Prognostic and predictive DNA methylated biomarkers in colorectal cancer in the era of emerging technologies

**Metilazione del DNA come biomarcatore prognostico e predittivo nelle neoplasie colon-rettali nel contesto delle tecnologie emergenti**

**Emanuela Massi**, **Valentina Summa**<sup>1,2</sup>, **Mariangela De Robertis**<sup>1</sup>, **Antonella Romanelli**<sup>3</sup>, **Francesco Picardo**<sup>1</sup>, **Manuela Costantini**<sup>1</sup>, **Luisa Loiacono**<sup>2</sup>, **Maria Luana Poeta**<sup>1,3</sup>

<sup>1</sup> Laboratory for Molecular Medicine and Biotechnology; CIR, University Campus Bio-Medico of Rome (Rome), Italy; <sup>2</sup> Laboratory of Oncology, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, FG (Italy); <sup>3</sup> Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari (Italy)

Colorectal cancer (CRC) is the third most common cancer worldwide, and despite the care improvement of the last years it represents the fourth most common cancer-related cause of death globally. CRC results from the progressive accumulation of genetic and epigenetic alterations. Epigenetics refers to heritable and reversible changes associated with gene expression without alterations in the DNA sequence. Epigenetic modifications include aberrant DNA methylation, histone modifications, non-coding RNA deregulation, chromatin remodeling and nucleosome positioning. Over the past few decades an increasing number of high-throughput platforms for genome-wide methylation studies has been developed favoring the identification of novel epigenetic targets. These epigenetic alterations have been noted to play crucial roles not only in cancer progression, but also in cancer initiation. In this perspective aberrantly methylated genes have indeed attracted considerable interest as potential biomarkers for the early detection of disease onset, prognosis and choice of treatment, and the monitoring of disease after therapy. Despite the rapidly evolving of sequencing technologies is giving an unprecedented opportunity for identification of cancer biomarkers, the large plethora of generated data not always provide new tools which can be efficaciously translated in clinical setting. The progress in diagnostics and therapeutics biomarker studies should not be conceived only in a deep pathway understanding and single target identification but it should encompasses a more comprehensive approach which aims to develop a much-needed framework integrating laboratory science, bioinformatics, histopathological and clinical expertise.

**Key words:** Methylation, Biomarkers, Colorectal Cancer
Le neoplasie colon-rettali (CRC) rappresentano complessivamente la quarta più comune causa di morte per neoplasia nonostante negli ultimi anni siano stati compiuti notevoli progressi sia in ambito diagnostico che terapeutico. I tumori del colon-retto derivano da un accumulo progressivo di alterazioni genetiche ed epigenetiche. Le modificazioni epigenetiche consistono in cambiamenti ereditari e reversibili associati a variazioni dell’espressione genica senza alterazioni nella sequenza del DNA. Le modificazioni epigenetiche includono metilazione del DNA, modificazioni post-traduzionali degli istoni, disregolazione dei non-coding RNAs, ed eventi di rimodellamento della cromatina. È ormai noto che le alterazioni epigenetiche rivestono un ruolo cruciale non solo nella progressione tumorale, ma anche nelle fasi più precoci di iniziazione. In questa prospettiva e anche grazie allo sviluppo e all’utilizzo delle nuove e numerose piattaforme high-throughput (Next Generation Sequencing) diversi geni con aberrante stato di metilazione sono stati proposti come potenziali biomarcatori per la diagnosi precoce, la prognosi, il follow-up della malattia e la scelta del trattamento. Nonostante l’enorme quantità di dati generati e generabili mediante l’utilizzo delle nuove e altamente performanti tecnologie (tempi di analisi, accessibilità delle metodiche etc.) solo pochi targets epigenetici presentano una reale utilità clinica. Un’efficace traslatabilità in ambito clinico di tali informazioni richiede, infatti, un approccio multidisciplinare che ponga le basi per una cornice di lavoro che integri diverse competenze e figure professionali per una corretta comprensione e contestualizzazione del dato molecolare nel setting diagnostico e terapeutico.

Parole chiave: Metilazione, Biomarkers, Tumori del colon-retto
Colorectal cancer

Colorectal cancer (CRC) is the third most common cancer worldwide, and accounts for 13% of all cancers in European Union countries (van de Velde et al., 2014). Despite the last twenty years improvement in colon and rectal cancers care, they still represent the fourth most common cancer-related cause of death globally, accounting for 600,000 deaths per year (Brenner et al., 2012). The proportion of rectal cancer cases in Europe can be variable, depending on the cancer registry and classification of recto-sigmoid tumours, and according to a recent study it can range from 27% to 58% (Brenner et al., 2012). Incidence strongly increases with age, with a median age at diagnosis of about 70 years in developed countries.

Most colon cancers start as non-neoplastic lesions with slow growth patterns (polyps); if polyps are removed in early stages, cancer evolution can be prevented and early detection of colorectal lesions plays an important role in survival rate improvement (Shah et al., 2014). As a consequence, a cornerstone in CRC management is represented by adequate screening and diagnostic programs. Screening modalities include fecal occult blood testing (FOBT), flexible sigmoidoscopy and colonoscopy. Despite pooled meta-analysis of randomized trials have shown a significant reduction in CRC mortality due to FOBT and flexible sigmoidoscopy application (16% and 30%, respectively) (Bretthauer, 2011), some limitations in these tests are still present. Concerning FOBT, patients adherence is low, sensitivity is limited in case of polyps and distal colon cancers and there is the possibility of false-positive results. Flexible sigmoidoscopy, although being a fast, low risk procedure that can be performed without patient sedation, has compliance limitations, and its diagnostic power can be limited by the quality of patient’s preparation. Colonoscopy represents the gold standard in CRC diagnosis; it has high sensitivity and specificity, but, on the other hand, is an expensive invasive tests. Research to identify new biomarkers as an ideal screening test with good compliance, high sensitivity and specificity, low costs and without risks for the patients is ongoing, but further validation is still required.

In CRC development recognized risk factors usually co-occur and interact; the ones established in epidemiological studies include age, male sex, family history of colorectal cancer, inflammatory bowel disease, smoking, excessive consumption of alcohol and red and processed meat, obesity and diabetes. People with first-degree affected relatives and inflammatory bowel disease are the ones with the strongest risk increase. Familial CRC is defined by the presence of two or more first-degree relatives affected with CRC and accounts for more than 20% of all cases. All CRC syndromes caused by known high-penetrance CRC genes collectively account for 2-6% of all cases (Valle, 2014) and despite technological progresses the genetic etiology of some familial forms still remains unknown. The two most common syndromes associated with CRC are Lynch Syndrome, an autosomal dominant inherited disorder caused by germline mutations or epimutations in a DNA mismatch repair gene (MLH1, MSH2, MSH6, PMS2) and Familial Adenomatous Polyposis, which in its classic form is autosomal dominantly inherited, is characterized by the development of up to thousands of colorectal adenomatous polyps and in most of the cases arises because of a germline heterozygous mutation in the tumour suppressor APC gene.

World Health Organization (WHO) has recently proposed a revised classification for colorectal cancers (Hamilton et al., 2010), and according to current international guidelines (Sobin et al., 2010), CRCs are also classified on the basis of local invasion depth (T stage), lymph node involvement (N stage) and presence of distant metastases (M stage). Therapeutic options are evaluated considering the overall stage definition, and important prognostic information are acquired thanks to this staging system application. In fact, five-year relative survival in colorectal cancer patients has reached almost 65% in high income countries (Siegel et al., 2012), remaining less than 50% in low income ones (Sankaranarayanan et al., 2010); for localized stages five-year relative survival rate is about 90%, it drops to 69.2% in case of regional spread and falls to 11.7% in advanced metastatic disease.

TNM staging system has a particularly relevant prognostic power especially for patients with early and advanced stage diseases, while, on the other hand, in case of intermediate stage disease its ability to predict outcome is less solid. Therefore, much effort has been provided for the identification of additional prognostic markers, in particular in this patients’ subgroup. Among the suggested prognostic markers, histological features can offer some support. In a recent review (Schneider and Langner, 2014) Signet Ring-cell Carcinoma and Micropapillary Carcinoma variants, as well as venous, lymphatic and perineural invasion and a high degree of tumour budding and necrosis have been proposed as markers of high risk for recurrence, while a favourable outcome has been associated to the Medullary variant and to the presence of a high degree of anti-tumor host response (as overall inflammation at the invasion margin, lymphocytic infiltration and tumour-associated eosinophils), in addition to the documented absence of markers indicating poor outcome. In recent years also molecular biomarkers have generated interest as prognostic and predictive markers. The most relevant example of molecular marker entered in clinical routine settings is KRAS mutation study in metastatic CRC, since mutations in this oncogene determine unresponsiveness to treatment with anti-EGFR antibodies. Novel proposed molecular classification models aimed to detect cancers at an early stage, determine prognosis and monitor therapeutic response are based on analysis of Microsatellite Instability, Chromosomal Instability, CpG island Methylator Phenotype and their correlation with other significant mutations as KRAS and BRAF (Sideris and Papagrorigiadiis, 2014).
Considering that in some occasions CRC patients with localized neoplasia die of disease relapse and metastases, the role of occult disease not recognised during conventional staging has been explored. Occult disease can be defined as the presence of Circulating Tumour Cells in the bloodstream, Disseminated Tumour Cells in the bone marrow or Isolated Tumour Cells and Micrometastases in lymph nodes. All these potential neoplastic infiltrations have been widely examined and their role as minimal residual disease defined. These parameters have been proposed as very promising in order to identify patients who could need additional treatment and as strong prognostic markers for patients’ survival (Bork et al., 2014).

Therapeutic strategies in CRC include surgery, chemotherapy and radiotherapy. In case of rectal cancer the standard surgical procedure is represented by total mesorectal excision, which involves removal of the rectum together with the mesorectum and the mesorectal fascia; since the introduction of this strategy the rate of local recurrences has dropped radically, considering that mesorectum contains the majority of potentially involved lymph nodes. Extension of colon surgery depends on tumour localization and supplying blood vessels; following current guidelines, the tumour and the corresponding lymph vessels are removed. Chemotherapy and radiotherapy can be administered as neoadjuvant (before surgery) therapies. In case of rectal cancer, benefit of neoadjuvant chemotherapy is clear in patients with Stage III disease, while is questioned for Stage II disease patients. Radiotherapy plays a more effective role as neoadjuvant strategy than as adjuvant (post-surgery), with reduced rates of toxic effects and local recurrences. Long-course radiotherapy combined with chemotherapy is privileged in some contexts over short-course protocols, especially in T3 distal rectal tumours. On the other hand, in locally advanced colon cancer the role of neoadjuvant treatment is not well defined, yet. Adjuvant chemotherapy, on the contrary, plays a pivotal role in Stage III colon cancer, which has a high risk of recurrence. Finally, in patients with distant metastases surgery should be proposed in case of resectable lesions, and palliative chemotherapy in all the other circumstances.

Epigenetic modifications

Colorectal cancer (CRC) results from the progressive accumulation of genetic and epigenetic alterations. Epigenetics refers to heritable and reversible changes associated with gene expression without alterations in the DNA sequence. Epigenetic modifications include aberrant DNA methylation, histone modifications, non-coding RNA deregulation, chromatin remodeling and nucleosome positioning (Lao and Grady, 2011).

DNA Methylation. Identification of aberrantly methylated genes since early stages of CRC (aberrant crypt foci and adenomas) has suggested that DNA methylation might play a role in driving CRC initiation and progression (Lao and Grady, 2011). DNA methylation in its two forms (hyper- and hypomethylation) occurs almost exclusively in the context of 5’CpG3’ dinucleotides. DNA hypomethylation, first described in 1983 from Feinberg and Vogelstein, is predominantly observed at CpG dinucleotides in repetitive sequences and endogenous transposons. Hypomethylation also occurs early in carcinogenesis and increases in age-dependent manner (Feinberg and Vogelstein, 1983). Loss of methylation in repeat elements has been postulated to be linked with chromosome instability. In the process of cytosine methylation, a methyl group is transferred from the donor S-adenosylmethionine (SAM) to the 5’-C cytosine carbon atom. This reaction is catalyzed by a family of enzymes, DNA methyltransferase (DNMTs) which are involved in de novo DNA methylation (DNMT3A and DNMT3B) and in maintaining methylation (DNMT1). During DNA replication DNMT1 acts in a complex, recognizes hemimethylated DNA and catalyzes the addition of a methyl group to the unmethylated strand. DNMT3A and DNMT3B are mainly involved in the novo methylation, although a role in maintaining methylation pattern in some cell types has been suggested (Okano et al., 1999). Depending on the genomic location, aberrant DNA methylation may have different biological functions. CpG dinucleotides can be located through the body genome or concentrated in CpG rich regions (referred as “CpG islands”). CpG islands occur in the flanking promoter region of genes or in regions such as centromeres and retrotransposon elements where are located repetitive sequences. Gene body hypermethylation is associated with transcriptional activation, and regulation of tissue-specific expression from alternative promoter (Hellman and Chess, 2007; Maunakea et al., 2010). A distinctive methylation pattern identified around exons and exon-intron borders, suggests that chromatin structure is also important for the exon selection (Anastasiadou et al., 2011), so DNA methylation may also result in a dysregulation of the splicing process (Maunakea et al., 2013). Hypermethylation of CpG islands in the promoter region is commonly associated with gene silencing through chromatin conformational modifications and inhibition of the access of the transcriptional machinery (Gius et al., 2004). methylated cytosine residue can also be subjected to deoxygenation to 5-hydroxymethylcytosine (5-hmC) by Ten-eleven translocation proteins (TET) (Tahiliani et al., 2009; Ko et al., 2010). 5-hmC is an epigenetic modification of DNA prevalently observed in embryonic stem cells (ES cells) and in different human tissues. Abundance of 5-hmC varies in human tissues and it is significantly decreased in numerous solid tumors, including colon carcinoma (Haffner et al., 2011). Although the exact function of 5hmC is still unclear, recent studies underline its important role in the epigenetic machinery, particularly in the chromatin remodeling and DNA methylation and in maintaining the normal status of the cells.
Histone modifications. Histone modifications include acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deamination and proline isomerization of specific core histone residues (H2A, H2B, H3, H4). These post-translational modifications are involved in the mechanism of transcriptional control and in various physiological processes and play an important role in arising the aberrant genetic and epigenetic profiles, which occur in numerous types of tumors (as breast, colorectal and hematological malignancies) (Dawson and Kouzarides, 2012; You and Jones, 2012). Among histone modifications, acetylation and methylation are the most extensively characterized in CRC. Acetylation of key lysine residues (K) at histone tail is a reversible modification mediated by histone acetyltransferase (HATs) and histone deacetylases (HDACs) (Dawson and Kouzarides, 2012). HDACs, including three families of enzymes, are often aberrantly regulated during CRC development. Methylation occurs prevalently at lysine residues, but also involves arginine and histidine ones, in which methyl groups are added or removed by specific enzymes, histone methyltransferases (HMTs) and histone demethylases (HDMDTs) respectively (Migheli and Migliore, 2012; Dawson and Kouzarides, 2012). Lysine residue can be mono-, di- or tri-methylated and these different histone methylated states exhibit a cell-specific pattern that results in specific differences in gene activation (Nguyen and Zhang, 2011). For histone H3, methylation at H3K36 and H3K79, and H3K4 trimethylation (H3K4me3), are strictly associated with transcriptional activation (Vakoc et al., 2006) while H3K27 trimethylation (H3K27me3) is frequently linked with gene silencing (Mikkelsen et al., 2007; Kouzarides, 2007).

miRNA. MiRNAs are short endogenous noncoding RNAs (~22 nucleotides) with a significant gene regulatory function. MiRNAs induce cleavage of target mRNA or translational repression binding to a complementary site on the 3’-untranslated region of a target mRNA. MiRNAs dysregulation plays an important role in tumor development and progression, and migration (Suzuki et al., 2012). Most of the miRNAs (as well as miR-34 and miR-17) act as tumor-suppressors and are often downregulated in CRC (Liang et al., 2013) while few miRNAs, called oncomiRNAs are upregulated. MiR-34 family members are direct targets of p53, and their ectopic expression in CRC cells induces cell cycle arrest and apoptosis (Bommer et al., 2007). Two important components of epigenetic machinery, DNA methylation and histone modifications, are involved in a “reciprocal” constitutive interplay, in the positive or negative regulation of miRNA gene expression. miR-140 and miR-449 target members of HDAC family and their downregulation results in aberrant histone induction (Iorio, 2010). DNMT3a is targeted by miR-143, and suppression of miR-143 in colorectal cancer induces an overexpression of DNMT3a (Liu and Chen, 2010; Suzuki et al., 2012). Methylation in the promoter region decreases transcription levels of corresponding miRNAs affecting the expression of miRNA target genes. Hypermethylation of miR-34b/c has been observed in the vast majority of primary CRCs as well as hypermethylation-induced silencing of miR-1-1, miR-9-1, miR-129-2 and miR-124 family (Suzuki et al., 2012).

High-throughput approach for DNA methylation- NGS platform analysis

Over the past few decades, there have been an increasing number of methods devoted to generate genome-wide methylation signatures, each with its own advantages, disadvantages, and areas of applicability. Sequencing and microarray-based platforms are two main approaches that have given rise to a wide range of techniques for global DNA methylation analyses. These approaches are high-throughput strategies with regard to the number of loci that can be analyzed at one time. Particularly, DNA methylation profiling using high-density microarrays is commonly used to identify broad differences between groups of samples. This method is less time consuming and less costly than sequencing. Additionally, microarrays allow for simultaneous analysis of a larger number of samples with a wider coverage of CpG islands (CGI). Nevertheless, microarray platforms lack reliable quantitation and are limited by hybridization efficiency, hybridization artifacts and probe design.

On the contrary, sequencing analyses provide quantitative information about the methylation status of every CpG and allow for the analysis of methylation in repeat sequences and rare methylation variants, which are not frequently detectable by microarrays. Moreover by sequencing techniques it is possible to analyze DNA methylation of regions with no prior knowledge of the sequence. The main weaknesses of sequencing strategies are cost, availability, library bias, and difficulties in the analysis and management of data, although the cost of massive sequencing technologies is continuously and rapidly decreasing.

Since DNA methylation is lost during PCR amplification, the majority of methylation analysis methods rely on treatment of DNA prior to amplification. These assays can be classified into three main groups: a) restriction enzyme (RE) digestion, b) affinity-based analysis, and c) bisulphite modification.

a) Restriction enzyme-based methods explore the property of methylation-sensitive enzymes which only digest unmethylated DNA and methylation-dependent enzymes able to cut only methylated DNA. These enzymes are used to enrich for methylated or unmethylated sequences and provide a read-out of DNA methylation. The most common RE-based methods include the restriction landmark genome scanning (RLGS) (Hayashizaki et al., 1993; Hatada et al., 1991; Smiraglia and Plass, 2002), that was the first RE-based method used for global DNA methylation profiling, although its
use is currently decreasing since it requires the use of radioactive materials and gel electrophoresis; the HpaII tiny fragment enrichment by ligation-mediated PCR (HELP) (Khulan et al., 2006), that has been combined with massively parallel sequencing (HELP-Seq) and/or array-based platforms (Khulan et al., 2006; Oda et al., 2009); the Methyl-Seq (Brunner et al., 2009), where the digestion products are size fractioned and the selected fragments are sequenced; the luminoimetric methylation assay (LUMA) (Brunner et al., 2009; Karimi et al., 2006), based on DNA digestion, bioluminometric polymerase extension, and pyrosequencing; the methylation-sensitive cut counting (MSCC) (Ball et al., 2009); the methylated CGI amplification (MCA) (Toyota et al., 1999) and the methylation amplification DNA chip (MAD) (Hatada et al., 2002) both utilizing Smal (methylation sensitive) and Xmal (methylation insensitive) digestion; the promoter-associated methylated DNA amplification DNA chip (PMAD) assay which incorporates the HpaII andMspI enzymes (Fukasawa et al., 2006; Huang et al., 2010).

Alternative strategies are based on the employ of methylation-dependent enzymes as McrBC, an enzyme that recognizes closely spaced methylated cytosines and so is high sensitive to densely methylated regions of DNA (Sutherland et al., 1992). The principal techniques that utilize this enzyme are the comprehensive high-throughput arrays for relative methylation (CHARM) (Irizarry et al., 2008); the microarray-based methylation assessment of single samples (MMASS), and MethylScope (Huang et al., 2010; Ibrahim et al., 2006; Ordway et al., 2006; Ordway et al., 2007). The advantage of McrBC-based approach relies on the fact that it does not require prior methylation information from a control reference genome. Other variations of RE-based methods include those that comprise a combination of methylation-sensitive enzymes as the differential methylation hybridization (DMH) (Huang et al., 1999) based on methylation-sensitive enzymes such as BstUI, HhaI, and HpaII; and the methylation single-nucleotide polymorphisms (MSNP) (Kerbel et al., 2008; Yuan et al., 2006) that has the advantage of providing information about copy number variations in XbaI fragments, SNPs in HpaII cutting sites in XbaI fragments, and methylation in HpaII cutting sites from one microarray.

RE-based genome-wide DNA methylation analysis is a relatively simple and cheap method used for genome-wide screening to identify frequently methylated CpGs (Kane et al., 2006; Lee et al., 1994; Pieper et al., 1991). However it presents the limit of interrogating a subset of methylation sites, moreover they are unable to distinguish 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) (Nestor et al., 2010). To address this issue, new enzymatic approaches have been developed for specific detection of hydroxymethylated cytosines (Kinney et al., 2011; Janosi et al., 1994).

b) Different strategies use affinity purification to enrich for methylated DNA. For example, in the methylated DNA immunoprecipitation (MeDIP) single-stranded DNA fragments are immunoprecipitated with one or more monoclonal anti-5-mC or anti-5-hmC antibodies. The collected DNA is enriched for methylated sequences and is then amplified and analyzed using sequencing (MeDIP-Seq) or microarray platforms (Weber et al., 2005; Down et al., 2008). However, the MeDIP method requires the use of single-stranded DNA which may be difficult to achieve in high CpG content regions; then specific antibody must be used. An alternative method is the methylated CGI recovery assay (MIRA) which utilizes methyl binding domain proteins (MBDs) and MBD column chromatography (Cross et al., 2012; Rauch and Pfeifer, 2005; Gebhard et al., 2006). The MIRA captured DNA is then PCR amplified and analyzed using microarrays or sequencing. Affinity-based methods allow for rapid and specific assessment of the mean methylation levels of large DNA regions. Relevant limits of this assay are the necessity of high-DNA input, the lack of information on distinct CpG dinucleotides and the sequence bias that can derive from MBD or antibody interaction with DNA.

c) Bisulphite treatment-based strategies of methylation analysis are the most widely accepted and most widely used. The principle of sodium bisulphite modification is based on the differential reaction of methylated and unmethylated cytosines with the reagent, such that only unmethylated cytosines are converted into uracils (Hayatsu et al., 1970). With this methodology it is possible to perform quantitative DNA methylation analysis almost anywhere in the genome, and to detect strand-specific methylation, with single CpG resolution. However, since analysis depends on the complete bisulphite conversion of unmethylated cytosines to uracil, incomplete or inappropriate conversion could be erroneously interpreted. Furthermore, the bisulphite conversion can cause significant DNA degradation. This can affect sequencing and array analysis.

The “oxidative bisulphite” sequencing (oxBS-Seq) approach as well as bisulphite-independent strategies involving alternative chemical pretreatments have been recently developed (Booth et al., 2012) to distinguish genomic loci identified as methylated from that hydroxymethylated. The bisulphite-dependent methylation methods are based on three technical approaches:

1. Sequencing, that is achieved by the following methodologies: a) whole genome shotgun bisulphite sequencing (WGSGS) which provides a genome-wide methylation profile at single base-pair resolution and is therefore the most comprehensive high-throughput method (Bormann Chung et al., 2012; Lister et al., 2009); b) reduced representation bisulphite sequencing (RRBS), that enriches for CpG-rich regions using RE such as BglII orMspI to reduce genome complexity and sequence redundancy (Meissner et al., 2005; Meissner et al., 2008) c) denaturing HPLC (DHPLC) (Baum et al., 2001) which allows to identify methylation profiles on the basis of different retention times of DNA fractions in HPLC under partially denaturing conditions. HPLC is a simple, cost-effec-
tive and rapid technique. However, it requires relatively high DNA quantities and has limited sensitivity, especially when analyzing tissue specimens.

2. Microarray-based DNA methylation profiling, that can be applied to a technique, known as bisulphite methylation profiling (BIMP), where bisulphite-treated DNA first undergoes a whole genome amplification (WGA) and then the enzymatic fragmentation and microarray hybridization (Reinders et al., 2008). The microarray is designed using differentially labeled oligonucleotide pairs complementary to the unchanged, methylated sequence. Therefore, methylation is detected as a signal and mismatches caused by the conversion of unmethylated cytosines do not result in signal. This approach results in overall low hybridization signal and may not be applicable to regions of sparse methylation. Similarly, the Infinium approach entails a sample preparation that involves bisulphite modification of genomic DNA followed by whole genome amplification (WGA) (Bibikova et al., 2006; Weisenberger et al., 2008). However, in this methodology the DNA is then hybridized to BeadChip microarrays, which are designed with oligonucleotide pairs targeting CpG sites of interest, with one complementary to the unchanged, methylated sequence and the other to the converted unmethylated sequence. Next, a PCR reaction is performed with fluorescently labeled universal PCR primers and the methylation levels can be determined by comparing the proportion of fluorescence emitted by each dye. Most microarray platforms contain a standard array of probes covering a library of CGIs.

3. Massively parallel DNA sequencing, which includes next-generation sequencing (NGS). Even if earlier methylation analyses relied exclusively on BS and, subsequently, array-based profiling approaches were leading the field of DNA methylation-based biomarker discovery, NGS-based approaches have quickly emerged as the platform of choice at the present.

The development of NGS platforms enables sequencing and mapping of millions of DNA fragments in parallel, thus significantly increasing throughput and decreasing cost per base thereby providing new opportunities for comprehensive, highly sensitive, genome-wide mapping of methylation sites at a more affordable price (Taylor et al., 2007). Innovation continues to fuel the NGS revolution so that these methodologies are gradually replacing conventional Sanger sequencing (Frommer et al., 1992). However, to extract signals from high-dimensional NGS data and make valid statistical inferences and predictions, novel data analytical and statistical techniques have to be constantly developed.

Roche 454 sequencing consists in clonal amplification of library fragments bound on beads and in sequencing of individual beads by pyrosequencing. It can generate up to one million reads per run at read lengths of up to 1 kbp. It provides the fastest time per run and longest read length compared with other NGS platforms, offering several advantages for methylation analysis.

Applied Biosystems SOLiDTM is also based on the PCR generation of clonally amplified sequencing fragments with small beads attached to a solid surface, but in this assay the sequencing is achieved using sequencing by synthesis driven by a ligase. It can generate up to 700 million reads per run at read lengths of up to 75 bp. In this platform each base position is examined twice; thus, miscalls can be more readily identified. In MethylSeqTM, bisulphite-modified DNA is also amplified by microdroplet emulsion PCR using a primer library targeting a large number of genes. The resulting PCR library is sheared, ligated, and subjected to massively parallel clonal sequencing (Herrmann et al., 2011; Komori et al., 2011). Differently, the Illumina Solexa genome analyzer is based on in situ bridge template clonal amplification on a solid surface with amplicons remaining clustered in a single physical location. Up to eight independent amplicon libraries are then sequenced in parallel using sequencing-by-synthesis technology that employs reversible terminators with removable fluorescent dyes. The Illumina Solexa genome analyzer can generate over 300 million reads per run at read lengths of up to 2 x 150 bp. It is the most widely used NGS strategy for DNA methylation analysis. In general, both Applied Biosystems SOLiDTM and Illumina Solexa genome analyzer offer higher throughput and lower cost compared to Roche 454 but are more limited in alignment of bisulphite-converted sequences.

As extensively described, there are still many challenges to the effective implementation of DNA methylation-based biomarkers. Emerging innovative methylation detection strategies are focused on addressing the remaining gaps in the field of epigenetics. In this scenery, the development of NGS-based DNA methylation assay is an exciting and rapidly evolving area of research that holds promise for potential applications in diverse clinical settings.

**DNA methylation as prognostic and predictive biomarker of colorectal cancer.**

Investigation of aberrantly methylated genes as markers is useful to identify novel tumor-suppressor genes and methylation biomarkers for cancer classification (Kaneda et al., 2002; Kaneda et al., 2004; Toyota et al., 1999; Toyota et al., 1999; Ushijima et al., 1997; Ushijima, 2005). Aberrant DNA methylation associated with certain genes has indeed attracted considerable interest as a potential biomarker for the early detection of disease onset, prognosis and choice of
treatment, and the monitoring of disease after therapy (Laird, 2003; Levenson, 2007; Tost, 2010). In Table I a list of examples of aberrantly methylated genes is shown. Several methods for genome-wide methylation analysis have been developed as shown in previous paragraphs. Epigenetic alterations have been noted to play crucial roles not only in cancer progression, but also in cancer initiation, since the alterations have been identified in the pre-cancerous “normal” tissues that could modify cancer risk (Kaneda et al., 2002; Baylin and Ohm, 2006; Feinberg et al., 2006; Sakatani et al., 2005). Importantly, gene silencing resulting from aberrant DNA methylation cooperates with other genetic mechanisms to disrupt the key molecular pathways critical in colorectal carcinogenesis (Baylin and Ohm, 2006).

Currently, the best cancer biomarkers are those that are not methylated in normal healthy tissues and are relatively homogeneously methylated both at early stages and as the tumor progresses. Furthermore, the biomarkers should be analyzed in both tumor and adjacent normal tissues.

Some DNA methylation biomarkers have already been adopted as disease markers in various studies. DNA methylation of MLH1 and BRCA1 in normal tissues indicates a predisposition to the development of colorectal and breast cancer, respectively (Gazzoli et al., 2002; Snell et al., 2008). CDKN2A (p16) in sputum has been used to screen patients at increased risk of lung cancer (Palmisano et al., 2000) and GSTP1 is an exceptionally clean biomarker for early detection of prostate cancer (Brooks et al., 1998). Our group has conducted a cancer methylome discovery study in cancer cell line and in 13 different cancer histotype including colorectal cancer (Hoque et al., 2008). The approach coupled probabilistic search algorithms in the entire human genome with an established pharmacologic unmasking strategy in cancer cell lines for unbiased and precise global localization of tumor-specific methylated genes. The computational approach followed by microarray analysis allowed to identify 200 genes predicted to be methylated. 25 were previously reported as harboring cancer specific promoter methylation after a literature search. The remaining 175 genes were tested for promoter methylation by bisulphite sequencing (BS) or methylation-specific PCR (MSP) in one or more cell line that exhibited re-expression after demethylating treatment. Promoter methylation of 82 genes (82/175) (47%) was documented in cell lines based on identification of ≈50% methylated CpG sites in the CpG island. Out of 82 genes which showed methylation in cell lines, promoter methylation was detected in 53 (65%) genes in primary tumor tissues. After testing corresponding age matched normal tissues, 28 of these genes were identified to be methylated in a cancerspecific manner. Thus, across multiple cancer types 28/175 (16%) new cancer-specific methylated genes were identified through the combination of a computational approach and empiric studies. For some of these genes no previous methylation in human cancers was discovered but they had been linked to cancer through functional studies; for other no previous connection with neoplasia was found. Particularly based on preliminary data of the discovery study we validated in more than one independent colorectal cancer sample sets a novel cancer-specific methylated gene: B4GALT1 (beta 1,4 galactosyltrasferase I). This gene has shown a methylation frequency 54% and a cancer-specificity > 90% (Poeta et al., 2012). In another genome-wide analysis (Khamas et al., 2012) 15 CRC cell lines and 23 paired tumor and normal samples from CRC patients were used to identify a set of methylation silenced genes in CRC. Gene expression studies were then used to confirm whether the methylated genes were really regulated by their methylation status. The results of this study revealed that 139 genes showed greater than 1.5-fold up-regulation in at least one 5-aza-2′-deoxycytidinetreated cell line and no less than a 1.2-fold change in other treated CRC cell lines. Among them 20 genes with poor an-

<table>
<thead>
<tr>
<th>BIOMARKER</th>
<th>SOURCE</th>
<th>Tissue of CRC</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWIST1</td>
<td>tissue CRC</td>
<td>Valdes-Mora F et al., 2009</td>
<td></td>
</tr>
<tr>
<td>IGFBP3</td>
<td>tissue CRC</td>
<td>Kawasaki T et al., 2007</td>
<td></td>
</tr>
<tr>
<td>GAS7</td>
<td>tissue CRC</td>
<td>Kim YH et al., 2011</td>
<td></td>
</tr>
<tr>
<td>WNT5A</td>
<td>tissue CRC</td>
<td>Rawson JB et al., 2011</td>
<td></td>
</tr>
<tr>
<td>SLIT2</td>
<td>Stool of CRC patients</td>
<td>Azzara D et al., 2013</td>
<td></td>
</tr>
<tr>
<td>PHACTR3</td>
<td>Stool of CRC patients</td>
<td>Bosch LJ et al., 2012</td>
<td></td>
</tr>
<tr>
<td>SPG20</td>
<td>Stool of CRC patients</td>
<td>Lind GE et al., 2011</td>
<td></td>
</tr>
<tr>
<td>3OST2</td>
<td>Stool of CRC patients</td>
<td>Tokuyama Y et al., 2010</td>
<td></td>
</tr>
<tr>
<td>OSMR</td>
<td>Stool of CRC patients</td>
<td>Kim MS et al., 2009</td>
<td></td>
</tr>
<tr>
<td>ITGA4</td>
<td>Stool of CRC patients</td>
<td>Auc R et al., 2009</td>
<td></td>
</tr>
<tr>
<td>HIC1</td>
<td>Stool of CRC patients</td>
<td>Lenhard K, et al., 2005</td>
<td></td>
</tr>
<tr>
<td>TFF12</td>
<td>Stool of CRC patients</td>
<td>Zhang J et al., 2012</td>
<td></td>
</tr>
<tr>
<td>NDRG4</td>
<td>Stool of CRC patients</td>
<td>Melotte V et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Vimentin</td>
<td>Stool of CRC patients</td>
<td>Li M et al., 2009</td>
<td></td>
</tr>
<tr>
<td>GATA4</td>
<td>Stool of CRC patients</td>
<td>Hellebrekers DM et al., 2009</td>
<td></td>
</tr>
<tr>
<td>DLC1</td>
<td>Blood of CRC patients</td>
<td>Peng H et al., 2013</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>Blood of CRC patients</td>
<td>Leung WK et al., 2005</td>
<td></td>
</tr>
<tr>
<td>HLT8</td>
<td>Blood of CRC patients</td>
<td>Wallner M et al., 2006</td>
<td></td>
</tr>
<tr>
<td>ALX4</td>
<td>Blood of CRC patients</td>
<td>Ebert MP et al., 2006</td>
<td></td>
</tr>
<tr>
<td>TMEF2</td>
<td>Blood of CRC patients</td>
<td>Lofton-Day C et al., 2008</td>
<td></td>
</tr>
<tr>
<td>NGFR</td>
<td>Blood of CRC patients</td>
<td>Lofton-Day C et al., 2008</td>
<td></td>
</tr>
<tr>
<td>TPEF/HPP1</td>
<td>Blood of CRC patients</td>
<td>Wallner M et al., 2006</td>
<td></td>
</tr>
<tr>
<td>SDC2</td>
<td>Blood of CRC patients</td>
<td>Oh T et al., 2013</td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>Blood of CRC patients</td>
<td>Leung WK et al., 2005</td>
<td></td>
</tr>
<tr>
<td>SEPT9</td>
<td>Blood of CRC patients</td>
<td>Ahquist DA et al., 2012</td>
<td></td>
</tr>
<tr>
<td>RUNX3</td>
<td>Blood of CRC patients</td>
<td>Nishio M, et al., 2010</td>
<td></td>
</tr>
<tr>
<td>NEUROG1</td>
<td>Blood of CRC patients</td>
<td>Herbst A et al., 2011</td>
<td></td>
</tr>
<tr>
<td>CDKN2A(p16)</td>
<td>Blood of CRC patients</td>
<td>Ahn JB et al., 2011</td>
<td></td>
</tr>
<tr>
<td>SFRP2</td>
<td>Blood of CRC patients</td>
<td>Tang D et al., 2011</td>
<td></td>
</tr>
</tbody>
</table>

Legend: CRC, Colorectal Cancer
notation, 20 genes located on the X chromosome, 16 genes with duplicated probes, 2 genes with no CpG islands, 8 genes with unknown function, 23 without a relevant function in tumorigenesis and 22 genes with potential oncogenic activity were also excluded, leaving 20 candidates (CAMK2B, CHAC1, THSD1, CSTA, COL1A1, GADD45B, DMRB1, COL5A1, GASS, GP RC5A, GPSM1, KLLH35, LTBP2, NA11, RBP4, SEMA7A, SYCP3, TBRG1, TNFSF9 and TXNIP) that had not been previously reported to be affected by epigenetic mechanisms in CRC. Therefore, from the 54613 genes analyzed, a much smaller set of genes was isolated as potential biomarkers for CRC. An interesting hypermethylated target THSD1 (Thrombospondin type-1 domain containing protein 1) seems to have the potential for diagnostic, prognostic or therapeutic use. THSD1 is located in a region that is strongly associated with the progression of colorectal adenoma to carcinoma and encodes a transmembrane molecule containing a thrombospondin type 1 repeat that might be involved in cell adhesion and angiogenesis. High THSD1 expression positively correlated with better distant metastasis survival in breast cancer. Therefore, its loss may be associated with metastatic tumor spread. Additionally another study has shown that THSD1 was expressed in Duke D CRC (Levenson, 2007).

Molecular Subgroups of CRC: Although all CRCs are characterized by the presence of hypermethylation, a specific subgroup of them, denoted as the CpG island methylator phenotype, shows extensive levels of methylated genes (Toyota et al., 1999). CRC is not a homogeneous disease, but rather a heterogeneous disorder in which different subtypes are characterized by distinct genetic, cytogenetic and epigenetic alterations (Kaneda et al., 2004). For colorectal cancers, the acquisition of genomic instability is considered a key hallmark. Three major molecular subtypes can be recognized: MIN (or MSI, for “microsatellite instability”), CIN (for “chromosomal instability”) and CIMP (for “CpG island methylator phenotype”). These three patterns (MIN, CIN, and CIMP) are not mutually exclusive, and it is believed that a tumor can occasionally show features of multiple pathways, although the extent and nature of this overlap remains to be determined (Kaneda et al., 2002). By the epigenetic point of view, CRCs can be divided into CIMP (CIMP+), and non-CIMP tumors, and three subgroups can be identified: (a) CIMP1 tumors, showing often microsatellite instability (80%) and BRAF mutations (53%); (b) CIMP2 tumors that present often K-RAS mutations (92%), but rarely have microsatellite instability or BRAF or TP53 mutations. (c) Non-CIMP tumors displaying a high frequency of TP53 mutations (71%) (Shen et al., 2007). Another genome-scale analysis of aberrant DNA methylation identified four DNA methylation-based subgroups of CRC, including (1) A CIMP-high subgroup with a very high frequency of cancer-specific DNA hypermethylation associated with MLH1 methylation and the BRAF V600E mutation. (2) A CIMP-low subgroup presenting mutations of K-RAS and methylation of a subset of CIMP-high associated genes, rather than a unique group of CpG islands. (3) A non-CIMP subgroup characterized by the presence of TP53 mutations and frequent occurrence in the distal colon. (4) A non-CIMP subgroup displaying a low frequency of cancer-specific gene mutation and hypermethylation, and enriched of rectal tumors (Hinoue et al., 2012).

Diagnostic and prognostic value of DNA methylation in CRC: At least six genes (SLC5A8, SFRP1, SFRP2, CDH13, CRBP1, and RUNX3) and two loci (MINT1 and MINT31) have been consistently found to be methylated in the passage from a normal colon epithelium to an aberrant crypt focus. Other genes (p14, HLTf, ITGA4, CDKN2A/p16, CDH1 and ESR1) are frequently methylated in the passage from an aberrant crypt focus to poly/pseud/adenoma. Four additional genes (TIMP3, CXCL12, ID4, and IRF8) are frequently methylated in late CRC stages and could have a role in CRC progression and metastasis (Lao and Grady, 2011). Other genes whose methylation seems to be involved in the progression from adenomas to cancer are the DNA repair genes MGMT and hMLH1 (Lao and Grady, 2011). It has been also shown that lymphovascular invasion of CRC is related to methylation of the gene SPARC (encoding the secreted protein acidic and rich in cysteine) in stromal cells (Yoshimura et al., 2011). The methylation analysis of hMLH1 is currently useful to distinguish between sporadic CRC with microsatellite instability from familial forms such as the Lynch syndrome (Bouzourene et al., 2010). It has been recently shown that the analysis of DNA methylation in mucosal wash fluid from patients undergoing colonoscopy may be a good molecular marker for predicting the invasiveness of colorectal tumors, and showed that mir-34b/c methylation had the greatest correlation with invasive tumors (Kaminue et al., 2011). It was also observed that CRC that have silenced (methylated) genes in the extracellular matrix (ECM) remodeling pathway, such as IGFBP3, EVL, CD109 and FLNC, show worse survival, suggesting that methylation of this pathway might represent a prognostic signature in CRC (Yi et al., 2012). Furthermore, hypomethylation of the IGF2 (Insulin growth Factor II) differentially methylated region in colorectal tumors was associated with poor prognosis (Baba et al., 2010).

Biomarkers of DNA methylation in blood: Biomarkers detected in patient blood samples would provide a very useful screening tool for CRC because blood specimens can be easily collected. It is indeed well known that genetic material can shed from tumor cells, and aberrant DNA methylation can be specifically quantified in blood despite the large amounts of normal DNA in circulation. A blood biomarker with a high sensitivity and specificity for CRC can not only be used to discriminate high-risk patients for further clinical tests but may also be an excellent tool to monitor recurrence in patients who have undergone surgical resection (Leung et al., 2005; Wallner et al., 2006; Ebert et al., 2006; Lofton-Day et al., 2008; Grützmann et al., 2008; Tänzer et
al., 2010; Ahlquist et al., 2012; Warren et al., 2011; Herbst et al., 2011; Tang et al., 2011; Tan et al., 2007). The SEPT9 gene, encoding a guanosine triphosphate enzyme involved in cytokinesis and cell cycle control, has been reported to be associated with several cancers. The v2 region of the Septin 9 (SEPT9) promoter has been shown to be methylated in CRC tissue compared with normal colonic mucosa. Methylated SEPT9 was first detected in the plasma of CRC patients with an overall sensitivity of 72% and a specificity of 90% (Grützmann et al., 2008). Significant validation has been performed for this methylation biomarker. Based on these results, blood SEPT9 methylation appears to have the highest probability of distinguishing between CRC cases and cancer-free subjects. Currently, two CRC detection kits using plasma SEPT9 methylation analysis are in the market for clinical application. Other genes, such as APC, hMLH1, ALX4, TMEFF2, NGFR, NEUROG1, SFRP2, CDKN2A/p16, TPEF/HPP1 and RUNX3, have also emerged as serum methylation markers for CRC, with sensitivities ranging from 6% to 83% and specificities ranging from 69% to 100%. Among them, ALX4, TMEFF2 and NEUROG1 showed better performance relative to the others, and the use of these markers in combination can improve detection accuracy (Tänzer et al., 2010; Ahlquist et al., 2012; Warren et al., 2011; Herbst et al., 2011; Tang et al., 2011; Tan et al., 2007). Methylation of HLTf (helicase-like transcription factor) was detectable in the stool of almost half of all patients with hyperplastic polyps or colorectal adenomas (Oberwalder et al., 2008), further supporting its use in the detection of premalignant lesions. Methylation level of fecal SFRP2 drops dramatically after surgery [postoperative: 8.7% (6/69) vs preoperative: 87% (60/69)] (Wang and Tang, 2008), suggesting its possible utility as a recurrence biomarker. The vimentin gene, which encodes an intermediate filament protein involved in cell attachment, migration and signaling, is detectable in the stool of 83% of CRC patients with a specificity of 90%. Follow-up studies have been shown similar results (Li et al., 2009; Ahlquist et al., 2008; Itzkowitz et al., 2008; Zou et al., 2007). Single-gene stool kit for CRC detection based on vimentin methylation is on the market since few years ago. More recently, vimentin methylation has been used in combination with other methylation markers as reported by Ahlquist et al. (Ahlquist et al., 2012). In addition to SFRP2 and vimentin, several other methylation biomarkers have been identified; these include GATA4, HIC1, ITG4, NDRG4, OSMR, TPFP2, ESR1, SLIT2, PHACTR3, SPG20, 3OST2 and MGMT. These genes have sensitivities for CRC ranging from 38% to 89% and specificities ranging from 79% to 100%. However, more clinical studies are required to confirm these results.

**Predictive Value of DNA methylation in CRC.** Epigenetic alterations in colorectal cancer are also of interest for their possible interactions with chemotherapeutic agents. Indeed, the epigenetic silencing of a particular gene might result in chemosensitivity or chemoresistance toward a particular therapeutic agent. Little is still known concerning these interactions. Increased TYMS expression is one of the major mechanisms of 5-FU chemoresistance, and there is indication that histone acetylation/deacetylation processes, rather than DNA methylation of the promoter, might be of relevance in epigenetically regulating TYMS expression in CRC. Several other genes are involved in pyrimidine metabolism and might represent potential molecular determinants of 5-FU chemoresistance, including those coding for dihydopyrimidine dehydrogenase (DYPD gene), thymidine phosphory-
lase (TYMP gene), and uridine monophosphate/cytidine monophosphate kinase (UMPK gene). Some evidence suggests that these genes might be regulated by promoter DNA methylation. However the majority of these studies have been conducted in cancer cell lines, and their potential epigenetic link to 5-FU resistance in CRC patients is still not clear (Crea et al., 2011). It has been also shown that decreased levels of SPARC expression are associated with a decreased sensitivity to chemotherapy, including 5-FU and irinotecan therapy, and the SPARC gene resulted methylated in CRC tissues but not in the normal colon (Cheetham et al., 2008). Stages II or III patients with CIMP-high CRCs, undergoing surgery with fluoropyrimidine-based adjuvant chemotherapy, showed significantly better recurrence-free survival than patients treated with surgery alone (Min, 2011). Irinotecan is an inhibitor of topoisomerase I. The UDP-glucuronosyltransferase (UGT) 1A1 enzyme is the major enzyme involved in irinotecan detoxification. Irinotecan pharmacogenomics is mainly based on the UGT1A1 genetic profile and in vitro data shown that the gene silencing occurs by DNA methylation, indicating that epigenetic regulation of this gene might be relevant to irinotecan chemosensitivity in CRC (Crea et al., 2011).

Conclusion

A variety of DNA methylation, biomarkers from tissue, stool, and blood has been reported for CRC detection. Identified markers are usually derived from tumor cells and others are derived from non-tumor cells in the tumor microenvironment or blood. DNA biomarkers mostly shed from tumor cells, and so, theoretically, these molecules should be more specific than protein biomarkers such as carcinoembryonic antigen (CEA), carbohydrate antigen (CA) 19-9, CA242 and CA724, which are currently and widely used in the clinic. Since nucleic acid-based markers can be amplified thus producing a stronger signal, they show a greater sensitivity. In comparison with DNA, protein biomarkers are less specific because tumors often induce inflammatory reactions, and some of the biomarkers that initially showed promise for cancer detection now appear to also detect a wide range of bowel diseases, such as ulcerative colitis and Crohn’s disease. Further, protein biomarkers are often altered in more than one type of cancer, such as CEA which has been reported as a biomarker for various malignancies, including breast cancers, renal, pancreatic, lung and colorectal. In spite of these issues, protein biomarkers may still be useful for large-scale screening for CRC because proteins can be observed through assays in small sample volumes with relatively simple and cheap assays. In relation to DNA methylation, the most recent research aims to identify epigenetic signatures that could be used for CRC diagnosis, staging, metastasis prediction, prognosis, and response to treatment. Rapidly evolving sequen-

References


Min BH, Bae JM, Lee EI, et al. The CpG island methylator phenotype may confer a survival benefit in patients with stage II or III colorectal carcinomas receiving fluoropyrimidine-based adjuvant chemotherapy. BMC Cancer 2011;9:344.


Supported by: AIRC MFA g 10520