomas and cancer (128).

A multicenter study showed also that the incidence of extraintestinal malignancies among MAP cases is almost twice that of the general population, with a significant increase in the incidence of ovarian, bladder, and skin cancers, and a trend of increased risk of breast tumors (128).

Regarding clinical management of MAP patients, the genetic testing of *MUTYH* in patient with phenotypic features suggestive of this disease is essential in planning for the surveillance needs in the extended family. Indeed, there seems to be a consensus in the literature to suggest surveillance in MAP patients, according to AFAP protocols. Colonic surveillance should start at age 18–20 years and gastroduodenal surveillance at age 25–30 years (129).

3.2 FAMILIAL ADENOMATOUS POLYPOSIS (FAP)

FAP is an autosomal dominant disease with almost complete penetrance by age of 40 years (ORPHA 733; OMIM 175100). In 1991, germline mutations in the *APC* gene were identified as the cause of FAP that is characterized by the development of hundreds to thousands of premalignant adenomas in the gastrointestinal tract, mostly in the colon and the rectum, at a young age (130).

APC is a gene with 15 exons located in the long arm of chromosome 5 in band q22.2 that encodes a 2843 amino acid protein. More than 4000 different mutations of the *APC* gene causing FAP have been reported in the LOVD database resulting in a truncated protein where approximately one third of germline mutations in *APC* lie between codons 1061 and 1309 in the center of the gene (131-132).

FAP accounts for less than 1% of all CRC cases and affects approximately 1 in 10,000 people. Patients with FAP develop CRC at an average age of 35 years if left untreated, although there are differences within and between families, some of which can be explained by specific germline mutations in *APC* (133).

An attenuated form of Familial Adenomatous Polyposis (AFAP) is characterized by mutations at the 5'-and 3' ends of *APC* gene and the development of less than 100 polyps, mainly located in the left colon. In AFAP, the average age for the development of adenomas and CRCs is older compared to that associated to the classic FAP (134). FAP patients frequently develop benign extracolonic lesions including polyps of the gastric fundus and duodenum, osteomas, dental anomalies, congenital hypertrophy of the retinal pigment epithelium (CHRPE), soft tissue tumors, and desmoid tumors.

In addition, several extracolonic cancers occur with a higher incidence in FAP than in the general population. These cancers include tumors of the upper gastrointestinal tract, liver, thyroid and adrenal gland, pancreas, and brain (135-136).

Genetic testing is routinely used for detection of FAP. Flexible sigmoidoscopy at the age of 10-12 years old is recommended for screening for polyps in *APC* gene mutation carriers. Once polyps are detected, annual colonoscopy for polyp screening is recommended and when the polyp burden increases, prophylactic colectomy is offered. If the rectum is left, annual endoscopy is needed because of adenoma's development risk. Screening for polyps and adenomas in the upper gastrointestinal tract with gastroduodenoscopy is preferred to initiate at the age of 25-30 years, every 1-3 years depending on the poly burden (137).

3.3 LYNCH SYNDROME (LS)

Lynch Sindrome (LS) is the most common hereditary colon cancer syndrome with autosomal dominant transmission and high penetrance (80-85%) accounting for 2-4% of all CRCs (138-139).

LS is characterized by germline mutations in DNA MMR genes *MLH1*, *MSH2*, *MSH6* and *PMS2*, which cause inability to correctly repair errors occurring during DNA replication, especially in the repeat sequences like microsatellites. For this reason, most Lynch syndrome tumors arise through the MSI pathway.

Mutations in these genes not only cause development of CRC but also an increased risk of extracolonic cancers including endometrial, ovarian, gastric, small bowel, upper urologic tract, pancreatic cancer, hepato-biliary tract and brain (140).

The LOVD database has collected so far 2360 MMR germline variants; more frequently mutated genes are *MLH1* (50%) *MSH2* (40%), followed by *MSH6* (10%) and *PMS2* (less than 5%) (141).

The germline mutation is present in every cell of these individuals and loss of functional MMR occurs when the wild-type allele is inactivated by a somatic mutation, due to different genetic mechanisms (deletion, mitotic recombination, point mutation, genetic conversion, loss of heterozygosity), or by an epimutation even in absence of germline mutation in MMR genes (142).

Many criteria have been proposed for identifying patients with Lynch syndrome, and are mostly based on age at diagnosis, presence of multiple tumours and number of

affected family members. The syndrome is clinically defined by the Amsterdam and the revised Bethesda guidelines for selecting patients for further genetic analysis (143).

The main clinical feature is early age of diagnosis and the occurrence of multiple tumours. The average age at diagnosis of CRC in Lynch syndrome is between 42 and 61 years, which is lower than the general population (144).

Periodic examination by colonoscopy is recommended for detecting CRC in an early stage and a 63% risk reduction in CRC development can be achieved reducing the cancer related mortality (140).

4. IDENTIFICATION OF LOW-PENETRANCE ALLELES IN CRC

Over 29,000 Single nucleotide polymorphisms (SNPs) associated with numerous traits and diseases have been discovered thus far through genome-wide association studies (GWAS) (145). SNPs can exert their influence on disease pathogenesis in a variety of ways. If located within a gene a SNP may have direct or indirect consequences on the function or structural stability of a protein if it changes the primary structure (146). Exonic SNPs resulting in amino acid substitutions, called non-synonymous SNPs, are the most well characterized genetic polymorphisms. They are subject to detection bias and can usually be assayed for their functional effects (147). Synonymous exonic SNPs that do not alter protein structure may still affect mRNA stability and alter splicing signals (148). SNPs located in introns, promoters, enhancers, or any other non-coding regions can also be functionally important through alteration of gene regulation. Methodologies are currently being developed for predicting the function of SNPs located in introns or regulatory regions (149-150).

SNPs may also disrupt or create CpG dinucleotides, causing altered methylation patterns (151).

It is likely that the accumulation of a number of low-risk, low-penetrance alleles contributes to significant CRC risk.

SNPs identified as being associated with CRC or other disease may be risk-associated variants, or modifier variants/alleles. Modifier variants are coding or non-coding regulatory elements that interact with the genome, and modifier variants can act together to modulate a phenotype in complex diseases (152).

30 published studies established an association between SNPs and colorectal cancer, for a total of 193 SNPs at 145 genetic loci (153).

4.1 THE ROLE OF SINGLE NUCLEOTIDE POLYMORFISM rs1800734 OF *MLH1* PROMOTER IN PREDISPOSITION TO CRC

Several CRC susceptibility SNPs have been identified within MMR genes included *MLH1*.

The importance of *MLH1* in CRC and its propensity for hypermethylation has been known for some time (154), however there is few detailed analyses of how, where and when *MLH1* epimutations occur during sporadic carcinogenesis, and no good explanation for why somatic *MLH1* mutations are very rare.

In 1999 Deng et al. correlated *MLH1* hypermethylation with the absence of gene expression and examined the methylation status of its promoter in four regions: A region (from -711 to -577, containing 23 CpG sites), B (from −552 to −266, 19 CpG sites), C (from −248 to −178, 8 CpG sites), and D (from −109 to +15, 7 CpG sites) (155). They finally concluded that methylation status of CpG sites in region C provided the best correlation and prediction of MLH1 expression differently from regions B and D while methylation in region A seemed not to be critical in silencing the gene expression.

A common G/A single nucleotide polymorphism (SNP), rs1800734 (*MLH1*-93 G>A), is located in the core of the promoter of *MLH1,* 93 bases upstream of the transcription start site in a region that is required for maximal transcriptional activity (156) (Figure 5).

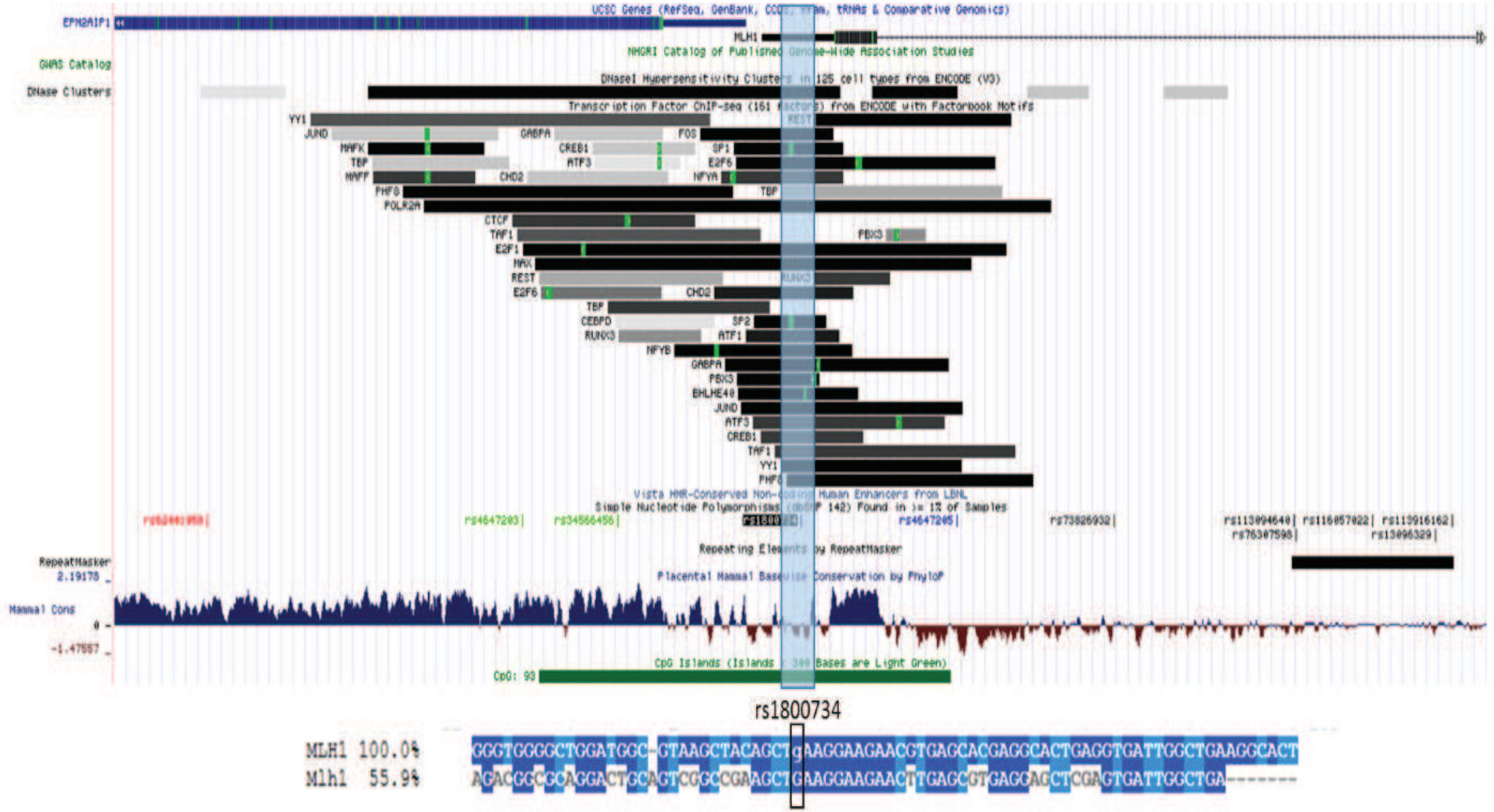


Figure 5. The *MLH1* promoter region and TFs binding site around rs1800734.

This SNP was initially analysed as a candidate for CRC susceptibility and has since then been assessed in a number of data sets (157-161).

Indeed, the AA homozygotes and AG heterozygotes have an increased risk of MSI CRC compared with GG homozygotes.

Published studies have found associations between rs1800734 genotype, *MLH1* methylation and protein absence in cancers (159-161). By contrast, there are very few information on the mechanism of CRC susceptibility caused by rs1800734 genotype. Data such as associations between SNP alleles and *MLH1*expression or methylation in normal tissue or MSS cancers are absent or deficient.

Two different studies have performed pilot luciferase reporter assay to investigate the allele-specific effects of rs1800734 on transcription and both groups have found the risk allele to be associated with significantly lower MLH1 mRNA expression, despite the reporter system lacking the endogenous chromatin environment and, most likely being essentially unmethylated. The findings indicate that in vitro differential expression is likely to be driven primarily by transcription factor binding rather than methylation, suggesting dynamic interplay between transcription and methylation as a potential mechanism of rs1800734-associated *MLH1* repression (162-163).

There is evidence that binding of the transcription factor TFAP4 (AP-4) is modified by rs1800734 (164). However, Liu et al. detected no difference in *MLH1* allele-specific expression as a result of TFAP4 allelic bias. They showed instead an effect on the expression of the gene encoding the protein kinase DCLK3 and long range chromatin interactions between rs1800734 and the DCLK3 promoter.

Although the majority of DNA methylation research focused on CpG islands within *MLH1* promoter, Savio et al. have observed recently a modulation of transciption factor binding and differential DNA methylation patterns at the *MLH1* shore in blood mononuclear cells (PBMC) DNA of CRC cases and controls in association with variant rs1800734 genotype.

Therefore, these data suggested a dynamic genotype-associated epigenetic regulation of the *MLH1* CpG island and shore (165-166).

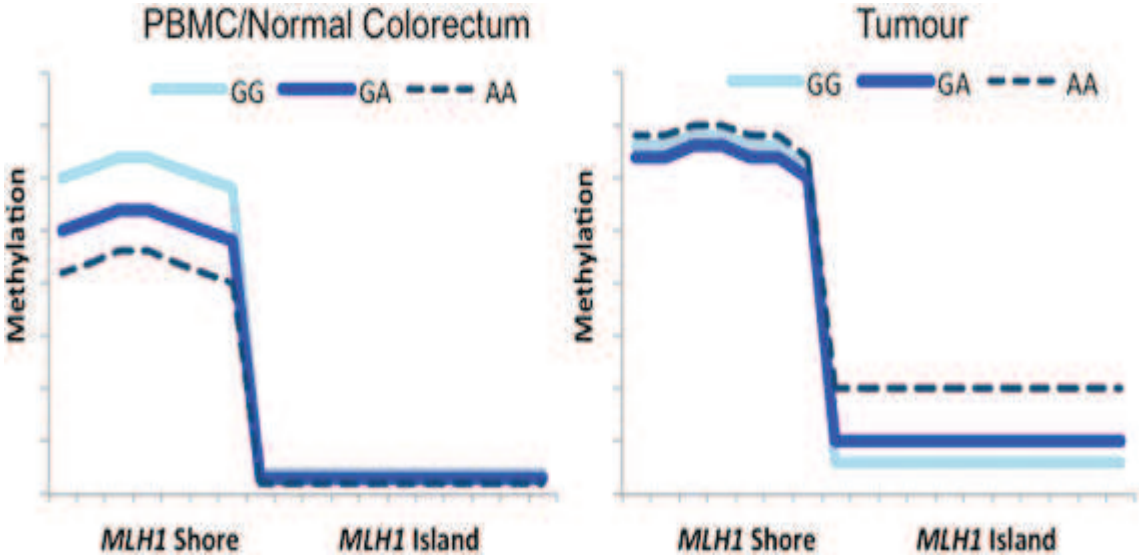


Figure 6. Schematic model of DNA methylation at the *MLH1* CpG island and shore. In PBMCs and normal colorectal tissue (left panel) the *MLH1* shore incurs hypomethylation in association with variant SNP genotype of rs1800734. No methylation is present at the CpG island in these DNA sources. In colorectal tumour (right panel), DNA methylation at the CpG shore loses its association with rs1800734 genotype whereas the CpG island incurs hypermethylation in association with variant SNP genotype of rs1800734 (166).

AIM

The operative work of my PhD training in Experimental and Traslational Medicine has been carried out at the laboratory of molecular pathology of the Anatomic Pathology Unit of Varese Hospital and at laboratory of Cancer Biology of the Wellcome Trust Centre for Human Genetics in Oxford.

Therefore, the thesis can be divided in two parts.

In the first part we focused on the aberrant hypomethylation of DNA and the main aims were:

1. To investigate the early genetic and epigenetic features in colorectal premalignant lesions characterized by oxidative DNA damage and to assess their potential involvement in driving colorectal carcinogenesis. To this purpose, we analyzed a cohort of colorectal adenomas and carcinomas derived from MAP subjects, sporadic adenomas and carcinomas and a control set of FAP/AFAP for their L1 and L1-*MET* methylation status and for their mutational profiles;

2. To characterize the L1-*MET* transcript induced by L1 hypomethylation.

In the second part we studied the promoter of *MLH1* in order to elucidate the relationship between methylation and risk/protective allele at rs1800734 during CRC progression.

In detail, the specific aims were the following:

1. To analyze in a cohort of normal colorectal samples and CRCs the associations between rs1800734 alleles, MLH1 mRNA expression and *MLH1* methylation;

2. To evaluate in vitro the dynamic of *MLH1* epigenetics and expression altering promoter methylation.

MATERIALS AND METHODS

FIRST PART
1. PATIENTS AND SAMPLES

We studied 168 formalin fixed and paraffin embedded (FFPE) adenomas and 26 CRCs cancers collected in the last 20 years from the files of the Department of Pathology of the Ospedale di Circolo-University of Insubria, Varese, and from the archives of the Pathology Unit of Candiolo Cancer Institute (FPO-IRCCS).

Overall we analyzed 52 adenomas and 11 carcinomas from 18 MAP patients carrying different biallelic *MUTYH* germline mutations (patients M1-M18; 9 females, 8 males; median age of 47 years, range 39-79 years), 36 adenomas from 17 FAP/AFAP patients with different *APC* germline alterations (patients F1–F17; 10 females and 7 males; median age of 37 years, range 21–61 years) and 80 sporadic adenomas from 62 individuals (28 females, 34 males; median age of 66 years, range 44-84 years) (Table 1 and Supplementary Tables S1-S3).

After genetic counseling, MAP and FAP/AFAP patients underwent *MUTYH* and *APC* germline mutation analysis. Germline mutations are reported in Supplementary Table S1-S2). Seventeen out 36 *MUTYH* mutations were the most common pathogenic p.Y179C and p.G396D; 11 were truncating and 8 were missense whose pathogenicity was checked in the Leiden Open Variation Database (http: //www.lovd.nl/MUTYH). Regarding *APC* mutations, all were truncating (N=14) expect for three large rearrangements: a exon 2 deletion, a whole gene deletion and a translocation t(5;7) (q22;p15).

In most of MAP and FAP/AFAP cases, multiple adenomas from the same patient were investigated. In particular, for six MAP patients, we could analyze adenomas and carcinomas derived from the same six subjects (cases M1, M6, M16, M18 with adenomas and one carcinoma each, and cases M4 and M9 with adenomas and two carcinomas each), for nine MAP patient only adenomas (M2, M3, M5, M7, M8, M10, M12, M14 and M15) and in three patients only carcinoma samples (M11, M13 and M17) (Supplementary Tables S4).

For the sporadic cases, we enrolled those individuals who had undergone the endoscopic screening without a previous report of adenoma detection. In addition, familiarity for polyposis was excluded according to their clinical history by interview.

In details, in this subset we distinguished two different groups of cases: 1) 56 sporadic adenomas (S-Ads) diagnosed in 45 patients who have never developed a CRC during a

follow-up of ten-years; 2) 24 sporadic adenomas (S-AdsC) from 17 individuals who developed a CRC at least one year after polyp removal. For 15 of these patients we could also analyze for comparison the matched CRC (Supplementary Tables S3, S5).

The histopathological revision was performed by an expert pathologist according to the WHO classification of CRCs (167).

High-grade dysplasia was observed in 8% of MAP adenomas, 6% of FAP/AFAP adenomas, 24% of sporadic adenomas, respectively. The remaining cases showed low-grade dysplasia. A tubular histology was observed in 91 adenomas (32 MAP, 27 FAP/AFAP and 32 sporadic polyps), while 73 samples were tubulovillous (20 MAP, 9 FAP/AFAP and 44 sporadic polyps) and 4 villous (Supplementary Tables S1-S3).

Additional clinico-pathological data were collected for sporadic samples and included: site, size and number of sporadic adenomas at diagnosis.

Table 1. Clinicopatholgical data of patients and adenomas included in the study.

	MAP*	**FAP/AFAP**	**SPORADIC****	
PATIENTS	18	17	62	
Male	9	7	34	
Female	9	10	28	
Mean age (min-max)	47 (39-79)	37 (21-61)	66 (44-84)	
	Adenomas	Adenomas	SAds	SAdsC
SAMPLES	52	36	56	24
Histology				
V	0	0	0	4
TV	20	9	36	8
TB	32	27	20	12
Dysplasia				
Low	47	33	43	18
High	3	2	13	6

Legend: MAP, *MUTYH*-associated-polyposis; FAP/AFAP, classical/attenuated adenomatous polyposis; V, villous adenoma; TBV, tubulovillous adenoma; TB, tubular adenoma;
MAP patients developed 11 carcinomas; ** Fifteen patients with sporadic adenomas developed also a metachronous CRC.

2. GENE MUTATIONAL ANALYSIS

All the 168 adenomas and the 26 CRCs (11 MAP and 15 sporadic cases) were analysed for mutations in *KRAS* (NM_004985.4/LRG_344), *NRAS* (NM_002524.4/LRG_92), *BRAF* (NM_00433.4/LRG_299), and *PIK3KA* (NM_006218.2/LRG_310) genes using the Sequenom MassArray system (Diatech Pharmacogenetics, Jesi, Italy), based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This analysis was performed using the Myriapod Colon Status Kit (Diatech Pharmacogenetics) that includes a series of PCR assays designed to interrogate a total of 153 non-synonymous hotspot mutations in the four genes. This method is based on primer extension and offers two levels of specificity. First, a locus-specific PCR reaction, followed by a locus-specific primer extension reaction (iPLEX) in which an oligonucleotide primer anneals immediately upstream of the polymorphic site being genotyped (Figure 7)

This system is particularly suitable for diagnostics use for the minimal amounts of required DNA utilized and for the robustness and reproducibility of protocols.

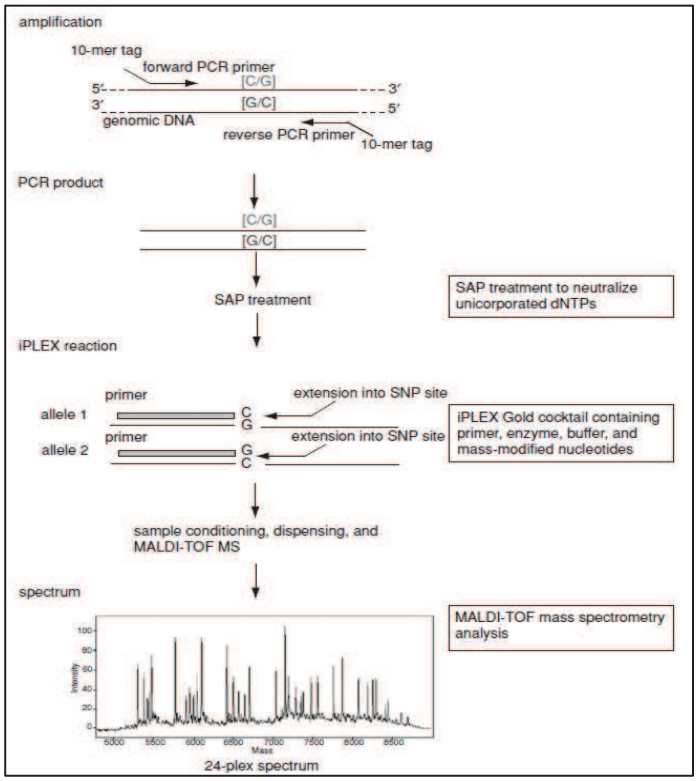


Figure 7. Workflow of mutation analysis using Sequenom MassArray system.

Through the use of MALDI-TOF MS, the mass of the extended primer is determined. The primer's mass indicates the sequence and, therefore, the alleles present at the

polymorphic site of interest. Sequenom supplies software (SpectroTYPER) that automatically translates the mass of the observed primers into a genotype for each reaction. Briefly, sequenom analysis was done performing 5 µl PCR reaction mixture containing from 10 to 20 ng of DNA (Table 2). After PCR, terminal nucleotides were dephosphorylated by Shrimp Alkaline Phosphatase (SAP) reaction (Table 3) and this step was followed by the iPLEX single base extension reaction (Table 4).

The extension products (analytes) were desalted using clean resin and spotted in nanoliter volumes onto a matrix-arrayed silicon SpectroCHIP with 96 elements using the MassARRAY Nanodispencer (Diatech Pharmacogenetics). The chip is placed into the mass spectrometer and each spot is then shot with a laser under vacuum by the MALDI-TOF method. A laser beam serves as desorption and ionization source in MALDI mass spectrometry. Once the sample molecules are vaporized and ionized, they are transferred electrostatically into a time-of-flight mass spectrometer (TOF-MS), where they are separated from the matrix ions, individually detected based on their mass-to-charge (*m/z*) ratios, and analized. Detection of an ion at the end of the tube is based on its flight time, which is proportional to the square root of its *m/z*.

Table 2. PCR reaction mix and thermic profile.

Reaction mix	
Water	1,3 µl
10x PCR Buffer	0,5 µl
MgCl2	0,4 µl
dNTP Mix	0,1 µl
Primer Mix	0,5 µl
PCR Enzyme	0,2 µl
DNA	2 µl
Total volume	5 µl

95°C	120"	
95°C	30"	
56°C	30"	X 45
72°C	60"	
72°C	300"	
4°C	300"	
10°C	Hold	

Table 3. SAP reaction mix and thermic profile.

Reaction mix	
Water	1,53 µl
SAP buffer	0,17 µl
SAP enzyme	0,3 µl
Total volume	2 µl

37°C	2400"
85°C	300"
4°C	300"
10°C	Hold

Table 4. IPLEX reaction mix and thermic profile.

Reaction mix	
Water	0,56 µl
Buffer Plus	0,20 µl
Termination mix	0,20 µl
Thermosequenase	0,04 µl
Primer Mix	1 µl
Total volume	2 µl

94°C	30"	
94°C	5"	
52°C	5"	X 40
80°C	5"	
52°C	5"	
		X 5
80°C	5"	
72°C	180"	
4°C	300"	
10°C	Hold	

3. L1 AND L1-*MET* METHYLATION STUDY IN HEREDITARY AND SPORADIC ADENOMAS BY BISULPHITE PIROSEQUENCING ANALYSIS

The methylation status of global and local L1 was evaluated by bisulfite-PCR and pyrosequencing in 15 samples of histologically normal colonic mucosa derived from healthy individuals obtained from the files of the Department of Pathology of Ospedale di Circolo-University of Insubria and in all MAP, FAP, sporadic adenomas and MAP and sporadic carcinomas.

Bisulfite modification of genomic DNA (300 ng) was performed with an EpiTect Bisulfite Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations.

Bisulfite-modified DNA was amplified and sequenced by using L1 and L1-*MET* primers and protocol reported below (Table 5, Table 6 and Table 7).

Pyrosequencing was carried out with PyroGold reagents on a PyroMark Q96 ID system (Qiagen, Hilden, Germany). Pyrogram outputs were analyzed by the Pyromark Q24 software using the Allele Quantification software (Qiagen, Hilden, Germany) to determine the percentage of methylated alleles at each CpG site examined.

Global L1 assay was designed toward a consensus L1 sequence (GenBank accession number M80343.1) and allowed to quantify the percentage of 5-methylated cytosines (%5mC) in four consecutive CpG sites as previously reported (Figure 8) (168).

Intragenic levels of L1 methylation were analyzed using the L1-*MET* assay: the forward PCR primer was located inside the L1 promoter, and the reverse primer was designed within the *MET* gene intron between exons 2 and 3. The sequencing primer is immediately upstream a sequence that includes three CpG sites whose mean methylation percentage was quantified (GenBank accession number: NG_0089961).

Fully methylated and unmethylated DNA (Millipore, Billerica MA, USA) were used as positive and negative controls in each experiment. Reproducibility was confirmed by analyzing all the samples in duplicate with a maximum of within-sample coefficients of variation equal to 5% (range 2%-5%).

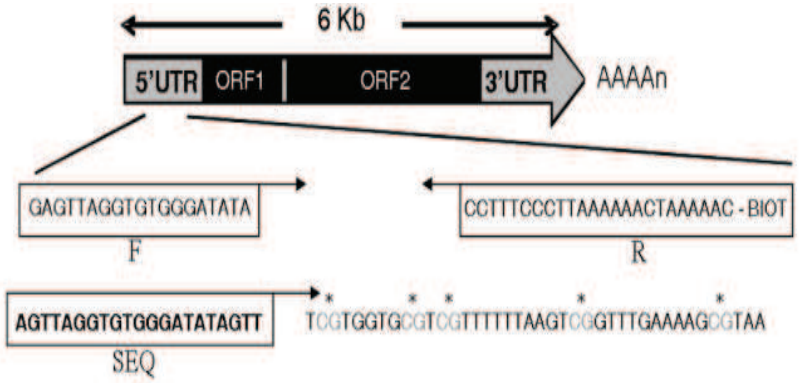


Figure 8. Diagram of the pyrosequencing assay used to measure L1 promoter methylation. Arrows indicate the sequence and the position of the primers used for the bisulfate PCR and pyrosequencing. The percentage of 5-methylated cytosines was measured in four consecutive CpG sites.

Table 5. Primer sequences of LINE-1 and L1-*MET* pyrosequencing analyses.

	Primer sequence
L1	Fw 5'-gagttaggtgtgggatatagt-3' Rev 5'-biot-caaaaatcaaaaaattccctttcc-3' Seq-5'-agttaggtgtgggatatagtt-3'
L1-*MET*	Fw 5'-gagatgaatttagtattttagatggaaatg-3' Rev 5'-biot-acaactcccatctacaactccca-3' Seq 5'-tttagatggaaatgtagaaattat-3'

Table 6. L1 reaction mix and thermic profile.

Reaction mix	
Buffer	10 µl
$MgCl_2$	3 µl
dNTPSs	5 µl
Primer For	1 µl
Primer Back	1 µl
GoTaq Promega	0,25 µl
Water MQ	22,25 µl
DNA	5 µl
Final volume	50 µl

95°C	3'	
95°C	25''	
50°C-40°C	25''	X 40 (TD)
72°C	25''	
72°C	5'	

Table 7. L1-*MET* reaction mix and thermic profile.

Reaction mix	
Buffer	5 µl
$MgCl_2$	4 µl
dNTPSs	6 µl
Primer For	1,5 µl
Primer Back	1,5 µl
Epitaq Takara	0,25 µl
Water MQ	24,25 µl
DNA	5 µl
Final volume	50 µl

95°C	3’	
95°C	25’’	
67°C-62°C	25’’	X 40 (TD)
72°C	30’’	
95°C	25"	
62°C	25"	X 34
72°C	25"	
72°C	3'	

4. *MET* AND L1-*MET* EXPRESSION STUDY IN CRCs AND PRENEOPLASTIC LESIONS

4.1 RNA SEQ ANALYSIS

RNA SEQ analysis of 13 FFPE samples of histologically normal colonic mucosa and 32 FFPE CRCs, previously examined in our lab for their clinico-pathological and molecular profiles (169), was performed at Laboratory of Molecular Medicine and Genomics, University of Salerno. This cohort of sporadic and hereditary CRCs were selected because they were stratified into four groups showing significant difference of L1 methylation levels: L1>60.1%, L2 (54.1%-60%), L3 (45.8%-54%), L4 (<45.6%).
Total RNA concentration was measured using Qubit® RNA HS Assay Kit on a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Integrity was assessed using Agilent RNA 6000 Nano Kit on a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) and the percentages of fragments larger than 200 nucleotides were calculated. RNA sequencing libraries were prepared using TruSeq RNA Access library kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer`s protocol. In brief, RNA samples (100 ng total RNA) were fragmented at 94°C for 8 minutes on a thermal cycler. First strand cDNA syntheses were performed at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes, using random hexameres and SuperScript II Reverse Transcriptase (Thermo Fisher Scientific Inc., Waltham, MA, USA). In a second strand cDNA synthesis the RNA templates were removed and a second replacement strand was generated by incorporation dUTP (in place of dTTP, to keep strand information) to generate ds cDNA. The 3`ends of the cDNA were then adenylated to facilitate adaptor ligation in the next step. In a first PCR amplification step, PCR (15 cycles of 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds) were used to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. After validation of the libraries, using Agilent DNA 1000 kit on a 2100 Bioanalyzer instrument, the first hybridization step were performed using exome capture probes. Before hybridization a 4-plex pool of libraries were made, by combining 200 ng of each DNA library. The hybridization was performed by 18 cycles of 1 minute incubation, starting at 94°C, and then decreasing 2°C per cycle. Then streptavidin coated magnetic beads were used to capture probes hybridized to the target regions. The second

hybridization (18 cycles of 1 minute incubation, starting at 94°C, and then decreasing 2°C per cycle) were required to ensure high specificity of the capture regions. A second capture with streptavidin-coated beads were performed, followed by two heated wash procedures to remove non-specific binding form the beads. The amplification step was performed by 10 cycles (98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds). Finally, the libraries were quantitated using Qubit dsDNA HS (High Sensitivity) Assay Kit on a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and validated using Agilent High Sensitivity DNA Kit on a Bioanalyzer. The size range of the DNA fragments were measured to be in the range of 200–650 bp and peaked around 270 bp. Libraries were normalized to 2 pM and sequenced on a NextSeq500 instrument 2x75 bp (Illumina, Inc. San Diego, CA, USA), according to the manufacturer's instructions.

For RNA sequencing bioinformatics analysis reads were quality filtered and aligned to the human genome hg19 (Homo sapiens Ensembl GRCh37) using STAR v.2.5.2a (170). HTSeq (171) was used to compute read counts across each gene annotated in UCSC, which were then used as input to R package DESeq2 (172). DESeq2 was used to normalize read counts for library size and dispersion followed by tests for differential gene expression. Significant differentially expressed genes were determined using false discovery rate (FDR) cutoff ≤ 0.05 and at least 1.5-fold change between conditions. Functional analyses were performed with Ingenuity Pathway Analysis suit (Ingenuity Systems). Finally, TopHat-Fusion (173) was used to detect new gene fusion events.

4.2 CHARACTERIZATION OF L1-*MET* CHIMERIC ISOFORM

After a critical analysis of the literature, we analyzed the structure of the protoncogene *MET* focusing on the L1 sequence residing within its intronic regions, between exon 2 and exon 3, in order to elucidate the regulatory mechanism of L1 hypomethylation-induced chimeric L1-*MET* isoform activation (Figure 8).

Previous studies have characterized the chimeric L1-MET isoform, describing the same number of exons of *MET* transcript, i.e. 21, but different exon 1 and exon 2, that are located within intron 2 of *MET* gene. From exon 3, the sequence of both transcripts is the same.

In order to analyse by qRT-PCR the expression levels of the two different transcripts, distinguishing the physiological variant of *MET* transcript from the aberrant isoform, we designed two primer sets using the sequence alignment software APE. (Table 8).

The forward PCR primer for *MET* was located inside the exon 1 and the reverse primer was designed within the exon 2. On the contrary, the forward PCR primer for L1-*MET* was located inside the exon 1 of L1-*MET* that contains also the L1 promoter sequence, and the reverse primer was designed within the exon 2 of LI-*MET*.

To evaluate the specificity of L1-*MET* amplicon and confirm the presence of a fusion transcript containing the L1 sequence, we sequenced the PCR-amplified products of four CRCs after selecting the specific band on 3% agarose gel. These fragments were sequenced using BigDye Terminator v1.1 kit (ThermoFisher scientific, Waltham, USA) following the protocol reported below (Table 9), purified using DyEx kit (Qiagen) and subjected to automated sequencing on ABI PRISM 310 (Applied Biosystems, Foster City, Ca).

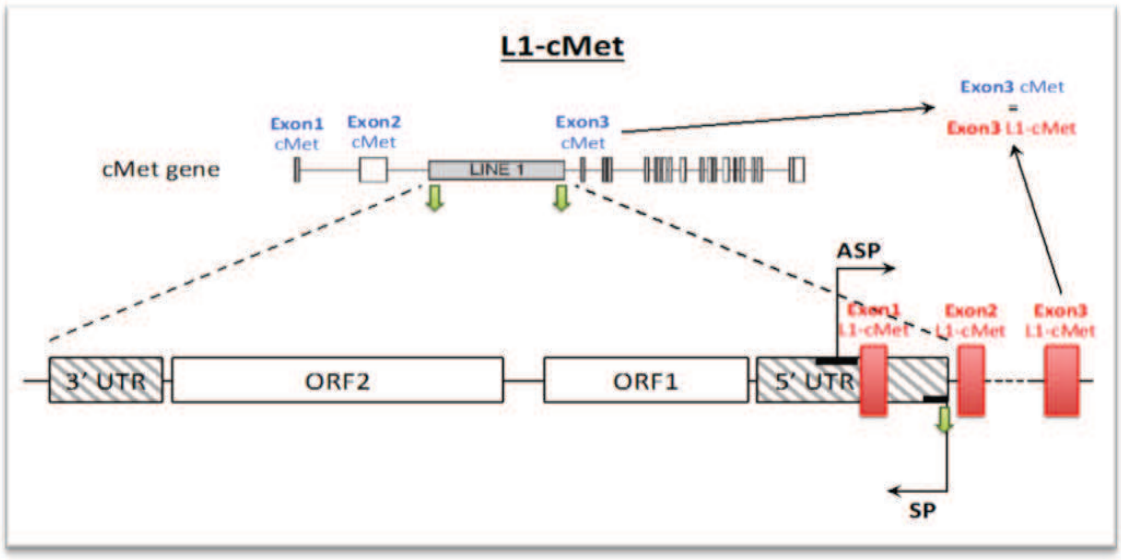


Figure 9. Schematic illustration of the local LINE-1 sequence in the *MET* gene between the exons 2 and 3. The black arrows indicate the LINE-1 sense and antisense promoter. The red boxes represent the first three exons of *L1-MET* chimeric isoform.

Table 8. Primer sequence of *MET* and L1-*MET*.

	Primer sequence
MET	Fw 5'-acttctccactggttcctgg-3' Rev 5'-gcaccaaggtaaacaggagc-3'
L1-*MET*	Fw 5'-ctgctgtgctagcaatcagc-3' Rev 5'-tcttcacctccaatctagcaca-3'

Table 9. Sequencing reaction mix and thermic profile.

Reaction mix	
Ready reaction Mix	2 µl
5X Sequecing Buffer	1 µl
Primer (For/ Back)	6 µl
DNA	2 µl
Water	3,4 µl
Total volume	10

96°C	1'	
96°C	10"	
50°C	5''	X 25
60°C	4'	
4°C	Hold	

4.3 *MET* AND L1-*MET* GENE EXPRESSION ANALYSIS

MET and L1-*MET* expression analyses were possible for 8 specimens of normal colonic mucosa, 85 adenomas and 42 CRCs. In details, we analyzed 13 MAP adenomas, 16 FAP adenomas, 56 sporadic adenomas, 7 MAP CRCs, 15 sporadic CRCs from patients with

S-AdsC and 20 CRCs previously stratified into the two groups showing the most significant differences of L1 methylation levels (namely the two groups: L1>60,1% and L4<45,6%) (169).

Total RNA was extracted from three 8-µm sections of FFPE tissue samples by manual microdissection and then automated purification on Maxwell® instrument using the 16 LEV RNA FFPE Purification Kit (Promega) according to the manufacturer's recommendations.

Purified RNA was quantified using Qubit® RNA Assay kit and Qubit® 2.0 Fluorometer. A total of 800 ng of total RNA was reverse-transcribed with random primers using the High-Capacity complementary DNA (cDNA) Archive kit (Applied

Biosystems; Foster City, Ca). The protocol of the retrotranscription and the corresponding thermic profile are reported below (Table 10).

Expression of *MET* and L1-*MET*, together with the housekeeping gene β2-microglobulin (β2-M), was evaluated using the Fast Start Universal Sybr Green Master kit (Roche Diagnostic, Mannheim, Germany) and qRT-PCR reactions reported below (Table 11) were performed in duplicates in 96-well plates on ABI-Prism 700 detection System Instrument using RQ-software (version 1.1, Applied Biosystems). After normalization to the endogenous control, the amount of each target gene in tumour samples was compared to the normalized expression of the same gene in 8 normal colonic mucosae and it was measured by the equation FC $2^{-\Delta\Delta Ct}$ (ie, mRNA fold change (fc), where ΔCt = Ct target genes – Ct of endogenous control and ΔΔCt = ΔCt of samples for target gene – ΔCt of 8 normal colonic mucosae average for the target gene).

For all the samples included in this study, we observed a very small variation between the reactions performed in duplicates with Ct standard deviations lower than the acceptable limit of ΔCt0.5

Table 10. Retrotranscription reaction mix and thermic profile.

Reaction mix	
10X RT Buffer	6 µl
10X RT Random Primers	6 µl
dNTPS Mix	2,4 µl
Multiscribe Reverse Transcriptase	3 µl
Rnase Inhibitor	1 µl
RNA (40 ng/ul)	20 µl
Water	21,6 µl
Total volume	60 µl

25°C	10'
37°C	120'
85°C	5'

Table 11. qRT-PCR reaction mix and thermic profile of *MET* (a) and L1-*MET* (b).

Reaction mix	
Master Mix	8.4 µl
Primer For	0,5 µl
Primer Back	0,5 µl
Water	2,6 µl
cDNA	5 µl
Total volume	17 µl

a			
95°C	10'		
95°C	30"		
57°C	45"	X 45	
72°C	1'		
95°C	15"		
60°C	20"		
95°C	15"		

b		
95°C	10'	
95°C	30"	
58°C	45"	X 45
72°C	1'	
95°C	15"	
60°C	20"	
95°C	15"	

5. STATISTICAL ANALYSIS

The statistical analysis was performed using Graph Pad PRISM v.5. Univariate comparisons of continuous data were carried out using Student's t-test and discrete variables were compared with $\chi 2$ test or Fisher's exact test. All comparisons were two-sided and a p-value <0.05 was considered to be significant.

SECOND PART

6. GENOTYPING OF *MLH1* PROMOTER rs1800734 BY KOMPETITIVE ALLELE-SPECIFIC PCR (KASP)

The Kompetitive Allele Specific PCR system (KASP™, LGC Genomics, Teddington, Middlesex, UK) is a fluorescence-based genotyping variant of PCR that is based on allele-specific oligo extension and fluorescence resonance energy transfer (FRET) for signal generation. This method is very useful and cost-effective for the bi-allelic discrimination of SNPs at specific loci.

The KASP genotyping reaction comprises three components: sample DNA with the sequence of interest, KASP assay and the KASP Master Mix (Figure 10). The KASP Assay mix is specific to the SNP to be targeted and contains three assay specic non-labelled primers: two allele specific forward primers plus one common reverse primer.

In the first and second rounds of PCR, allele-specific primers match the target SNP and with the common reverse primer, amplify the target region. Each forward primer incorporates an additional tail sequence that corresponds with one of two universal FRET cassettes present in the KASP Master mix. In further rounds of PCR, levels of allele-specific tail increase and the fluorescent signalling becomes stronger as more fluorescent primers are used in the amplification process.

Following completion of the KASP PCR, reaction mix are read and the data analysed using any cluster analysis viewing software. Detected signals are plotted as a graph, with samples of the same genotype clustering together.

The evaluation of the polymorphism A/G within the promoter of *MLH1* was performed in a 10µl reaction mixture and the oligonucleotides and the thermic profile used are reported below (Table 12-13).

In detail, the genotyping of rs1800734 was performed on the following samples:

- 33 fresh-frozen normal colorectal biopsies from patients diagnosed without pre-neoplastic or neoplastic lesions during screening colonoscopy at John Radcliffe Hospital in Oxford;
- 35 FFPE sporadic CRCs previously characterized at Anatomic Pathology Unit in Varese for *BRAF* mutation and for MMR defects. All these tumors were MSI CRCs with MLH1 immunohistochemical loss and showed *BRAF* ^{V600E} mutation in 23 out of 35 cases (66%).

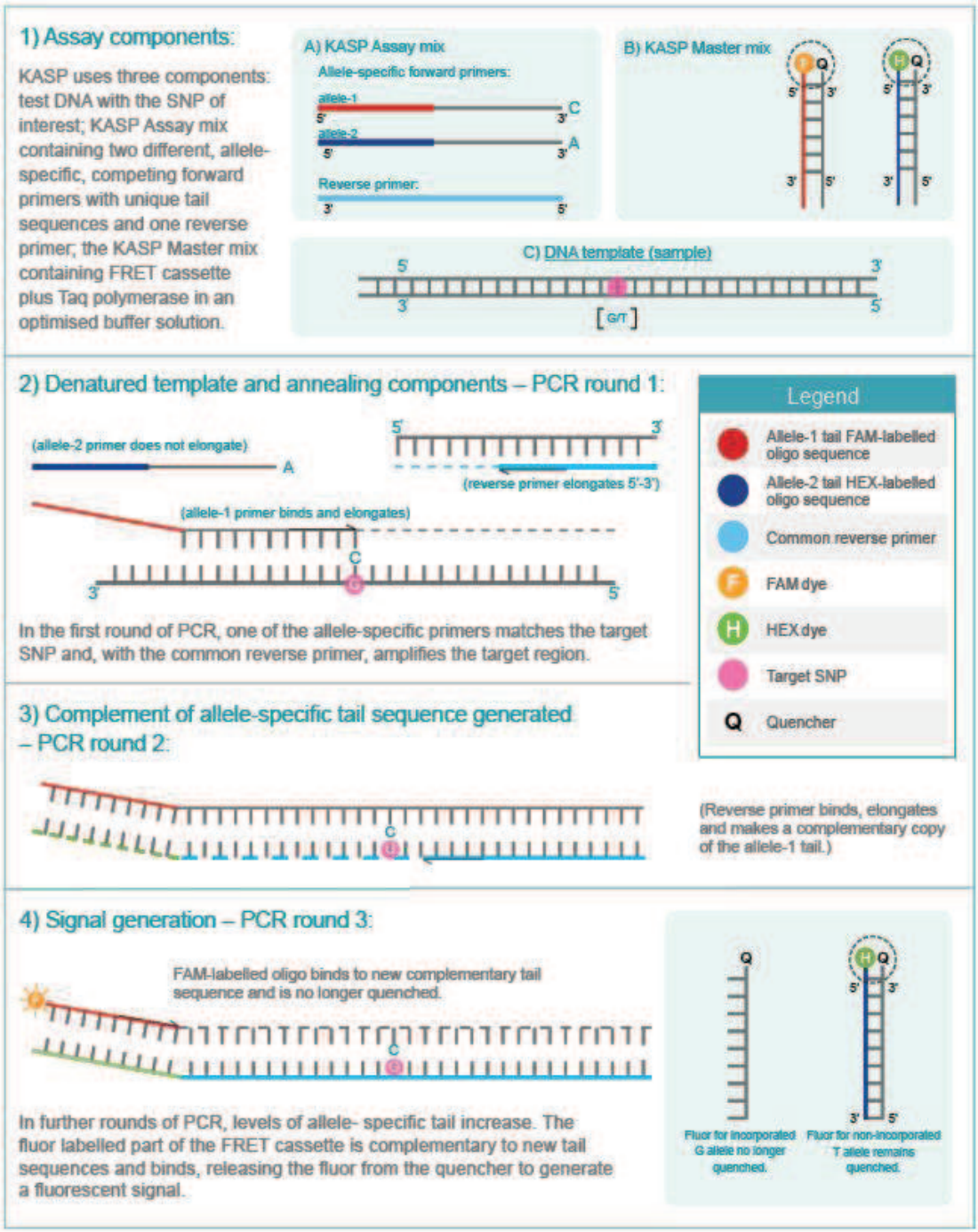


Figure 10. Schematic illustration of the KASP genotyping technology from LGC web site.

Table 12. Genotyping reaction mix and thermic profile.

Reaction mix	
KASP Master Mix	5 µl
Primer Mix	0,14 µl
Water	3,86 µl
DNA	1 µl
Total volume	10

94°C	15'	
61°C-55 °C	60"	X 10 (TD)
94°C	20"	
		X 36
55°C	60"	

Table 13. Primer sequence of rs1800734 allele A and rs1800734 allele G.

	Primer sequence
rs1800734 Allele A	5'-gaaggtgaccaagttcatgctggatggcgtaagctacagcta-3'
rs1800734 Allele G	5'-gaaggtcggagtcaacggattggatggcgtaagctacagctg-3'
Reverse Primer	5'-tcaacggaagtgccttcagccaat-3'

7. *MLH1* METHYLATION STUDY

7.1 MISEQ BARCODED AMPLICON BISULPHITE SEQUENCING OF *MLH1* CpG ISLANDS AND SHORES

Amplicon bisulphite sequencing is a highly targeted approach that enables to analyze DNA modifications in the form of cytosine methylation in specific genomic regions of interest in a large number of samples with high quantitative accuracy. This method uses elements of existing methods (bisulfite conversion and region-specific PCR amplification) and combines them with simple next-generation library construction followed by next-generation sequencing (NGS) using different platforms such as MiSeq System Illumina.

We designed tagged Illumina compatible primers to amplify 10 overlapping regions (200-300 bp) covering *MLH1* CpG islands and shores in order to perform a comprehensive assessment of association between *MLH1* methylation and rs1800734 genotype in 33 fresh-frozen normal colorectal biopsies (Table 12). We also used the same system to investigate the alteration of *MLH1* promoter methylation induced by 5-azacytidine (azaC) in CO-115 and SW48 cell lines.

Bisulfite modification of genomic DNA (300 ng) was performed with EZ DNA methylation Kit (Zymo Research) according to the manufacturer's recommendations.

Bisulfite-modified DNA was amplified (Table 13) and PCR amplicons were diluted 1:100 and barcoded with Illumina compatible primers provided by High Throughput Genomics Unit at Wellcome Trust Centre for Human genetics (Table 14).

The barcoded products were then pooled and gel purified (Qiaquick Gel Purification, Qiagen) to generate a library for sequencing. The purified library was quantified by Qubit Fluorometer 2.0 (Thermo Fisher) and further quantified and assessed using a 2200 Tapestation System (Agilent). The library was diluted to 2nM with sterile water prior to beginning the preparation to run on a MiSeq sequencer (Illumina).

The MiSeq library preparation was done according to manufacturer's instructions: briefly 5 µl of the 2nM library was denatured using 5 µl of 2M NaOH for 5 minutes. This was then diluted to 10pM with HT1 buffer (Illumina). PhiX control DNA (Illumina) was denatured and diluted to 10pM as above. The library and PhiX were mixed in equal volumes to give a 50% PhiX Spike-In and this step was advised by Illumin technicians due to the low complexity of the bisulphite amplicon library.

The reagents of a MiSeq v2 500 Nano sequencing kit (Illumina) were defrosted and prepared for use according to the manufacturer's instructions. The library was loaded into the indicated well of the cartridge and the run initiated by following the on-screen instructions. The output was saved and initial quality control checks carried out using the BaseSpace software (Illumina).

FastQ files were then download and analysed using a bsQC pipeline developed at Ludwig Institute for Cancer Research in Oxford (Fig. 11).

FastQ reads from Basespace (Illumina) were exported and the first step was the estimation of conversion's estimation efficiencies: samples which showed levels of unconverted cytosines > 2% were excluded from further analysis.

Trimming of low quality ends of reads and adapters was performed using Trim Galore, a tool which is used for quality control of the input reads (before and after trimming).

The trimmed reads were mapped to a reference genome using Bismark and the quality of the mapped reads was evaluated using Picard tools and SAMtools.

PCR and optical duplicates were removed using Picard tools and Bismark was used to extract the coverage and the percentage of unconverted reads in individual positions.

Finally, all results were visualised using Imagescale (Matlab) and/or plotted by individual CpG per sample.

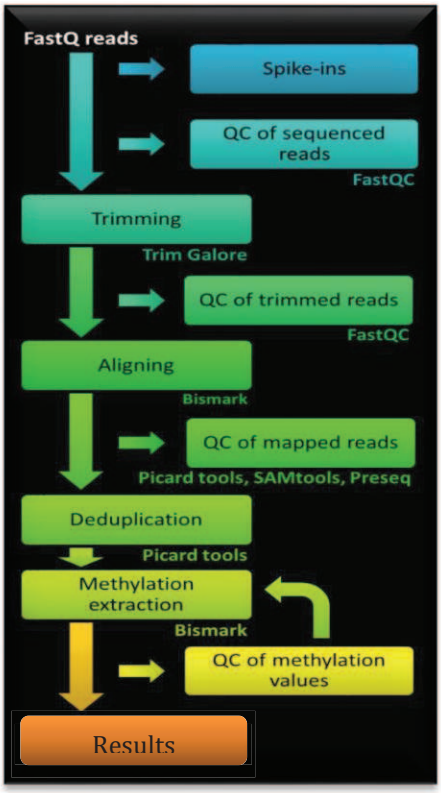


Figure 11. Overview of bsQC pipeline and tools used of the individual steps.

Table 14. Primer sequence of 10 amplicons covering CpG islands and shore within *MLH1* promoter.

	Primer sequence
MLH1 long	Fw 5'-acactctttccctacacgacgctcttccgatctgggaggttataagagtagggttaa-3' Rev 5'-agacgtgtgctcttccgatctaccttcaaccaatcacctca-3'
MLH1 2	Fw 5'-acactctttccctacacgacgctcttccgatctgttattgtttgtttgagaagtgga-3' Rev 5'-agacgtgtgctcttccgatctattcaaaatttcttcacttaaaac-3'
MLH1 3	Fw 5'-acactctttccctacacgacgctcttccgatctagttttaagtgaagaaattttgaa-3' Rev 5'-agacgtgtgctcttccgatcttctaaccataatcaacctaactca-3'
MLH1 4	Fw 5'-acactctttccctacacgacgctcttccgatctttgttgtttgtagggattttagga-3' Rev 5'-agacgtgtgctcttccgatcttccaaatatcacaaaacaaaaaatcct-3'
MLH1 5	Fw 5'-acactctttccctacacgacgctcttccgatcttgtttaggatttttgttttgtga-3' Rev 5'-agacgtgtgctcttccgatcttaacctccttcactcctaaa-3'
MLH1 6	Fw 5'-acactctttccctacacgacgctcttccgatcttttaggagtgaaggaggtta-3' Rev 5'-agacgtgtgctcttccgatcttctcaaactcctcctctcc-3'
MLH1 8	Fw 5'-acactctttccctacacgacgctcttccgatcttgttattaaagagatgattgagaa-3' Rev 5'-agacgtgtgctcttccgatctcacttacactccaaacaaccctta-3'
MLH1 9	Fw 5'-acactctttccctacacgacgctcttccgatctttaagggttgtttggagtgtaagt-3' Rev 5'-agacgtgtgctcttccgatctacccaaaatacatcaacctatcct3'
MLH1 SNP1	Fw 5'-acactctttccctacacgacgctcttccgatctgtatttttgtttttattggttggat-3' Rev 5'-agacgtgtgctcttccgatctaataccttcaaccaatcacctcaa-3'
MLH1 SNP2	Fw 5'-acactctttccctacacgacgctcttccgatctagagttgagaaatttgattggtat-3' Rev 5'-agacgtgtgctcttccgatctaccaattctcaatcatctctttaa-3'

Table 15. *MLH1* amplicons reaction mix and thermic profile.

Reaction mix	
Master Mix	12,5 µl
Primer For	0,5 µl
Primer Back	0,5 µl
Water	10 µl
DNA	1 µl
Total volume	15 µl

94°C	15'	
94°C	30"	
56°C	45"	X 44
72°C	30'	
72°C	10'	
4°C	Hold	

Table 16. DNA barcoding reaction mix and thermic profile.

Reaction mix	
Buffer	2,5 µl
Universal Primer	1 µl
MgCl2	0,75 µl
dNTP Mix	2 µl
Taq Bioline	1 µl
Index Primer	1 µl
DNA	1 µl
Water MQ	16,25 µl
Total Volume	25 µl

95°C	10'	
95°C	15"	
60°C	30"	X 11
72°C	1'	
72°C	3'	
4°C	Hold	

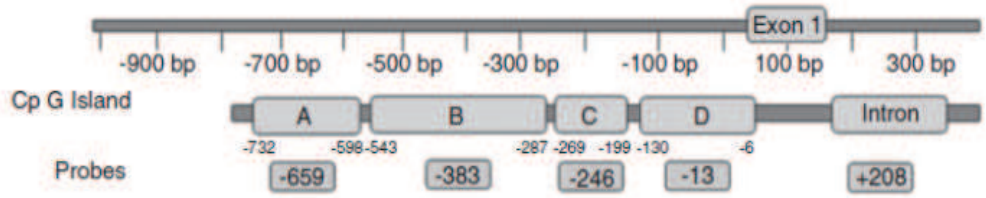
7.2 *MLH1* PROMOTER METHYLATION ANALYSIS USING MS-MLPA

The SALSA MS-MLPA ME011 Mismatch Repair genes (MMR) Kit (MRC-Holland, Amsterdam, The Netherlands) was used to perform *MLH1* promoter methylation analysis on 35 MLH1 immunonegative CRCs previously described (Matherial and Methods, Section 6). ME011 Mismatch Repair genes probemix was developed to detect aberrant CpG island methylation and included five informative probes for *MLH1* (3p22.1). The most important methylation region for *MLH1* expression, the Deng C-region, is from -248 nt to -178 nt before the transcription start site and the second most important region, the Deng D-region, is from -9 nt to +15 nt (Figure 11) (155).

All MS-MLPA reactions were done according to the manufacturer's instructions, using 100-150ng of DNA. The probemix is added to 5µl of denatured DNA and allowed to hybridize for 16 hours at 60°C. Subsequently, the sample is divided in two: one half is ligated by adding 10µl of ligase-mix, whereas in the other half ligation is combined with digestion by adding 10µl of ligation-digestion mix. These samples are incubated for 30 minutes at 54°C, then the HhaI enzyme is inactivated by denaturation at 95°C for one minute. Since the unmethylated sequences are cut by the restriction HhaI enzyme (Promega), this process results in the ligation of the methylated sequences only. Eight microliters of the two aliquots are then amplified in a 25µl PCR reaction using Veriti thermocycler (Applied Biosystems, Foster City, USA) with the following thermal protocol: 33 cycles of denaturation at 95°C for 20 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 1 minute with a final extension of 20 minute at 72°C. In order to assess MS-MLPA reliability two replicates were performed for each sample and positive and negative controls using fully methylated DNA (CpGenome Universal Methylated DNA, Millipore) and unmethylated DNA (CpGenome Universal UnMethylated DNA, Millipore) were included in each MS-MLPA experiment. Aliquots of 1.5µl of the PCR reaction were combined with 0.5µl TAMRA internal size standard (Applied Biosystems, Foster City, USA) and 13.5µl of deionized formamide. After denaturation, fragments were separated and quantified by electrophoresis on an ABI 310 capillary sequencer and analyzed with GeneMapper 4.0. (Applied Biosystems). Values corresponding to peak size (base pairs) and peak height were used for further data processing by Coffalyser V7 software (MRC-Holland).

Methylation dosage ratio (MR) was obtained by the following calculation: MR= (Px/Pctrl)Dig / (Px/Pctrl)Undig where Px is the peak height of a given probe, Pctrl is the sum of the peak heights of all control probes, Dig stands for HhaI digested sample, and Undig stands for undigested sample.

A methylation ratio (MR) for a given gene may range from 0 (0% of alleles methylated) to 1.0 (100% of alleles methylated) and threshold values of 0.3 and 0.7 are suggested by the manufacturer to consider a locus as hemi-methylated or fully-methylated, respectively.



Length (nt)	SALSA MLPA probe	Gene	HhaI site	Ligation site (Distance to ATG start)	Partial sequence with HhaI site	Distance to next probe
237	7187-L7710	MLH1	+	- 659 (Deng, A-region)	TCCGCCACATACCGCTCGTAGTAT-TCGTGCTCAGCCTCGTAGTGGCGCCTGAC	0.1 kb
148	1685-L1265	MLH1	-	- 518 (Deng, B-region)	GGGTCCACTCGGGCCGGAAAACT-AGAGCCTCGTCGACTTCCATCTTGCTTCTTT	0.1 kb
265*	6221-L1747	MLH1	+	- 383 Reverse (Deng, B-region)	CTGCTGAGGTGATCTGGCGCAGA-GCGGAGGAGGTGCTTGGCGCTTCTCAGGC	0.1 kb
198	6222-L7712	MLH1	+	- 246 (Deng, C-region)	AGAGCGGACAGCGATCTCTAACGCGCAA-GCGCATATCCTTCTAGGTAGCGGGCAGTA	0.2 kb
166	1686-L1266	MLH1	+	- 13 (Deng, D-region)	CGTTGAGCATCTAGACGTTTCCTTGGCTCT-TCTGGCGCCAAAATGTCGTTCGTGGCAGG	0.2 kb
292	2258-L1745	MLH1	+	+ 208 (Intron 1, 93 nt after exon 1)	CGGACACGCCTCTTTGCCCGGGCAGA-GGCATGTACAGCGCATGCCCACAACGGCG	

Figure 12. Representation of *MLH1* promoter region and list of MS-MLPA probes (ME-011 MRC-Holland, Amsterdam, The Netherlands). Restriction sites are highlighted in grey.

8. *MLH1* TRANSCRIPTIONAL ANALYSIS

Total RNA extraction of three 8-µm sections of 9 FFPE normal colonic mucosae and 35 FFPE CRCs previously described was performed by manual microdissection using a High Pure FFPE RNA Micro Kit (Roche Applied Science, Mannheim, Germany) according to the manufacturer's protocol.

Purified RNA was quantified using Nanodrop and RNAs were then treated with DNase I to degrade residual DNA and complementary DNA (cDNA) synthesis was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) (Table 17).

Limited-cycle preamplification has been performed before qRT-PCR using the TaqMan PreAmp (Applied Biosystems) kit to increase the amount of specific gene targets.

Expression of *MLH1* (Hs00179866, Applied Biosystems) was evaluated using the Fast Start Universal Taq Man Probe Master kit (Roche Diagnostic, Mannheim, Germany) on the ABI 7900HT cycler (Applied Biosystems) with GAPDH serving as an endogenous control (4332649, Applied Biosystems). Each sample was analyzed performing qRT-PCR reactions in duplicates in 96-well plates (Table 18).

After normalization to the endogenous control, the relative expression (RE) of *MLH1* transcript in tumour samples was obtained by comparing the normalized expression values in tumours versus those observed in normal colonic mucosae, calculated by the equation $FC=2^{-\Delta\Delta Ct}$, as previously described.

The same method was applied to investigate MLH1 expression in CO-115 and SW48 cell lines subjected to 5-azacytidine (azaC) treatment.

Table 17. Retrotranscription reaction mix and thermic profile.

10 X RT Buffer	2 µl	25°C	10'
10X RT Random Primers	2 µl	37°C	120'
dNTPS Mix	0,8 µl	85°C	5'
Multiscribe Reverse Transcriptase	1 µl		
RNA (40 ng/ul)	10 µl		
Water	4,2 µl		
Total volume	20 µl		

Table 18. qRT-PCR reaction mix and thermic profile.

Reaction mix	
Master Mix	10 µl
Probe	1 µl
Water	8 µl
cDNA	1 µl
Total volume	20 µl

95°C	20"	
95°C	1'	
60°C	20"	X 40

9. DEMETHYLATION ASSAYS BY 5-AZA-2' DEOXYCYTIDINE

Cell lines were grown in Dulbecco Modified Eagles Medium or RPMI-1640 supplemented with 10% fetal bovine serum and 1% penicillin streptyomycin (Sigma) at 37°C in 5% CO2.

Adherent semiconfluent MSI+ CO-115 and SW48 cells in exponential growth were treated with 5uM of 5-Aza-2′-deoxycytidine (AzaC) in a standard medium (AzaC, Sigma A3656) for 48 hours (with replenishment of AzaC after 24 hours). AzaC was removed, cells washed with PBS, and then cultured in standard medium for 4, 7 and 11 days. RNA and DNA were extracted simultaneously using the AllPrep kit (Qiagen) and MLH1 mRNA expression and promoter methylation assessed as described above.

RESULTS

FIRST PART

1. GENE MUTATION ANALYSIS

1.1 *KRAS*, *NRAS*, *BRAF* AND *PI3KCA* MUTATION ANALYSIS IN COLORECTAL ADENOMAS

A total of 168 adenomas were analyzed for mutations in *KRAS, NRAS, BRAF,* and *PI3KCA* genes. The series included 52 adenomas from 18 MAP patients, 36 adenomas from 17 FAP/AFAP patients and 80 sporadic adenomas from 62 individuals.
As reported in Material and Methods section, the subset of sporadic lesions was specifically selected to include 45 patients with 56 adenomas (S-Ads) without CRC development after ten-year follow-up and 17 patients with 24 adenomas (S-AdsC) who had developed a CRC at least one year after polyp removal.

As expected *KRAS* was the most frequently mutated gene: *KRAS* mutations were more common in MAP (40%, 21/52 cases) and sporadic adenomas (39%, 31/80 cases) compared with FAP/AFAP (22%, 8/36 cases); *NRAS* mutations were observed in only 8% of MAP adenomas (4 cases) and 1% of sporadic adenomas (1 case) while no mutations were found in *BRAF* and *PI3KCA* genes (Figure 13A and Supplementary Tables S1–S3).
Taking into consideration all the alterations, MAP and sporadic adenomas were significantly more mutated than FAP/AFAP lesions ($p=0.02$ and $p=0.09$, respectively).
In addition, within the subset of sporadic patients, gene mutation frequencies were slightly higher in S-AdsC compared with S-Ads (46%, 11/24 cases *versus* 38%, 21/56, respectively) although this difference was not statistically significant.

The spectrum of *KRAS/NRAS* mutations was different in the three sets of premalignant lesions: 80% of the MAP-mutated adenomas exhibited the c.34G>T transversion (p.G12C) in *KRAS* or *NRAS* (90% of *KRAS* mutations and 50% of *NRAS* mutations in MAP adenomas). By contrast, this mispair was totally absent in FAP/AFAP and sporadic groups ($p<0.0001$), which were both enriched for *KRAS* p.G12V, p.G12D and p.G13D mutations (Figure 13B and Supplementary Tables S1–S3).

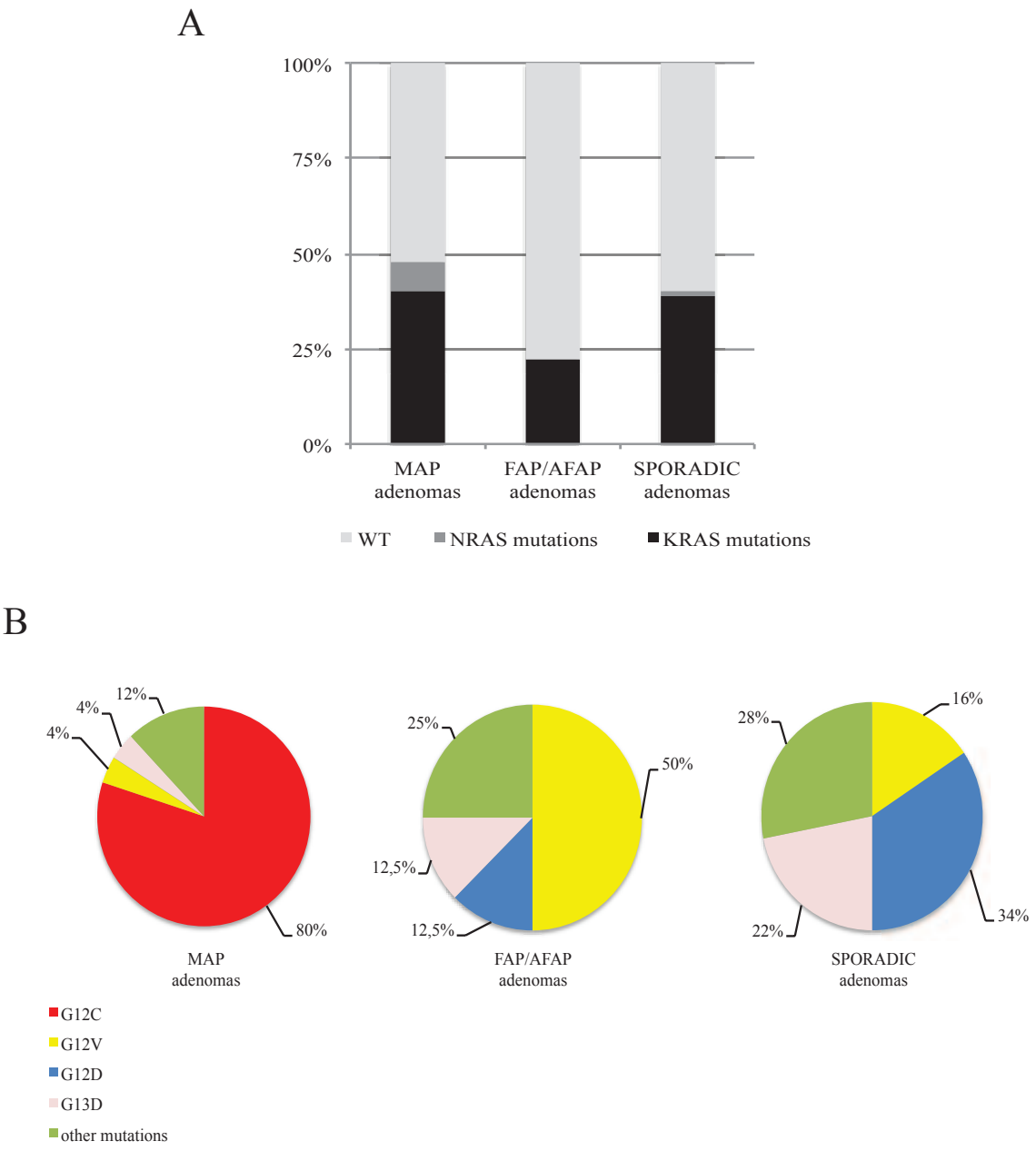


Figure 13. *KRAS/NRAS* mutations in the adenoma cohort. (A) Percentage of mutations in MAP FAP/AFAP and sporadic adenomas; columns represent the mutation frequencies and *KRAS/NRAS* mutations or wild-type (WT) status are indicated in gray scale as reported as per the legend below. (B) Spectrum of *KRAS/NRAS* mutations in different sets of adenomas; the adenoma groups (MAP, FAP/AFAP and sporadic) are reported in pie charts and colours show the different types of *KRAS/NRAS* alterations as per the legend below.

In addition, evaluating both number and type of *KRAS/NRAS* mutations in multiple adenomas of the same patient, 80% of the MAP patients (12/15 cases with adenomas) showed at least one mutated polyp compared with 35% of the FAP/AFAP (6/17 cases) (p=0.02) (Figure 14A-B). As expected, 10 out of 15 MAP patients exhibited one or more *KRAS/NRAS* p.G12C-mutated lesions (Supplementary Table S1).

Finally, considering the cohort of patients with sporadic adenomas, 23% (14/62 cases) showed multiple adenomas that ranged from two to four lesions and within this group

we observed that 64% of patients (9/14 cases) exhibited at least one mutated adenoma (Figure 15A-B).

KRAS/NRAS alterations were found in both tubular and tubulovillous MAP adenomas. However, MAP adenomas were more frequently mutated than FAP/AFAP or sporadic adenomas with the same morphology (P=0.003 and P=0.03, respectively; Supplementary Figure 1).

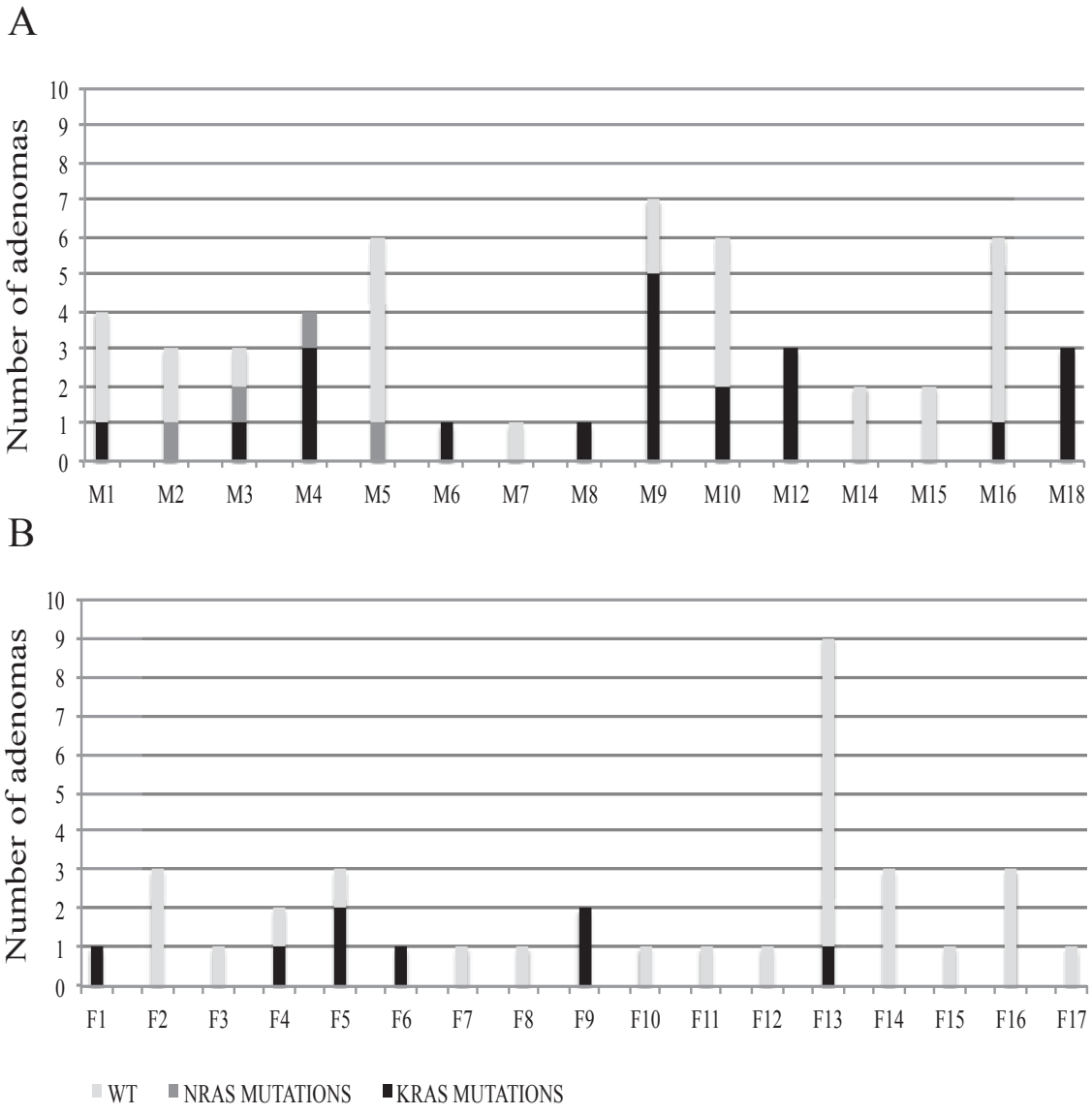


Figure 13. (A) Mutational status of adenomas in MAP, (B) FAP/AFAP. Columns represent the number of analysed adenomas (y axis) for each patient (x axis), whereas types of mutations and wild-type condition are reported in grayscale as per the legend below. WT, wild type.

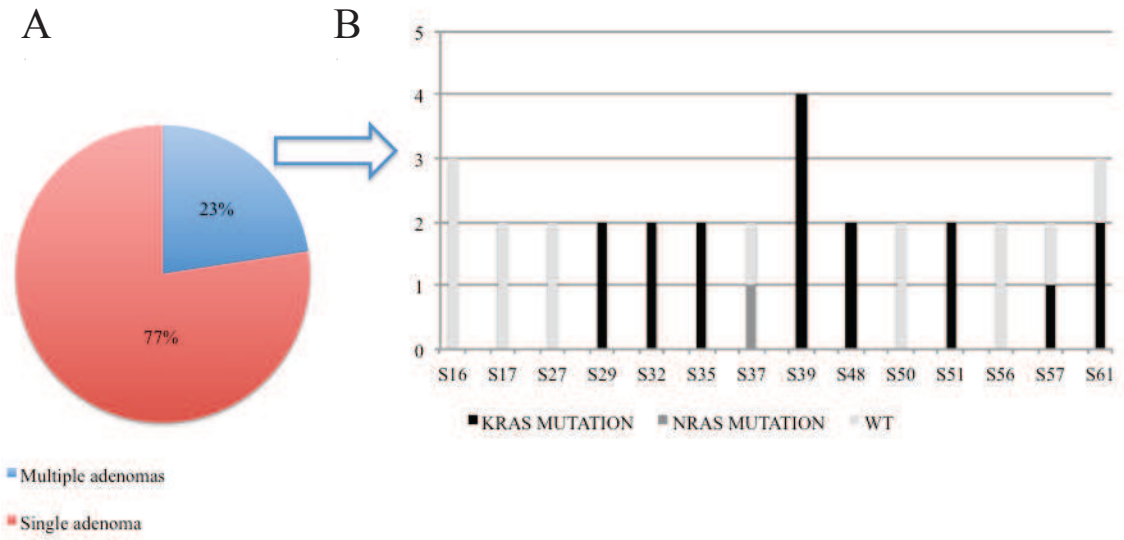


Figure 15. (A) Pie chart shows the percentage of patients with single or multiple sporadic adenomas; (B) Mutational status of multiple sporadic adenomas. Columns represent the number of analysed adenomas (y axis) for each patient (x axis), whereas types of mutations and wild-type condition are reported in grayscale as per the legend below. WT, wild type.

1.2 *KRAS*, *NRAS*, *BRAF* AND *PI3KCA* MUTATION ANALYSIS IN MATCHED ADENOMAS AND CRC OF THE SAME PATIENTS

We investigated the occurrence of metachronous or synchronous CRCs in all patients included in the study: none among FAP/AFAP patients developed a CRC, whereas within the MAP subset, 50% of the patients (9/18 cases) exhibited at least one CRC (p=0.001) (Figure 16) (Supplementary table S1, S4). Interestingly, three MAP patients were diagnosed with CRC in the absence of adenomas and *KRAS* p.G12C mutation was observed in two out of these three CRCs. In the remaining six MAP patients we observed at least one CRC in presence of multiple adenomas and G>T transversion in codon 12 (p.G12C or p.G12V) of *KRAS* gene was the only substitution observed (Supplementary Figure 2).

In the subset of sporadic patients, we could examine gene mutations in 15 CRCs and matched adenomas of the same patients and we found that 67% of CRCs (10 of 15 tumors) showed *KRAS* mutations but also that 47% of these patients showed at least one *KRAS* mutated adenoma.

As expected, the spectrum of *KRAS* mutations in the 10 mutated CRCs was extremely heterogeneous: 20% (2/10 cases) showed p.G12D mutation, 10% p.G13D (1/10 case), 10% p.G12V (1/10 case) and 60% (6/10 cases) other mutations (Supplementary table S5 and Supplementary Figure S2).

None of the MAP CRC exhibited *NRAS* or *BRAF* mutations, while one *NRAS* mutation and one *BRAF* mutation were observed among the sporadic CRCs (Supplementary table S4-S5).

Finally, we detected *PIK3CA* mutations exclusively in MAP CRCs (M1, M6, M17 and M18) and always concurrently with *KRAS* pG12C (Supplementary table S4).

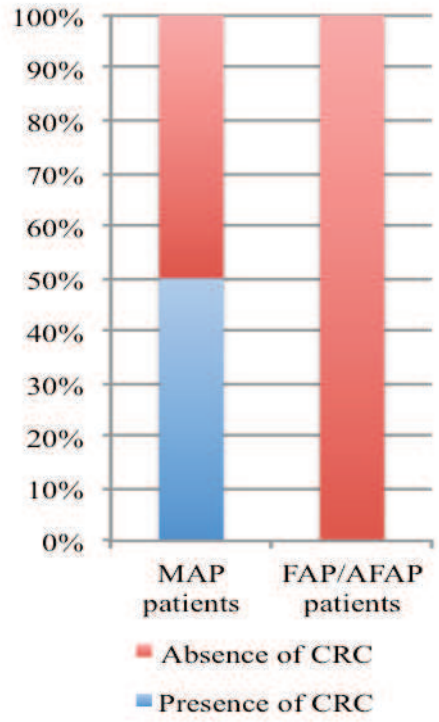


Figure 16. Columns represent the percentage of MAP and FAP/AFAP patients with at least one CRC.

2. L1 AND L1-*MET* HYPOMETHYLATION LEVELS IN COLORECTAL ADENOMAS AND CRCs

As second approach to identify valuable risk biomarkers for colorectal carcinogenesis, we then evaluated hypomethylation levels of the L1 sequences in the three subsets of patients with adenomas previously analyzed. This analysis was performed in a total of 168 adenomas using the two assays L1 and L1-*MET* and a linear correlation between them was also demonstrated (Pearson r=0.6, p<0.0001) (Figure 17).

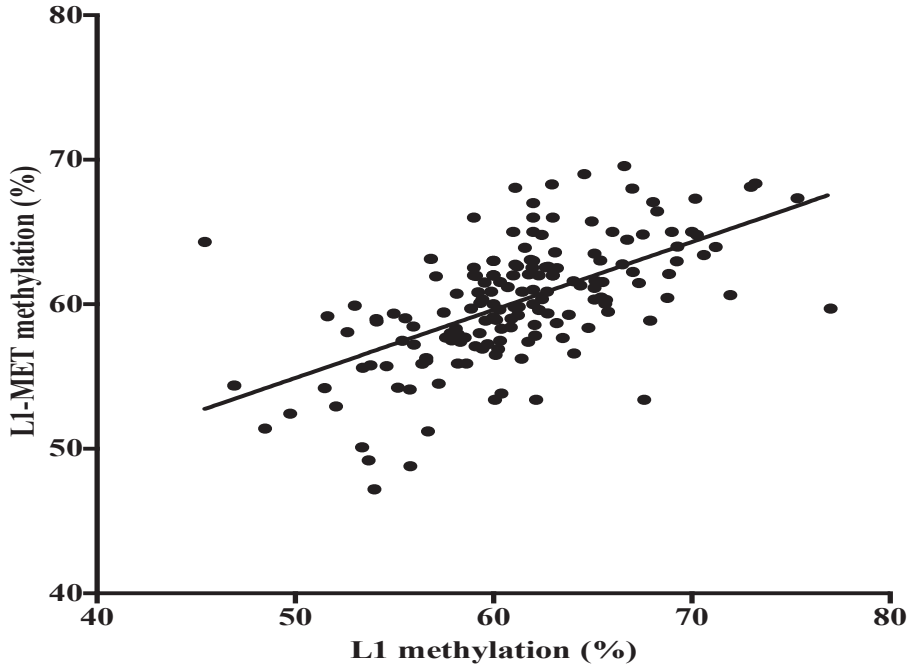


Figure 17. Scatterplot shows a linear correlation between the values of L1 and L1-*MET* methylation of the all 168 adenomas analyzed.

By using the lowest methylation value of normal mucosa (60%) as the methylation threshold, MAP and sporadic adenomas showed a similar hypomethylation profiles and exhibited a significantly higher frequency of hypomethylated samples compared to FAP/AFAP adenomas (p=0.034, p=0.005 and p<0.0001, p=0.0002 respectively) (Figure 18). Analyzing the levels of L1 and L1-*MET* methylation within the subsets of sporadic adenomas, we also observed than S-AdsC were more frequently hypomethylated (54,2% and 66,7% of cases with L1 and L1-*MET* analyses, respectively) than S-Ads (26,8% and 53,4% of cases with L1 and L1-*MET* analyses, respectively) with both assays.

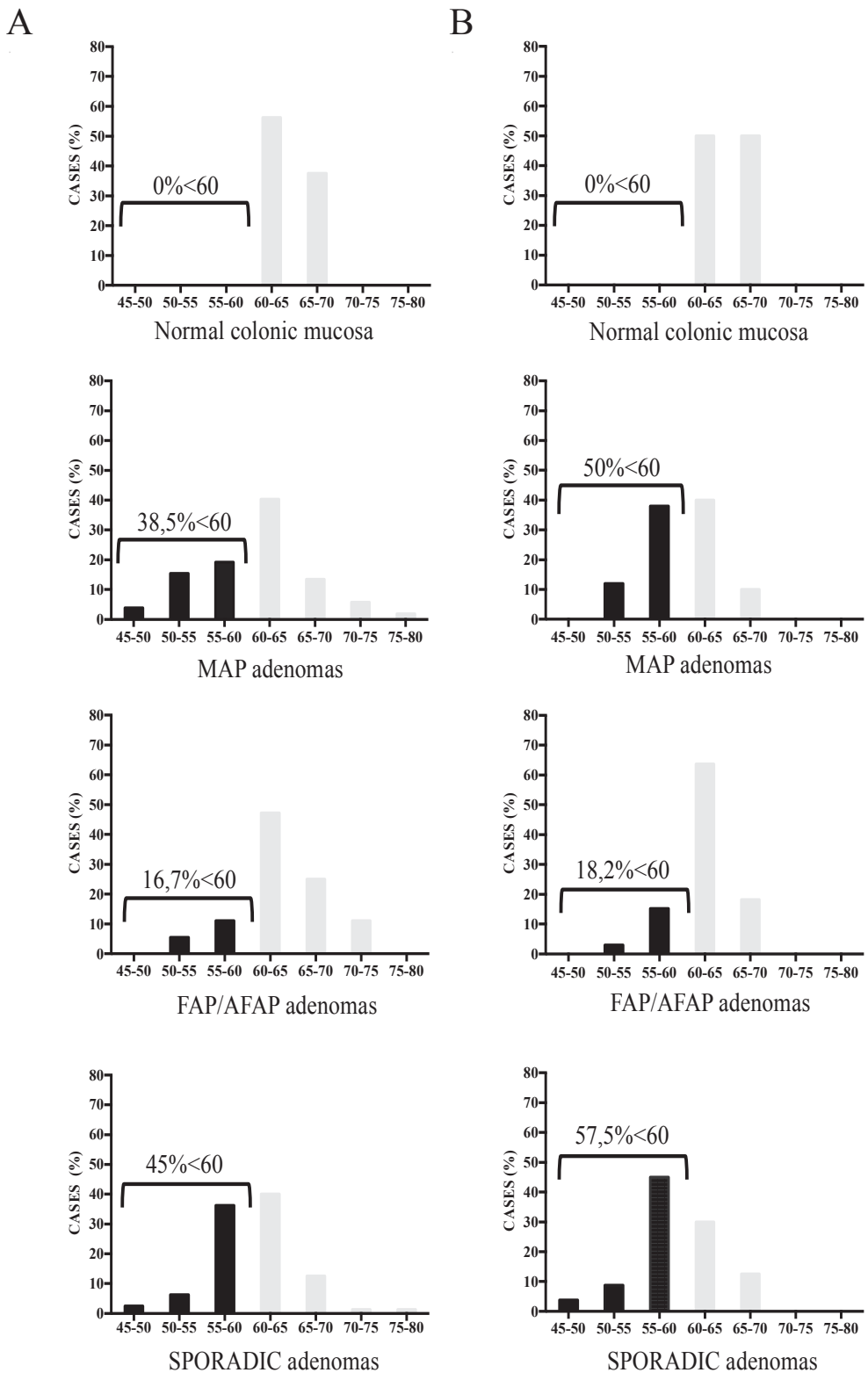


Figure 18. (A) L1 and (B) L1-*MET* methylation percentages in normal colorectal mucosa, different adenoma groups (MAP, FAP/AFAP and sporadic); the methylation threshold was set at 60% as the lowest methylation value of normal mucosa for both L1 and L1-*MET*; black columns represent hypomethylated samples (<60%), while grey columns identify samples with a methylation level ≥60%.

To investigate the role of DNA hypomethylation during CRC cancerogenesis, we analyzed the levels of L1 and L1-*MET* methylation in 11 MAP and 15 sporadic CRCs included in this study.

As expected, these two groups exhibited a significantly higher frequency of hypomethylated samples with respect to adenomas showing an increase of hypomethylation during tumor progression (63,6% and 73,3%, 90,9% and 78,6% of cases with L1 and L1-*MET* analyses, respectevely (Figure 19).

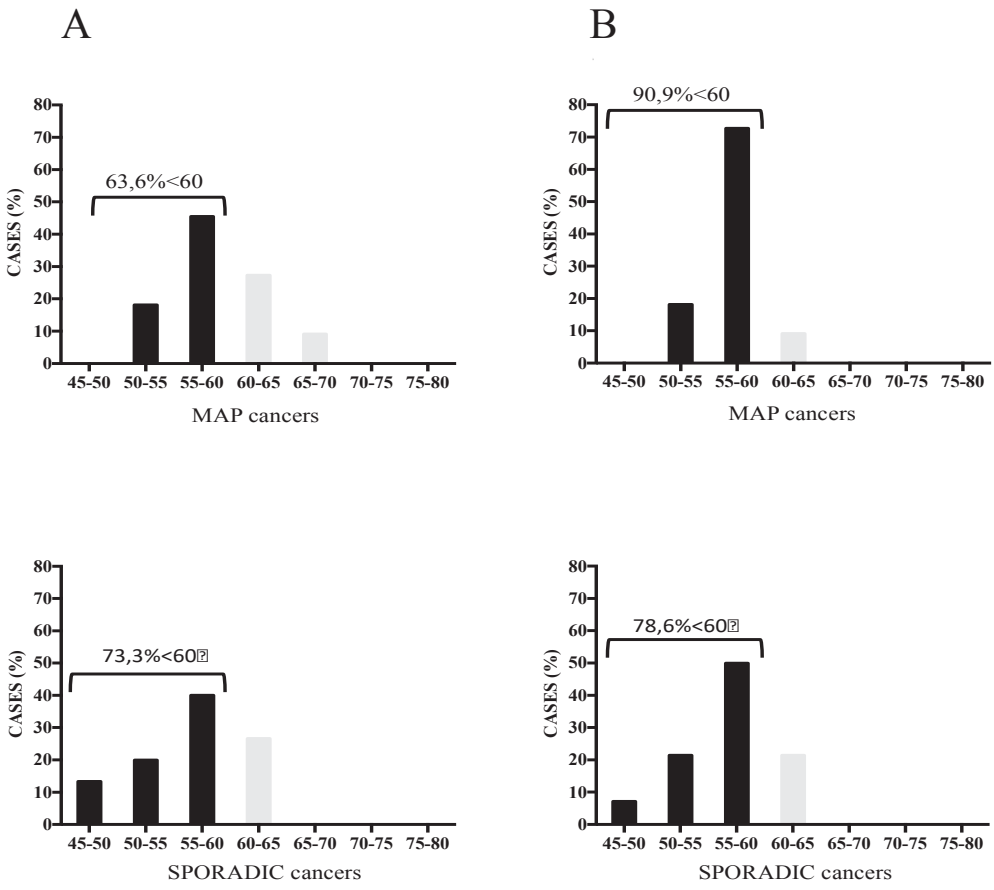


Figure 19. (A) L1 and (B) L1-*MET* methylation percentages in MAP and sporadic cancers; the methylation threshold was set at 60% as the lowest methylation value of normal mucosa for both LINE-1 and L1-MET; black columns represent hypomethylated samples (<60%), while grey columns identify samples with a methylation level >60%.

Considering the distribution of L1 and L1-*MET* percentages in the multiple adenomas of each patient, methylation variability was higher in MAP and sporadic cases than FAP/AFAP patients.

Notably, we found that at least one hypomethylated adenoma, evaluated with L1 assay, was detected in most of the MAP and sporadic patients (83%, 10/12 cases and 79%,

11/14 cases, respectively) respect to FAP/AFAP (43%, 3/7 cases) individuals (Figure 20).

Similar results were also obtained with L1-*MET* assay and are reported in Supplementary Figure S3: in detail, MAP and sporadic patients showed at least one hypomethylated adenoma in 81% (9/11 cases) and 79% (11/14 cases), while FAP patients in 60% of individuals (3/5 cases).

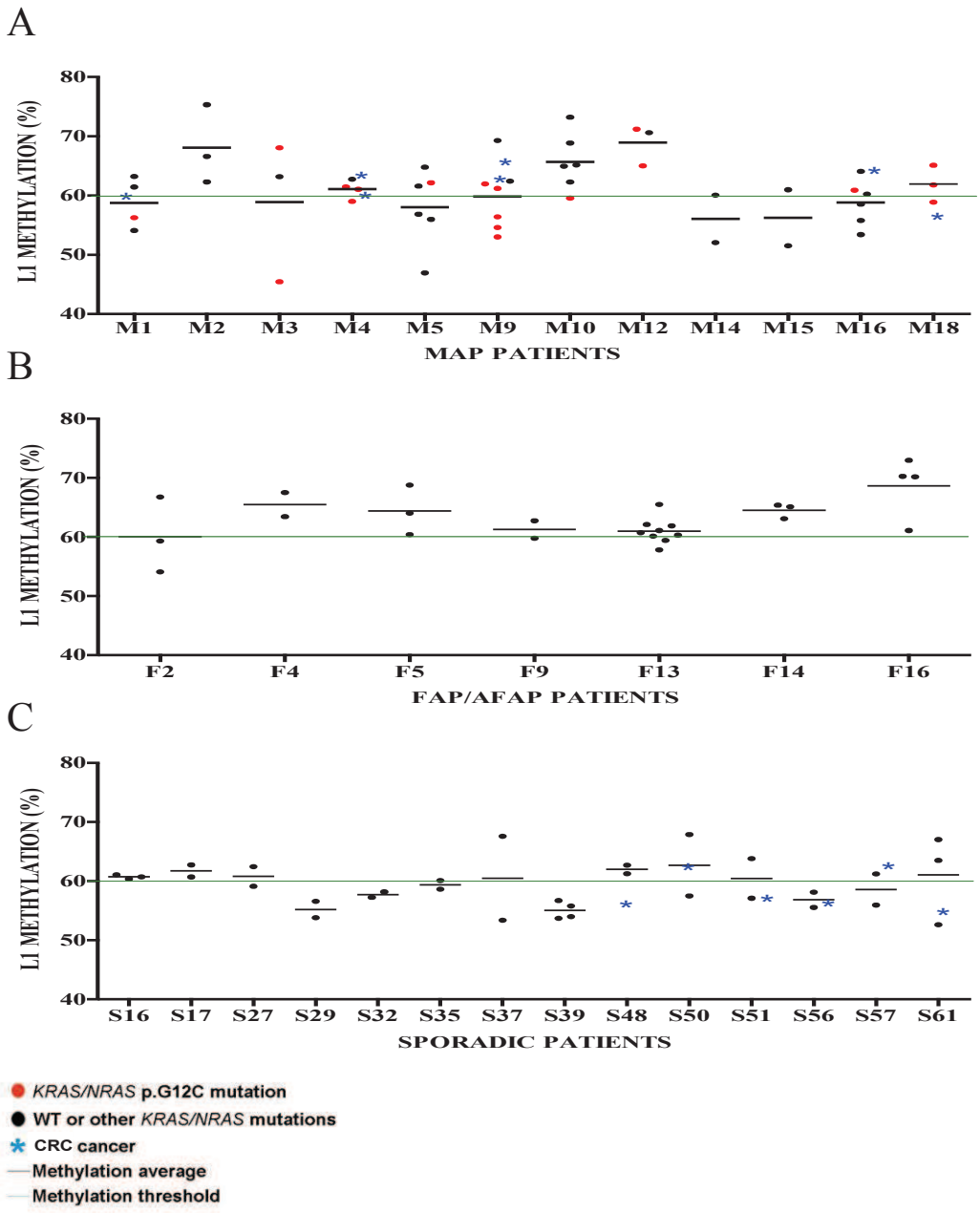


Figure 20. (A) Dot plots represent L1 methylation percentage in the multiple adenomas of MAP (A), FAP/AFAP (B) and sporadic (C) patients; black dots identify adenomas and red dots indicate *KRAS*/*NRAS* p.G12C mutated polyps; the green line symbolizes the methylation threshold (60%) for L1 assay; bars show the mean values of L1 methylation percentage among adenomas of the same patients and light blu asterisks show the methyilation level of the corresponding carcinomas; symbols and colours are reported as per the legend below. WT, wild type

3. EXPRESSION ANALYSIS AND CHARACTERIZATION OF L1-CHIMERIC TRANSCRIPTS

3.1 RNA-SEQ ANALYSIS ON CRCs CHARACTERIZED FOR L1 HYPOMETHYLATION LEVELS

We have investigated the causal link between local L1-hypomethylation and aberrant L1 chimeric transcripts (LCT) activation in genes harboring specific intronic L1 sequences.

In silico analysis have revealed up to 911 new putative LCT across the genome, but only a few of these elements are implicated in cancer (81).

In colorectal tumorigenesis, previous studies have demonstrated that the LCT expression may cause inadvertent activation of multiple proto-oncogenes including *MET*, *RAB3IP*, *ACVR1C* or blockade of onco-suppressor genes like *TFPI-2* (84, 174). Therefore, in this work we focused on these four LCTs potentially involved in CRC development.

RNA-Seq analyses of different transcript variants of these four genes were performed on 13 FFPE samples of histologically normal colonic mucosa and 32 FFPE CRCs, previously characterized in our lab for their clinico-pathological and molecular profiles (Sahnane 2015). This cohort of 32 CRCs were divided in four distinct groups based on L1 methylation levels: L1-group (L1>60.1%), L2-group (54.1%-60%), L3-group (45.8%-54%) and L4-group (<45.6%) and tumor RNA from cancers belonging to the same group were pooled.

Using the UCSC Genome Browser Annotation and through specific bioinformatic tools, we found 4 transcript variants of *ACVR1C*, 8 for *RAB3IP*, 3 for *TFPI-2* and 16 for *MET*.

As the Heat Map shows (Figure 20), no significant difference of mRNA expression of *ACVR1C, TFPI-2* and *RAB3IP* variants was observed when comparing L1/L2/L3/L4 CRC groups. Only for the oncogene *MET*, expression levels of 3 transcript variants have surprisingly shown to be correlated significantly and in a linear fashion with L1 hypomethylation levels among the 4 CRC groups.

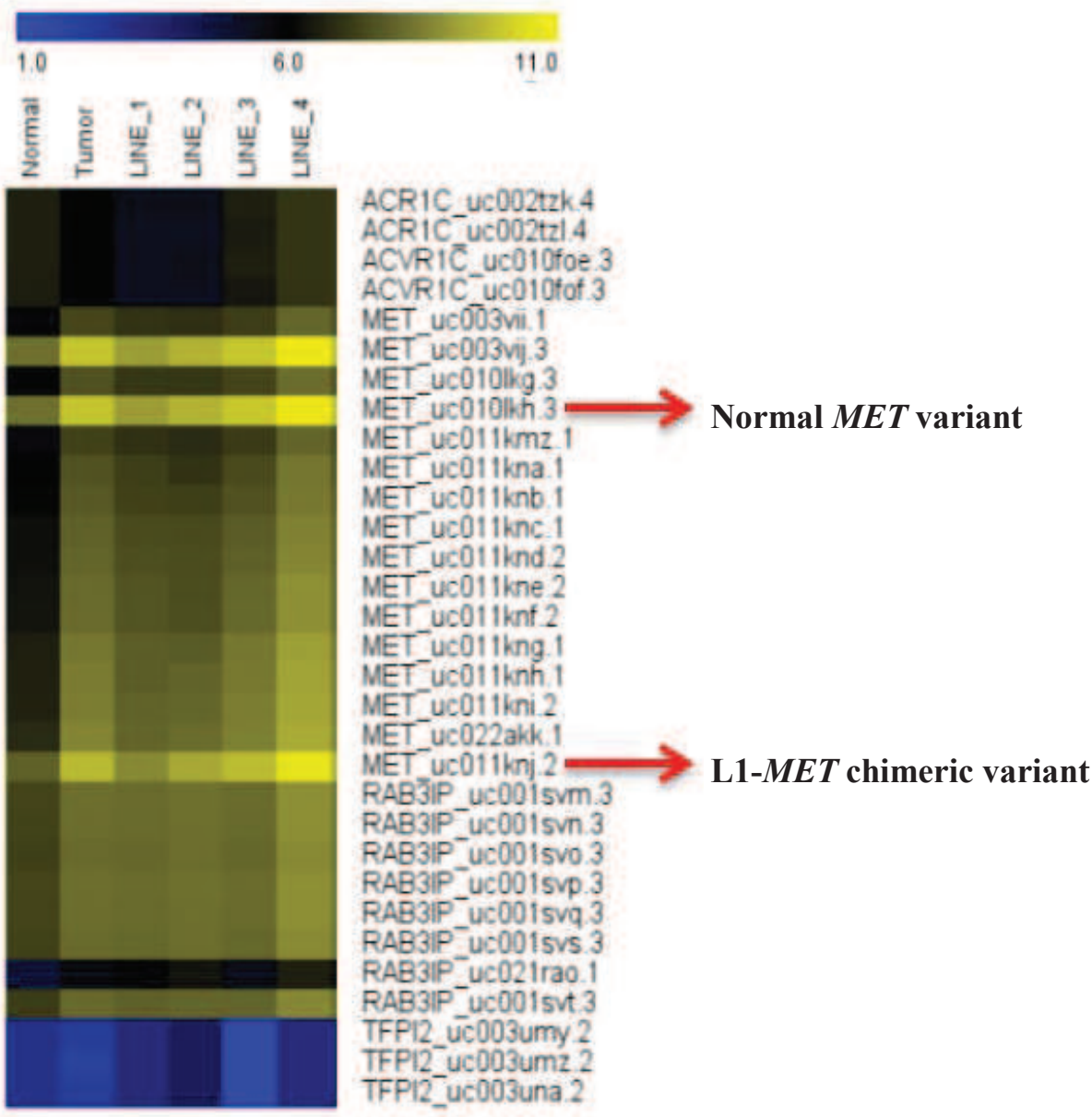


Figure 21. Heat Map of RNA-SEQ about *ACVR1C*, *MET*, *RAB3IP* and *TFPI2* transcript variants in colorectal tissue samples (Normal), CRCs pool (Tumor), CRCs L1 (LINE_1), CRCs L2 (LINE_2), nei CRCs L3 (LINE_3), CRCs L4 (LINE_4).

In detail, these 3 variants corresponded to: normal *MET* transcript and variant 2 of UCSC annotation with the same length respect to normal transcript; L1-*MET* fusion transcript, composed by L1 element within the intron 2 and the *MET* sequence from exon 3 to exon 21 (Figure 22).

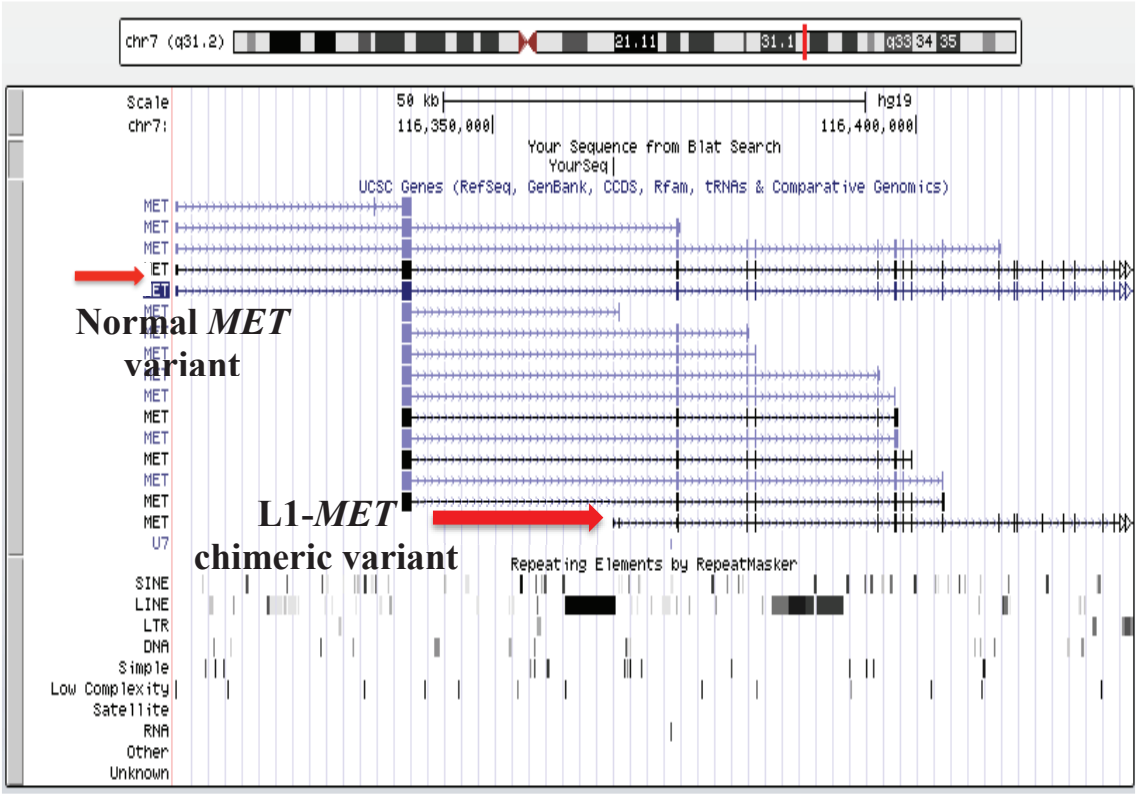


Figure 22. UCSC Genome Browser analysis of 16 *MET* transcript variants.

3.2 *IN SILICO* CHARACTERIZATION OF THE L1-*MET* TRANSCRIPT

In-silico analyses confirmed the presence of the ORF0-containing L1 element, as reported (175-176), within the second intron of *MET*, between nucleotides g.51092C (c.1201-13172C) and g.57084C (c.1201-7191C) (GenBank Accession; #M80343.1) (GenBank Accession; #NG_008996.1) (Figure 23). The first methionine of ORF0 mapped to nucleotide g.56654A (c.1201-7610A) of *MET* and its first nucleotide was taken as +1 of the L1- *MET* reference numbering.

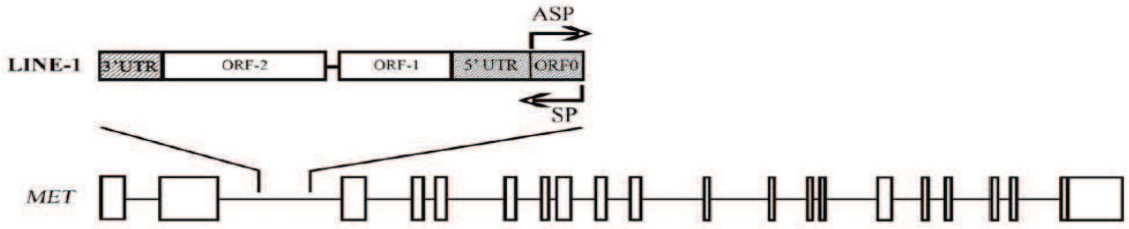


Figure 23. Graphical representation of the L1 element located within intron 2 of MET; L1 is inserted in the opposite orientation of MET with the sense promoter (SP) and the 5'-UTR encompassing the recently identified ORF0 sequence; ORF contains the antisense promoter (ASP) leading to the development of a chimeric L1-MET transcript (176).

In order to evaluate the *MET* and L1-*MET* transcript expression, we designed a new assay to discriminate normal and chimeric variant.

The forward PCR primer for L1-*MET* transcript was located inside the exon 1 of L1-*MET* that contains also the L1 promoter sequence, and the reverse primer was designed within the exon 2 of LI-*MET* (Figure 24).

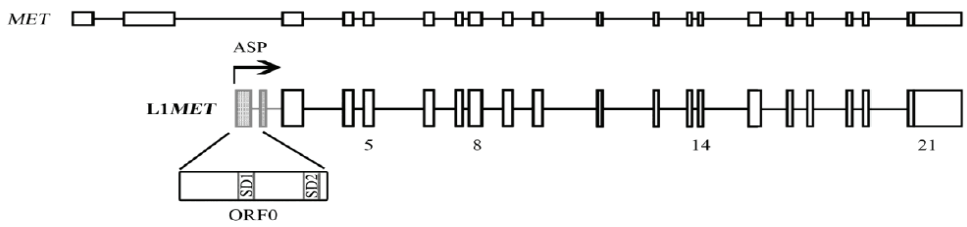


Figure 24. Graphical representation of the L1-*MET* transcipt arising from the L1 element located within intron 2 of *MET* (176).

The assay's specificity was tested by running the specific 81 bp amplicon on 3% agarose gel and by sequencing the specific band taking advantage of analysis of4 CRC samples already tested by RNA-Seq and resulted positive for *L1-MET* transcript (Figure 25).

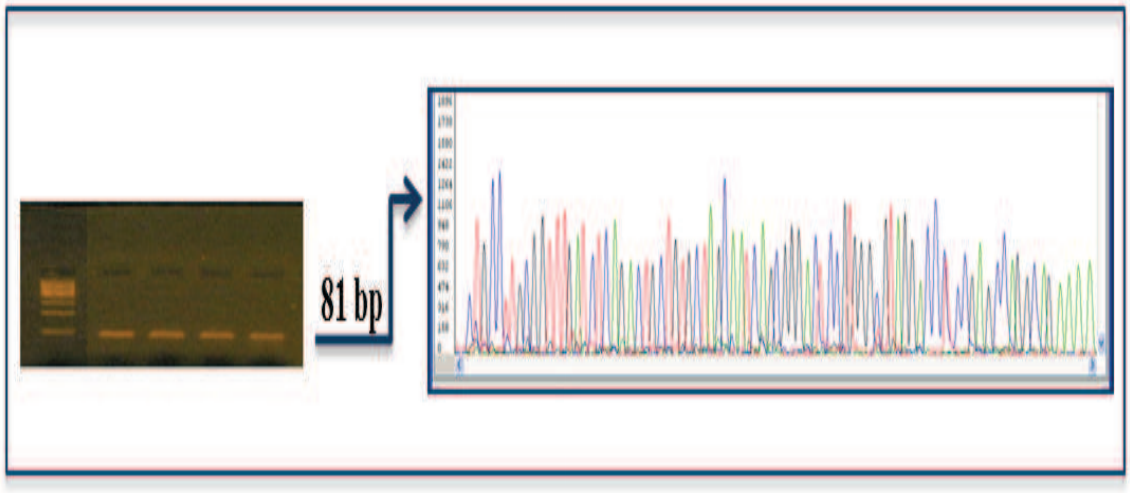


Figure 25. *L1-MET* transcript amplification of 4 CRCs run on 3% agarose gel and the corresponding sequences.

3.3 HIGH LEVELS OF *MET* AND L1-*MET* TRANSCRIPTS IN HYPOMETHYLATED CRCs

In order to validate RNA-Seq results, we performed qRT-PCR analysis to evaluate the expression levels of *MET* and L1-*MET* in cDNA samples from 9 L1 (>60,1%) and 11 L4 (<45,6%) hypomethylated CRC and the results of these two tumor subsets with those obtained from 8 normal mucosa cDNA samples (Figure 26A).

Interestingly, we observed a significant up-regulation of both transcripts ($p<0.0001$) in L4 *vs* L1 CRCs, confirming RNA-Seq data previously obtained (Figure 26B) (Supplemntary table S6).

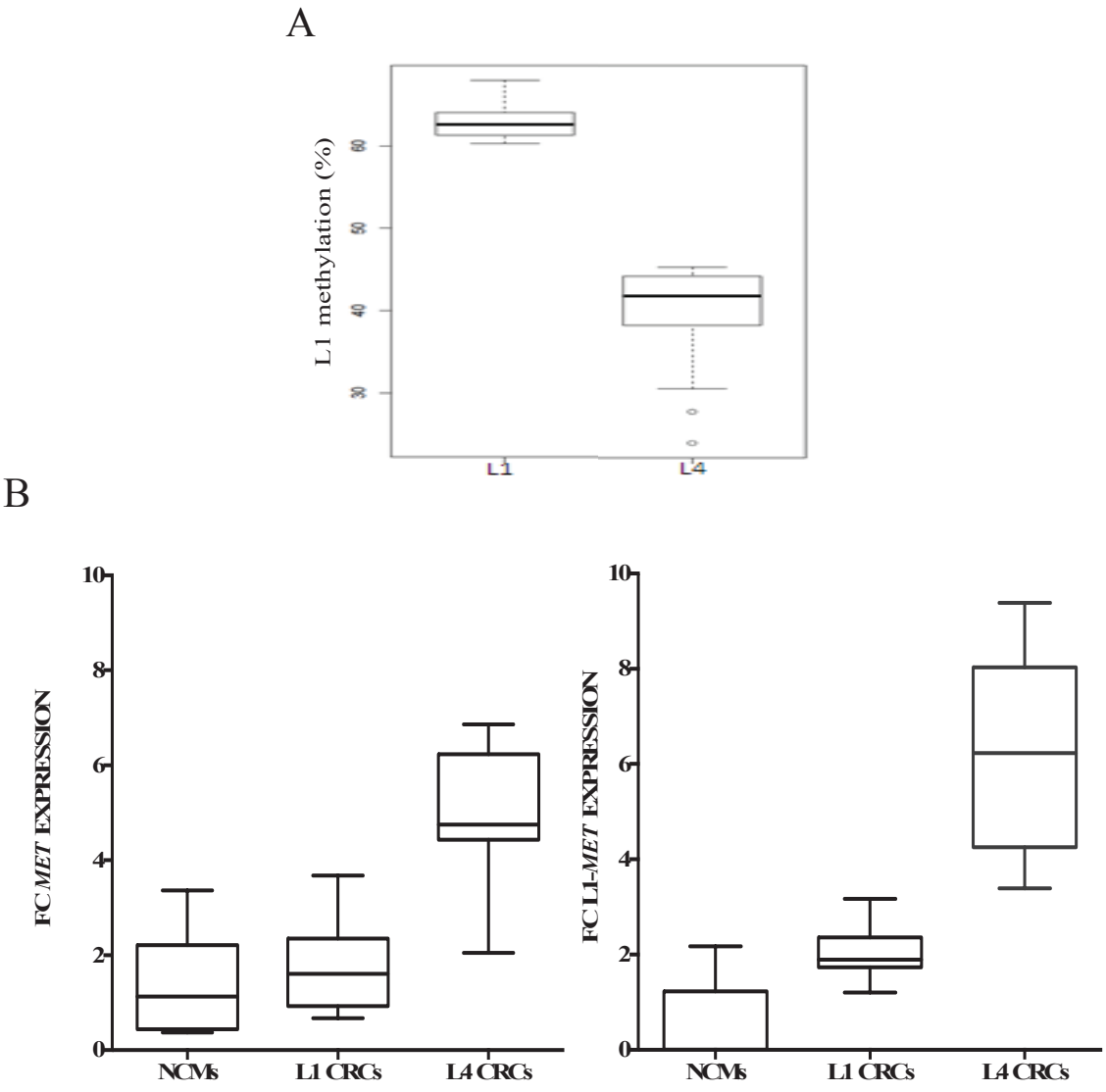


Figure 26. (A) Clustering analysis of L1 and L4 groups stratified for L1 methylation levels: L1 (9 CRCs, mean 63,1), L4 (11 CRCs, mean 40,2). (B) Boxplots show a significantly up-regulation of *MET* and *L1-MET* expression level in L4 group compared with those observed in L1 and NCMs (normal colon mucosa samples). FC, Fold Change.

3.4 HIGH LEVELS OF *MET* AND L1-*MET* TRANSCRIPTS IN HYPOMETHYLATED COLORECTAL ADENOMAS

In order to confirm the in *silico* results also in preneoplastic lesions, *MET* and L1-*MET* transcripts were evaluated in a cohort of 85 adenomas (13 MAP, 16 FAP and 56 sporadic) and 8 normal colorectal mucosa samples that we used as calibrator to calculate the fold change expression.

As box plot shows in Figure 27A, *MET* transcript was expressed at minimal basal level in all colorectal mucosa samples while an up-regulation was observed in 53% (45/85 cases) of colorectal adenomas using as cut-off the 2X fold change (Supplementary table S6).

Likewise, all colorectal mucosa samples exhibited levels of L1-*MET* transcript below the cut off whereas an up-regulation of the transcript was detected in 36% (31/85 cases) of colorectal adenomas (Figure 27B) (Supplementary table S7).

Notably, a significant inverse correlation between L1-*MET* hypomethylation and *MET*/L1-*MET* transcript levels was observed in 75% and 51% of cases ($p<0.0001$) (Figure 28A and Figure 28B).

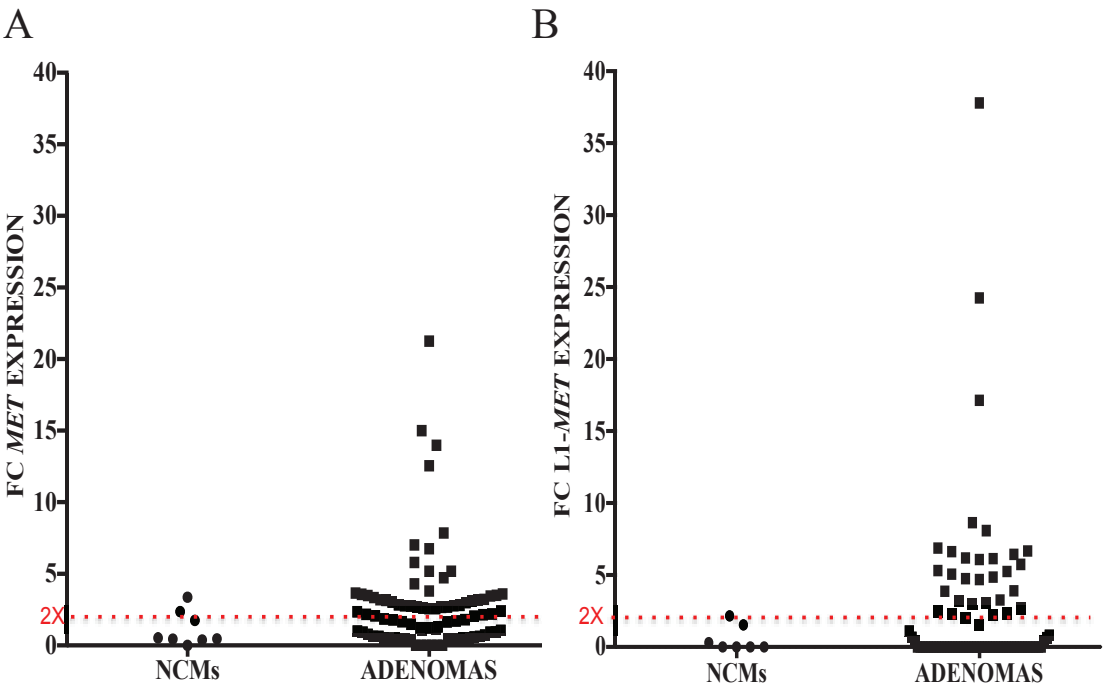


Figure 27. (A) Level of *MET* and (B) L1-*MET* transcripts expression in 8 normal colon mucosa samples (NCMs) and 85 adenomas. FC, Fold Change.

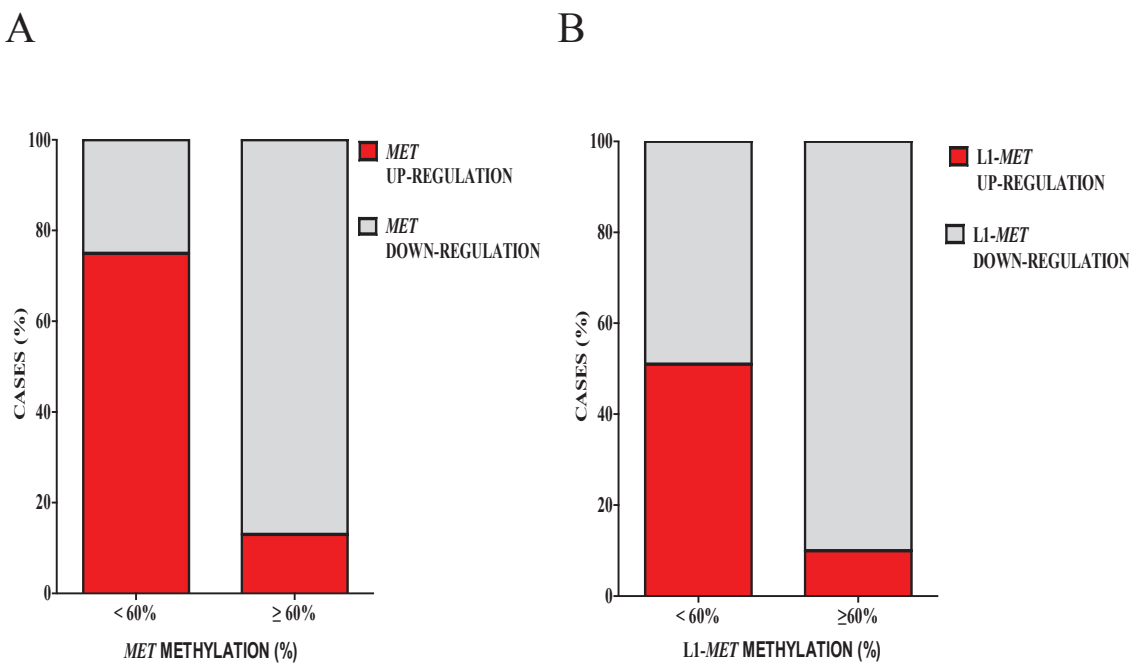


Figure 28. Columns represent the percentage of (A) *MET* and (B) L1-*MET* up-regulated and down-regulated samples with L1-*MET* methylation<60% and L1-*MET* ≥60%. FC, Fold Change.

Moreover, 42% (23/55) of L1-*MET* hypomethylated adenomas showed up-regulation of both transcripts, while 42% (23/55) of these cases exhibited upregulation of either one or the other transcript. The remaining L1-*MET* hypomethylated adenomas (9/55; 16% of cases) exhibited down-regulation of both transcripts (Figure 29).

Among samples with normal L1-*MET* methylation levels, we observed that 17% of cases showed up-regulation of either one or the other transcript (5/30) while only 1 case exhibited down-regulation of both transcripts; the remaining cases (24/30) showed down-regulation of MET and L1-*MET* expression ($p<0.0001$).

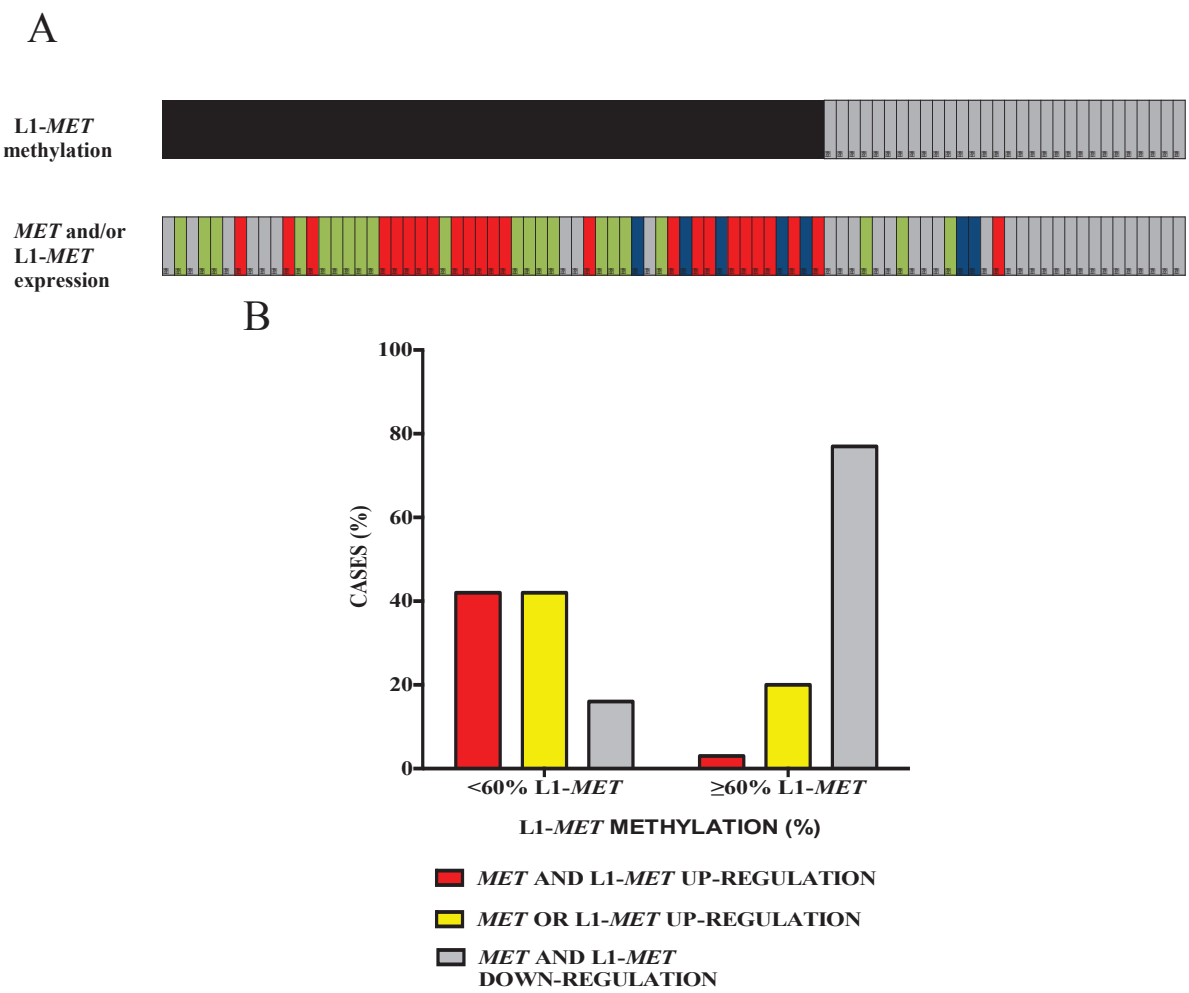


Figure 29. (A) Top panels indicated samples with L1-hypomethylation (<60%; black) or without L1-hypomethylation (grey) ≥*60*% and samples with *MET* up-regulation (green), L1-*MET* up-regulation (blue), both transcritps up-regulation (red) and down-regulation (grey). (B) Colums represent the percentage of *MET* and/or L1-*MET* up-regulated samples with L1-*MET* methylation<60% and L1-*MET* ≥60%.

4. CORRELATION BETWEEN CLINICO-PATHOLOGICAL AND MOLECULAR ANALYSIS IN SPORADIC ADENOMAS

Correlation analysis between molecular and clinico-pathological markers was possible for the subset of 80 sporadic adenomas. In detail, we correlated *KRAS*/*NRAS* mutational status, L1/L1-*MET* methylation levels and *MET*/L1-*MET* transcript expression with the 5 clinico-pathological features currently used for classification of high risk adenomas.

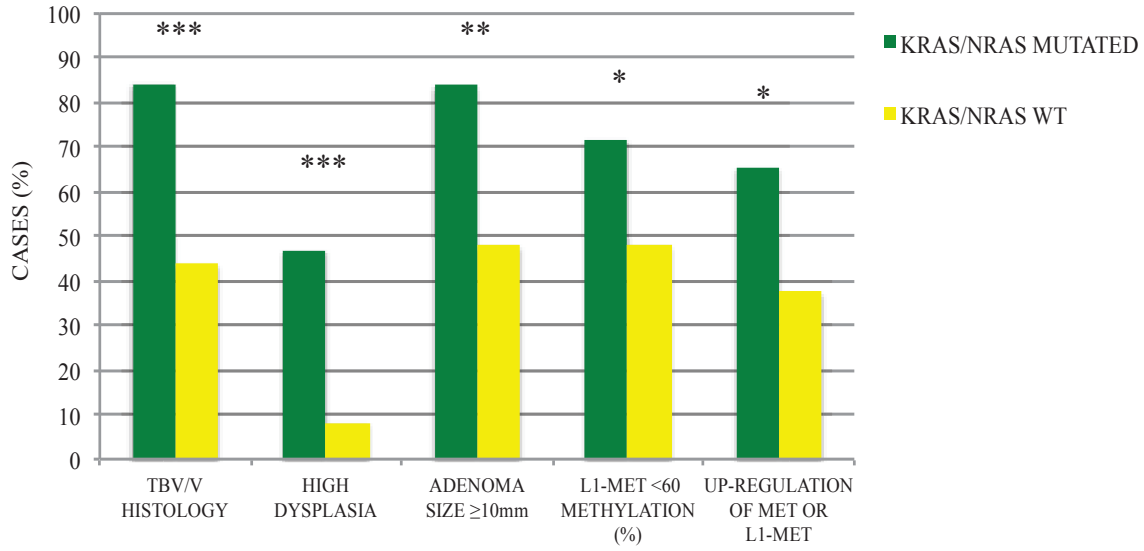
These characteristics include histology (TB/TBV/V), grade of dysplasia (Low/High), adenoma size (<10mm/≥10mm), adenoma site (right/left colon) and number of synchronous adenomas at diagnosis (<3/≥3).

As reported in the Figure 29A, presence of *KRAS/NRAS* mutation was significantly correlated with TBV/V histology (p=0.0004), high grade of dysplasia (p=0.0001), adenoma size ≥10mm (p=0.002), L1-*MET* hypomethylation (p=0.04) and *MET* and/or L1-*MET* up-regulation (p=0.02) (Supplementary Table S3).

The same correlation analysis was performed stratifying sporadic adenomas according to L1-*MET* methylation levels. Again we found that L1-*MET* hypomethylation was significantly correlated with TBV/V histology (p=0.02), adenoma size ≥10mm (p=0.03), presence of *KRAS*/*NRAS* mutations (p=0.04), L1 hypomethylation (p<0.0001) and *MET* or/and L1-*MET* up-regulation (p<0.0001) (Figure 29B and Supplementary Table S6).

Finally, stratifying 56 sporadic adenomas according to *MET* and/or L1-*MET* transcripts expression levels, we observed that only L1-*MET* hypomethylation was significant correlated with *MET* and/or L1-*MET* up-regulation (p=0.001) (Supplementary Figure S4 and Supplementary Table S3,S7).

A



B

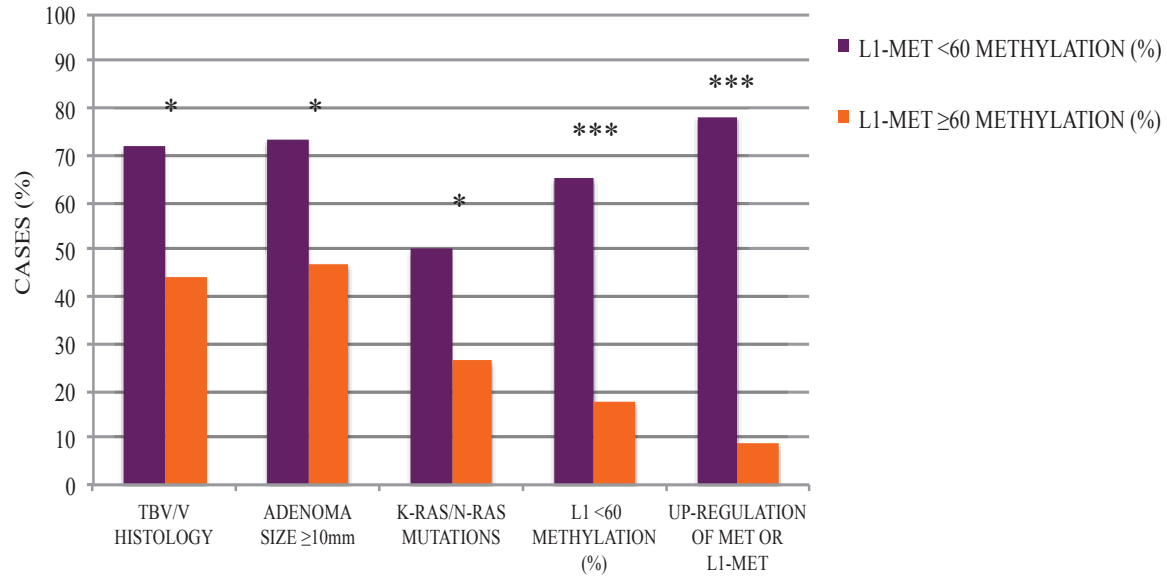


Figure 30. (A) Colums represent the percentage of cases with TBV/V histology, high grade of dysplasia, adenoma size ≥10 mm, L1-*MET* methylation <60% and up-regulation of *MET* or L1-*MET* stratifying sporadic adenomas according to mutational KRAS/NRAS status. (B) Colums represent the percentage of cases with TBV/V histology, high grade of dysplasia, adenoma size ≥10 mm, L1-*MET* methylation <60% and up-regulation of *MET* or L1-*MET* stratifying sporadic adenomas according to L1-*MET* methylation levels.
Legend:TBV, tubulovillous adenoma; V villous adenoma; * $0.1<p\leq 0.05$; ** $0.001<p\leq 0.01$; *** $0.0001<p\leq 0.001$.

SECOND PART

5. CORRELATION BETWEEN rs1800734 ALLELES AND *MLH1* METHYLATION IN NORMAL COLORECTAL SAMPLES

The single nucleotide polymorphism (SNP) rs1800734, that lies between the transcription start site and the ATG codon in the 5' untranslated region of *MLH1,* has been assessed as a candidate for microsatellite unstable CRC susceptibility in a number of MSI CRC data sets while no evidences were found for unselected CRCs. Therefore, it was hypothesized that SNP rs1800734 may play a mechanistic role in the silencing of *MLH1* during MSI cancer development (Allan et al., 2008; Campbell et al., 2009; Raptis et al., 2007; Whiffin et al., 2011).

In order to detect potential differences of CpG island methylation patterns of *MLH1* gene and clarify the primary event leading to loss of MLH1 protein in MSI cancers, we devised a comprehensive analysis of the relationship between rs1800734 genotype and *MLH1* methylation in normal colorectal samples.

First of all, we analyzed by KASP assay the distribution of three genotypes AA/AG/GG of SNP rs1800734 in a cohort of 33 normal colorectal mucosa samples from patients undergoing colonoscopy without malignancies; in detail, the GG, AG and AA genotypes were found respectively in 52% (17/33), 45% (15/33) and 3% (1/33) of cases (Figure 30).

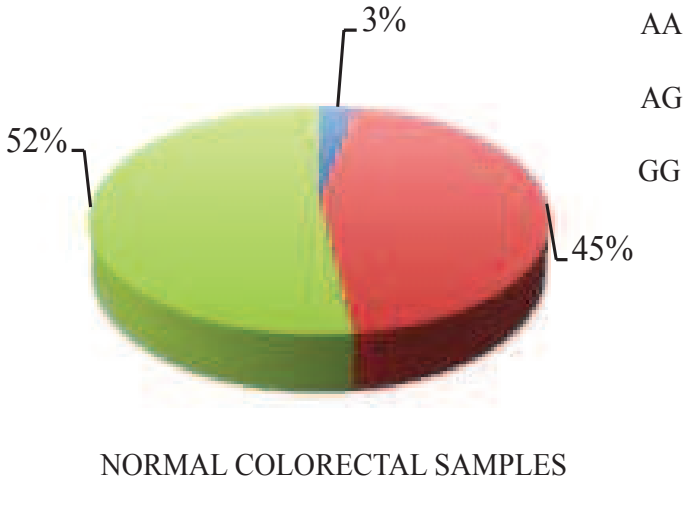


Figure 31. Pie chart shows the frequencies of the three SNP genotypes in our cohort of normal colorectal mucosa samples.

Subsequently, in the same normal samples we determined the methylation profiles across *MLH1* CpG island and CpG shore by Miseq NGS analysis of bisulphite treated DNA amplicons.

As expected, little or no methylation close to rs1800734 in Deng region C frequently involved in *MLH1* silencing, was observed in any of the normal samples (Figure 32).

There were increasing levels of methylation towards the upstream CpG island shore and interestingly these were higher in low-risk GG homozygotes than in the AG heterozygotes (p=0.011; insufficient AA samples were present for statistical analysis).

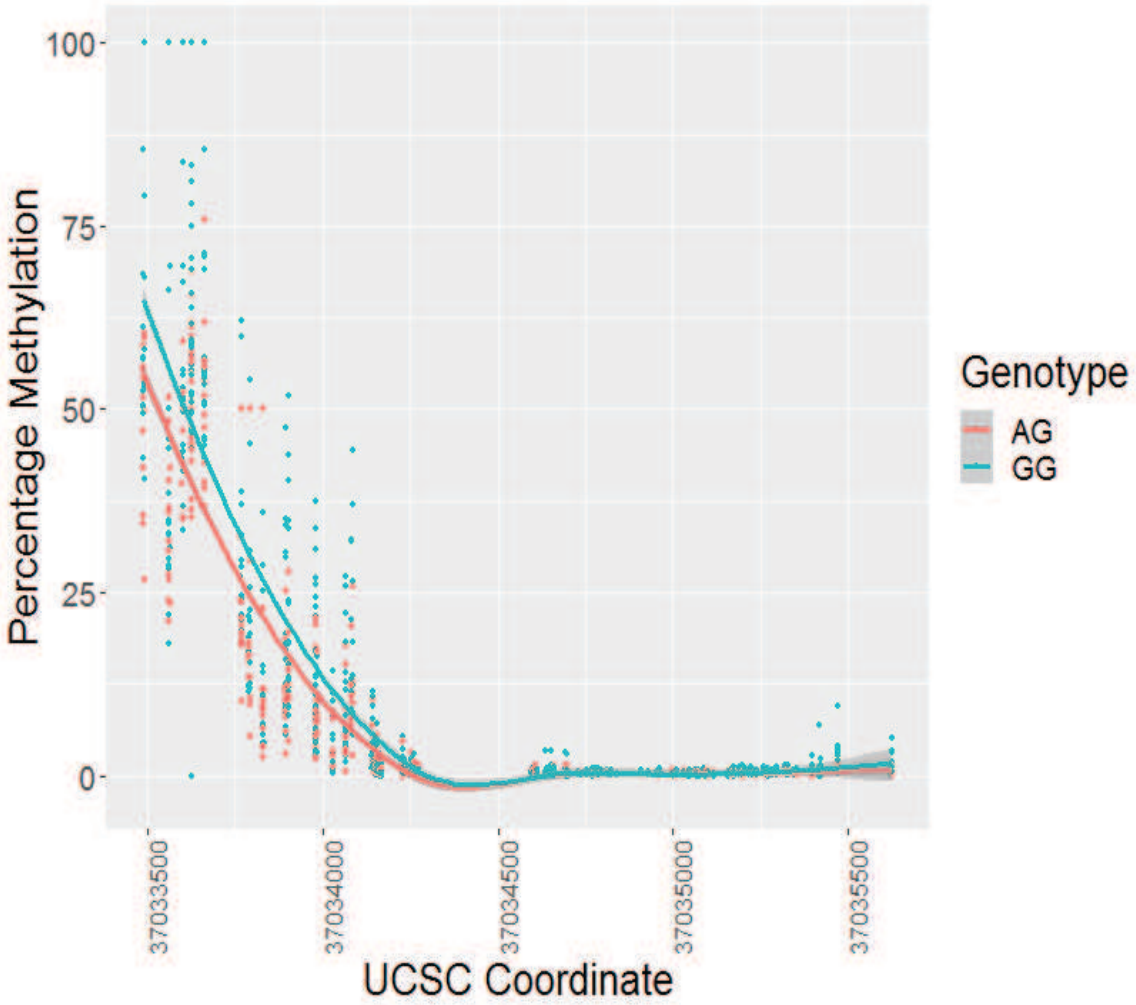


Figure 32. Scatter plot showing total methylation levels across all CpGs of *MLH1* gene with 33 normal colorectal mucosa samples grouped by genotype.

6. CORRELATION BETWEEN rs1800734 ALLELES, *MLH1* METHYLATION AND *MLH1* EXPRESSION IN MSI CRCs

The *MLH1* methylation analysis was also carried out in a cohort of 35 FFPE MSI CRCs in order to correlate the methylation status of promoter with the genotype of rs1800734. Due to fragmentation of DNA extracted from formaline-fixed tissues, this analysis was possible only in the CpG islands of the gene and was performed using MS-MLPA ME-011 (Figure 10).

In this cohort, the distribution of the three genotypes was the following: GG, AG and AA genotypes in 43% (15/35), 34% (12/35) and 23% (8/35) of cases, respectively (Figure 32).

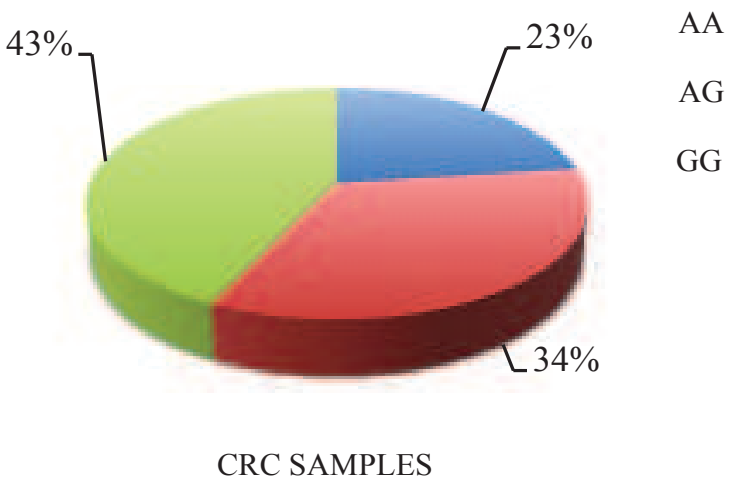


Figure 33. Pie chart shows the frequencies of the three SNP genotypes in our cohort of MSI CRCs.

MLH1 promoter methylation levels were variable but significantly higher in AA and AG patients than GG patients across the CpG island (P=0.0002) (Figure 35).

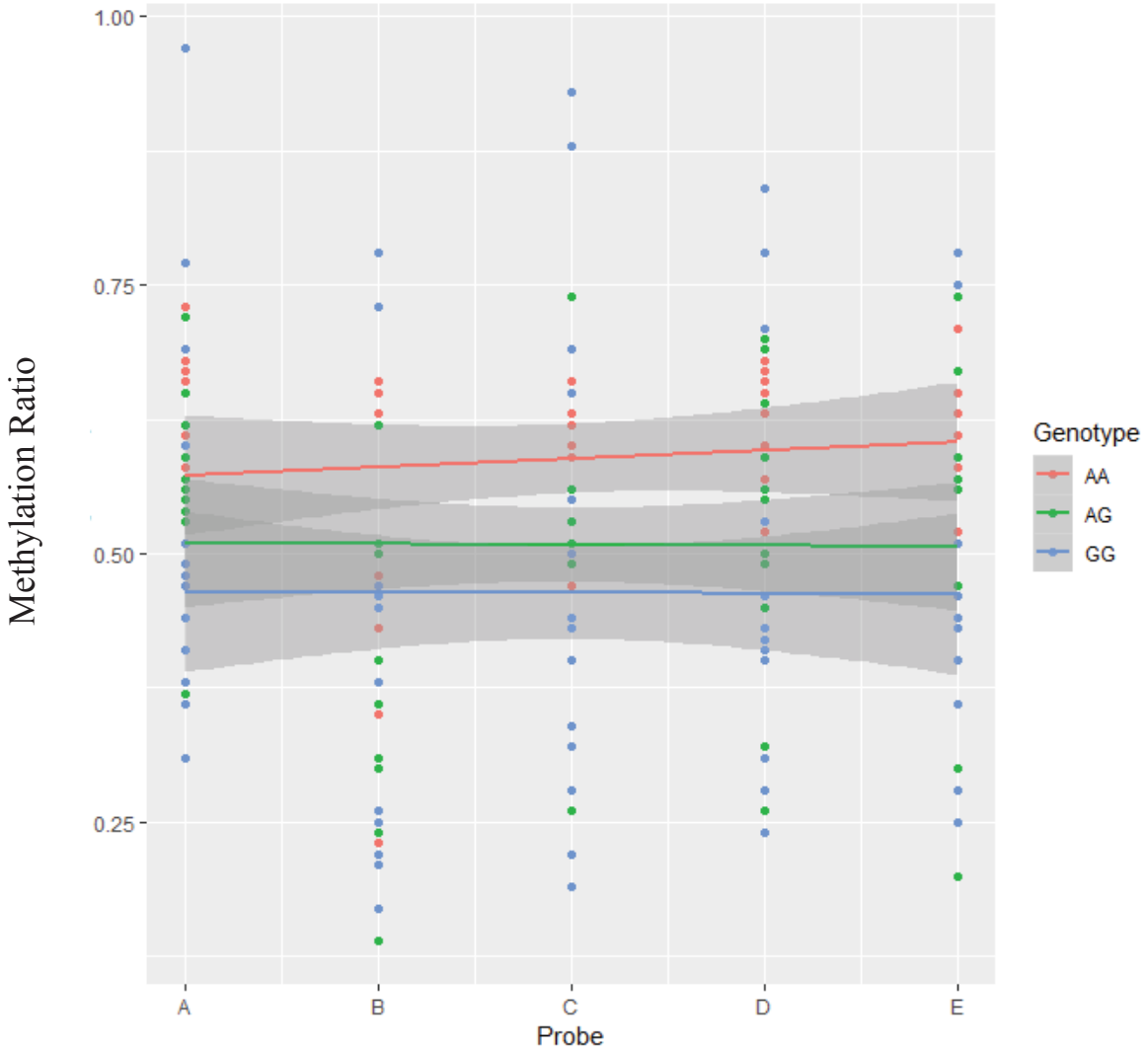


Figure 36. Scatter plot showing methylation ratio of 5 probes used by MS-MLPA with 35 MSI tumours grouped by genotype.

We interrogated these samples further to determine if these differences were specifically due to increase levels on the A allele in heterozygous patients. This analysis was performed by using bisulphite-treated DNA amplified by PCR, followed by long paired end reads and Miseq NGS to allow phasing of rs1800734 allele with methylation in region C.

Despite variable levels of methylation between patients, we found in 15 patients that methylation on the A allele was significantly greater than the G allele (P=0.03) (Figure 36).

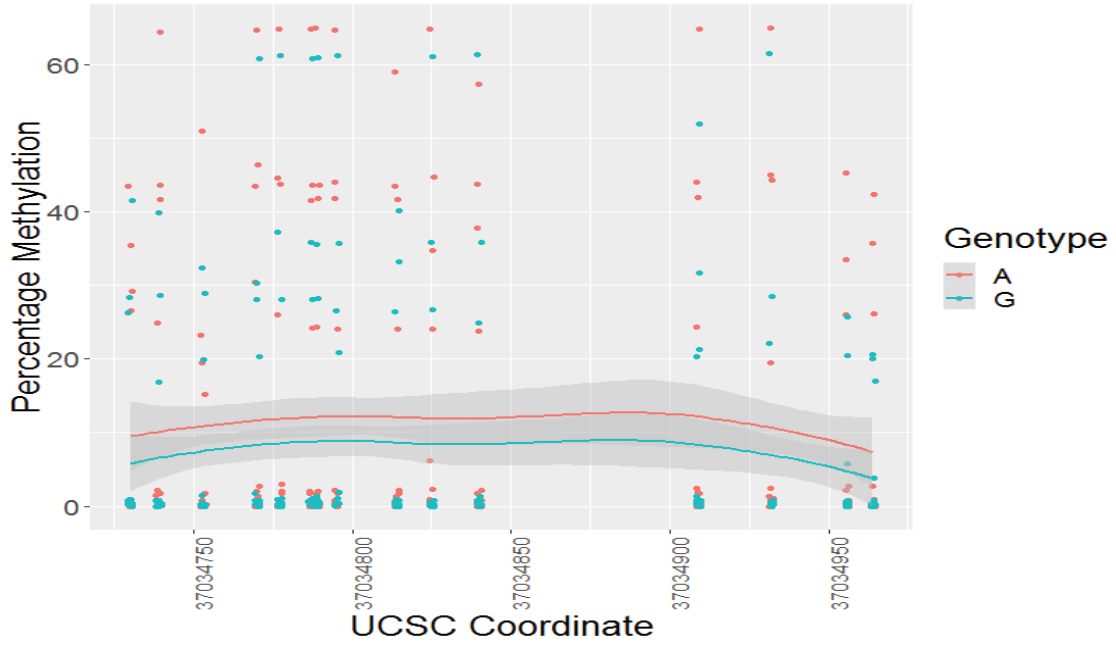


Figure 36. Scatter plot showing allele specific methylation levels across CpGs close to rs1800734 with 15 CRC samples grouped by genotype.

Finally, in order to investigate the potential correlation between *MLH1* methylation and expression, we analyzed MLH1 mRNA expression levels on the same 35 CRCs and we observed that results varied significantly with genotype (p=0.0001) (Figure 37).

Notably, MLH1 mRNA expression were significantly correlated with methylation levels ($P=1.67 \times 10^{-5}$), confirming an interdependent relationship.

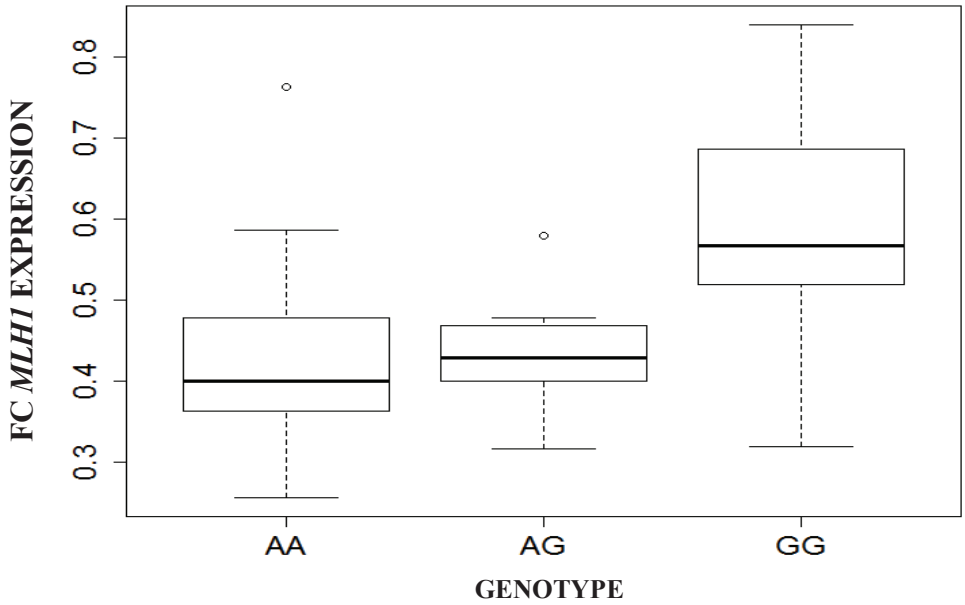


Figure 37. Boxplot showing total MLH1 mRNA levels in all 35 CRC samples grouped by genotype. FC, Fold Change.

We carried out a similar analysis on colorectal cancer with matched normal data from The Cancer Genome Atlas (TCGA COADREAD, http://cancergenome.nih.gov/). We found that methylation in region C and at rs1800734 was significantly higher in tumours with AA and AG genotypes (Supplementary Figure 5A, n= 432, p=0.000133). When stratified by MSI status, the MSI tumours alone still showed a genotype-specific significant difference in methylation levels (Figure 38, n=157, p=0.00115), whilst MSS tumours showed no allele specific differences (Supplementary Figure 5B, n=275, p=0.627).

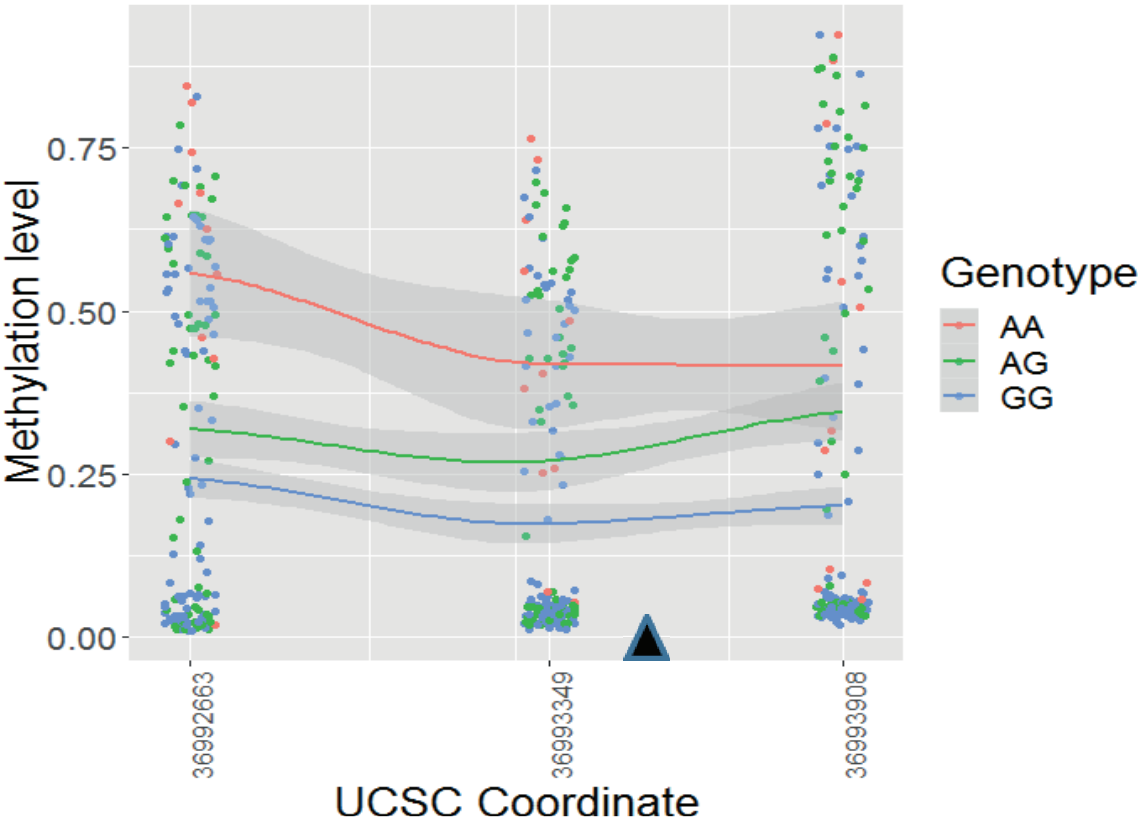


Figure 38. Scatter plot showing allele specific methylation levels across CpGs close to rs1800734 (black arrow) with 157 CRC samples grouped by genotype.

Analysis of mRNA expression levels on the same data sets showed that these also varied significantly with genotype in all tumour samples (Supplementary Figure 5C, n=432, p=0.00153, ANOVA). When stratified by MSI status the MSI tumour samples alone showed a highly significant difference between the genotypes (Figure 39, n=157, p=0.0006) whilst the MSS tumour samples showed no effect of genotype on expression (Supplementary Figure 5D, n=275, p=0.627).

As in our MSI cancer data set, expression and methylation showed a highly significant correlation (p-value < 10-15, Pearsons).

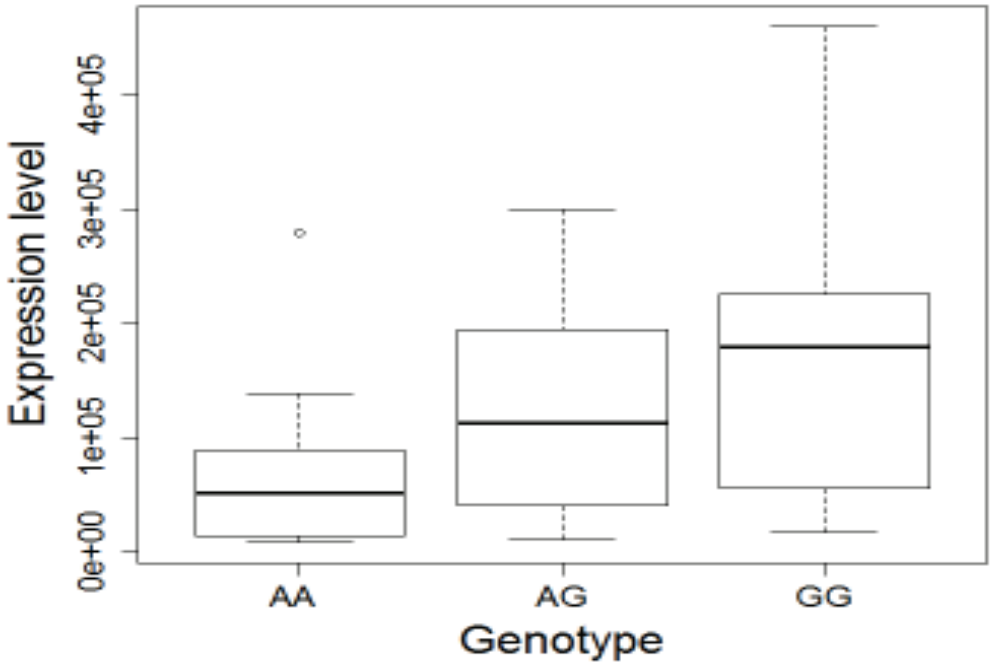


Figure 39. Boxplot showing total MLH1 mRNA levels in all 157 CRC samples grouped by genotype.

7. CORRELATION BETWEEN rs1800734 ALLELES, *MLH1* METHYLATION AND *MLH1* EXPRESSION *IN VITRO*

We wished to take a dynamic approach to determine if DNA methylation is the primary and causal event leading to MLH1 transcriptional silencing by analysing two MSI colorectal cancer cell lines, CO-115 and SW48, both with rs1800734 G/A genotype. Thus, we subjected the two cell lines to a treatment inducing global demethylation and measured allele-specific methylation and gene expression at baseline and specific time-points (4, 7 and 11 days after treatment).

5-azacytidine and 5-aza-2'-deoxycytidine (AzaC) are chemical analogues of cytosine and inhibitors of DNA methylation and they cause a global loss of methylation when used to treat cell lines (177).

At baseline, MSI CRC cell line CO-115 has very high levels of methylation at all CpGs analysed in the *MLH1* promoter by MiSeq NGS (Mean=91%) and expression of *MLH1* is undetectable by qRT-PCR (Figure 40 and Figure 42).

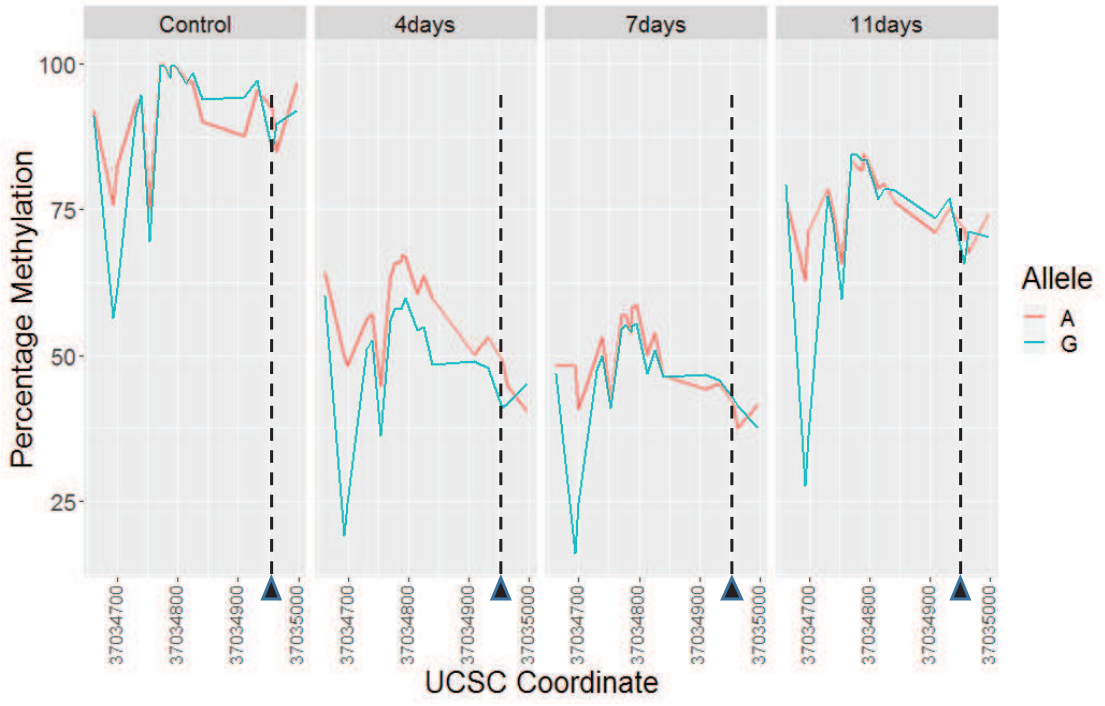


Figure 40. Line graphs showing methylation percentage levels at individual CpGs in the rs1800734 region grouped by genotype. Each panel shows a control or time-point post AzaC treatment. The position of the SNP is marked (black triangle).

Similar results were obtained by MS-MLPA analysis that confirmed biallelic methylation in untreated cells: for all the five probes (Figure 11) we observed MR (Methylation Ratio) ranging from 0.72 to 1.08 with a mean value of 0.96.

After 4 days from the treatment, we observed a 50% reduction in DNA methylation (MR mean value=0.51) and progressive increase after 7 days (MR mean value=0.60) and 11 days (MR mean value=0.78) (Figure 41).

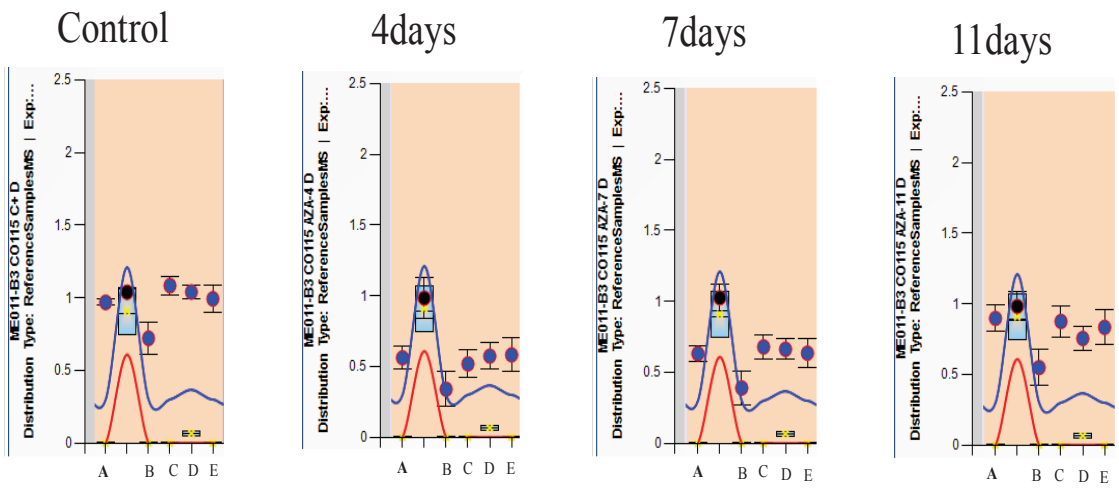


Figure 41. MS-MLPA analysis of the *MLH1* gene in control cells and timepoints post AzaC treatment. A, B, C, D and E (intron) are the 5 probes used to analyze the *MLH1* methylation.

This loss of *MLH1* methylation is accompanied by an equally dramatic re-expression of MLH1 mRNA peaking at 4 days post-AzaC treatment (P=0.007) and then starting to reduce by day 11 post-treatment (Figure 42). Again, there are allele specific differences at each stage with the risk (A) allele expressed at lower levels that the protective (G). Re-expression of *MLH1* to similar levels was also seen in a second MSI CRC cell line, SW48, following 5AzaC-induced loss of methylation (Supplementary Figure 6).

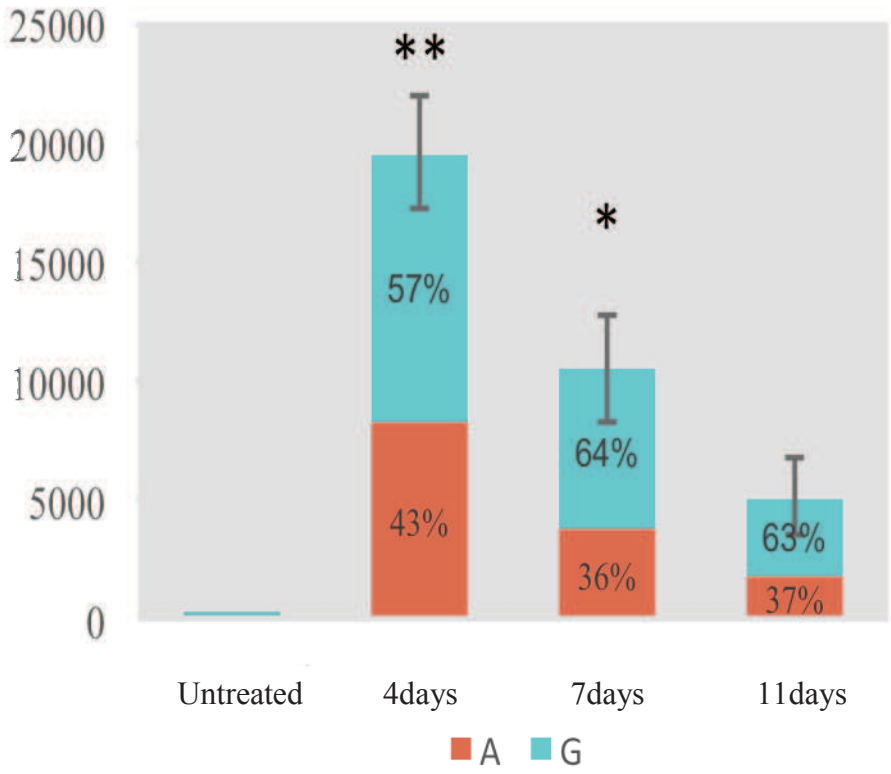


Figure 42. Bar graph shows total MLH1 mRNA expression and the allelic components of this expression in control cells and timepoints post AzaC treatment. Error bars show the standard error of the mean of replicates. Asteriks denote significant (1, $p<0.05$) or highly significant (2, $p<0.01$) increase in expression due to AzaC treatment. FC, Fold Change.

DISCUSSION

CRC is a heterogeneous disease developing through different pathogenetic mechanisms. Although mortality has declined progressively in many Western countries mainly due to cancer screening programs and early detection of cancerous lesions, the discrimination of high risk adenomas is still challenging, both in hereditary syndromes and in sporadic cases. Actually, there is a strong need to establish clinically useful biomarkers for risk assessment and early detection of CRC.

In the first part of this thesis, we focused on identification of new valuable molecular risk factors for the onset of CRC, studying both genetic and epigenetic characteristics, such as *KRAS/NRAS/BRAF/PIK3CA* mutations and DNA global hypomethylation in the early events of colorectal carcinogenesis. Our working hypothesis was that MAP, a rare hereditary colorectal polyposis associated to a rapid cancer progression due to inactive BER system and consequent accumulation of oxidative DNA damage (121), could be the appropriate model to validate new potential risk factors for colorectal cancerogenesis and also to study the interplay between genetic and epigenetic markers, when oxidative stress is the main etiological factor linked to cancer transformation.

To verify this hypothesis, we initially selected 18 MAP patients carrying biallelic *MUTYH* germline mutations and a control set of 17 FAP/AFAP patients with pathogenic *APC* germline alterations. Interestingly, considering only the clinical data of these patients at the moment of FAP or MAP diagnosis, 50% of MAP patients showed at least one CRC with or without syncronous adenomas whereas no FAP patients exhibited a co-occurence of CRC during adenoma removal. This observation in our cohort is in line with previous data suggesting a more accelerated carcinogenesis in MAP than in FAP patients (125-126) and that a pathogenetic mechanism linked to oxidative DNA damage may be generally associated to a higher risk of CRC.

As reported by two recent studies (178-179), we detected a two-fold higher level of somatic mutations in MAP adenomas respect to FAP/AFAP suggesting that a deficiency of *MUTYH* gene could lead to a mutator phenotype. As expected, in our cohort the *KRAS* p.G12C mutation was by far the most frequent alteration in MAP adenomas, and never detectable in comparable lesions from FAP/AFAP subset. *KRAS* p.G12C was frequently identified in both adenomas and carcinomas derived from the same MAP patients, confirming that the adenoma cell clones carrying this specific mutation are positively selected for the adenoma to carcinoma transition as well as that

KRAS p.G12C may be used as the somatic hallmark of oxidative DNA damage in MAP disease (127, 180-181).

Our findings reflect the well-known repair specificity of the MUTYH enzyme and the increased susceptibility to oxidative damage at the first base of the GGT sequence in *KRAS* codon 12. The peculiarity of the p.G12C mutation in this genomic region is further supported by the same transversion in the homologous BLAST-aligned nucleotide sequence of *NRAS* gene being detected in four of our MAP adenomas. So far, no other reports have documented the presence of *NRAS* mutations in MAP lesions. Although these alterations can be found in 10% of sporadic CRCs, they were shown to regulate homeostasis of colonic cells differently to *KRAS* alterations and were only occasionally detected in sporadic colorectal adenomas (182-183).

In addition, we detected the co-occurrence of both *KRAS* p.G12C and *PI3KCA* mutations in 80% of the examined MAP carcinomas, suggesting that key mechanisms for cells to regulate cell survival, proliferation, and motility, such as Ras-ERK and PI3K-mTORC1 signaling, are often simultaneously inactivated in these type of CRCs. This result may have an important clinical impact since co-inhibition of both pathways has been successful in reducing tumor growth in xenograft cancer models and, importantly, also in genetically engineered mouse models (184-185). Moreover, the recent development of *KRAS* p.G12C specific inhibitors potentially opens new promising prospects for the treatment of patients affected by CRC with this somatic alteration (186-187).

Regarding the histological phenotype, *KRAS/NRAS* mutations were more frequent in MAP adenomas with tubular architecture than FAP/AFAP lesions with the same morphology. Also this feature can be distinctive of MAP carcinogenesis since *KRAS* mutations were recently proposed as a risk factor for adenomas because they were mainly observed in tubulovillous and villous polyps rather than in tubular lesions (188-189). These data suggest that mutator phenotype in MAP could lead gene mutations also in preneoplastic lesions without aggressive morphological features.

In order to verify if somatic *KRAS/NRAS/BRAF/PIK3CA* mutations in polyps may generally represent a potential molecular marker for the risk of developing CRC (190-191), we chose to include in this study two groups of sporadic colorectal adenomas with different clinico-pathological aggressiveness, namely a first subset of 56 sporadic adenomas (S-Ads) diagnosed in 45 patients who never developed a CRC during a follow-up of at least ten-years and a second subset of 24 sporadic adenomas

(S-AdsC) from 17 individuals who developed a CRC at least one year after polyp removal.

This analysis revealed that gene mutation frequencies were slightly higher in S-AdsC compared with S-Ads (46%, *versus* 38%, respectively) although this difference was not statistically significant. Globally, we found a higher frequency of *KRAS/NRAS* mutations in sporadic adenomas (40%) compared to previous works reporting a mutation frequency ranging from 10% to 20% (189,192). A possible explanation for this different result is that these studies analyzed consecutive series of unselected sporadic adenomas. By contrast, our series includes a large number of sporadic adenomas characterized by high-risk features, such as TBV/V hystotype, high grade of dysplasia, large size (≥10mm), multiple adenomas (ranging from three to six syncronous or metachronous adenomas), right colon site and development of a metachronous CRC (S-AdsC). In fact, *KRAS/NRAS* mutation frequency in our series is in line with previous works examining cohorts entirely composed of TBV/V adenomas or of *in situ* carcinomas (188). Moreover, not surprisingly, these frequencies were not very different from those generally reported in invasive CRC (193).

An additional observation from our work was that 64% of patients showing multiple adenomas exhibited at least one mutated polyp, confirming that a condition of few synchronous/metachronous benign lesion may be *per se* relevant to detect high risk patients. Finally, contrary to MAP, the spectrum of *KRAS/NRAS* mutations in sporadic adenomas and carcinomas was extremely heterogeneous suggesting that different etiological factors may be involved in the acquisition of individual driver mutations (194).

The second aspect that we considered was the role of genome-wide DNA hypomethylation as an early epigenetic alteration since it was observed both in hereditary and sporadic CRCs (169, 195-196). Likewise somatic gene mutations, L1 hypomethylation has been reported to occur early in colorectal carcinogenesis, including aberrant crypt foci and preneoplastic lesions as adenomas (72, 77, 197-199). However, although this marker has been recognized as a poor prognostic factor for increased cancer-related mortality and overall mortality in CRC patients (70-71), currently no studies have evaluated whether L1 hypomethylation in polyps may generally represent a potential molecular marker for the risk of developing CRC.

According to the working hypothesis presented above, we initially chose MAP adenomas as a model in which L1 hypomethylation rates were expected to be higher in comparison with FAP/AFAP lesions.

Although no studies have so far evaluated this marker in MAP polyps, the rationale of this analysis came from several data published about the mechanisms of how oxidative stress decreases DNA methylation (89). Briefly, oxidized DNA lesions that are induced by ROS, such as 8-OHdG in CpG dinucleotides have been shown to strongly inhibit methylation of adjacent cytosine residues (91). Secondly, an unfixed 8-OHdG may introduce a G>T transversion resulting in the loss of CpG dinucleotides (95). Moreover, in an oxidative stress condition, with decreased availability of S-adenosylmethionine (SAM), depletion of the methyl pool in folate-deficient models has been shown to cause DNA hypomethylation (94). Recently, Kloypan et al. demonstrated that in bladder cancer and human kidney cell lines, L1 hypomethylation is associated with the oxidative stress-mediated activation of the ten-eleven translocation (TET) hydroxylase enzymes which cooperate with BER in controlling the formation and replacement of 5-hydroxymethylcytosine (5hmC) (200). Finally, ROS are considered to be responsible for aberrant DNA methylation in different cancer models (200-204) and during chronic inflammation and inflammation-associated carcinogenesis, such as ulcerative colitis and Crohn's disease (205-206).

In agreement with our hypothesis, MAP adenomas were found to be more often hypomethylated than FAP/AFAP lesions, for both global genomic and intragenic sequences, as measured by L1 and L1-*MET* assays. Therefore, for the first time, we proposed a novel link between a BER defect and DNA hypomethylation in precursor lesions of CRC suggesting that unfixed 8-OHdG may strongly correlate not only with genetic lesions but also with aberrant DNA hypomethylation.

In order to assess if L1 hypomethylation may generally represent a potential risk factor in adenomas, we used L1 and L1-*MET* assays to analyse this marker in the 80 sporadic adenomas both globally and stratyfing S-Ads and S-AdsC. This analysis revealed that about 50% of sporadic adenomas were L1-hypomethylated showing similar DNA hypomethylation frequencies of MAP adenomas. Considering the two subsets of sporadic adenomas, we observed that S-AdsC were more frequently hypomethylated than S-Ads, although this difference was not statistically significant. Moreover, in line with what observed for *KRAS/NRAS* mutations, the level of L1 and L1-*MET* methylation in MAP CRCs and sporadic CRCs decreased in comparison with the

matched adenomas of the same patients, suggesting that DNA demethylation is acquired early and retained in the adenoma-carcinoma transition.

Finally, an interesting observation regards the different distributions of L1 and L1-*MET* methylation percentages in the multiple MAP or sporadic adenomas of the same patient compared with FAP/AFAP polyps. In this last group methylation levels were significantly less heterogeneous with respect to the first two subsets of patients. Globally, the stronger genetic and epigenetic heterogeneity among multiple MAP or sporadic adenomas suggests the presence of a field defect, the process of carcinogenesis initiating from multiple alterations in the colorectal mucosa which can lead to the clonal expansion of premalignant daughther cells in a particular field (207). In the colorectal mucosa, the identification of a field defect may indicate a higher risk of metachronous neoplastic lesions and coul help to identify which patients require a more radical surgery. Moreover, the most exciting use of field cancerisation theory is its potential application in chemoprevention (208).

Taken together all these findings may suggest that DNA hypomethylation may play an early role in colorectal carcinogenesis characterizing a more aggressive clinical behavior of precursor lesions and cancers. Further fuctional analysis in vitro could be useful to understand the molecular mechanism linking inactive BER and DNA hypomethylation and further research is needed to better define which specific etiologic factors may be involved in sporadic tumorigenic mechanisms associated with DNA hypomethylation.

Currently, few studies focused on the consequences of L1 hypomethylation in driving malignant progression (77, 82, 85, 174, 176, 209-211). L1 hypomethylation has been suggested to be a key event in cancer development because it results in retrotransposition throughout the genome leading to CIN (212-213).

Moreover, in the context of CRC, Hur et al. in 2013 provided the first evidence for increased L1 hypomethylation in CRC metastasis, and systematically uncovered its relationship with locus-specific L1 hypomethylation within the proto-oncogene *MET*. More specifically, they demonstrated that global L1 hypomethylation correlates with hypomethylation of a specific L1 sequence which permit transcriptional activation of the aberrant chimeric transcript L1-*MET* with negative prognostic value and finally that higher expression of MET protein specifically correlates with higher levels of L1-*MET* transcription during development metastasis in CRC (84). Based on this study, we wondered if this aberrant mechanism that was well-known in invasive and metastatic

CRCs, may also be observed in preneoplastic lesions, characterizing the earlier phases of colorectal carcinogenesis.

Before moving to adenomas, firstly we confirmed by RNA-Seq the strong association between L1 hypomethylation and *MET/L1-MET* expression in CRCs according to L1 methylation levels. In line with Hur et al. (84), our data revealed that L1 methylation percentages of L1/L2/L3/L4 CRC groups were inversely correlated with MET/L1-MET transcript levels and that the more L1-*MET* sequence was hypomethylated, the more chimeric L1-*MET*/*MET* were overexpressed.

Secondly, we designed a new assay to specifically distinguish L1-*MET* chimeric transcript from normal *MET* transcript and we validated RNA-seq results in L1 and L4 CRC groups by using qRT-PCR analysis on the same tumor samples, applying this technique on small amounts of FFPE samples. For the first time, we demonstrated that the aberrant chimeric L1-*MET* transcript has already been expressed in a subset of colorectal adenomas through the same epigenetic mechanism demonstrated in CRC.

Indeed our data confimed that more that 80% of L1-*MET* hypomethylated colorectal adenomas showed an abnormal upregulation of L1-*MET* and/or *MET* transcript suggesting that already in the early phases of colorectal carcinogenesis the molecular mechanism regulating *MET* expression may be altered through L1-hypomethyaltion.

In conclusion, our data indicate that, in addition to well-assessed specific mutations, the early steps of oxidative DNA damage-induced colorectal carcinogenesis are characterised by decreased DNA global methylation and specific L1-*MET* hypomethylation. Moreover, our results emphasise the idea that genetic and epigenetic mechanisms strengthen each other in driving colorectal tumourigenesis. Because of the clinical recommendations for recently established aspirin-based chemoprevention strategies (214), these results appear to be interesting in improving the identification of individuals who are most likely to benefit from a prophylactic aspirin regimen.

Finally, we conclude that both genetic and epigenetic markers such as *KRAS/NRAS* mutations, L1-hypomethylation and transcriptional activation of aberrant chimeric transcripts such *L1-MET* deserve to be evaluated in a future diagnostic algorithm in which both clinico-pathological and molecular factors may be integrated for a classification of patients with colorectal adenomas into different categories of risk.

A preliminary correlation analysis between these three molecular markers and the main clinico-pathological factors currently used for high risk colorectal adenomas (10) reveals that both *KRAS/NRAS* mutations and/or L1 hypomethylation are significantly

associated to clinical and morphological features suggestive of aggressive behavior in precursor lesions of CRC.

Future perspectives of this study will include the validation of these results in an independent and larger series of sporadic adenomas in order to assess a robust multivariable analysis that may evaluate the independent contribution of each factor.

- -

In the second part of this thesis, we focused on the correlation between *MLH1* SNP rs1800734 and *MLH1* mehylation-driven silencing in MSI CRCs.

The accumulation of a number of low-risk and low-penetrance variants contributes to significant cancer risk. However, despite the fact that the majority of SNPs are located in non-coding regions of the genome, their diverse role in disease pathogenesis, including CRC, are steadely becoming established through both experimental and computational methods (149, 150, 215).

Of particular interest to my research project was the *MLH1* promoter SNP rs1800734 (-93G>A), located in the CpG island of *MLH1* promoter. The variant A allele of this SNP shows a strong association with *MLH1* CpG island hypermethylation, loss of *MLH1* expression, MMR deficiency, and MSI sporadic CRC (157-161). Further, the allelic variant of rs1800734 has a functional consequence in that it decreases the transcriptional activity of *MLH1* in CRC cell lines (163). This SNP has also demonstrated to be associated with increased risk of glioblastoma, gastric cancer, lung cancer, and ovarian cancer (216-219).

These epidemiological data collected in the last years about SNP rs1800734 suggested us to investigate the mechanistic basis of promoter hypermethylation and transcriptional silencing of *MLH1*. To date, the relationship between the initiating genetic events responsible for tumorigenesis (e.g.,acquisition of activating mutations in oncogenes or presence of specific SNP) and the subsequent promoter hypermethylation and transcriptional silencing remain to be determined. To this aim, we decided to examine in-depth the mechanism of *MLH1* silencing in order to hypothesize a general mechanism of gene inactivation due to epigenetic alterations.

We have confirmed that the SNP rs1800734 in the promoter of *MLH1* is associated with the risk of sporadic MSI CRC. This strong influence on the mismatch repair pathway makes it likely that the causal functional effect of the rs1800734 risk allele is to reduce

expression of the mismatch repair pathway protein, MLH1.

Our results demonstrate that a significant allele specific effects are seen on both methylation and mRNA expression in MSI cancers in our own and publically available data sets. As expected the risk (A) allele leads to significantly higher methylation levels and this strongly correlates with lower mRNA expression.

Conversely, the risk (A) rs1800734 allele has no measurable repressive effect on *MLH1* in normal colon tissues, even using highly sensitive techniques, and in fact associates with reduced DNA methylation at the upstream CpG island shore, in line with previous observations (165).

We have also clearly demonstrated that DNA methylation is necessary for *MLH1* transcriptional silencing by treating with AzaC, removing methylation and demonstrating re-expression of *MLH1* in MSI cells. This result implies that methylation is indeed the primary cause of *MLH1* silencing in sporadic MSI cancers. Interestingly, even in this engineered situation the risk (A) allele is more prone to acquire methylation and its mRNA is re-expressed at lower levels. The reason could be that the A allele is more readily methylated could be due to the lack of binding of the TFAP4 transcription factor as shown by Lui et al. (164).

Premilary data not shown confirmed that there is indeed a strong bias in TFAP4 binding in unmethylated CRC cells and that this binding is inversely correlated with DNA methylation post AzaC treatment in MSI cells.

However, it is unlikely that TFAP4 may be the only transcription factor binding across rs1800734 and publically available genome wide ChIP-seq experiments show multiple proteins binding in the region (ENCODE, UCSC).

Indeed, TFAP4 has been shown to belong to a class of enhancer binding factors that are important for co-factor recruitment and activation (220). This suggests that it could be a major factor in determining protein binding and potentially chromatin landscape across the region. The fact that TFAP4 also binds allele specifically to another disease associated SNP (rs12722522, Type 1 diabetes (221) suggests it might play a more generalised role in a subset of SNP trait associations, acting to recruit activating factors in an allele specific manner.

Liu et al. suggest that allele specific TFAP4 binding at rs1800734 may exert an effect on the cancer pathway via long range interactions with the promoter of the DCLK3 gene, causing enhanced expression of genes related to epithelial-to-mesenchymal transition (164). However we were unable to corroborate this finding (data not shown).

We failed to detect significant DCLK3 expression in either our MSS or MSI cell lines using sensitive Q-PCR based techniques.

We have clearly demonstrated that rs1800734 is only associated with an increased risk of MSI cancers and only modifies methylation in these cancers. Thus, it seems unlikely that a gene with no known role in MMR plays the primary or causative role in rs1800734 associated cancer risk.

Our data, taken together with other studies described above, support the prevailing hypothesis that *MLH1* repression is the main mechanism by which rs1800734 confers cancer risk. Since the majority of our MSI cancers (23/35) carry a *BRAF* V600E mutation we suggest it is likely that rs1800734 influences the acquisition of methylation via the BRAF/MAFG pathway described by Fang et al. (103).

Figure 43 represents our proposed model in which TFAP4, with co-factors binds within the promoter region on the G protective allele. This restricts the access of the BRAF activated MAFG complex and consequently reduces the spread of DNA methylation spreading from the CpG island shore. On the A allele TFAP4 binding occurs less frequently or with lower affinity allowing MAFG access and methylation spreading.

While the functional role of rs1800734 in the pathways of CRC development is becoming clearer, it is also interesting to note how readily the accumulation of *MLH1* promoter methylation could be reversed resulting in re-expression of *MLH1*. The importance of *MLH1* promoter methylation in other cancer types is less well understood however it is frequently found in endometrial, gastric cancers as well as lung, bladder, and some haematological malignancies (222-226). Drugs such as Azacitidine that inhibit DNA methylation are already approved for the treatment of some cancers. However with the advent of CRISPR technology, more precise demethylation is now possible (227) and could be harnessed in the design of future therapies.

Figure 43 reports a schematic presentation of the relevant points of this work.

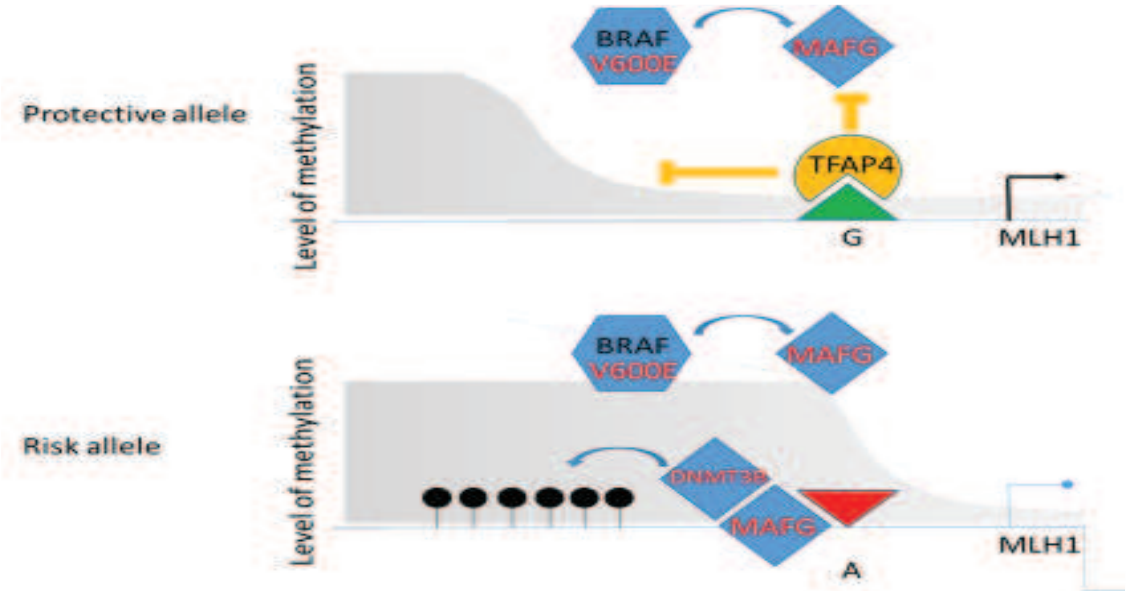


Fig 43. (Upper Panel) The protective allele (G, green triangle, upper panel) binds TFAP4 (yellow) which protects the promoter (black arrow) from BRAF and MAFG (blue) directed methylation and/or methylation spreading from the CpG island shore. The grey shaded area represents methylation levels across the region. (Lower Panel). The risk allele (A, red inverted triangle) does not bind TFAP4 allowing MAFG and cofactors to mediate DNMT3B methylation in the promoter region causing transcriptional repression.

HIGHLIGHTS

- The early steps of oxidative DNA damage in MAP carcinogenesis are characterized by a specific pattern of somatic mutations;
- MAP adenomas and carcinomas show a decreased DNA global methylation and specific L1-*MET* hypomethylation;
- DNA hypomethylation and expression of L1-*MET* chimeric transcript may play an early role in colorectal carcinogenesis characterizing a subset of more aggressive precursor lesions and cancers;
- Both genetic and epigenetic markers such as *KRAS/NRAS* mutations, L1-hypomethylation and transcriptional activation of *L1-MET* deserve to be evaluated in a future diagnostic algorithm in which both clinico-pathological and molecular factors may be integrated for a classification of patients with colorectal adenomas into different categories of risk.

- - - - - - - - - - - - - -

- In normal colon tissue, allele-specific differences about rs1800734 exist in methylation of *MLH1* CpG shore;
- Allele-specific differences in both *MLH1* promoter methylation and expression are present in MSI cancers;
- The *MLH1* transcriptional repression is dependent on DNA methylation and can be reversed by a methylation inhibitor;
- We proposed a new model to explain how the rs1800734 may regulate the mechanism of *MLH1* promoter methylation induced by *BRAF* mutation.

Figure 44. Schematic presentation of the relevant points.

SUPPLEMENTARY TABLES AND FIGURES

Supplemtary Table S1. Histopathological and molecular features of 52 colorectal adenomas from 15 MAP patients.

MAP patients	*MUTYH* germline mutations		Age at diagnosis	Adenoma	Histological features	Site	Dysplasia	% Methylation		Mutation	
	DNA change	Protein						L1	LI-*MET*	*KRAS*	*NRAS*
M1	c.1187G>A/ c.1187insGG	p.G396D/ p.Gly396GlyfsX12	51	A1	TB	Cecum	low	61.43	56.23	WT	WT
				A2	TB	NA	NA	63.20	62.50	WT	WT
				A3	TB	NA	NA	54.08	58.97	WT	WT
				A4	TBV	Right colon	low	56.25	N.A.	p.G12C	WT
M2	c.1187G>A/ c.1187insGG	p.G396D/ p.Gly396GlyfsX12	48	A1	TB	Sigma	low	75.33	67.33	WT	p.G12S
				A2	TBV	Sigma	low	66.60	69.57	WT	WT
				A3	TBV	Sigma	low	62.3	59.6	WT	WT
M3	c.1103delCC/ c.1187G>A	p.Pro367GlnfsX160 / p.G396D	79	A1	TB	Cecum	low	68.05	67.07	p.G12C	WT
				A2	TBV	Rectum	low	45.45	64.3	WT	p.G12C
				A3	TBV	Rectum	low	63.18	58.7	WT	WT
M4	c.536A>G/ c.536A>G	p.Y179C/ p.Y179C	47	A1	TB	Colon NOS	low	59.00	62.53	p.G12C	WT
				A2	TB	Colon NOS	low	61.05	59.80	p.G12C	WT
				A3	TBV	Colon NOS	low	61.45	60.87	p.G12C	WT
				A4	TB	Colon NOS	low	62.73	59.37	WT	p.G12V
M5	c.536A>G/ c.536A>G	p.Y179C/ p.Y179C	44	A1	TBV	Sigma	high	46.93	54.37	WT	WT
				A2	TB	Right colon	low	56.85	63.13	WT	WT
				A3	TBV	Right colon	high	55.98	57.2	WT	WT
				A4	TB	Right colon	low	62.15	53.4	WT	p.G12C
				A5	TB	Right colon	high	61.58	63.9	WT	WT
				A6	TB	Right colon	low	64.78	58.36	WT	WT
M6	c.536A>G/ c.1105G>T	p.Y179C/ p.E369X	42	A1	TBV	Colon NOS	low	51.63	59.17	p.G12C	WT
M7	c.312C>A/ c.325C>T	p.Y104X/ p.R109W	50	A1	TB	Right colon	low	54.97	59.36	WT	WT
M8	c.536A>G/ c.536A>G	p.Y179C/ p.Y179C	47	A1	TB	Sigma	low	59.88	60.86	p.G13D	WT

M9	c.1187insGG / c.1187insGG	p.Gly396GlyfsX12/ p.Gly396GlyfsX12	40	A1	TBV	Rectum	low	53	59.9	p.G12C	WT
				A2	TB	Rectum	low	62.43	64.8	WT	WT
				A3	TBV	Rectum	low	69.28	64	WT	WT
				A4	TBV	Rectum	low	54.6	55.73	p.G12C	WT
				A5	TBV	Rectum	low	56.4	55.88	p.G12C	WT
				A6	TBV	Rectum	low	61.95	62.53	p.G12C	WT
				A7	TB	Rectum	low	61.2	62.63	p.G12C	WT
M10	692G>A/ c.1395_1397 delGGA	p.R231H/ p.Glu466del.	43	A1	TB	Right colon	low	62.28	62	WT	WT
				A2	TB	Left colon	low	68.85	62.1	WT	WT
				A3	TB	Right colon	low	59.55	61.5	p.G12C	WT
				A4	TB	Right colon	low	73.2	68.36	WT	WT
				A5	TB	Rectum	low	64.95	65.73	WT	WT
				A6	TBV	Right colon	low	65.15	61.63	p.G12A	WT
M12	c.733C>T/ c.1187G>A	p.R245C/ p.G396D	50	A1	TB	Right colon	low	71.2	63.96	p.G12C	WT
				A2	TBV	Rectum	low	70.61	63.4	p.K117N	WT
				A3	TB	Right colon	low	65.1	63.5	p.G12C	WT
M14	c.494A>G/ c.1145G>A	p.E77X/ p.E393X	65	A1	TB	Colon NOS	low	52.05	52.93	WT	WT
				A2	TB	Colon NOS	low	60.07	53.4	WT	WT
M15	c.1187G>A/ c.1187G>A	p.G396D/ p.G396D	39	A1	TBV	NA	low	60.98	N.A.	WT	WT
				A2	TB	NA	low	51.5	54.2	WT	WT
M16	c.536A>G/ c.933 +3A>C	p.Y179C/ p.Gly264TrpfsX7	54	A1	TB	Hepatic flexure	low	60.23	56.9	WT	WT
				A2	TBV	Cecum	low	55.78	54.1	WT	WT
				A3	TB	Colon NOS	low	64.06	56.6	WT	WT
				A4	TB	Cecum	low	60.9	59	p.G12C	WT
				A5	TB	Colon NOS	low	58.55	57.7	WT	WT
				A6	TBV	Colon NOS	low	53.4	55.6	WT	WT
M18	c.1012C>T/ 1187G>A	p.Gln338*/G396D	50	A1	TBV	Right colon	high	58.88	59.7	WT	WT
				A2	TB	Right colon	low	65.1	60.33	WT	WT
				A3	TB	Right colon	low	61.78	62.07	WT	WT

Legend: MAP, *MUTYH*-associated-polyposis; TB, tubular adenoma; TBV, tubulovillous adenoma; WT, wild type; NA, Not Available; NOS, Not Otherwise Specified.

Supplemtary Table S2. Histopathological and molecular features of 36 adenomas from 17 FAP/AFAP patients.

FAP/AFAP patients	*APC* germline mutations		Age at diagnosis	Adenoma	Histological features	Site	Dysplasia	% Methylation		Mutation	
	DNA change	Protein						L1	L1-*MET*	*KRAS*	*NRAS*
F1	c.3577_3578 delCA	p.Gln1193ValfsX14	29	A1	TB	Rectum	low	68.25	66.43	WT	WT
F2	c.505delATAGA	p.Ile169X	31	A1	TB	Sigma	low	54.10	58.80	WT	WT
				A2	TB	Rectum-Sigma	low	66.75	64.47	WT	WT
				A3	TB	Colon NOS	low	59.33	60.10	WT	WT
F3	c.505delATAGA	p.Ile169X	38	A1	TB	Right colon	low	60.33	61.53	WT	WT
F4	c.2524_2525 insA	p.Asp842 GlufsX74	37	A1	TB	Sigma	low	67.53	64.83	p.G12V	WT
				A2	TB	Right colon	low	63.43	NA	WT	WT
F5	c.4147_4148 delAT	p.Met1383ValfsX2	30	A1	TB	Left colon	NA	68.78	60.43	WT	WT
				A2	TBV	Left colon	low	60.40	53.83	p.G12S	WT
				A3	TBV	Sigma	low	64.03	61.60	p.G12V	WT
F6	Whole gene del.		48	A1	TBV	NA	low	67.33	61.47	p.G12D	WT
F7	c.2734delT	p.Leu912 TyrfsX3	21	A1	TB	Colon NOS	low	71.95	60.63	WT	WT
F8	c.2734delT	p.Leu912 TyrfsX3	23	A1	TB	Sigma	low	64.58	69.00	WT	WT
F9	c.677_684 delAGGACATAinsTTTC	p.Lys226 IlefsX66	61	A1	TBV	NA	low	62.75	62.60	p.G12V + p.A146P	WT
				A2	TBV	NA	low	59.8	NA	p.G12V	WT
F10	c.3495-3496insA	p.Tyr1166 IlefsX1	19	A1	TB	Sigma	low	64.38	61.30	WT	WT
F11	c.745A>T	p.Lys249X	37	A1	TB	Rectum	low	66.50	62.77	WT	WT
F12	c.2626C>T	p.R876X	37	A1	TBV	Rectum-Sigma	high	53.53	NA	p.G13D	WT

F13	c.3926del5	p.Glu1309AspfsX3	42	A1	TB	Left colon	low	60.73	61.2	WT	WT
				A2	TB	Left colon	low	61.88	63.07	p.G13C	WT
				A3	TB	Colon NOS	low	60.3	59.63	WT	WT
				A4	TBV	Colon NOS	low	57.83	58	WT	WT
				A5	TB	Colon NOS	low	60.13	58.93	WT	WT
				A6	TBV	Ampulla of Vater	low	65.5	61.53	WT	WT
				A7	TBV	Colon NOS	high	61.1	62.73	WT	WT
				A8	TB	Left colon	low	59.43	60.33	WT	WT
				A9	TB	Right colon	low	62.1	57.83	WT	WT
F14	c.4549_455 delCAGA	p.Gln1517LysfsX5	57	A1	TB	Duodenum	low	63.1	63.6	WT	WT
				A2	TB	Ampulla of Vater	low	65.38	63.03	WT	WT
				A3	TB	Ampulla of Vater	low	65.1	61.13	WT	WT
F15	t(5;7) (q22;p15)		33	A1	TB	Rectum	low	62.63	62.56	WT	WT
F16	c.637C>T	p.R213X	22	A1	TB	Rectum	low	70.18	67.3	NA	NA
				A2	TB	Right colon	low	72.98	68.13	WT	WT
				A3	TB	Left colon	low	70.25	64.8	WT	WT
				A4	TB	Rectum	low	61.1	68.06	WT	WT
F17	Exon 2del.		40	A1	TB	Sigma	low	62.95	68.3	WT	WT

Legend: FAP/AFAP, classical/attenuated adenomatous polyposis; TB, tubular adenoma; TBV, tubulovillous adenoma; WT, wild type; NA, Not Avaiable; NOS, Not Otherwise Specified.

Supplemtary Table S3. Histopathological and molecular features of 80 sporadic colorectal adenomas from 62 patients.

Sporadic patient	Age at dignosis	Adenoma	Histological features	Site	Dysplasia	Size (mm)	N° of adenomas at diagnosis	% Methylation		Mutation	
								L1	L1-*MET*	*KRAS*	*NRAS*
S-Ads											
S1	62	A1	TBV	Rectum-Sigma	high	18	2	66.53	65.13	p.G12D	WT
S2	66	A1	TB	Rectum	low	6	3	61.38	65.60	WT	WT
S3	66	A1	TBV	Sigma	low	18	1	59.48	62.37	WT	WT
S4	77	A1	TBV	Sigma	NA	18	1	60.88	63.03	WT	WT
S5	54	A1	TBV	Sigma	high	7	2	63.03	66.40	WT	WT
S6	82	A1	TBV	Sigma	low	18	1	62.03	65.67	WT	WT
S7	75	A1	TBV	Sigma	low	5	1	62.03	66.03	WT	WT
S8	72	A1	TBV	Sigma	low	10	3	59.05	66.37	p.G12V	WT
S9	61	A1	TBV	Rectum-Sigma	high	17	2	60.25	62.17	p.G12V	WT
S10	71	A1	TBV	Sigma	high	6	2	62.83	63.40	p.G12D	WT
S11	50	A1	TBV	Colon NOS	low	2	2	60.55	63.43	p.G12D	WT
S12	67	A1	TB	Colon NOS	low	1	2	62.4	61.2	WT	WT
S13	68	A1	TB	Ascending colon	low	5	4	60.05	59	WT	WT
S14	84	A1	TB	Hepatic flexure	low	4	2	63.13	62.9	WT	WT
S15	62	A1	TB	Ascending colon	low	8	2	63	62.1	WT	WT
S16	59	A1	TB	Colon NOS	low	2	3	60.38	62.23	WT	WT
		A2	TB	Colon NOS	low	2		60.73	62.56	WT	WT
		A3	TB	Colon NOS	low	NA		61.06	62.06	WT	WT
S17	57	A1	TB	Sigma	low	2	4	62.75	60.06	WT	WT
		A2	TB	Left colon	low	2		60.07	60.36	WT	WT
S18	61	A1	TB	Colon NOS	low	4	2	67.3	68.6	WT	WT

S19	61	A1	TB	Ascending colon	low	3	6	62.9	67	p.G12A	WT
S20	73	A1	TB	Colon NOS	low	4	2	70	65.03	WT	WT
S21	74	A1	TB	Rectum-Sigma	low	7	2	69.23	65.07	WT	WT
S22	68	A1	TBV	Left colon	low	9	1	60.88	58.40	p.G12E	WT
S23	64	A2	TBV	Left colon	low	12	1	59.43	56.93	WT	WT
S24	57	A1	TBV	Left colon	low	5	1	65.43	60.47	WT	WT
S25	52	A1	TBV	Left colon	low	12	1	58.10	57.57	WT	WT
S26	64	A1	TBV	Right colon	low	6	1	59.60	58.87	WT	WT
S27	73	A1	TBV	Left colon	low	10	2	59.13	61.97	WT	WT
		A2	TBV	Left colon	low	15		62.45	60.37	WT	WT
S28	61	A1	TBV	Left colon	low	10	1	62.08	58.57	p.G12S	WT
S29	38	A1	TBV	Right colon	high	10	2	56.60	56.27	p.G13D	WT
	44	A2	TB	Right colon	high	16		53.80	55.77	p.G13D	WT
S30	50	A1	TBV	Right colon	low	6	1	60.33	57.47	WT	WT
S31	69	A1	TBV	Left colon	low	>20	1	58.33	57.40	WT	WT
S32	64	A1	TBV	Right colon	low	22	2	57.25	54.5	p.G12D	WT
		A2	TBV	Right colon	low	22		58.2	55.9	p.G12D	WT
S33	49	A1	TBV	Left colon	low	3	1	60.4	58.3	WT	WT
S34	63	A1	TBV	Right colon	low	5	1	59.3	58.0	WT	WT
S35	75	A1	TBV	Left colon	low	20	2	60.1	56.5	p.G13D	WT
		A2	TBV	Left colon	low	20		58.63	55.9	p.G13D	WT
S36	59	A1	TB	Left colon	low	5	1	61.75	57.4	WT	WT
S37	74	A1	TBV	Right colon	low	5	2	53.38	50.1	WT	p.G12D
		A2	TB	Right colon	low	3		67.6	53.4	WT	WT
S38	66	A1	TB	Left colon	low	5	2	58.1	58.3	WT	WT
S39	76	A1	TBV	Left colon	high	30	4	56.7	51.2	p.G12D	WT
		A2	TBV	Left colon	high	30		55.8	48.8	p.G12S	WT
		A3	TBV	Left colon	high	30		54.0	47.2	p.G12D	WT

		A4	TBV	Left colon	high	30		53.7	49.2	p.G12S	WT
S40	70	A1	TB	Right colon	low	10	1	59.1	57.1	WT	WT
S41	55	A1	TBV	Left colon	high	15	2	57.9	57.5	WT	WT
S42	63	A1	TBV	Left colon	low	8	1	57.6	57.7	WT	WT
S43	59	A1	TBV	Left colon	low	11	1	60.0	59.1	WT	WT
S44	64	A1	TB	Left colon	low	13	1	56.6	56.1	p.A146T	WT
S45	62	A1	TBV	Left colon	high	15	1	77.0	59.7	p.G12V	WT
S-AdsC											
S46	65	A1	V	Left colon	low	15	1	58.20	57.93	WT	WT
S47	60	A1	V	Left colon	high	30	3	55.18	54.23	p.G13D	WT
S48	72	A1	TBV	Right colon	high	30	3	62.70	60.87	p.Q61L	WT
		A2	TBV	Right colon	high	18		61.28	59.83	p.G12A	WT
S49	58	A1	TBV	Right colon	high	13	1	59.23	60.83	WT	WT
S50	76	A1	TBV	Left colon	low	13	3	67.90	58.87	WT	WT
		A2	TB	Right colon	low	>20		57.50	59.43	WT	WT
S51	78	A1	TBV	Right colon	low	20	2	63.80	59.27	p.G12D	WT
		A2	TB	Right colon	low	20		57.10	61.93	p.G12D	WT
S52	67	A1	TB	Left colon	low	>20	1	65.78	59.47	WT	WT
S53	82	A1	TB	Left colon	low	15	1	69.25	62.97	p.G12D	WT
S54	69	A1	TB	Left colon	low	NA	1	65.68	60.30	WT	WT
S55	68	A1	TBV	Right colon	low	23	1	49.75	52.43	p.G13D	WT
S56	82	A1	TB	Right colon	low	>20	3	55.55	59.03	WT	WT
		A2	TB	Right colon	low	> 20		58.15	60.73	WT	WT
S57	66	A1	TB	Right colon	low	<20	2	48.48	51.40	WT	WT
	69	A2	TB	Right colon	low	>20	4	61.23	59.23	p.G12A	WT
S58	74	A1	TB	Left colon	low	6	1	55.95	58.47	WT	WT
S59	61	A1	TBV	Left colon	low	<20	2	59.70	57.23	WT	WT

S60	63	A1	TB	Left colon	low	>20	1	65,63	60,03	WT	WT
S61	72	A1	TB	Left colon	low	>20	2	67.03	62.23	WT	WT
		A2	V	Left colon	high	55		63.50	57.67	p.G12V	WT
		A3	V	Left colon	high	15		52.63	58.07	p.G12V	WT
S62	58	A1	TBV	Left colon	low	NA	1	55.40	57.47	p.G13D	WT

Legend: S-Ads, patients who have never developed a CRC during a follow-up of ten years; S-AdsC, patients who developed a CRC at least one year after polyp removal; TB, tubular adenoma; TBV, tubulovillous adenoma; V, villous adenoma; WT, wild type; NA, Not Avaiable; NOS, Not Otherwise Specified.

Supplemtary Table S4. Molecular features of the 11 MAP adenocarcinomas.

MAP patient	*MUTYH* germline mutations		Age at diagnosis	Adenocarcinoma	Site	% Methylation		Mutation			
	DNA change	Protein				L1	L1-*MET*	*KRAS*	*NRAS*	*BRAF*	*PI3KCA*
M1	c.1187G>A/ c.1187insGG	p.G396D/ p.Gly396 GlyfsX12	51	CRC	NA	58.70	57.73	p.G12C	WT	WT	p.Q546K
M4	c.536A>G / c.536A>G	p.Y179C/ p.Y179C	47	CRC	Ascending colon	63.00	59.97	p.G12C	WT	WT	WT
				CRC	Ascending colon	59.55	59.54	p.G12C	WT	WT	WT
M6	c.536A>G/ c.1105G>T	p.Y179C/ p.E369X	42	CRC	Colon NOS	52.10	51.90	p.G12C	WT	WT	p.Q546K
M9	c.1186-1187 insGG/ c.1186-1187 insGG	p.Gly396 GlyfsX12/ p.Gly396GlyfsX12	40	CRC	Rectum	62.52	58.90	p.G12C	WT	WT	WT
				CRC	Right colon	65.23	61.47	p.G12V	WT	WT	WT
M11	c.1145G>A/ c.1145G>A	p.G382D/ p.G382D	47	CRC	Rectum	56. 33	57.13	p.G12C	WT	WT	WT
M13	c.733C>T/ c.1187G>A	p.R245C/ p.G396D	57	CRC	Right colon	53.03	51.46	WT	WT	WT	WT
M16	c.536A>G / c.933 +3A>C	p.Y179C/ p.Gly264 TrpfsX7	54	CRC	Rectum	64.45	59.63	p.G12C	WT	WT	WT
M17	c.536A>G/ c.721C>T	p.Y179C/ R241W	41	CRC	Right colon	59.60	58.50	p.G12C	WT	WT	p.R38H
M18	c.1012C>T/ 1187G>A	p.Gln338*/ G396D	50	CRC	Right Colon	56.83	59.73	p.G12C	WT	WT	p.Q546K

Legend: MAP, *MUTYH*-associated-polyposis; CRC, colorectal cancer; WT, wild type; NA, Not Available; NOS, Not Otherwise Specified.

Supplemtary Table S5. Molecular features of the 15 sporadic adenocarcinomas.

Sporadic patient	Age at diagnosis	Adenocarcinoma	Site	% Methylation		Mutation			
				L1	L1-*MET*	*KRAS*	*NRAS*	*BRAF*	*PI3KCA*
S48	72	CRC	Right colon	55.6	56.4	p.G12V	WT	WT	WT
S49	58	CRC	Right colon	49.0	49.7	p.G12A	WT	WT	WT
S50	76	CRC	Right colon	62.4	62.5	p.G12A	WT	WT	WT
S51	78	CRC	Right Colon	57.0	54.9	WT	WT	WT	WT
S52	67	CRC	Left Colon	61.3	55.4	p.G12D	WT	WT	WT
S53	82	CRC	Left Colon	59.6	60.2	p.G12D	WT	WT	WT
S54	69	CRC	Left Colon	58.3	53.5	p.G12S	WT	WT	WT
S55	68	CRC	Right Colon	55,1	56.4	p.G13D	WT	WT	WT
S56	82	CRC	Right Colon	56.5	57.9	WT	WT	WT	WT
S57	72	CRC	Right Colon	61.1	60.2	p.G12A	WT	WT	WT
S58	74	CRC	Left Colon	62.9	58.2	WY	WT	p.V600E	WT
S59	61	CRC	Left Colon	49.5	50.7	WT	WT	WT	WT
S60	63	CRC	Left Colon	54.8	55.1	WT	WT	WT	WT
S61	72	CRC	Left Colon	54.6	NA	p.G12S	WT	WT	WT
S62	58	CRC	Left Colon	52.5	55.2	WT	p.A146T	WT	WT

Legend: CRC, colorectal cancer; WT, wild type; NA, Not Available; NOS, Not Otherwise Specified.

Supplemtary Table S6. *MET*/L1-*MET* expression level of 9 L1 and L4 group CRCs.

Patient	Adenocarcinoma	% Methylation	Expression levels (FC)	
		L1-*MET*	*MET*	L1-*MET*
L1-1	CRC	61.4	1.02	1.87
L1-2	CRC	60.9	084	1.9
L1-3	CRC	63.97	0.67	1.79
L1-4	CRC	60.47	1.61	2.32
L1-5	CRC	61.73	2.26	1.98
L1-6	CRC	64.4	3.68	1.68
L1-7	CRC	63.4	1.85	1.2
L1-8	CRC	55.7	2.44	3.17
L1-9	CRC	55.9	1.45	2.4
L4-1	CRC	54	6.86	7.26
L4-2	CRC	57.43	6.23	8.93
L4-3	CRC	52.8	4.43	6.23
L4-4	CRC	56.4	6.19	4.56
L4-5	CRC	50.26	4.75	7.26
L4-6	CRC	56.3	4.59	6.1
L4-7	CRC	53.55	6.23	9.38
L4-8	CRC	40.8	5.48	8.03
L4-9	CRC	53.1	3.86	3.87
L4-10	CRC	48.5	2.05	3.39
L4-11	CRC	58.8	4.7	4.25

Legend: CRC, colorectal cancer; L1 group CRCs; L4 group CRCs; FC, Fold Change.

Supplemtary Table S7. *MET*/L1-*MET* expression levels of 13 MAP, 16 FAP and 56 sporadic adenomas analyzed.

Patient	Adenoma	% Methylation	Expression levels (FC)	
		L1-*MET*	*MET*	L1-*MET*
M2	A2	69.57	0.38	0.00
M4	A1	62.53	0.76	2.23
M4	A2	59.8	2.04	6.19
	A3	60.87	0.42	2.68
	A4	59.37	1.72	6.45
M9	A1	59.9	5.17	8.63
	A2	64.8	1.29	0
	A4	55.73	3.51	5.24
	A5	55.88	2.60	6.15
	A6	62.53	0.65	0
	A7	62.63	1.16	0
M18	A1	59.7	2.69	6.68
	A2	60.33	2.38	3.05
F13	A1	61.2	0.93	0
	A2	63.07	0.50	0
	A3	59.63	1.85	5.31
	A4	58	2.66	5.74
	A5	58.93	1.66	5.06
	A6	61.53	0.45	0
	A7	62.73	0.93	0
	A8	60.33	0.99	0
	A9	57.83	3.34	4.86
F14	A1	63.6	0.63	0
	A2	63.03	0.52	0
	A3	61.13	0.49	0

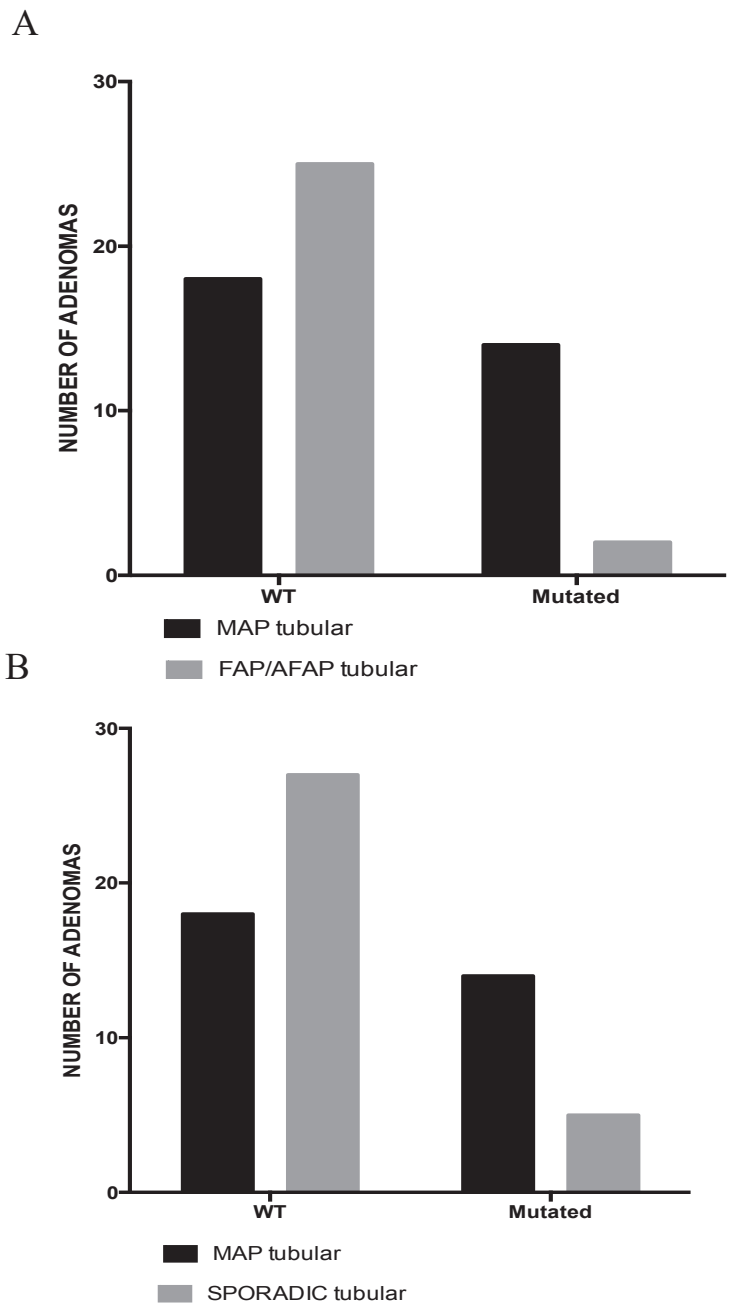
F15	A1	62.56	0.63	0
F16	A2	68.13	0.51	0
	A3	64.8	0.44	0
	A4	68.06	0.52	0
S22	A1	58.40	0	0
S23	A2	56.93	13.98	0
S24	A1	60.47	0	0
S25	A1	57.57	0	0
S26	A1	58.87	15	0
S27	A1	61.97	0	0
	A2	60.37	0	0
S28	A1	58.57	12.55	0
S29	A1	56.27	1.23	0
	A2	55.77	5.80	17.15
S30	A1	57.47	0	0
S31	A1	57.40	1.12	0
S32	A1	54.5	1.8	0.44
	A2	55.9	2.2	2.31
S33	A1	58.3	3.01	0
S34	A1	58.0	2.81	2.46
S35	A1	56.5	2.84	1.53
	A2	55.9	2.79	0.67
S36	A1	57.4	2.65	0.81
S37	A1	50.1	3.48	0
	A2	53.4	3.03	0
S38	A1	58.3	4.31	2.36
S39	A1	51.2	2.56	3.89
	A2	48.8	3.22	3.20
	A3	47.2	3.14	3.27

	A4	49.2	2.07	3.01
S40	A1	57.1	3.81	0.43
S41	A1	57.5	3.18	4.75
S42	A1	57.7	3.45	6.63
S43	A1	59.1	4.72	4.69
S44	A1	56.1	3.58	2.02
S45	A1	59.7	3.65	3.91
S46	A1	57.93	6.75	0
S47	A1	54.23	21.26	0
S48	A1	60.87	2.17	0
	A2	59.83	2.78	1.10
S49	A1	60.83	0.60	0
S50	A1	58.87	2.15	0
	A2	59.43	1,05	0
S51	A1	59.27	1.48	0
	A2	61.93	0.97	0
S52	A1	59.47	7.01	6.87
S53	A1	62.97	3.08	0
S54	A1	60.30	0.71	0.66
S55	A1	52.43	3.16	0
S56	A1	59.03	5.17	0
	A2	60.73	0.45	0
S57	A1	51.40	3.59	0
	A2	59.23	2.35	0
S58	A1	58.47	1.78	0
S59	A1	57.23	1.64	6.08
S60	A1	60.03	1.23	0
S61	A1	62.23	2.57	0
	A2	57.67	2.76	37.79

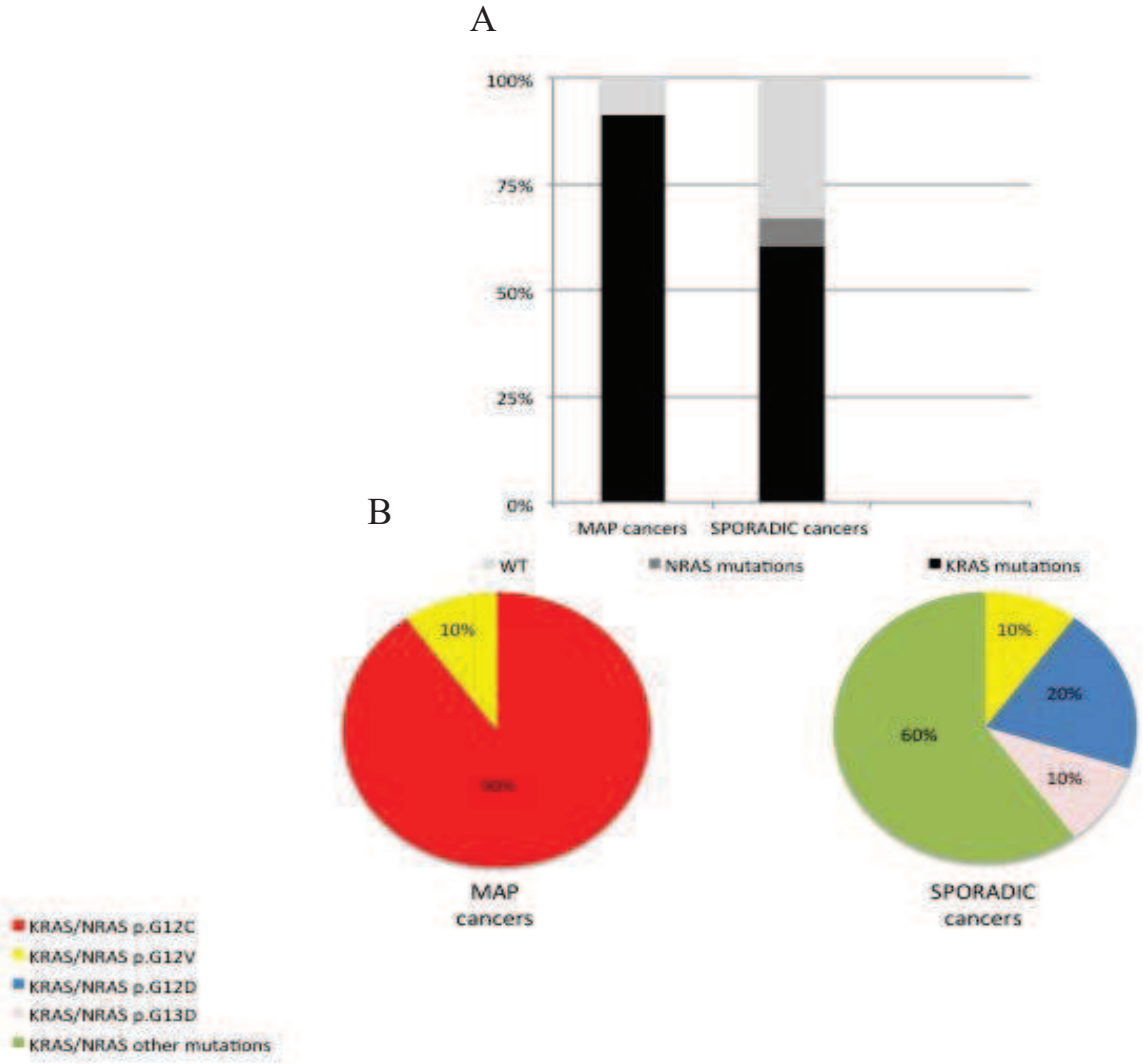
	A3	58.07	0.42	8.08
S62	A1	57.47	7.86	24.25

Legend: M, MAP patient; F, FAP patient; S, sporadic patient; FC, Fold change.

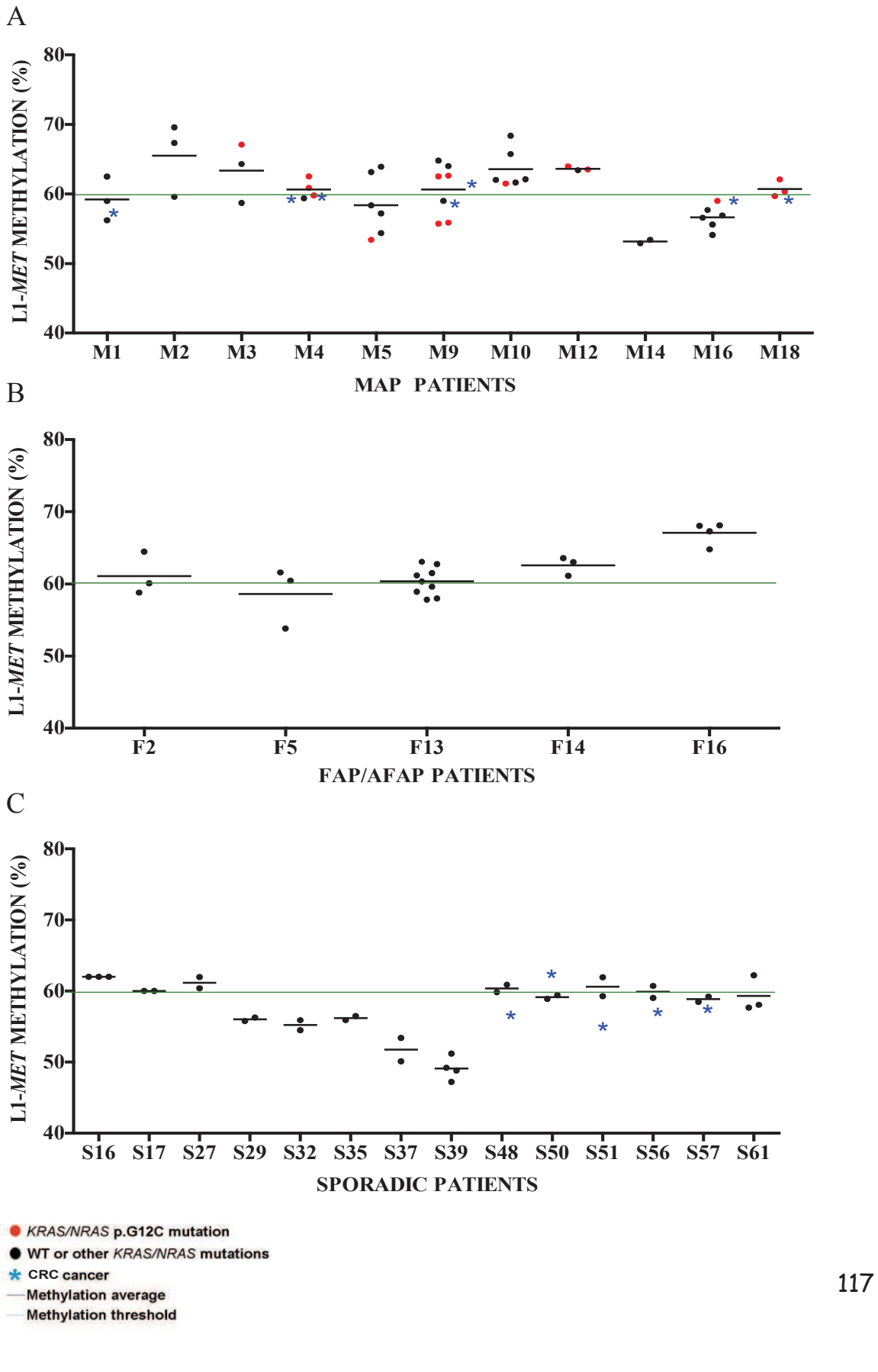
Supplemtary Figure S1. Correlation between mutational status and tubular architecture of MAP adenomas compare to that FAP/AFAP and sporadic (B). Columns represent the number of analysed adenomas (y axis) respect to mutational status (x axis).



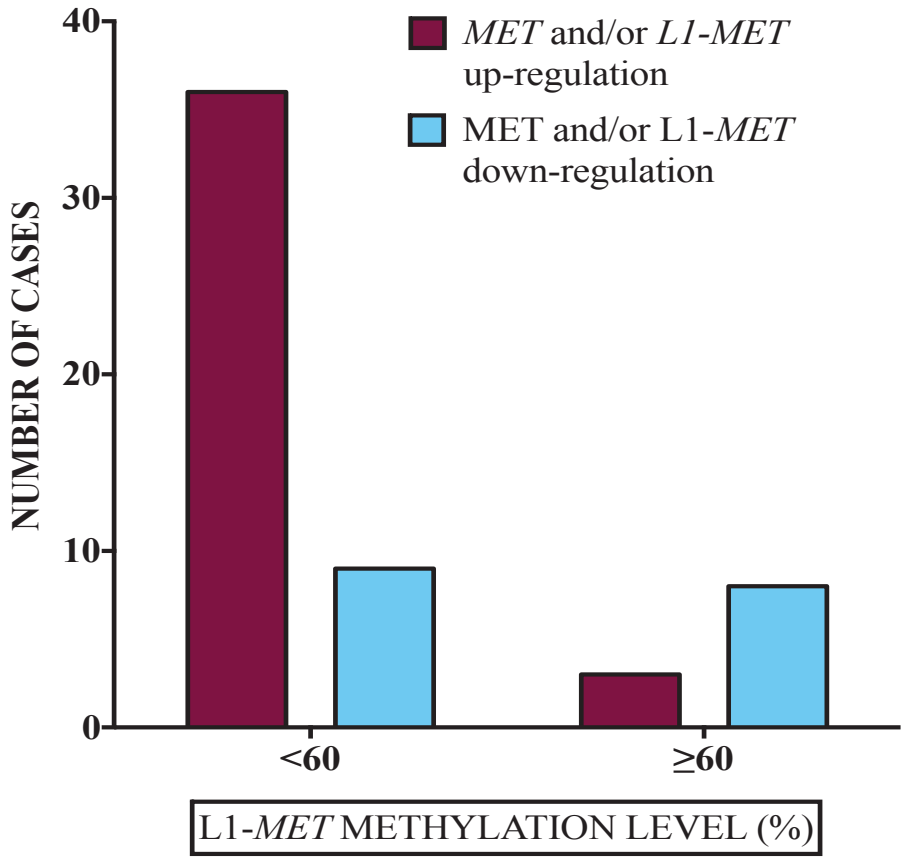
Supplemtary Figure S2. *KRAS/NRAS* mutations in the MAP and sporadic CRC cohort. (A) Percentage of mutations in MAP and sporadic CRCs; columns represent the mutation frequencies and *KRAS/NRAS* mutations or wild-type (WT) status are indicated in gray scale as reported as per the legend below.
(B) Spectrum of *KRAS/NRAS* mutations in two groups of CRC (MAP and sporadic) are reported in pie charts and colours show the different types of *KRAS/NRAS* alterations as per the legend below.



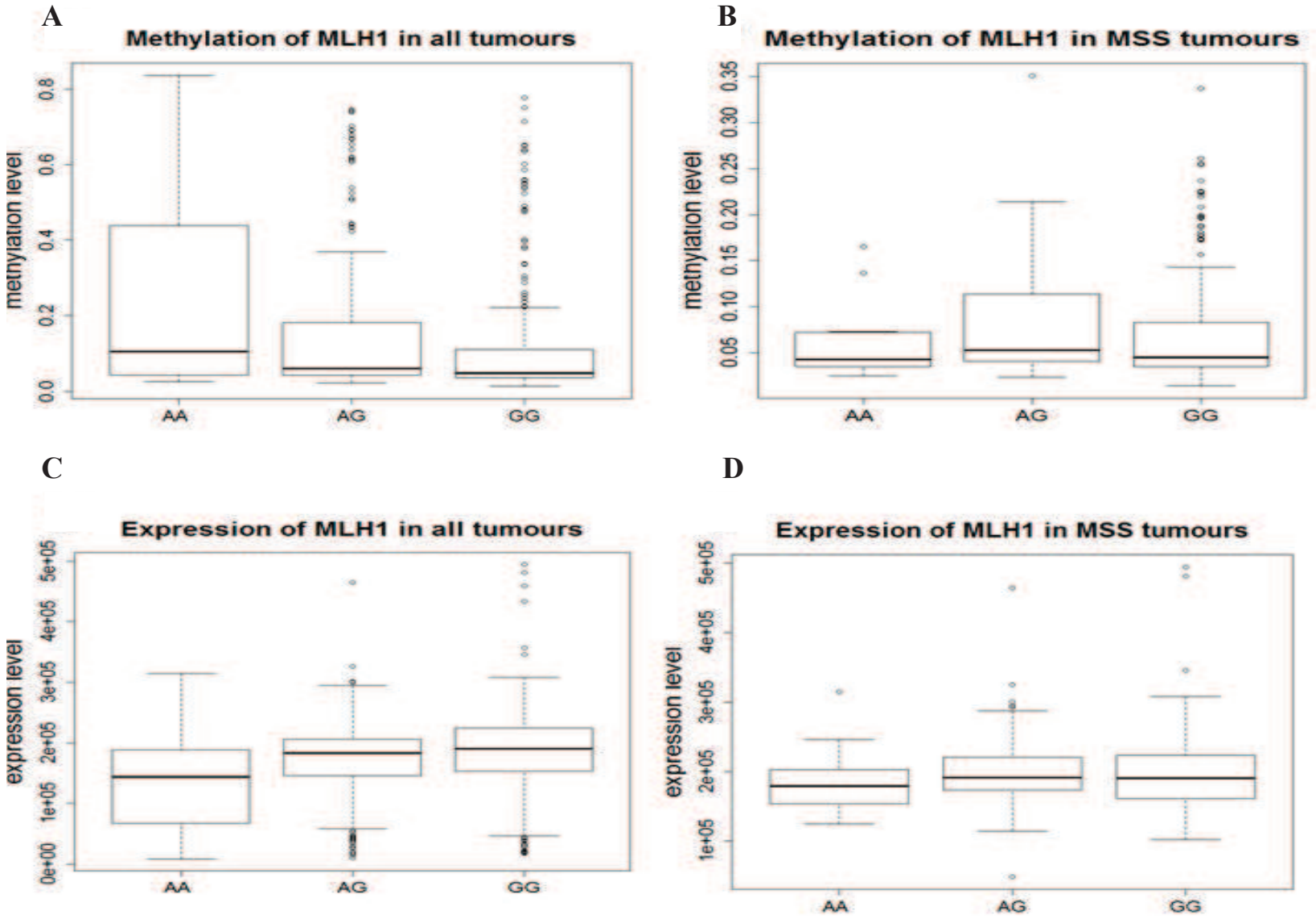
Supplemtary Figure S3. (A) Dot plots represent L1-*MET* methylation percentage in the multiple adenomas of MAP (A), FAP/AFAP (B) and sporadic (C) patients; black dots identify adenomas and red dots indicate *KRAS*/*NRAS* p.G12C mutated polyps; the green line symbolizes the methylation threshold (60%) for L1 assay; bars show the mean values of L1 methylation percentage among adenomas of the same patients and light blu asterisks show the methyilation level of the corresponding carcinomas. symbols and colours are reported as per the legend below. WT, wild type.



Supplemtary Figure S4. (A) Colums represent the number of cases with L1-*MET* methylation <60% and ≥60% stratifying sporadic adenomas according to *MET* and/or L1-*MET* expression status.



Supplemtary Figure S5. Methylation (A-B) and expression (C-D) of *MLH1* in TCGA COADREAD samples stratified by rs1800734 genotype.



Supplemtary Figure S6. Increase in *MLH1* expression levels in SW48 (MSI) cells after treatment with AzaC. Bar chart showing total MLH1 mRNA expression in control untreated cells and timepoints post AzaC treatment.

BIBLIOGRAPHY

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.

2. Cho KR, Vogelstein B. Genetic alterations in the adenoma--carcinoma sequence. Cancer 1992;70:1727-31.

3. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135:1079-99.

4. Pretlow TP, Barrow BJ, Ashton WS, O'Riordan MA, Pretlow TG, Jurcisek JA, Stellato TA. Aberrant crypts: putative preneoplastic foci in human colonic mucosa. Cancer Res 1991;51:1564-7.

5. Smith AJ, Stern HS, Penner M, Hay K, Mitri A, Bapat BV, Gallinger S. Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. Cancer Res 1994;54:5527-30.

6. Shpitz B, Bomstein Y, Mekori Y, Cohen R, Kaufman Z, Neufeld D, Galkin M, Bernheim J. Aberrant crypt foci in human colons: distribution and histomorphologic characteristics. Hum Pathol 1998;29:469-75.

7. Takayama T, Ohi M, Hayashi T, Miyanishi K, Nobuoka A, Nakajima T, Satoh T, Takimoto R, Kato J, Sakamaki S, Niitsu Y. Analysis of K-ras, APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis. Gastroenterology 2001;121:599-611.

8. Alrawi SJ, Schiff M, Carroll RE, Dayton M, Gibbs JF, Kulavlat M, Tan D, Berman K, Stoler DL, Anderson GR. Aberrant crypt foci. Anticancer Res 2006;26:107-19.

9. Heitman SJ, Ronksley PE, Hilsden RJ, Manns BJ, Rostom A, Hemmelgarn BR. Prevalence of adenomas and colorectal cancer in average risk individuals: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2009;7:1272-8.

10. Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol 2012;3:153-73.

11. Silva P, Albuquerque C, Lage P, Fontes V, Fonseca R, Vitoriano I, Filipe B, Rodrigues P, Moita S, Ferreira S, Sousa R, Claro I, et al. Serrated polyposis associated with a family history of colorectal cancer and/or polyps: The preferential location of polyps in the colon and rectum defines two molecular entities. Int J Mol Med 2016;38:687-702.

12. Fu X, Li L, Peng Y. Wnt signalling pathway in the serrated neoplastic pathway of the colorectum: possible roles and epigenetic regulatory mechanisms. J Clin Pathol 2012;65:675-9.

13. Marmol I, Sanchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. Int J Mol Sci 2017;18.

14. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol 2001;96:2992-3003.

15. Rustgi AK. The genetics of hereditary colon cancer. Genes Dev 2007;21:2525-38.

16. Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. Eur J Cancer 2006;42:216-27.

17. Song X, Gong X, Zhang T, Jiang W. Height and risk of colorectal cancer: a meta-analysis. Eur J Cancer Prev 2018;27:521-9.

18. Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, Chadwick RB, Kaariainen H, Eskelinen M, Jarvinen H, Mecklin JP, de la Chapelle A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338:1481-7.

19. Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Koessler T, Pharoah P, Schafmayer C, Hampe J, Volzke H, Chang-Claude J, Hoffmeister M, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut 2013;62:871-81.

20. Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, Church JM, Dominitz JA, Johnson DA, Kaltenbach T, Levin TR, Lieberman DA, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on Colorectal Cancer. Dis Colon Rectum 2014;57:1025-48.

21. Tezcan G, Tunca B, Ak S, Cecener G, Egeli U. Molecular approach to genetic and epigenetic pathogenesis of early-onset colorectal cancer. World J Gastrointest Oncol 2016;8:83-98.

22. Davis DM, Marcet JE, Frattini JC, Prather AD, Mateka JJ, Nfonsam VN. Is it time to lower the recommended screening age for colorectal cancer? J Am Coll Surg 2011;213:352-61.

23. Munkholm P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment Pharmacol Ther 2003;18 Suppl 2:1-5.

24. Freeman HJ.Colorectal cancer risk in Crohn's disease.World J Gastroenterol 2008;14:1810-1.

25. Lakatos PL, Lakatos L. Risk for colorectal cancer in ulcerative colitis: changes, causes and management strategies. World J Gastroenterol 2008;14:3937-47.

26. Tariq H, Kamal MU, Patel H, Patel R, Ameen M, Elona S, Khalifa M, Azam S, Zhang A, Kumar K, Baiomi A, Shaikh D, et al. Predicting the presence of adenomatous polyps during colonoscopy with National Cancer Institute Colorectal Cancer Risk-Assessment Tool. World J Gastroenterol 2018;24:3919-26.

27. Lin SH, Raju GS, Huff C, Ye Y, Gu J, Chen JS, Hildebrandt MAT, Liang H, Menter DG, Morris J, Hawk E, Stroehlein JR, et al. The somatic mutation landscape of premalignant colorectal adenoma. Gut 2018;67:1299-305.

28. Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2001;10:725-31.

29. Moskal A, Norat T, Ferrari P, Riboli E. Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. Int J Cancer 2007;120:664-71.

30. Alexander DD, Weed DL, Cushing CA, Lowe KA. Meta-analysis of prospective studies of red meat consumption and colorectal cancer. Eur J Cancer Prev 2011;20:293-307.

31. Boyle T, Keegel T, Bull F, Heyworth J, Fritschi L. Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. J Natl Cancer Inst 2012;104:1548-61.

32. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH, Panish JF, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993;329:1977-81.

33. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. CA Cancer J Clin 2008;58:130-60.

34. Moiel D, Thompson J. Early detection of colon cancer-the kaiser permanente northwest 30-year history: how do we measure success? Is it the test, the number of tests, the stage, or the percentage of screen-detected patients? Perm J 2011;15:30-8.

35. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, Church TR. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-14.

36. Lieberman DA. Clinical practice. Screening for colorectal cancer. N Engl J Med 2009;361:1179-87.

37. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 1993;328:1365-71.

38. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, James PD, Mangham CM. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-7.

39. Mandel JS, Church TR, Ederer F, Bond JH. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.

40. Mandel JS, Church TR, Bond JH, Ederer F, Geisser MS, Mongin SJ, Snover DC, Schuman LM. The effect of fecal occult-blood screening on the incidence of colorectal cancer. N Engl J Med 2000;343:1603-7.

41. Garborg K, Holme O, Loberg M, Kalager M, Adami HO, Bretthauer M. Current status of screening for colorectal cancer. Ann Oncol 2013;24:1963-72.

42. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of colonoscopy and death from colorectal cancer. Ann Intern Med 2009;150:1-8.

43. Kahi CJ, Imperiale TF, Juliar BE, Rex DK. Effect of screening colonoscopy on colorectal cancer incidence and mortality. Clin Gastroenterol Hepatol 2009;7:770-5; quiz 11.

44. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 2011;154:22-30.

45. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, Shi W, Bond JH, Schapiro M, Panish JF, Stewart ET, Waye JD. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366:687-96.

46. Bacchus CM, Dunfield L, Gorber SC, Holmes NM, Birtwhistle R, Dickinson JA, Lewin G, Singh H, Klarenbach S, Mai V, Tonelli M. Recommendations on screening for colorectal cancer in primary care. CMAJ 2016;188:340-8.

47. Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, Jr., Garcia FAR, Gillman MW, Harper DM, Kemper AR, Krist AH, Kurth AE, Landefeld CS, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315:2564-75.

48. Compton CC, Greene FL. The staging of colorectal cancer: 2004 and beyond. CA Cancer J Clin 2004;54:295-308.

49. Hari DM, Leung AM, Lee JH, Sim MS, Vuong B, Chiu CG, Bilchik AJ. AJCC Cancer Staging Manual 7th edition criteria for colon cancer: do the complex modifications improve prognostic assessment? J Am Coll Surg 2013;217:181-90.

50. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin 2014;64:104-17.

51. Gonzalez RS, Washington K, Shi C. Current applications of molecular pathology in colorectal carcinoma. Applied Cancer Research 2017.

52. Sinicrope FA, Rego RL, Halling KC, Foster N, Sargent DJ, La Plant B, French AJ, Laurie JA, Goldberg RM, Thibodeau SN, Witzig TE. Prognostic impact of microsatellite instability and DNA ploidy in human colon carcinoma patients. Gastroenterology 2006;131:729-37.

53. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci U S A 2007;104:18654-9.

54. Cheng YW, Pincas H, Bacolod MD, Schemmann G, Giardina SF, Huang J, Barral S, Idrees K, Khan SA, Zeng Z, Rosenberg S, Notterman DA, et al. CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. Clin Cancer Res 2008;14:6005-13.

55. Levine AJ, Phipps AI, Baron JA, Buchanan DD, Ahnen DJ, Cohen SA, Lindor NM, Newcomb PA, Rosty C, Haile RW, Laird PW, Weisenberger DJ. Clinicopathologic Risk Factor Distributions for MLH1 Promoter Region Methylation in CIMP-Positive Tumors. Cancer Epidemiol Biomarkers Prev 2015;25:68-75.

56. Grady WM. Genomic instability and colon cancer. Cancer Metastasis Rev 2004;23:11-27.

57. Roberts DM, Pronobis MI, Poulton JS, Kane EG, Peifer M. Regulation of Wnt signaling by the tumor suppressor adenomatous polyposis coli does not require the ability to enter the nucleus or a particular cytoplasmic localization. Mol Biol Cell 2012;23:2041-56.

58. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330-7.

59. Rubinfeld B, Robbins P, El-Gamil M, Albert I, Porfiri E, Polakis P. Stabilization of beta-catenin by genetic defects in melanoma cell lines. Science 1997;275:1790-2.

60. Segditsas S, Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene 2006;25:7531-7.

61. Li XL, Zhou J, Chen ZR, Chng WJ. P53 mutations in colorectal cancer - molecular pathogenesis and pharmacological reactivation. World J Gastroenterol 2015;21:84-93.

62. Muller MF, Ibrahim AE, Arends MJ. Molecular pathological classification of colorectal cancer. Virchows Arch 2016;469:125-34.

63. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 1983;11:6883-94.

64. Lengauer C, Kinzler KW, Vogelstein B. DNA methylation and genetic instability in colorectal cancer cells. Proc Natl Acad Sci U S A 1997;94:2545-50.

65. Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. Biochim Biophys Acta 2007;1775:138-62.

66. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics 2009;1:239-59.

67. Miousse IR, Koturbash I. The Fine LINE: Methylation Drawing the Cancer Landscape. Biomed Res Int 2015;2015:131547.

68. Xiao-Jie L, Hui-Ying X, Qi X, Jiang X, Shi-Jie M. LINE-1 in cancer: multifaceted functions and potential clinical implications. Genet Med 2016;18:431-9.

69. Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat Commun 2014;5:5011.

70. Figueiredo JC, Grau MV, Wallace K, Levine AJ, Shen L, Hamdan R, Chen X, Bresalier RS, McKeown-Eyssen G, Haile RW, Baron JA, Issa JP. Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. Cancer Epidemiol Biomarkers Prev 2009;18:1041-9.

71. Baba Y, Huttenhower C, Nosho K, Tanaka N, Shima K, Hazra A, Schernhammer ES, Hunter DJ, Giovannucci EL, Fuchs CS, Ogino S. Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors. Mol Cancer 2010;9:125.

72. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, Cuatrecasas M, Bujanda L, Giraldez MD, Takahashi M, Cabanne A, Barugel ME, et al. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS One 2012;7:e45357.

73. Murata A, Baba Y, Watanabe M, Shigaki H, Miyake K, Ishimoto T, Iwatsuki M, Iwagami S, Sakamoto Y, Miyamoto Y, Yoshida N, Nosho K, et al. Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. Br J Cancer 2013;109:408-15.

74. Baba Y, Murata A, Watanabe M, Baba H. Clinical implications of the LINE-1 methylation levels in patients with gastrointestinal cancer. Surg Today 2014;44:1807-16.

75. Sheaffer KL, Elliott EN, Kaestner KH. DNA Hypomethylation Contributes to Genomic Instability and Intestinal Cancer Initiation. Cancer Prev Res (Phila) 2016;9:534-46.

76. Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, Kuchiba A, Kim SA, Mima K, Sukawa Y, Jung S, Zhang X, Wu K, et al. Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. J Natl Cancer Inst 2014;106.

77. Goel A, Xicola RM, Nguyen TP, Doyle BJ, Sohn VR, Bandipalliam P, Rozek LS, Reyes J, Cordero C, Balaguer F, Castells A, Jover R, et al. Aberrant DNA methylation in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. Gastroenterology 2010;138:1854-62.

78. Ogino S, Nishihara R, Lochhead P, Imamura Y, Kuchiba A, Morikawa T, Yamauchi M, Liao X, Qian ZR, Sun R, Sato K, Kirkner GJ, et al. Prospective study of family history and colorectal cancer risk by tumor LINE-1 methylation level. J Natl Cancer Inst 2013;105:130-40.

79. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res 2005;15:1073-8.

80. Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic Acids Res 2006;34:1512-21.

81. Criscione SW, Theodosakis N, Micevic G, Cornish TC, Burns KH, Neretti N, Rodic N. Genome-wide characterization of human L1 antisense promoter-driven transcripts. BMC Genomics 2016;17:463.

82. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet 2010;6:e1000917.

83. Weber B, Kimhi S, Howard G, Eden A, Lyko F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene 2010;29:5775-84.

84. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, Goel A. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut 2014;63:635-46.

85. Zhu C, Utsunomiya T, Ikemoto T, Yamada S, Morine Y, Imura S, Arakawa Y, Takasu C, Ishikawa D, Imoto I, Shimada M. Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. Ann Surg Oncol 2014;21 Suppl 4:S729-35.

86. Nones K, Waddell N, Song S, Patch AM, Miller D, Johns A, Wu J, Kassahn KS, Wood D, Bailey P, Fink L, Manning S, et al. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. Int J Cancer 2014;135:1110-8.

87. Zhang J, Babic A. Regulation of the MET oncogene: molecular mechanisms. Carcinogenesis 2016;37:345-55.

88. Liou GY, Storz P. Reactive oxygen species in cancer. Free Radic Res 2010;44:479-96.

89. Franco R, Schoneveld O, Georgakilas AG, Panayiotidis MI. Oxidative stress, DNA methylation and carcinogenesis. Cancer Lett 2008;266:6-11.

90. Wu Q, Ni X. ROS-mediated DNA methylation pattern alterations in carcinogenesis. Curr Drug Targets 2015;16:13-9.

91. Weitzman SA, Turk PW, Milkowski DH, Kozlowski K. Free radical adducts induce alterations in DNA cytosine methylation. Proc Natl Acad Sci U S A 1994;91:1261-4.

92. Valko M, Rhodes CJ, Moncol J, Izakovic M, Mazur M. Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chem Biol Interact 2006;160:1-40.

93. Turk PW, Laayoun A, Smith SS, Weitzman SA. DNA adduct 8-hydroxyl-2'-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase. Carcinogenesis 1995;16:1253-5.

94. Miller JW, Nadeau MR, Smith J, Smith D, Selhub J. Folate-deficiency-induced homocysteinaemia in rats: disruption of S-adenosylmethionine's co-ordinate regulation of homocysteine metabolism. Biochem J 1994;298 (Pt 2):415-9.

95. Kuchino Y, Mori F, Kasai H, Inoue H, Iwai S, Miura K, Ohtsuka E, Nishimura S. Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent residues. Nature 1987;327:77-9.

96. Pfeifer GP. Mutagenesis at methylated CpG sequences. Curr Top Microbiol Immunol 2006;301:259-81.

97. Kasymov RD, Grin IR, Endutkin AV, Smirnov SL, Ishchenko AA, Saparbaev MK, Zharkov DO. Excision of 8-oxoguanine from methylated CpG dinucleotides by human 8-oxoguanine DNA glycosylase. FEBS Lett 2013;587:3129-34.

98. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 2007;67:946-50.

99. Bhutani N, Burns DM, Blau HM. DNA demethylation dynamics. Cell 2011;146:866-72.

100. Iyama T, Wilson DM, 3rd. DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) 2013;12:620-36.

101. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999;21:163-7.

102. Miyakura Y, Sugano K, Akasu T, Yoshida T, Maekawa M, Saitoh S, Sasaki H, Nomizu T, Konishi F, Fujita S, Moriya Y, Nagai H. Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. Clin Gastroenterol Hepatol 2004;2:147-56.

103. Fang M, Ou J, Hutchinson L, Green MR. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. Mol Cell 2014;55:904-15.

104. Kim IJ, Kang HC, Park JH, Shin Y, Ku JL, Lim SB, Park SY, Jung SY, Kim HK, Park JG. Development and applications of a beta-catenin oligonucleotide microarray: beta-catenin mutations are dominantly found in the proximal colon cancers with microsatellite instability. Clin Cancer Res 2003;9:2920-5.

105. Fernandez-Peralta AM, Nejda N, Oliart S, Medina V, Azcoita MM, Gonzalez-Aguilera JJ. Significance of mutations in TGFBR2 and BAX in neoplastic progression and patient outcome in sporadic colorectal tumors with high-frequency microsatellite instability. Cancer Genet Cytogenet 2005;157:18-24.

106. Goldstein NS. Small colonic microsatellite unstable adenocarcinomas and high-grade epithelial dysplasias in sessile serrated adenoma polypectomy specimens: a study of eight cases. Am J Clin Pathol 2006;125:132-45.

107. Soreide K, Janssen EA, Soiland H, Korner H, Baak JP. Microsatellite instability in colorectal cancer. Br J Surg 2006;93:395-406.

108. des Guetz G, Mariani P, Cucherousset J, Benamoun M, Lagorce C, Sastre X, Le Toumelin P, Uzzan B, Perret GY, Morere JF, Breau JL, Fagard R, et al. Microsatellite instability and sensitivitiy to FOLFOX treatment in metastatic colorectal cancer. Anticancer Res 2007;27:2715-9.

109. Tougeron D, Fauquembergue E, Rouquette A, Le Pessot F, Sesboue R, Laurent M, Berthet P, Mauillon J, Di Fiore F, Sabourin JC, Michel P, Tosi M, et al. Tumor-infiltrating lymphocytes in colorectal cancers with microsatellite instability are correlated with the number and spectrum of frameshift mutations. Mod Pathol 2009;22:1186-95.

110. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-20.

111. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 1999;96:8681-6.

112. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. Gastroenterology 2005;129:837-45.

113. Nazemalhosseini Mojarad E, Kuppen PJ, Aghdaei HA, Zali MR. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterol Hepatol Bed Bench 2013;6:120-8.

114. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002;3:415-28.

115. van Engeland M, Derks S, Smits KM, Meijer GA, Herman JG. Colorectal cancer epigenetics: complex simplicity. J Clin Oncol 2011;29:1382-91.

116. Ogino S, Odze RD, Kawasaki T, Brahmandam M, Kirkner GJ, Laird PW, Loda M, Fuchs CS. Correlation of pathologic features with CpG island methylator phenotype (CIMP) by quantitative DNA methylation analysis in colorectal carcinoma. Am J Surg Pathol 2006;30:1175-83.

117. Patai AV, Molnar B, Tulassay Z, Sipos F. Serrated pathway: alternative route to colorectal cancer. World J Gastroenterol 2013;19:607-15.

118. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.

119. Armelao F, de Pretis G. Familial colorectal cancer: a review. World J Gastroenterol 2014;20:9292-8.

120. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet 2002;30:227-32.

121. Michaels ML, Miller JH. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). J Bacteriol 1992;174:6321-5.

122. Boiteux S, Radicella JP. The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. Arch Biochem Biophys 2000;377:1-8.

123. Ohtsubo T, Nishioka K, Imaiso Y, Iwai S, Shimokawa H, Oda H, Fujiwara T, Nakabeppu Y. Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria. Nucleic Acids Res 2000;28:1355-64.

124. Nielsen M, Joerink-van de Beld MC, Jones N, Vogt S, Tops CM, Vasen HF, Sampson JR, Aretz S, Hes FJ. Analysis of MUTYH genotypes and colorectal phenotypes in patients With MUTYH-associated polyposis. Gastroenterology 2009;136:471-6.

125. Nieuwenhuis MH, Vogt S, Jones N, Nielsen M, Hes FJ, Sampson JR, Aretz S, Vasen HF. Evidence for accelerated colorectal adenoma--carcinoma progression in MUTYH-associated polyposis? Gut 2012;61:734-8.

126. Cardoso J, Molenaar L, de Menezes RX, van Leerdam M, Rosenberg C, Moslein G, Sampson J, Morreau H, Boer JM, Fodde R. Chromosomal instability in MYH- and APC-mutant adenomatous polyps. Cancer Res 2006;66:2514-9.

127. Venesio T, Balsamo A, Errichiello E, Ranzani GN, Risio M. Oxidative DNA damage drives carcinogenesis in MUTYH-associated-polyposis by specific mutations of mitochondrial and MAPK genes. Mod Pathol 2013;26:1371-81.

128. Vogt S, Jones N, Christian D, Engel C, Nielsen M, Kaufmann A, Steinke V, Vasen HF, Propping P, Sampson JR, Hes FJ, Aretz S. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 2009;137:1976-85 e1-10.

129. Nielsen M, Morreau H, Vasen HF, Hes FJ. MUTYH-associated polyposis (MAP). Crit Rev Oncol Hematol 2011;79:1-16.

130. Nakamura Y, Nishisho I, Kinzler KW, Vogelstein B, Miyoshi Y, Miki Y, Ando H, Horii A, Nagase H. Mutations of the adenomatous polyposis coli gene in familial polyposis coli patients and sporadic colorectal tumors. Princess Takamatsu Symp 1991;22:285-92.

131. Trimbath JD, Giardiello FM. Review article: genetic testing and counselling for hereditary colorectal cancer. Aliment Pharmacol Ther 2002;16:1843-57.

132. Strate LL, Syngal S. Hereditary colorectal cancer syndromes. Cancer Causes Control 2005;16:201-13.

133. Miyoshi Y, Nagase H, Ando H, Horii A, Ichii S, Nakatsuru S, Aoki T, Miki Y, Mori T, Nakamura Y. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet 1992;1:229-33.

134. Knudsen AL, Bulow S, Tomlinson I, Moslein G, Heinimann K, Christensen IJ. Attenuated familial adenomatous polyposis: results from an international collaborative study. Colorectal Dis 2010;12:e243-9.

135. Galle TS, Juel K, Bulow S. Causes of death in familial adenomatous polyposis. Scand J Gastroenterol 1999;34:808-12.

136. Chen CS, Phillips KD, Grist S, Bennet G, Craig JE, Muecke JS, Suthers GK. Congenital hypertrophy of the retinal pigment epithelium (CHRPE) in familial colorectal cancer. Fam Cancer 2006;5:397-404.

137. Gallagher MC, Phillips RK, Bulow S. Surveillance and management of upper gastrointestinal disease in Familial Adenomatous Polyposis. Fam Cancer 2006;5:263-73.

138. Peltomaki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10:735-40.

139. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 2009;76:1-18.

140. Vasen HF. Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). Aliment Pharmacol Ther 2007;26 Suppl 2:113-26.

141. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I, Capella G, den Dunnen JT, du Sart D, Fabre A, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014;46:107-15.

142. Peltomaki P. Epigenetic mechanisms in the pathogenesis of Lynch syndrome. Clin Genet 2014;85:403-12.

143. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261-8.

144. Weissman SM, Bellcross C, Bittner CC, Freivogel ME, Haidle JL, Kaurah P, Leininger A, Palaniappan S, Steenblock K, Vu TM, Daniels MS. Genetic counseling considerations in the evaluation of families for Lynch syndrome--a review. J Genet Couns 2011;20:5-19.

145. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014;42:D1001-6.

146. Yates CM, Sternberg MJ. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J Mol Biol 2013;425:3949-63.

147. Timofeeva MN, Kinnersley B, Farrington SM, Whiffin N, Palles C, Svinti V, Lloyd A, Gorman M, Ooi LY, Hosking F, Barclay E, Zgaga L, et al. Recurrent Coding Sequence Variation Explains Only A Small Fraction of the Genetic Architecture of Colorectal Cancer. Sci Rep 2015;5:16286.

148. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, Trembath RC. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. Hum Mol Genet 2004;13:2361-8.

149. Spisak S, Lawrenson K, Fu Y, Csabai I, Cottman RT, Seo JH, Haiman C, Han Y, Lenci R, Li Q, Tisza V, Szallasi Z, et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat Med 2015;21:1357-63.

150. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. Nucleic Acids Res 2016;44:8641-54.

151. Zhi D, Aslibekyan S, Irvin MR, Claas SA, Borecki IB, Ordovas JM, Absher DM, Arnett DK. SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics 2013;8:802-6.

152. Harper AR, Nayee S, Topol EJ. Protective alleles and modifier variants in human health and disease. Nat Rev Genet 2015;16:689-701.

153. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2016;45:D896-D901.

154. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res 1997;57:808-11.

155. Deng G, Chen A, Hong J, Chae HS, Kim YS. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res 1999;59:2029-33.

156. Ito E, Yanagisawa Y, Iwahashi Y, Suzuki Y, Nagasaki H, Akiyama Y, Sugano S, Yuasa Y, Maruyama K. A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. Biochem Biophys Res Commun 1999;256:488-94.

157. Raptis S, Mrkonjic M, Green RC, Pethe VV, Monga N, Chan YM, Daftary D, Dicks E, Younghusband BH, Parfrey PS, Gallinger SS, McLaughlin JR, et al. MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. J Natl Cancer Inst 2007;99:463-74.

158. Allan JM, Shorto J, Adlard J, Bury J, Coggins R, George R, Katory M, Quirke P, Richman S, Scott D, Scott K, Seymour M, et al. MLH1 -93G>A promoter polymorphism and risk of mismatch repair deficient colorectal cancer. Int J Cancer 2008;123:2456-9.

159. Campbell PT, Curtin K, Ulrich CM, Samowitz WS, Bigler J, Velicer CM, Caan B, Potter JD, Slattery ML. Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. Gut 2009;58:661-7.

160. Mrkonjic M, Roslin NM, Greenwood CM, Raptis S, Pollett A, Laird PW, Pethe VV, Chiang T, Daftary D, Dicks E, Thibodeau SN, Gallinger S, et al. Specific variants in the MLH1 gene region may drive DNA methylation, loss of protein expression, and MSI-H colorectal cancer. PLoS One 2010;5:e13314.

161. Whiffin N, Broderick P, Lubbe SJ, Pittman AM, Penegar S, Chandler I, Houlston RS. MLH1-93G > A is a risk factor for MSI colorectal cancer. Carcinogenesis 2011;32:1157-61.

162. Mei M, Liu D, Dong S, Ingvarsson S, Goodfellow PJ, Chen H. The MLH1 -93 promoter variant influences gene expression. Cancer Epidemiol 2010;34:93-5.

163. Perera S, Mrkonjic M, Rawson JB, Bapat B. Functional effects of the MLH1-93G>A polymorphism on MLH1/EPM2AIP1 promoter activity. Oncol Rep 2011;25:809-15.

164. Liu NQ, Ter Huurne M, Nguyen LN, Peng T, Wang SY, Studd JB, Joshi O, Ongen H, Bramsen JB, Yan J, Andersen CL, Taipale J, et al. The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression. Nat Commun 2017;8:14418.

165. Savio AJ, Mrkonjic M, Lemire M, Gallinger S, Knight JA, Bapat B. The dynamic DNA methylation landscape of the mutL homolog 1 shore is altered by MLH1-93G>A polymorphism in normal tissues and colorectal cancer. Clin Epigenetics 2017;9:26.

166. Savio AJ, Bapat B. Modulation of transcription factor binding and epigenetic regulation of the MLH1 CpG island and shore by polymorphism rs1800734 in colorectal cancer. Epigenetics 2017;12:441-8.

167. Hamilton SR BF BP, Ilyas M, Morreau H, Nakamura S I, Quirke P, Riboli E, Sobin LH. WHO Classification of Tumours of the Digestive System, Chapter 8 Tumours of the Colon and Rectum. IARC Press 2010.

168. Stefanoli M, La Rosa S, Sahnane N, Romualdi C, Pastorino R, Marando A, Capella C, Sessa F, Furlan D. Prognostic relevance of aberrant DNA methylation in g1 and g2 pancreatic neuroendocrine tumors. Neuroendocrinology 2014;100:26-34.

169. Sahnane N, Magnoli F, Bernasconi B, Tibiletti MG, Romualdi C, Pedroni M, Ponz de Leon M, Magnani G, Reggiani-Bonetti L, Bertario L, Signoroni S, Capella C, et al. Aberrant DNA methylation profiles of inherited and sporadic colorectal cancer. Clin Epigenetics 2015;7:131.

170. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.

171. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-9.

172. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

173. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 2011;12:R72.

174. Cruickshanks HA, Vafadar-Isfahani N, Dunican DS, Lee A, Sproul D, Lund JN, Meehan RR, Tufarelli C. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. Nucleic Acids Res 2013;41:6857-69.

175. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MC, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, et al. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell 2015;163:583-93.

176. Miglio U, Berrino E, Panero M, Ferrero G, Coscujuela Tarrero L, Miano V, Dell'Aglio C, Sarotto I, Annaratone L, Marchio C, Comoglio PM, De Bortoli M, et al. The expression of LINE1-MET chimeric transcript identifies a subgroup of aggressive breast cancers. Int J Cancer 2018.

177. Mossman D, Kim KT, Scott RJ. Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists. BMC Cancer 2010;10:366.

178. Rashid M, Fischer A, Wilson CH, Tiffen J, Rust AG, Stevens P, Idziaszczyk S, Maynard J, Williams GT, Mustonen V, Sampson JR, Adams DJ. Adenoma development in familial adenomatous polyposis and MUTYH-associated polyposis: somatic landscape and driver genes. J Pathol 2016;238:98-108.

179. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, Policicchio E, Urso ED, Agostini M, Genuardi M, Lucci-Cordisco E, Venesio T, et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine 2017;20:39-49.

180. van Puijenbroek M, Nielsen M, Tops CM, Halfwerk H, Vasen HF, Weiss MM, van Wezel T, Hes FJ, Morreau H. Identification of patients with (atypical) MUTYH-associated polyposis by KRAS2 c.34G > T prescreening followed by MUTYH hotspot analysis in formalin-fixed paraffin-embedded tissue. Clin Cancer Res 2008;14:139-42.

181. Aime A, Coulet F, Lefevre JH, Colas C, Cervera P, Flejou JF, Lascols O, Soubrier F, Parc Y. Somatic c.34G>T KRAS mutation: a new prescreening test for MUTYH-associated polyposis? Cancer Genet 2015;208:390-5.

182. Haigis KM, Kendall KR, Wang Y, Cheung A, Haigis MC, Glickman JN, Niwa-Kawakita M, Sweet-Cordero A, Sebolt-Leopold J, Shannon KM, Settleman J, Giovannini M, et al. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nat Genet 2008;40:600-8.

183. Vagaja NN, Parry J, McCallum D, Thomas MA, Bentel JM. Are all RAS mutations the same? Coexisting KRAS and NRAS mutations in a caecal adenocarcinoma and contiguous tubulovillous adenoma. J Clin Pathol 2015;68:657-60.

184. Kinkade CW, Castillo-Martin M, Puzio-Kuter A, Yan J, Foster TH, Gao H, Sun Y, Ouyang X, Gerald WL, Cordon-Cardo C, Abate-Shen C. Targeting AKT/mTOR and ERK MAPK signaling inhibits hormone-refractory prostate cancer in a preclinical mouse model. J Clin Invest 2008;118:3051-64.

185. Engelman JA, Chen L, Tan X, Crosby K, Guimaraes AR, Upadhyay R, Maira M, McNamara K, Perera SA, Song Y, Chirieac LR, Kaur R, et al. Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. Nat Med 2008;14:1351-6.

186. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 2013;503:548-51.

187. Patricelli MP, Janes MR, Li LS, Hansen R, Peters U, Kessler LV, Chen Y, Kucharski JM, Feng J, Ely T, Chen JH, Firdaus SJ, et al. Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. Cancer Discov 2016;6:316-29.

188. Zauber P, Marotta S, Sabbath-Solitare M. KRAS gene mutations are more common in colorectal villous adenomas and in situ carcinomas than in carcinomas. Int J Mol Epidemiol Genet 2013;4:1-10.

189. Juarez M, Egoavil C, Rodriguez-Soler M, Hernandez-Illan E, Guarinos C, Garcia-Martinez A, Alenda C, Giner-Calabuig M, Murcia O, Mangas C, Paya A, Aparicio JR, et al. KRAS and BRAF somatic mutations in colonic polyps and the risk of metachronous neoplasia. PLoS One 2017;12:e0184937.

190. Chen H, Lefferts JA, Schwab MC, Suriawinata AA, Tsongalis GJ. Correlation of polypoid colorectal adenocarcinoma with pre-existing adenomatous polyps and KRAS mutation. Cancer Genet 2011;204:245-51.

191. Yadamsuren EA, Nagy S, Pajor L, Lacza A, Bogner B. Characteristics of advanced- and non advanced sporadic polypoid colorectal adenomas: correlation to KRAS mutations. Pathol Oncol Res 2012;18:1077-84.

192. Yamane LS, Scapulatempo-Neto C, Alvarenga L, Oliveira CZ, Berardinelli GN, Almodova E, Cunha TR, Fava G, Colaiacovo W, Melani A, Fregnani JH, Reis RM, et al. KRAS and BRAF mutations and MSI status in precursor lesions of colorectal cancer detected by colonoscopy. Oncol Rep 2014;32:1419-26.

193. Cox AD, Fesik SW, Kimmelman AC, Luo J, Der CJ. Drugging the undruggable RAS: Mission possible? Nat Rev Drug Discov 2014;13:828-51.

194. Temko D, Tomlinson IPM, Severini S, Schuster-Bockler B, Graham TA. The effects of mutational processes and selection on driver mutations across cancer types. Nat Commun 2018;9:1857.

195. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, Voravud N, Sriuranpong V, Mutirangura A. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. Oncogene 2004;23:8841-6.

196. Suter CM, Martin DI, Ward RL. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis 2004;19:95-101.

197. Belshaw NJ, Pal N, Tapp HS, Dainty JR, Lewis MP, Williams MR, Lund EK, Johnson IT. Patterns of DNA methylation in individual colonic crypts reveal aging and cancer-related field defects in the morphologically normal mucosa. Carcinogenesis 2010;31:1158-63.

198. Kamiyama H, Suzuki K, Maeda T, Koizumi K, Miyaki Y, Okada S, Kawamura YJ, Samuelsson JK, Alonso S, Konishi F, Perucho M. DNA demethylation in normal colon tissue predicts predisposition to multiple cancers. Oncogene 2012;31:5029-37.

199. Jiang AC, Buckingham L, Barbanera W, Korang AY, Bishesari F, Melson J. LINE-1 is preferentially hypomethylated within adenomatous polyps in the presence of synchronous colorectal cancer. Clin Epigenetics 2017;9:25.

200. Kloypan C, Srisa-art M, Mutirangura A, Boonla C. LINE-1 hypomethylation induced by reactive oxygen species is mediated via depletion of S-adenosylmethionine. Cell Biochem Funct 2015;33:375-85.

201. Campos AC, Molognoni F, Melo FH, Galdieri LC, Carneiro CR, D'Almeida V, Correa M, Jasiulionis MG. Oxidative stress modulates DNA methylation during melanocyte anchorage blockade associated with malignant transformation. Neoplasia 2007;9:1111-21.

202. Lim SO, Gu JM, Kim MS, Kim HS, Park YN, Park CK, Cho JW, Park YM, Jung G. Epigenetic changes induced by reactive oxygen species in hepatocellular carcinoma: methylation of the E-cadherin promoter. Gastroenterology 2008;135:2128-40, 40 e1-8.

203. Ziech D, Franco R, Pappa A, Panayiotidis MI. Reactive oxygen species (ROS)--induced genetic and epigenetic alterations in human carcinogenesis. Mutat Res 2011;711:167-73.

204. Quan X, Lim SO, Jung G. Reactive oxygen species downregulate catalase expression via methylation of a CpG island in the Oct-1 promoter. FEBS Lett 2011;585:3436-41.

205. Iborra M, Moret I, Rausell F, Bastida G, Aguas M, Cerrillo E, Nos P, Beltran B. Role of oxidative stress and antioxidant enzymes in Crohn's disease. Biochem Soc Trans 2011;39:1102-6.

206. Hartnett L, Egan LJ. Inflammation, DNA methylation and colitis-associated cancer. Carcinogenesis 2012;33:723-31.

207. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6:963-8.

208. Patel A, Tripathi G, Gopalakrishnan K, Williams N, Arasaradnam RP. Field cancerisation in colorectal cancer: a new frontier or pastures past? World J Gastroenterol 2015;21:3763-72.

209. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, Fuchs CS. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. Int J Cancer 2008;122:2767-73.

210. Igarashi S, Suzuki H, Niinuma T, Shimizu H, Nojima M, Iwaki H, Nobuoka T, Nishida T, Miyazaki Y, Takamaru H, Yamamoto E, Yamamoto H, et al. A novel correlation between LINE-1 hypomethylation and the malignancy of gastrointestinal stromal tumors. Clin Cancer Res 2010;16:5114-23.

211. Sunami E, de Maat M, Vu A, Turner RR, Hoon DS. LINE-1 hypomethylation during primary colon cancer progression. PLoS One 2011;6:e18884.

212. Yamada Y, Jackson-Grusby L, Linhart H, Meissner A, Eden A, Lin H, Jaenisch R. Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. Proc Natl Acad Sci U S A 2005;102:13580-5.

213. Ewing AD, Gacita A, Wood LD, Ma F, Xing D, Kim MS, Manda SS, Abril G, Pereira G, Makohon-Moore A, Looijenga LH, Gillis AJ, et al. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res 2015;25:1536-45.

214. Burn J, Gerdes AM, Macrae F, Mecklin JP, Moeslein G, Olschwang S, Eccles D, Evans DG, Maher ER, Bertario L, Bisgaard ML, Dunlop MG, et al. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. Lancet 2011;378:2081-7.

215. Biancolella M, Fortini BK, Tring S, Plummer SJ, Mendoza-Fandino GA, Hartiala J, Hitchler MJ, Yan C, Schumacher FR, Conti DV, Edlund CK, Noushmehr H, et al. Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23.1. Hum Mol Genet 2014;23:2198-209.

216. Lo YL, Hsiao CF, Jou YS, Chang GC, Tsai YH, Su WC, Chen KY, Chen YM, Huang MS, Hsieh WS, Chen CJ, Hsiung CA. Polymorphisms of MLH1 and MSH2 genes and the risk of lung cancer among never smokers. Lung Cancer 2011;72:280-6.

217. Rodriguez-Hernandez I, Perdomo S, Santos-Briz A, Garcia JL, Gomez-Moreta JA, Cruz JJ, Gonzalez-Sarmiento R. Analysis of DNA repair gene polymorphisms in glioblastoma. Gene 2014;536:79-83.

218. Niu L, Li S, Liang H, Li H. The hMLH1 -93G>A Polymorphism and Risk of Ovarian Cancer in the Chinese Population. PLoS One 2015;10:e0135822.

219. Zhu H, Li X, Zhang X, Chen D, Li D, Ren J, Gu H, Shu Y, Wang D. Polymorphisms in mismatch repair genes are associated with risk and microsatellite instability of gastric cancer, and interact with life exposures. Gene 2016;579:52-7.

220. Grossman SR, Engreitz J, Ray JP, Nguyen TH, Hacohen N, Lander ES. Positional specificity of different transcription factor classes within enhancers. Proc Natl Acad Sci U S A 2018;115:E7222-E30.

221. Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, Mann M. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet 2012;8:e1002982.

222. Matsushita M, Takeuchi S, Yang Y, Yoshino N, Tsukasaki K, Taguchi H, Koeffler HP, Seo H. Methylation of the MLH1 gene in hematological malignancies. Oncol Rep 2005;14:191-4.

223. Zighelboim I, Goodfellow PJ, Gao F, Gibb RK, Powell MA, Rader JS, Mutch DG. Microsatellite instability and epigenetic inactivation of MLH1 and outcome of patients with endometrial carcinomas of the endometrioid type. J Clin Oncol 2007;25:2042-8.

224. Moura Lima E, Ferreira Leal M, Cardoso Smith Mde A, Rodriguez Burbano R, Pimentel de Assumpcao P, Bello MJ, Rey JA, Ferreira de Lima F, Casartelli C. DNA mismatch repair gene methylation in gastric cancer in individuals from northern Brazil. Biocell 2008;32:237-43.

225. Gomes A, Reis-Silva M, Alarcao A, Couceiro P, Sousa V, Carvalho L. Promoter hypermethylation of DNA repair genes MLH1 and MSH2 in adenocarcinomas and squamous cell carcinomas of the lung. Rev Port Pneumol 2014;20:20-30.

226. Wojtczyk-Miaskowska A, Presler M, Michajlowski J, Matuszewski M, Schlichtholz B. Gene Expression, DNA Methylation and Prognostic Significance of DNA Repair Genes in Human Bladder Cancer. Cell Physiol Biochem 2017;42:2404-17.

227. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. Editing DNA Methylation in the Mammalian Genome. Cell 2016;167:233-47 e17.

ACADEMIC RESULTS ACHIEVED DURING THE PhD TRAINING

PUBBLICATIONS

- Furlan D, **Trapani D**, Berrino E, Debernardi C, Panero M, Libera L, Sahnane N, Riva C, Tibiletti MG, Sessa F, Sapino A, Venesio T. Oxidative DNA damage induces hypomethylation in a compromised base excision repair colorectal tumourigenesis. Br J Cancer. 2017 Jan 31.

- Vanoli A, Di Sabatino A, Furlan D, Klersy C, Grillo F, Fiocca R, Mescoli C, Rugge M, Nesi G, Fociani P, Sampietro G, Ardizzione S, Luinetti O, Calabrò A, Tonelli F, Volta U, Santini D, Caio G, Giuffrida P, Elli L, Ferrero S, Latella G, Ciardi A, Caronna R, Solina G, Rizzo A, Ciacci C, D'Armiento FP, Salemme M, Villanacci V, Cannizzaro R, Canzoneri V, Reggiani Bonetti L, Biancone L, Monteleone G, Orlandi A, Santeusanio G, Macciomei MC, D'Incà R, Perfetti V, Sandri G, Silano M, Florena AM, Giannone AG, Papi C, Coppola L, Usai P, Maccioni A, Astegiano M, Migliora P, Manca R, Martino M, **Trapani D**, Cerutti R, Alberizzi P, Riboni R, Sessa F, Paulli M, Solcia E, Corazza GR. Small Bowel Carcinomas in Coeliac or Crohn's Disease: Clinico-pathological, Molecular, and Prognostic Features. A Study From the Small Bowel Cancer Italian Consortium. Journal of Crohn's and Colitis, 2017, 1–12.

- Thomas R*, T**rapani D***, Goodyer-Sait L, Tomkova M, Fernandez-Rozadilla C, Sahnane N, Woolley C, Davide H, Leedham S, Palles C, Furlan D, Tomlinson I, Lewis A. The polymorphic variant rs1800734 influences methylation acquisition and allele-specific TFAP4 binding in the MLH1 promoter leading to differential mRNA expression. Paper submitted.

WORK EXPERIENCE

- May 30-June 10 2016. EPIGEN Training Program at Lab of Molecular and Genomics Medicine in Baronissi (SA). I had the opportunity to see the preparation and sequencing of libraries from DNA and RNA sources through several high-throughput platforms, such as Illumina HiSeq2500, NextSeq 500 and MiSeq and Life technologies Ion Torrent PGM.

- February-November 2017. Training abroad as Ph.D. Visiting Student at Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom.

Project title: Understanding how an *MLH1* promoter polymorphism predisposes to gene silencing and colorectal cancer.

ABSTRACTS

- Berrino E, Panero M, **Trapani D**, Libera L, Sahnane N, Tibiletti MG, Debernardi C, Furlan D, Venesio T. "MAPK pathway mutations and DNA hypomethylation are both involved in the onset of MUTYH associated colorectal carcinogenesis" XIII AIFEG Conference, Naples October 2015.

- **Trapani D**, Furlan D, Berrino E, Debernardi C, Panero M, Libera L, Sahnane N, Riva C, Tibiletti MG, Venesio T. "MAPK pathway mutations and DNA hypomethylation are involved in early step of oxidative DNA damage driven colorectal carcinogenesis". 28°Symposium Pezcoller, Initial Steps on the Route to Tumorigenesis, Trento 20-21 June 2016.

- Rinaldi A, Rizzo F, Sellitto A, Memoli D, **Trapani D**, Sahnane N, Furlan D, Giurato G, Weisz A. "Involvement of PIWI-interacting RNAs in different molecular and clinicopathological subclasses of colon cancer". 14th Annual Meeting of the Bioinformatics Italian Society. Cagliari 5-8 July 2017.

ORAL PRESENTATION

EPIGEN Workshop. "Dna global demethylation and specific somatic mutation patterns as potential biomarkers of risk for colorectal neoplasm". Rimini 5 May 2016.

CONFERENCE, COURSE AND SEMINARY

- Course of "BIOSTATISTICA" UNINSUBRIA, Varese 9-10-16-17-23-24 March 2015

- Course "Targeting Lung Cancer" Ospedale di Circolo Fondazione Macchi, Varese 30 March 2015

- Seminary "Watching at the D side: D-amino acids in physiology and pathology" UNINSUBRIA, Varese 15 April 2015

- Course "Focus on clinical: Next Generation Sequencing and Digital PCR" organized by ThermoFisher Scientific. Milan 13 October 2015

- Course of "BIOINFORMATICS" UNINSUBRIA, Varese September-December 2015

- Seminar “DNA methylation and demethylation in cancer” organized by EPIGEN. Milan 11 December 2015
- Seminary “PhD thesis: General and regulatory aspect”, Varese 29 February 2016
- Course “Droplet Digital PCR” organized by Bio-Rad. Milan 18 June 2016
- Online course “WEBINAR: MEET.IMMUNO PATHOLOGY IN CLINICAL PRACTISE 2016, i processi immunologici alla base dello sviluppo dei tumori”, 12 October 2016
- Course of “Animal Experimentation” UNINSUBRIA, Varese 3 December 2016
- Course “Exploring the human proteome in health and disease using nexprot and semantic web technologies” UNINSUBRIA, Varese 29 January 2018
- Practical course at Anatomic Pathology Unit about the application of NGS technology in cancer research organized by DIATECH SRL. Varese 6-9 March 2018
- Course “Metodi Alternativi e Scienza: Approcci Integrati” organized by University of Insubria, Varese 20 Septmber 2018
- Course “Seconda Giornata Giovanni Tosi-Recenti acquisizioni nel campo dell’Oncologia e Malattie Infettive” UNINSUBRIA, Varese 27 September 2018
- Course “News in Oncolgy: Focus on Gastrointestinal Tumors” organized by Fondazione IRCCCS Istituto Nazionale dei Tumori. Milan, 26 October 2018
- I° International meeting “Present and future in Forensic Genetics” organized by Ordine Nazionale dei Biologi and Arma dei Carabinieri at Italian Minister of Health. Rome, 30 November and 1 December 2018

ACKNOWLEDGMENT

Varese, Italy

Unit of Anatomic Pathology:

Prof. Fausto Sessa

Prof. Cristina Riva

Dr. Anna Maria Chiaravalli

Molecular Biology Lab

Prof. Daniela Furlan

Dr. Nora Sahnane

Dr. Roberta Cerutti

Dr. Laura Libera

Dr. Chiara Alberni

Dr. Lucio Lettig

Dr. Daniele Sabatino

Giuseppe Borgia

Oncological Genetic Counseling Service

Dr. Maria Grazia Tibiletti

Dr. Ileana Carnevali

Candiolo (Torino), Italy

Candiolo Cancer Institute, FPO-IRCCS

Prof. Anna Sapino

Dr Tiziana Venesio

Dr. Enrico Berrino

Dr. Carla Debernardi

Dr. Para Panero

Dr. Umberto Miglio

Oxford, United Kingdom

Wellcome Trust Centre for Human Genetics

Dr. Annabelle Lewis

Dr. Rachel Thomas

Dr. Lily Goodyer-Sait

Dr. Hayley Davis

Dr Simon Leedham

Ludwig Institute for Cancer Research Ltd

Dr. Marketa Tomkova

Birmingham, United Kingdom

Institute of Cancer and Genomic Sciences

Prof. Ian Tomlinson

Dr. Claire Pelles

Dr. Connor Woolley

Dr. Laura Chegwidden