Università degli Studi di Cagliari

PHD DEGREE

MOLECULAR AND TRANSLATIONAL MEDICINE

Cycle XXII

TITLE OF THE PHD THESIS

DNA METHYLATION ALTERATIONS AS FUNCTIONALLY RELEVANT MECHANISMS, PROMISING TUMOUR BIOMARKERS AND THERAPEUTIC TARGETS

Scientific Disciplinary Sector

BIO/13

PhD Student:	Eleonora Loi
Coordinator of the PhD Programme	Sebastiano Banni
Supervisor	Patrizia Zavattari

Final exam. Academic Year 2018 – 2019
Thesis defence: February 2020 Session


Eleonora Loi gratefully acknowledges Sardinian Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. - Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis III Education and training, Thematic goal 10, Investment Priority 10ii), Specific goal 10.5.

PUBLICATIONS PRODUCED FROM THIS THESIS PROJECT

Vega-Benedetti AF*, **Loi E***, Moi L, Blois S, Fadda A, Antonelli M, Arcella A, Badiali M, Giangaspero F, Morra I, Columbano A, Restivo A, Zorcolo L, Gismondi V, Varesco L, Bellomo SE, Giordano S, Canale M, Casadei-Gardini A, Faloppi L, Puzzoni M, Scartozzi M, Ziranu P, Cabras G, Cocco P, Ennas MG, Satta G, Zucca M, Canzio D, Zavattari P. Clustered protocadherins methylation alterations in cancer. Clin Epigenetics. 2019 Jul 9;11(1):100. doi: 10.1186/s13148-019-0695-0.

* equal contribution

Loi E, Moi L, Fadda A, Satta G, Zucca M, Sanna S, Amini Nia S, Cabras G, Padoan M, Magnani C, Miligi L, Piro S, Gentilini D, Ennas MG, Southey MC, Giles GG, Wong Doo N, Cocco P, Zavattari P. Methylation alteration of *SHANK1* as a predictive, diagnostic and prognostic biomarker for chronic lymphocytic leukemia. Oncotarget. 2019 Aug 13;10(48):4987-5002. doi: 10.18632/oncotarget.27080. eCollection 2019 Aug 13.

Loi E, Zavattari P. CpG Islands Methylation Alterations in Cancer: Functionally Intriguing Security Locks, Useful Early Tumor Biomarkers. In: Springer, editor. DNA, RNA, Histone Methylomes, 2019 p.53–62.

ABSTRACT

DNA methylation alterations are frequent early events in cancer. A global loss of DNA methylation and a focal hypermethylation of CpG islands (CGIs), mostly located at gene promoter regions, occur in cancer cells. Cancer DNA methylation changes present several features making them potential biomarkers for different applications, such as cancer risk definition, early tumour detection, tumours stratification, prognosis, prediction of therapy response and monitoring of disease evolution including early detection of tumour recurrence. Methylation alterations can be also trace in cell-free circulating tumour DNA, allowing their detection in non-invasive matrices such as urine, stool, serum or plasma. Although several studies have been focused on the identification of methylation-based biomarkers, very few of them have been introduced in clinical practice. Therefore, biomarkers of proven clinical utility that can support or even replace the current diagnostic methods of many cancers consisting in invasive procedures, such as tumour biopsy, are still lacking.

DNA methylation alterations are associated with changes in gene expression patterns but the relationship between these two mechanisms is still not clear and needs to be elucidated. In fact, while promoter hypermethylation of some genes, defined as tumour suppressor genes, is associated to gene downregulation, the majority of genes targeted for DNA methylation in cancer are genes expressed at background or very low levels in the normal tissues from which tumours originate. In this last case, some studies have not found a decrease in gene expression levels, while others have detected a further gene downregulation. Moreover, while hypermethylation of promoter-associated CGIs is usually associated with gene downregulation, hypermethylation of CGIs located at gene bodies have been either negatively or positively associated with gene expression.

Moreover, since DNA methylation is a reversible epigenetic change, it represents a potential cancer therapeutic target. In fact, methylation status can be edited and reverted to a DNA methylation pattern characteristic of normal cells, possibly restoring a normal cell phenotype.

This thesis project aims to identify new potential methylation-based biomarkers in different cancers, to investigate the association between methylation and gene expression and to test the application of CRISPR-dCas9 tool for targeted-methylation editing.

A genome-wide approach for the selection of the most informative biomarkers have been employed in all the tumours analysed, including two solid cancers, colorectal cancer (CRC) and biliary tract cancer (BTC) and one blood cancer, chronic lymphocytic leukaemia (CLL). Potential methylation biomarkers have been identified in all the cancers analysed and validated in publicly available datasets.

In colorectal cancer, the potential utility of two selected methylation alterations for CRC detection through non-invasive methods has been demonstrated.

In chronic lymphocytic leukemia, methylation of a CGI located in the gene body of *SHANK1* has been proposed as a biomarker for prediction of the disease years prior diagnosis, CLL diagnosis and prognosis.

Finally, specific-BTC methylation alterations showing high specificity and sensitivity have been identified in the BTC study.

The analysis of different types of cancer has allowed to discover that methylation alterations of CGIs associated with protocadherin (*PCDH*) gene cluster are frequent common events in different solid tumours but not in CLL.

Gene expression analyses have confirmed a negative correlation between hypermethylation and gene expression, also for the *SHANK1*-associated CGIs located in the gene body. Moreover, our results showed that normally low-expressed genes that are hypermethylated in cancer are further downregulated.

Finally, preliminary results of CRISPR-dCas9 tool confirmed its utility for targeted methylation editing and its possible use as a cancer therapeutic strategy.

INDEX

1. INTRODUCTION

1.1 DNA methylation

DNA methylation is a covalent modification of cytosines consisting in the addition of a methyl group to the fifth carbon atom of the pyrimidine ring leading to 5-methylcytosine (5mC). This modification predominantly occurs in the context of CpG dinucleotides defined as CpG sites. CpG density shows a bimodal distribution within the genome: high CpG density regions, termed as CpG islands (CGIs), usually overlap transcription start sites (TSSs), while the rest of the genome shows low CpG density. DNA methylation is a reversible process regulated by several proteins that write and erase DNA methylation patterns. Two DNA methyltransferases (DNMTs), DNMT3A and DNMT3B, are responsible for *de novo* DNA methylation, while DNMT1 needs a hemimethylated DNA and copies methylation patterns of the parental strand onto the newly synthetized strand during DNA replication. In the absence of a functional DNA methylation maintenance system, passive de-methylation during replication occurs. Active DNA de-methylation, consisting in the removal of methyl groups, is mediated by enzymes belonging to Ten-eleven translocation (TET) family. These enzymes oxidize 5mC to 5-hydroximethylcitosine (5hmC) and then 5hmC to 5-formylcytosine and 5-carboxylcytosine (Ito et al., 2011).

DNA methylation patterns of a new organism are established during development. In fact, after embryo implantation, DNA methylation patterns from gametes are erased and *de novo* methylation occurs at almost all CpG sites in the genome except for promoter-associated CGIs. Seventy-five percent of promoters are within CGIs (Bestor et al., 2015). These CGIs are somehow protected from DNA methylation, probably by the presence at these sites of the transcription machinery, H3K4 methyltransferases and MLL proteins (Smith and Meissner, 2013). The established bimodal DNA methylation pattern is maintained throughout all the subsequent cell divisions by DNMT1. During organogenesis, some genes become methylated in particular cell types or at particular developmental stages while others are de-methylated. The establishment of these patterns is finely regulated by proteins able to recruit DNMT and TET enzymes and are maintained for the whole organism life. Moreover, a stochastic process slowly alters the basal methylation pattern and the rate of methylation changes increases in function of age (Maegawa et al., 2010).

1.2 DNA methylation and gene expression

Epigenetic modifications, including DNA methylation, regulate gene expression profiles in different cell types and tissues.

Depending on the context where DNA methylation occurs, it affects gene expression in different ways. Methylation of intergenic and repetitive regions of the genome ensures repression of transposable elements that can have deleterious effects if activated. On the other hand, CGIs,

especially those associated with promoters of housekeeping genes, are normally not methylated. Unmethylated CGIs allow DNA accessibility and favor transcription factors binding, thus promoting gene expression.

CGIs methylation has been linked to stable silencing of gene expression. Some emblematic examples are genes located on the inactivate X-chromosome (Xi), imprinted genes and genes expressed in germinal cells that must be repressed in somatic cells. In the '80s, promoter-centric studies have investigated the correlation between methylation and gene expression, finding that promoter hypermethylation is associated to a decreased gene expression (Challen et al., 2012). These results contributed to the definition of the dogma associating CGIs hypermethylation with gene repression. However, the development of technologies to study genome-wide methylation patterns, have demonstrated that the relationship between methylation and gene expression is more complex. In fact, it can be expected that if methylation was linked to gene repression, promoter-associated CGIs would have to be methylated in a tissue specific manner. On the contrary, promoter-associated CGIs are rarely methylated and although decreased expression of some genes is associated with DNA methylation, the timing of CGIs methylation in respect to gene repression is not clear (Jones, 2012). The first theories suggested that DNA methylation directly induces gene silencing (Holliday and Pugh, 1975; Riggs, 1975). Recent studies have questioned these theories demonstrating that a transient reduction of gene expression induces the activation of silencing pathways involving firstly histone modifications and DNA methylation only in a second step (Oyer et al., 2009). Therefore, DNA methylation seems unnecessary to induce gene repression, but important to stabilize and maintain a transcriptional inactive state. The function of DNA methylation as a "lock" for gene repression, was already highlighted for genes located on Xi. In fact, Lock et al. showed that *Hprt* gene methylation occurs only after X-chromosome inactivation (Lock et al., 1987). The transcription factor *Oct3/4* is also silenced before being methylated during differentiation and hypermethylation seems to have a role in its stable repression (Feldman et al., 2006).

This "locking" model has been put into discussion by the finding that *Dnmt3A* is fundamental for hematopoietic stem cell (HSC) differentiation suggesting that it is involved in the epigenetic silencing of HSC regulatory genes (Challen et al., 2012).

However, the finding that the majority of genes targeted by DNA methylation in cancer cells are genes already repressed by Polycomb complex, adds support to the idea that methylation occurs after gene inactivation (Schlesinger et al., 2007; Gal-Yam et al., 2008).

Moreover, it has been demonstrated that cells expressing a catalytically inactive form of DNMT, DNMT3L, require the formation of a nucleosome, which is normally not present at active TSS, to recruit DNMT3A and *de novo* methylate DNA (Ooi et al., 2007). These nucleosomes flanking TSS

must not contain histone marks associated with active transcription such as H3K4me2 or H3K4me3 to allow DNA methylation.

Gene bodies are CpG-poor regions but are usually methylated in a tissue-specific fashion. It has been demonstrated that methylation of these intragenic CGIs does not block transcript elongation, but it is rather positively associated with gene expression. In fact, nucleosomes bearing H3K36me3 mark, associated with transcript elongation, recruit DNMTs promoting methylation of intragenic CGIs (Hahn et al., 2011). The role of intragenic methylation is still not clear, but it appears to have a role in several molecular processes such as alternative splicing (Shukla et al., 2011), expression of non-coding RNA (Saito et al., 2006; Lujambio et al., 2007) and transposable elements (Saied et al., 2012), regulation of intragenic alternative promoters (Maunakea et al., 2010) and intragenic enhancer activation (Schmidl et al., 2009).

In contrast to methylation of promoter-associated CGIs, which is linked to a decrease in gene expression, methylation of intragenic CGIs can be either be positively and negatively associated with gene expression.

Therefore, the correlation between methylation and gene expression changes is more complex than initially thought and the role of methylation in the regulation of gene expression must still be elucidated.

1.3 DNA methylation changes in cancer

Cancer cells undergo a series of early epigenetic alterations, including a dramatic change in DNA methylation patterns. A global loss of DNA methylation (hypomethylation) occurs and promoter-associated CGIs, which are usually unmethylated, are targeted for *de novo* methylation (focal hypermethylation) (Figure 1).

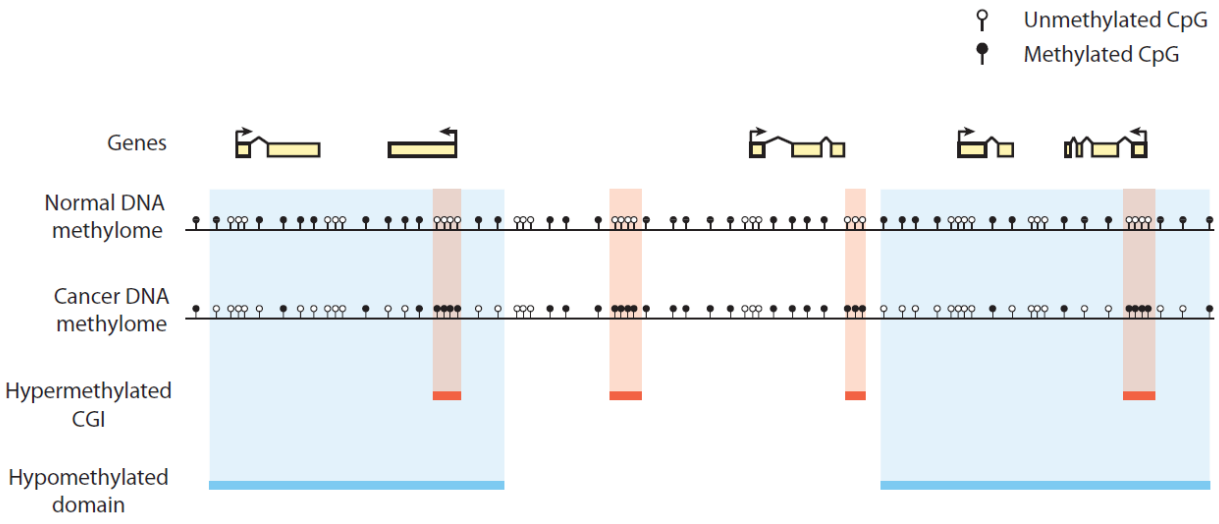


Figure 1. Changes in DNA methylation patterns in cancer
Figure from (Reddington et al., 2014)

1.3.1 Global hypomethylation

A cancer methylome is characterized by a wide hypomethylation, affecting intergenic regions including repetitive and transposable elements, gene deserts, introns and CpG-poor promoters. Loss of DNA methylation may be linked to re-activation of retrotransposons, repeats and oncogenes. Activation of repeats and retrotransposon may lead to chromosomal rearrangements and thus to genomic instability, one of the tumour hallmarks (Jones and Baylin, 2007).

DNA hypomethylation is not restricted to retrotransposons. In fact, whole genome bisulfite sequencing technologies have revealed that it is actually concentrated in genomic blocks larger from 28kb to 10Mb (Hansen et al., 2011; Berman et al., 2012; Hon et al., 2012). In normal embryonic stem (ES) and differentiated cells, these large genomic domains, termed as partially methylated domains, are heavily methylated (about 80% methylation) across different tissue types. In cancer, these genomic blocks lose about 20-40% of methylation, leading to hypomethylated domains(Hansen et al., 2011; Berman et al., 2012; Bert et al., 2013), usually overlapping with lamina-associated domains at nuclear periphery (Berman et al., 2012; Timp and Feinberg, 2013).

Although the exact mechanism by which DNA methylation is lost in cancer cells has not been fully elucidated, it can be consequent to deregulated activation of TETs or partial loss of function of DNMTs.

1.3.2 Focal hypermethylation

Aberrant methylation of CGIs within gene promoters is the other feature of a cancer methylome.

Two models explain DNA hypermethylation in cancer: according to the classical models, a tumour can originate in any cells of the organ and *de novo* methylation occurs as a result of cancer transformation; while, according to the alternative model, these methylation aberrations, originating during the process of aging, are already present at some levels in the founding cells that will give rise to tumours and are selected and clonally expanded during transformation (Nejman et al., 2014). This last model agrees with the "cancer stem cells (CSCs) model" suggesting that the epigenetic changes occurring in normal stem or progenitor cells are the earliest event in tumorigenesis, preceding cancer mutations (Feinberg et al., 2006). DNA hypermethylation and repression of genes involved in the regulation of stem cells self-renewal capacity, such as *p16* and *APC,* have been observed in different cancers (Jones and Baylin, 2007). Silencing of these genes may cause the acquisition of an infinite renewal capacity of the affected cells. These immortal cells, selected and expanded, undergo further genetic mutations contributing to tumour development (Jones and Baylin, 2007). These CSCs, which show similar self-renewal and multipotency features of adult stem cells, are responsible for tumour initiation and growth in contrast to the bulk of tumour cells with a differentiated phenotype (Visvader, 2011).

The promoter-centric studies of the '80s have led to the discovery that promoters of many tumour-suppressor genes (TSGs) are hypermethylated in cancer contributing to the theory that DNA methylation is responsible for gene silencing and can act as a driver for neoplastic transformation (Herman and Baylin, 2003). Several genes implied in apoptosis, cell cycle regulation, cell adhesion, DNA repair and angiogenesis are in fact hypermethylated and downregulated in many hereditary cancer syndromes as well as in sporadic cancer. In hereditary cancer syndromes, these genes are mutated predisposing to cancer in specific tissues. DNA methylation acts as a "second hit" (Knudson's hypothesis) for the complete inactivation of the affected genes (Grady et al., 2000). In sporadic cancer, these genes show tissue-specific hypermethylation and phenocopy the equivalent genetic mutations. Some emblematic examples are *MLH1* hypermethylation in colorectal cancer (Herman et al., 1998) or *RB1* in retinoblastomas (Ohtani-Fujita et al., 1997). In other tumours, such as breast cancer, this correlation is less clear. In fact, although *BRCA1* mutated individuals develop estrogen receptor negative (ER-) tumours, *BRCA1* hypermethylation has been observed in both ER+ and ER- tumours (Turner et al., 2007).

Moreover, treatment of cells with a demethylating agent, 5-azacytidine (5-AZA), caused gene reactivation, supporting the hypothesis that methylation causes gene silencing (Baylin and Jones, 2016). However, these studies did not examine the temporal sequence of events leading to gene silencing. In fact, although 5-AZA treatment is also able to cause re-activation of genes in the inactive X chromosome, it is know that silencing of genes in Xi precedes their methylation (Lock et al., 1987). TSGs that are expressed before cancer transformation and become hypermethylated and repressed in cancer are defined as epigenetic drivers (Kalari and Pfeifer, 2010; Baylin and Jones, 2011). However, their repression may not be consequent to hypermethylation but rather DNA methylation may follow gene downregulation by other means.

The majority of hypermethylated genes in cancer are normally repressed and marked by Polycomb complex (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). These genes are repressed from embryogenesis by trimethylation on lysine 27 on histone H3 (H3K27me3) and their epigenetic silencing is maintained by a histone-lysine methyltransferase, called Enhancer of Zeste homolog 2 (EZH2), constituting the Polycomb repressive complex 2 (PRC2). In normal cells, the removal of the complex and of this histone mark allows the expression of these genes during differentiation to the mature cell type. In fact, genes targeted by PRC are usually transcription factors involved in lineage commitment (Bracken et al., 2006). On the other hand, in cancer-prone progenitor cells, the recruitment of DNMTs leads to *de novo* methylation of these sites (Viré et al., 2006). This tumour "epigenetic switching" allows a stable repression of these genes and leads to a permanent proliferative state of these cells unable to differentiate (Gal-Yam et al., 2008). It has been

hypothesized that hypermethylation of PRC-targets maintains cancer cells in a stem-cell-like aggressive state (Widschwendter et al., 2007). However, cancer hypermethylator phenotypes are usually associated with a better prognosis. Therefore, a stable gene silencing by hypermethylation may be a mechanism of inhibition against tumour progression (Sproul and Meehan, 2013). It is possible that hypermethylation may have different roles in different cancers and subtypes as well as different stages of tumour development, initially promoting tumour growth and later preventing metastasis (Sproul and Meehan, 2013).

Since many genes hypermethylated in cancer are frequently repressed prior methylation, it is not surprising if the majority of cancer studies has found low correlation between promoter hypermethylation and gene expression. A recent meta-analysis of a large dataset of 672 matched cancerous and healthy methylomes, gene expression, and copy number profiles across 3 types of tissues from The Cancer Genome Atlas (TCGA), showed that cancer-promoter hypermethylation was not linked to a decreased expression of the affected genes (Moarii et al., 2015). Gene expression analysis of *RUNX3*, a gene frequently hypermethylated in gastric cancer, revealed that is never expressed in normal gastrointestinal epithelial cells supporting the hypothesis that genes aberrantly hypermethylated in cancer are fully repressed in the normal cells where tumours originate (Levanon et al., 2011).

It has also been suggested that genes targeted by hypermethylation in cancer are not fully repressed in the normal cells but expressed at low levels (Berman et al., 2012). A colon cancer study found that only 7% of genes downregulated compared to adjacent normal tissues were also hypermethylated in CpG island methylation phenotype (CIMP) negative tumours and a proportion of these genes was also downregulated in CIMP+ colorectal cancer (CRC) samples not showing DNA hypermethylation (Hinoue et al., 2012). However, the majority of gene expression studies have been performed using gene expression microarray and the background levels of hybridization to probes do not to allow to capture small changes in gene expression (Sproul and Meehan, 2013).

In fact, a colon cancer study conducted in our laboratory has shown that a transcriptome analysis by gene expression arrays of CRC and normal tissues did not show any dysregulation of genes whose promoter-associated CGIs were hypermethylated in CRC. On the other hand, the employment of a more sensitive method, such as qRT-PCR, revealed a significant downregulation of the tested genes in CRC samples compared to normal samples (Fadda et al., 2018).

Moreover, another recent study has shown that methylation is significantly associated with gene downregulation of a subset of genes enriched in common cancer pathways in CRC (Klett et al., 2018). Environmental factors, such as local tissue inflammation and oxidative stress, may also contribute to the establishment of tumour methylation pattern (Niwa and Ushijima, 2010; O'Hagan et al., 2011).

Finally, genetic mutations of genes involved in DNA methylation process, such as TET mutations in myeloid malignancies (Abdel-Wahab et al., 2009) and DNMT3A mutation in acute myeloid leukemia, contribute to some methylation alterations occurring later in cancer (Shlush et al., 2014).

1.4 Common cancer pathways affected by DNA methylation

DNA methylation aberrations cause the disruption of many important signaling pathways in cancer. The main pathways affected by DNA methylation in cancer are: DNA repair and genomic stability, RB1/CDK4 cell cycle regulation, WNT/β-catenin, TGF-β, cellular differentiation, induction of apoptosis and cell growth pathways.

In several types of cancer, control of cell cycle regulation is lost due to hypermethylation and downregulation of cyclin-dependent kinase inhibitor 2A (*CDKN2A*), encoding for a CDK inhibitor protein ($p16^{INK4A}$) (Morsczeck et al., 2018). p16 protein regulates the transition of the cell cycle from G1 to S phase by inhibiting RB1 phosphorylation mediated by cyclin-dependent kinases CDK4 and CDK6. Lack of p16 expression by deletion or promoter hypermethylation of *CDKN2A* p16 allows the cell to bypass the G1/S checkpoint.

Another signaling pathway commonly dysregulated in cancer, most notably in colorectal cancer, is WNT/β-catenin pathway. DNA promoter hypermethylation and silencing of genes encoding for one class of WNT inhibitors, secreted frizzled-related proteins (SFRPs) (Surana et al., 2014) or *APC* gene (Liang et al., 2017) represent two common mechanisms to constitutively activate WNT signaling in cancer.

DNA repair and genomic stability mechanisms are loss in several cancers. In sporadic colorectal cancer, hypermethylation of *MLH1*, a DNA mismatch repair gene, is very frequent and increases genome susceptibility to mutations and microsatellite instability (Poynter et al., 2008). In ovarian and breast cancers, *BRCA1* inactivation by hypermethylation lead to impaired DNA repaired by homologous recombination and consequent genomic instability (Catteau et al., 1999). *MGMT* gene encoding for O6-methylguanine methyltransferase has been found hypermethylated and downregulated in several cancers, most prominently in gliomas. *MGMT* silencing leads to a diminished efficiency of O6-alkylguanine repair (Poynter et al., 2008).

Hypermethylation of proapoptotic genes, such as death-associated protein kinase (*DAPK*) and caspase 8 (*CASP8*) genes has been reported in many tumours (Hervouet et al., 2013).

Methylation of the *RASSF1A* gene encoding for one member of Ras association domain family proteins can be detected in almost all cancers including pre-neoplastic lesions. These proteins positively regulate Hippo growth control pathways (Richter et al., 2009).

The most frequently methylated genes in cancer are homeobox gene family members. These genes encode for transcriptional factors important during development and cell and organ differentiation.

These genes are targeted by PRC1 and are lowly expressed in adult tissues. In cancer, methylation of these genes ensures a permanent silencing of PRC target genes maintaining cells in a undifferentiated state (Rodrigues et al., 2016).

Many genes encoding for ion channel and transporters, aquaporins, membrane receptors and cell adhesion molecules have been found hypermethylated in cancer (Nunna et al., 2014; Lastraioli et al., 2015; Khatami et al., 2017; Dajani et al., 2018; Xie et al., 2018).

A gene enrichment analysis of 74 CGIs aberrantly methylated in CRC and adenomas in our study, identified genes encoding for solute transporters, G-protein coupled receptors, protocadherins, integrins, confirming that the pathways most affected in cancer are involved in the crosstalk between tumour cells and surrounding environment (Fadda et al., 2018).

1.5 DNA methylation: a promising cancer biomarker

After the discovery of DNA methylation aberrations in cancer, a multitude of studies have started to consider DNA methylation-based biomarkers for several applications such as: cancer risk stratification, diagnosis, prognosis, prediction of therapy response and monitoring of disease.

A biomarker is any biological characteristics that can be objectively measured as an indicator of a normal biological process, a pathogenic process or a response to an exposure or an intervention (Naylor, 2003).

In fact, cancer-specific DNA methylation changes present several features making them promising useful biomarkers. Firstly, these methylation alterations are early and frequent events in cancerogenesis, some of them even present in precancerous lesions, making them potential biomarkers for early cancer diagnosis. Moreover, they can be also found in cell-free circulating tumour DNA (ctDNA) in various body fluids making possible their detection trough non-invasive methods, usually termed as liquid biopsies. Finally, DNA methylation is a stable epigenetic mark and a multitude of well-established techniques can be used for its detection. In fact, it can be detected even in Guthrie neonatal blood spots, formalin-fixed paraffin-embedded (FFPE) samples and microscopic preparations (Mikeska and Craig, 2014).

A clinically significant difference in DNA methylation patterns must be present between two groups of interest to consider a DNA methylation alteration as a useful biomarker. The ideal biomarker should have 100% specificity and 100% sensitivity. Specificity describes the proportion of individuals without the disease having a negative test results, while sensitivity describes the proportions of patients having a positive test result (Mikeska and Craig, 2014).

Although, an immense number of studies suggesting DNA methylation-based candidate biomarkers have been published, few are commercially available and have been included in clinical guidelines

and only two (Cologuard and Epi proColon) have been approved by the Food and Drug Administration (FDA) to be introduced in clinical practice for CRC screening (Koch et al., 2018).

1.5.1 Risk stratification

Mutational inactivation of some genes can predispose to a variety of cancer. These genes have been defined as hereditary cancer genes, among which *RB1*, *MLH1*, *BRCA1* and *APC* genes are some emblematic examples. It has been found that these same genes undergo methylation in the corresponding sporadic form of cancer.

Methylation of one allele of these genes can occur in a mosaic form in the normal somatic cells of one or more tissues, an event defined as constitutional methylation (Mikeska et al., 2012).

One example is represented by constitutional methylation of *MLH1* gene, a situation predisposing to colorectal cancer. In fact, cases of early-onset colorectal cancer, had one of the alleles methylated in normal tissues and loss of the normal allele in the tumour (Gazzoli et al., 2002).

Constitutional methylation can be detected in peripheral blood leukocytes. In fact, *BRCA1* methylation has been found in blood samples of more than 30% of women with *BRCA1*-like cancer and only in a small percentage of other breast cancer-type patients and controls (Wong et al., 2011).

Moreover, methylation alterations have been detected in pre-diagnostic blood samples more than 10 years before diagnosis of mature B-cell neoplasms (Wong Doo et al., 2016; Georgiadis et al., 2017).

Therefore, a population screening through non-invasive methods analyzing peripheral blood samples, may allow to predict patients at risk of early-onset cancer forms. Consequently, monitoring of these individuals and specific interventions could be applied to prevent tumour development in these subjects.

1.5.2 Early cancer detection

The development of screening tests for early cancer diagnosis is a major goal of cancer research. In fact, many types of cancer are diagnosed in advanced stages when therapeutic interventions or surgical removal of the tumour are no longer an option and patient's outcomes are often poor. DNA methylation represents a promising biomarker for early cancer diagnosis since it occurs in the first stages of the cancerogenesis and can be detected even in precancerous lesions.

Moreover, since DNA methylation can be detected in several body fluids, such as blood, urine, sputum and stool samples, it represents a new non-invasive screening method that may potentially support or even replace invasive diagnostic methods. In fact, the gold standard for cancer diagnosis is tissue biopsy, a highly invasive method that cannot always be applied because of the anatomical location of the tumours or the high risk of post-biopsy infection.

1.5.2.1 Colorectal cancer: an emblematic example

The gold-standard diagnostic method for colorectal cancer is colonoscopy that allows the detection and the following removal of the precursor benign lesions of the tumour, i.e. adenomas or polyps. However, this method is highly invasive, costly and has scarce compliance (Taylor et al., 2011). Less-invasive screening methods include fecal occult blood test (FOBT) and fecal immunochemical test (FIT). Although these tests offer possible indications for CRC, their specificity and sensitivity are still limited (Morikawa et al., 2005). For this reason, screening tests based on the analysis of molecular biomarkers in stool and plasma samples have been developed. Currently, carcinoembryonic antigen (CEA) and CA19-9 are the two biomarkers used in clinical practice for colorectal cancer diagnosis but present scarce accuracy (Tham et al., 2014).

Three DNA-methylation based blood tests are commercially available for colorectal cancer detection: Epi proColon (Epigenomics), ColoVantage (Quest Diagnostics) and RealTime mS9 (Abbott) (Payne, 2010). All these three tests are based on *SEPT9* methylation analysis but only Epi proColon has been recently approved by FDA for the detection of colorectal cancer (Mikeska et al., 2012).

SEPT9 promoter hypermethylation detection in plasma samples has shown improved specificity and sensitivity compared to the available FOBT and FIT (Molnár et al., 2015).

However, the effects of demographic characteristics and pathological features on *SEPT9* methylation need to be still evaluated. In fact, positive results have been found in patients with other cancers and in controls. Finally, the power of *SEPT9* methylation-based test on detecting adenomas and early stage CRCs is limited (Wang et al., 2018).

Commercial tests based on the detection of methylation biomarkers in stool samples have been also developed. In fact, since colonocytes are constantly shed from the tumour into the lumen and passed into the stools, these samples represent a good source for detecting CRC-specific biomarkers. The first DNA methylation-based stool test commercially available was ColoSure (LabCorp), evaluating *VIM* methylation. This test showed a specificity between 82 and 100% and a sensitivity of 33-81% and 15-45% for CRC and adenomas respectively (Ned et al., 2011; Amiot et al., 2014).

Multi-biomarkers panels represent more reliable biomarkers to capture tumour heterogeneity.

FDA has approved Cologuard (Exact Science) for colorectal cancer detection. This kit detects hypermethylation of *NDRG4* and *BMP3* genes, *KRAS* mutations, beta actin and includes a hemoglobin immunoassay. A large study determined that this test has a sensitivity for adenomas (42%) higher than FIT, and a specificity of 87% (Imperiale et al., 2014).

Our research group has recently identified 74 early methylation alterations shared in both CRC and adenomas samples (Fadda et al., 2018). To explore the possible use of these biomarkers for cancer detection through non-invasive methods, three selected CGIs, associated with *GRIA4*, *SLC8A1* and *SYN3* genes, have been selected to detect their methylation levels in DNA extracted from stool

samples collected at the time of tumour surgical resection in CRC patients. 87.5% of analysed patients showed more than 1% of methylation for at least one of the three biomarkers. Therefore, our panel guaranteed an overall sensitivity of 87.5%, even higher of other multi-biomarker panels, such as one of three selected markers (*AGTR1*, *WNT2*, *SLIT2*) that reached a sensitivity of 78%, based on the criteria that at least one of marker was methylated (Carmona et al., 2013).

1.5.3 Prognosis and prediction of therapy response

Prognostic biomarkers provide information about patients' overall survival before and after tumour resection. Methylation alterations have often shown correlation with patients' survival and have been proposed as prognostic biomarkers. Moreover, patients' stratification based on molecular markers is important to define cancer therapy strategies. In fact, select patients who can positively respond to a specific cancer treatment may increase treatment efficacy, diminish toxicity and consequently increase patients' overall survival. Finally, provide the treatment only to patients who can benefit from a specific therapy would allow a reduction of such treatment's costs.

Although many candidate methylation-based biomarkers for cancer prognosis and prediction of response have been proposed very few have been implemented into clinics. One of these tests, Predict MDx Brain Cancer (MDxHealth), based on evaluation of *MGMT* methylation, allows prediction of response to alkylating agents in glioblastoma patients. In fact, patients with promoter hypermethylation and inactivation of *MGMT*, show increase survival when treated with temozolomide (Hegi et al., 2005).

1.5.4 Monitoring of patients

After surgical removal of the tumour and/or during cancer treatments, patients should be monitored to assess surgery outcome, predict tumour recurrence and check response to therapy. Classification of patients at risk of recurrence would allow to intensify routine monitoring and specific biomarkers can allow to detect patients experiencing a relapse. Biomarkers for early treatment response would allow an immediate adjustment of treatment regimens.

Marker investigations at multiple time points can provide information for patient management. Since serial sampling of the tumour is not feasible because of the invasiveness of the procedure, analysis of ctDNA in different body fluids would allow patients' monitoring through a non-invasive procedure.

Changes in DNA methylation patterns represent useful biomarkers for these purposes. ctDNA present the same genetic and methylation alterations of the solid tumours and thus can be distinguished from circulating DNA from healthy cells. Furthermore, since ctDNA in the blood has a half-life of about 2 hours (Diehl et al., 2005), serum and plasma can provide a real-time measure of the tumour status. The levels of ctDNA in plasma depends on the location, size and vascularity of the tumours. The persistence of ctDNA after surgical removal of the tumour, reflects residual tumour tissue in the

body and is linked to prognosis (Diehl et al., 2005). A list of tumour-specific methylated genes has shown to decrease after surgery. Some example include: *CDKN2A* in plasma samples of hepatocarcinoma patients (Wong et al., 2003), *APC* in serum samples of esophageal cancer patients (Hoffmann et al., 2009), *RARb2*, *MSH2* and *ESR1B* genes in plasma samples of breast cancer patients (Liggett et al., 2011) and *RUNX3* in serum of gastric cancer patients (Sakakura et al., 2009).

Methylation changes including a decrease or an increase in methylation of some genes in serum or plasma samples has been shown to be indicator of therapy response. Similarly, methylation detection of specific biomarkers may allow traceability of minimal residual disease and prediction of recurrence. One commercially available test for early CRC recurrence based on methylation analysis of *BCAT1* and *IKZF1*genes is Colvera (Clinical Genomics) with sensitivity and specificity of 73.1% and 89.3%, respectively (Murray et al., 2017).

Our laboratory has also assessed methylation levels of the three biomarkers selected for detection in in ctDNA isolated from plasma samples of CRC patients. One group of patients were receiving post tumour resection adjuvant therapy and showed no radiological evidence of disease (NED), and another group of patients was still bearing a lesion. Median methylation of *GRIA4* and *SLC8A1* was significantly higher in metastatic patients compared to NED patients confirming the high specificity of these two markers (Fadda et al., 2018). Patients still bearing a lesion were divided according to CEA levels (threshold 5 ng/ml) into CEA-high and CEA-low. *GRIA4* and *SLC8A1* showed significantly higher methylation levels in CEA-high patients compared to NED patients. Moreover, NED patients with high levels of CEA did not show methylation for the tested markers, suggesting that methylation alterations displayed a higher negative predictive value than CEA (Fadda et al., 2018).

Another study has recently shown that a multi-marker panel including *EYA4*, *GRIA4*, *ITGA4*, *MAP3K14-AS1* and *MSC* can be used to monitor tumour burden in liquid biopsy in CRC patients under different therapy regimens (Barault et al., 2018).

1.5.5 Topographical biomarkers

In some cancers, anatomical location of the tumour can influence treatment strategies, degree of resection and prognosis. For example, tumours in the proximal colon (right side) and distal colon (left side) display different molecular alterations and histology. These differences affect both therapeutic choice and prognosis. In fact, left-sided CRC patients are treated with adjuvant chemotherapies and targeted therapies and usually have better prognosis; while right-sided CRC patients show low response rates to conventional chemotherapies and are treated with immunotherapy (Baran et al., 2018).

In brain tumours, the identification of location-specific alterations may provide potential biomarkers predictive of tumour behavior.

Several studies have identified different gene expression and methylation profiles in brain tumours having different localizations (Wong et al., 2005; Sharma et al., 2007; Palm et al., 2009; Tchoghandjian et al., 2009; Lambert et al., 2013; Bergthold et al., 2015; Zakrzewski et al., 2015; Jeyapalan et al., 2016; Sexton-Oates et al., 2018).

A methylome analysis of pilocytic astrocytoma (PA) samples conducted in our laboratory has confirmed that different methylation patterns characterized tumours from two different localizations (supratentorial and infratentorial localization) (Antonelli et al., 2018). Moreover, two genes, *IRX2* and *TOX2*, whose associated CGIs were hypermethylated in supratentorial PAs compared to infratentorial PAs, showed also a decreased expression in supratentorial PAs. However, while expression levels of *IRX2* were in line with those observed in normal brain (i.e. higher expression in infratentorial localization), *TOX2* was normally more expressed in supratentorial brain region. These results suggested that *IRX2* lower expression in supratentorial PAs is probably more a reflection of the brain region gene expression pattern where the tumour originates while *TOX2* decreased expression in supratentorial tumours may be related to tumour development (Antonelli et al., 2018).

Other studies had already suggested that different methylation and gene expression profiles among tumours from different brain regions may be related to the specific brain site where the tumour arises (Sharma et al., 2007). Therefore, *IRX2* may have been targeted by DNA methylation in supratentorial tumours because of its low expression in this brain location and it may be a potential topographical biomarker. This hypothesis is supported by the discovery of methylation alterations that may be potential markers of cell lineages in other cancers. Sproul et al. (Sproul et al., 2011) have identified several genes showing a significant association between methylation and gene expression in different breast cancer cell lines. However, the majority of these genes were already repressed in normal cells of the same lineages (Sproul et al., 2011). In another study, the analysis of data derived from more than one thousand tumours arising in seven different human tissues confirmed that hypermethylation of genes repressed in the normal tissues of tumour origin is a common event in different cancers (Sproul et al., 2012).

In conclusion, in many cases, aberrant methylation observed in cancer is a marker of cell lineage rather than tumour development.

1.5.6 Rules to select good methylation-based biomarkers

Despite an increasing number of cancer studies is focused on identifying methylation-based biomarkers, only few have been introduced into clinical practice.

Before clinical implementation, a biomarker needs to fulfil a series of parameters. In fact, a biomarker identified in a single study, defined as potential biomarker, needs to be validated with a different method and replicated in an independent cohort. Subsequently, after a systematic review and meta-analysis of multiple studies, they can enter into clinical trials as candidate clinical biomarkers and, if they have proven a clinical benefit, are approved and become proven clinical biomarkers (Mikeska and Craig, 2014).

Very few potential biomarkers pass all these steps from initial discovery to clinical implementation. The reasons include lack of validation and/or replication, lack of clinical utility, lack of consensus for an appropriate methodology for DNA methylation detection and lack of clinical relevance of the genomic location of DNA methylation (Koch et al., 2018).

Definition of the region where DNA methylation occurs is very important. In cancer-biomarker discovery studies, genomic regions, usually CGIs, showing a statistically significant difference between two groups of interest, for example between tumour and control samples, are selected. In the majority of these studies, only few CpG sites covering these regions are analysed. In fact, methylation analysis of all single CpG sites is rarely suitable and the CpG dinucleotides analysed are considered representative of the methylation status of the entire region. Furthermore, definition of a baseline to consider a sample positive for methylation of the interrogated sites needs to be defined. In fact, apart from some promoter CGIs, the majority of CGIs are methylated at background levels also in healthy individuals. Therefore, a good methylation-based biomarker should be unmethylated or methylated at very low levels in normal samples and highly methylated in tumours (Mikeska et al., 2012).

Another important criterion that a good biomarker must fulfil is cell-type specificity. A study has detected methylation of many potential cancer methylation-based biomarkers in normal peripheral blood mononuclear cells of healthy individuals (Kristensen et al., 2012), highlighting the importance of verify biomarker cell-type specificity.

The optimal number of CpG sites to interrogate in the methylation assay should be critically evaluated. In fact, while in some cases methylation interrogation of one CpG site may be sufficient to discriminate with high power tumour and control samples, in other cases multiple CpG sites must be analysed. Technical limitations, such as PCR primer design, sometimes contribute to the definition of the region to analyze (Mikeska et al., 2012).

It has also been suggested that the genomic context of a potential methylation-based biomarker is important for its clinical value (Koch et al., 2018). The simplest case is hypermethylation of a promoter-associated CGI linked to gene downregulation. However, some studies have shown that methylation alterations occurring outside promoters may have a clinical value (Hu et al., 2014; Kang et al., 2015). Moreover, since DNA methylation targets genes that are already repressed or expressed

at very low levels in tissues where tumour occurs, this alteration is not always associated to a measurable change in gene expression levels.

A systematic procedure for methylation-based biomarker development has been proposed (Bock, 2009). This multi-step approach involves the use of genome-wide DNA methylation analysis method for the initial discovery phase in order to maximize genomic coverage. Several methods allow a global analysis of DNA methylation such as Illumina arrays, of which last version (Infinium MethylationEPIC BeadChip) allows to analyze more than 850 000 CpG sites, and whole-genome bisulfite sequencing (WGBS). The analysis of such massive amount of data consists in the application of computational methods with the final aim to select a small number of DNA methylation alterations that show high predictive power for the disease condition of interest and thus may represent potential biomarkers. The performance of these candidate biomarkers needs to be validated in a second independent cohort using targeted assays. This step involves the selection of the most informative CpG sites representative of the entire region methylation pattern. In fact, methylation pattern of adjacent CpG sites is highly correlated. Several robust and cost-efficient methods for a targeted methylation analysis of selected regions exist, such as Pyrosequencing, MethyLight, quantitative methylation specific PCR (qMSP) and clonal bisulfite sequencing (Bock, 2009).

Moreover, the deposition of a large amount of molecular and clinical data of different cancers publicly available allows validation of the potential biomarkers identified in larger datasets in an easy and cost-free way (Koch et al., 2018). The main databases are The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov) and International Cancer Genome Consortium (ICGC) (https://icgc.org). Moreover, raw and elaborated data of thousands of published studies are available in the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/gds).

After this validation step, employment of statistical classifier models and performance prediction models allows the selection of the most promising candidates for confirmation in additional independent validation cohorts. Finally, if these biomarkers are validated in several cohorts, a prospective study can be conducted to prove the clinical utility of these biomarkers (Bock, 2009).

1.6 DNA methylation as a potential therapeutic target: editing DNA methylation

DNA methylation is a reversible epigenetic modification. Therefore, aberrant methylation alterations observed in cancer may represent potential therapeutic targets. Several DNMT inhibitors (DNMTi) have been developed for this purpose. Two DNMTi, 5-azacytidine (5-Aza-CR, Vidaza) and 5-aza-2'-deoxycytidine (5-Aza-CdR, decitabine, Dacogen), have been approved by FDA for treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) patients (Christman, 2002; Issa et al., 2005). A new epigenetic inhibitor, S110 (AzapG), has shown promising results for treatment of both MDS and AML as well as for some solid cancers (Chuang et al., 2010).

Although these drugs have shown clinical efficacy, their mechanisms of action are still not clear. Studies on patients treated with these drugs have shown global gene-specific demethylation and demethylation of repetitive elements such as Alu and long interspersed nucleotide elements (LINEs) (Yang et al., 2006). Long-term effects after interruption of treatment on specific genes and repetitive elements were different. In fact, repetitive elements are quickly re-methylated after some days, while demethylation of some genes can persist even over several weeks (Yang et al., 2006; Kantarjian et al., 2007).

DNMTi have global genome effects and do not allow a targeted methylation of the locus of interest. Epigenome editing represent a promising approach for a personalized cancer therapy since it allows a specific methylation/demethylation of the target of interest (Vojta et al., 2016).

A tool for epigenome editing requires a DNA-binding targeting domain and a functional domain. The first tools employed zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) as DNA-binding targeting domains (Li et al., 2007; Bernstein et al., 2015). These two have been exceeded by CRISPR-Cas9 system that can be modified to be employed for genome and epigenome editing with the possibility of multiple guide RNAs (gRNAs) designed for the targets of interest and insensitivity to CpG methylation in contrast to TALEs (Perez-Pinera et al., 2013). In this system, a modified version of a bacterial adaptive immune system, Clustered regularly interspaced palindromic repeats (CRISPR), has been adapted to target the Cas9 nuclease to the genomic site of interest through the design of sequence-specific guide RNAs (Mali et al., 2013). The first twenty nucleotides at 5' end of the gRNAs ensures the binding of the gRNAs to the complementary target sequence only if is followed by a protospacer-adjacent motif (PAM) (Mojica et al., 2009). A catalytically inactive form of Cas9 (dCas9) was generated to bring effector proteins to the site of interest without cutting the DNA molecules. The functional domain is the catalytic domain of a DNMT for targeted methylation or TET functional domain for targeted demethylation (Figure 2). The most used DNMT domain is that of DNMT3A, which is a *de novo* DNMT that has shown enzymatic activity also in transfected cells. DNMT3A domain combined with DNMT3L can increase DNMT3A enzymatic activity (Chen et al., 2005).

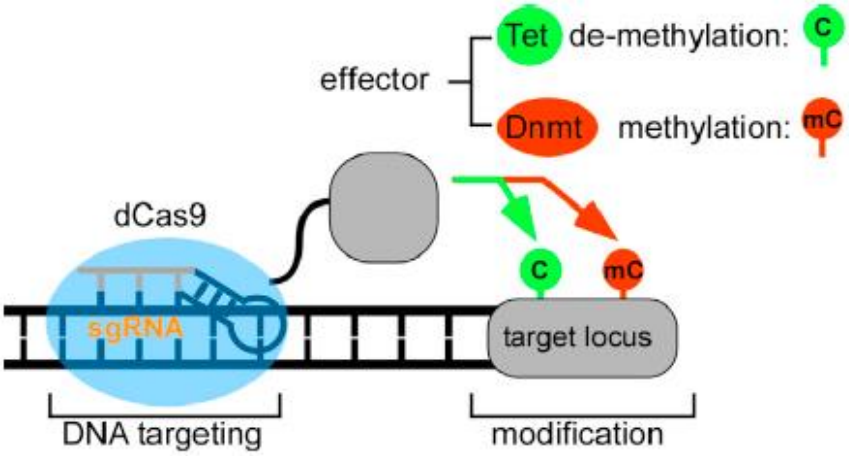


Figure 2. Schematic representation of the modified CRISPR-Cas9 system for methylation editing
Figure adapted from (Liu et al., 2016)

Several variants of this system have been developed and have shown encouraging results for DNA methylation editing (Amabile et al., 2016; Choudhury et al., 2016; Liu et al., 2016; McDonald et al., 2016; Xu et al., 2016; Lei et al., 2017; Saunderson et al., 2017; Kang et al., 2019). Vojta et al. (2016) realized a system in which dCas9 is fused with DNMT3A catalytic domain though a flexible Gly4Ser linker. They demonstrated a targeted methylation of a genomic region of about 35 bp and of a wider region using multiple gRNAs (Vojta et al., 2016). Liu et al. (2016) were able to edit DNA methylation both *in vitro* and *in vivo* using dCas9 fused with Tet1 or Dnmt3a (Liu et al., 2016). Efficient targeted DNA methylation was also obtained in both human cells and in mouse embryos by fusing dCas9 with an engineered prokaryotic DNA methyltransferase MQ (Lei et al., 2017). Amabile et al. (2016) have developed a system for targeted gene repression by fusing engineered transcriptional repressors (ETRs) with the Kruppel associated box containing zinc-finger proteins (KRAB-ZFPs) and DNMT3A. They demonstrated that gene silencing was highly specific and only due to DNA demethylation (Amabile et al., 2016). Long-term targeted DNA methylation and silencing of multiple genes in in primary breast cells isolated from healthy human tissue were obtained by transiently transfecting a fusion plasmid containing the catalytic domain of Dnmt3a and C-terminal domain of Dnmt3l (3A3L) coupled to a dCas9 (Saunderson et al., 2017; Stepper et al., 2017).

1.7 Technologies to identify methylation alterations

1.7.1 PCR-based methods

DNA methylation alterations have been initially discovered using a candidate-gene approach.

The most diffuse methods for detecting DNA methylation at single loci are PCR-based methods using bisulfite-converted DNA as template. Bisulfite conversion allows to maintain DNA methylation information that would be otherwise lost during PCR amplification. In fact, DNA polymerase is not

able to distinguish methylated cytosines from unmethylated ones and guanines would be incorporated in the newly synthetized DNA filament in both situations. Treatment with bisulfite sodium deaminates unmethylated cytosines to uracils, while methylated cytosines remain unaffected. At this point, the sense and antisense DNA strands are no longer complementary and PCR primers are designed for one strand. During subsequent PCR cycles, uracils are replaced by thymines. Several commercial kits have been developed to perform DNA bisulfite conversion and conversion rates are usually higher that 99% depending on DNA quality (Warnecke et al., 2002).

Two strategies can be used for PCR primer designing of bisulfite-converted DNA: primers can be designed on DNA sequence without CpG sites to allow a PCR amplification independent from methylation status (methylation-independent PCR or MIP) or specific primers for methylated and unmethylated sequences can be designed to perform a methylation-specific PCR (MSP).

MIP methods have usually a bias towards unmethylated sequences because of the difference in GC content after bisulfite conversion. Inclusion of few CpG sites in primer sequences and optimization of annealing temperatures may overcome the PCR-bias for unmethylated sequences (Wojdacz and Lotte Hansen, 2006; Shen et al., 2007).

To obtain information of methylation status at the level of single CpG site, PCR products are traditionally sequenced. Sequencing PCR clones instead of directly sequencing amplified DNA has the advantage to provide information on individual molecules and to calculate the ratio of methylated to unmethylated molecules. Unfortunately, this method is too time-consuming to be introduced in clinical practice. Another approach, digital bisulfite genomic sequencing, requires the use of multiple PCR reactions and drastic sample dilution to minimize the occurrence of more than two PCR template molecules in each reaction well(Weisenberger et al., 2008).

Traditional Sanger sequencing can be replaced by pyrosequencing based on the detection of pyrophosphate. In fact, when a pyrophosphate is released after the incorporation of a nucleotide in the DNA strand, is converted to ATP by an ATP sulfurylase. ATP is used by an enzyme called luciferase to oxidase luciferin causing the emission of light. The addition of one nucleotide at time allows a sequential base calling and thus provide information at single CpG site level (Colella et al., 2003).

Instead of sequencing PCR products, they can be digested by specific restriction enzymes allowing to discriminate between methylated and unmethylated sequences, an approach called combined bisulfite restriction analysis (COBRA) (Sadri and Hornsby, 1996). However, only specific sequences are recognized by restriction enzymes and therefore not all CpG sites can be analysed. Moreover, incomplete conversion of unmethylated cytosines may lead to the formation of heteroduplex between

strands containing the restriction sites and site without them compromising the accurate quantification of methylation using this method.

In another method, methylation-sensitive single-nucleotide primer extension (MS-SnuPE), PCR products are isolated from an agarose gel and subsequently hybridized with a primer terminating immediately 5' of the CpG site of interest (Gonzalgo and Jones, 1997). This primer is then extended by DNA polymerase that uses radioactive labelled dCTP or dTTP. The relative of amount of the two nucleotides is quantified with phosphor-imaging analysis. However, this method is quite laborious and requires the use of radioactive. Variants of this method not requiring radioactive labelled nucleotides have been developed including a microarray-based version (Wu et al., 2008).

Different melting properties of methylated and unmethylated sequences can be exploited in a methylation-sensitive melting curve analysis (Worm et al., 2001). By performing MIP in presence of a fluorescent dye that intercalates double strand DNA, it is possible to monitor the emitted fluorescence when temperature is increased. The major transition in fluorescence can be observed as peak in a melting curve. In case of a population of fully methylated and unmethylated molecules, two peaks would be generated. However, some molecules can present an heterogenous pattern of methylation and melting pattern can be difficult to interpret. High-resolution version of this technique (MS-HRM) allows to detect subtle differences within the amplicon with a moderate high-throughput.

MALDI-TOFF mass spectrometry can be also used for DNA methylation analysis after base specific cleavage or primer extension (van den Boom and Ehrich, 2009). Base specific cleavage determines the generation of cleavage products analysed by mass spectrometry. In primer extension, the use of four different types of terminators determines the termination of the primer-extension reaction on different nucleotides with the production of different signals analysed by mass spectrometry.

In another method, HeavyMethyl, oligonucleotide blockers are used to discriminate between methylated and unmethylated sequences. MIP primers are designed to anneal next to a CpG-rich region for which oligonucleotide blockers are designed to hybridize only if the sequence is not methylated. When the sequence is methylated, the oligonucleotide blockers leave free the DNA sequence allowing MIP primers to hybridize. The binding of a fluorescent probe complementary to the CpG-rich region allows the emission of a signal proportional to the amount of the amplicon. False-positive rate of this method is extremely low since oligonucleotide blockers ensures the selective amplification of the methylated sequences at each cycle of the reaction (Cottrell et al., 2004). On the other hand, the use of MSP primers may cause false priming events because of the annealing of primers in unmethylated sequences despite the mismatches. Higher annealing temperature, the use of negative controls and a limited number of amplification cycles may help to avoid false priming. Moreover, in traditional MSP, a second set of primers specific for the corresponding unmethylated

sequence is designed. Gel electrophoresis of the PCR products generated with MSP primers and primers for unmethylated sequence allows an approximative estimate of the amount of the relative methylation levels with a high cost-effective method. (Kristensen et al., 2009).

Several quantitative variants of traditional MSP have been developed.

In MethyLight, a fluorescent probe allows to monitor the amplification in real-time (Eads et al., 2000). Amplification is detected when the probe hybridizes eliminating any signal from unspecific primer hybridization (Figure 3A). Moreover, inclusion of multiple non-CpG sites within the probe sequence prevents false positives due to incomplete bisulfite conversion. Detection of methylation alteration of ctDNA requires the use of ultra-sensitive techniques. In fact, ctDNA concentrations are extremely low and it usually represent less than 0.01% of total cfDNA (Schwarzenbach et al., 2008). For this reason, a compartmentalized version of MethyLight, digital MethyLight (dMethyLight) has been developed. In this method, the sample is diluted and distributed over a large number of wells in order to obtain one or no template molecules in each well and making possible an absolute quantification of the methylated locus of interest. Sensitivity and reproducibility depend on the number of reaction wells (Weisenberger et al., 2008). In Droplet digital MethyLight (ddMethyLight), the sample is distributed into thousands to millions of droplets through a microfluidic system allowing sensitivity even 25-fold higher than conventional MethyLight (Yu et al., 2015). Each single droplet would be positive or negative for methylation of the target.

Another ultra-sensitive method, Beads, Emulsion, Amplification, Magnetics digital PCR (BEAMing) combines emulsion PCR with magnetic beads and flow cytometry with a sensitivity of 0.01% (Li et al., 2009).

In another version of quantitative MSP, the probe is substituted by an intercalating dye such as SYBR Green. Although this method avoids the problem of a complicated probe designing, does not allow to avoid possible false positive events due to primer dimers formation (Chan et al., 2004). Sensitive melting analysis after real-time MSP (smart-MSP) exploits HRM technology for the detection of false positive results (Kristensen et al., 2008) (Figure 3B).

The use of probes and intercalating dyes can be avoided in Methylation-specific fluorescent amplicon generation (MS-FLAG). In this method, MSP primers present a 5' tail with a fluorophore and a quencher separated by a recognition site for an endonuclease. When the polymerase has synthesized the new DNA strand, the endonuclease recognizes the cutting site and the fluorophore is released giving a fluorescent signal (Figure 3C). Since primers can be differentially labeled, multiplexing MS-FLAG is also possible (Bonanno et al., 2007).

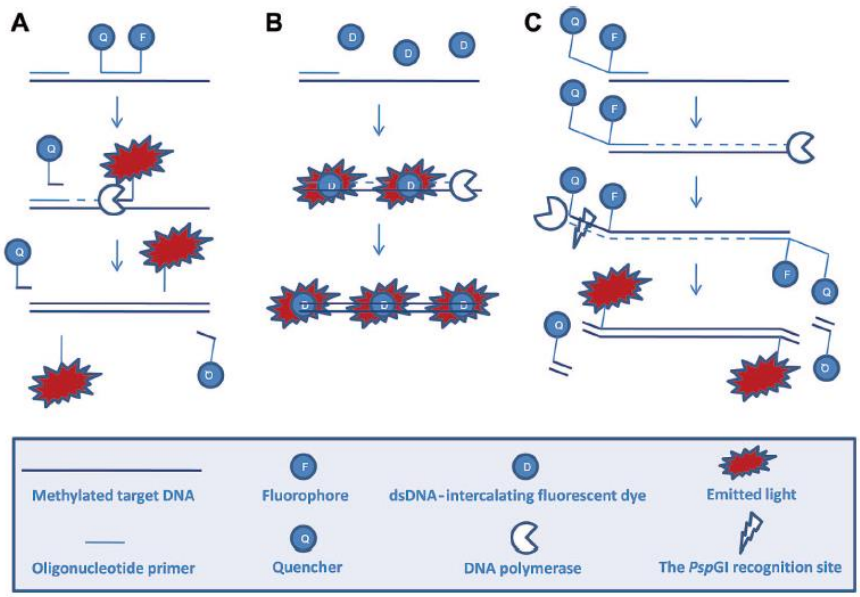


Figure 3. Three different quantitative MSP methods. (A) MethyLight, (B) Smart-MSP, (C) MS-FLAG. Figure from (Kristensen et al., 2009)

1.7.2 Genome-wide methods

The advent of high-throughput technologies has completely revolutionized genomic and epigenomic studies allowing a broad genomic coverage of the methylation landscape. These methods are based on next generation sequencing (NGS) or genome-wide hybridization. NGS technologies allow to perform bisulfite sequencing at genome-wide level (whole-genome bisulfite sequencing or WGBS) with a single-base resolution. In this method, following genomic DNA shearing, DNA fragment ends are repaired by adding an adenine to the 3' end. Sequencing adapters are ligated to the DNA fragments that are then size selected and purified by gel electrophoresis. Library preparation involves bisulfite conversion of the selected DNA fragment and PCR amplification using primers designed on sequencing adapter. At this point, the library is ready for cluster generation and sequencing (Urich et al., 2015). The greatest limitations of this method are the high cost and the significant technical expertise needed to analyze the large amount of data produced. However, since only a fraction of the genome can be differentially methylated between cancer and normal tissues, sequencing of the 5-mC-enriched fraction of the genome allows to reduce costs and increase the sequencing coverage. In this sequencing variant, reduced representation bisulfite sequencing (RRBS), enrichment of CpG-rich region is performed by treating DNA fragments with an enzyme, MspI, that recognizes CCGG sites and cut independently of DNA methylation status. The rest of the protocol is identical to classical WGBS (Meissner et al., 2005). Several commercial kits are available to perform this enrichment step using bait sequences. A costumed version of these kits allows to enrich only regions of interest allowing a targeted bisulfite sequencing (Chatterjee et al., 2012).

Illumina Infinium BeadChips represent a cost-effective, easy-to-use, popular alternative to WBGS.

This technology was first introduced with HumanMethylation27K (HM27) BeadChip in 2008 and led to the realization of the first large epigenome-wide association studies. These arrays interrogated over 25000 CpG predominantly within the proximal promoter regions of about 14000 consensus coding sequence (CCDS) genes and well-described cancer genes (Bibikova et al., 2009).

Subsequently, HumanMethylation450K (HM450) BeadChips, allowing the analysis of more than 450000 CpG sites, replaced HM27K arrays. The new genomic content of this array includes: CpG islands, shores and shelves, the 5′UTR, 3′UTR and bodies of RefSeq genes, FANTOM4 promoters, the MHC region and some enhancer regions (Bibikova et al., 2011). The use of these arrays has allowed the generation of a large amount of global methylation data of several cancers produced by ICGC, the International Human Epigenome Consortium (IHEC) and TCGA. However, WGBS data have highlighted the importance of methylation in regulatory regions such as enhancers that were not covered in 450K arrays. For this reason, Illumina released HumanMethylationEPIC (EPIC) BeadChips arrays containing more than 850000 probes targeting 90% of the sites interrogated by HM450 arrays and 350000 CpGs at regions identified as potential enhancers by FANTOM5 (Lizio et al., 2015) and the ENCODE project (Siggens and Ekwall, 2014).

In Illumina Infinium assays, bisulfite-converted DNA is amplified on whole-genome level and PCR products are fragmented using restriction enzymes. Finally, purified DNA fragments are hybridized onto the chips. In HM450 and EPIC arrays, two types of probes, Infinium I and Infinium II probes, are designed to hybridize a 50-bp sequence downstream the targeted CpG site. DNA methylation measurement with Infinium I probes is carried out by two beads: one, the unmethylated (U) bead measures the unmethylated signal and the other one, methylated (M) bead, measures the methylated signal (Figure 4). The hybridization of unmethylated fragment to the U-bead allows the single base extension and incorporation of a labelled ddNTP matching the nucleotide preceding the cytosine of the target CpG site. In case of hybridization of a methylated fragment to this bead, the mismatch at 3’ inhibits single base extension. Signal detection produced by the hybridization of a methylated fragment on the M bead works on the same way. Both bead types incorporate the same type of labelled ddNTP and therefore will be detected in the same color channel. In Infinium II probes, methylated and unmethylated signals are measured by the same bead (Figure 4). In fact, the probe is designed to bind DNA sequence immediately next to the interrogated CpG site that become the site for single base extension. If this site is methylated, a ddNTP labelled G nucleotide will be incorporated and the signal will be detected in the green channel, while if the site is unmethylated, a ddNTP labelled A will be incorporated and the signal will be detected in the red channel. The other cytosines of the CpG sites within the probe are replaced by R degenerated bases to hybridize both to C and T bases (Pidsley et al., 2016).

After extension, the arrays are fluorescently stained, scanned and the signal intensities are measured to generate beta values. Beta value (β) is a measure of the degree of methylation at one CpG locus and it is calculated using the formula: β =intensity of the methylated signal/ (intensity of the unmethylated signal + intensity of the methylated signal + 100). β-values range from 0, a completely unmethylated CpG site, to 1, representing a fully methylated CpG site (Pidsley et al., 2016).

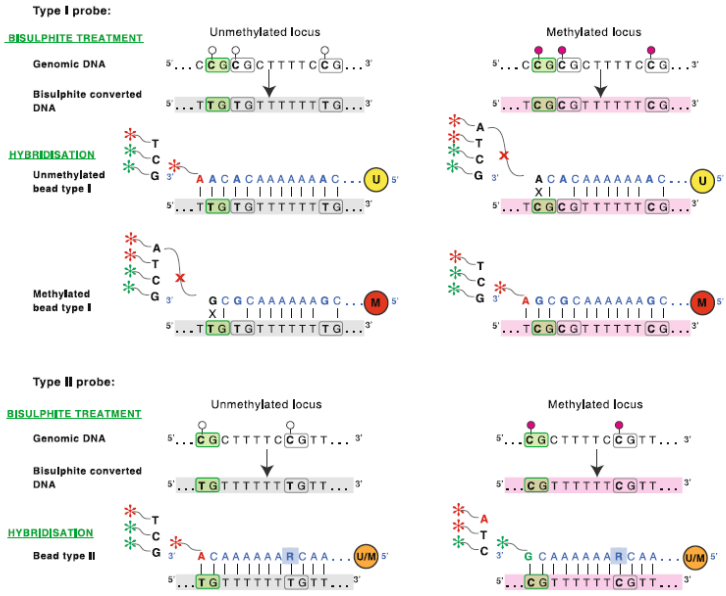


Figure 4. Infinium methylation probe design
Figure adapted from (Pidsley et al., 2016)

1.7.2.1 Analysis methods for methylation Infinium array data

The bioinformatic pipeline for processing and analysis of Infinium array methylation data consists in several steps summarized in Figure 5 (Wilhelm-Benartzi et al., 2013).

Several packages for a comprehensive analysis of methylation data have been developed, including methylumi, minfi, wateRmelon, ChAMP and RnBeads (Morris and Beck, 2015).

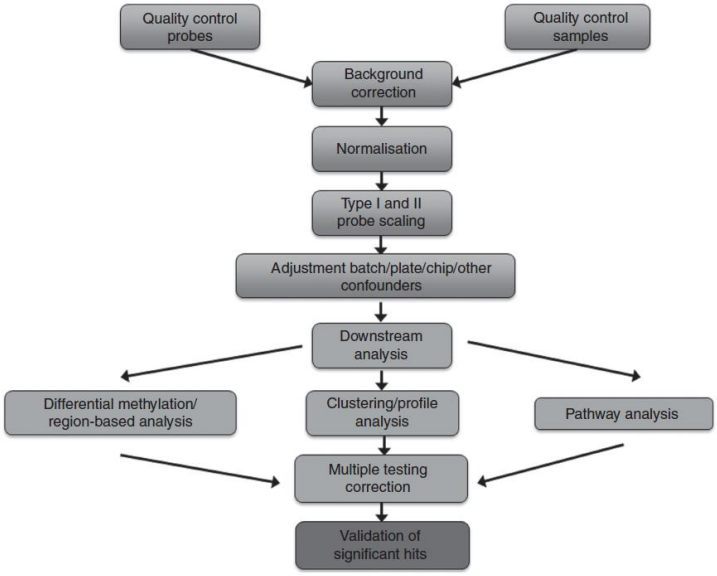


Figure 5. Pipeline for DNA methylation array data analysis
Figure from (Wilhelm-Benartzi et al., 2013)

- Quality control

Infinium methylation arrays contain different control probes to evaluate the quality of the experiment. Sample independent-controls include control probes to monitor efficiency of staining, extension, target removal and overall hybridization performance. Sample-dependent controls include probes to evaluate efficiency of bisulfite conversion and hybridization. Control probe intensity values of poor-performing samples usually deviate from intensity values of the other samples. A comparison of beta value density plots can identify poor-performing samples based on a large deviation from beta values distribution of the rest of the samples (Figure 6). Methylation data of poor-quality samples may be inaccurate, therefore these should be excluded from the analysis (Wright et al., 2016).

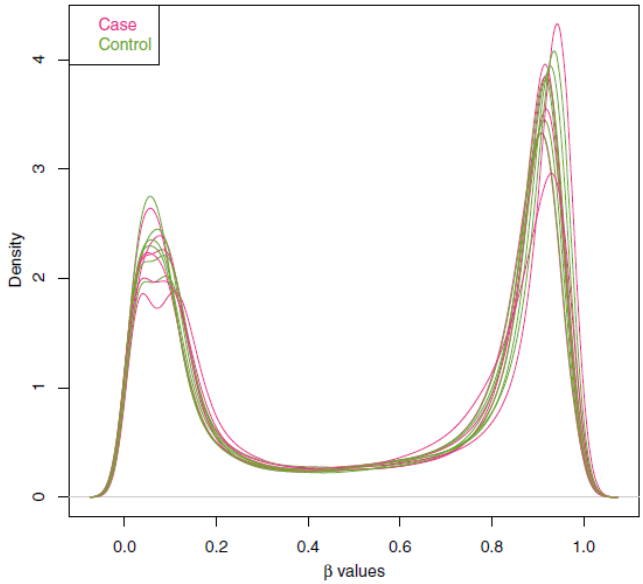


Figure 6. Beta values density plot by experimental groups (cases and controls)
Figure adapted from (Wright et al., 2016)

Similarly, probes that do not meet certain criteria: intensity levels near or at background intensity, probes that fail to measure DNA methylation in a proportion of samples (e.g. 25% of samples), probes with a mean detection p-value higher than a certain threshold (e.g. 0.05) and probes overlapping with single nucleotide polymorphisms (SNPs) should be filter out (Wilhelm-Benartzi et al., 2013; Wright et al., 2016).

- Background correction

Background correction allows to remove nonspecific signal from total signal to obtain true intensity values. Several methods for background corrections, implemented in the different packages, are available (Wilhelm-Benartzi et al., 2013; Wright et al., 2016).

- Normalization

Two types of normalization exist: a within-array normalization, correcting for technical dye-based biases and a between-array normalization to remove technical artifacts between samples in different arrays (Wilhelm-Benartzi et al., 2013).

- Type I and type II probes scaling

The last two versions of Infinium methylation arrays, HM450 and EPIC arrays, used two different types of probes: Infinium type I and Infinium type II probes. Type I probes interrogate more CpG sites mapping on CpG islands than type II probes (Bibikova et al., 2011) leading to a biased detection of differentially methylated regions enriched for type I probes. Moreover, beta values obtained from type II probe has a smaller range and measures of methylation display a larger variance between replicates (Dedeurwaerder et al., 2011) (Figure 7).

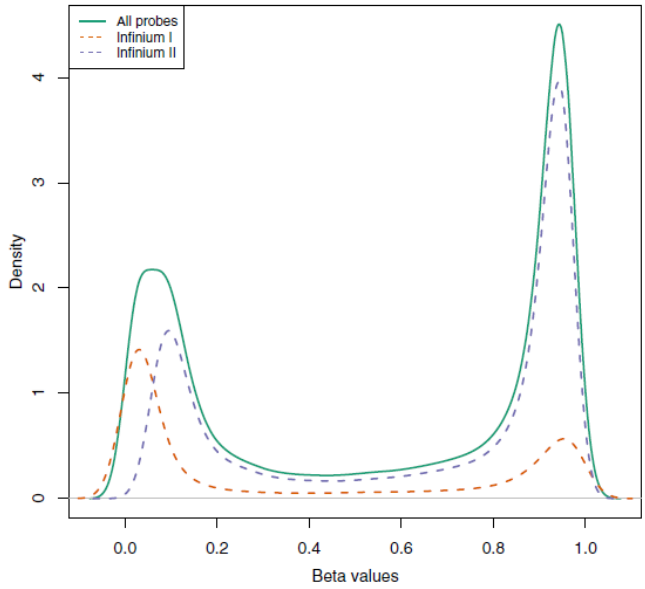


Figure 7. Beta values density plots by probe type
Figure adapted from (Wright et al., 2016)

Different correction methods to rescale the difference between the two types of probes are available. Peak-based correction method implemented in IMA was the first proposed method. In this approach, type II data are rescaled based one type I data assuming a bimodal shape of methylation density profiles (Dedeurwaerder et al., 2011). However, this method requires two distinct peaks for the methylation density.

Subset-quantile within-array normalization (SWAN), implemented in minfi, determines an average quantile distribution using a subset of probes considered as biologically similar based on the number of CpG sites within the sequence of the probe and performed a simultaneous normalization of type I and type II probes (Maksimovic et al., 2012).

In another method, probes are divided into different subgroup on the basis of genomic location of CpG sites. Type I reference quantiles are used for type II signals normalization (Touleimat and Tost, 2012).

In a "data driven" normalization approach, implemented in wateRmelon package, known methylation patterns were used to derive three metrics to test different schemes of correction and normalization (Pidsley et al., 2013).

Another normalization approach implemented in wateRmelon package, Beta MIxture Quantile dilation (BMIQ), adjusts beta values of type II probes into a statistical distribution characteristic of type1 probes using a three-state beta-mixture model to assign probes to methylation states and then transformation probabilities of methylation state membership into quantiles (Teschendorff et al., 2013).

- Adjustment for batch effects

A rational study design is fundamental to reduce batch effects. Surrogate variable analysis (SVA) and independent surrogate variable analysis (ISVA) are two statistical models that can be used when source of batch effects is unknown (Leek and Storey, 2007; Teschendorff et al., 2011). SVA estimates the sources of batch effects from the array data that are used as covariate in the statistical model (Leek and Storey, 2007). ISVA identifies features correlating with the phenotype of interest in the presence of potential confounders, which are modelled as statistically independent surrogate variables (Teschendorff et al., 2011).

- Downstream analyses

Two different measures of methylation are commonly used: beta values and M-values. M-value is calculated as $M= \log2 \frac{\text{Max (M,0)}}{\text{Max (U,0)}}$ (Du et al., 2010). M-values can be also transformed in beta values and *vice versa* using this formula $M= \log2 \frac{\beta}{1-\beta}$ (Du et al., 2010).

One common methylation analysis consists in the calculation of a differential methylation, termed as delta beta (or $\Delta\beta$), between cases and controls. A genomic region with a delta beta between cases and controls higher than 0.2 is usually considered a differentially methylated region (DMR). High correlation between neighbor CpG sites exists and decreases as pairwise distance increases. The identification of a differentially methylated regions is considered as a robust finding when this regional difference is detected across several probes (Wright et al., 2016).

Differential methylation patterns between cases and controls can simply reflect cell heterogeneity differences between samples. Peripheral blood samples, which are often used as samples for differential methylation analysis, contain different cell types that have characteristic methylation patterns. Therefore, if a sample has an abnormal cell-type proportion, the identified DNA methylation

change can simply reflect a different cell-type composition. Statistical correction methods can be used to estimate cell-type proportion based on methylation patterns identified in purified blood cell lines and used as covariates in a differential methylation analysis (Houseman et al., 2012). A new reference-free approach for cell mixture adjustment has been developed (Houseman et al., 2014)

Functional enrichment analyses are usually used to interpret methylation data from a biological and clinical point of view. However, the association of a differentially methylated region to a particular gene is a simplistic and potentially wrong view. In fact, it is difficult to identify which genes are affected by a specific DNA methylation change. A new approach consists in a regulatory enrichment analysis that evaluate enrichment of DMR in functional regulatory regions (Wright et al., 2016).

Multiple testing multiple testing correction is necessary to reduce the likelihood of false positives by adjusting statistical confidence measures by the number of tests performed. Several correction methods are available such as Bonferroni correction and adjustment of the false discovery rate (FDR) (Wilhelm-Benartzi et al., 2013).

- Clustering analysis

Clustering analysis is usually performed using DNA methylation data to identify different subgroups of samples. Samples belonging to a specific subgroup may be enriched for a specific clinical or molecular characteristic. Clustering include non-hierarchical and hierarchical methods. Non-hierarchical methods, such as K-means, require the specification of the number of classes. On the other hand, hierarchical methods build a binary tree by including similar samples or probes in one tree based on a measure of similarity. Recursive partitioning mixture model (RPMM) uses a beta-mixture model to divide samples into different subgroups and provides an estimate for the number of clusters (Houseman et al., 2008).

2. AIMS

Cancer represents the second cause of death worldwide (World Health Organization, 2018). Effective cancer prevention strategies, global screening programs and methods for early cancer diagnosis are still lacking. A great effort for the identification of new biomarkers for early detection of cancer, prediction of disease outcome and therapy response has been done in the last decades. In fact, all these factors have the power to reduce cancer mortality.

Global epigenetic changes, including changes in DNA methylation pattern, represent one of the cancer hallmarks. DNA methylation alterations occur early during tumour formation and can be detected in cell free circulating DNA from different biological matrices.

The first aim of this thesis was to identify methylation alterations that may be potential biomarkers for prediction of cancer occurrence, early diagnosis, prognosis and monitoring of disease course.

In particular, we evaluated DNA methylation status of two selected methylation alterations, previously identified in colorectal cancer by our research group (Fadda et al., 2018), in tumour and matched-normal tissues of CRC patients and we investigated whether these methylation alterations can be detected in stool samples of CRC patients. Moreover, we performed a comprehensive methylation analysis of one blood cancer, chronic lymphocytic leukemia, and one solid tumour, biliary tract cancer.

DNA methylation changes are closely associated to gene expression changes. Although promoter DNA hypermethylation is usually associated to gene downregulation, gene-body hypermethylation can be either positively and negatively associated to gene expression. Moreover, the temporal succession of events leading to gene expression and methylation aberrations is not clear and many evidences suggest that DNA methylation targets genes that are already repressed or lowly expressed in normal tissues where the tumour arises.

The second aim of this thesis was to investigate the relationship between methylation and gene expression. We performed a differential gene expression analysis between tumour and normal samples of selected genes whose associated-CGIs were aberrantly methylated in tumours.

Many cancer studies have found that methylation alterations affect common pathways and gene families across different types of cancer. In this thesis, we investigated the methylation status of CGIs associated to clustered *PCDH* genes in both solid cancers (pilocytic astrocytoma, colorectal cancer, gastric cancer and biliary tract cancer) and blood cancer (chronic lymphocytic leukemia).

Finally, since DNA methylation is a reversible epigenetic modification, it represents a possible therapeutic target. A modification of CRISPR-Cas9 system represents a cutting-edge method to edit DNA methylation and restore normal methylation pattern in cancer cells. The final aim of this thesis

was to employ this sophisticated system to selectively de-methylated a target of interest in a colorectal cancer cell line.

3. MATERIALS AND METHODS

3.1 Exploratory datasets

Different types of cancers have been analysed in this thesis (Figure 8)

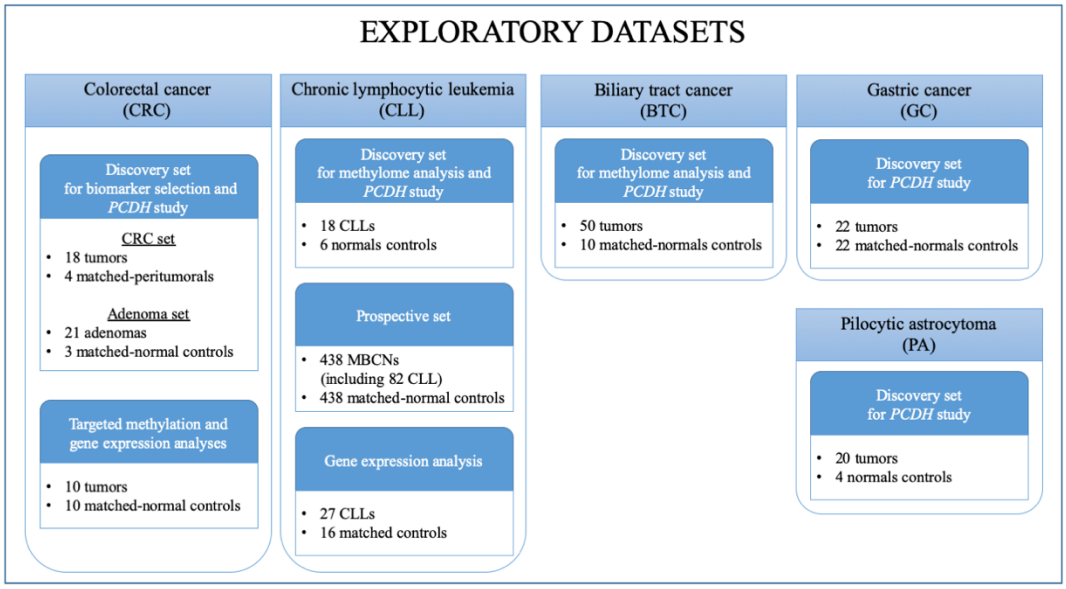


Figure 8. Exploratory datasets

3.1.1 Samples and data collection

3.1.1.1 Colorectal cancer (CRC)

Discovery set for biomarker selection and PCDH study

Methylation data were obtained from a previous genome-wide methylation study of 18 primary CRCs and four matched peritumoural samples, 21 colorectal adenomas and three matched-normal intestinal mucosa samples, performed by our research group (Fadda et al., 2018). Patients' clinical information including tumour location, CIMP status, microsatellite instability (MSI) status, *KRAS* mutational status and Dukes staging classification were previously collected (Fadda et al., 2018).

Discovery set for targeted methylation and gene and protein expression analyses

Tumour and matched-normal fresh-frozen (FF) tissue samples of 10 CRC patients were collected from the Department of General Surgery of the University of Cagliari (Italy). Normal samples were taken at a distance >10 cm from the neoplastic tissue. After molecular analyses, tissue samples have been analysed by a histopathologist. Frozen section slides underwent standard hematoxylin and eosin (H&E) staining. Microscope images were acquired of each individual slide.

Stool samples from the same patients were collected intraoperatively during tumour resection and immediately frozen at -80°C until being processed.

Patients clinical data are reported in Table 1.

Sample ID	Tumour location	Stage at diagnosis	Mucinous histology	Lymphovascular invasion	Grade	Ulcerative neoplasia
CRC_2	Left colon	I	NO	NO	G2	NO
CRC_3	Right colon	III	NO	YES	G2	YES
CRC_8	Rectum	IV	YES	YES	G2	NO
CRC_12	Right colon	0	NO	YES	G1	NO
CRC_14	Rectum	III	YES	YES	G3	NO
CRC_19	Right colon	II	YES	YES	G2	YES
CRC_21	Transversal colon	II	NO	YES	G2	NO
CRC_29	Right colon	II	NO	YES	G2	NA
CRC_33	Right colon	IV	NO	YES	G2	NA
CRC_34	Rectum	III	NO	YES	G2	YES

Table 1. Clinical characteristics of CRC patients

3.1.1.2 Chronic lymphocytic leukemia (CLL)

Discovery set for genome-wide methylation analysis and PCDH study

Blood samples of 18 CLL patients (10 men and eight women, mean age at diagnosis: 65.3±12.3) and six normal controls (mean age: 51.3±20.2) were collected from the Hematology Department of the Businco Oncology Hospital in Cagliari (Italy).

Clinical and immunophenotypic characteristics of CLL samples are reported in Table 2.

Sample ID	CD5+ (%)	CD5+/ CD19+ (%)	CD23+ (%)	CD38+ (%)	*IGHV* mutational Status	Lymphocyte count/mm^3
304012 007	93.30	71.50	72.10	42	Positive	5050
304012 030	98.60	92	92	2.6	Negative	32380
304012 048	94.1	78.70	75.40	15	Negative	9580
304012 088	98.2	91	91.4	4.7	Negative	45330
304012 092	88.7	64.8	63.7	10	Positive	21830
304012 112	34	6	70.3	14	Negative	5270
304012 114	Negative/ weak	NA	65	16	Negative	24360

304012 188	97	87	88	9	Positive	5410
304012 193	56	NA	NA	NA	Negative	58880
304012 196	68	54	76	NA	Negative	9060
304012 475*	97.9	94.60	Partially expressed	Weak	Positive	94100

Table 2. Clinical and immunophenotypic characteristics of CLL patients
Notes: For 7/18 patients clinical and immunophenotypic data were not available.
* this patient was firstly diagnosed as Follicular lymphoma

Discovery set for gene expression analysis

Blood samples of 27 CLL cases and 16 normal controls were collected from Italian hospitals in Novara, Florence, Perugia and Cagliari.

Prospective set

The prospective set included 438 incident mature B-cell neoplasms (MBCN), among which 82 were CLL and small lymphocytic lymphoma (SLL) cases, and 438 normal controls (individually matched to cases at 1:1 ratio based on age at enrollment, gender, ethnicity and DNA source) from the Melbourne Collaborative Cohort Study (Wong Doo et al., 2016). Peripheral blood samples were collected prior to any cancer diagnosis with a mean time between collection and diagnosis of 10.6 years (range 0.2-20 years) for the entire MBCN cohort and 9.5 years (range 0.6-17.8 years) for the CLL subset cohort (Wong Doo et al., 2016). Methylation data of this cohort constitutes our prospective set.

3.1.1.3 Biliary tract cancer (BTC)

Discovery set for genome-wide methylation analysis and PCDH study

Formalin-fixed paraffin-embedded (FFPE) tissue samples from 50 BTCs (25 men and 25 women, mean age at diagnosis: 70.4±10.9) and 10 matched-normal controls were obtained from the Department of Oncology, University of Cagliari (Italy) and the Scientific Institute Romagnolo for the Study and Treatment of Tumours (IRST) Srl – IRCCS, Meldola, FC (Italy). Clinical characteristics of these patients are reported in Table 3.

Sample ID	Tumour location	Stage at diagnosis	Grade
06 26142-1	Intrahepatic	II	G2
06-31850-1	Intrahepatic	III	G2
06-7929-1	Intrahepatic	III	G3
06B1315-2	Intrahepatic	IV	G2
07-21-665-6	Intrahepatic	III	G3
07B5152	Intrahepatic	IV	G2
08-4421-1	Intrahepatic	III	G3

08-5028-2	Intrahepatic	III	G3
09_26274_1	Intrahepatic	II	G1
09-7104-1	Intrahepatic	IV	G3
10_20382_8	Intrahepatic	II	G3
10_4472_2	Extrahepatic	III	G2
10-14035-3	Extrahepatic	III	G2
10-26765-1	Gallbladder	IV	G2
10-5669-3	Intrahepatic	IV	G3
1014532-14	Intrahepatic	III	G2
11-11445-5	Extrahepatic	IV	G2
11-12043-1	Intrahepatic	IV	G3
11-1668-1	Extrahepatic	IV	G3
11-24627-1	Gallbladder	IV	G3
11-28627-1	Extrahepatic	IV	G3
11B3614	Intrahepatic	IV	G2
12_16875_6	Intrahepatic	III	G2
12_20714_1	Intrahepatic	IV	G3
12B1108-B	Intrahepatic	II	G2
12B5263	Intrahepatic	IV	G3
12B9086	Intrahepatic	IV	G2
13_161_1	Intrahepatic	IV	G3
13_24812_1	Intrahepatic	II	G2
13_3421_1	Extrahepatic	I	G3
13-2544-5	Extrahepatic	III	G2
BTC_011GP_T	Gallbladder	IV	G2
BTC_022GP_T	Extrahepatic	II	G2
BTC_023GP_T	Gallbladder	III	G2
BTC_027GP_T	Gallbladder	IV	G2
BTC-006-BM_T	Gallbladder	III	G3
BTC-009-FAM_T	Gallbladder	III	G2
BTC-017-LL_T	Gallbladder	III	G2
BTC-019-MM_T	Gallbladder	III	G3
BTC-028-DMGC_T	Gallbladder	III	G1
BTC-030-GV_T	Gallbladder	IV	G3
BTC-032-MA_T	Gallbladder	II	G3
BTC-036-PM_T	Gallbladder	III	G2
BTC-038-PL_T	Gallbladder	IV	G2
BTC-039-RS_T	Gallbladder	III	G1
BTC-041-PG_T	Gallbladder	I	G1
BTC-042-RR_T	Gallbladder	III	G3
BTC-043-SA_T	Gallbladder	II	G3
BTC-047-ZL_T	Gallbladder	IV	G3
BTC-049-ZA_T	Gallbladder	I	G2

Table 3. Clinical characteristics of BTC patients

3.1.1.4 Gastric cancer (GC)

Discovery set for PCDH study

Paired tumour and normal FF tissue samples of 22 gastric cancer patients were collected from the Candiolo Cancer Institute-FPO, IRCCS, University of Turin (Italy). Tumour localization included: gastroesophageal junction, antrum/pylorus, fundus, body, body/fundus, antrum. GC samples were classified into three molecular subtypes: microsatellite instable (MSI), chromosomal instable (CIN) or genomic stable (GS). A further classification included Epstein–Barr virus (EBV) positivity.

3.1.1.5 Pilocytic astrocytoma (PA)

Discovery set for PCDH study

Methylation data were obtained from a genome-wide methylation analysis of 20 pediatric PA samples and four normal brain controls previously performed by using IlluminaMethylation27 BeadChips (27K) in our laboratory (Antonelli et al., 2018).

3.1.2 Experimental assays

3.1.2.1 Nucleic acids extraction and quantification

DNA was extracted from tissues using DNeasy Blood & Tissue Kit (Qiagen) for FF samples and by QIAamp DNA FFPE Tissue kit (Qiagen) for FFPE samples.

DNA from stool samples were extracted using QIAamp Fast DNA Stool Mini Kit (Qiagen).

DNA was isolated from peripheral blood lymphocytes using the DNA extraction 500 arrow® Kit (DiaSorin Ireland Ltd).

DNA samples were quantified by spectrophotometric reading (NanoDrop Products, Thermo Scientific) and by fluorometric reading (Quant-iT™ PicoGreen® dsDNA Assay Kit) and their quality was evaluated by electrophoresis in a 0.8% agarose gel.

RNA extraction from CRC and matched-normal tissue samples, peripheral blood mononuclear cells (PBMCs) of CLL and normal controls, was performed by RNeasy Mini Kit (Qiagen).

RNA samples were quantified using NanoPhotometer (NanoPhotometer™Pearl).

3.1.2.2 DNA bisulfite conversion

DNA samples were bisulfite converted using EZ DNA Methylation Gold Kit (Zymo Research).

3.1.2.3 Targeted methylation assay

3.1.2.3.1 MethyLight

Methylation of two selected biomarkers, *GRIA4* and *VIPR2*, was assessed by MethyLight (Eads et al., 2000) in 10 CRC and matched-normal tissue samples and in 10 stool samples from the same patients. The reference repetitive element Alu was used in a methylation-independent control reaction to normalize the amount of DNA input. Primers and probes were designed using Beacon Designer™ (Premier Biosoft) and their sequences are reported in Table 4. Each probe was labelled with the 6-

Carboxyfluorescein (6-FAM) fluorophore at the 5′ end. A primer-probe mix containing 300 nM of each primer and 100 nM of the probe was prepared. Each assay was performed in triplicate using: 15 µl of TaqMan Genotyping Master mix (Applied Biosystems), 4.5 µl of primer-probe mix, 5 µl of bisulfite-converted DNA (10 ng/µl) and 5.5 µl of RNase-free water. A fully-methylated DNA (Human Methylated & Non-Methylated (WGA) DNA Set; Zymo Reasearch) was used as a positive control for the reaction. The experiment was conducted on a DNA Engine Opticon 2 Real-Time Cycler (Bio-Rad) using the following thermal conditions: initial PCR activation step at 95 °C for 10 minutes (min), followed by 50 cycles of denaturation step at 95 °C for 15 seconds (sec) and annealing/extension step at 60 °C for 60 sec.

Biomarker	Forward primer (5'-3')	Reverse primer (5-3')	Probe (5'-3')
GRIA4	GGGTTGGTGTAGGTTTGTT	CTCCCCCCTTACTTTCTCAC ATACACACAA	AACGCCGCGACCGCCACAC
VIPR2	TCGGTTTCGAGTAGAGAGA ATTGG	AAACAAATACAAACGACC GCAAAA	CCCTTCCGAACGCACACCT AACCC
Alu	GGTTAGGTATAGTGGTTTA TATTTGTAATTTTAGAT	ATTAACTAAACTAATCTTA AACTCCTAACCTCA	CCTACCTTAACCTCCC

Table 4. Primers and probe sequences for methyLight assay

3.1.2.3.2 Droplet digital PCR

GRIA4 and *VIPR2* methylation was also evaluated by droplet digital PCR (ddPCR) in the 10 stool samples from CRC patients.

ddPCR reactions containing 2 × ddPCR Supermix for probes (Bio-Rad), forward and reverse primer (900 nM), probe (250 nM) and 2 to 5 µl bisulfite-converted DNA in a final volume of 20 µl were partitioned into ~20,000 oil-emulsified droplets per well and replicated in three wells using a Bio-Rad QX200 droplet generator. The droplets were transferred into 96-well plates and PCR was performed using the following conditions: 10 minutes at 95 °C, 40 cycles of 30 seconds at 95 °C followed by 60 seconds at 60 °C, then 5 minutes at 4 °C, 5 minutes at 95 °C. Plates were subsequently read on a Bio-Rad QX200 droplet reader.

3.1.2.4 Gene expression assay (qRT-PCR)

RNA samples were retro-transcribed using the High Capacity Kit (Applied Biosystems). The obtained cDNA samples were used for gene expression analyses of selected genes by qRT-PCR conducted on a DNA Engine Opticon 2 Real-Time Cycler (Bio-Rad), using iQ™ SYBR® Green Supermix (Bio-Rad). qRT-PCR conditions were: primary denaturation at 95°C for 2 min followed by 50 cycles of denaturation at 95°C for 15 sec and annealing/extension at 60°C for 1 min. After the amplification cycles, melting curves are produced by increasing the temperature from 65°C to 95°C holding each temperature for 5 sec and reading fluorescence every 0.5°C.

CRC gene expression assay

Gene expression levels of four selected genes, *GRIA4*, *VIPR2*, *SLC6A3* and *SPOCK1*, were evaluated in CRC and matched-normal samples. *TFRC* was used as reference gene.

CLL gene expression assay

SHANK1 gene expression analysis was performed in CLL and normal control samples. *ACTB* was used as reference gene. A subset of these samples was also re-analysed using *GUSB* as reference gene.

Primer sequences used for the two gene expression studies can be found in Table 5.

	Gene	Forward primer (5'-3')	Reverse primer (5-3')
CRC gene expression assay	*GRIA4*	TCATGTGGACAACATTGAGACA	ATCATAGAGTCCAAAAATGGCAAA
	VIPR2	GTCTCTTGCAACAGGAAGCA	TCTCAGGATGAAGGACAGGAA
	SLC6A3	CCATACTGAAAGGTGTGGGCT	AGAAGAGATAGTGCAGCGCC
	SPOCK1	AGGTAAAATGCAGCCCTCACA	TTCCCCTTCTTTTGCCTGGG
	TFRC	GGCACAGCTCTCCTATTGAAAC	CAAAGTCTCCAGCACTCCAACT
CLL gene expression assay	*SHANK1*	AGACCATCAGTGCAAGCGAA	GGGATCGAAGCTCGACTCAG
	ACTB	AAATCTGGCACCACACCTTC	AGCACAGCCTGGATAGCAAC
	GUSB	CACCTAGAATCTGCTGGCTACT	AGAGTTGCTCACAAAGGTCACA

Table 5. Primers sequences for qRT-PCR assay

3.1.2.5 Protein expression assay (western blot)

The protein expression levels of long canonical and short GluR4 isoforms, VIPR2, SPOCK1, and SLC6A3 were evaluated by western blot in CRC and matched-normal samples. Proteins extraction from tissue was performed Membrane Protein Extraction Kit (Mem-PERTM Plus, ThermoFisher Scientific). Protein extracts were electrophoretically separated on a 10% SDS-polyacrylamide gel at 100V for 90 min and transferred onto 0.45µm nitrocellulose membrane. The membrane was blocked in 5% milk overnight at 4ºC and then incubated with primary antibody GluR4 (0.5µg/ml, PA5-18931, ThermoFisher Scientific), SPOCK1 (1µg/ml, MA5-24039, ThermoFisher Scientific), VIPR2 (0.5µg/ml, AB2266, Millipore) and SLC6A3 (1µg/ml, SAB2502027, Sigma-Aldrich) for 2h. After three washes with Tris-buffered saline containing Tween-20, the membrane was incubated with horseradish peroxidase-labeled secondary antibody (Jackson ImmunoResearch) for 1 h at room temperature and washed again with Tris-buffered saline containing Tween-20 for three times. Each assay was performed in duplicate. Final detection was performed with ECL Chemiluminescent Western blotting reagents (Bio-Rad). An antibody against NaK ATPase (1:40000, Ab76020, Abcam) was used for gel-loading control.

3.1.2.6 Genome-wide methylation assay (Infinium assay)

The quality of DNA extracted from FFPE samples was evaluated prior to bisulfite conversion using Infinium HD FFPE QC Assay (Illumina). DNA samples that passed this quality control step were bisulfite converted and subjected to a DNA restoration process using Infinium FFPE DNA Restore Kit (Illumina). This kit repairs degraded FFPE DNA in preparation for the whole-genome amplification step in the Infinium assay.

Bisulfite converted DNA samples were processed according to the Illumina Infinium® HD Methylation protocol (Illumina) and the Illumina HiScan was used to scan and record high-resolution images of the emitted fluoresce.

CLL and normal control samples were analysed by Illumina Infinium HumanMethylation450 BeadChips (450K), which allows the interrogation of more than 450 000 CpG sites.

BTC and GC samples along with their relative normal samples were analysed by Illumina Infinium HumanMethylationEPIC BeadChips (EPIC) interrogating over 850 000 CpG sites.

3.1.2.7 sgRNA design and cloning

Three sgRNAs for *GRIA4* target region including eight CpG loci have been designed using an online tool (http://crispr.mit.edu). sgRNAs consist in a 20-bp sequence that targets the region of interest and a 3-bp sequence which is the PAM sequence to recruit Cas9 (Table 6).

***GRIA4* guide RNAs**	**Sequence (5'-3')**
sgRNA3	AGCGTCTAGTGGCTGCTCGC**AGG**
sgRNA 5	AGGGAGTGCGCGCTCGAGGA**GGG**
sgRNA 19	CCGGGCTGGTGCAGGCTTGC**TGG**

Table 6. gRNAs sequences
Notes: in bold PAM sequences

Two DNA fragments (one containing the first gRNA sequence, gRNA scaffold, U6 promoter and the second gRNA sequence and the other one containing gRNA scaffold, U6 promoter and the third gRNA sequence) were designed, synthetized and cloned separately into two pU57 vectors by Gene Universal (Gene Universal Inc.). The two DNA fragments were then subcloned into a lentiGuide-Puro plasmid (Addgene plasmid #52963, a gift from Feng Zhang) (Sanjana et al., 2014). In particular, plasmids obtained from Gene Universal were digested with BsmBI and XbaI enzymes (New England Biolabs) and lentiGuide-Puro plasmid was digested with BsmBI enzyme (New England Biolabs). The digestion fragments were separated by gel electrophoresis and DNA fragments of interest were extracted from the agarose gel using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel). Ligation reaction was performed using T4 DNA ligase (Takara Biotechnology). A lentiGuide-Puro plasmid with sgRNAs sequences cloned between U6 promoter and gRNA scaffold sequences in the original plasmid were obtained (Figure 9).

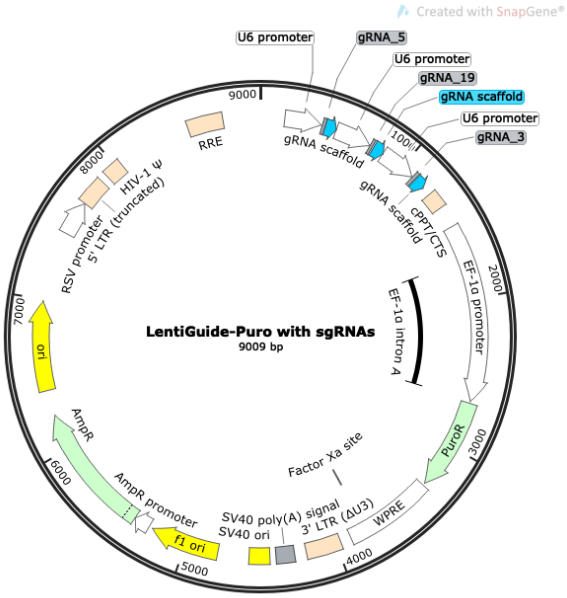


Figure 9. LentiGuide-Puro plasmid with sgRNAs

3.1.2.8 Cell culture

HCT116 colorectal cancer cell lines and HEK293T cell line were originally obtained from the American Type Culture Collection (ATCC). Both cell lines were cultured in Dulbecco's modified eagle medium (DMEM, ThermoFisher Scientific) supplemented with 10% fetal bovine serum (FBS, ThermoFisher Scientific) and 5 ml L-Glutamine (100X stock). Cells were incubated at 37 °C in a humidified 5% CO_2 environment.

3.1.2.9 Plasmid for editing DNA methylation

A previously generated Lenti-dCas9-Tet1 plasmid was available for targeted de-methylation. This plasmid contains dCas9 domain, HumanTet1 catalytic domain (CD) and enhanced green fluorescent protein *(eGFP*)-encoding gene under control of EF1α promoter (Figure 10).

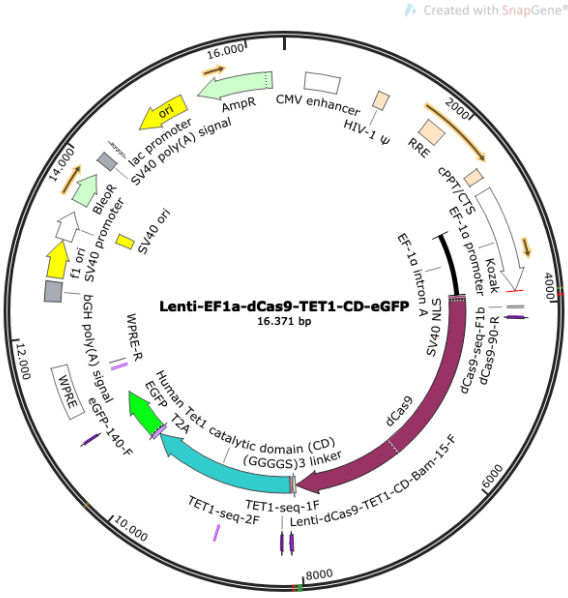


Figure 10. Lenti-dCas9-Tet1 plasmid

3.1.2.10 Lentivirus production and purification

Lentiviruses were produced by transfecting 80% confluent HEK293T cells, plated in a 10-cm dish, with a three-plasmids system. A lentiviral plasmid mix containing the lentiviral vector plasmid (Lenti-dCas9-Tet1 or lentiGuide-Puro with sgRNAs plasmid) and the lentiviral packaging vectors (pVSVg and psPAX2) in a 1:1:0.5 ratio was prepared in 2 ml Opti-MEM (Invitrogen Life Technologies). Lentiviral plasmid mix was combined to a reagent mix containing 2 ml of Opti-MEM and 100 µl of Lipofectamine 2000 (ThermoFisher Scientific) and incubated at room temperature for 5 min. After aspiration of old media, 6 ml of fresh media and 4 ml transfection mix was added to HEK293T cells. Cells were incubated at 37°C in a 5% CO_2 humidified atmosphere. Transfection media was replaced with fresh DMEM media after 24h. Viral supernatants were collected 48h and 72h after transfection. Lentiviral particles were concentrated using 100KDa Amicon Ultra-15 centrifugal filter devices (Millipore Corporation).

3.1.2.11 Viral infection of cells

The experimental plan is summarized in Figure 11.

HCT116 cells were infected with lentiviral particles expressing Lenti-dCas9-Tet1. pUltra-Chili-Luc was used as positive control. 1 ml transduction medium (120 µl virus + 10 µl Polybrene (stock 400 µg/ml) + 870 µl medium) was added to each well of cells seeded in a 6-well plate. The plate was centrifuged 600 xg for 1h at room temperature and then incubated at 37°C in a 5% CO_2 humidified atmosphere. After 24h, transduction medium was replaced with fresh medium. Cells were transferred to two T75 flasks two days after transfection. GFP-positive cells (cells efficiently transduced and consequently expressing GFP) were visible under the microscope 48h after transduction. After additional 96h, 80% confluent cells were harvested for FACS sorting of GFP-positive cells. FACS sorting was performed at ACRF Flow and Laser Scanning Cytometry Facility (SAMHRI, Adelaide). GFP-positive sorted cells were plated on T75 flask and incubated at 37°C in a 5% CO_2 humidified atmosphere. After few days, GFP-positive sorted cells were plated on a 6-well plate and transduced with lentiviral particles expressing lentiGuide-Puro with sgRNAs or lentiGuide-Puro alone. Transduced cells were harvested for DNA extraction after 24h and 48h post transduction. Puromycin selection was performed after 48h by adding 1 µg/ml Puromycin (1 mg/ml stock) to the cells. The following day, cells were harvested for DNA extraction.

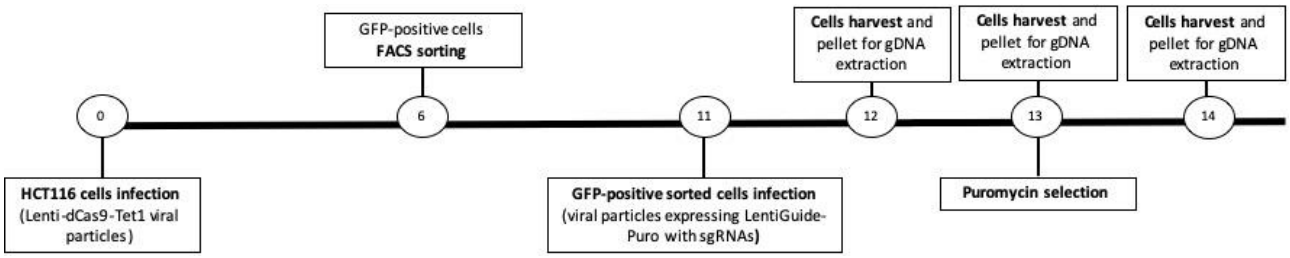


Figure 11. Schematic representation of the experimental plan. Number inside the circles indicate days.

3.1.2.12 5-AZA treatment

HCT116 cells were treated with 1 µM 5-aza-cytidine (AZA) (Sigma-Aldrich) and were harvested for DNA extraction after 72h.

3.1.2.13 DNA samples preparation and pyrosequencing

DNA was extracted from cell pellets using DNeasy Blood & Tissue Kit (Qiagen) and quantified by spectrophotometric reading (NanoDrop Products Thermo Scientific). DNA samples were bisulfite converted using EZ DNA Methylation Gold Kit (Zymo Research). *GRIA4* methylation levels of untreated and treated HCT116 cells were assessed by pyrosequencing. Primers are reported in Table 7. Bisulfite PCR was performed in a 25 µl reaction volume containing: 12.5 µl 2X GoTaq Master Mix (Promega), 1 µl forward primer (5 µM) and 1 µl reverse primer (5 µM), 2 µl bisulfite converted DNA (40 ng/µl) and 8.5 µl RNase-free water**.** Each PCR reaction included a fully methylated bisulfite-converted DNA (Human Methylated & Non-Methylated (WGA) DNA Set; Zymo Reasearch) as positive control and unmethylated bisulfite-converted DNA (Human Methylated & Non-Methylated (WGA) DNA Set; Zymo Reasearch), unmethylated unconverted DNA (Human Methylated & Non-Methylated (WGA) DNA Set; Zymo Reasearch) as negative controls and a no template control (NTC). The thermocycling program was set as follow: 2 min denaturation at 95°C followed by 45 cycles of 95°C for 45 sec, 62°C for 45 sec, 72°C for 45 sec and a final extension of 5 min at 72°C. PCR products were checked on a 2% gel before sending to AGFR Facility (Australia) for pyrosequencing.

Gene	Primer/probe	Sequence (5'-3')
GRIA4	Forward	GGGTTGGTGTAGGTTTGTT
	Reverse	[btn]CTCCCCCCTTACTTTCTCACATACACACAA
	Sequence	GTGTAGGTTTGTTGGG

Table 7. Primers and probes for pyrosequencing

3.1.3 Data analyses

3.1.3.1 Biomarkers selection

A restricted number of biomarkers was selected for targeted methylation and gene expression analyses from the panel of 74 CGIs altered in CRC and adenomas by following the steps summarized in Figure 12. Firstly, CGIs for which mean beta value was higher than 0.25 in peritumoural and normal samples were filtered out. Secondly, only CGIs with a beta value higher than 0.45 in at least

75% of tumours and a beta value higher than 0.25 in not more than 25% of peritumoural and normal samples were selected. ROC curves using methylation data of the selected biomarkers were performed by R "ROCR" package using methylation data from our discovery dataset and methylation data from TCGA-COAD validation dataset. CGIs with an AUC higher than 0.95 in both datasets were selected obtaining 24 CGIs. Finally, additional filtering criteria including feasibility of the assay, functional evidences and literature information about these biomarkers led to the selection of four biomarkers for targeted methylation and gene expression analyses.

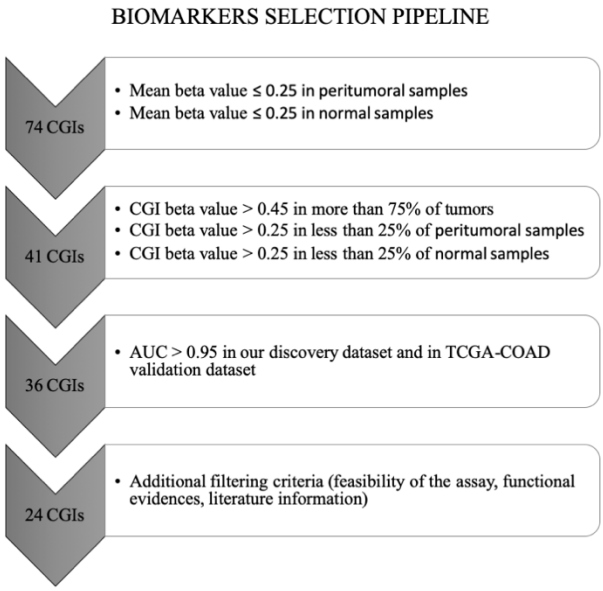


Figure 12. Biomarkers selection pipeline

3.1.3.2 Genome-wide methylation data analysis

Raw DNA methylation data (.idat files) were analysed using a Bioconductor software package, RnBeads (Assenov et al., 2014), installed in R environment. RnBeads allows to analyze both methylation array data and sequencing data. The analysis is divided into several modules (Figure 13): quality control, preprocessing, tracks and tables, exploratory analysis and differential methylation.

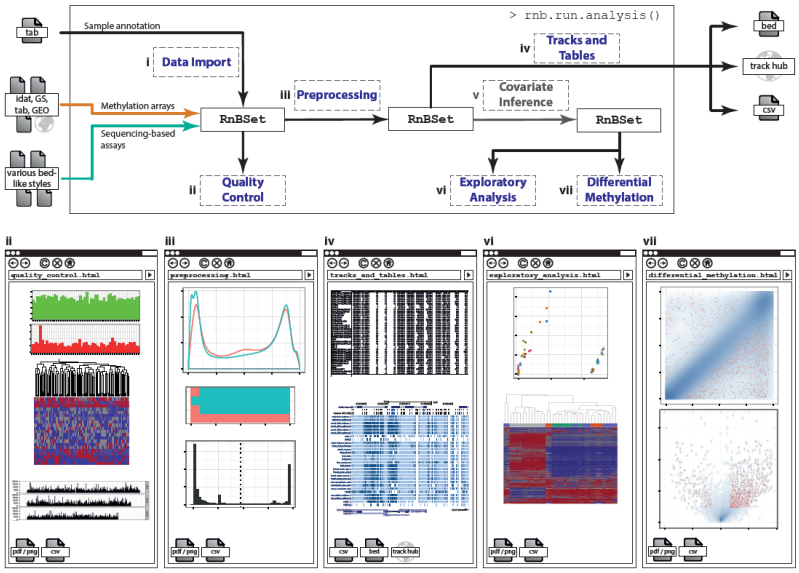


Figure 13. RnBeads workflow and modules
Figure from (Müller et al., 2016)

Two types of analysis can be performed: a "Vanilla" analysis, consisting in the specification of some parameters and execution of all analysis modules and a "Tailored" analysis to perform each module individually. A "Vanilla" analysis was performed with the specification of the type of data to analyze and the location of methylation data, sample annotation file and final output directory. Input files are IDAT intensity files generated by IlluminaHiScan and a sample annotation file including sample identifiers, sample position in the arrays and phenotypic information. Set options included: removal of probes on sex chromosomes, background subtraction and normalization algorithms, specification of the column with sample identifiers and of the column name to be used for group definition in the differential methylation analysis.

Quality control

Quality control is performed using control probes present in the array and several quality control plots are generated. Moreover, an analysis of values of SNP-based probes can be performed to help in the identification of sample mix-ups.

Preprocessing

Preprocessing includes a series of filtering and normalization steps. Firstly, probes and samples that can bias the normalization procedure are removed. This step includes the removal of SNP-enriched probes as well probes and samples with high fraction of unreliable measures using the Greedycut algorithm. Background subtraction was performed using the methylumi package (Triche et al., 2013). Type I and type II scaling was performed using the SWAN normalization method (Maksimovic et al., 2012). In BTC study data, beta values normalization was also carried out using BMIQ method (Teschendorff et al., 2013). RnBeads generates a beta values density plot to estimate the effects of the applied normalization procedure.

Tracks and Tables

Methylation data are exported in tables of the specified file (.csv) format.

Exploratory analysis

Exploratory analysis includes principal component analysis (PCA) and multidimensional scaling (MDS), clustering analyses and evaluation of batch effects.

Differential methylation analysis

Differential methylation analysis between tumours and normal controls was conducted both on CpG site level and on genomic region (tailing, genes, promoters and CpG Islands) level. p-values were computed for each CpG site using linear models employed in the limma package (Smyth, 2004) and adjusted for multiple testing using the false discovery rate (FDR) method. p-values of all CpG sites associated with a genomic region were combined using a generalization of Fisher's method (Makambi, 2003) and corrected using FDR.

Following analyses were focused on differentially methylated CGIs.

Combined rank scores were generated combining p-values, absolute and relative effect sizes into a single score. These scores are inversely correlated to the degree of differential methylation and are used to generate a list of top-ranked differentially methylated CGIs.

Additional analyses in CLL study

An additional methylome analysis was performed using absolute lymphocyte count data as a covariate in the limma analysis of differential DNA methylation. Moreover, Houseman method (Houseman et al., 2012) was applied to estimate cell-type contributions of whole-blood samples by using methylation profiles of a sorted blood cell types as reference (Reinius et al., 2012).

Additional analyses in BTC study

Type I and type II scaling was repeated using BMIQ normalization method implemented in R "watermelon" package.

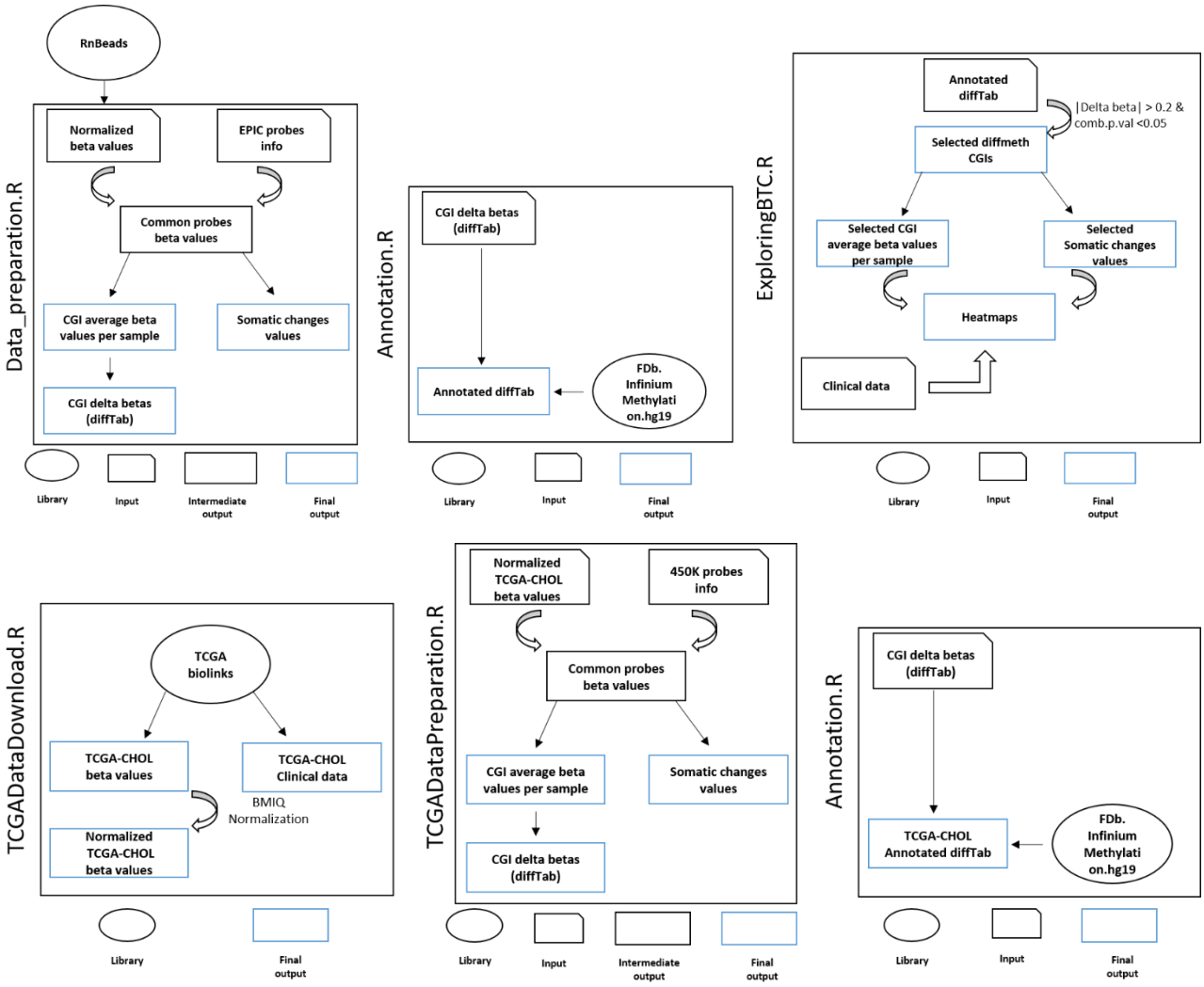
Intensity information were extracted from IDAT files using R function "readIDAT". Mean intensity values were used to generate density plots in R environment.

A second methylome analysis was performed on a restricted dataset including 17 BTCs and nine normal controls.

Data analysis workflow consisted in: methylation data preparation, CGIs annotation, data exploration, data validation and selection of BTC-specific methylation alterations. All the analyses were carried using custom R scripts. Figure 14 summarized the steps performed by each script. ROC curves using methylation data of the selected biomarkers were performed by R "ROCR" package

Gene annotation

The selected CGIs were annotated to nearest genes and to the nearest transcript using R annotation package FDb.InfiniumMethylation.hg19 (Triche, 2014).



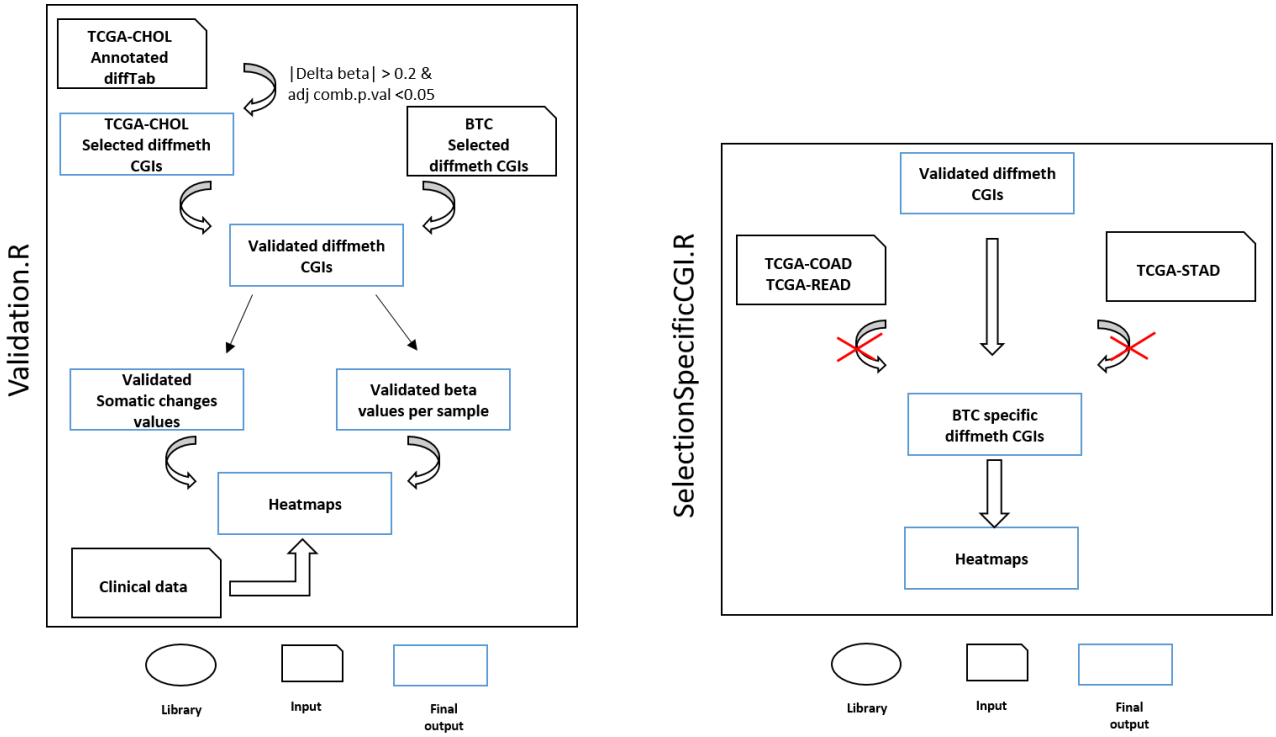


Figure 14. Summary of the steps performed in each script

3.1.3.3 Targeted DNA methylation (MethyLight) data analysis

3.1.3.3.1 MethyLight data analysis

A comparison between CRC and their matched-normal samples have been performed. CRC samples were considered as hypermethylated if delta Ct (mean Ct normal – mean Ct tumour) > 2 and as hypomethylated if delta Ct (mean Ct normal – mean Ct tumour) < -2.

3.1.3.3.2 ddPCR data analysis

Data were analysed using the QuantaSoft 1.7.4 software (Bio-Rad). The droplet counts (positive or negative) from all replicated wells were combined to yield a 'merged' well. Concentration values (number of copies/µl) and Poisson confidence intervals were computed for each "merged" well.

3.1.3.4 Gene expression (qRT-PCR) data analysis

Cycle threshold (Ct) data were analysed using the ΔΔCT method (Schmittgen and Livak, 2008). The statistical difference between average ΔCt of the sample-groups of interest was calculated by unpaired t-test.

3.1.3.5 Protein expression (western blot) data analysis

Western blot signals were quantified using ImageJ program. The intensity of each band was normalized respect to that of NaK ATPase. Statistical differences between tumour and normal samples were calculated using a Welch's t-test.

3.1.3.6 Pyrosequencing analysis

GRIA4 average methylation percentage was calculated by averaging methylation percentages of the eight CpG site analysed.

3.1.3.7 *PCDH* cluster methylation analysis

The methylation status of clustered *PCDH*s was evaluated in solid cancers (CRC, BTC, GC and PA) and in one blood cancer (CLL).

PCDH-associated differentially methylated CGIs (Δβ values ≥ 0.2 or ≤ − 0.2 and an adjusted p-value < 0.05) were selected in methylation data from CRC, BTC, GC and CLL studies. Hypermethylation was defined as Δβ values ≥ 0.2 and adjusted p-value < 0.05, while hypomethylation was defined as Δβ values ≤ -0.2 and adjusted p-value < 0.05. The nominal threshold (p-values < 0.05) was used in the differential methylation analysis between adenomas and normal samples since methylation analysis results were less robust (Fadda et al., 2018). Unsupervised hierarchical clustering (UHC) analysis was carried out using the mean methylation values of each altered CGI for each sample. These methylation values were visualized in a heatmap generated by a Bioconductor package "ComplexHeatmap" (Gu et al., 2016).

27K methylation data of the PA study were analysed as follow: differentially methylated probes (Δβ values ≥ 0.2 or ≤ − 0.2, i.e. 20% differential methylation level) annotated in *PCDH* gene clusters with a p-value threshold < 0.001 were selected. Hypermethylation was defined as Δβ values ≥ 0.2 and p-value threshold < 0.001, while hypomethylation was defined as Δβ values ≤ -0.2 and p-value threshold < 0.001.

To investigate whether *PCDH*-associated altered CGIs were associated with CTCF binding sites, CTCF genomic coordinates were downloaded from ENCODE database (https://www.encodeproject.org/). These two genomic regions were considered associated if their distance was lower than 1000 bp.

3.2 Validation datasets

Processed methylation data from different databases were used to validate the methylation alterations identified in our experimental datasets (Figure 15).

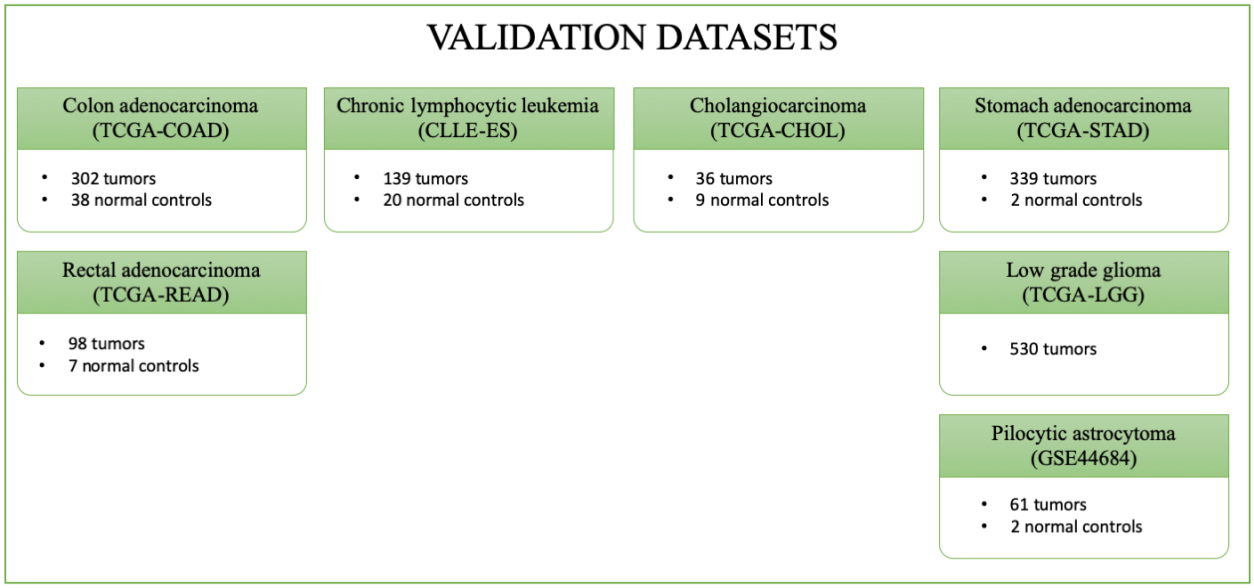


Figure 15. Validation datasets

TCGA validation datasets

Processed Illumina 450K methylation data of colon adenocarcinoma (TCGA-COAD) and rectal adenocarcinoma (TCGA-READ), cholangiocarcinoma (TCGA-CHOL), stomach adenocarcinoma (TCGA-STAD) and low-grade glioma (TCGA-LGG) were downloaded using the Bioconductor package "TCGAbiolinks" (Colaprico et al., 2016) .

GEO portal validation datasets

Methylation data from a PA study (Lambert et al., 2013) were retrieved from the NCBI GEO Portal under the accession number GSE44684.

ICGC validation dataset

Methylation data from a CLL study (Kulis et al., 2012) were retrieved from ICGC Data Portal (DCC Data Release 27, DCC Project Code: CLLE-ES).

4. RESULTS

4.1 DNA methylation as a promising biomarker

Targeted methylation and genome-wide methylation analyses were conducted using experimental datasets of three different types of cancer: CRC, CLL and BTC. Methylation alterations that can be represent potential tumour biomarkers have been identified in all cancer analysed and have been validated in different TCGA datasets.

4.1.1 Colorectal cancer

4.1.1.1 Biomarkers selection

A restricted number of CGIs were selected from a panel of 74 CGIs previously identified by our research group as aberrantly methylated in both CRCs and adenomas (Fadda et al., 2018). All the identified methylation alterations were successfully validated in TCGA-COAD dataset (Fadda et al., 2018). For 36 CGIs, which fulfilled our selection criteria (explained in Materials and Methods), area under the ROC curve (AUC) was calculated using methylation data of our discovery set and methylation data from TCGA-COAD validation set. Twenty-four CGIs showed an AUC higher than 0.95 in both datasets (Table 8).

CGI	Gene	Δβ Discovery set	AUC Discovery set	Δβ Validation set	AUC Validation set
chr2:182321761-182323029	*ITGA4*	0.37	1.00	0.35	0.96
chr4:156129168-156130209	*NPY2R*	0.27	1.00	0.3	0.97
chr4:157997166-157997686	*GLRB*	0.30	1.00	0.36	0.97
chr4:107956555-107957453	*DKK2*	0.32	0.97	0.33	0.97
chr5:136834016-136835146	*SPOCK1*	0.29	1.00	0.33	0.98
chr5:140864527-140864748	*PCDHGA4*	0.35	1.00	0.38	0.97
chr5:1444678-1446648	*SLC6A3*	0.26	1.00	0.29	0.98
chr5:178016558-178017670	*COL23A1*	0.29	0.98	0.32	0.96
chr5:159399004-159399928	*ADRA1B*	0.25	0.97	0.31	0.97
chr6:159589636-159591319	*FNDC1*	0.33	1.00	0.33	0.97
chr6:73330942-73333109	*KCNQ5*	0.36	1.00	0.33	0.96
chr7:28448716-28450028	*CREB5*	0.27	1.00	0.27	0.96
chr7:158936507-158938492	*VIPR2*	0.33	0.96	0.35	0.96
chr8:97505747-97507607	*SDC2*	0.29	1.00	0.36	0.96
chr8:75896528-75897116	*CRISPLD1*	0.21	0.96	0.27	0.97
chr10:15761423-15762101	*ITGA8*	0.30	0.96	0.35	0.97
chr11:105481126-105481422	*GRIA4*	0.40	1.00	0.41	0.96
chr11:133938850-133939681	*JAM3*	0.30	1.00	0.29	0.97
chr12:117798076-117799448	*NOS1*	0.25	0.97	0.27	0.96
chr13:110958891-110960590	*COL4A1*	0.33	0.96	0.37	0.98
chr16:23846941-23848102	*PRKCB*	0.25	0.98	0.34	0.97

chr19:48918115-48918340	*GRIN2D*	0.33	1.00	0.38	0.96
chr21:28337856-28340237	*ADAMTS5*	0.29	1.00	0.31	0.99
chr22:33453892-33454505	*SYN3*	0.28	0.97	0.32	0.97

Table 8. Selected CpG islands showing an AUC > 0.95 in our discovery set and in TCGA-COAD validation set

Based on the feasibility of the assay, functional evidences and literature information, two CGIs were selected among the 24 CGIs for a targeted methylation analysis. These two CGIs (chr11:105481126-105481422 and chr7:158936507-158938492), hypermethylated in CRC compared to normal controls, respectively map on promoter regions of *GRIA4* and *VIPR2* genes (Figure 16). *VIPR2* was selected for its functional role and the involvement of vasoactive intestinal peptide receptors in cancer (Moody et al., 2016), while *GRIA4* was chosen because our and other previous studies have detected methylation of this marker both in CRC tissues but also in CRC stool and plasma samples (Barault et al., 2018; Fadda et al., 2018).

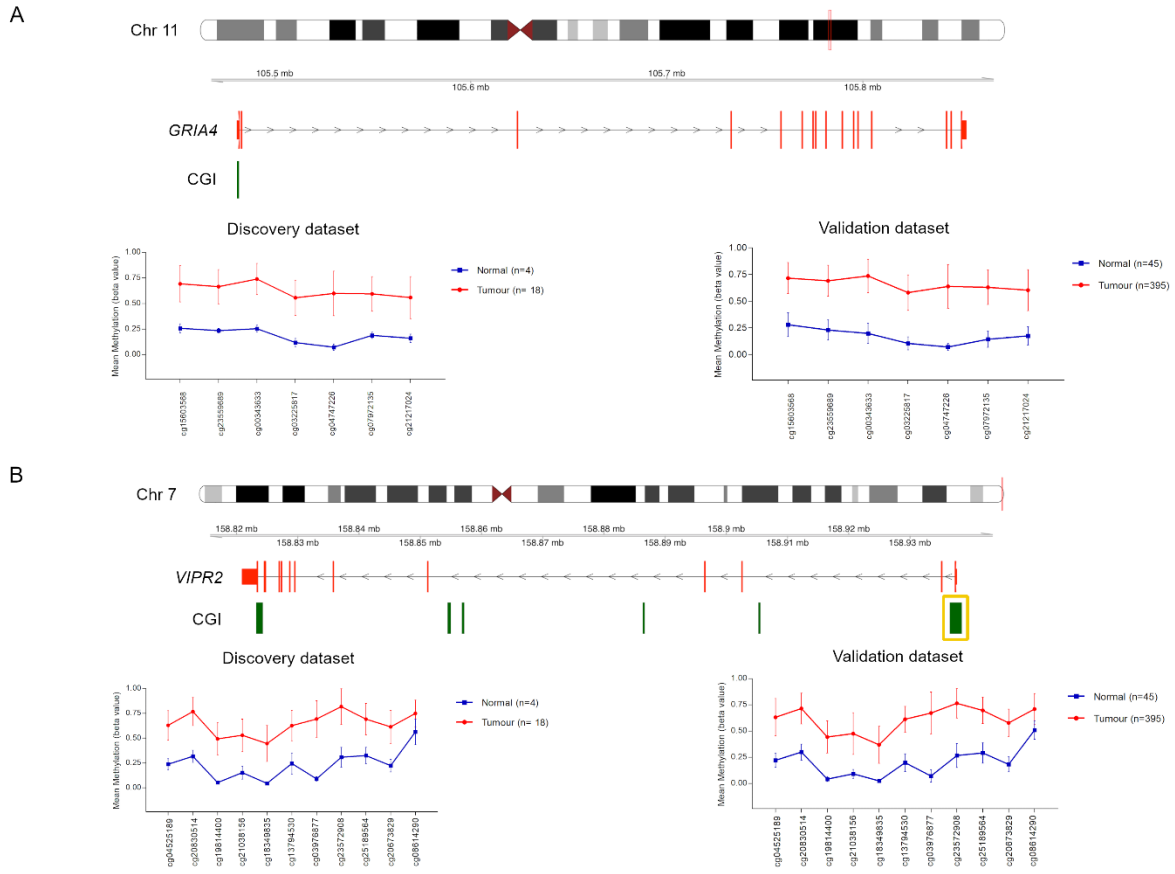


Figure 16. CGIs position relative to *GRIA4* (A) and *VIPR2* (B) genes and their methylation levels in our discovery set and in TCGA-COAD validation dataset
Notes: *VIPR2*-associated CGI is enclosed in an orange box

4.1.1.2 Targeted methylation analyses

GRIA4 and *VIPR2* methylation analyses were performed in DNA extracted from tumour and matched-normal tissue samples of 10 CRC patients by MethyLight. Methylation levels of tumour samples were compared to methylation levels of their respective normal samples.

GRIA4 resulted hypermethylated in 6/10 tumour samples, while *VIPR2* was hypermethylated in 7/10 tumour samples. Some tumour samples showed *GRIA4* (2/10 samples) or *VIPR2* (2/10 samples) methylation levels similar to their respective normal samples, while for other tumour and matched-normal samples, methylation of *GRIA4* (2/10 samples) and *VIPR2* (1/10 sample) was not detected (Table 9).

To investigate the possible causes of these methylation differences among the analysed samples, we performed a histopathological examination of the tissue specimens. Low content of tumour cells was observed in histologic slides from tumour samples with undetectable methylation or methylation levels similar to the respective normal samples. On the other hand, histologic slides from tumour samples with high methylation levels showed a high content of tumour cells.

To explore the possible usability of these markers for CRC early detection through non-invasive methods, *GRIA4* and *VIPR2* methylation was also evaluated in stool samples from the same 10 patients by MethyLight.

GRIA4 methylation was detected in 4/10 samples, while *VIPR2* methylation was detected in 7/10 samples (Table 9).

Since MethyLight might not be enough sensitive for the detection of low targets such as methylated DNA from stool samples, the same samples were analysed using a much more sensitive technique as ddPCR. This method allowed to detect methylation of both *GRIA4* and *VIPR2* in 9/10 samples (Table 9). Higher levels of *VIPR2* methylated DNA were observed in ulcerating tumours compared to non-ulcerating tumours.

	GRIA4			***VIPR2***		
	Tumour tissue sample	**Stool sample MethyLight**	**Stool sample ddPCR**	**Tumour tissue sample**	**Stool sample MethyLight**	**Stool sample ddPCR**
CRC_2	Hypermethylated	Methylated	Methylated	Hypermethylated	Methylated	Methylated
CRC_3	Hypermethylated	Methylated	Methylated	Hypermethylated	Methylated	Methylated
CRC_8	Undetectable methylation	Undetectable methylation	Methylated	Undetectable methylation	Methylated	Methylated
CRC_12	Undetectable methylation	Undetectable methylation	Methylated	Hypermethylated	Undetectable methylation	Methylated

CRC_14	Not differentially methylated	Undetectable methylation	Methylated	Hypermethylated	Methylated	Methylated
CRC_19	Hypermethylated	Methylated	Methylated	Not differentially methylated	Methylated	Methylated
CRC_21	Hypermethylated	Undetectable methylation	Undetectable methylation	Hypermethylated	Undetectable methylation	Undetectable methylation
CRC_29	Not differentially methylated	Methylated	Methylated	Hypermethylated	Methylated	Methylated
CRC_33	Hypermethylated	Undetectable methylation	Methylated	Hypermethylated	Methylated	Methylated
CRC_34	Hypermethylated	Undetectable methylation	Methylated	Not differentially methylated	Undetectable methylation	Methylated

Table 9. *GRIA4* and *VIPR2* methylation analyses results

4.1.2 Chronic lymphocytic leukemia

4.1.2.1 Genome-wide methylation analysis

A genome-wide methylation analysis on 18 CLL cases and six normal controls was conducted using 450K arrays. CLL methylome resulted characterized by a widespread hypomethylation and a focal hypermethylation of regulatory regions, especially CGIs (Figure 17).

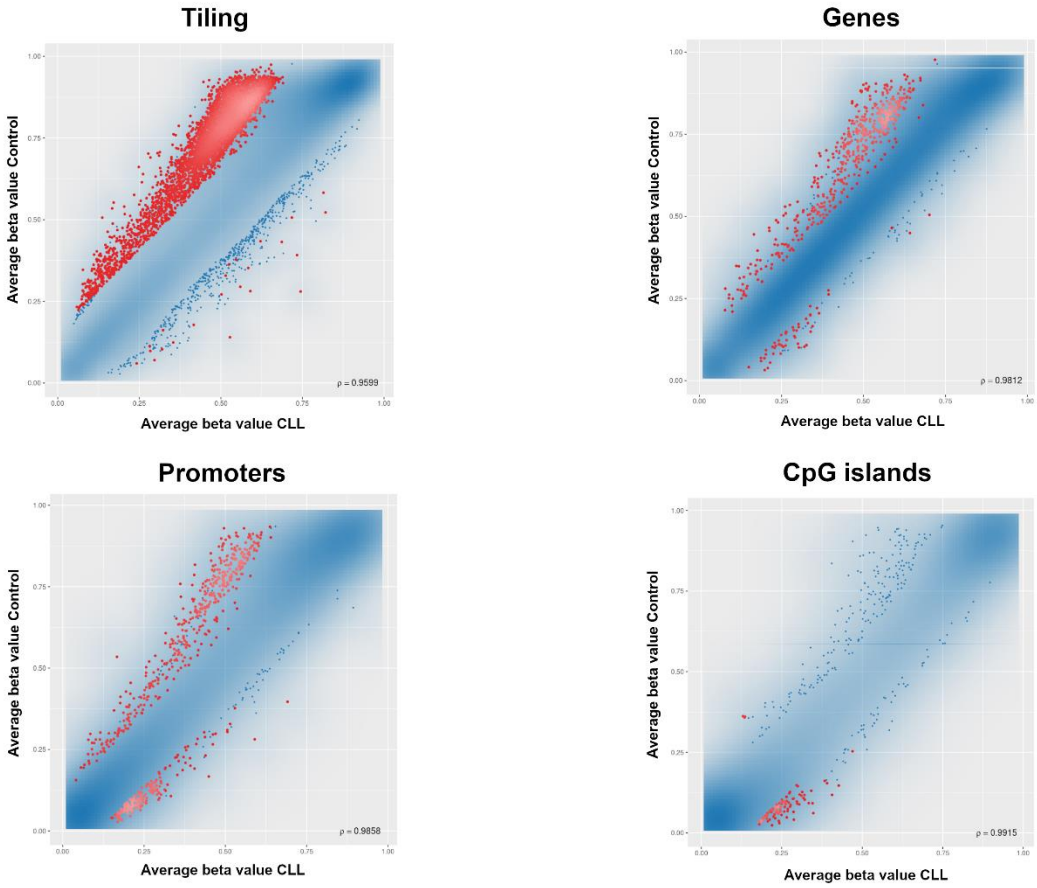


Figure 17. Average beta values distribution of CLL and normal control samples at different region levels (tiling, promoters, genes and CpG Islands). Red dots represent statistically significant differentially methylated CpG loci (automatically selected rank cutoff)

Differential methylation analysis did not reveal any CGIs significantly hypermethylated after p-value correction in CLL samples compared to normal controls, probably due to the fact that methylation assay was carried out on DNA isolated from whole blood samples, characterized by high cellular heterogeneity. Sample cellular heterogeneity was explored by estimating the relative contributions of each principal immune components of whole blood (B cells, granulocytes, monocytes, NK cells, and T cells subsets) using Houseman method (Houseman et al., 2012). Granulocytes contributed for most of the relative cell type composition in whole blood control samples.

A second differential methylation analysis was performed using absolute lymphocyte count data as a covariate in the limma model. In fact, absolute lymphocytic counts, obviously higher in CLL samples than in normal controls, might represent potential confounding factors that should be taken into consideration when computing the site-specific p-values. There was no substantial difference in the order of differentially methylated CGIs in the 100-top ranked lists generated in the two differential methylation analyses suggesting that the identified methylation differences were not due to a different number of circulating lymphocytes between CLL cases and controls (Table 10).

DIFFERENTIAL METHYLATION ANALYSIS BEFORE COVARIATE ADJUSTMENT								DIFFERENTIAL METHYLATION ANALYSIS AFTER COVARIATE ADJUSTMENT							
CpG Islands	**Mean β CLL**	**Mean β CTRL**	**Mean Δβ**	**Comb. P-value**	**Comb. P-value adj.**	**Distance**	**Gene Symbol**	**CpG Islands**	**Mean β CLL**	**Mean β CTRL**	**Mean Δβ**	**Comb. P-value**	**Comb. P-value adj.**	**Distance**	**Gene Symbol**
chr19:51198144-51198460	0.41	0.11	0.30	0.0013	0.2518	0	*SHANK1*	chr19:51198144-51198460	0.41	0.12	0.29	0.0023	0.2641	0	*SHANK1*
chr9:116860474-116860695	0.35	0.06	0.29	0.0016	0.2746	0	*KIF12*	chr9:116860474-116860695	0.35	0.06	0.29	0.0024	0.2641	0	*KIF12*
chr4:5894072-5895116	0.45	0.16	0.29	0.0016	0.2746	0	*CRMP1*	chr4:5894072-5895116	0.43	0.15	0.28	0.0025	0.2671	0	*CRMP1*
chr2:200524002-200524204	0.38	0.10	0.28	0.0012	0.2518	101054	*FTCDNL1*	chr2:200524002-200524204	0.35	0.09	0.25	0.0014	0.2435	101054	*FTCDNL1*
chr12:127210779-127211651	0.43	0.15	0.27	0.0044	0.2933	3595	*LINC00944*	chr9:23850911-23851522	0.29	0.04	0.25	0.0018	0.2599	24847	*ELAVL2*
chr11:125774293-125774584	0.35	0.10	0.24	0.0017	0.2803	0	*DDX25*	chr11:125774293-125774584	0.33	0.09	0.24	0.0021	0.2641	0	*DDX25*
chr9:23850911-23851522	0.29	0.05	0.24	0.0019	0.2803	24847	*ELAVL2*	chr19:52452317-52452543	0.39	0.15	0.24	0.0056	0.2811	3305	*ZNF613*
chr20:13975769-13976287	0.34	0.10	0.24	0.0057	0.3057	0	*SEL1L2*	chr8:35092680-35093559	0.38	0.16	0.22	0.0052	0.2786	0	*UNC5D*
chr19:52452317-52452543	0.39	0.15	0.23	0.0064	0.3057	3305	*ZNF613*	chr19:30865684-30866490	0.29	0.07	0.22	0.0012	0.2425	0	*ZNF536*
chr8:35092680-35093559	0.40	0.17	0.23	0.0030	0.2834	0	*UNC5D*	chr8:120220389-120221222	0.32	0.10	0.22	0.0019	0.2605	0	*MAL2*
chr19:30865684-30866490	0.31	0.08	0.23	0.0015	0.2746	0	*ZNF536*	chr9:90589210-90589807	0.27	0.05	0.22	0.0001	0.1000	0	*CDK20*
chr8:120220389-120221222	0.33	0.11	0.23	0.0012	0.2518	0	*MAL2*	chr19:54485304-54486322	0.47	0.25	0.22	0.0015	0.2477	0	*CACNG8*
chr3:170303533-170303768	0.37	0.14	0.23	0.0063	0.3057	0	*SLC7A14*	chr8:24770909-24772547	0.30	0.09	0.22	0.0008	0.2062	0	*NEFM*
chr9:90589210-90589807	0.29	0.06	0.23	0.0002	0.1151	0	*CDK20*	chr11:134201785-134202407	0.24	0.03	0.21	0.0006	0.1869	0	*GLB1L2*
chr8:24770909-24772547	0.32	0.10	0.23	0.0007	0.2184	0	*NEFM*	chr13:24121204-24121465	0.34	0.13	0.21	0.0006	0.1777	23043	*TNFRSF19*
chr19:54485304-54486322	0.48	0.26	0.23	0.0011	0.2505	0	*CACNG8*	chr14:38724255-38725537	0.32	0.11	0.21	0.0031	0.2684	0	*CLEC14A*
chr13:24121204-24121465	0.36	0.14	0.22	0.0009	0.2379	23043	*TNFRSF19*	chr5:16179065-16180420	0.30	0.09	0.21	0.0049	0.2754	0	*MARCH11*

chr14:38724255-38725537	0.34	0.12	0.22	0.0024	0.2803	0	*CLEC14A*	chr6:1378446-1379318	0.24	0.04	0.20	0.0036	0.2706	10750	*FOXF2*
chr11:134201785-134202407	0.25	0.04	0.22	0.0007	0.2161	0	*GLB1L2*	chr5:174151479-174152364	0.32	0.12	0.20	0.0024	0.2641	0	*MSX2*
chr5:174151479-174152364	0.34	0.12	0.21	0.0024	0.2803	0	*MSX2*	chr1:61508643-61509282	0.27	0.07	0.20	0.0005	0.1721	33663	*NFIA*
chr5:72746990-72747587	0.31	0.10	0.21	0.0008	0.2318	46662	*BTF3*	chr1:65991002-65991811	0.30	0.10	0.20	0.0059	0.2811	0	*LEPR*
chr1:61508643-61509282	0.30	0.08	0.21	0.0006	0.2161	33663	*NFIA*	chr19:5827755-5828405	0.22	0.02	0.20	0.0001	0.1000	0	*NRTN*
chr11:30038523-30038823	0.33	0.12	0.21	0.0021	0.2803	0	*KCNA4*	chr11:30038523-30038823	0.31	0.12	0.20	0.0037	0.2706	0	*KCNA4*
chr5:16179065-16180420	0.30	0.09	0.21	0.0040	0.2933	0	*MARCH11*	chr4:164252955-164253471	0.27	0.07	0.19	0.0056	0.2811	0	*NPY1R*
chr7:32467435-32467948	0.34	0.14	0.21	0.0036	0.2927	28561	*LOC100130673*	chr5:72746990-72747587	0.28	0.08	0.19	0.0009	0.2062	46662	*BTF3*
chr14:85997469-85998637	0.28	0.08	0.21	0.0020	0.2803	0	*FLRT2*	chr2:176993480-176995557	0.25	0.06	0.19	0.0014	0.2435	0	*HOXD8*
chr19:5827755-5828405	0.23	0.03	0.20	0.0001	0.1151	0	*NRTN*	chr14:85997469-85998637	0.26	0.07	0.19	0.0018	0.2599	0	*FLRT2*
chr2:176993480-176995557	0.26	0.06	0.20	0.0013	0.2518	0	*HOXD8*	chr11:132952539-132953307	0.28	0.09	0.19	0.0010	0.2247	0	*OPCML*
chr20:11871375-11872207	0.26	0.06	0.20	0.0050	0.3018	0	*BTBD3*	chr9:79630969-79631749	0.31	0.12	0.19	0.0024	0.2641	2821	*FOXB2*
chr1:65991002-65991811	0.31	0.11	0.20	0.0065	0.3057	0	*LEPR*	chr17:15686219-15686474	0.31	0.13	0.19	0.0031	0.2684	3689	*MEIS3P1*
chr6:1378446-1379318	0.24	0.04	0.20	0.0039	0.2933	10750	*FOXF2*	chr15:100880958-100882438	0.23	0.05	0.18	0.0002	0.1178	0	*ADAMTS17*
chr2:119606039-119606313	0.32	0.12	0.20	0.0029	0.2834	279	*EN1*	chr2:119606039-119606313	0.29	0.11	0.18	0.0032	0.2687	279	*EN1*
chr11:132952539-132953307	0.31	0.11	0.20	0.0019	0.2803	0	*OPCML*	chr9:98273679-98273908	0.25	0.07	0.18	0.0020	0.2635	0	*PTCH1*
chr9:79630969-79631749	0.32	0.13	0.20	0.0023	0.2803	2821	*FOXB2*	chr8:106330519-106332120	0.25	0.07	0.18	0.0014	0.2435	0	*ZFPM2*
chr8:106330519-106332120	0.27	0.08	0.20	0.0013	0.2518	0	*ZFPM2*	chr5:175223610-175224679	0.24	0.06	0.18	0.0028	0.2684	0	*CPLX2*
chr15:28753440-28753940	0.31	0.11	0.19	0.0026	0.2803	10816	*GOLGA8G*	chr4:7940564-7941853	0.28	0.11	0.18	0.0012	0.2421	0	*AFAP1*
chr15:100880958-100882438	0.24	0.05	0.19	0.0002	0.1464	0	*ADAMTS17*	chr10:100992157-100992687	0.26	0.08	0.18	0.0007	0.1917	0	*HPSE2*

chr4:4873263-4873613	0.28	0.09	0.19	0.0008	0.2318	7602	*MSX1*	chr4:128544032-128544903	0.27	0.09	0.18	0.0046	0.2718	9183	*INTU*
chr17:15686219-15686474	0.32	0.13	0.19	0.0027	0.2805	3689	*MEIS3P1*	chr4:118006539-118006859	0.22	0.05	0.17	0.0020	0.2641	0	*TRAM1L1*
chr2:130737206-130737718	0.26	0.07	0.19	0.0053	0.3057	0	*RAB6C-AS1*	chr1:236227683-236228817	0.24	0.06	0.17	0.0050	0.2754	0	*NID1*
chr9:98273679-98273908	0.27	0.08	0.19	0.0029	0.2834	0	*PTCH1*	chr20:55500348-55501102	0.25	0.07	0.17	0.0035	0.2706	242706	*BMP7*
chr10:100992157-100992687	0.26	0.08	0.19	0.0009	0.2379	0	*HPSE2*	chr4:4873263-4873613	0.26	0.09	0.17	0.0011	0.2347	7602	*MSX1*
chr4:7940564-7941853	0.30	0.11	0.18	0.0011	0.2505	0	*AFAP1*	chr4:55092962-55093242	0.21	0.04	0.17	0.0047	0.2727	0	*PDGFRA*
chr4:128544032-128544903	0.28	0.10	0.18	0.0043	0.2933	9183	*INTU*	chr19:54483022-54483572	0.21	0.04	0.17	0.0013	0.2432	0	*CACNG8*
chr20:55500348-55501102	0.27	0.08	0.18	0.0037	0.2933	242706	*BMP7*	chr2:183902403-183903625	0.23	0.06	0.17	0.0005	0.1681	0	*NCKAP1*
chr5:175223610-175224679	0.25	0.07	0.18	0.0033	0.2861	0	*CPLX2*	chr19:54393040-54393300	0.26	0.09	0.17	0.0052	0.2786	0	*PRKCG*
chr2:183902403-183903625	0.25	0.07	0.18	0.0003	0.1548	0	*NCKAP1*	chr5:87437096-87437505	0.23	0.06	0.17	0.0014	0.2432	53517	*TMEM161B*
chr5:87437096-87437505	0.26	0.08	0.18	0.0024	0.2803	53517	*TMEM161B*	chr19:31839636-31843049	0.24	0.07	0.17	0.0022	0.2641	0	*TSHZ3*
chr1:236227683-236228817	0.25	0.07	0.18	0.0042	0.2933	0	*NID1*	chr1:152487979-152488270	0.25	0.08	0.17	0.0059	0.2811	0	*CRCT1*
chr4:164252955-164253471	0.26	0.08	0.18	0.0044	0.2933	0	*NPY1R*	chr18:55094826-55096310	0.27	0.10	0.17	0.0052	0.2786	6606	*ONECUT2*
chr11:12398966-12399863	0.27	0.09	0.17	0.0018	0.2803	0	*PARVA*	chr11:12398966-12399863	0.25	0.08	0.17	0.0014	0.2435	0	*PARVA*
chr6:19837505-19839314	0.28	0.11	0.17	0.0025	0.2803	0	*ID4*	chr6:19837505-19839314	0.26	0.10	0.16	0.0030	0.2684	0	*ID4*
chr4:55092962-55093242	0.22	0.05	0.17	0.0063	0.3057	0	*PDGFRA*	chr9:1042418-1042973	0.25	0.09	0.16	0.0028	0.2684	7372	*DMRT2*
chr19:54483022-54483572	0.21	0.04	0.17	0.0016	0.2746	0	*CACNG8*	chr2:223161532-223161919	0.20	0.04	0.16	0.0015	0.2435	0	*PAX3*
chr18:55094826-55096310	0.29	0.11	0.17	0.0062	0.3057	6606	*ONECUT2*	chr16:54962423-54967805	0.25	0.09	0.16	0.0031	0.2684	0	*CRNDE*
chr19:31839636-31843049	0.25	0.08	0.17	0.0022	0.2803	0	*TSHZ3*	chr1:94702691-94703344	0.23	0.07	0.16	0.0031	0.2684	0	*ARHGAP29*
chr4:118006539-118006859	0.23	0.05	0.17	0.0024	0.2803	0	*TRAM1L1*	chr2:45511-46559	0.24	0.08	0.16	0.0023	0.2641	0	*FAM110C*

chr9:1042418-1042973	0.27	0.10	0.17	0.0031	0.2834	7372	*DMRT2*	chr7:27203916-27206462	0.23	0.07	0.16	0.0046	0.2718	0	*HOXA10-HOXA9*
chr18:75611919-75612142	0.20	0.03	0.17	0.0002	0.1151	629822	*GALR1*	chr8:41165853-41167140	0.24	0.08	0.16	0.0043	0.2718	0	*SFRP1*
chr2:45511-46559	0.26	0.09	0.17	0.0012	0.2518	0	*FAM110C*	chr1:50489418-50489846	0.24	0.08	0.16	0.0042	0.2718	0	*AGBL4*
chr8:41165853-41167140	0.26	0.09	0.17	0.0037	0.2933	0	*SFRP1*	chr20:48098513-48099560	0.30	0.14	0.15	0.0052	0.2786	0	*KCNB1*
chr16:54962423-54967805	0.27	0.10	0.17	0.0029	0.2834	0	*CRNDE*	chr7:20817456-20818227	0.24	0.08	0.15	0.0028	0.2684	3666	*SP8*
chr3:45187027-45187946	0.23	0.06	0.17	0.0060	0.3057	0	*CDCP1*	chr4:147576110-147576762	0.21	0.06	0.15	0.0067	0.2848	12486	*POU4F2*
chr15:26327496-26327896	0.22	0.05	0.17	0.0002	0.1464	29228	*LINC02346*	chr18:75611919-75612142	0.18	0.03	0.15	0.0002	0.1158	629822	*GALR1*
chr7:27212417-27214396	0.26	0.09	0.17	0.0030	0.2834	0	*HOXA10-HOXA9*	chr3:145878431-145879287	0.22	0.06	0.15	0.0039	0.2718	0	*PLOD2*
chr7:27203916-27206462	0.25	0.08	0.17	0.0039	0.2933	0	*HOXA10-HOXA9*	chr13:53419898-53422872	0.26	0.11	0.15	0.0044	0.2718	0	*PCDH8*
chr1:91183241-91184540	0.24	0.07	0.17	0.0058	0.3057	446	*BARHL2*	chr20:590223-591222	0.25	0.10	0.15	0.0046	0.2718	0	*TCF15*
chr20:4228533-4230496	0.31	0.15	0.16	0.0027	0.2805	0	*ADRA1D*	chr4:85503547-85504893	0.19	0.04	0.15	0.0036	0.2706	0	*CDS1*
chr2:223161532-223161919	0.22	0.05	0.16	0.0016	0.2746	0	*PAX3*	chr12:59313560-59314452	0.20	0.05	0.15	0.0026	0.2671	0	*LRIG3*
chr1:94702691-94703344	0.24	0.08	0.16	0.0034	0.2861	0	*ARHGAP29*	chr2:177052958-177054350	0.24	0.09	0.15	0.0060	0.2811	0	*HAGLR*
chr5:35617856-35618339	0.20	0.04	0.16	0.0003	0.1536	0	*SPEF2*	chr5:35617856-35618339	0.19	0.04	0.15	0.0005	0.1655	0	*SPEF2*
chr12:59313560-59314452	0.22	0.06	0.16	0.0031	0.2834	0	*LRIG3*	chr18:73167403-73167920	0.27	0.12	0.15	0.0057	0.2811	27813	*SMIM21*
chr1:50489418-50489846	0.25	0.09	0.16	0.0049	0.2994	0	*AGBL4*	chr19:51227662-51228883	0.22	0.07	0.15	0.0014	0.2432	0	*CLEC11A*
chr19:51227662-51228883	0.24	0.08	0.16	0.0017	0.2803	0	*CLEC11A*	chr7:27212417-27214396	0.23	0.08	0.15	0.0038	0.2706	0	*HOXA10-HOXA9*
chr3:145878431-145879287	0.23	0.07	0.16	0.0038	0.2933	0	*PLOD2*	chr1:48937305-48937683	0.20	0.05	0.15	0.0018	0.2599	0	*SPATA6*
chr2:177052958-177054350	0.26	0.10	0.16	0.0038	0.2933	0	*HAGLR*	chr3:39851026-39851820	0.18	0.03	0.15	0.0014	0.2435	0	*MYRIP*
chr19:54409967-54410200	0.20	0.04	0.16	0.0046	0.2933	0	*PRKCG*	chr20:4228533-4230496	0.30	0.15	0.15	0.0054	0.2801	0	*ADRA1D*

chr18:73167403-73167920	0.28	0.13	0.16	0.0055	0.3057	27813	*SMIM21*	chr15:96873409-96877721	0.28	0.13	0.15	0.0053	0.2786	0	*NR2F2*
chr2:132121264-132121762	0.20	0.04	0.16	0.0034	0.2861	0	*RAB6D*	chr10:118927022-118928132	0.23	0.08	0.15	0.0042	0.2718	0	*MIR3663*
chr4:85503547-85504893	0.20	0.04	0.16	0.0027	0.2805	0	*CDS1*	chr6:80656745-80657593	0.19	0.04	0.15	0.0004	0.1400	0	*ELOVL4*
chr20:590223-591222	0.26	0.11	0.16	0.0045	0.2933	0	*TCF15*	chr12:88973461-88974666	0.21	0.07	0.15	0.0053	0.2786	0	*KITLG*
chr7:20817456-20818227	0.26	0.10	0.16	0.0036	0.2899	3666	*SP8*	chr7:28995306-28998541	0.25	0.10	0.15	0.0026	0.2672	0	*TRIL*
chr10:47008085-47008410	0.22	0.07	0.16	0.0033	0.2861	3345	*ANXA8*	chr2:132121264-132121762	0.19	0.04	0.15	0.0046	0.2718	0	*RAB6D*
chr8:24812947-24814299	0.24	0.08	0.16	0.0038	0.2933	0	*NEFL*	chr16:31580560-31581023	0.19	0.04	0.15	0.0051	0.2764	0	*YBX3P1*
chr10:8091375-8098329	0.22	0.07	0.16	0.0018	0.2803	0	*GATA3-AS1*	chr14:85999533-86000478	0.22	0.08	0.15	0.0044	0.2718	0	*FLRT2*
chr7:28995306-28998541	0.27	0.11	0.16	0.0025	0.2803	0	*TRIL*	chr10:47008085-47008410	0.20	0.06	0.15	0.0031	0.2684	3345	*ANXA8*
chr12:88973461-88974666	0.24	0.08	0.16	0.0056	0.3057	0	*KITLG*	chr12:20521617-20523122	0.23	0.09	0.15	0.0039	0.2718	0	*PDE3A*
chr3:39851026-39851820	0.19	0.04	0.16	0.0015	0.2746	0	*MYRIP*	chr2:211089414-211090176	0.20	0.05	0.15	0.0027	0.2684	0	*ACADL*
chr12:20521617-20523122	0.26	0.10	0.16	0.0042	0.2933	0	*PDE3A*	chr1:65775019-65775746	0.19	0.05	0.15	0.0036	0.2706	0	*DNAJC6*
chr10:118927022-118928132	0.25	0.09	0.15	0.0048	0.2994	0	*MIR3663*	chr8:72468561-72469561	0.23	0.08	0.15	0.0066	0.2848	194093	*EYA1*
chr2:211089414-211090176	0.22	0.06	0.15	0.0022	0.2803	0	*ACADL*	chr7:79081566-79081879	0.24	0.09	0.15	0.0067	0.2848	0	*MAGI2*
chr13:53419898-53422872	0.27	0.11	0.15	0.0045	0.2933	0	*PCDH8*	chr6:39692744-39692966	0.18	0.04	0.15	0.0043	0.2718	0	*KIF6*
chr20:8112885-8113592	0.21	0.05	0.15	0.0016	0.2803	0	*PLCB1*	chr10:8091375-8098329	0.21	0.06	0.14	0.0021	0.2641	0	*GATA3-AS1*
chr6:80656745-80657593	0.20	0.04	0.15	0.0004	0.1652	0	*ELOVL4*	chr20:8112885-8113592	0.19	0.05	0.14	0.0017	0.2594	0	*PLCB1*
chr8:72468561-72469561	0.24	0.09	0.15	0.0051	0.3044	194093	*EYA1*	chr8:24812947-24814299	0.21	0.07	0.14	0.0044	0.2718	0	*NEFL*
chr8:12990091-12990914	0.20	0.05	0.15	0.0039	0.2933	0	*DLC1*	chr22:17083385-17083628	0.18	0.04	0.14	0.0047	0.2727	0	*TPTEP1*
chr7:79081566-79081879	0.26	0.11	0.15	0.0065	0.3057	0	*MAGI2*	chr6:125283125-125284389	0.20	0.06	0.14	0.0056	0.2811	0	*RNF217-AS1*

chr1:27961560-27961810	0.15	0.37	-0.22	3.52E-06	0.015163	0	*FGR*	chr5:139040820-139041028	0.13	0.36	-0.22	4.37E-06	0.010331	0	*CXXC5*
chr5:139040820-139041028	0.15	0.40	-0.25	4.52E-06	0.016705	0	*CXXC5*	chr1:27961560-27961810	0.14	0.36	-0.22	2.66E-06	0.009826	0	*FGR*
chr11:47399789-47400006	0.16	0.41	-0.25	3.26E-06	0.015163	0	*SLC39A13*	chr11:47399789-47400006	0.13	0.36	-0.23	1.59E-06	0.008215	0	*SLC39A13*

Table 10. 100-top ranked differentially methylated CpG islands in CLL discovery set

4.1.2.2 *SHANK1*-associated CGI methylation in different datasets

A CGI (chr19:51198144-51198460, referred to hg19) located in the gene body of *SHANK1* showed the highest differential methylation value (Δβ = 0.29) between CLL and control samples (Table 10). A statistically significant positive correlation (r = 0.78, p-value = 0.0045) (Figure 18A) between *SHANK1* methylation percentage and absolute lymphocyte count data was observed. Moreover, CLL samples with high lymphocyte counts, showed high CD19+ B-cells contributions and higher *SHANK1* methylation values, compared to samples with low CD19+ B-cells contributions (Table 11). Correlation analysis confirmed a strong positive correlation (r = 0.91, p-value < 0.0001) between CD19+ B-cells contribution and *SHANK1* methylation (Figure 18B).

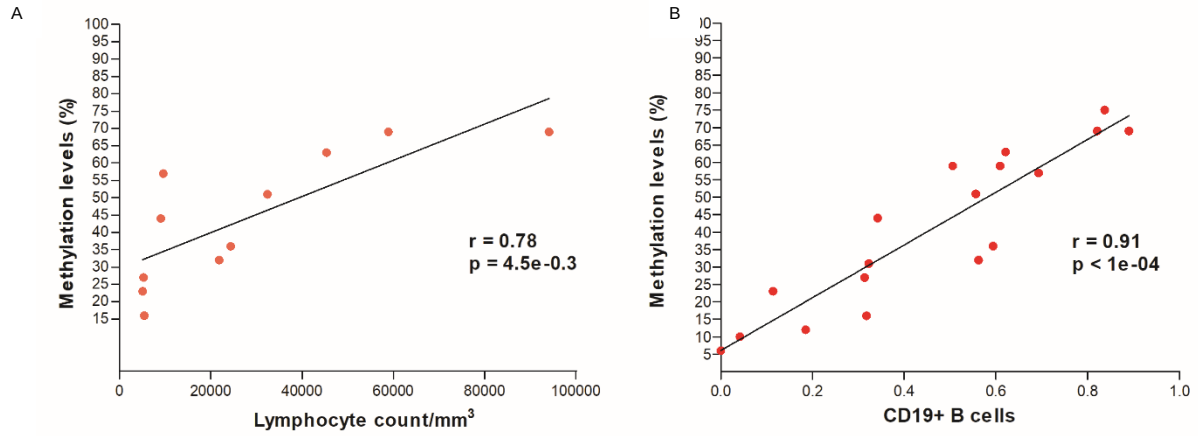


Figure 18. Correlation analysis between methylation percentage and absolute lymphocyte count. *SHANK1* methylation levels (%) are plotted against absolute lymphocyte count values (A); Correlation analysis between methylation percentage and CD19+ B-cells. *SHANK1* methylation levels (%) are plotted against CD19+ cells contribution values (B)

	CD14+ monocytes	CD19+ B cells	CD4+ T cells	CD56+ nk cells	CD8+ t cells	Granulocytes	Lymphocyte count/mm^3	*SHANK1* methylation value (%)
304012_002	0,104	0,609	-0,010	0,001	0,000	0,245	NA	59
304012_007	0,076	0,114	-0,001	0,011	0,000	0,797	5050	23
304012_030	0,061	0,556	0,000	0,040	0,000	0,355	32380	51
304012_048	0,074	0,693	0,000	0,049	0,000	0,122	9580	57
304032_083	0,084	0,837	0,038	0,000	0,000	0,002	NA	75
304012_088	0,062	0,621	-0,008	0,050	0,000	0,289	45330	63
304012_092	0,099	0,562	0,000	0,146	0,000	0,154	21830	32
304032_100	0,061	0,506	0,000	0,150	0,000	0,323	NA	59
304032_104	0,019	0,323	0,002	0,323	0,000	0,313	NA	31
304012_112	0,073	0,314	0,000	0,056	0,000	0,627	5270	27
304012_114	0,057	0,594	0,000	0,092	-0,001	0,291	24360	36
304032_132	0,000	0,000	-0,016	0,000	0,000	0,988	NA	6
304032_134	0,070	0,185	0,000	0,049	0,000	0,720	NA	12
304012_188	0,089	0,318	0,000	0,104	0,000	0,484	5410	16
304012_193	0,047	0,890	-0,028	0,000	-0,006	0,015	58880	69
304012_196	0,090	0,342	0,000	0,099	0,000	0,549	9060	44
304012_198	0,065	0,042	0,096	0,157	0,000	0,672	NA	10

304012_475	0,068	0,821	0,000	0,001	0,000	0,000	94100	69

Table 11. Comparison between estimated relative leucocyte contributions, lymphocyte count and *SHANK1* methylation values in CLL samples

SHANK1-associated CGI methylation alteration was successfully validated in a publicly available dataset including 139 CLLs (≥95% neoplastic cells) and 20 normal samples (normal B-cells from peripheral blood including total B cells and various subtypes of B-cells) (Kulis et al., 2012). This CGIs showed a statistically significant differential methylation (Δβ = 0.26, p-value = 8.66e-12, adjusted p-value = 2.43e-10) between CLL and normal controls (Table 12). Moreover, the same CGI resulted hypermethylated (false discovery rate <0.05) in a subgroup of 59 CLL with a low or absent IGHV mutational load (U-CLLs) compared to 6 naïve B cells (CD5+NBC/NBC) (Kulis et al., 2012) (Table 12).

To investigate whether this methylation alterations may be a potential biomarker to predict the risk of developing CLL or other MBCN diseases, a cohort of 438 MBCN cases, including 82 CLL/SLL, and matched-normal controls was analysed. A statistically significant gain of methylation in the *SHANK1*-associated CGI was detected both in the prospective set including 82 CLL and SLL (Δβ = 0.047, p = 0.00863, adjusted p-value = 0,0921) and in the larger dataset of 438 MBCN cases, including the 82 CLL/SLL (Δβ = 0.03, p-value < 10-7, adjusted p-value = 7.47e-05) between cases and controls (Table 12, Figure 19).

	Mean ΔB	**p-value**	**Adjusted p-value**
18 CLLs vs 6 normal controls (experimental dataset)	0.29	0.0023	0.2641
139 CLLs vs 20 normal controls (validation dataset) (Kulis et al., 2012)	0.26	8.66e-12	2.43e-10
59 u-CLLs vs 6 CD5+NBC/NBC (validation dataset) (Kulis et al., 2012)	0.38		FDR<0.05 (Kulis et al., 2012)
82 CLL/SLLs yr before diagnosis vs matched-normal controls (experimental predictive dataset)	0.047	0.00863	0.0921
438 MBCN cases yr before diagnosis vs matched-normal controls (experimental predictive dataset)	0.03	4.97e-07	7.47e-05

Table 12. *SHANK1* differential methylation data in the datasets analysed

Notes: adj: adjusted; CLL: chronic lymphocytic leukemia; NBC: naïve B cells; SLL: small lymphocytic lymphoma; yr: years; MBCN: mature B-cell neoplasms

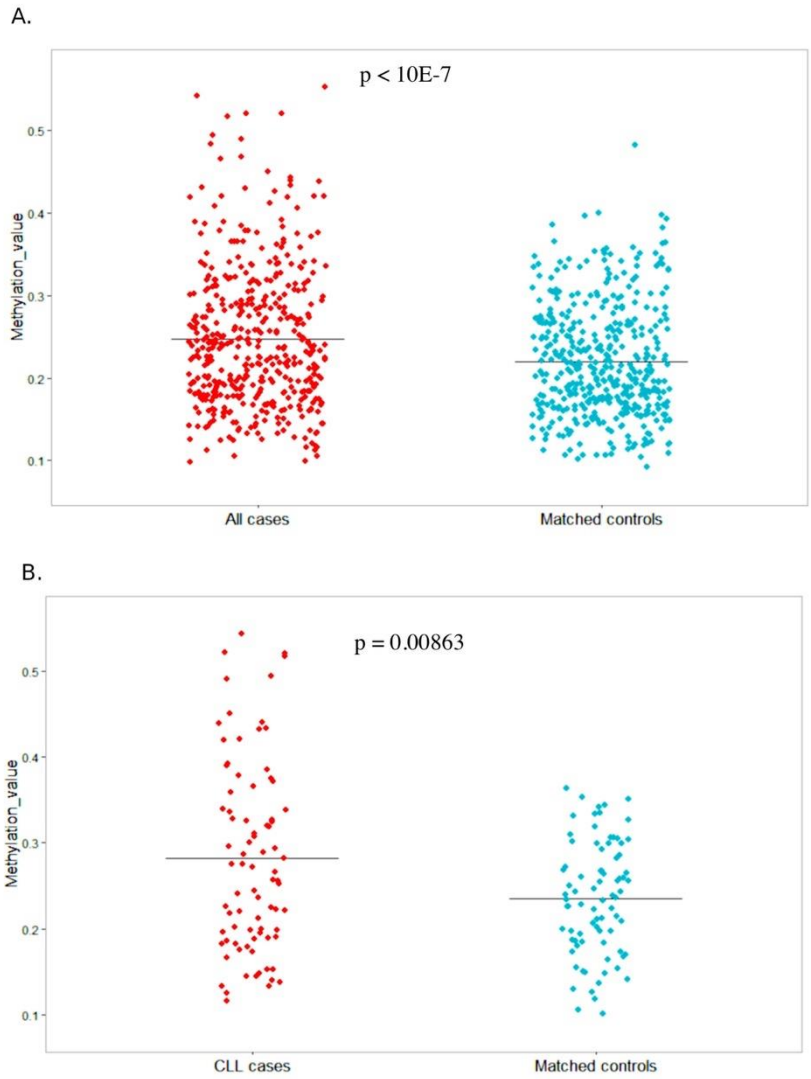


Figure 19. *SHANK1* methylation values in 438 MBCN cases/controls (A) and in the 82 CLL/SLL cases and controls (B)

4.1.3 Biliary tract cancer

4.1.3.1 Genome-wide methylation analysis

A genome-wide methylation analysis of 50 BTCs and 10 matched-normal samples was conducted by EPIC arrays. CpG methylation values were normalized using SWAN method (Maksimovic et al., 2012). An unusual density distribution of beta-values was observed, particularly evident after SWAN normalization indicating a non-optimal adjustment (Figure 20).

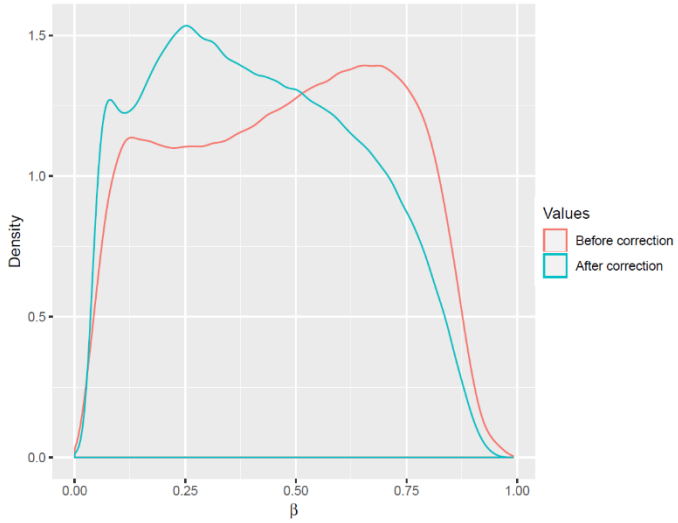


Figure 20. Density plot of the beta values distribution before and after correction (SWAN)

For this reason, methylation raw data were normalized using another normalization method (BMIQ). Both SWAN and BMIQ normalization algorithms performed a within-array normalization but, while BMIQ directly uses the type I probes as a reference in order to normalize beta values for type II probes to the distributions of type I probes belonging the same estimated state (methylated, hemimethylated and unmethylated), SWAN modifies both type I probe and type II probe values.

For samples with well-defined peaks in the unmethylated and methylated states, the tail ends of the BMIQ type2 probes distribution better matched those of the type1 distribution indicating an improved normalization using BMIQ than SWAN method (Figure 21).

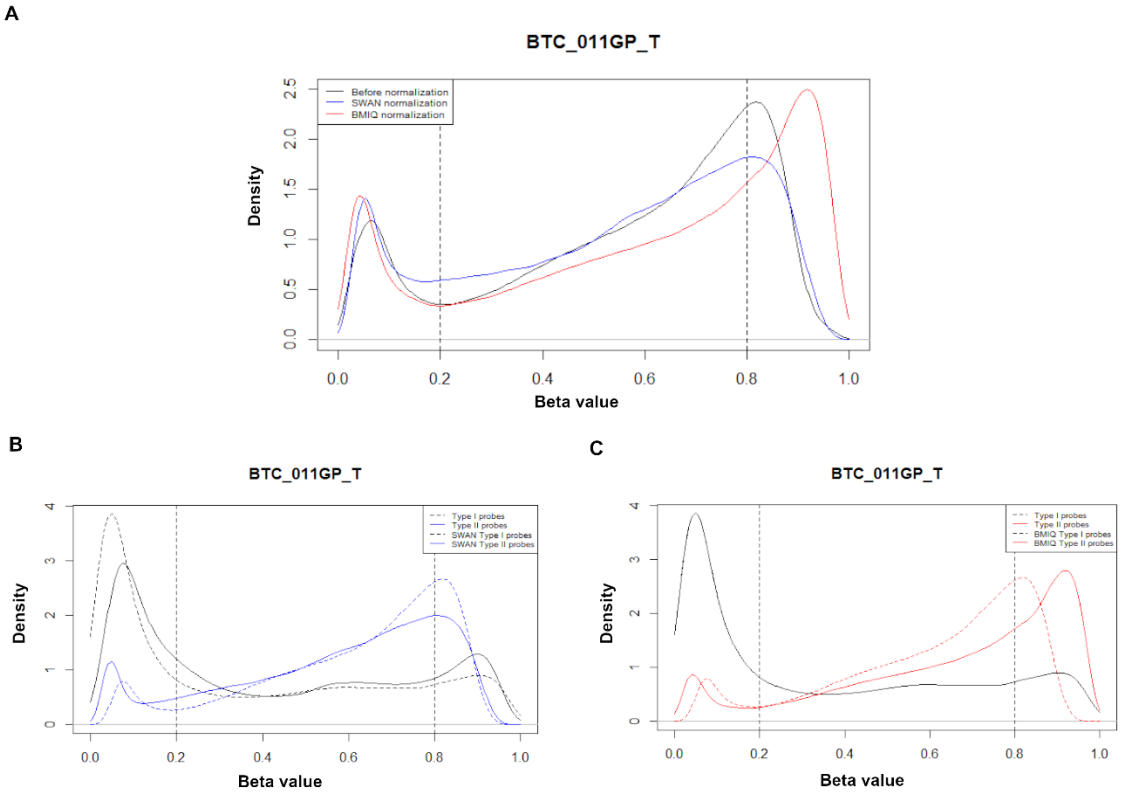


Figure 21. Density plot of beta value distribution before normalization and after SWAN and BMIQ normalization for one sample (A), Density plots of beta value distribution for Type I and Type II probes before and after SWAN (B) and BMIQ (C) correction.

As expected, for samples with an aberrant distribution of raw (not normalized) beta values, normalization methods do not work, and these samples should be removed from the analyses (Figure 22).

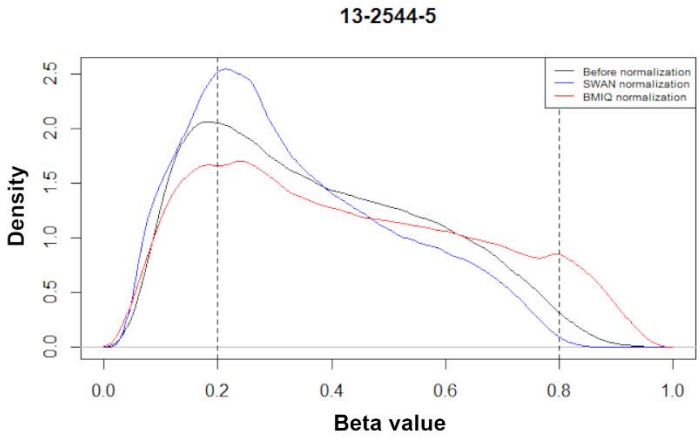


Figure 22. Density plot of beta value distribution before normalization and after SWAN and BMIQ normalization

Density plots of the average intensity values for each bead type revealed that, as expected, samples with abnormal beta values distribution showed also abnormal average intensity values distribution compared to samples with a normal beta value distribution (Figure 23).

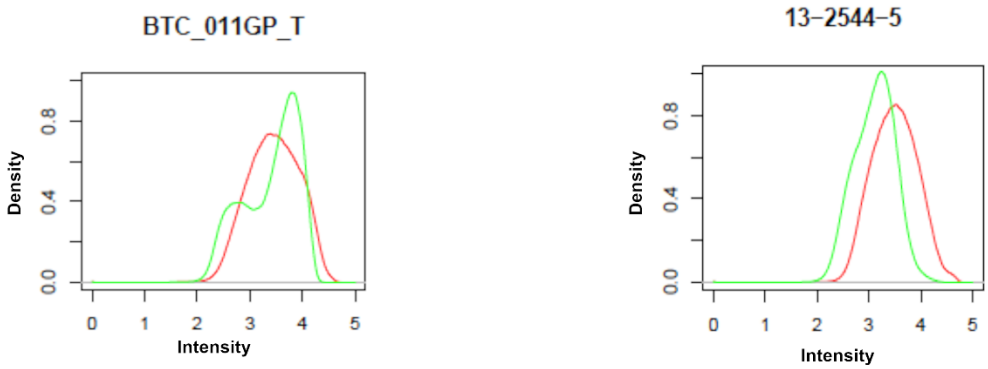


Figure 23. Example of probe intensity value distribution in two samples

Based on the two analysed parameters (beta values distribution before and after normalization and intensity values distribution), 26 good-quality samples were selected for the following analyses. This restricted dataset included 17 BTCs and nine normal control samples.

RnBeads pipeline analysis was repeated for this restricted dataset using BMIQ normalization method. The influence of this normalization of beta values distribution can be observed in Figure 24.

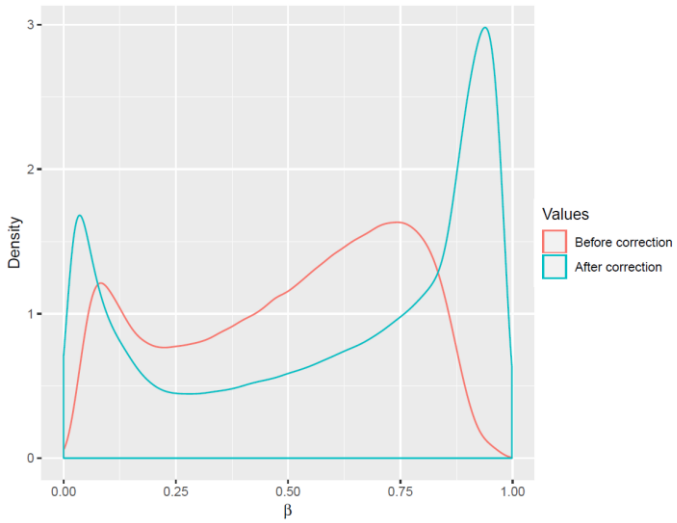


Figure 24. Beta values distribution before and after correction (BMIQ)

Differential methylation analysis revealed that, as expected, the number of differentially methylated CGIs identified analyzing this restricted dataset was higher than that identified using the total dataset including also low-quality samples (Figure 25).

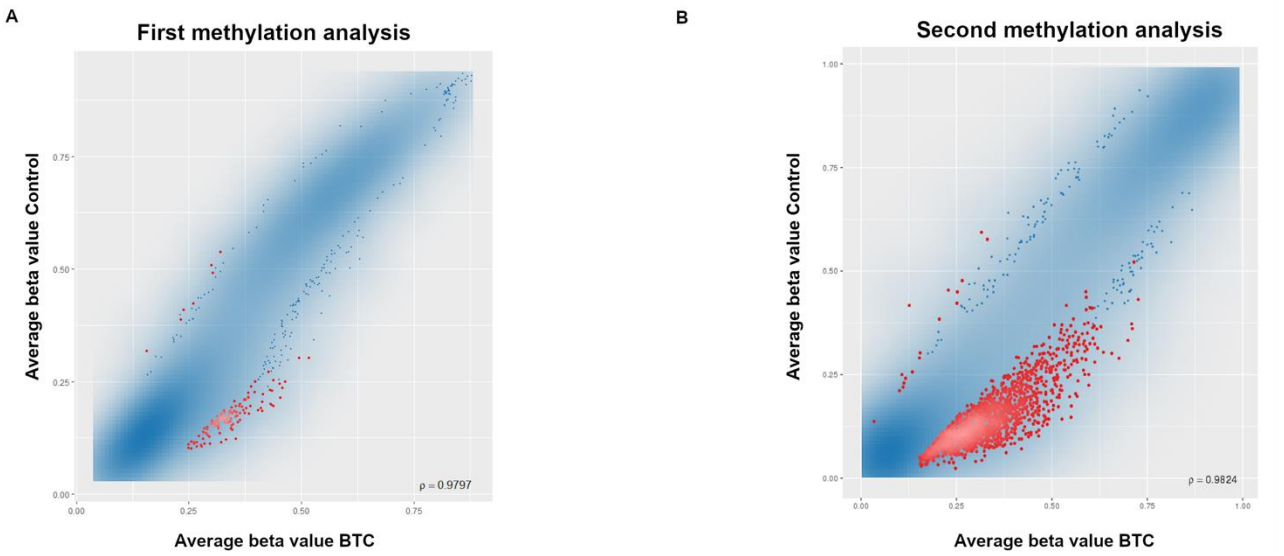


Figure 25. Average beta values distribution of BTC and normal control samples at CpG Island level in the first (A) and in the second (B) methylation analysis. Red dots represent statistically significant differentially methylated CpG islands (automatically selected rank cutoff)

648 CGIs were identified as differentially methylated ($|\Delta\beta| > 0.20$, combined p-value < 0.05) between BTC and normal control samples (Figure 26). After p-value correction, no statistically significant differentially methylated CGIs were detected, probably for the reduced sample size analysed.

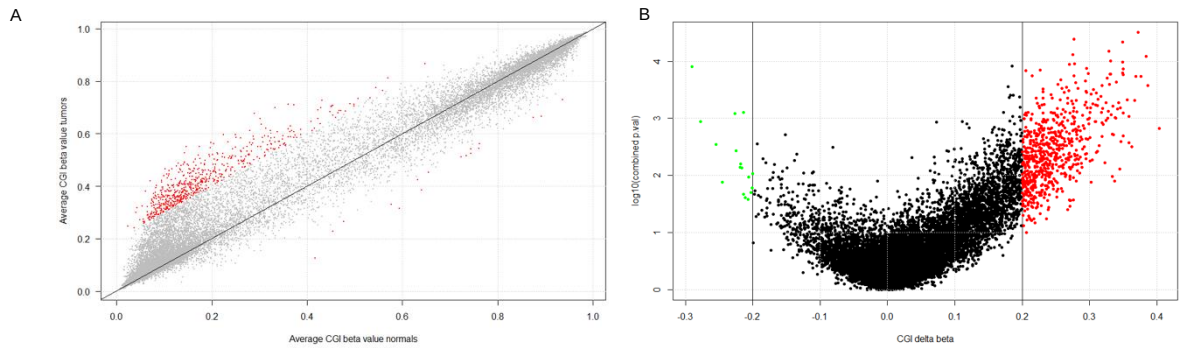


Figure 26. Average CGI beta values distribution in normal and tumour samples of our exploratory dataset. Red dots indicate statistically significant differentially methylated CGIs (A); Volcano plot distribution of CGIs delta betas of our exploratory dataset. Red dots indicate hypermethylated CGIs ($\Delta\beta > 0.2$) and green dots represent hypomethylated CGIs ($\Delta\beta < -0.2$) (B)

Heatmaps were generated using CGI methylation values (Figure 27A) and somatic change values (calculated as the difference between CGI methylation value of a sample and the average CGI methylation value of normal samples) of each sample (Figure 27B). Unsupervised hierarchical clustering analysis (UHC) showed two main sample clusters, including a first restricted cluster of the most hypermethylated tumour samples and a second cluster further divided into a cluster of tumours showing intermediate methylation values and a cluster including normal and three tumour samples with low methylation values (Figure 27). No association was observed between methylation values and tumour location or grade. Tumour samples of lower stages (I and II) were predominantly in the same cluster of normal samples compared to tumour samples of higher stages (III and IV) that were in the cluster including only tumours (p-value 3.11e-03, calculated using contingency table 2x2).

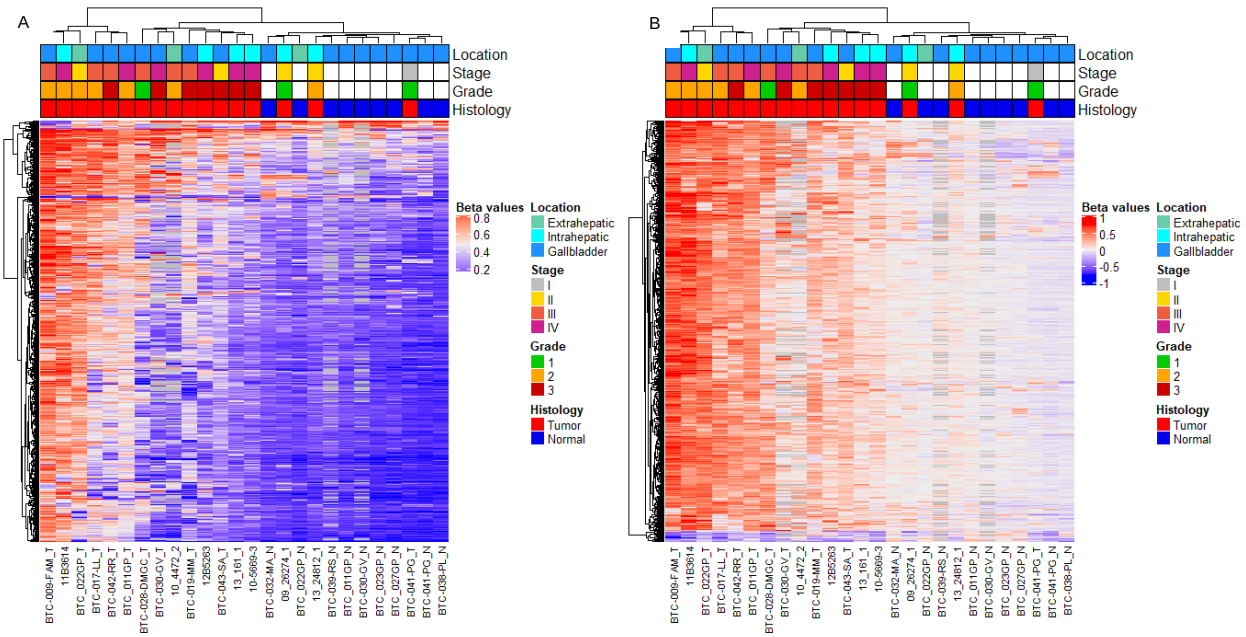


Figure 27. BTC discovery set unsupervised hierarchical clustering analysis based on the average CGI β values (A) or somatic changes (B) for each of the aberrantly methylated CGI.

CGIs showing β values higher than 0.20 in normal samples were filtered out obtaining 171 differentially methylated CGIs.

4.1.3.2 Methylation alterations validation

To validate and increase the robustness of the identified methylation alterations, methylation data from TCGA-CHOL dataset (including 36 tumour samples and nine normal samples) were analysed. Differential methylation analysis revealed 2203 differentially methylated CGIs (($|\Delta\beta| > 0.20$, adjusted combined p-value < 0.05) (Figure 28). This dataset showed a higher number of both hypermethylated and hypomethylated CGIs compared to our exploratory dataset.

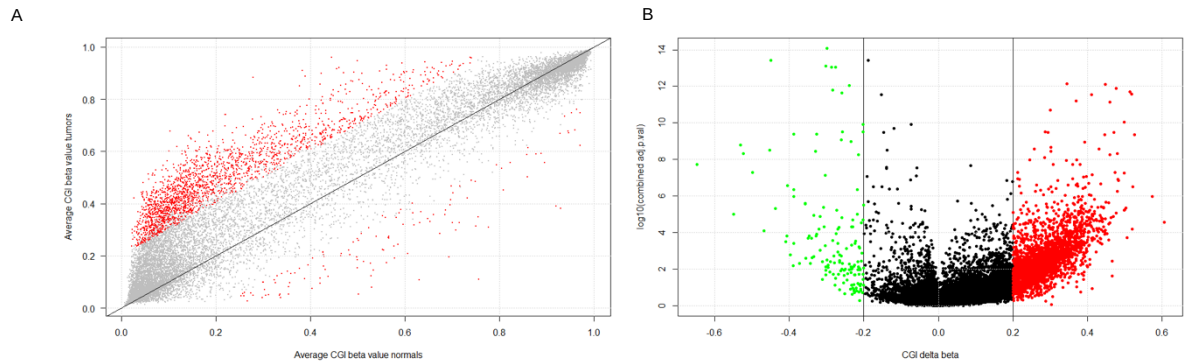


Figure 28. Average CGI beta values distribution in normal and tumour samples of TCGA-CHOL validation dataset. Red dots indicate statistically significant differentially methylated CGIs (A); Volcano plot distribution of CGIs delta betas of TCGA_CHOL validation dataset. Red dots indicate hypermethylated CGIs ($\Delta\beta > 0.2$) and green dots represent hypomethylated CGIs ($\Delta\beta < -0.2$) (B)

Heatmaps were generated using CGI methylation values (Figure 29A) and somatic change values (Figure 29B) of each sample. UHC showed a clusterization of the samples similar to that observed for in our discovery dataset, with only one tumour sample clusterizing in the sample subgroup. No association with clinical data was observed.

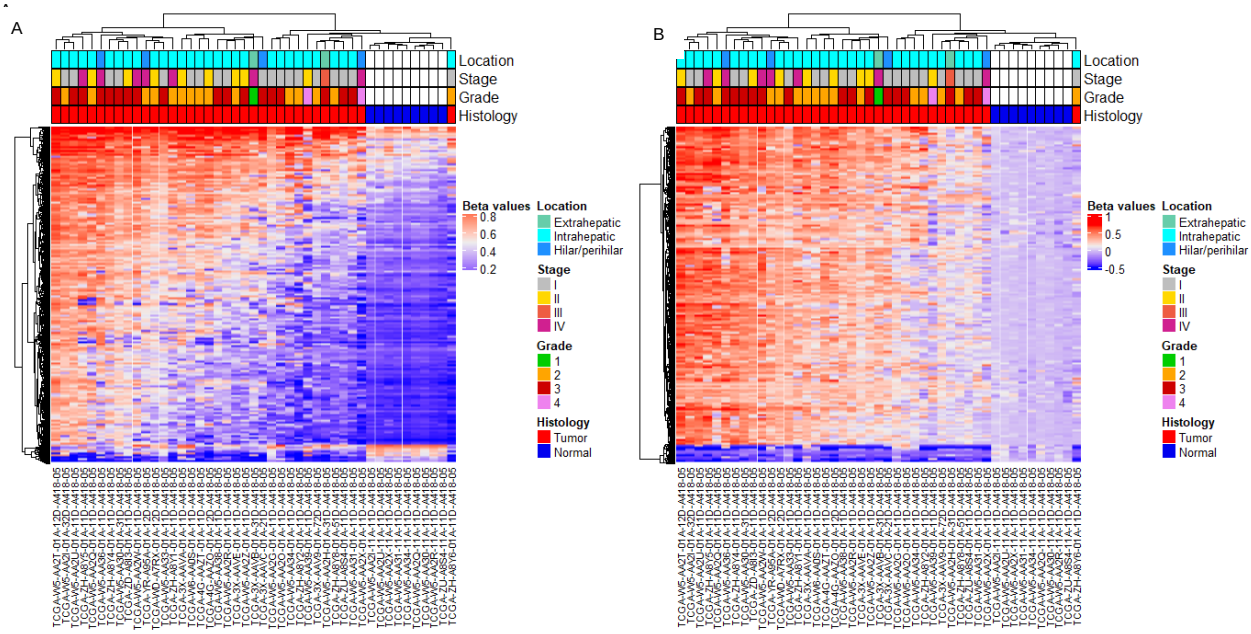


Figure 29. TCGA-CHOL validation set unsupervised hierarchical clustering analysis based on the average CGI β values (A) or somatic changes (B) for each of the aberrantly methylated CGI.

CGIs showing β values higher than 0.20 in normal samples were filtered out obtaining 998 differentially methylated CGIs.

Methylation alterations of 125 CGIs identified in our discovery dataset were validated in TCGA-CHOL dataset.

4.1.3.3 Tumours clustering

To identify tumour subgroups and explore possible correlation with clinical data, recursively partitioned mixture modeling (RPMM) was employed. In our discovery dataset, RPMM resulted in five clusters (Figure 30). Tumours of low grade and stage were mainly in the clusters of tumours with the lowest methylation values (rRLL), however no significant correlation was observed between tumour clusters and clinical data.

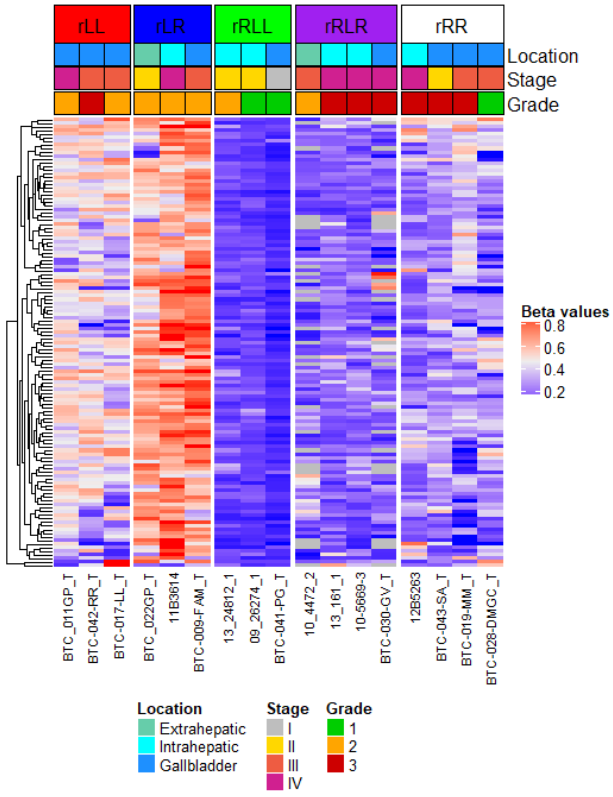


Figure 30. RPMM of validated CGIs in our discovery set

In TCGA-CHOL, RPMM resulted in seven classed of tumour samples (Figure 31). No correlation with clinical data was observed.

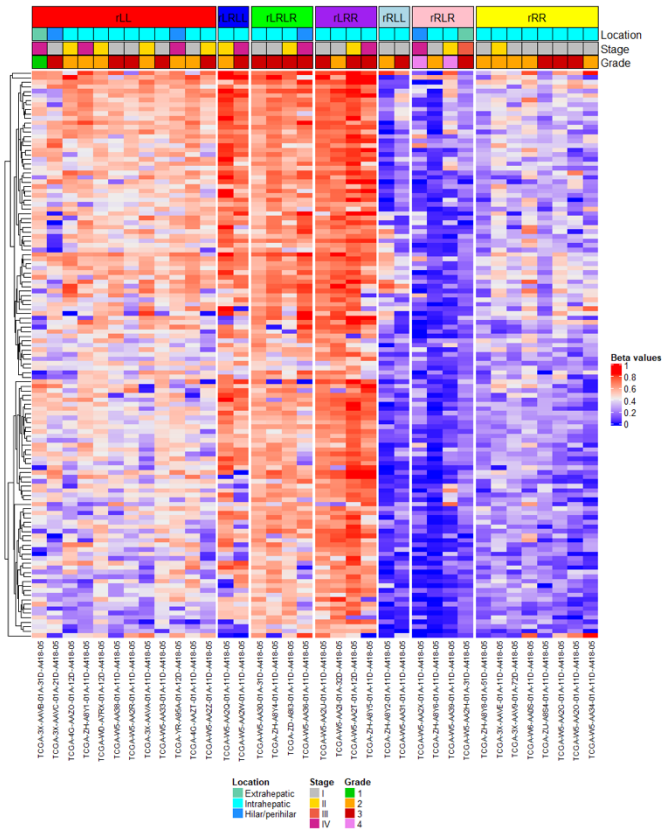


Figure 31. RPMM of validated CGIs in TCGA-CHOL discovery dataset

4.1.3.4 Selection of BTC-specific CGIs

To select only BTC-specific alterations, CGIs differentially methylated ($|\Delta\beta| > 0.20$) in TCGA-COAD and TCGA-STAD datasets were filtered out, leading to the identification of 30 CGIs specifically altered in BTC. Heatmaps generated with methylation values of these 30 CGIs for each sample of our dataset (Figure 32A) and TCGA-CHOL (Figure 32B) showed the same clusterization of the samples observed with the differentially methylated CGIs identified in our discovery dataset (Figure 27) and in TCGA-CHOL dataset (Figure 29).

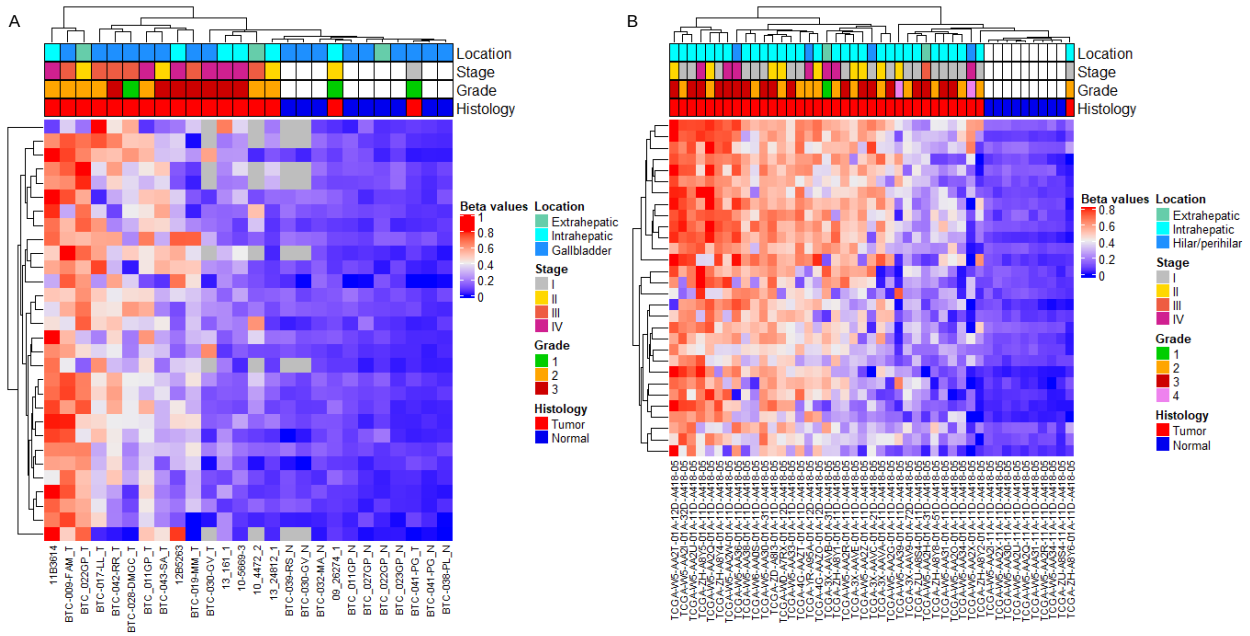


Figure 32. Discovery set (A) and TCGA-CHOL set (B) unsupervised hierarchical clustering analysis based on the average CGI β values for the 30 BTC-specific altered CGIs

Specificity and sensitivity of these 30 BTC-specific biomarkers were evaluated by generating ROC curves. In our exploratory dataset, 11 CGIs showed an AUC equal or higher than 0.90 (Figure 33), while 21 CGIs had an AUC > 0.90 in TCGA-CHOL dataset. Seven CGIs had an AUC higher than 0.90 in both datasets (Table 13).

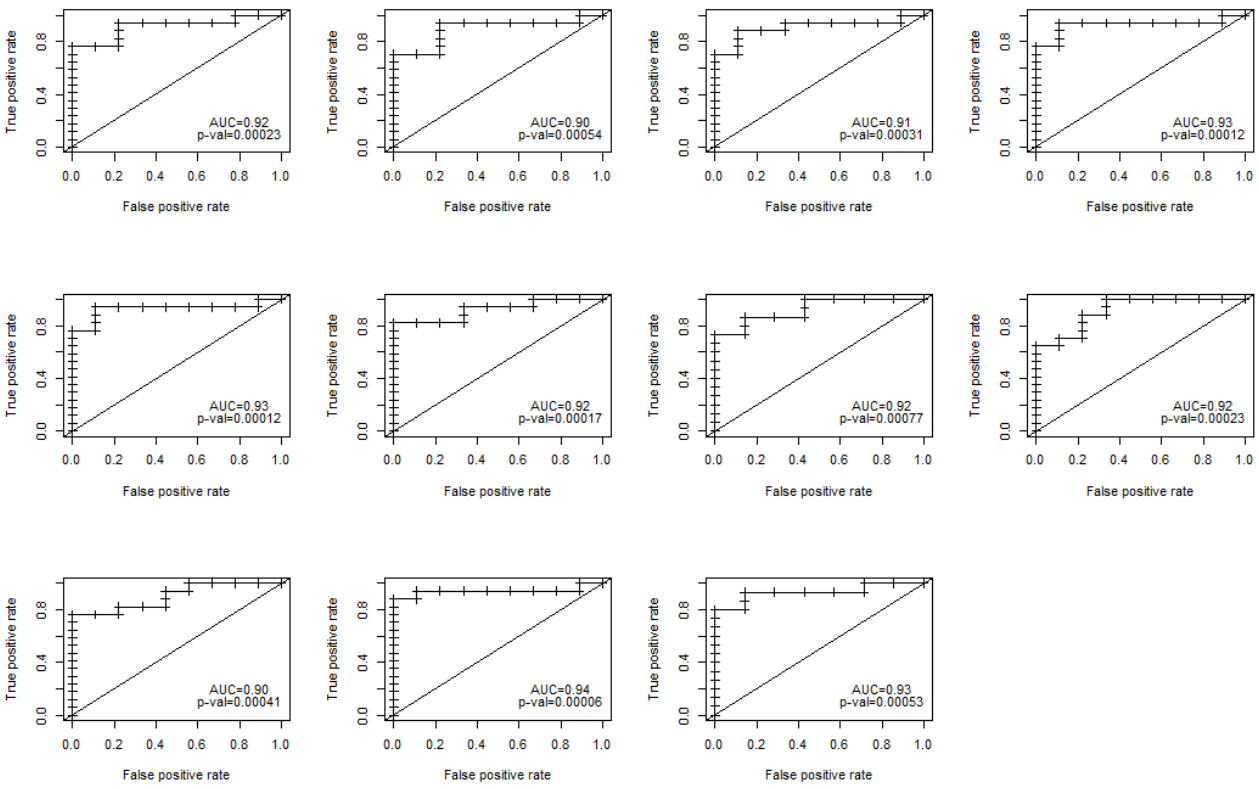


Figure 33. ROC curves for the 11 CGIs showing an AUC = or > 0.90 in our discovery dataset

CGI	Δβ Discovery set	AUC Discovery set	Δβ Validation set	AUC Validation set
CGI 1	0.22	0.92	0.31	0.93
CGI 2	0.21	0.90	0.41	1.00
CGI 3	0.25	0.93	0.38	0.94
CGI 4	0.21	0.93	0.24	0.97
CGI 5	0.27	0.92	0.41	0.95
CGI 6	0.22	0.92	0.29	0.93
CGI 7	0.25	0.90	0.38	1.00

Table 13. CGIs showing an AUC = or > 0.90 in our discovery set and in TCGA-CHOL validation set

4.1.3.5 BTC-specific altered CGIs in the excluded samples

To explore the possibility to rescue methylation data from the 33 excluded tumour samples, UHC was carried out using CGI methylation values for the 30 BTC-specific altered CGI, revealing a clusterization of these samples similar to that of good-quality selected tumour samples (Figure 34).

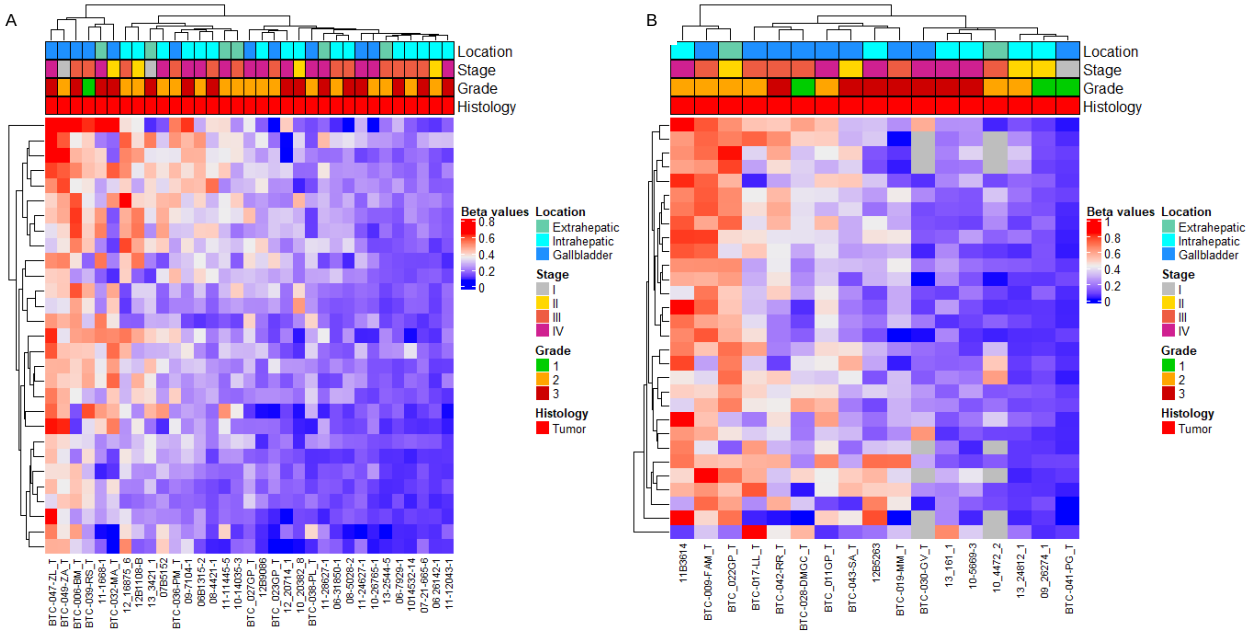


Figure 34. Excluded samples (A) and good-quality selected samples (B) unsupervised hierarchical clustering analysis based on the average CGI β values for the 30 BTC-specific altered CGIs

Principal component analysis (PCA) was performed in order to investigate whether some excluded tumour samples clusterized in a cluster along with good-quality samples and may be rescued and

integrated in the methylation analysis. However, PCA did not reveal a clear separation between excluded samples and good-quality samples making difficult to select possible samples to rescue (Figure 35).

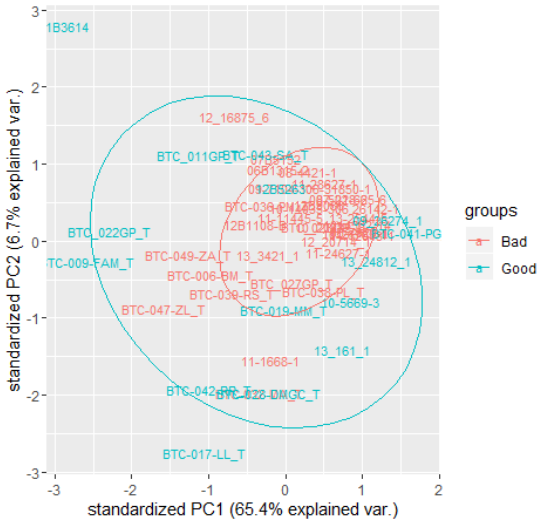


Figure 35. PCA of bad and good-quality samples

4.2 DNA methylation alterations are linked to gene expression changes

Gene expression analysis of selected genes whose associated CGIs were hypermethylated in tumors were performed in order to explore the association between methylation and gene expression.

4.2.1 Colorectal cancer

4.2.1.1 mRNA expression analysis

To verify whether hypermethylation of CGIs in the promoter regions of *GRIA4* and *VIPR2* was associated to a change in gene expression levels of these genes, a gene expression analysis on the 10 CRC and matched-normal samples analysed by MethyLight, was conducted by qRT-PCR. This sample cohort included the same samples analysed for DNA methylation status.

Gene expression results revealed a statistically significant downregulation of these genes in tumour samples compared to matched-normal controls (Figure 36A and D).

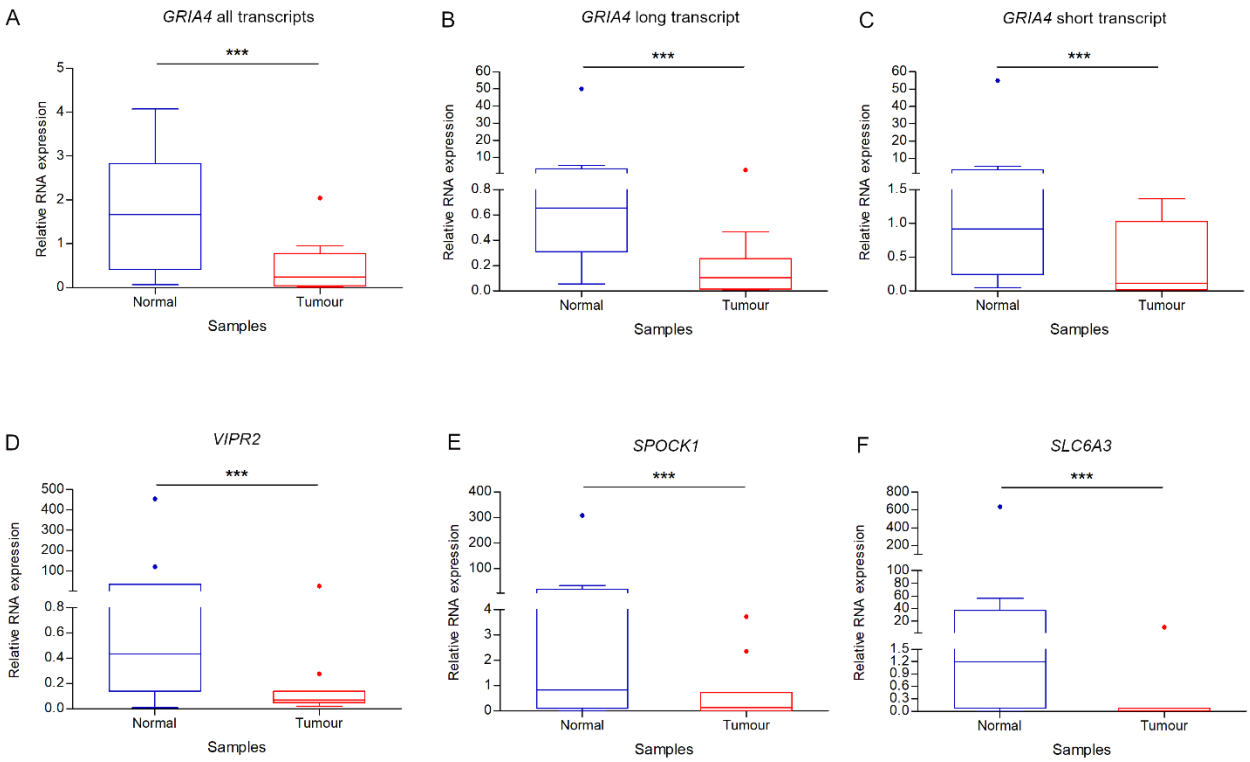


Figure 36. Differential gene expression analyses of *GRIA4* (A), *GRIA4* long canonical (B) and short (C) alternative transcripts, *VIPR2* (D), *SPOCK1* (E) and *SLC6A3* (F) genes between CRC and normal-matched control samples. Box plots show fold change values
*** indicates p-value < 0.0001

Since DNA methylation alterations can also affect the expression of alternative transcripts, gene expression analysis of the two main alternative transcripts of *GRIA4* gene, respectively encoding for the long canonical protein isoform and for a shorter isoform of Glutamate receptor 4 (GluR4) (Figure 37), was performed. A statistically significant reduction of both *GRIA4* alternative transcripts expression levels was observed in CRC samples compared to matched normal controls (Figure 36B and C).

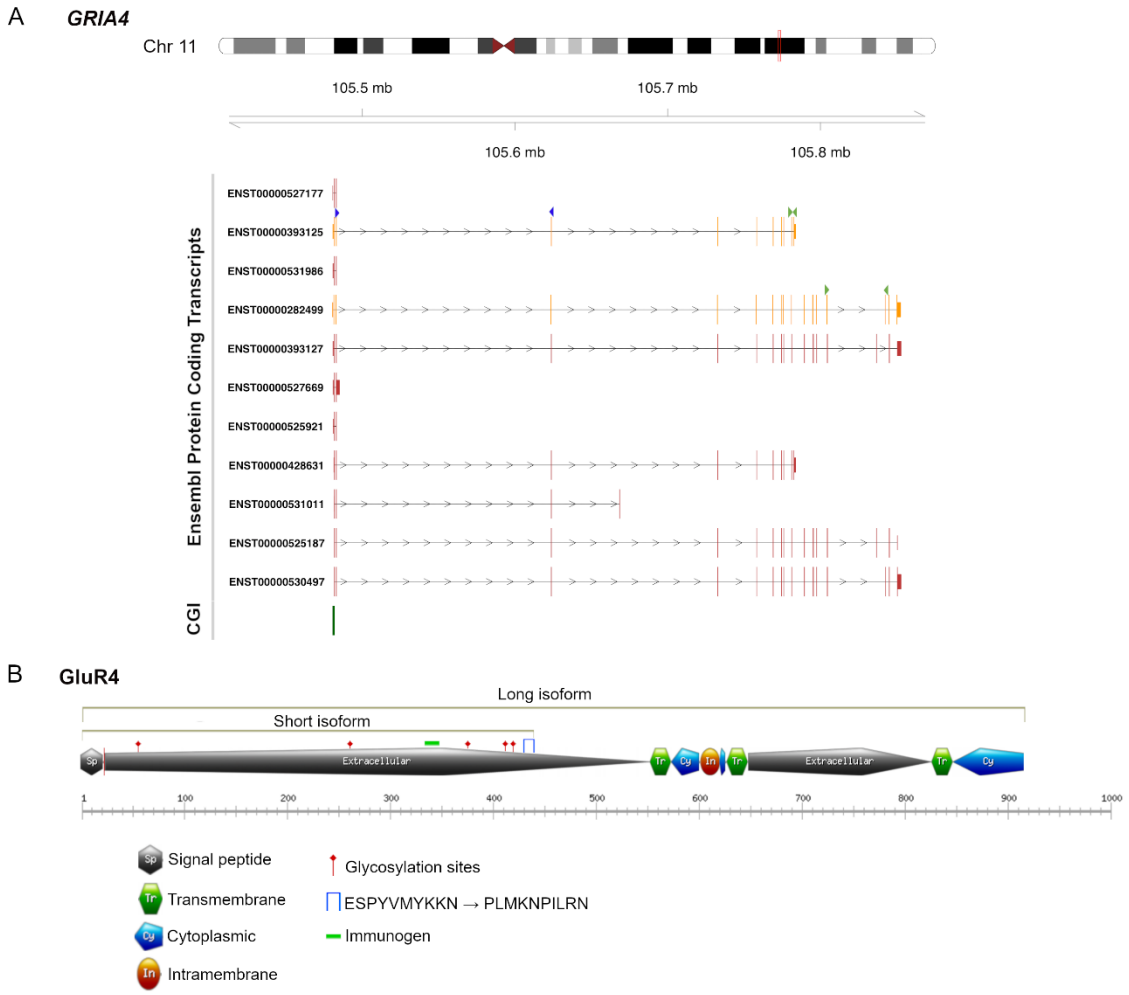


Figure 37. *GRIA4* alternative transcripts and primers position for the gene expression assays (long canonical and short alternative transcripts are shown in orange; blue arrows indicate primers for estimating gene expression levels of all transcripts, green arrows indicate primers for the long and short transcripts) (A); GluR4 protein isoforms and protein domains (green dash indicates the immunogen)

Moreover, gene expression analyses of *SPOCK1* and *SLC6A3* genes were performed in the same sample cohort. CGIs associated to these two genes have been found hypermethylated in CRC and were selected among the biomarkers showing an AUC > 0.95 both in our discovery dataset and in TCGA-COAD dataset. These genes resulted statistically significantly downregulated in tumour samples compared to matched-normal samples (Figure 36E and F).

4.2.1.1.1 Gene expression changes validation

To validate our finding, RNAseq data from TCGA-COAD dataset were explored. This analysis confirmed the statistically significant lower expression levels of *GRIA4* and *VIPR2* in CRC samples than in normal ones. In disagreement with our experimental data, *SLC6A3* resulted more expressed in CRC samples, while *SPOCK1* showed similar expression levels in both CRC and matched-normal samples (Figure 38). However, even if *SLC6A3* showed statistically significant higher levels in CRC samples of this dataset, tumour expression values largely overlap with normal expression values.

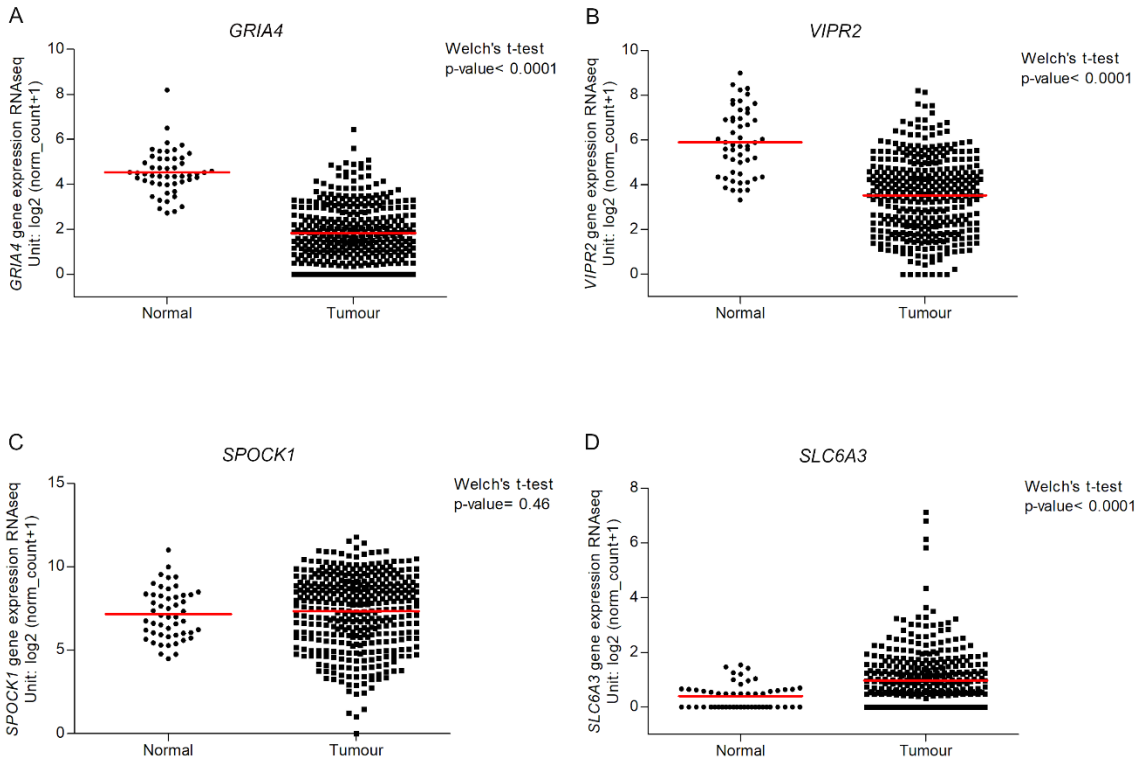


Figure 38. Differential gene expression analyses of *GRIA4* (A), VIPR2 (B), *SPOCK1* (C) and *SLC6A3* (D) genes between CRC and normal control samples from TCGA-COAD dataset. Box plots show log2(normalized count +1)

4.2.1.2 Protein expression analysis

To evaluate whether the identified gene expression alterations are translated in protein expression changes, western blot analyses of GluR4, VIPR2, SPOCK1 and SLC6A3 were performed in the 10 paired samples analysed by qRT-PCR.

The long GluR4 canonical isoform, SPOCK1 and SLC6A3 showed statistically significant lower protein levels in CRC samples, following the same pattern of the mRNA expression levels (Fig. 39). The short GluR4 and VIPR2 did not show statistically significant differences of expression between CRC and matched-normal samples, although a trend towards downregulation in CRC can be observed (Figure 39).

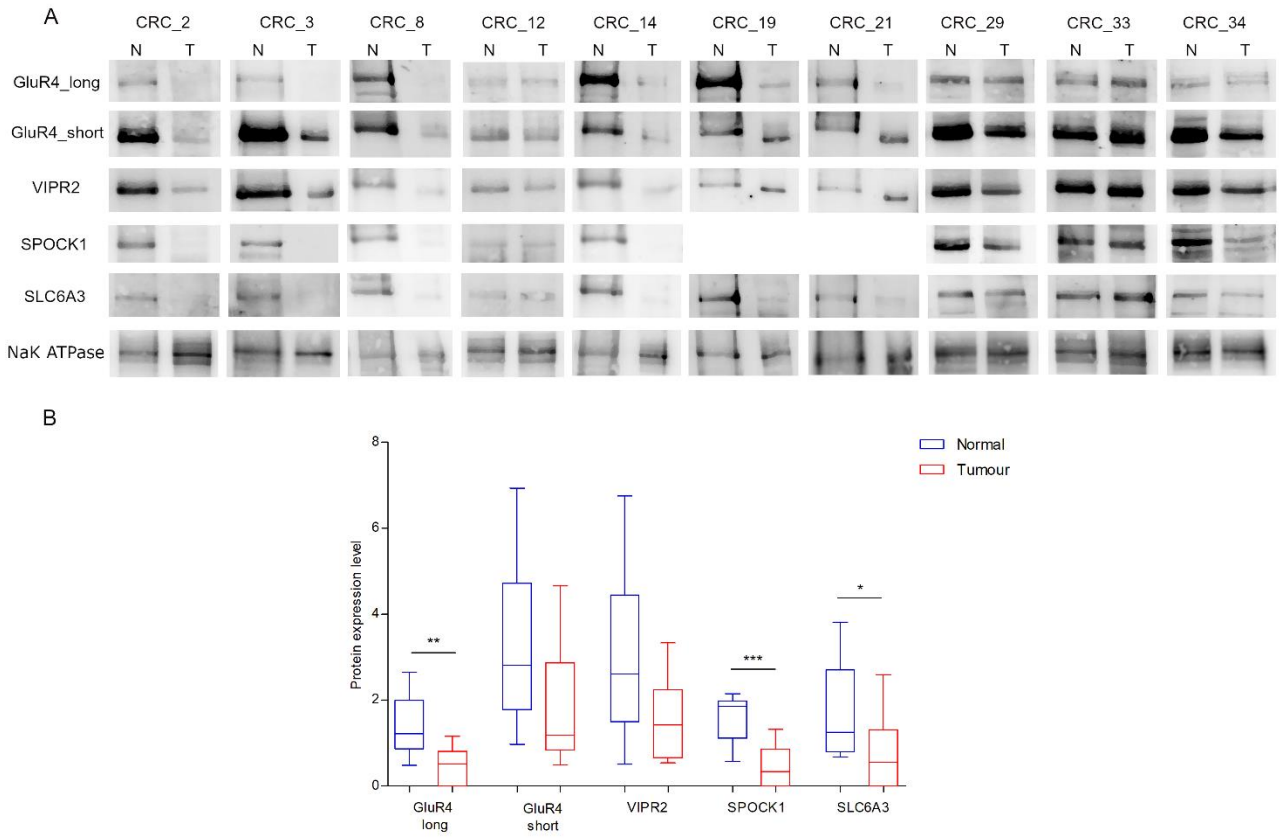


Figure 39. GluR4 long, GluR4 short, VIPR2, SPOCK1, SLC6A3 and NaK ATPase immunoblots of CRC and their matched-normal samples (A); Protein expression levels in tumour and matched-normal controls. Box plots show normalized protein levels
* indicates p-value < 0.05 ** indicates p-value < 0.001, *** indicates p-value < 0.0001

A comparison between transcript expression levels and their respective protein isoform expression levels revealed that *GRIA4* long transcript was statistically significantly more expressed than the short one in both tumour and normal samples (Figure 40). This result was not reflected at protein level where GluR4 long isoform was statistically significantly less expressed compared to the short isoform in both tumour and normal samples (Figure 40).

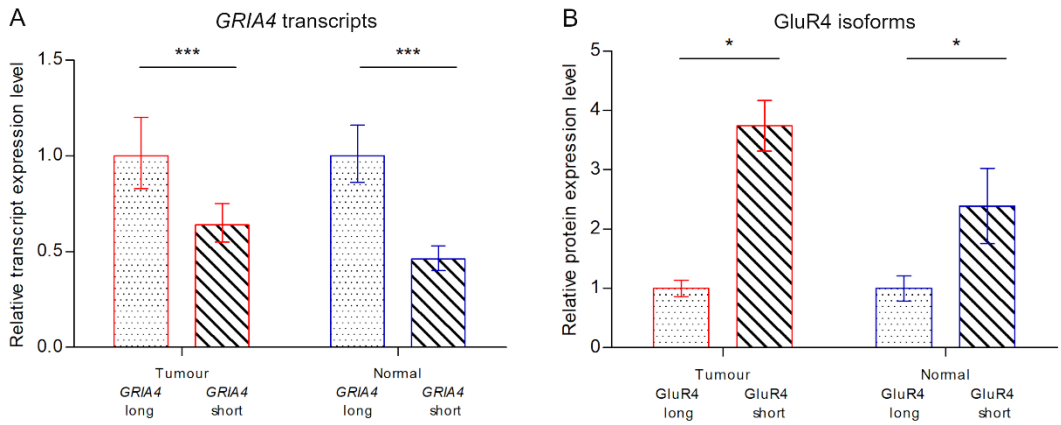


Figure 40. *GRIA4* transcripts levels in tumour and normal samples (A), GluR4 isoforms levels in tumour and normal samples (B). *GRIA4* short transcript expression level are shown relative to *GRIA4* long transcript expression level (A); GluR4 short isoform expression level are shown relative to GluR4 long isoform expression level (B)
* indicates p-value < 0.05, *** indicates p-value < 0.0001

4.2.2 Chronic lymphocytic leukemia

4.2.2.1 *SHANK1* gene expression analysis

To investigate the impact of *SHANK1*-associated CGI hypermethylation on gene expression, a gene expression analysis was conducted on 27 CLL samples and 16 normal controls by qRT-PCR. Hypermethylation of the CGI located on the gene body of *SHANK1* gene (Figure 41A) was associated to a statistically significant down-regulation of *SHANK1* gene in CLL samples (Figure 41B).

To exclude that the detected *SHANK1* different expression levels could be related to a different expression of the reference gene *ACTB* in CLL whole blood samples, *SHANK1* downregulation was also confirmed using *GUSB* as reference gene in a subgroup of samples (data not shown).

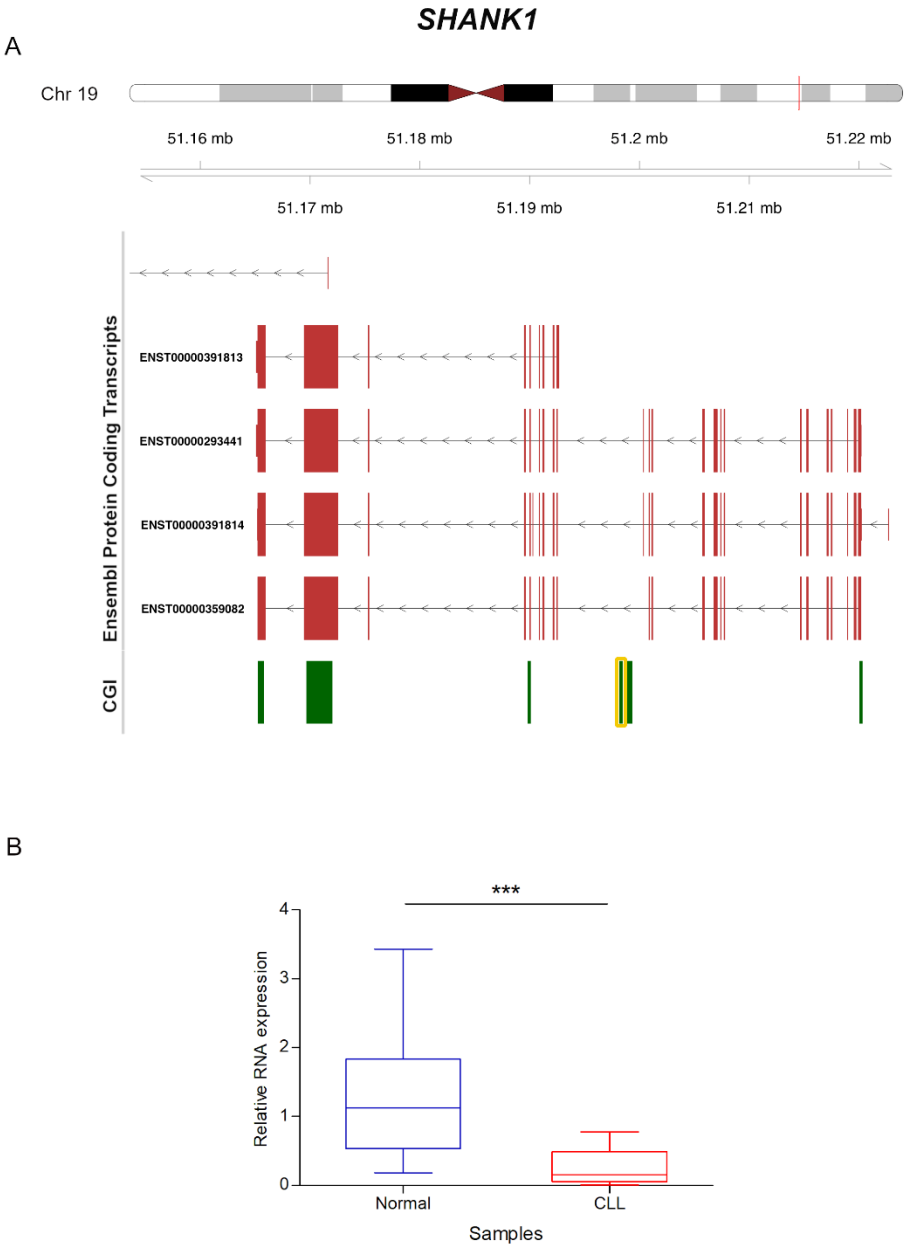


Figure 41. *SHANK1* protein coding transcripts and associated CGIs (the identified hypermethylated CGI is enclosed in an orange box) (A); differential gene expression analysis of *SHANK1* gene between CLL and normal control samples. Box plot shows fold change values
*** indicates p-value < 0.0001

4.2.2.1.1 Gene expression validation

To validate our result in an independent cohort, RNA-seq data from a dataset (GEO accession number: GSE70830) including 10 CLL samples and five normal peripheral blood CD19+ B cells were analysed. In this dataset, *SHANK1* gene did not show a statistically significant reduced expression in CLL samples compared to control ones. Gene expression data of a large dataset (Ferreira et al., 2014) including 98 CLL and three subtypes of normal B cells (naïve, memory IgM/IgD, and memory IgG/IgA), were unsuitable for estimating *SHANK1* gene expression levels since the majority of the samples have zero reads.

This was not surprising since *SHANK1* resulted a gene weakly expressed in whole-blood in a whole-transcript analysis (GTEx Portal data: median Transcripts Per Kilobase Million (TPM): 0.030, number of samples: 407).

4.3 *PCDH* cluster region is aberrantly methylated region in solid cancers

Methylation of CGIs associated to *PCDH* cluster was evaluated in different exploratory datasets including both solid cancers (PA, CRC, GC and BTC) and a blood cancer (CLL). *PCDH*-associated CGIs alterations were among the most significant alterations identified in all the solid tumours analysed. In CRC, the most altered CGI associated with *PCDH* gamma cluster (*PCDHG@*) was the fourth most hypermethylated CGI and the fifth most significantly altered CGI among a panel of 74 CGI alterations previously identified in CRC (Fadda et al., 2018). In the other two gastrointestinal tumours analysed, the most altered CGIs associated with *PCDHG@* were among the statistically significant hypermethylated CGIs: 122/522 and 40/48, considering the Δβ, and 28/522 and 13/48, considering the p-value, in GC and BTC respectively. On the other hand, no CGI resulted hypermethylated in PA, but a *PCDH*-associated region resulted one of the most hypomethylated regions (12/208) identified in PA.

4.3.1 *PCDH*-associated CGI alterations in colorectal cancer

Results from a previous study conducted in our laboratory have shown that *PCDH*-associated CGIs were some of the most prominent alterations shared between CRC and adenomas (Fadda et al., 2018). For this reason, an analysis of the methylation status of the *PCDH*-associated CGIs was conducted using methylation data of the same sample cohort analysed in our previous study. Four CGIs, associated with *PCDHG@*, whose three mapped one promoter regions (Table 14, Figure 42A), showed a statistically significant hypermethylation in CRC compared to normal controls (Table 14, Figure 42B). Three of these CGIs were also hypermethylated in adenomas but this hypermethylation was more pronounced in CRC, except for one CGI that resulted hypermethylated only in adenomas (chr5:140750050-140750264, CpG 16).

Since DNA methylation and CCTC binding factor (CTCF) play a fundamental role in *PCDH* promoter choice and transcription regulation, the association between the identified altered CGIs and CTCF binding sites was investigated. Three out of the four altered CGIs resulted associated with CTCF binding sites (Table 14, Fig. 38A).

UCSC CGI	UCSC CGI name	CTCF binding site	Δβ value (CRC-Control)	Δβ value (CRA-Control)	Genes within region	Gene promoter-associated CGI
chr5:140750050-140750264	CpG 16	chr5:140748521-140750945	0.157*	0.200	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*	*PCDHGB3*
chr5:140855386-140856620	CpG 95	chr5:140854218-140856648	0.200*	0.105	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA12*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*, *PCDHGC3*	*PCDHGC3*
chr5:140857864-140858065	CpG 19	chr5:140856882-140859319	0.310	0.259	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA12*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*, *PCDHGC3*	---
chr5:140864527-140864748	CpG 22	---	0.435	0.277	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA12*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*, *PCDHGC3*, *PCDHGC4*	*PCDHGC4*
chr5:140892914-140893189	CpG 20	chr5:140890901-140893291 chr5:140891594-140893806	0.302	0.200	---	---

Table 14. Altered CGIs in colorectal cancer (CRC) and colorectal adenoma (CRA)
* nominal threshold (p-value<0.05).

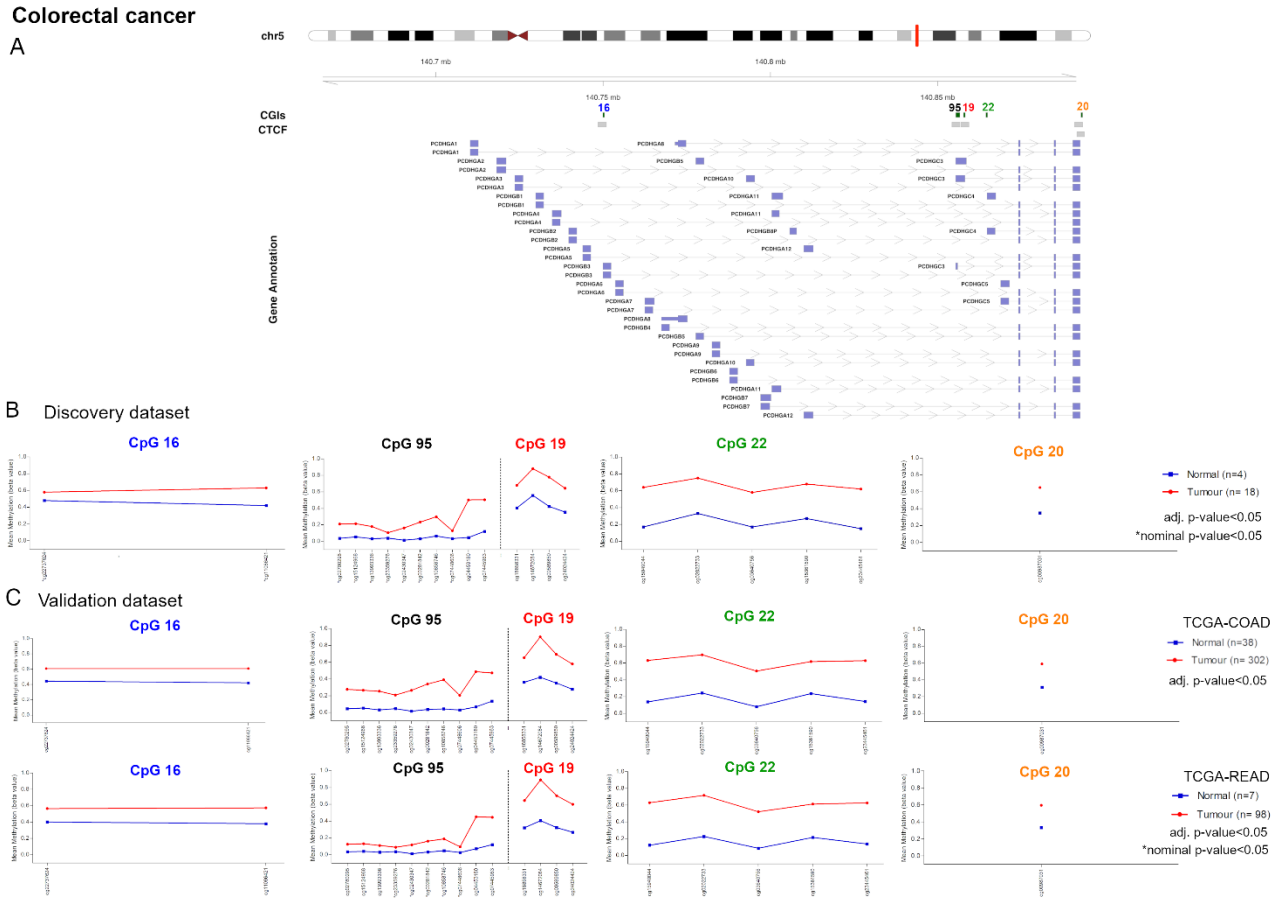


Figure 42. Genomic organization of *PCDHG@*, including the localization of exons, CGIs (annotated with the UCSC CGI names) and CTCF binding sites (A); β values, resulting from the average of our discovery samples set of each probe mapping to the altered CGIs (B); β values, resulting from the average of TCGA-COAD and TCGA-READ validation sample sets of each probe mapping to the altered CGIs (C)

CGI methylation values of the altered CGIs were visualized in a heatmap and UHC showed two main clusters: one including all, expect one (293T), CRC samples and 12 adenomas, and the other including normal samples, nine adenomas and one CRC sample. No association was observed between methylation values and clinical data (Figure 43).

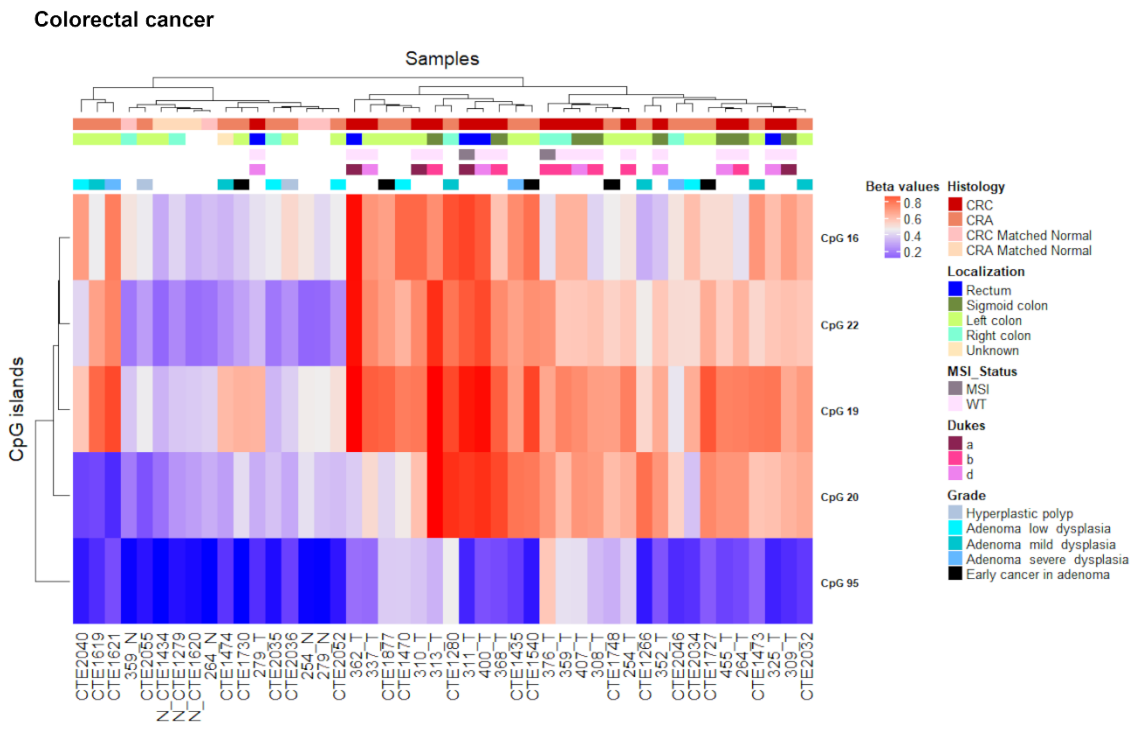


Figure 43. Colorectal cancer discovery set unsupervised hierarchical clustering analysis based on the average CGI β values for the five altered *PCDH*-associated CGIs

All the identified CGI alterations were successfully validated using methylation data from TCGA-COAD and TCGA-READ datasets (Figure 42C). One CGI (chr5:140855386-140856620, CpG 95) resulted hypermethylated only in COAD samples, with the exception of the most telomeric part of the CGI (cg04453180, cg07445963) showing hypermethylation also in READ samples (Figure 42C).

4.3.2 *PCDH*-associated CGI alterations in chronic lymphocytic leukemia

To investigate whether *PCDH*-associated CGI alterations may also be present in blood cancer, methylation data of our CLL discovery set and CLLE-ES validation set were explored.

No significant methylation alterations in CGI associated with *PCDH* clusters were detected (Table 15)

UCSC CGI	UCSC CGI name	CTCF binding site	Discovery set Δβ value (CLL-Control)	Validation set Δβ value (CLL-Control)	Genes within region
chr5:140174573-140174888	CpG 28	chr5:140173111-140175577 chr5:140173795-140176243	0.158	0.189	*PCDHA1*, *PCDHA2*
chr5:140762401-140762768	CpG 28	chr5:140761029-140763470	0.163	0.135	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*
chr5:140180844-140181082	CpG 21	chr5:140179374-140181807 chr5:140180022-140182456	0.165	0.182	*PCDHA1*, *PCDHA2*, *PCDHA3*
chr5:140186792-140187268	CpG 40	chr5:140185348-140187798 chr5:140186021-140188437	0.160	0.135	*PCDHA1*, *PCDHA2*, *PCDHA3*, *PCDHA4*

Table 15. *PCDH*-associated CGIs in CLL discovery and validation datasets

4.3.3 *PCDH*-associated CGI alterations in biliary tract cancer

To explore whether *PCDH*-associated CGI alterations are frequent events in other gastrointestinal tumours, data from our first methylome analysis of 50 BTCs and 10 matched-normal controls were analysed. This study did not reveal any CGI differentially methylated between BTC and matched-normal samples according to our selection criteria. However, two CGIs (chr5:140787447-140788044, CpG 45 and chr5:140797162-140797701, CpG 41), mapping on promoter regions and associated with CTCF binding sites (Table 16, Figure 44A), showed statistically significant differential methylation levels between BTC and matched-normal control samples (Table 16, Figure 44B). Moreover, a statistically significant difference among BTC tumours from different localization was observed (Table 16, Figure 44B). In fact, extrahepatic/gallbladder tumours showed higher differential methylation values relative to matched-normal controls than intrahepatic tumours.

UCSC CGI	UCSC CGI name	CTCF binding site	Δβ value (BTC-Control)	Δβ value (Intrahepatic BTC-Control)	Δβ value (Extrahepatic BTC-Control)	Δβ value (Gallbladder BTC-Control)	Δβ value (Extrahepatic /Gallbladder BTC-Control)	Genes within region	Gene promoter-associated CGI
chr5:140787447-140788044	CpG 45	chr5:140786247-140788714	0.175	0.104	0.212	0.235	0.229	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*	*PCDHGB6*
chr5:140797162-140797701	CpG 41	chr5:140795962-140798360	0.130	0.108	0.200	0.215	0.209	*PCDHGA1*, *PCDHGA10*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*	*PCDHGB7*

Table 16. Altered CGIs in biliary tract cancer (BTC)
Note: CGI: CpG island; BTC: biliary tract cancer

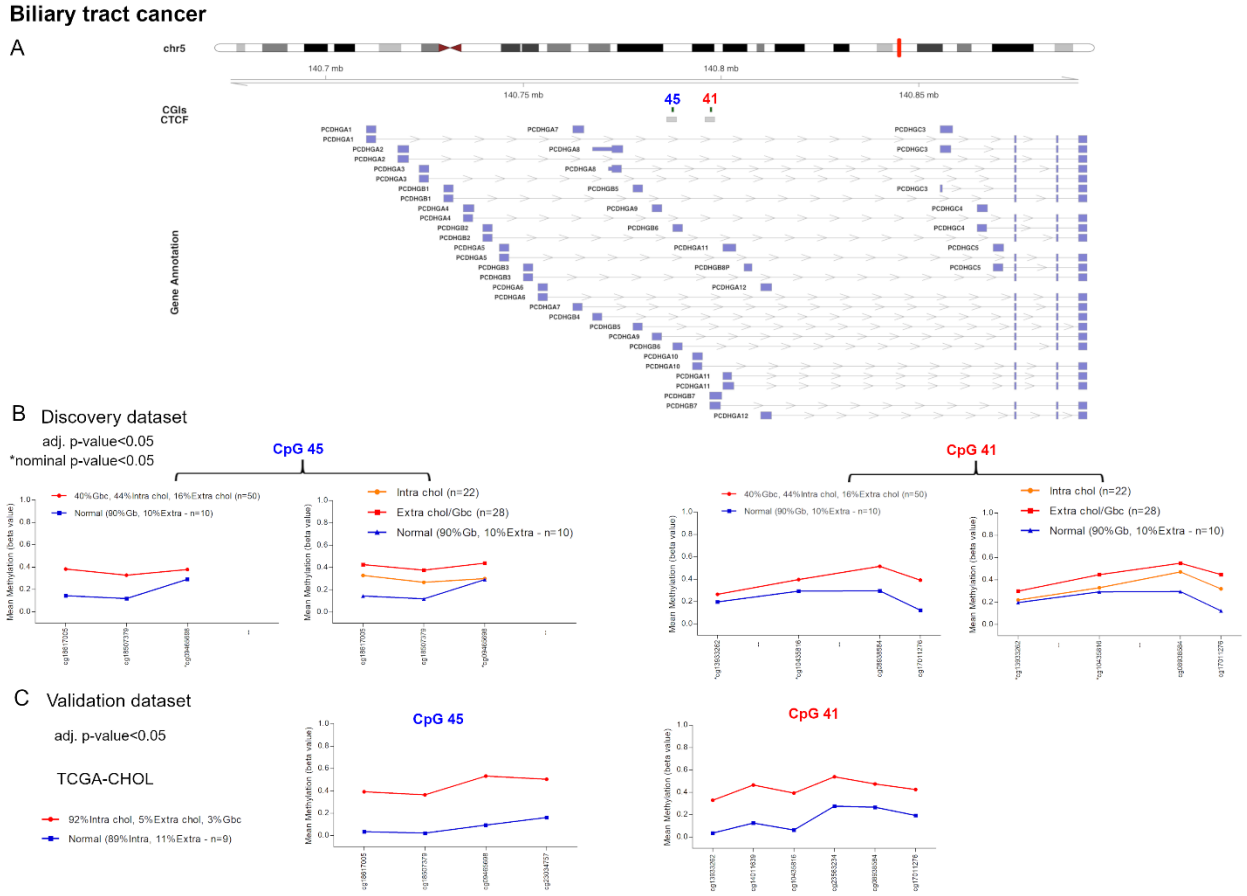


Figure 44. Genomic organization of *PCDHG@*, including the localization of exons, CGIs (annotated with the UCSC CGI names) and CTCF binding sites (A); β values, resulting from the average of our discovery samples set of each probe mapping to the altered CGIs (B); β values, resulting from the average of TCGA-CHOL validation sample sets of each probe mapping to the altered CGIs (C)

UHC reflected this difference among tumour localizations since most of gallbladder (85%) and extrahepatic (87.5%) tumours were in a distinct cluster respect to normal samples, while intrahepatic tumours were distributed almost equally between the two main clusters (Figure 45).

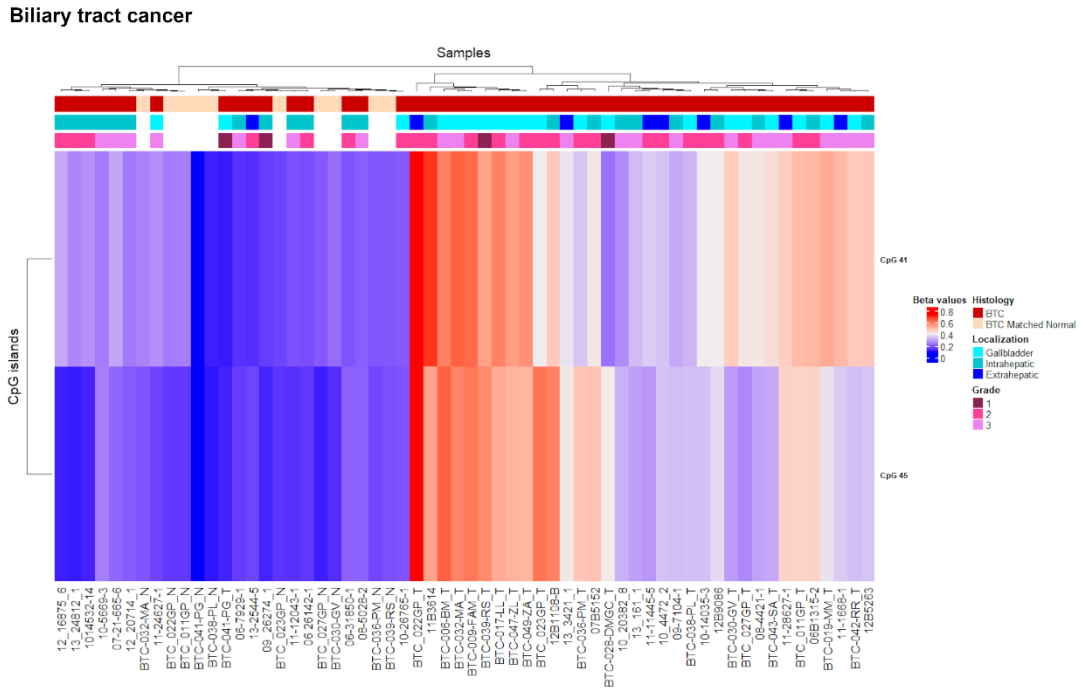


Figure 45. Biliary tract cancer discovery set unsupervised hierarchical clustering analysis based on the average CGI β values for the two altered *PCDH*-associated CGIs

Methylation alterations of these two *PCDH*-associated CGIs were confirmed in TCGA-CHOL validation dataset where differential methylation levels were even more pronounced (Figure 46C). However, this dataset included mostly intrahepatic tumours and also the 10 normal samples were mostly collected form intrahepatic tissues (8/10) showing average methylation values of 0.078 (CpG 45), while normal samples from our discovery set included nine gallbladders and one extrahepatic tissue.

Data obtained from the second methylome analysis of the restricted data were explored in order to verify a possible detection of additional altered *PCDH*-associated CGIs. This analysis indeed allowed the identification of additional 11 *PCDH*-associated CGIs hypermethylated in our BTC discovery set and validated in TCGA-CHOL dataset (Table 17). Moreover, the two CGIs previously identified (chr5:140787447-140788044, CpG 45 and chr5:140797162-140797701, CpG 41) showed higher Δβ values in this restricted dataset. TCGA-CHOL results showed higher Δβ values also for these CGIs (Table 17).

UCSC CGI	UCSC CGI name	CTCF binding site	Discovery set Δβ value (BTC-Control)	Validation set Δβ value (BTC-Control)	Genes within region
chr5:140214068-140214464	CpG 35	chr5:140212561-140214991	0.211	0.310	*PCDHA1*, *PCDHA2*, *PCDHA3*, *PCDHA4*, *PCDHA5*, *PCDHA6*, *PCDHA7*
chr5:140221007-140221381	CpG 31	chr5:140219505-140221977,	0.208	0.256	*PCDHA1*, *PCDHA2*, *PCDHA3*, *PCDHA4*,

		chr5:140220141-140222571			*PCDHA5*, *PCDHA6*, *PCDHA7*, *PCDHA8*
chr5:140235737-140236033	CpG 23	chr5:140234229-140236690, chr5:140234845-140237272	0.252	0.258	*PCDHA1*, *PCDHA10*, *PCDHA2*, *PCDHA3*, *PCDHA4*, *PCDHA5*, *PCDHA6*, *PCDHA7*, *PCDHA8*, *PCDHA9*
chr5:140620864-140621698	CpG 76	-	0.244	0.323	*PCDHB19P*
chr5:140711797-140712542	CpG 71	-	0.229	0.283	*PCDHGA1*
chr5:140734720-140735028	CpG 27	chr5:140733354-140735959	0.203	0.318	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGB1*
chr5:140741174-140741738	CpG 65	-	0.254	0.384	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGB1*, *PCDHGB2*
chr5:140750050-140750264	CpG 16	chr5:140748521-140750945	0.268	0.3229	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*
chr5:140767196-140767695	CpG 35	chr5:140765966-140768414	0.211	0.268	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*
chr5:140777442-140777938	CpG 36	chr5:140776215-140778639, chr5:140777214-140779496	0.253	0.315	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*
chr5:140787447-140788044	CpG 45	chr5:140786247-140788714	0.277	0.383	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*
chr5:140797162-140797701	CpG 41	chr5:140795962-140798360	0.219	0.300	*PCDHGA1*, *PCDHGA10*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*
chr5:140800760-140801072	CpG 29	chr5:140799360-140801740	0.261	0.323	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*,

					PCDHGB5, PCDHGB6, PCDHGB7

Table 17. Altered CGIs in the restricted BTC dataset and TCGA-CHOL dataset

UHC analysis using methylation data for the altered *PCDH*-associated CGIs showed a clusterization similar to that obtained with the BTC-specific altered CGIs (Figure 46).

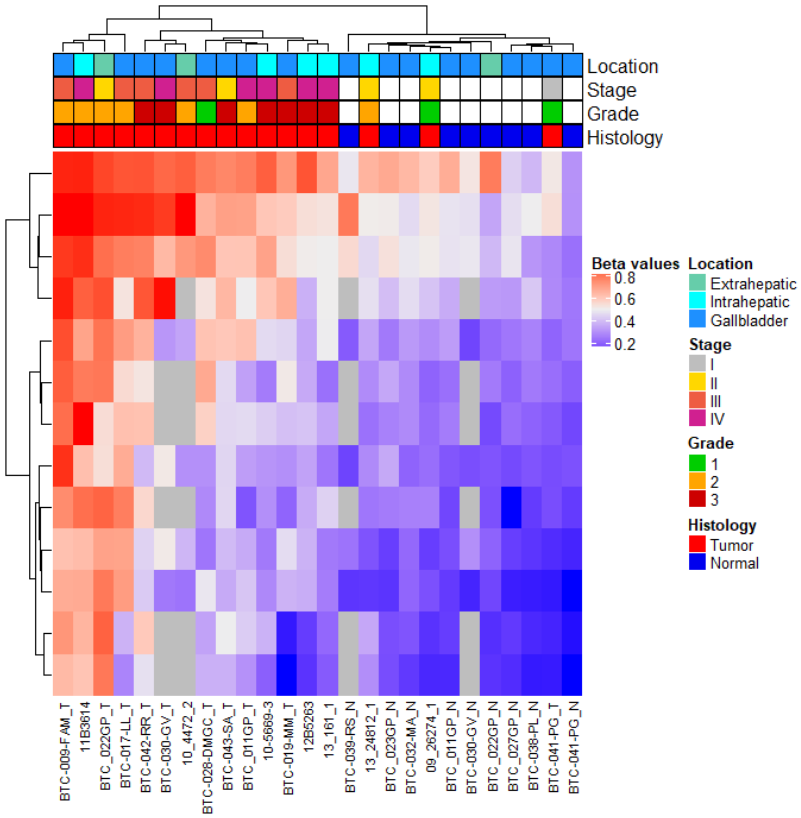


Figure 46. Unsupervised hierarchical clustering analysis based on the average CGI β values for the 13 *PCDH*-associated CGIs altered in our restricted discovery dataset

4.3.4 *PCDH*-associated CGI alterations in gastric cancer

Methylation status of *PCDH*-associated CGIs was also investigated in 22 gastric tumours and their matched-normal samples revealing four statistically significantly hypermethylated CGIs, mapping on promoter regions and of which three were associated with CTCF binding sites, (Table 18, Figure 47A, B). Lower differential methylation values were observed in body/fundus tumour samples compared to their respective normal samples (CpG 28 = 0.119, CpG45 = 0.106, CpG 95 = 0.067 and CpG 22 = 0.130).

UCSC CGI	UCSC CGI name	CTCF binding site	Δβ value (GC-Control)	Genes within region	Gene promoter-associated CGI
chr5:140762401-140762768	CpG 28	chr5:140761029-140763470	0.241	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*	*PCDHGA7*
chr5:140787447-140788044	CpG 45	chr5:140786247-140788714	0.210	*PCDHGA1*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*	*PCDHGB6*
chr5:140855386-140856620	CpG 95	chr5:140854218-140856648	0.212	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA12*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*, *PCDHGC3*	*PCDHGC3*
chr5:140864527-140864748	CpG 22	---	0.243	*PCDHGA1*, *PCDHGA10*, *PCDHGA11*, *PCDHGA12*, *PCDHGA2*, *PCDHGA3*, *PCDHGA4*, *PCDHGA5*, *PCDHGA6*, *PCDHGA7*, *PCDHGA8*, *PCDHGA9*, *PCDHGB1*, *PCDHGB2*, *PCDHGB3*, *PCDHGB4*, *PCDHGB5*, *PCDHGB6*, *PCDHGB7*, *PCDHGC3*, *PCDHGC4*	*PCDHGC4*

Table 18. Altered CGIs in gastric cancer

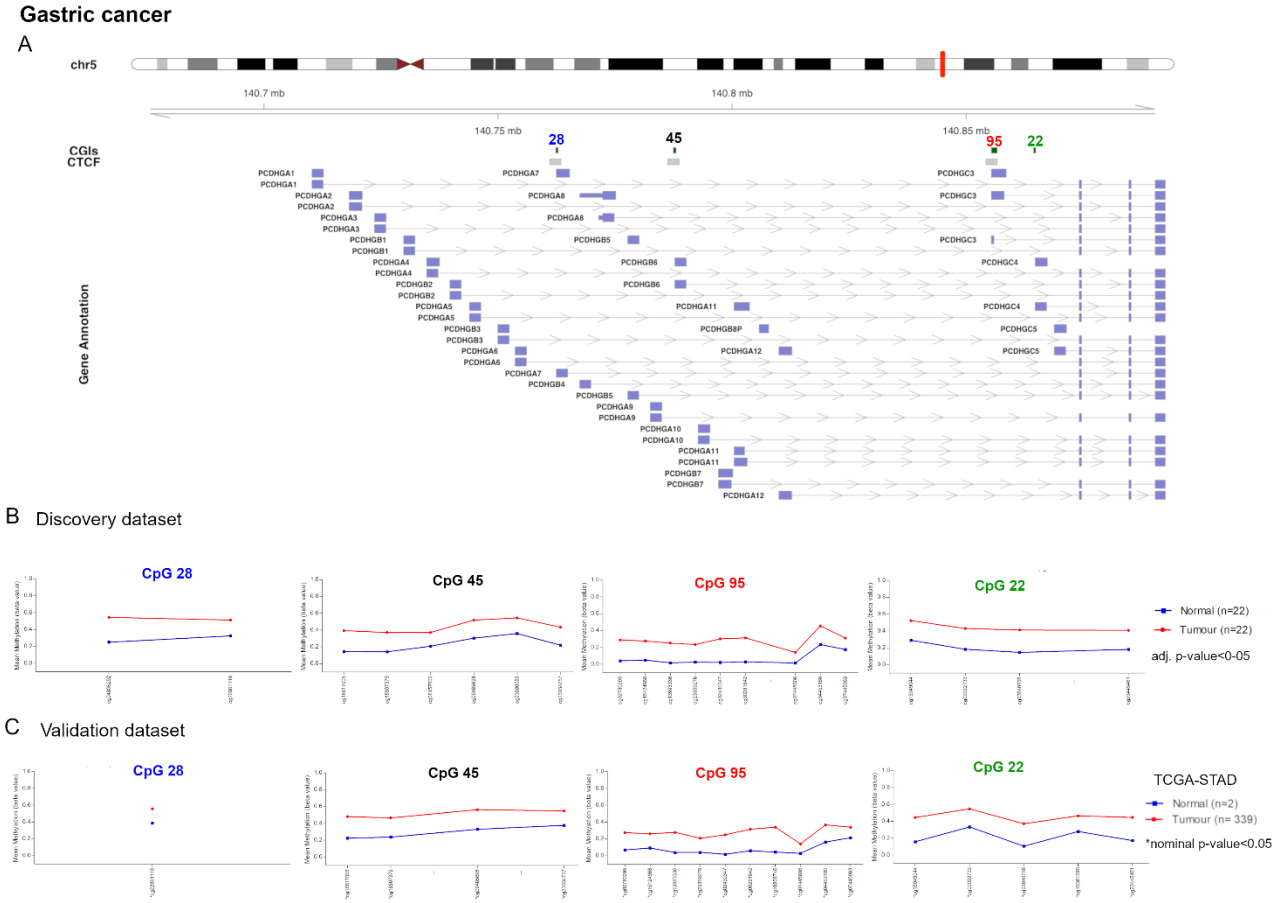


Figure 47. Genomic organization of *PCDHG@*, including the localization of exons, CGIs (annotated with the UCSC CGI names) and CTCF binding sites (A); β values, resulting from the average of our discovery samples set of each probe mapping to the altered CGIs (B); β values, resulting from the average of TCGA-STAD validation sample sets of each probe mapping to the altered CGIs (C)

UHC analysis revealed three clusters of tumours: one (N = 7) characterized by high methylation values in all the altered CGIs, a second group clustering (N = 5) along with normal samples and a third group (N = 10) showing intermediate methylation values compared to the other groups (Figure 48). Moreover, a statistically significant enrichment (p-value= 2.0E− 02) of MSI tumours (8/9) were observed in the group showing high methylation values (Figure 48).

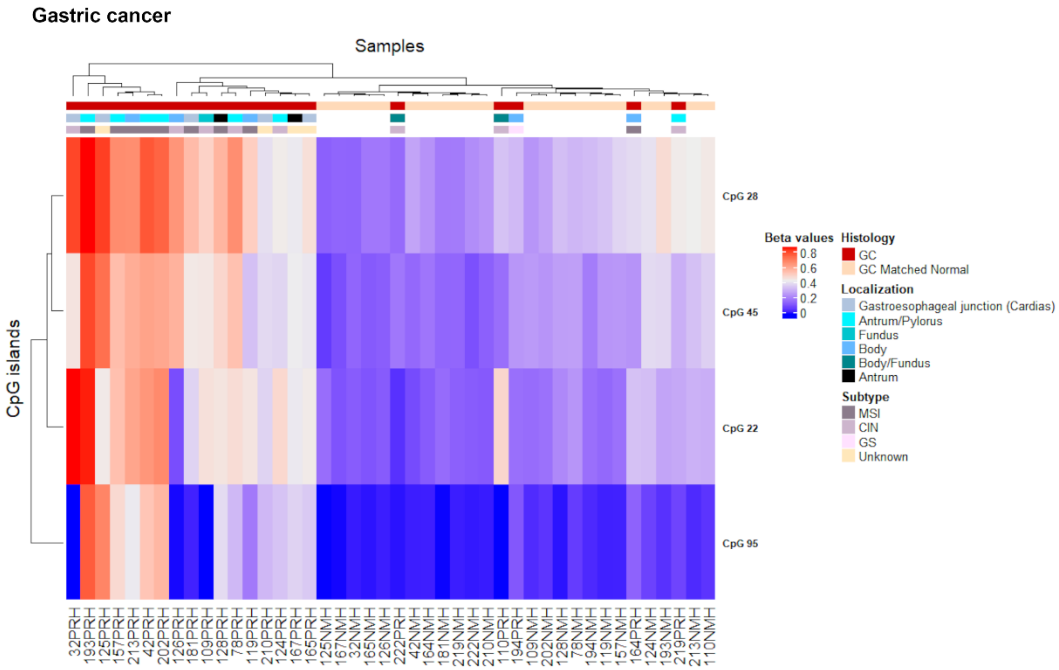


Figure 48. Gastric cancer discovery set unsupervised hierarchical clustering analysis based on the average CGI β values for the four altered *PCDH*-associated CGIs

The identified methylation alterations were successfully validated in TCGA-STAD dataset (Figure 47C). UHC analysis of this validation samples set with available molecular subtype categorization (N = 248), revealed two clusters of tumours characterized by different methylation levels for these altered CGIs and confirmed that MSI tumours clusterized mainly (47/49) in the group of tumours with higher methylation values (p-value = 1.3E− 10) (Figure 49). Moreover, the subgroup of tumours displaying high β values in all CGIs (within the dashed box) mainly included MSI samples (16/24). A statistically significant higher frequency tumours (22/25) positive for Epstein-Barr virus (EBV) infection clustered was observed in the cluster of highly methylated tumours (p-value = 1.4E− 04) (Figure 49).

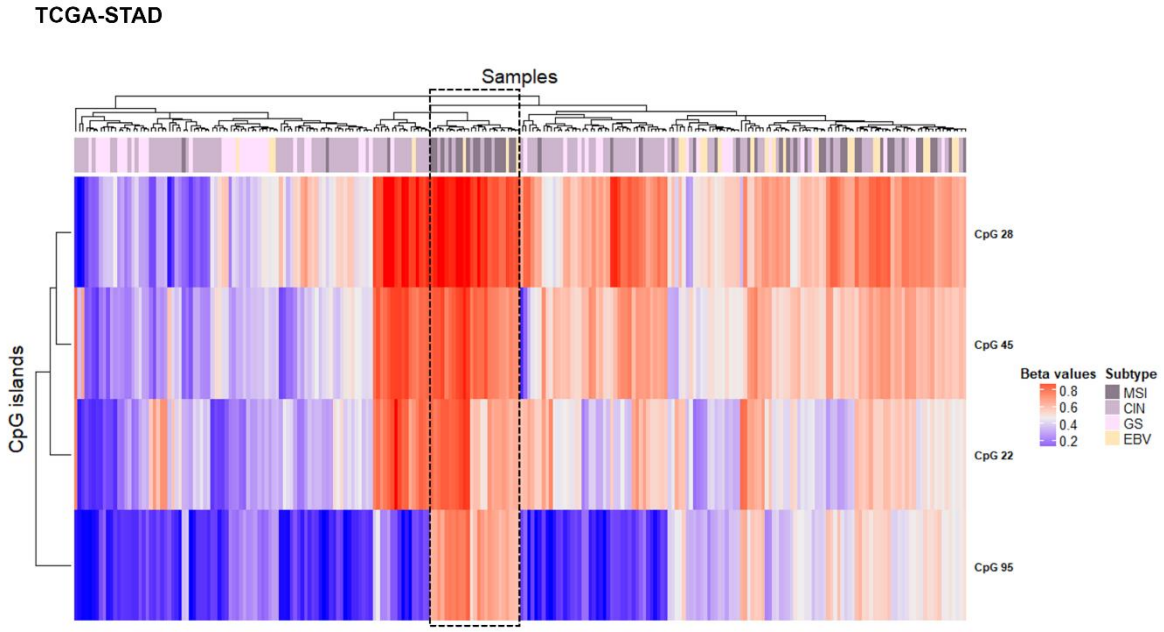


Figure 49. TCGA-STAD validation set unsupervised hierarchical clustering analysis based on the average CGI β values for the four altered *PCDH*-associated CGIs

4.3.5 *PCDH*-associated CGI alterations in pilocytic astrocytoma

Methylation data obtained from the analysis of a completely different type of solid cancer, pilocytic astrocytoma revealed DNA hypomethylation (Δβ value = − 0.285) of a flanking CGI region (chr5:140871064-140872335, CpG 122) mapping on *PCDHGC5* promoter and associated with the *PCDHG* cluster and two CTCF binding sites (Table 19, Figure 50A, B).

Data analysis of a validation dataset (GSE44684) confirmed hypomethylation of this region that was extended in the flanking CGI (chr5:140871064-140872335, CpG 122) that could not be investigated in our discovery dataset for the lack of epigenome coverage using Illumina 27K arrays (Figure 50C).

UCSC CGI	UCSC CGI name	CTCF binding site	Relation to UCSC CGI	Δβ value (PA-Control)	Genes within region	Gene promoter-associated N-shelf
chr5:140871064-140872335	CpG 122	chr5:140870147-140872480 chr5:140871308-140873492	N-shelf	-0.285	*PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PCDHGB1, PCDHGB2, PCDHGB3, PCDHGB4, PCDHGB5, PCDHGB6, PCDHGB7, PCDHGC3, PCDHGC4, PCDHGC5*	PCDHGC5

Table 19. Altered CGIs in pilocytic astrocytoma

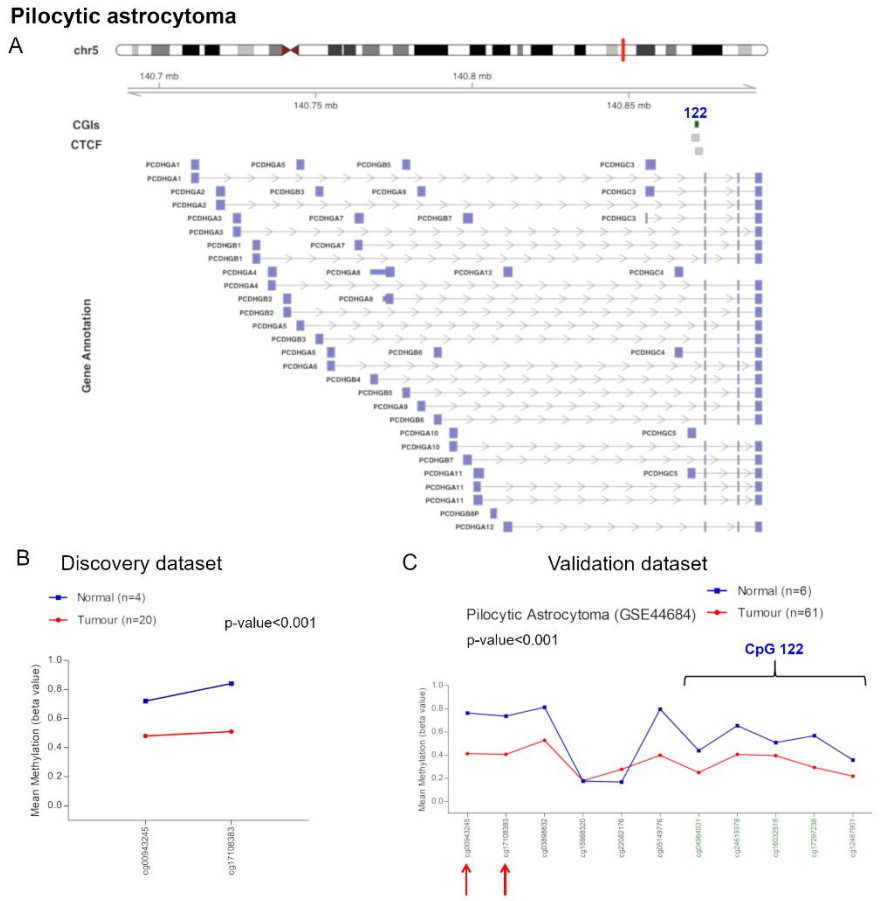


Figure 50. Genomic organization of *PCDHG@*, including the localization of exons, CGIs (annotated with the UCSC CGI names) and CTCF binding sites (A); β values, resulting from the average of our discovery samples set of each probe mapping to the altered N-shelf (B); β values, resulting from the average of pilocytic astrocytoma validation sample set of each probe mapping to the altered N-shelf and flanking CGI (C)

4.4 Editing DNA methylation: DNA methylation alterations as possible therapeutic targets

The potentiality of methylation alterations as therapeutic targets was explored by editing methylation status of a selected CGI, *GRIA4*-associated CGI, in HCT116, a 2D-colon cancer cell line. This CGI is hypermethylated in human colorectal cancer tissues and it is also known to be highly methylated in this cell line (ENCODE/HAIB data). Targeted de-methylation of the most hypermethylated region of *GRIA4*-associated CGI, including eight CpG sites, was carried out by the CRISPR-dCas9 system, using a plasmid for the expression of the catalytic domain of Tet1 demethylase and a plasmid expressing sgRNAs for targeting the region of interest.

Pyrosequencing analysis revealed an almost 15% reduction of the targeted region in HCT116 cells using this system (Figure 51). As expected, the eight CpG sites analysed showed similar methylation levels (Table 20). A reduction in methylation levels was observed already at 24h after sgRNA plasmid transduction and was still observed after 72h. This methylation decrease was similar to that observed by treating HCT116 cells with 5-AZA (Figure 51). However, de-methylation seemed to be independent of the presence of sgRNAs, suggesting that de-methylation was not specific for the region of interest (Figure 51).

	CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	CpG 6	CpG 7	CpG 8
DMSO	99	90	93	88	84	98	87	87
5-AZA	79	76	83	77	73	83	77	75
CRISPR-dCas9 (24h)	77	74	80	75	70	80	73	74

Table 20. Pyrosequencing results for the 8 CpG sites analysed in HCT116 cells

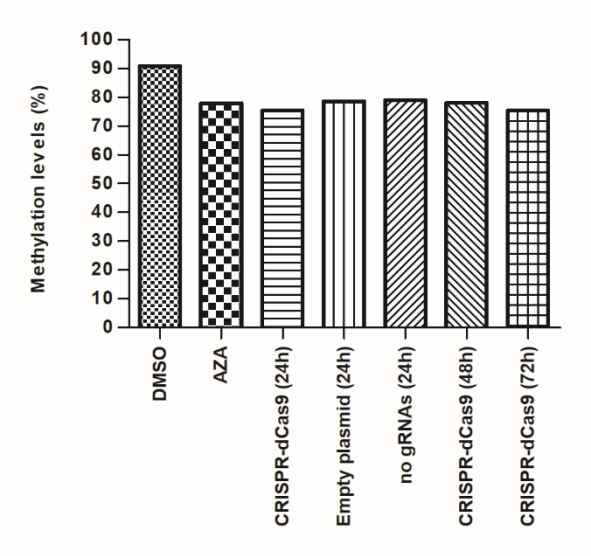


Figure 51. Pyrosequencing results. Bar plots show HCT116 methylation levels (%)

5. DISCUSSION

Epigenetic dysregulation is one of the tumour hallmarks. DNA methylation alterations, one of the several epigenetic aberrations in cancer, are early events in tumorigenesis and might represent powerful biomarkers for prediction of cancer risk, early diagnosis, prognosis, prediction of therapy response and tumor relapse.

DNA methylation changes have been associated with changes in gene expression. However, the relationship between these two events is not yet clear and needs to be elucidated.

Moreover, since DNA methylation is a reversible epigenetic modification, different strategies can be employed to edit DNA methylation in tumours in order to restore methylation patterns of normal cells, possibly reverting the tumour phenotype.

In this project thesis, the roles of DNA methylation as tumour biomarker, functional mechanism and therapeutic target have been explored.

In the last decades, several cancer DNA methylation-based biomarkers have been identified and some, such as *MGMT* and *SEPT9*, respectively for therapy response prediction in glioblastoma multiform and colorectal cancer detection, have been introduced in clinical practice.

However, even in cancers such as CRC for which a multitude of biomarkers have been proposed, specific-cancer biomarkers able to recognize the disease in the early phases are still lacking.

In this thesis work, methylation analyses of colorectal cancer, chronic lymphocytic leukemia and biliary tract cancer have been performed to identify potential tumour biomarkers.

Methylation data obtained in our laboratory from a previous study constitute the basis for the current CRC study. Two CGIs, mapping on promoter regions of *GRIA4* and *VIPR2* genes, have been selected from a panel of 74 CGIs altered in both CRC and adenomas previously identified in our laboratory (Fadda et al., 2018). These two CGIs showed high specificity and sensitivity in both our discovery set and TCGA-COAD validation set. Targeted-methylation analysis confirmed hypermethylation of these two CGIs in tumour tissues compared to matched-normal tissues. In cases where DNA methylation differences between tumour and matched-normal samples have not been detected, the histopathological analysis of the tissue specimens has revealed a low content of tumour cells confirming the high specificity of our biomarkers for tumour tissues.

The possible use of these biomarkers as non-invasive tools for CRC detection was explored by analyzing stool samples from the same CRC patients collected intraoperatively from the bowel resection specimens. Some samples were positive for DNA methylation of these biomarkers in the analysis conducted by MethyLight. However, the use of a more sensitive method, ddPCR, allowed to detect methylation in all, except one, samples for both biomarkers. It has been shown that ddPCR is able to overcome the technical challenges (i.e. poor DNA quality, presence of contaminating DNA

and PCR inhibitors) responsible for a reduced performance of MethyLight for the detection of low copies of methylated DNA in such samples (Yu et al., 2015). This method resulted 25-fold more sensitive than conventional MethyLight (Yu et al., 2015).

Interestingly, patients with ulcerative neoplasia showed more copies of methylated *VIPR2*-associated CGI, possibly related to the increased shedding of tumour cells from the ulcerated tumours.

Overall, our results reinforce the potential of *GRIA4*-associated CGI as non-invasive biomarker for CRC (Barault et al., 2018; Fadda et al., 2018) and proposed for the first time *VIPR2*-associated CGI as an additional CRC biomarker.

Further studies are needed to select the most powerful biomarkers and define a multi-biomarker panel for a non-invasive early detection of CRC. In fact multi-biomarkers panels have shown to improve diagnostic accuracy (Carmona et al., 2013; Imperiale et al., 2014; Barault et al., 2018; Fadda et al., 2018; Freitas et al., 2018). The future aim of this project is to test this panel in a larger dataset including stool samples from patients who have already received a diagnosis of CRC in a first step and samples from FOBT screening in a second phase.

In the CLL study, a methylome analysis of CLL and normal-control samples have been carried out to identify methylation patterns characteristics of CLL. Our results confirmed a global hypomethylation in CLL (Kulis et al., 2012) and identified several CGIs differentially methylated between CLL and controls. Although three CGIs resulted statistically significant hypomethylated in CLL, we focused our attention on hypermethylation events. In fact, in order to propose a biomarker, it must be taken into account that it is easier to define cut-off values to consider a sample as hypermethylated compared to a normal methylation condition.

A CGI, located in the gene body of *SHANK1*, was the most hypermethylated CGI in CLL, although its statistical significance did not resist correction for multiple testing, probably for the small sample size of our discovery set.

Although we are aware of the importance of significance threshold for epigenome-wide studies (EWAS) (Saffari et al., 2018), it is important to consider the cell heterogeneity of the samples analysed. Our previous work has already shown that statistically significant methylation alterations detected in CRC samples did not reach statistical significance after multiple testing correction in heterogenous samples such as adenomas. In fact, while many tumors cells harbor the epimutations in CRC tissues, adenomas include a mixture of cells with and without methylation alterations.

However, the validation of *SHANK1*-associated CGI hypermethylation in a larger CLL cohort including CLL samples with ≥95% neoplastic cells and normal control B cells samples, confirmed that this methylation alterations occurs in neoplastic cells (Kulis et al., 2012). The detection of this methylation alteration in whole blood samples, such as those of our discovery set, reinforces its

potential use of *SHANK1*-associated CGI as a diagnostic biomarker to be introduced in clinical practice without performing an expensive method such as cell sorting.

Moreover, a statistically significant differential methylation between U-CLL and normal naïve B cells was detected in the validation dataset (Kulis et al., 2012) suggesting that *SHANK1*-associated CGI hypermethylation might be correlated to a worst patient prognosis. In fact, CLL patients with a high level of IGHV mutation (M-CLL) have a favorable prognosis, while U-CLL patients have generally poor outcomes. Also in our discovery set, IGHV-mutated samples had an average methylation value of 35.0%, while IGHV-unmutated samples showed an average methylation value of 49.6 %. Moreover, the exclusion of an outlier (69.0% methylation), an IGHV-mutated sample firstly diagnosed as follicular lymphoma, determines a drop of the average methylation level to 23.7%.

Interestingly, methylation levels of *SHANK1*-associated CGIs positively correlated with the sample lymphocyte count and with CD19+ B-cells contributions. The increased lymphocyte count at diagnosis in a CLL patient is mainly due (> 80%) to the proliferation of neoplastic cells. In fact, in patients with an absolute lymphocyte count of at least 5×10^9/L, monoclonal B-cells represented more than 86% of B-cells in contrast to polyclonal B-cells representing only a small fraction (Shanafelt et al., 2009). This suggests that *SHANK1*-associated CGI methylation might be a biomarker of tumour aggressiveness in CLL.

Furthermore, the identification of a statistically significant gain of methylation (differential methylation of 3%) of this CGI in peripheral blood collected many years prior to MBCN diagnosis compared to peripheral blood samples of matched controls, suggested that this methylation alteration might be a marker also for other MBCNs although methylation data at diagnosis are needed to confirm this hypothesis.

Interestingly, an even more pronounced gain of methylation (differential methylation of 4.7%) was observed by analyzing CLL/SLL cases within the prospective MBCN cohort.

Methylation levels of this marker are in line or even higher than those of other methylation markers identified in studies analyzing pre-diagnostic blood samples years prior to diagnosis of B-cell neoplasm (Wong Doo et al., 2016; Georgiadis et al., 2017)

Nevertheless, since cases methylation values overlap with control methylation values, this methylation alteration cannot represent a predictive biomarker at individual level. However, 210/438 (48%) MBCN samples showed a differential methylation greater than the average $\Delta\beta$ (0.03), ranging from 0.031 to 0.36, and 38/82 (46%) CLL/ SLL samples showed a differential methylation greater than the average $\Delta\beta$ (0.047), ranging from 0.06 to 0.33. It would be interesting to perform an analysis of the *SHANK1*-associated CGI methylation status prior diagnosis, at diagnosis and during follow-up to confirm its role as a CLL cancer risk biomarker.

The BTC study consisted in a methylome analysis of a cohort of BTC and their matched-normal samples, even larger than the TCGA-CHOL cohort, with the aim of identifying methylation alterations in this rare, highly malignant tumour. In fact, very few studies have been focused on the identification of biomarkers for BTC that, due to its silent course in the first stages of the disease, is currently diagnosed in the late stages when the therapeutic solutions are limited, and the prognosis is poor.

The total number of differentially methylated CGIs identified in this study were markedly lower than that observed by analyzing other gastrointestinal tumors such as colorectal cancer and gastric cancer, probably for the use of DNA extracted from FFPE samples rather than from FF samples as in the other studies. In fact, an abnormal beta value distribution was observed for many of the analysed samples that were therefore excluded from the analysis. The methylome analysis conducted on a restricted dataset of good-quality samples allowed the detection of a higher number of differentially methylated CGIs between tumour and matched-normal controls. Although our analysis did not detect any CGIs statistically significantly methylated after multiple testing correction, probably for the reduced sample size, the detected methylation alterations were successfully validated in TCGA-CHOL dataset.

The analysis of TCGA-CHOL dataset revealed a higher number of both hypermethylated but especially hypomethylated CGIs compared to our discovery set. This might be due to a better quality of the DNA samples in the validation set or to the fact that this TCGA-CHOL set mainly included intrahepatic samples in contrast to our dataset including all the tumor localizations. However, our analyses did not reveal any difference between tumors from different location as observed for other molecular markers in other studies (Sohal et al., 2016) but this might be related to the small sample size or to the fact that our selected differentially methylated CGIs are independent from tumour localization as already observed for methylation alterations detected in our CRC study.

UHC analysis showed a sample distribution into two main clusters: one cluster of tumours displaying high methylation values and a second cluster, furtherly branched into two additional clusters, respectively including tumors with intermediate methylation values and a cluster of normal and three BTCs with low methylation values. A similar clusterization was observed in TCGA-CHOL dataset.

A statistically significant enrichment of low-stage tumours (I and II) was observed in the cluster of tumours including also normal-matched samples compared to the other clusters in our dataset. This was not observed in TCGA-CHOL dataset including mostly stage I tumours.

RPMM tumours clusterization identified different tumour clusters in both our discovery set and TCGA-CHOL validation set. However, no correlation with the considered clinical information was observed. Further information including therapy response, viral infection, genetic mutations might be

interesting to be collected in order to investigate the correlation with tumour classification and provide molecular markers for patients' stratification. In fact, in the study conducted by the TCGA consortium, an association between methylation profiles and genetic mutations was observed (Farshidfar et al., 2017). An association between methylation profiles and genetic mutations was also observed in a large study including nearly 500 BTCs from 10 countries (Jusakul et al., 2017). Moreover, the authors of this study have found that a cluster of tumours displaying hypermethylation of CpG islands was enriched in fluke-positive tumours, while another cluster of tumours showing hypermethylation of CpG shores was enriched in fluke-negative tumours (Jusakul et al., 2017).

The specificity of these methylation alterations for BTC was tested by analyzing methylation data from TCGA-COAD and TCGA-STAD datasets. 30 CGIs were identified as specifically methylated in BTC and not in the other two gastrointestinal tumours analysed. UHC analysis performed using methylation data of these 30 CGIs showed samples clusterization similar to that obtained using methylation data including also CGIs not specifically altered in BTC.

Among these 30 CGIs, seven showed high specificity and sensitivity values in both our discovery set and TGCA-CHOL.

The identified BTC-specific methylation alterations might be potential biomarkers for an early diagnosis of BTC through non-invasive methods. In fact, BTC diagnosis is currently carried out by the histopathological analysis of tissue biopsy, a highly-invasive procedure. Very few studies have been focused on the identification of non-invasive biomarkers for BTC early diagnosis. A study has identified two serum methylation-based biomarkers, *OPCML* and *HOXD9* that might be allow to distinguish the diagnosis of BTC from the diagnosis of other biliary diseases (Wasenang et al., 2019). However, these biomarkers are not specific for BTC since they have been found hypermethylated in other cancers (Wasenang et al., 2019). To our knowledge, our study is the first study that has identified potential specific methylation-based biomarkers for BTC. Of course, other studies are needed to validate the identified biomarkers in a large independent BTC cohort and the test the presence of these biomarkers in non-invasive matrices such as serum and plasma and potentially in stool samples since bile is released into the intestine.

The association between selected methylation alterations identified in the CRC and CLL studies and gene expression of the associated genes was investigated. Gene expression analyses results showed a negative correlation between methylation and gene expression.

A statistically significant downregulation of GRIA long, SLC6A3 and SPOCK1 both at mRNA and protein level was observed in CRC samples compared to matched-normal controls. *VIPR2* and *GRIA4* short showed a statistically significant downregulation in CRC samples only at mRNA levels, while although, VIPR2 and GluR4 short protein levels did not show statistically significant differences

between tumour and matched-normal samples, probably for the high variance among the samples, a trend towards lower levels in CRC samples was evident. The identified hypermethylated CGIs mapped on promoter regions confirming the association between promoter hypermethylation and gene downregulation described in a multitude of studies and considered as a mechanism for gene transcriptional inactivation in cancer (Baylin, 2005; Jones, 2012). Moreover, the current and other studies (Antonelli et al., 2018; Fadda et al., 2018) conducted by our research groups also confirmed that methylation in cancer targets genes that are barely expressed in the respective normal tissue where tumour arises (Sproul et al., 2011; Moarii et al., 2015). This is in agreement with the concept of "epigenetic switching" guaranteeing a stable and permanent repression of these genes and important for cell proliferation (Gal-Yam et al., 2008; Loi and Zavattari, 2019). It has been hypothesized that hypermethylation of PRC-targets maintains cancer cells in a stem-cell-like aggressive state (Widschwendter et al., 2007). In addition, our results showed that although these genes are already expressed at very low levels in normal tissues, promoter hypermethylation in cancer is associated to a further reduction of their expression levels that can be observed only by means of a targeted gene expression analysis using methods such as qRT-PCR. In fact, the background levels of hybridization to probes in gene expression microarray do not allow to detect small gene expression changes of low-expressed genes (Sproul and Meehan, 2013; Fadda et al., 2018).

qRT-PCR results confirmed *GRIA4* reduced expression in CRC (Fadda et al., 2018), also validated in TCGA-COAD RNA-seq data. *GRIA4* gene encodes for the Glutamate Ionotropic Receptor AMPA Type Subunit 4 (GluR4) of the AMPA tetrameric receptor complex, a cationic ion channel, mainly expressed in glutamatergic synapses. Different *GRIA4* transcript variants and two protein isoforms, a long canonical isoform of 902 amino acids and a short isoform of 433 amino acids including only the extracellular protein domain, have been described.

The transcript variant encoding for the canonical isoform resulted higher expressed than the short transcript in both tumour and matched-normal samples. However, an opposite situation was observed at protein level, where the short isoform resulted more expressed than the canonical isoform. This might be related to the lifetime of each isoform. In fact, although the degradation mechanism of GluR4 is not well elucidated, it has been previously reported that caspases cleave GluR4 at the C-terminal domain contributing to selective proteolysis (Chan et al., 1999). This process might increase the susceptibility of GluR4 long isoform to degradation but would not affect the short one that lacks the C-terminal region. Another hypothesis is the post-transcriptional regulation mediated by microRNAs (miRNAs). In fact, using TargetScan (http://www.targetscan.org/vert_72/) bioinformatic tool, four miRNAs, miR-506-3p, miR-124-3p.1, miR-124-3p.2 and miR-137, highly conserved among vertebrates, with the strictest matching site types and high preferentially conserved targeting

scores, have been identified to target the long *GRIA4* transcript variant. These interactions were also reported by miRanda (http://www.microrna.org/microrna/home.do) and DIANA micro-T (http://diana.imis.athena-innovation.gr/DianaTools/). In contrast, no miRNAs interactions were predicted for the short transcript using the same parameters, suggesting a specific transcript regulation and possibly explaining the higher expression of the short protein isoform.

VIPR2 downregulation was also validated in TCGA-COAD RNA-seq data. This gene belongs to a class of genes encoding for vasoactive intestinal peptide receptors (VIPRs). VIPRs, members of the G-protein-coupled receptors (GPCRs) superfamily, comprise VPAC1 and VPAC2, showing a similar affinity for the vasoactive intestinal peptide (VIP) and pituitary adenylate cyclase-activating polypeptide (PACAP) and PAC1 with a high affinity for PACAP and a low affinity for VIP (Dickson and Finlayson, 2009). VIP is usually secreted by nerve endings and immune cells in the gastrointestinal tract and is involved in gut motility acting as a vasoregulatory hormone, whereas PACAP has hypophysiotropic effect on pituitary hormone secretion (Sherwood et al., 2000; Vaudry et al., 2000; Waschek, 2013). VIPRs expression have been detected in a variety of tumors. In particular, VPAC1 receptors are expressed in malignant epithelial neoplasms, while expression of VPAC2 receptors has mainly been found in some leiomyomas and gastrointestinal stromal tumors (Oka et al., 1998; Reubi et al., 2000; Fernández-Martínez et al., 2012; Liu et al., 2014). On the other hand, our results showed that VIPR2 expression can be also detected in colorectal cancer. However, we did not detect a statistically significant downregulation of VIPR2 at protein level, probably for the large variability among the samples.

A study has found that expression of VPAC1 in CRC tissues is negatively correlated with the differentiation of colon cancer suggesting that it is associated with the malignancy of the tumour. Moreover, the results of this study have shown higher levels of VPAC1 in blood vessels surrounding colon cancer and in tumor-associated macrophages (TAMs) than in normal colon mucosa (Liu et al., 2014). Our results, although obtained from the analysis of a restricted number of samples, did not show an association between *VIPR2* expression and tumour grade.

Gene expression analysis of *SPOCK1*, encoding for a Ca^{2+}-binding matricellular glycoprotein, showed a statistically significant average reduced expression in CRC studies compared to matched-normal samples, that was also observable by the paired comparison between CRC to its respective normal tissue in most cases. This reduced expression was also observable at protein level. These results are in contrast with other studies that detected a higher expression of this gene in CRC (Zhang et al., 2017) and in other tumour types (Miao et al., 2013; Wang et al., 2016). On the other hand, TCGA-COAD RNA-seq data showed a similar *SPOCK1* expression between CRC and normal control tissue samples. The paired comparison between tumours and their normal-matched samples

of our discovery set has also shown a higher expression in matched-normal samples for few cases, indicating that *SPOCK1* expression is highly variable. A gene expression analysis conducted by microarray, showed that *SPOCK1* was upregulated in the normal mucosa adjacent to CRC tissue (minimum distance of 10 cm) compared to colon mucosa from healthy donors but its expression was similar to that of the tumour (Sanz-Pamplona et al., 2014). Protein-protein network analysis showed that SPOCK is a protein secreted by adjacent mucosa and interact with a receptor in tumor (Sanz-Pamplona et al., 2014). Therefore, further studies to clarify these discordant results are needed.

SLC6A3, a gene encoding for the dopamine transporter was downregulated in our dataset but it showed a higher expression in normal samples in TCGA-COAD RNA-seq data. However, an overlapping of gene expression values of tumour and normal samples can be observed in TCGA-COAD dataset. The discrepancy with our results might be related to the low expression of this gene in colon (GTEx data: about 0.8 TPM) and to even reduced expression in tumours that might be not detected by techniques such as whole exome / whole transcriptome NGS / microarrays. Therefore, this result should be validated in a larger sample cohort by ultra-sensitive methods such as ddPCR.

In CLL study, gene expression analysis of *SHANK1* showed a statistically significant downregulation in our CLL cohort. *SHANK1* is one of the three members of the *SHANK* (SH3 And Multiple Ankyrin Repeat Domains) gene family. SHANK proteins (SHANK1, SHANK2 and SHANK3) act as scaffold proteins and have a fundamental role in the formation, development and function of neuronal synapses. A study investigating the association between *SHANK* genes methylation and their expression revealed that, although all these genes present several methylated CpG sites, only *SHANK3* was highly methylated in tissues where its expression was low or absent, suggesting that methylation might regulate tissue-specific *SHANK3* expression. In contrast, our study identified an association between methylation and expression for *SHANK1*.

SHANK1 downregulation in CLL was not validated in RNA-seq data, possibly, also in this case, related to the very low expression of this gene in whole-blood and to the consequent inability to detect its downregulation by this method. Interestingly, the identified hypermethylated CGI is located in the gene body of *SHANK1* gene confirming a possible negative association between gene body methylation and gene expression previously reported in CLL (Kulis et al., 2012). The identified altered CGI is located downstream of the main transcript but upstream of a short alternative transcript possibly regulating its expression. In fact, methylation of intragenic CGIs has been negatively correlated to expression of alternative transcripts and positively associated with expression of the main transcripts (Kulis et al., 2012). A study to quantify each *SHANK1* transcript expression should be conducted to elucidate this hypothesis.

From the methylome analyses of our discovery sets emerged that frequent methylation alterations can be observed in *PCDH* cluster in cancer, confirming that the epigenetic dysregulation of protocadherins observed in different human diseases, including cancer (Waha et al., 2005; Novak et al., 2008; Dallosso et al., 2012; Wang et al., 2015; El Hajj et al., 2017).

In particular, our results showed *PCDH* methylation alterations in gastrointestinal tumors and gliomas but not in blood cancer (CLL).

The absence of *PCDH*-associated CGIs methylation alterations in CLL suggested that they are not targeted by methylation during tumorigenesis in haematological neoplasms in contrast to solid tumours. A possible explanation might be related to the normal function of PCDH in cell-cell adhesion (Murata et al., 2004; Shan et al., 2016) that is not essential in blood cancer in contrast to its importance in tumour mass formation. Studies of other types of hematological neoplasms are needed to confirm this hypothesis, including a study of multiple myeloma to investigate if, in this type of tumour affecting plasma cells originating in the blood marrow, there is a different methylation pattern.

Moreover, our results showed that in tissues where *PCDH* are normally low expressed, *PCDH*-associated CGIs are hypermethylated in tumors, confirming that methylation in tumors targets genes lowly expressed in the respective normal tissue where tumour arises (Sproul et al., 2011; Moarii et al., 2015; Fadda et al., 2018).

On the other hand, hypomethylation of a CGI flanking region mapping to *PCDHG* was observed in gliomas. Since PCDH are essential during neuronal development and their knockdown or deficiency leads to loss of different neuronal cell types, synapse decrease or dendritic arborization decline (Lefebvre et al., 2008; Garrett et al., 2012; Chen and Maniatis, 2013), their hypomethylation in tumors could be associated to an increased expression of PCDH suggesting that tumour cells can revert to the conditions required during development.

However, it is also possible that tumour methylation status mirrors the cell of origin pattern clonally expanded (Sharma et al., 2007; Sproul et al., 2011, 2012; Antonelli et al., 2018). It is important to consider the cellular heterogeneity of the tumour that contains both cells sharing the same genetic and epigenetic alterations that are acquired by the early precursors cells clonally expanded but also sub-clones of cells that harbor different genetic and epigenetic modifications that are acquired during or after the clonal expansion of these cells (Mazor et al., 2016).

Many altered *PCDH*-associated CGIs resulted associated to CTCF binding sites. CTCF are involved in PCDH transcription through the formation of DNA loops (Guo et al., 2012, 2015) and it has been found that methylation regulated CTCF binding to DNA (Golan-Mashiach et al., 2012). Therefore, it is plausible that the identified methylation alterations identified in our project, may avoid or modify CTCF binding to DNA, dysregulating *PCDH* transcription.

PCDHs in CRC are the most highly hypermethylated among the gastrointestinal tumours analysed. These methylation alterations can also be frequently detected in adenomas independently of the disease grade, suggesting that they are early events in tumorigenesis.

Interestingly, some CGI alterations were common in the different gastrointestinal cancers, while others were specific for each cancer-type. Nevertheless, the majority of these alterations are concentrated in the *PCDHG* cluster.

A correlation between hypermethylation and tumour location was also observed for some CGIs. For examples, CpG 95 was hypermethylated in colon cancer samples but not in rectal cancer samples; gastric cancer altered CGIs showed lower $\Delta\beta$ by comparing paired samples localized in body/fundus than samples from other localizations; higher β values were observed in BTCs from gallbladder/extrahepatic localization than intrahepatic tumours.

Thus, our study identified common methylation alterations that may provide broad-spectrum biomarkers as well as methylation alterations associated with tumour location that may be potential tissue-specific tumours biomarkers.

Correlation with other clinical information were observed analyzing gastric cancer methylation data. A correlation between MSI status and hypermethylation was observed, confirming previous reports (Ottini et al., 2006; Loh et al., 2014). Moreover, TCGA-STAD methylation data for the altered *PCDH*-associated CGIs showed that EBV-positive samples displayed high methylation levels as already observed in the TCGA study where EBV-positive samples were also CIMP positive (Cancer Genome Atlas Research Network, 2014).

The final aim of this thesis project was to test CRISPR-dCas9 system for specific epigenetic editing. This tool might represent a possible cancer therapeutic strategy and may provide a tool for the study of the functional effects of methylation alterations.

In this preliminary study, dCas9-Tet1 tool was employed to specifically de-methylate a restricted genetic region of *GRIA4*-associated CGI in HCT116 colon cancer cell lines. Preliminary results showed a decreased in methylation levels of the target region similar to that obtained by 5-AZA, a potent DNMT inhibitor, that persisted after 72h. However, this de-methylation was independent of the presence of sgRNAs suggesting that it was not specific for the target of interest. A possible explanation could be the high constitutive expression of dCas9-Tet1 that may be prevented by employing an inducible system. Future experiments will be carried out by using such a system and methylation of multiple gene regions will be tested to verify the specificity of this system. Moreover, the possible use of this system in a better translation model such as 3D cultures will be tested. Finally, dCas9-DNMT tool will be applied to induce methylation of specific targets in cell lines generated

from healthy donors to identify driver methylation changes that may lead to the acquisition of tumorigenic features.

6. CONCLUSIONS

In conclusion, this thesis work confirmed that methylation alterations are frequent and early events in cancer. In particular, it has allowed the identification of potential biomarkers for colorectal cancer, chronic lymphocytic leukemia and biliary tract cancer and the discovery that methylation alterations in *PCDH* cluster are frequent events in gastrointestinal tumours (colorectal cancer, gastric cancer and biliary tract cancer) and gliomas suggesting their possible importance in solid cancers and providing additional tumour biomarkers. Moreover, it explored and confirmed the association between hypermethylation and downregulation, supporting the idea that methylation may represent a functional mechanism to inactivate specific genes in cancer in a defined program. This work also highlights that genes targeted by methylation in cancer are lowly expressed in the tissue of origin but their hypermethylation is associated to a further downregulation of these genes suggesting its possible role in inducing their permanent silencing, as in agreement with the so-called "epigenetic switching". In fact, as mentioned, epigenetic modifications, characteristic of stem cells, occur very early in tumour cells, which therefore undergo an undifferentiation towards progenitor cells. Finally, the possible use of methylation as a cancer therapeutic target was investigated by employing CRISPR-dCas9 tool to edit DNA methylation patterns in tumour cells, suggesting that this approach represents an innovative and powerful system that can revolutionize cancer therapeutic strategies.

7. BIBLIOGRAPHY

Abdel-Wahab, O. et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144–147.

Amabile, A. et al. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219-232.e14.

Amiot, A. et al. (2014). The detection of the methylated Wif-1 gene is more accurate than a fecal occult blood test for colorectal cancer screening. PLoS One *9*, e99233.

Antonelli, M. et al. (2018). Integrated DNA methylation analysis identifies topographical and tumoral biomarkers in pilocytic astrocytomas. Oncotarget *9*, 13807–13821.

Assenov, Y. et al. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat. Methods *11*, 1138–1140.

Baran, B. et al. (2018). Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. Gastroenterol. Res. *11*, 264–273.

Barault, L. et al. (2018). Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer. Gut *67*, 1995–2005.

Baylin, S.B. (2005). DNA methylation and gene silencing in cancer. Nat. Clin. Pract. Oncol. *2*, S4–S11.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nat. Rev. Cancer *11*, 726–734.

Baylin, S.B., and Jones, P.A. (2016). Epigenetic Determinants of Cancer. Cold Spring Harb. Perspect. Biol. *8*.

Bergthold, G. et al. (2015). Expression profiles of 151 pediatric low-grade gliomas reveal molecular differences associated with location and histological subtype. Neuro. Oncol. *17*, 1486–1496.

Berman, B.P. et al. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. Nat. Genet. *44*, 40–46.

Bernstein, D.L. et al. (2015). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J. Clin. Invest. *125*, 1998–2006.

Bert, S.A. et al. (2013). Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell *23*, 9–22.

Bestor, T.H. et al. (2015). Notes on the role of dynamic DNA methylation in mammalian development. Proc. Natl. Acad. Sci. U. S. A. *112*, 6796–6799.

Bibikova, M. et al. (2009). Genome-wide DNA methylation profiling using Infinium (R) assay. Epigenomics *1*, 177–200.

Bibikova, M. et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Bock, C. (2009). Epigenetic biomarker development. Epigenomics *1*, 99–110.

Bonanno, C. et al. (2007). MS-FLAG, a Novel Real-Time Signal Generation Method for Methylation-Specific PCR. Clin. Chem. *53*, 2119–2127.

van den Boom, D., and Ehrich, M. (2009). Mass Spectrometric Analysis of Cytosine Methylation by Base-Specific Cleavage and Primer Extension Methods. In: Methods in Molecular Biology (Clifton, N.J.), 207–227.

Bracken, A.P. et al. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209.

Carmona, F.J. et al. (2013). DNA Methylation Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. Cancer Prev. Res. *6*, 656–665.

Catteau, A. et al. (1999). Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics. Oncogene *18*, 1957–1965.

Challen, G.A. et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat.

Genet. *44*, 23–31.
Chan, M.W.Y. et al. (2004). Quantitative detection of methylated SOCS-1, a tumor suppressor gene, by a modified protocol of quantitative real time methylation-specific PCR using SYBR green and its use in early gastric cancer detection. Biotechnol. Lett. *26*, 1289–1293.
Chan, S.L. et al. (1999). Evidence for caspase-mediated cleavage of AMPA receptor subunits in neuronal apoptosis and Alzheimer's disease. J. Neurosci. Res. *57*, 315–323.
Chatterjee, A. et al. (2012). Technical Considerations for Reduced Representation Bisulfite Sequencing with Multiplexed Libraries. J. Biomed. Biotechnol. *2012*, 1–8.
Chen, W. V., and Maniatis, T. (2013). Clustered protocadherins. Development *140*, 3297–3302.
Chen, Z.-X. et al. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J. Cell. Biochem. *95*, 902–917.
Choudhury, S.R. et al. (2016). CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget *7*.
Christman, J.K. (2002). 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483–5495.
Chuang, J.C. et al. (2010). S110, a 5-Aza-2'-Deoxycytidine-Containing Dinucleotide, Is an Effective DNA Methylation Inhibitor In vivo and Can Reduce Tumor Growth. Mol. Cancer Ther. *9*, 1443–1450.
Colaprico, A. et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71–e71.
Colella, S. et al. (2003). Sensitive and quantitative universal Pyrosequencing^{TM} methylation analysis of CpG sites. Biotechniques *35*, 146–150.
Cottrell, S.E. et al. (2004). A real-time PCR assay for DNA-methylation using methylation-specific blockers. Nucleic Acids Res. *32*, 10e – 10.
Dajani, S. et al. (2018). Water transport proteins-aquaporins (AQPs) in cancer biology. Oncotarget *9*, 36392–36405.
Dallosso, A.R. et al. (2012). Long-range epigenetic silencing of chromosome 5q31 protocadherins is involved in early and late stages of colorectal tumorigenesis through modulation of oncogenic pathways. Oncogene *31*, 4409–4419.
Dedeurwaerder, S. et al. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771–784.
Dickson, L., and Finlayson, K. (2009). VPAC and PAC receptors: From ligands to function. Pharmacol. Ther. *121*, 294–316.
Diehl, F. et al. (2005). Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc. Natl. Acad. Sci. U. S. A. *102*, 16368–16373.
Du, P. et al. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.
Eads, C.A. et al. (2000). MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. *28*, 32e – 0.
Fadda, A. et al. (2018). Colorectal cancer early methylation alterations affect the crosstalk between cell and surrounding environment, tracing a biomarker signature specific for this tumor. Int. J. Cancer *143*, 907–920.
Farshidfar, F. et al. (2017). Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH -Mutant Molecular Profiles. Cell Rep. *18*, 2780–2794.
Feinberg, A.P. et al. (2006). The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. *7*, 21–33.
Feldman, N. et al. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194.
Fernández-Martínez, A.B. et al. (2012). Overexpression of vasoactive intestinal peptide receptors and cyclooxygenase-2 in human prostate cancer. Analysis of potential prognostic relevance. Histol.

Histopathol. *27*, 1093–1101.
Ferreira, P.G. et al. (2014). Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. Genome Res. *24*, 212–226.
Freitas, M. et al. (2018). A novel DNA methylation panel accurately detects colorectal cancer independently of molecular pathway. J. Transl. Med. *16*, 45.
Gal-Yam, E.N. et al. (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc. Natl. Acad. Sci. *105*, 12979–12984.
Garrett, A.M. et al. (2012). γ-Protocadherins Control Cortical Dendrite Arborization by Regulating the Activity of a FAK/PKC/MARCKS Signaling Pathway. Neuron *74*, 269–276.
Gazzoli, I. et al. (2002). A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. Cancer Res. *62*, 3925–3928.
Georgiadis, P. et al. (2017). Evolving DNA methylation and gene expression markers of B-cell chronic lymphocytic leukemia are present in pre-diagnostic blood samples more than 10 years prior to diagnosis. BMC Genomics *18*, 728.
Golan-Mashiach, M. et al. (2012). Identification of CTCF as a master regulator of the clustered protocadherin genes. Nucleic Acids Res. *40*, 3378–3391.
Gonzalgo, M.L., and Jones, P.A. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res. *25*, 2529–2531.
Grady, W.M. et al. (2000). Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat. Genet. *26*, 16–17.
Gu, Z. et al. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.
Guo, Y. et al. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin promoter choice. Proc. Natl. Acad. Sci. *109*, 21081–21086.
Guo, Y. et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.
Hahn, M.A. et al. (2011). Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. PLoS One *6*, e18844.
El Hajj, N. et al. (2017). Epigenetic dysregulation of protocadherins in human disease. Semin. Cell Dev. Biol. *69*, 172–182.
Hansen, K.D. et al. (2011). Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. *43*, 768–775.
Hegi, M.E. et al. (2005). MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. N. Engl. J. Med. *352*, 997–1003.
Herman, J.G. et al. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc. Natl. Acad. Sci. *95*, 6870–6875.
Herman, J.G., and Baylin, S.B. (2003). Gene Silencing in Cancer in Association with Promoter Hypermethylation. N. Engl. J. Med. *349*, 2042–2054.
Hervouet, E. et al. (2013). DNA Methylation and Apoptosis Resistance in Cancer Cells. Cells *2*, 545.
Hinoue, T. et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. *22*, 271–282.
Hoffmann, A.-C. et al. (2009). Methylated DAPK and APC promoter DNA detection in peripheral blood is significantly associated with apparent residual tumor and outcome. J. Cancer Res. Clin. Oncol. *135*, 1231–1237.
Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226–232.
Hon, G.C. et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain

formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.
Houseman, E.A. et al. (2008). Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. BMC Bioinformatics *9*, 365.
Houseman, E.A. et al. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.
Houseman, E.A. et al. (2014). Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics *30*, 1431–1439.
Hu, C.Y. et al. (2014). Kidney Cancer Is Characterized by Aberrant Methylation of Tissue-Specific Enhancers That Are Prognostic for Overall Survival. Clin. Cancer Res. *20*, 4349–4360.
Imperiale, T.F. et al. (2014). Multitarget Stool DNA Testing for Colorectal-Cancer Screening. N. Engl. J. Med. *370*, 1287–1297.
Issa, J.-P.J. et al. (2005). Azacitidine. Nat. Rev. Drug Discov. *4*, 275–276.
Ito, S. et al. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science (80-.). *333*, 1300–1303.
Jeyapalan, J.N. et al. (2016). DNA methylation analysis of paediatric low-grade astrocytomas identifies a tumour-specific hypomethylation signature in pilocytic astrocytomas. Acta Neuropathol. Commun. *4*, 54.
Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.
Jones, P.A., and Baylin, S.B. (2007). The Epigenomics of Cancer. Cell *128*, 683–692.
Jusakul, A. et al. (2017). Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma. Cancer Discov. *7*, 1116–1135.
Kalari, S., and Pfeifer, G.P. (2010). Identification of Driver and Passenger DNA Methylation in Cancer by Epigenomic Analysis. 277–308.
Kang, J.-Y. et al. (2015). Identification of Long-Range Epigenetic Silencing on Chromosome 15q25 and Its Clinical Implication in Gastric Cancer. Am. J. Pathol. *185*, 666–678.
Kang, J.G. et al. (2019). Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. Sci. Rep. *9*, 11960.
Kantarjian, H. et al. (2007). Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. Blood *109*, 52–57.
Khatami, F. et al. (2017). Meta-analysis of promoter methylation in eight tumor-suppressor genes and its association with the risk of thyroid cancer. PLoS One *12*.
Klett, H. et al. (2018). Robust prediction of gene regulation in colorectal cancer tissues from DNA methylation profiles. Epigenetics *13*, 386–397.
Koch, A. et al. (2018). Analysis of DNA methylation in cancer: location revisited. Nat. Rev. Clin. Oncol. *15*, 459–466.
Kristensen, L.S. et al. (2008). Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. Nucleic Acids Res. *36*, e42–e42.
Kristensen, L.S. et al. (2009). PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin. Chem. *55*, 1471–1483.
Kristensen, L.S. et al. (2012). Methylation profiling of normal individuals reveals mosaic promoter methylation of cancer-associated genes. Oncotarget *3*, 450–461.
Kulis, M. et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. *44*, 1236–1242.
Lambert, S.R. et al. (2013). Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. Acta Neuropathol. *126*, 291–301.
Lastraioli, E. et al. (2015). Ion channel expression as promising cancer biomarker. Biochim. Biophys. Acta - Biomembr. *1848*, 2685–2702.

Leek, J.T., and Storey, J.D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. *3*, e161.
Lefebvre, J.L. et al. (2008). Gamma-Protocadherins regulate neuronal survival but are dispensable for circuit formation in retina. Development *135*, 4141–4151.
Lei, Y. et al. (2017). Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat. Commun. *8*, 16026.
Levanon, D. et al. (2011). Absence of Runx3 expression in normal gastrointestinal epithelium calls into question its tumour suppressor function. EMBO Mol. Med. *3*, 593–604.
Li, F. et al. (2007). Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. Nucleic Acids Res. *35*, 100–112.
Li, M. et al. (2009). Sensitive digital quantification of DNA methylation in clinical samples. Nat. Biotechnol. *27*, 858–863.
Liang, T.-J. et al. (2017). APC hypermethylation for early diagnosis of colorectal cancer: a meta-analysis and literature review. Oncotarget *8*.
Liggett, T.E. et al. (2011). Methylation patterns in cell-free plasma DNA reflect removal of the primary tumor and drug treatment of breast cancer patients. Int. J. Cancer *128*, 492–499.
Liu, S. et al. (2014). VPAC1 overexpression is associated with poor differentiation in colon cancer. Tumor Biol. *35*, 6397–6404.
Liu, X.S. et al. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247.e17.
Lizio, M. et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.
Lock, L.F. et al. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*, 39–46.
Loh, M. et al. (2014). DNA methylation subgroups and the CpG island methylator phenotype in gastric cancer: a comprehensive profiling approach. BMC Gastroenterol. *14*, 55.
Loi, E., and Zavattari, P. (2019). CpG Islands Methylation Alterations in Cancer: Functionally Intriguing Security Locks, Useful Early Tumor Biomarkers. In: The DNA, RNA, and Histone Methylomes, Springer (ed). 53–62.
Lujambio, A. et al. (2007). Genetic Unmasking of an Epigenetically Silenced microRNA in Human Cancer Cells. Cancer Res. *67*, 1424–1429.
Maegawa, S. et al. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res. *20*, 332–340.
Makambi, K. (2003). Weighted inverse chi-square method for correlated significance tests. J. Appl. Stat. *30*, 225–234.
Maksimovic, J. et al. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol. *13*, R44.
Mali, P. et al. (2013). RNA-Guided Human Genome Engineering via Cas9. Science (80-.). *339*, 823–826.
Maunakea, A.K. et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.
Mazor, T. et al. (2016). Intratumoral Heterogeneity of the Epigenome. Cancer Cell *29*, 440–451.
McDonald, J.I. et al. (2016). Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol. Open *5*, 866–874.
Meissner, A. et al. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.
Miao, L. et al. (2013). SPOCK1 is a novel transforming growth factor-β target gene that regulates lung cancer cell epithelial-mesenchymal transition. Biochem. Biophys. Res. Commun. *440*, 792–797.
Mikeska, T. et al. (2012). DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert Rev. Mol. Diagn. *12*, 473–487.

Mikeska, T., and Craig, J.M. (2014). DNA Methylation Biomarkers: Cancer and Beyond. Genes (Basel). *5*, 821.
Moarii, M. et al. (2015). Changes in correlation between promoter methylation and gene expression in cancer. BMC Genomics *16*, 873.
Mojica, F.J.M. et al. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.
Molnár, B. et al. (2015). Plasma methylated septin 9: a colorectal cancer screening marker. Expert Rev. Mol. Diagn. *15*, 171–184.
Morikawa, T. et al. (2005). A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. Gastroenterology *129*, 422–428.
Morris, T.J., and Beck, S. (2015). Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods *72*, 3–8.
Morsczeck, C. et al. (2018). The cell cycle regulator protein P16 and the cellular senescence of dental follicle cells. Mol. Cell. Biochem. *439*, 45–52.
Müller, F. et al. (2016). RnBeads – Comprehensive Analysis of DNA Methylation Data. 52.
Murata, Y. et al. (2004). Interaction with Protocadherin-γ Regulates the Cell Surface Expression of Protocadherin-α. J. Biol. Chem. *279*, 49508–49516.
Murray, D.H. et al. (2017). Validation of a Circulating Tumor-Derived DNA Blood Test for Detection of Methylated *BCAT1* and *IKZF1* DNA. J. Appl. Lab. Med. An AACC Publ. *2*, 165–175.
Naylor, S. (2003). Biomarkers: current perspectives and future prospects. Expert Rev. Mol. Diagn. *3*, 525–529.
Ned, R.M. et al. (2011). Fecal DNA testing for Colorectal Cancer Screening: the ColoSure™ test. PLoS Curr. *3*, RRN1220.
Nejman, D. et al. (2014). Molecular Rules Governing De Novo Methylation in Cancer. Cancer Res. *74*, 1475–1483.
Niwa, T., and Ushijima, T. (2010). Induction of Epigenetic Alterations by Chronic Inflammation and Its Significance on Carcinogenesis. In: Advances in Genetics, 41–56.
Novak, P. et al. (2008). Agglomerative Epigenetic Aberrations Are a Common Event in Human Breast Cancer. Cancer Res. *68*, 8616–8625.
Nunna, S. et al. (2014). Targeted Methylation of the Epithelial Cell Adhesion Molecule (EpCAM) Promoter to Silence Its Expression in Ovarian Cancer Cells. PLoS One *9*, e87703.
O'Hagan, H.M. et al. (2011). Oxidative Damage Targets Complexes Containing DNA Methyltransferases, SIRT1, and Polycomb Members to Promoter CpG Islands. Cancer Cell *20*, 606–619.
Ohm, J.E. et al. (2007). A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat. Genet. *39*, 237–242.
Ohtani-Fujita, N. et al. (1997). Hypermethylation in the retinoblastoma gene is associated with unilateral, sporadic retinoblastoma. Cancer Genet. Cytogenet. *98*, 43–49.
Oka, H. et al. (1998). Pituitary adenylate-cyclase-activating polypeptide (PACAP) binding sites and PACAP/vasoactive intestinal polypeptide receptor expression in human pituitary adenomas. Am. J. Pathol. *153*, 1787–1796.
Ooi, S.K.T. et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.
Ottini, L. et al. (2006). Patterns of genomic instability in gastric cancer: clinical implications and perspectives. Ann. Oncol. *17*, vii97–vii102.
Oyer, J.A. et al. (2009). Aberrant Epigenetic Silencing Is Triggered by a Transient Reduction in Gene Expression. PLoS One *4*, e4832.
Palm, T. et al. (2009). Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis. Cancer *115*, 3955–3968.
Payne, S.R. (2010). From discovery to the clinic: the novel DNA methylation biomarker mSEPT9 for the detection of colorectal cancer in blood. Epigenomics *2*, 575–585.

Perez-Pinera, P. et al. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat. Methods *10*, 973–976.
Pidsley, R. et al. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics *14*, 293.
Pidsley, R. et al. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. *17*, 208.
Poynter, J.N. et al. (2008). Molecular Characterization of MSI-H Colorectal Cancer by MLHI Promoter Methylation, Immunohistochemistry, and Mismatch Repair Germline Mutation Screening. Cancer Epidemiol. Biomarkers Prev. *17*, 3208–3215.
Reddington, J.P. et al. (2014). DNA methylation reprogramming in cancer: Does it act by re-configuring the binding landscape of Polycomb repressive complexes? BioEssays *36*, 134–140.
Reinius, L.E. et al. (2012). Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLoS One *7*, e41361.
Reubi, J.C. et al. (2000). Vasoactive intestinal peptide/pituitary adenylate cyclase-activating peptide receptor subtypes in human tumors and their tissues of origin. Cancer Res. *60*, 3105–3112.
Richter, A.M. et al. (2009). The RASSF proteins in cancer; from epigenetic silencing to functional characterization. Biochim. Biophys. Acta - Rev. Cancer *1796*, 114–128.
Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. *14*, 9–25.
Rodrigues, M.F.S.D. et al. (2016). Methylation status of homeobox genes in common human cancers. Genomics *108*, 185–193.
Sadri, R., and Hornsby, P.J. (1996). Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. Nucleic Acids Res. *24*, 5058–5059.
Saffari, A. et al. (2018). Estimation of a significance threshold for epigenome-wide association studies. Genet. Epidemiol. *42*, 20–33.
Saied, M.H. et al. (2012). Genome Wide Analysis of Acute Myeloid Leukemia Reveal Leukemia Specific Methylome and Subtype Specific Hypomethylation of Repeats. PLoS One *7*, e33213.
Saito, Y. et al. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell *9*, 435–443.
Sakakura, C. et al. (2009). Quantitative analysis of tumor-derived methylated RUNX3 sequences in the serum of gastric cancer patients. Anticancer Res. *29*, 2619–2625.
Sanjana, N.E. et al. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.
Sanz-Pamplona, R. et al. (2014). Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. Mol. Cancer *13*, 46.
Saunderson, E.A. et al. (2017). Hit-and-run epigenetic editing prevents senescence entry in primary breast cells from healthy donors. Nat. Commun. *8*, 1450.
Schlesinger, Y. et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236.
Schmidl, C. et al. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res. *19*, 1165–1174.
Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. *3*, 1101–1108.
Schwarzenbach, H. et al. (2008). Detection and Monitoring of Cell-Free DNA in Blood of Patients with Colorectal Cancer. Ann. N. Y. Acad. Sci. *1137*, 190–196.
Sexton-Oates, A. et al. (2018). Methylation profiling of paediatric pilocytic astrocytoma reveals variants specifically associated with tumour location and predictive of recurrence. Mol. Oncol. *12*, 1219–1232.
Shan, M. et al. (2016). Aberrant expression and functions of protocadherins in human malignant tumors. Tumor Biol. *37*, 12969–12981.
Shanafelt, T.D. et al. (2009). B-cell count and survival: differentiating chronic lymphocytic

leukemia from monoclonal B-cell lymphocytosis based on clinical outcome. Blood *113*, 4188–4196.
Sharma, M.K. et al. (2007). Distinct Genetic Signatures among Pilocytic Astrocytomas Relate to Their Brain Region Origin. Cancer Res. *67*, 890–900.
Shen, L. et al. (2007). Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. Biotechniques *42*, 48–58.
Sherwood, N.M. et al. (2000). The Origin and Function of the Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)/Glucagon Superfamily*. Endocr. Rev. *21*, 619–670.
Shlush, L.I. et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328–333.
Shukla, S. et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.
Siggens, L., and Ekwall, K. (2014). Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. J. Intern. Med. *276*, 201–214.
Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.
Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. *3*, 1–25.
Sohal, D.P.S. et al. (2016). Molecular characteristics of biliary tract cancer. Crit. Rev. Oncol. Hematol. *107*, 111–118.
Sproul, D. et al. (2011). Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proc. Natl. Acad. Sci. *108*, 4364–4369.
Sproul, D. et al. (2012). Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biol. *13*, R84.
Sproul, D., and Meehan, R.R. (2013). Genomic insights into cancer-associated aberrant CpG island hypermethylation. Brief. Funct. Genomics *12*, 174–190.
Stepper, P. et al. (2017). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. Nucleic Acids Res. *45*, 1703–1713.
Surana, R. et al. (2014). Secreted frizzled related proteins: Implications in cancers. Biochim. Biophys. Acta - Rev. Cancer *1845*, 53–65.
Taylor, D.P. et al. (2011). Comparison of compliance for colorectal cancer screening and surveillance by colonoscopy based on risk. Genet. Med. *13*, 737–743.
Tchoghandjian, A. et al. (2009). Pilocytic astrocytoma of the optic pathway: a tumour deriving from radial glia cells with a specific gene signature. Brain *132*, 1523–1535.
Teschendorff, A.E. et al. (2011). Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics *27*, 1496–1505.
Teschendorff, A.E. et al. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics *29*, 189–196.
Tham, C. et al. (2014). Postoperative serum methylation levels of *TAC1* and *SEPT9* are independent predictors of recurrence and survival of patients with colorectal cancer. Cancer *120*, 3131–3141.
Timp, W., and Feinberg, A.P. (2013). Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nat. Rev. Cancer *13*, 497–510.
Touleimat, N., and Tost, J. (2012). Complete pipeline for Infinium Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics *4*, 325–341.
Triche, J.T. (2014). FDb.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. R Packag. Version 2.2.0.
Triche, T.J. et al. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. *41*, e90–e90.
Turner, N.C. et al. (2007). BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene *26*, 2126–2132.

Urich, M.A. et al. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat. Protoc. *10*, 475–483.
Vaudry, D. et al. (2000). Pituitary adenylate cyclase-activating polypeptide and its receptors: from structure to functions. Pharmacol. Rev. *52*, 269–324.
Viré, E. et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.
Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.
Vojta, A. et al. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. *44*, 5615–5628.
Waha, A. et al. (2005). Epigenetic silencing of the protocadherin family member PCDH-gamma-A11 in astrocytomas. Neoplasia *7*, 193–199.
Wang, K.-H. et al. (2015). Global methylation silencing of clustered proto-cadherin genes in cervical cancer: serving as diagnostic markers comparable to HPV. Cancer Med. *4*, 43–55.
Wang, Y. et al. (2018). Advance in plasma SEPT9 gene methylation assay for colorectal cancer early detection. World J. Gastrointest. Oncol. *10*, 15–22.
Wang, Z. et al. (2016). SPOCK1 promotes tumor growth and metastasis in human prostate cancer. Drug Des. Devel. Ther. *Volume 10*, 2311–2321.
Warnecke, P.M. et al. (2002). Identification and resolution of artifacts in bisulfite sequencing. Methods *27*, 101–107.
Waschek, J. (2013). VIP and PACAP: neuropeptide modulators of CNS inflammation, injury, and repair. Br. J. Pharmacol. *169*, 512–523.
Wasenang, W. et al. (2019). Serum cell-free DNA methylation of OPCML and HOXD9 as a biomarker that may aid in differential diagnosis between cholangiocarcinoma and other biliary diseases. Clin. Epigenetics *11*, 39.
Weisenberger, D.J. et al. (2008). DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res. *36*, 4689–4698.
Widschwendter, M. et al. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.
Wilhelm-Benartzi, C.S. et al. (2013). Review of processing and analysis methods for DNA methylation array data. Br. J. Cancer *109*, 1394–1402.
Wojdacz, T.K., and Lotte Hansen, L. (2006). Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. Biotechniques *41*, 274–278.
Wong Doo, N. et al. (2016). Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics *8*, 55–66.
Wong, E.M. et al. (2011). Constitutional Methylation of the BRCA1 Promoter Is Specifically Associated with BRCA1 Mutation-Associated Pathology in Early-Onset Breast Cancer. Cancer Prev. Res. *4*, 23–33.
Wong, I.H.N. et al. (2003). Quantitative analysis of tumor-derived methylated p16INK4a sequences in plasma, serum, and blood cells of hepatocellular carcinoma patients. Clin. Cancer Res. *9*, 1047–1052.
Wong, K.-K. et al. (2005). Expression analysis of juvenile pilocytic astrocytomas by oligonucleotide microarray reveals two potential subgroups. Cancer Res. *65*, 76–84.
Worm, J. et al. (2001). In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin. Chem. *47*, 1183–1189.
Wright, M.L. et al. (2016). Establishing an analytic pipeline for genome-wide DNA methylation. Clin. Epigenetics *8*.
Wu, Z. et al. (2008). Microarray-based Ms-SNuPE: Near-quantitative analysis for a high-throughput DNA methylation. Biosens. Bioelectron. *23*, 1333–1339.
Xie, J. et al. (2018). Solute carrier transporters: potential targets for digestive system neoplasms. Cancer Manag. Res. *10*, 153.
Xu, X. et al. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov. *2*, 16009.

Yang, A.S. et al. (2006). DNA Methylation Changes after 5-Aza-2′-Deoxycytidine Therapy in Patients with Leukemia. Cancer Res. *66*, 5495–5503.
Yu, M. et al. (2015). MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics *10*, 803–809.
Zakrzewski, K. et al. (2015). Transcriptional profiles of pilocytic astrocytoma are related to their three different locations, but not to radiological tumor features. BMC Cancer *15*, 778.
Zhang, J. et al. (2017). SPOCK1 is up-regulated and promotes tumor growth via the PI3K/AKT signaling pathway in colorectal cancer. Biochem. Biophys. Res. Commun. *482*, 870–876.

8. ACKNOWLEDGEMENTS

The biggest thanks to Professor Patrizia Zavattari, my PhD tutor in these three years and my mentor. Thank you for having welcomed me into this small but very close-knit research group when, three years ago, I wrote to you after having passionately read your exciting research projects. Thank you for showing me what it means to do research in a rigorous but passionate way. I was extremely struck by your enthusiasm for any scientific result, your brilliant interpretation of the data and your ingenious problem solutions. Thank you for all our scientific talks and your teachings that really go beyond this project, giving me the opportunity to learn how the world of research works. I will never thank you enough for this.

Another big thanks to Lory, our laboratory technician, who taught me most of the methods applied in this thesis. Thank you for your patience and above all for your enormous generosity and availability.

Thanks to Florencia, my lab-mate, for all the help and big support in this project. I am so happy that you joined our lab, contributing to create a friendly and collaborative environment.

I would also like to thank all the supervisors in my abroad experiences for having accepted to host me in their research labs. All these experiences have been fantastic, and I found a stimulating and friendly environment in each of your research groups.

Thanks to Dr. Duncan Sproul, my supervisor at IGMM in Edinburgh. It has been a pleasure to compare some of our results with yours and to be introduced in one of your projects. Thanks to all your group for their help and especially to Francesca Taglini, my supervisor in all the experiments.

Thanks to Dr. Daniel Worthley and Dr. Susan Woods, it has been an honor to stay in your lab at SAHMRI in Adelaide. Thanks for having shown so much enthusiasm in our idea and having supported us in all the experimental phases of the methylation editing. A special thank you to Susan, my supervisor in the lab, for your immense contribute, your kindness and your precious advices. Thanks to all the members of your research group, especially to TongTong, Laura and Jo for their important help in the lab.

Thanks to Dr. Sergio Alonso, my supervisor at IGTP in Barcelona. Your contribution for the methylation analyses in thesis has been fundamental. I made treasure of all your suggestions and your great experience. I really found a friendly atmosphere in your research lab for which I thank everyone but especially Bea and Maria.

I also thank all collaborators of Patrizia Zavattari for their contribution in the different cancer studies and of course all patients who participate in these studies.

Finally, an immense thanks to my family and Stefano for having always believed in me, for having support all my choices and for their closeness even when I was distant. All of this would not be possible without all of you.