World Journal of Gastrointestinal Oncology

Submit a Manuscript: http://www.wjgnet.com/esps/ *World J Gastrointest Oncol* 2017 March 15; 9(3): 105-120

DOI: 10.4251/wjgo.v9.i3.105 ISSN 1948-5204 (online)

FRONTIER

CpG island methylator phenotype in adenocarcinomas from the digestive tract: Methods, conclusions, and controversies

Francisco Sánchez-Vega, Valer Gotea, Yun-Ching Chen, Laura Elnitski

Francisco Sánchez-Vega, Valer Gotea, Yun-Ching Chen, Laura Elnitski, Genomic Functional Analysis Section, National Human Genome Research Institute, National Institutes of Health, Rockville, MD 20852, United States

Author contributions: Sánchez-Vega F, Gotea V, Chen YC and Elnitski L have all read and approved the final manuscript; Elnitski L developed the idea and supervised the research; Sánchez-Vega F, Gotea V and Chen YC generated and analyzed data; and all authors contributed to the writing and editing of the final manuscript; Sánchez-Vega F and Gotea V contributed equally to this work.

Conflict-of-interest statement: None.

Manuscript source: Invited manuscript

Correspondence to: Laura Elnitski, PhD, Senior Investigator, Head, Genomic Functional Analysis Section, National Human Genome Research Institute, National Institutes of Health, 5625 Fishers Lane, Rockville, MD 20852, United States. elnitski@mail.nih.gov
Telephone: +1-301-4510265
Fax: +1-301-4356170

Received: June 2, 2016
Peer-review started: June 6, 2016
First decision: September 2, 2016
Revised: November 3, 2016
Accepted: January 2, 2017
Article in press: January 3, 2017
Published online: March 15, 2017

Abstract

Over the last two decades, cancer-related alterations in DNA methylation that regulate transcription have been reported for a variety of tumors of the gastrointestinal tract. Due to its relevance for translational research, great emphasis has been placed on the analysis and molecular characterization of the CpG island methylator phenotype (CIMP), defined as widespread hypermethylation of CpG islands in clinically distinct subsets of cancer patients. Here, we present an overview of previous work in this field and also explore some open questions using cross-platform data for esophageal, gastric, and colorectal adenocarcinomas from The Cancer Genome Atlas. We provide a data-driven, pan-gastrointestinal stratification of individual samples based on CIMP status and we investigate correlations with oncogenic alterations, including somatic mutations and epigenetic silencing of tumor suppressor genes. Besides known events in CIMP such as *BRAF V600E* mutation, CDKN2A silencing or MLH1 inactivation, we discuss the potential role of emerging actors such as Wnt pathway deregulation through truncating mutations in RNF43 and epigenetic silencing of WIF1. Our results highlight the existence of molecular similarities that are superimposed over a larger backbone of tissue-specific features and can be exploited to reduce heterogeneity of response in clinical trials.

Key words: CpG island methylator phenotype; CpG island; Promoter; DNA methylation; Hypermethylation; Gastrointestinal cancer

Core tip: Awareness of the CpG island methylator phenotype (CIMP) is growing for all adenocarcinomas. Here, we summarize previous work on the topic and discuss unanswered questions regarding commonalities and differences of CIMP tumors from esophageal, gastric, and

colorectal adenocarcinomas, where data has been made available from the Cancer Genome Atlas. Our analysis includes a review of our pan-cancer method to stratify tumors based on CIMP and addresses the most frequent mutations found in those samples. We include new data implicating truncating mutations in RNF43 and silencing of WIF1I. We also describe in detail the methylation of CpG sites within the MLH1 promoter across these tumor types.

Sánchez-Vega F, Gotea V, Chen YC, Elnitski L. CpG island methylator phenotype in adenocarcinomas from the digestive tract: Methods, conclusions, and controversies. *World J Gastrointest Oncol* 2017; 9(3): 105-120 Available from: URL: http://www.wjgnet.com/1948-5204/full/v9/i3/105.htm DOI: http://dx.doi.org/10.4251/wjgo.v9.i3.105

INTRODUCTION

Aberrant patterns of DNA methylation occur in human cancers[1-3], with the most notable being a widespread and pronounced gain of methylation at CpG islands in tumor cells[4]. A prominent increase in global levels of CpG island methylation observed across multiple samples was first reported in a subset of patients with colorectal cancer (CRC) and it is now a clinically recognized characteristic of many types of tumor, referred to as the CpG island methylator phenotype (CIMP)[5]. In this commentary, we discuss the classification and functional ramifications of CIMP across four types of gastrointestinal adenocarcinomas (esophageal, gastric, colon and rectal), using data from The Cancer Genome Atlas (TCGA) to address lingering questions and identify novel areas of inquiry to spur future investigation. Finally, we explore CIMP's potential application to cancer diagnostics and subtyping, while emphasizing that much remains unknown regarding the molecular mechanisms of tumor-associated DNA methylation, including CIMP generation and maintenance.

CpG islands play a crucial biological role in development and disease by acting as transcriptional regulatory elements in the genome and controlling the expression of ubiquitously expressed genes. Approximately 50% of all CpG islands are located within promoter regions[6], and approximately 70% of all annotated promoters are associated with a CpG island[7]. Hypermethylation of CpG dinucleotides within these regions results in the establishment or reinforcement of repressive chromatin and the steric occlusion of transcription factor binding[4,8], reducing gene expression. When promoters of tumor suppressor genes are methylated, repression can represent a critical "hit", in the terminology of the double-hit theory of gene inactivation, conferring a selective advantage to affected cancer cells[9]. For example, the heterozygous silencing of *BRCA1 via* DNA methylation plays a critical role in breast cancer oncogenesis and tumor proliferation[10]. Other well-known examples of silencing involve *MLH1* in CRCs[5,11] and *MGMT* silencing in gliomas[12]. In the case of *MLH1*, methylation-derived silencing inhibits DNA repair[11,13,14], which leads to microsatellite instability (MSI) and cascades into many other downstream functional consequences.

Researchers have identified reproducible, tissue-specific patterns of CpG island promoter hypermethylation in various types of tumors[15]. The specificity of hypermethylation appears to result from the precise targeting of CpG islands by polycomb repressors[16], resulting in the preferential deposition of DNA methyl groups during oncogenesis[17-19]. Because these patterns are frequently occurring in cancer patients, they have been used as novel, clinically relevant molecular markers for cancer diagnosis and prognosis[20]. To cite two examples, hypermethylation of the *GSTP1* promoter in more than 90% of prostate adenocarcinomas has been used to improve diagnosis of this disease[21], whereas hypermethylation of SET pseudogene 9 allows researchers to differentiate among different stages of CRC[22].

The demonstration that tumors exhibiting CIMP represent a distinct clinical subtype of CRC[5] provided the first evidence that, by subdividing cancers into methylation subclasses, clinicians could potentially refine treatment outcomes. Numerous studies have since demonstrated the presence of CIMP in additional cancer types[23-25]. However, little overlap has been detected among these CIMP incarnations, indicating the tissue-specific nature of the effect. Current models indicate that tumorigenesis affects DNA methylation at CpG islands where repressive H3K27me3 modifications are already present[26], providing a more permanent layer of suppression in differentiated cells and explaining the origin of tissue-specific patterns. According to such models, aberrant DNA methylation is not a stochastic outcome, but a targeted, albeit abnormal, process. In this light, it becomes reasonable to speculate that distinct tumor types could use similar cellular pathways to target their own characteristic CpG islands for DNA methylation. Mechanistic congruity among different tumor types would allow us to understand multi-cancer and pan-cancer processes from a unified molecular perspective. However, testing this hypothesis requires us to use consistent methods to assess DNA methylation across tumor types and to analyze large numbers of samples to provide statistical power. In the rest of this article, we provide examples of such analyses.

EVALUATING CIMP: FROM GENE PANELS TO GENOME-WIDE METHYLATION PROFILES

A quick overview of important milestones in the study of CIMP within the context of gastrointestinal cancers is provided in Table 1. Given the diversity of methods for assessing DNA methylation, profiling has been performed over a wide range of technical depths and breadths. Initially, the implementation challenges of wide-scale methylation profiling limited the scope of CIMP evaluation. Researchers working on CRC employed panels of genes

Table 1 Overview of previous studies of CpG island methylator phenotype in tumors from the gastrointestinal track

Year	Event	Ref.
1999	CIMP is first reported in a set of CRC patients	[5]
2004	Nature Reviews paper discussing CIMP in a variety of tumors besides CRC	[23]
2006	Refined molecular subtyping includes CIMP-low and CIMP-0 categories in CRC, with associations to *KRAS* mutations	[47]
	New insights are gained about the interplay between BRAF V600E mutations, MSI status, *MLH1* promoter methylation and CIMP in CRC	[14]
2006-2012	High throughput DNA methylation arrays become widely available, enabling the use of larger gene panels for CIMP characterization	[45,46]
2014	TCGA marker paper on gastric cancer highlights the biological relevance of CIMP for molecular subtyping, exploring associations with EBV infection	[64]
	A better mechanistic understanding of CIMP in CRC is gained through elucidation of the role of MAFG in the context of MLH1 silencing and *BRAF V600E* mutations	[76]
2015	Pan-cancer stratification of solid tumors reveals similarities in CIMP across a wide variety of cancer types	[51]

CIMP: CpG island methylator phenotype; CRC: Colorectal cancer; MSI: Microsatellite instability; TCGA: The Cancer Genome Atlas.

using a low-throughput approach, such as methylation-specific PCR. These panels varied in size from four[27] to several dozen genes[28], and invariably included subsets of the sequences originally employed by Toyota *et al*[5,29]. Although other CIMP-tumor characterizations have emerged, CRC remains the most heavily investigated tumor type with respect to CIMP subtypes. A variety of gene panels are still in use[30], some of which include *MLH1*[31-35] due to its aforementioned connections to MSI[36,37].

Following an increase in the scope of methylation studies, individual CpG sites started being used to detect aberrant methylation across multiple cancer types. For example, *CDKN2A* profiling has been used in at least 10 cancer types[24], and *MLH1* profiling has been extended to pancreatic cancer[38], leukemia[39], ovarian cancer[40], endometrial cancer[41], gastric cancer[42], and lung cancer[43]. Although these sites are consistently differentially methylated in multiple tumor types, none of them are informative enough to classify samples as CIMP in an independent manner.

The limitations of these early ascertainment methods and lack of extensive overlaps across tumor types, coupled with a variable range of methylation at any given CpG site, fueled a debate over the relevance of CIMP in cancer[44]. The advent of array-based platforms for measuring DNA methylation, such as the Illumina Infinium HumanMethylation27 and HumanMethylation450 arrays[45], helped end this debate[46]. Recent genome-wide experiments using high-throughput data have not only corroborated the biological relevance of CIMP to CRC diagnostics and survival rates, but have led to finer subdivisions of methylation levels, such as CIMP-low and CIMP-zero[47-49]. These classifications better reflect global patterns of hypermethylation, which often fail to fit within "high" or "low" classes in colorectal and other cancers. For example, our early studies of gynecological tumor epigenomes showed a finely increasing signal of CpG island hypermethylation among ovarian and endometrial tumors, rather than a binary methylation signature[50]. This signature represented an intermediate ranking between the fully methylated and unmethylated states, where the CIMP intermediate group corresponded to the serous subtype with TP53 mutations. This observation weighed heavily into our recently demonstrated method to stratify DNA methylation patterns of most cancer types collected by TCGA at the time for CIMP classification. Categories that we defined include CIMP+, CIMP-intermediate (CIMPi), and CIMP-[51]. Such broad-scale analyses provide a means of subtyping individual tumor collections into relatively homogenous methylation subgroups, notwithstanding the fact that each subgroup can contain a gradient in methylation levels. The absence of a highly dichotomous methylation pattern suggests that a complex interplay of factors determines CIMP status, including tumor heterogeneity and clonality[52], multiple somatic/germline mutations[53], copy number variation, and mutation heterozygosity[54].

Within the ongoing effort to better the understanding of cancer biology, we argue that evaluating methylation on an epigenome-wide scale should be favored over the analysis of a few, select loci. For example, large-scale analyses have revealed the now widely recognized phenomenon that DNA methylation occurs at genes with a role in early development and morphogenesis, leading to the discovery that polycomb binding is a precursor to aberrant DNA methylation[25,55,56]. Also, a number of recent studies have highlighted important similarities in terms of somatic alterations and epigenetic patterns across cancers of different organs and tissues[51,57-59]. This type of multi-cancer or pan-cancer approach benefits from increased statistical power compared with smaller studies of individual cancer types, which, however, are better suited to capturing tissue-specific features. Researchers can harness the advantages of both approaches by studying related cancer types that occur in tissues derived from common cell lineages. A good example of this approach is provided by previous multi-cancer analyses of tumors of the gastrointestinal tract[60].

GENOMIC CHARACTERISTICS ASSOCIATED WITH CIMP IN GASTROINTESTINAL TUMORS

TCGA has used patterns of mutation to classify colo-

Table 2 Gastrointestinal adenocarcinoma types, sample sizes, probe set sizes, and CpG island methylator phenotype status

Cancer type	Differentially methylated probes	Control samples	Tumor samples	CIMP-	CIMPi	CIMP+
EAC	6717	11	87	26	31	30
STAD	1110	2	260	109	95	56
COAD	2656	38	274	96	92	86
READ	1255	7	96	31	39	26

CIMP: CpG island methylator phenotype; CIMPi: CIMP intermediate; COAD: Colon adenocarcinoma; EAC: Esophageal adenocarcinoma; READ: Rectal adenocarcinoma; STAD: Stomach adenocarcinoma.

rectal sample genomes into two large groups, non-hypermutated and hypermutated[61]. Colon and rectal tumor samples in the former class largely possess CIMP-low phenotypes and have almost indistinguishable molecular signatures in terms of copy number variation, mRNAs, and miRNAs. By contrast, hypermutated samples are predominantly tumors of the colon. Roughly three-quarters exhibit CIMP-high status, as well as *MLH1* silencing and MSI, whereas the other quarter are characterized by mutations in other mismatch repair genes such as *MLH3* and mutations in *POLE*. The contrast between samples exhibiting high chromosomal instability (CIN) and samples exhibiting large mutational load is not unique to CRCs, as it has been described in other cancer types, including the endometrioid *vs* serous subtypes of both ovarian and endometrial cancers[50,62]. Consistent with these observations, the importance of CIMP as a mutually exclusive alternative to CIN has been underscored in describing dysfunctional events in tumor genomes[63]. As we have reported previously[51], the MSI *vs* CIN duality largely corresponds to a CIMP+ *vs* CIMP- dichotomy. This can be extrapolated to a pan-cancer dichotomization of tumors into a "mutator" class, characterized by a large number of somatic mutations, closer to CIMP+, and a "copy number" class, characterized by an abundance of copy-number alterations but lacking excessive mutations, closer to CIMP-. This duality has been previously referred to as the cancer genome hyperbola[57].

Even if it is conceptually helpful, the simple high-level dichotomy assessed by mutations or copy number alterations fails to adequately represent all of the mechanisms of diversity in gastric tumors. For example, a comprehensive molecular study carried out by TCGA subdivided gastric tumors into four distinct subgroups[64]. Two distinct CIMP-high tumor subgroups were identified: One associated with Epstein-Barr virus (EBV), and one associated with MSI. Among 10 different cancer types analyzed by TCGA, the EBV-CIMP subgroup exhibited the highest frequency of DNA hypermethylation at gene promoters, highlighting the interplay, causative or correlational, between environmental exposures such as viral infection and DNA methylation of the tumor genome. Studies involving other infectious agents also suggest potentially relevant associations between presence of pathogens, gastric cancer prognosis and CIMP status. For example, in patients infected with *Helicobacter pylori*, CIMP+ tumors exhibit higher rates of recurrence and metastasis than CIMP- tumors[65].

Of the four types of gastrointestinal cancer examined in the present article, esophageal cancers have been the least thoroughly studied with regards to CIMP stratification. However, CIMP and its associated driver mutations have been investigated in the context of some esophageal tumor subtypes[66]. In particular, subsets of tumors exhibiting high levels of methylation have been reported in both esophageal adenocarcinoma and Barrett's esophagus, a precursor lesion to esophageal adenocarcinoma[67]. Moreover, the overall amounts of DNA hypermethylation in Barrett's esophagus predict progression to esophageal adenocarcinoma[68,69]. Genes such as *CDKN2A*, *APC*, *CDH1*, *TAC1* and *MGMT* have been reported to exhibit increased methylation in esophageal adenocarcinomas, esophageal squamous cell carcinomas and Barrett's esophagus when compared to normal esophageal DNA[70]. By contrast, *MLH1* promoter methylation has been reported in esophageal squamous cell carcinomas, but not adenocarcinomas[70,71], confirming differences in methylation profiles between esophageal subtypes.

ANALYSES OF CIMP IN GASTROINTESTINAL CANCERS

Here, we investigated CIMP in four types of gastrointestinal adenocarcinoma (GIAD) samples provided by TCGA: Esophageal adenocarcinoma (EAC), which is a subset of esophageal carcinoma (or ESCA, using the TCGA nomenclature); stomach adenocarcinoma (STAD); colon adenocarcinoma (COAD); and rectal adenocarcinoma (READ). Using a previously described approach[51], we assessed mean methylation levels in tumor and healthy adjacent tissues and ranked samples using unsupervised clustering. Specifically, we measured DNA methylation levels at a set of informative probes (*i.e.*, sets of loci that were differentially methylated between tumor and normal samples at statistically significant levels) using statistical selection criteria applied independently for each tumor collection (Table 2; see research). We then evaluated CIMP status by classifying samples according to average methylation levels across the set of informative probes. This type of CIMP stratification, in which samples with similar methylation intensity levels are grouped together, reduces heterogeneity within the full tumor collection and facilitates the identification of functional somatic alterations that may play a shared role across different cancer types (and subtypes).

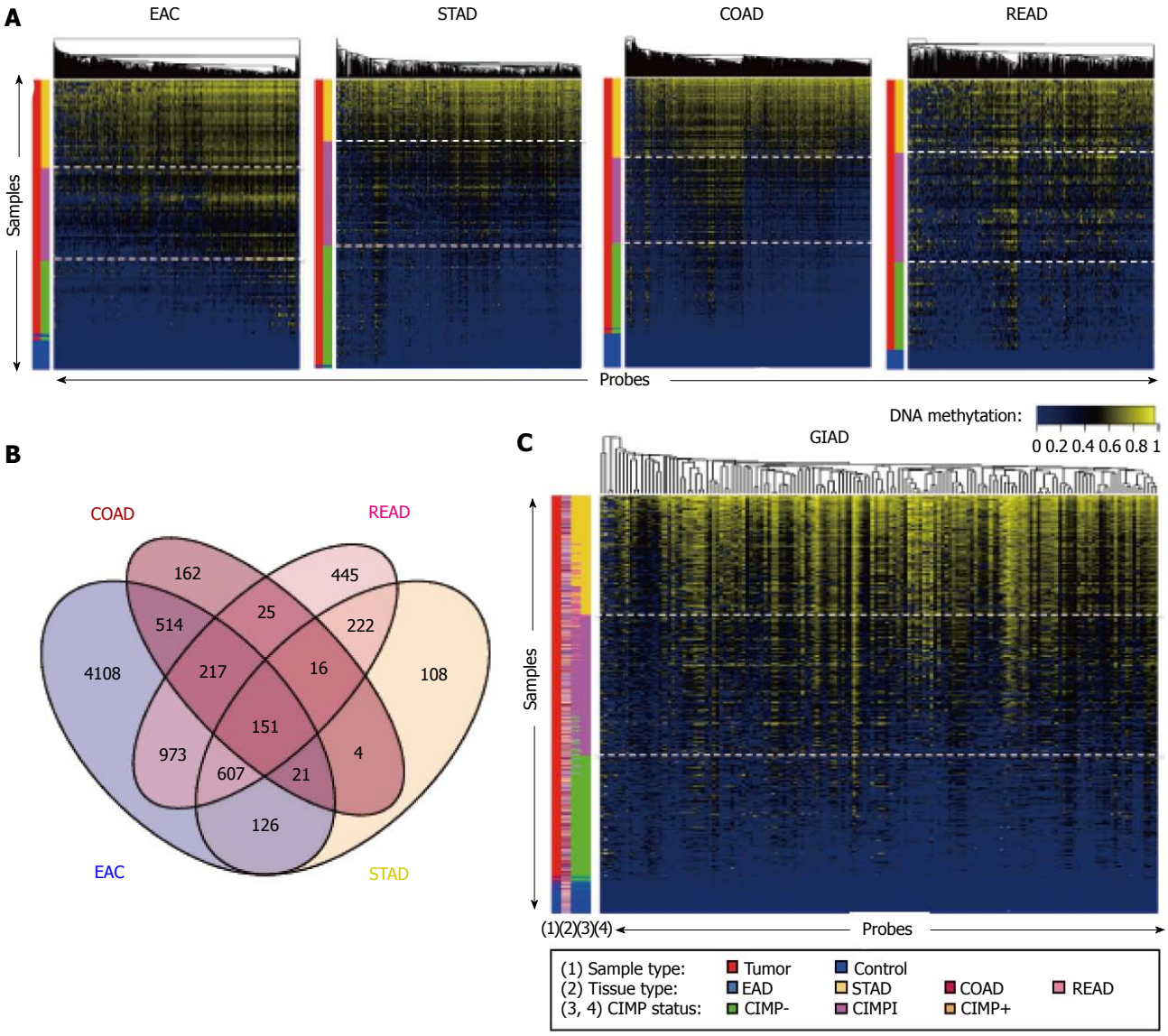

Figure 1 CpG island methylator phenotype analysis of gastrointestinal adenocarcinoma samples from the Cancer Genome Atlas. A: CIMP analysis for EAC, STAD, COAD, and READ samples. Each row represents a sample, and each column represents a probe. The two-color side bar shows tumor samples (red) and normal samples (blue). The four-color side bar indicates CIMP status: CIMP+ (gold), CIMP intermediate (CIMPi; magenta), CIMP- (green), and normal (blue). Samples were ranked vertically according to mean methylation levels across all of the probes shown in the heat map; B: Venn diagram showing the intersection of the selected, informative probes with regard to CIMP status across the four cancer types; C: CIMP analysis for the combined GIAD data set, in which samples from the four individual tumor types were pooled together. The side bars show (1) sample type (tumor *vs* adjacent tissue); (2) cancer type (EAC, STAD, COAD, or READ); (3) CIMP status based on the individual analyses shown in panel A; and (4) CIMP status based on the analysis using the pooled data set. CIMP: CpG island methylator phenotype; EAC: Esophageal adenocarcinoma; STAD: Stomach adenocarcinoma; COAD: Colon adenocarcinoma; READ: Rectal adenocarcinoma; GIAD: Gastrointestinal adenocarcinoma.

After clustering based on average methylation levels across the probes, samples were categorized into three distinct groups: CIMP+, CIMPi, and CIMP-. CIMP- samples had CpG island methylation profiles that were closer to those observed in normal samples, whereas CIMP+ samples showed a reproducible pattern of DNA hypermethylation with respect to non-cancer controls (Figure 1A). CIMPi samples displayed methylation levels that fell between the CIMP+ and CIMP- groups. In subsequent analyses, we compared CIMP- and CIMP+ samples and excluded the intermediate group, to avoid borderline cases and to guarantee that the tumors being compared were sufficiently different from a molecular point of view.

In a previous study, we showed that our CIMP+ and CIMP- assignments largely coincided with independent assignments by the TCGA for an overlapping sample set of CRC tumors[51]. Here, we compared our CIMP classification with the four molecular subtypes defined by TCGA for gastric tumors: (1) EBV+; (2) MSI; (3) genomically stable (GS); and (4) CIN[64] (Table 3). We observed a significant association between CIMP+ status and the EBV+ and MSI subtypes, in agreement with the extreme CIMP reported for these subtypes by TCGA. Highlighting the previously mentioned incompatibility of CIMP and CIN, CIN samples were significantly skewed

Table 3 Comparison between our CpG island methylator phenotype classification of stomach adenocarcinomas and the four subtypes defined by The Cancer Genome Atlas Research Network[1]

	CIN	EBV+	GS	MSI	Total
CIMP+	[2]12 (26.6)	[3]13 (5.4)	[2]2 (11.3)	[3]27 (10.7)	-54
CIMPi	43 (45.75)	12 (9.4)	20 (19.5)	18 (18.4)	93
CIMP-	[3]67 (49.68)	[2]0 (10.2)	[3]30 (21.2)	[2]4 (20.0)	101
Total	122	25	52	49	248

[1]Numbers outside parentheses correspond to actual sample counts, whereas numbers in parentheses show expected counts under the null model of independent classification; [2]Indicates under-represented counts; [3]Indicates cells with significantly over-represented counts ($P < 0.05$, based on Fisher's exact test). CIMP: CpG island methylator phenotype; CIMPi: CIMP intermediate; CIN: Chromosomal instability; EBV+: Epstein-Barr virus positive; GS: Genomically stable; MSI: Microsatellite instability.

toward CIMP- status. However, other samples also occupied the CIMP- category, including GS samples, which displayed few alterations in DNA methylation and lacked MSI.

In addition to evaluating CIMP in each of the four cancer types independently, we combined all of the data into a single set. Here, the intersection of the loci selected in the four previous, independent analyses was considered informative (n = 151, Figure 1B). In this new classification of samples (Figure 1C), CIMP labels remained largely consistent with the previously assigned labels. Importantly, when samples in the pooled data set were ranked according to their average level of DNA methylation across the set of selected probes, they tended to cluster by CIMP status rather than tissue of origin. This novel finding implies commonalities in the underlying generation of aberrant methylation across cancer types.

CIMP AND *MLH1* PROMOTER HYPERMETHYLATION

Early studies of CIMP established that the *MLH1* promoter is consistently hypermethylated in CRC[5]. This observation has since been extended to other cancer types[72], and its importance is highlighted by the inclusion of *MLH1* in many gene panels used to evaluate CIMP. The strong association between CIMP and *MLH1* promoter hypermethylation continues to be reinforced by recent studies with large sample sizes, such as a pan-cancer analysis performed by our group[51] using a catalog of 479 somatic functional events (Ciriello *et al*[57], 2013). In this previous work, we investigated a cohort of 3299 samples that spanned 9 different cancer types and found that *MLH1* promoter silencing was the single genomic functional event that displayed the strongest statistical association with CIMP.

Since promoter hypermethylation is usually associated with gene silencing[4,8], one could compare the effects of *MLH1* promoter hypermethylation and disabling gene mutations, addressing parallels with loss-of-function. Indeed, *MLH1* promoter silencing replicates the phenotype of *MLH1* loss-of-function mutations in hereditary nonpolyposis colon cancer, which displays dinucleotide repeat instability[73]. Moreover, research in cell lines demonstrates that reversing *MLH1* promoter hypermethylation increases transcription of the gene and restores mismatch repair capacity[11,74]. It is therefore tempting to hypothesize that *MLH1* promoter hypermethylation, which is strongly associated with CIMP and displays the functional hallmarks of a loss-of-function mutation, is a causal event in the onset of CIMP. However, previous studies, including our own, have shown that CIMP can be observed in the absence of *MLH1* promoter hypermethylation or mutation[51,61,75], implying either a relationship that is correlational but not causal, or multiple mechanisms underlying CIMP development.

Only recently has experimental evidence emerged to help elucidate the role of *MLH1* promoter hypermethylation in CIMP. In CRC, Fang *et al*[76] have shown that the common BRAF V600E mutation leads to elevated levels of the protein MAFG. In turn, MAFG binds to the promoter of *MLH1* and other genes, where it recruits a heterodimeric partner, BACH1; a chromatin remodeling factor, CHD8; and a DNA methyltransferase, DNMT3B - ultimately resulting in increased methylation at the target sites. These results suggest that mutations such as BRAF V600E orchestrate aberrant methylation patterns; therefore, *MLH1* promoter hypermethylation might be thought of as part of the CIMP onset process rather than an initiating event.

Many interesting genes may fit into a model in which, following the onset of somatic mutations, a cascade of downstream methylation events occurs. For instance, *CDKN2A* promoter hypermethylation is also linked to *BRAF* mutations, through increased expression of the DNA methyltransferase DMNT3B[77]. Similarly, hypermethylation and silencing of the *INK4-ARF* locus (also known as *CDKN2A* and *CDKN2B*) occurs through KRAS activation of ZNF304, which recruits the DNA methyltransferase, DNMT1[78].

MLH1 PROMOTER METHYLATION IN GASTROINTESTINAL TUMOR DATA FROM TCGA

We analyzed GIAD data supplied by TCGA to learn more about the relationship between *MLH1* promoter methylation and CIMP. First, we identified 41 probes from the Illumina Infinium HumanMethylation450 array

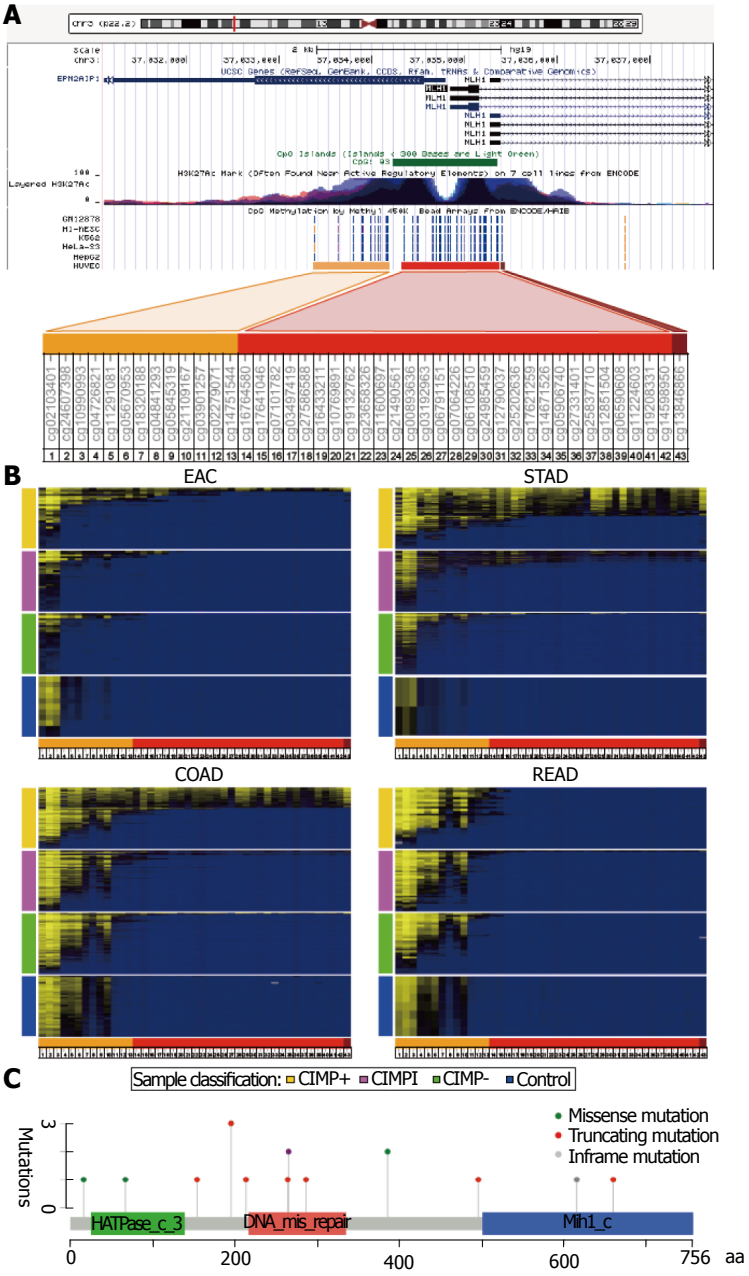

Figure 2 ***MLH1*** **promoter methylation and somatic mutations.** A: Diagram of the *MLH1* promoter region and the adjacent gene, *EPM2AIP1*, obtained from the UCSC Human Genome Browser. The probes in this region from the Illumina Infinium HumanMethylation450 array are shown with color bars relative to the CpG island present at this locus: The north shore (orange), the CpG island (red), and the south shore (dark red); B: Heat maps of GIADs showing DNA methylation status across a large genomic region that encompasses the *MLH1* promoter. Probes are displayed from left to right, and samples are ordered from top to bottom by average methylation across the region. Color side bars indicate CIMP status: CIMP+ (gold), CIMP intermediate (CIMPi; magenta), CIMP- (green), and control tissue (blue); C: Distribution of 16 somatic mutations in the coding region of *MLH1*. Color boxes correspond to different functional domains, as specified in the cBioPortal at MSKCC[99], and the vertical axis shows the number of mutations affecting a given codon. GIADs: Gastrointestinal adenocarcinomas; CIMP: CpG island methylator phenotype.

located in the extended *MLH1* promoter, operationally defined as 1.5 kb upstream and 500 bp downstream of the transcription start site (TSS) (Figure 2A). We then examined methylation levels for each cancer type, comparing CIMP+ to CIMP- samples, and found that COAD and STAD tumors displayed the strongest differences (Figure 2B). We next looked at the positions of differential methylation. A set of 24 probe sites were differentially methylated between the CIMP+ and CIMP- groups in COAD tumors, and an extended region of 38 probe sites were differentially methylated in STAD tumors (after Bonferroni correction for 41 positions). By contrast, we found no significantly differentially methylated positions in READ samples, and only three in EAC samples. The strongest association between *MLH1* promoter hypermethylation and CIMP occurred in COAD tumors (Figure 2B): One-third (20/60) of CIMP+ samples in COAD exhibited *MLH1* promoter hypermethylation, in contrast to less than 3% of CIMP- samples (2/71) ($P = 2.1 \times 10^{-6}$, Fisher's exact test). At the other extreme, no READ CIMP+ samples exhibited hypermethylation of the *MLH1* promoter.

We also examined the association between mutations that disable *MLH1* and the presence of CIMP. First, we collected all somatic mutations mapped to *MLH1* in samples whose CIMP status had been determined (Table 4 and Figure 2C). The most detrimental somatic alterations in *MLH1* are frameshift mutations, which render large fractions of the protein product nonfunctional. We observed frameshift mutations in all three CIMP classes (CIMP+, CIMPi and CIMP-), without a significant bias toward CIMP+ samples. In fact, several truncating mutations within the DNA mismatch repair functional domain of the protein occurred in CIMPi and CIMP- samples. These data suggest that loss of function alterations at *MLH1* might not be sufficient for the onset of CIMP.

Table 4 Somatic mutations found in the tumor suppressor gene *MLH1* in gastrointestinal adenocarcinoma samples

Sample	Cancer type	CIMP class	Mutation	Mutation type	AA pos.	Aff. AAs	VEST score[1]
TCGA-A6-6780-01	COAD	CIMP+	chr3:37038192.G>A	Missense substitution	67	1	0.994
TCGA-CA-6719-01	COAD	CIMP+	chr3:37067243.G>A	Missense substitution	385	1	0.701
TCGA-CM-6171-01	COAD	CIMP+	chr3:37070349.C>-	Frameshift deletion	495	262	-
TCGA-EI-6917-01	READ	CIMP+	chr3:37058999.C>T	Missense substitution	265	1	0.981
TCGA-BR-6452-01	STAD	CIMP+	chr3:37107356.A>G	3' UTR	-	-	-
TCGA-FP-A4BE-01	STAD	CIMP+	chr3:37090086.C>T	Nonsense substitution	659	98	-
TCGA-A6-6138-01	COAD	CIMPi	chr3:37035084.G>A	Missense substitution	16	1	0.901
TCGA-AD-6889-01	COAD	CIMPi	chr3:37053348.->A	Frameshift insertion	195	562	-
TCGA-AZ-6601-01	COAD	CIMPi	chr3:37067242.C>T	Missense substitution	385	1	0.952
TCGA-CM-4746-01	COAD	CIMPi	chr3:37059062.A>-	Frameshift deletion	286	471	-
TCGA-EI-6884-01	READ	CIMPi	chr3:37058995.A>G	Acceptor splice site	264	493	-
TCGA-BR-6802-01	STAD	CIMPi	chr3:37053348.A>-	Frameshift deletion	195	562	-
TCGA-F1-6874-01	STAD	CIMPi	chr3:37050312. ACCTTTTTTACAACATAGCC>-	Frameshift deletion	154	603	-
TCGA-A6-6781-01	COAD	CIMP-	chr3:37053348.A>-	Frameshift deletion	195	562	-
TCGA-CM-6674-01	COAD	CIMP-	chr3:37058999.C>-	Frameshift deletion	265	492	-
TCGA-F4-6856-01	COAD	CIMP-	chr3:37089123.GAA>-	In-frame deletion	615	1	-
TCGA-R6-A6KZ-01	EAC	CIMP-	chr3:37034874.T>C	5′ UTR	-	-	-
TCGA-CG-5723-01	STAD	CIMP-	chr3:37053550.G>-	Frameshift deletion	213	544	-

[1]Computed using the VEST tool[100], which evaluates only the effect of missense substitutions. AA pos.: Amino acid position; Aff. AAs: Number of affected amino acids (the MLH1 protein contains 756 residues); CIMP: CpG island methylator phenotype; CIMPi: CIMP intermediate; COAD: Colon adenocarcinoma; EAC: Esophageal adenocarcinoma; READ: Rectal adenocarcinoma; STAD: Stomach adenocarcinoma; VEST: Variant effect scoring tool.

CIMP AND PROMOTER HYPERMETHYLATION OF TUMOR SUPPRESSOR GENES

The evidence pointing to *MLH1* inactivation as a corollary to the appearance of CIMP suggests that other tumor suppressor genes could potentially be silenced through promoter hypermethylation and result in comparable functional vulnerabilities as well; moreover, the silencing of these genes could represent actionable clinical targets. We explored this concept by searching for known tumor suppressor genes that exhibited concerted promoter hypermethylation in all four GIAD cancer types. Using the TSGene database[79], we found that 26 of 634 tumor suppressor genes (4.1%) contained at least one probe site in the promoter region that exhibited methylation levels significantly different between CIMP+ and CIMP- samples across all four cancer types (Table 5). These genes included *ERBB4*, *WT1*, *WIF1*, and *RASSF2*. By contrast, only 2.4% of genes not included in the TSGene database exhibited concordant differential methylation in CIMP+ samples across the four cancer types (P = 0.007, hypergeometric test).

Furthermore, in affected tumor suppressor genes, such as *DFNA5, RASSF2* and *WIF1*, promoter methylation was significantly negatively correlated with mRNA expression across several tumor types (Table 5), which is indicative of epigenetic silencing. *DNFA5* is a tumor suppressor gene involved in apoptosis and response to DNA damage[80,81]. Its hypermethylation has been reported in colorectal and gastric cancer, where it is associated with EBV-positive status[82,83]. In addition, *WIF1* and *RASSF2*, whose methylation and expression levels were significantly correlated across all four cancer types in our study, have been described in the context of CIMP in gastrointestinal adenocarcinomas[60,65,84-86]. These data suggest that, in a subset of genes, selective pressure may favor loss-of-function events caused by DNA methylation, facilitating tumor growth.

CIMP AND ASSOCIATED SOMATIC MUTATIONS

An outstanding question that remains is the causal connection between somatic mutations and the onset of CIMP. Over the years, extensive association analyses in colon and rectal cancers have been performed to address this problem[30,87]. The results have highlighted the diverse mutation spectrum across tissues, which refutes the hypothesis of a universal driver mutation being responsible for altered DNA methylation levels[51]. Mutations associated with CIMP have been found in *CDKN2A*, *IDH1/2, TET2* and *RB1*, among other genes[25]. In addition, as discussed, mutations in *BRAF* directly lead to hypermethylation at specific loci[76,77], and their effects probably extend to myriad targets across the genome.

We further explored the association between somatic mutations and DNA methylation using data from TCGA. For this purpose, we compared the recurrence of somatic mutations in CIMP+ and CIMP- samples across the entire GIAD cohort. A decision tree analysis pointed to several alterations associated with CIMP+ status (Figure 3A and B). This approach ranks mutations in descending order of statistical significance based on their presence or absence in CIMP+ samples. The top-scoring mutation was a 1-bp deletion at chr17:56,435,161 (Figure 3A), which was present in 21 of 22 STAD CIMP+ samples (Figure 3B). This mutation causes a frameshift in the last exon of *RNF43*, a tumor suppressor that encodes a RING-type

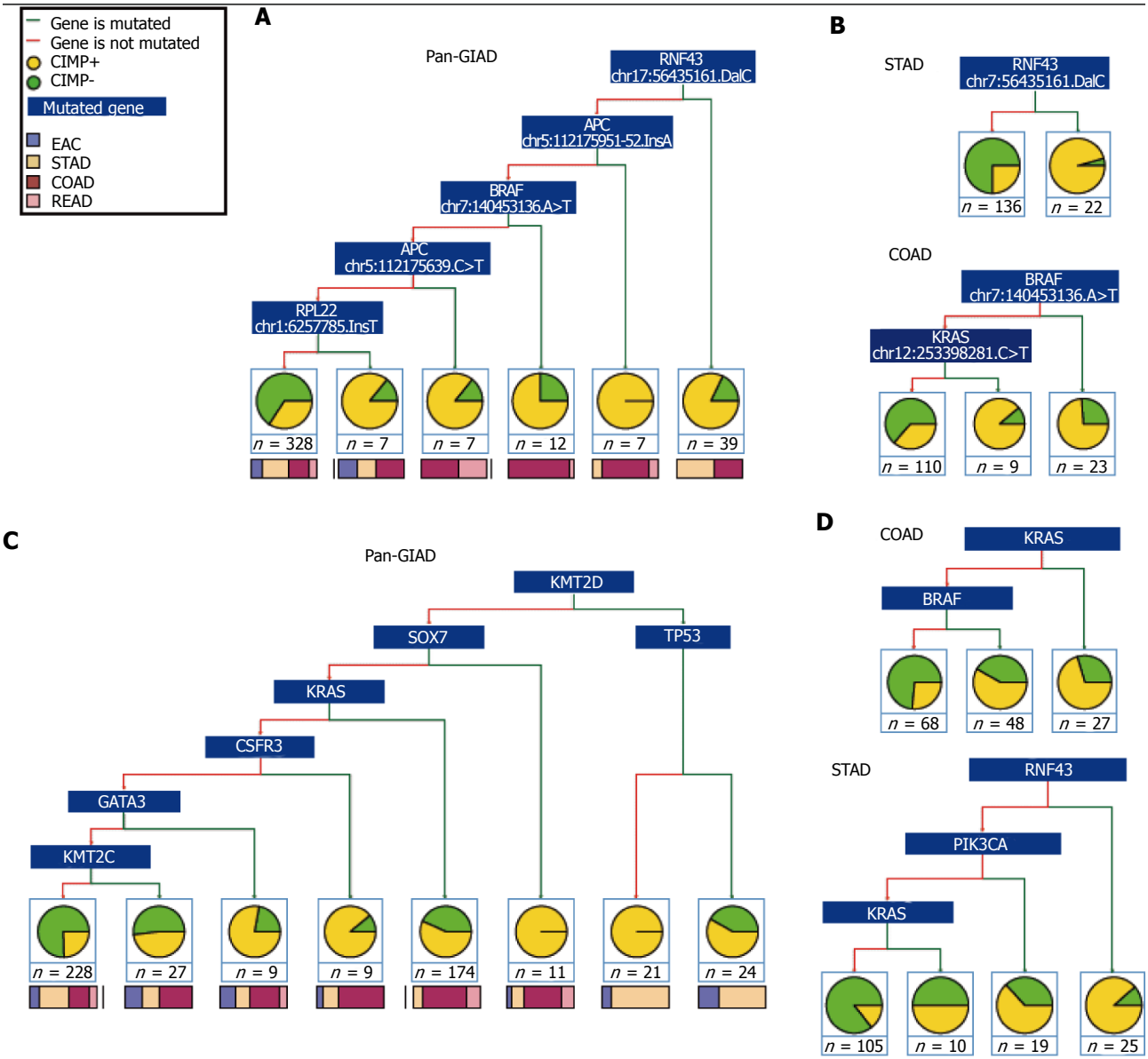

Figure 3 Binary decision trees for separating gastrointestinal adenocarcinomas into CpG island methylator phenotype categories. Recursive partitioning of GIADs from TCGA using binary classification trees based on CIMP status and mutational profiles. Results are provided for A: The combined GIAD data set at the individual mutation level; B: The STAD and COAD data sets at the individual mutation level; C: The combined GIAD data set at the mutated gene level; D: The STAD and COAD data sets at the mutated gene level. Red and green branches illustrate whether a specific mutation is present or absent (or whether a given gene is mutated or not) in the corresponding subset of tumors. Terminal nodes show the number of samples and the associated CIMP+ *vs* CIMP- fractions, as well as the proportion of different cancer types represented in each subset. GIADs: Gastrointestinal adenocarcinomas; TCGA: The Cancer Genome Atlas; CIMP: CpG island methylator phenotype.

E3 ubiquitin ligase (p.G659fs*41). *RNF43* is upregulated in colon cancer[88] and inhibits Wnt/β-catenin signaling in pancreatic cancer cells[89]. Two other top-scoring mutations affect *APC*, a tumor suppressor whose inactivation is associated with the onset of colon cancer. One was a nonsynonymous C-to-T substitution at chr5:112,175, 639, and the second was an AA insertion at chr5:112,175,951. Although these alterations were present in a relatively small number of samples (14 in total), they were observed almost exclusively in CIMP+ tumors (13 out of 14). Not surprisingly, we also found a BRAF V600E mutation (A-to-T change at chr7:140,453,136) that was significantly associated with CIMP+ status (Figure 3A). Together with a common *KRAS* mutation (C-to-T change at chr12:2,539,281; p.G13D), these represent the only two mutations significantly associated with CIMP+ in COAD samples; this is consistent with their already characterized presence in COAD[14,78]. Finally, a T insertion at chr1:6257785 affecting *RPL22* was also significantly associated with CIMP+ status across GIAD samples, although the number of affected samples was relatively small (6 out of 7 were CIMP+). In the future, these associations may be explored further to investigate their potential functional role in the context of aberrant DNA methylation.

We also compared mutations in CIMP+ and CIMP- samples by aggregating point mutations at the gene level (Table 6). Amid the top scorers in this analysis, we found

Table 5 Association between methylation and gene expression in tumor suppressor genes with significantly hypermethylated promoters in CpG island methylator phenotype + samples across four gastrointestinal adenocarcinoma types[1]

Gene symbol	Promoter probes	Differential methylation				Correlation with expression							
		Significant probes per cancer type				EAC		STAD		COAD		READ	
		COAD	EAC	READ	STAD	cor	p-val	cor	p-val	cor	p-val	cor	p-val
TP73	24	18	3	2	23	-0.34	4.E-01	-0.24	1.E-01	-0.10	1.E+00	-0.20	1.E+00
MAL	8	6	5	2	7	-0.37	3.E-01	-0.46	2.E-07	-0.47	7.E-09	-0.45	3.E-02
C2orf40	8	5	3	1	7	-0.51	2.E-03	-0.57	5.E-13	-0.39	2.E-05	-0.24	6.E-01
TMEFF2	7	7	7	7	7	-0.54	2.E-03	-0.49	2.E-08	-0.41	2.E-03	-0.32	5.E-01
ERBB4	6	7	7	2	7	-0.26	5.E-01	-0.15	5.E-01	NA	1.E+00	-0.30	6.E-01
TWIST2	5	5	4	1	4	-0.26	3.E-01	-0.33	3.E-04	-0.38	2.E-06	-0.37	3.E-02
LRRC3B	13	9	7	1	12	-0.36	5.E-01	-0.38	2.E-04	-0.41	4.E-03	-0.11	1.E+00
HTRA3	10	6	3	1	6	0.02	1.E+00	-0.03	1.E+00	-0.10	1.E+00	-0.07	1.E+00
UNC5C	13	13	13	8	13	-0.38	7.E-02	-0.34	5.E-04	-0.43	4.E-08	-0.40	3.E-02
FAT4	13	13	9	2	13	-0.33	2.E-01	-0.44	4.E-07	-0.34	4.E-05	-0.28	4.E-01
IRX1	5	3	4	3	4	-0.37	2.E-01	-0.37	3.E-04	NA	1.E+00	NA	1.E+00
SCGB3A1	9	9	4	2	9	-0.27	4.E-01	-0.37	3.E-05	-0.22	3.E-01	-0.12	1.E+00
AKAP12	10	9	5	1	10	-0.19	1.E+00	-0.42	1.E-06	-0.42	7.E-08	-0.27	6.E-01
DFNA5	12	10	8	1	9	-0.75	0.E+00	-0.56	1.E-12	-0.36	1.E-05	-0.34	1.E-01
TFPI2	15	22	14	19	22	-0.49	3.E-03	-0.54	2.E-11	-0.38	4.E-06	-0.22	1.E+00
NRCAM	7	6	1	1	6	-0.52	3.E-04	-0.47	2.E-08	-0.19	8.E-02	-0.17	1.E+00
CNTNAP2	14	14	11	1	14	-0.14	1.E+00	-0.27	2.E-02	-0.14	1.E+00	-0.12	1.E+00
PAX6	12	12	5	3	11	-0.22	1.E+00	-0.18	6.E-01	-0.04	1.E+00	-0.32	3.E-01
WT1	12	12	12	3	12	-0.25	1.E+00	-0.04	1.E+00	-0.26	8.E-03	-0.23	1.E+00
PHOX2A	11	11	6	5	11	-0.26	1.E+00	-0.13	1.E+00	-0.38	3.E-03	-0.26	1.E+00
WIF1	8	5	5	3	7	-0.57	2.E-03	-0.32	5.E-03	-0.44	9.E-07	-0.56	2.E-04
SLC5A8	9	11	10	4	12	-0.26	1.E+00	-0.28	2.E-01	-0.17	1.E+00	-0.33	9.E-01
TBX5	17	11	7	1	16	-0.32	5.E-01	-0.09	1.E+00	0.03	1.E+00	-0.16	1.E+00
ATP8A2	8	5	4	2	5	-0.28	1.E-01	-0.37	2.E-05	-0.24	5.E-03	-0.24	3.E-01
ADAMTS18	8	7	5	3	7	-0.27	3.E-01	-0.36	6.E-05	-0.34	5.E-05	-0.30	2.E-01
GALR1	29	27	5	8	27	-0.17	1.E+00	-0.47	2.E-04	0.00	1.E+00	-0.14	1.E+00
RASSF2	5	6	4	3	6	-0.52	5.E-04	-0.31	1.E-03	-0.41	1.E-07	-0.40	2.E-02
CDH4	3	2	2	2	4	-0.17	6.E-01	-0.09	8.E-01	-0.16	3.E-01	-0.32	1.E-01

[1]Promoter regions were designated as 2-kb regions encompassing 1.5 kb upstream and 0.5 kb downstream of the transcription start site. For each cancer type, probes were considered significant if the *P*-value after a one-sided Mann-Whitney *U* test with Bonferroni correction for multiple testing was < 0.05. The total number of probes considered across all four cancer types was 395814. COAD: Colon adenocarcinoma; EAC: Esophageal adenocarcinoma; READ: Rectal adenocarcinoma; STAD: Stomach adenocarcinoma.

chromatin remodeling genes such as *ARID1A*, which is an important member of the SWI/SNF complex, and histone methyltransferase genes such as *KMT2D* (*MLL2*) and *KMT2C* (*MLL3*). These two MLL complexes are involved in H3K27 demethylation and H3K4 methylation, which regulate the transcription of many developmental genes, including the *HOX* gene family[90]. The list of genes whose mutation levels were associated with CIMP status was also significantly enriched for genes from the RTK/RAS/PI(3)K signaling pathway (FDR < 4×10^{-8}), including *ERBB2*, *ERBB3, ERBB4, KRAS, PIK3CA, NRAS*, and *PTEN*. These results suggest that the cumulative signal of somatic mutations in coding genes could contribute to CIMP.

Finally, we applied binary decision trees to identify combinations of mutated genes that correlate with CIMP+ or CIMP- status (Figure 3C and D). Using the pooled GIAD data set, our tree shows that *KMT2D* mutations recur in gastroesophageal (*i.e.*, STAD and EAC) samples (Figure 3C). In fact, tumors with mutated *KMT2D* and wild-type *TP53* consist exclusively of CIMP+ samples (n = 21). We observed a second set of samples (including representatives from all four histologies) that contained *SOX7* mutations and lacked *KMT2D* mutations; all 11 of these tumors were CIMP+. Our trees from individual cancers (Figure 3D) show that *KRAS* and *BRAF* mutations in COAD, as well as *RNF43*, *PIK3CA,* and *KRAS* mutations in STAD, are associated with CIMP+ status.

CONFOUNDING FACTORS IN THE EVALUATION OF CIMP

Basing CIMP classification on mean methylation levels in tumor *vs* normal tissues allows us to separate cancer-related features from tissue-of-origin signals, but it also makes stratification vulnerable to a number of potential technical and biological artifacts. For example, our classification algorithm relies on the assumption of having a sufficiently large and sufficiently heterogeneous set of controls for each individual tumor type in order to guard against potentially confounding variables such as age, gender, race or anatomic location. Since only two non-tumor control samples were available for STAD, we may have encountered false positives in the probe selection process for this cancer type[51]. Another confounding effect may come from tumors' stimulation of the immune response, leading leukocytes (including T cells, NK cells, and macrophages) to infiltrate cancerous tissues and skew the methylation signature[91]. Additionally, tumor

Table 6 Genes differentially mutated between CpG island methylator phenotype+ and CpG island methylator phenotype- gastrointestinal adenocarcinoma samples[1]

Gene	Count CIMP+	% CIMP+	Count CIMP-	% CIMP-	P% Diff	*P*-value	FDR	Pathway
KMT2D	35	20.30%	10	4.30%	16.00%	6.22E-07	2.24E-05	Chromatin
ARID1A	60	34.90%	32	13.90%	21.00%	1.15E-06	2.24E-05	Chromatin
RNF43	42	24.40%	17	7.40%	17.10%	3.04E-06	3.79E-05	Wnt
CSF3R	19	11.00%	2	0.90%	10.20%	4.19E-06	3.79E-05	ERK
SOX7	14	8.10%	0	0.00%	8.10%	4.86E-06	3.79E-05	ERK
PIK3CA	48	27.90%	26	11.30%	16.70%	2.62E-05	1.70E-04	PI3K/RAS
PAX6	17	9.90%	2	0.90%	9.00%	3.96E-05	2.21E-04	Differentiation
ATM	37	21.50%	17	7.40%	14.20%	5.05E-05	2.46E-04	DNA damage
KRAS	52	30.20%	32	13.90%	16.40%	1.04E-04	4.53E-04	PI3K/RAS
EGR1	15	8.70%	2	0.90%	7.90%	1.63E-04	6.37E-04	Differentiation
GATA3	19	11.00%	5	2.20%	8.90%	2.22E-04	7.87E-04	NF-KB
KMT2C	38	22.10%	22	9.50%	12.60%	6.15E-04	2.00E-03	Chromatin
ALDH2	10	5.80%	1	0.40%	5.40%	1.18E-03	3.30E-03	Metabolic
CDK12	18	10.50%	6	2.60%	7.90%	1.18E-03	3.30E-03	PI3K/RAS
SAFB	15	8.70%	4	1.70%	7.00%	1.44E-03	3.73E-03	Chromatin
BCOR	19	11.00%	7	3.00%	8.00%	1.68E-03	4.09E-03	Chromatin
PTEN	24	14.00%	11	4.80%	9.20%	1.97E-03	4.32E-03	PI3K/RAS
AXIN2	21	12.20%	9	3.90%	8.30%	2.00E-03	4.32E-03	Wnt
CTCF	14	8.10%	4	1.70%	6.40%	2.73E-03	5.41E-03	Chromatin
PALB2	11	6.40%	2	0.90%	5.50%	2.77E-03	5.41E-03	DNA repair
ERBB3	18	10.50%	7	3.00%	7.40%	2.96E-03	5.49E-03	PI3K/RAS
ERBB4	29	16.90%	17	7.40%	9.50%	4.05E-03	6.97E-03	PI3K/RAS
FBXW7	32	18.60%	20	8.70%	9.90%	4.11E-03	6.97E-03	Notch
CIC	23	13.40%	12	5.20%	8.20%	6.55E-03	1.06E-02	Proliferation
HLA.A	17	9.90%	8	3.50%	6.40%	1.13E-02	1.71E-02	Immune
MSH6	19	11.00%	10	4.30%	6.70%	1.14E-02	1.71E-02	MMR
ERBB2	15	8.70%	8	3.50%	5.30%	2.98E-02	4.21E-02	PI3K/RAS
CASP8	13	7.60%	6	2.60%	5.00%	3.02E-02	4.21E-02	Apoptosis
SMAD4	27	15.70%	20	8.70%	7.00%	4.05E-02	5.45E-02	Wnt
TFE3	6	3.50%	1	0.40%	3.10%	4.53E-02	5.90E-02	Wnt
APC	82	47.70%	87	37.70%	10.00%	5.24E-02	6.60E-02	Wnt
NRAS	10	5.80%	5	2.20%	3.60%	6.55E-02	7.74E-02	PI3K/RAS
SMARCB1	10	5.80%	5	2.20%	3.60%	6.55E-02	7.74E-02	Chromatin
IGFBP7	3	1.70%	0	0.00%	1.70%	7.70E-02	8.65E-02	DNA Damage
TBL1XR1	6	3.50%	2	0.90%	2.60%	7.76E-02	8.65E-02	Wnt

[1]Results are based on a combined set of 179 CIMP- and 154 CIMP+ gastrointestinal adenocarcinoma samples. *P*-values were computed using a two-tailed Fishers' exact test. Only genes with FDR < 10% are shown. CIMP: CpG island methylator phenotype.

samples often consist of a heterogeneous mixture of cancer cells and non-cancer cells from adjacent tissues, the latter unwittingly included as a result of some biopsy collection procedures. As of today, there are no universally accepted methods to correct for tumor heterogeneity in DNA methylation studies; however, estimates of tumor heterogeneity can be computed from molecular data, such as copy number changes and mRNA expression, and these estimates can be used to discard problematic samples or to eliminate potential biases in downstream analyses[92,93]. As an alternative, future studies may benefit from improved sample collection requirements (*e.g.*, tumor microdissection) that lead to enhanced tumor purity and lower stromal contamination.

ASSESSING THE IMPACT OF TUMOR HETEROGENEITY ON CIMP CLASSIFICATION

We examined our CIMP classifications using the measure of tumor purity calculated with ABSOLUTE, a computational method based on the analysis of somatic DNA alterations[92]. As a proof of principle, we reclassified CIMP status for the STAD data set using only high-purity (*i.e.*, $\geqslant$ 50%) samples, as determined by the purity estimates available through TCGA[64]. We then compared sets of selected probes and CIMP designations before and after filtering for purity. After removing low-purity samples, the number of differentially methylated probes increased from 1110 to 1610. This result is consistent with the removal of samples that added background noise and masked the methylation signal of tumor cells. Since the new set of differentially methylated probes encompassed the original probe set, the inclusion of low-purity samples does not appear to have appreciably impacted our precision for feature selection, although it may have impoverished recall due to an increased number of false negatives. After using the new probe set, only five samples changed status from CIMP+ to CIMPi, and 11 samples changed status from CIMPi to CIMP-. However, no sample changed from CIMP+ to CIMP- or

vice versa. Thus, our CIMP classification system is robust in the presence of varying sample purity.

CONCLUSIONS AND PERSPECTIVES

Ever since the original study in CRC by Toyota *et al*[5,29], evaluation of CIMP status in cancer has been an active area of research. CIMP stratification has direct implications for patient treatment[24]. Because DNA methylation is potentially reversible, it represents an attractive target for therapies that can be tailored to individual cancer epigenomes[20,94]. Nucleoside analogs, such as 5-azacytidine, can be incorporated into DNA to reversibly block DNA methylation, and their effectiveness is being tested in numerous clinical trials.

In this commentary, we have provided evidence that supports refining the molecular profiles of gastrointestinal tumors based on CIMP status, to look beyond traditional tissue-of-origin interpretations. Our analysis of four types of gastrointestinal tumors not only confirms known CIMP associations but also leads to several new observations relevant to current models of DNA methylation and cancer. For example, we report recurrence of a frameshift mutation in *RNF43* that is significantly associated with CIMP status in stomach and, to a lesser extent, colon tumors. A recent study linked *RNF43* mutations to MSI in colorectal and endometrial tumors, which are Wnt-dependent[95]. The tumor suppressor function of this gene qualifies its mutations to be potential drivers of STAD, although mechanistic links to DNA methylation remain inconclusive. In addition, *RNF43* mutations had been identified in endometrioid and mucinous ovarian carcinomas[96]; we have shown the former tumor subtype is largely CIMP+[50]. The *RNF43* frameshift mutation that we highlighted in STAD samples in this paper is located within a 7-bp, CG-rich tract, and it may be created by the mismatch repair deficiency responsible for the MSI phenotype. Thus, the mutation's connection to CIMP status may occur downstream of MSI. However, *RNF43*-truncating mutations, which are common in MSI+ colorectal tumors, display mutual exclusivity with inactivating *APC* mutations[95], suggesting a more direct role in oncogenesis. Furthermore, our results point to additional events that could target the Wnt signaling pathway, such as epigenetic silencing of *WIF1*, which is consistently observed across the four GIAD types, or several of the somatic mutations highlighted in Table 6.

We believe that subdividing samples according to CIMP status has the potential to reduce heterogeneity within cancer subtypes and lead to more uniform molecular and phenotypic characteristics, thus producing more uniform response rates in clinical trials. Whether employed within cohort analyses or individual-level assessments, CIMP profiles have the potential to orient researchers and clinicians toward the biological properties of a tumor through their associations with MSI phenotypes, specific mutational profiles, and the repression of important tumor suppressor genes. Each of these avenues could potentially identify complementary therapeutic modalities. Guided in this way, researchers may identify new candidates for synthetic lethal therapeutic targeting, in which bottlenecks in pathways necessary for tumor cell survival can be targeted, resulting in more precise interventions than many of the current standard-of-care regimens.

RESEARCH

Data

We downloaded level 3 DNA methylation data from TCGA's data portal (https://tcga-data.nci.nih.gov/tcga/). Data had been acquired using the Illumina HumanMethylation 450K platform and pre-processed following TCGA standard protocols. We further normalized the data from each sample using the BMIQ method[97], which corrects for technical differences between type Ⅰ and type Ⅱ probes in the Illumina HumanMethylation platform. We also downloaded level 3 RNA-Seq data from the Broad Institute TCGA Genome Data Analysis Center (standard run dated 06/01/2015, http://dx.doi.org/10.7908/C1251HBG). For EAC, COAD, and READ, we used $\log_2$, normalized RSEM RNA-SeqV2 values. For STAD, we used $\log_2$ RPKM RNA-Seq values, since RSEM estimates were not available. In addition, somatic mutation data for all four cancer types were downloaded through the bulk download interface of the TCGA portal (https://tcga-data.nci.nih.gov/tcga/findArchives.htm). Finally, CpG island and transcript annotation data were downloaded from the UCSC genome browser (cpgIslandExt track for CpG islands, and refFlat and knownGene tracks for transcripts).

Algorithms and statistical analysis

All statistical analyses were done using the R statistical package. We used CpG island annotations from UCSC for hg19 and gene annotations provided by Illumina for the HumanMethylation 450K platform. Promoter regions were defined as 2-kb regions encompassing the TSS of protein-coding loci (1.5 kb upstream of the TSS and 500 bps downstream of the TSS). Our DNA methylation analysis was restricted to probes located within CpG islands. Within each individual cancer type, we discarded probes with low variance across samples (SD < 0.1, based on normalized β values between 0 and 1), as well as probes located on the X and Y chromosomes.

Discriminative probes were selected by requiring minimal methylation in control samples (average methylation in controls < 0.05) and increased methylation in tumor samples (average methylation in tumors > 0.25). After a set of discriminative probes had been chosen separately for each tumor type, samples were classified into CIMP categories using *k*-means clustering on the vector of average methylation values computed across the set of selected probes (k = 3, initial centroids chosen to match population quartiles). Binary decision trees were computed using the R package "partykit"[98-100].

Probe selection, CIMP classification, and decision tree analysis were performed as published in our previous

pan-cancer study[51]. We computed Spearman correlation values between expression values for each of the 28 genes in Table 5 and methylation values for probes in the corresponding TSSs. *P*-values were adjusted using the Bonferroni correction to account for the multiple probes associated with each gene.

ACKNOWLEDGMENTS

We thank Kristin Harper for editorial assistance. This work was funded by the Intramural program of the National Human Genome Research Institute, the National Institutes of Health.

REFERENCES

1 **Feinberg AP**, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983; **301**: 89-92 [PMID: 6185846 DOI: 10.1038/301089a0]
2 **Jones PA**, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002; **3**: 415-428 [PMID: 12042769 DOI: 10.1038/nrg816]
3 **Baylin SB**, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 2011; **11**: 726-734 [PMID: 21941284 DOI: 10.1038/nrc3130]
4 **Varley KE**, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ, Myers RM. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* 2013; **23**: 555-567 [PMID: 23325432 DOI: 10.1101/gr.147942.112]
5 **Toyota M**, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 1999; **96**: 8681-8686 [PMID: 10411935]
6 **Illingworth RS**, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 2010; **6**: e1001134 [PMID: 20885785 DOI: 10.1371/journal.pgen.1001134]
7 **Saxonov S**, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA* 2006; **103**: 1412-1417 [PMID: 16432200 DOI: 10.1073/pnas.0510310103]
8 **Herman JG**, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 2003; **349**: 2042-2054 [PMID: 14627790 DOI: 10.1056/NEJMra023075]
9 **Knudson AG**. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971; **68**: 820-823 [PMID: 5279523]
10 **Esteller M**. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007; **16** Spec No 1: R50-R59 [PMID: 17613547 DOI: 10.1093/hmg/ddm018]
11 **Herman JG**, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci USA* 1998; **95**: 6870-6875 [PMID: 9618505]
12 **Esteller M**, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, Baylin SB, Herman JG. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 2000; **343**: 1350-1354 [PMID: 11070098 DOI: 10.1056/NEJM200011093431901]
13 **Fleisher AS**, Esteller M, Wang S, Tamura G, Suzuki H, Yin J, Zou TT, Abraham JM, Kong D, Smolinski KN, Shi YQ, Rhyu MG, Powell SM, James SP, Wilson KT, Herman JG, Meltzer SJ. Hypermethylation of the hMLH1 gene promoter in human gastric cancers with microsatellite instability. *Cancer Res* 1999; **59**: 1090-1095 [PMID: 10070967]
14 **Weisenberger DJ**, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French AJ, Thibodeau SN, Jass J, Haile R, Laird PW. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 2006; **38**: 787-793 [PMID: 16804544 DOI: 10.1038/ng1834]
15 **Sproul D**, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol* 2012; **13**: R84 [PMID: 23034185 DOI: 10.1186/gb-2012-13-10-r84]
16 **Shen H**, Laird PW. Interplay between the cancer genome and epigenome. *Cell* 2013; **153**: 38-55 [PMID: 23540689 DOI: 10.1016/j.cell.2013.03.008]
17 **Ohm JE**, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 2007; **39**: 237-242 [PMID: 17211412 DOI: 10.1038/ng1972]
18 **Schlesinger Y**, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* 2007; **39**: 232-236 [PMID: 17200670 DOI: 10.1038/ng1950]
19 **Widschwendter M**, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW. Epigenetic stem cell signature in cancer. *Nat Genet* 2007; **39**: 157-158 [PMID: 17200673 DOI: 10.1038/ng1941]
20 **Heyn H**, Esteller M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* 2012; **13**: 679-692 [PMID: 22945394 DOI: 10.1038/nrg3270]
21 **Jerónimo C**, Usadel H, Henrique R, Oliveira J, Lopes C, Nelson WG, Sidransky D. Quantitation of GSTP1 methylation in non-neoplastic prostatic tissue and organ-confined prostate adenocarcinoma. *J Natl Cancer Inst* 2001; **93**: 1747-1752 [PMID: 11717336]
22 **Warren JD**, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, Fang JC, Samowitz WS, Heichman KA. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med* 2011; **9**: 133 [PMID: 22168215 DOI: 10.1186/1741-7015-9-133]
23 **Issa JP**. CpG island methylator phenotype in cancer. *Nat Rev Cancer* 2004; **4**: 988-993 [PMID: 15573120 DOI: 10.1038/nrc1507]
24 **Hughes LA**, Melotte V, de Schrijver J, de Maat M, Smit VT, Bovée JV, French PJ, van den Brandt PA, Schouten LJ, de Meyer T, van Criekinge W, Ahuja N, Herman JG, Weijenberg MP, van Engeland M. The CpG island methylator phenotype: what's in a name? *Cancer Res* 2013; **73**: 5858-5868 [PMID: 23801749 DOI: 10.1158/0008-5472.CAN-12-4306]
25 **Witte T**, Plass C, Gerhauser C. Pan-cancer patterns of DNA methylation. *Genome Med* 2014; **6**: 66 [PMID: 25473433 DOI: 10.1186/s13073-014-0066-6]
26 **Cedar H**, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 2009; **10**: 295-304 [PMID: 19308066 DOI: 10.1038/nrg2540]
27 **An C**, Choi IS, Yao JC, Worah S, Xie K, Mansfield PF, Ajani JA, Rashid A, Hamilton SR, Wu TT. Prognostic significance of CpG island methylator phenotype and microsatellite instability in gastric carcinoma. *Clin Cancer Res* 2005; **11**: 656-663 [PMID: 15701853]
28 **Roman-Gomez J**, Jimenez-Velasco A, Agirre X, Castillejo JA, Navarro G, Calasanz MJ, Garate L, San Jose-Eneriz E, Cordeu L, Prosper F, Heiniger A, Torres A. CpG island methylator phenotype redefines the prognostic effect of t(12; 21) in childhood acute lymphoblastic leukemia. *Clin Cancer Res* 2006; **12**: 4845-4850 [PMID: 16914570 DOI: 10.1158/1078-0432.CCR-05-2592]
29 **Toyota M**, Ho C, Ahuja N, Jair KW, Li Q, Ohe-Toyota M, Baylin SB, Issa JP. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 1999; **59**: 2307-2312 [PMID: 10344734]

30 **Curtin K**, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future. *Patholog Res Int* 2011; **2011**: 902674 [PMID: 21559209 DOI: 10.4061/2011/902674]

31 **Samowitz WS**, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* 2005; **129**: 837-845 [PMID: 16143123 DOI: 10.1053/j.gastro.2005.06.020]

32 **Chan AO**, Broaddus RR, Houlihan PS, Issa JP, Hamilton SR, Rashid A. CpG island methylation in aberrant crypt foci of the colorectum. *Am J Pathol* 2002; **160**: 1823-1830 [PMID: 12000733 DOI: 10.1016/S0002-9440(10)61128-5]

33 **Park SJ**, Rashid A, Lee JH, Kim SG, Hamilton SR, Wu TT. Frequent CpG island methylation in serrated adenomas of the colorectum. *Am J Pathol* 2003; **162**: 815-822 [PMID: 12598316 DOI: 10.1016/S0002-9440(10)63878-3]

34 **Ogino S**, Cantor M, Kawasaki T, Brahmandam M, Kirkner GJ, Weisenberger DJ, Campan M, Laird PW, Loda M, Fuchs CS. CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut* 2006; **55**: 1000-1006 [PMID: 16407376 DOI: 10.1136/gut.2005.082933]

35 **Shen L**, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci USA* 2007; **104**: 18654-18659 [PMID: 18003927 DOI: 10.1073/pnas.0704652104]

36 **Ang PW**, Loh M, Liem N, Lim PL, Grieu F, Vaithilingam A, Platell C, Yong WP, Iacopetta B, Soong R. Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer* 2010; **10**: 227 [PMID: 20492682 DOI: 10.1186/1471-2407-10-227]

37 **Kaneda A**, Yagi K. Two groups of DNA methylation markers to classify colorectal cancer into three epigenotypes. *Cancer Sci* 2011; **102**: 18-24 [PMID: 21159060 DOI: 10.1111/j.1349-7006.2010.01712.x]

38 **Ueki T**, Toyota M, Sohn T, Yeo CJ, Issa JP, Hruban RH, Goggins M. Hypermethylation of multiple genes in pancreatic adenocarcinoma. *Cancer Res* 2000; **60**: 1835-1839 [PMID: 10766168]

39 **Toyota M**, Kopecky KJ, Toyota MO, Jair KW, Willman CL, Issa JP. Methylation profiling in acute myeloid leukemia. *Blood* 2001; **97**: 2823-2829 [PMID: 11313277]

40 **Strathdee G**, Appleton K, Illand M, Millan DW, Sargent J, Paul J, Brown R. Primary ovarian carcinomas display multiple methylator phenotypes involving known tumor suppressor genes. *Am J Pathol* 2001; **158**: 1121-1127 [PMID: 11238060 DOI: 10.1016/S0002-9440(10)64059-X]

41 **Whitcomb BP**, Mutch DG, Herzog TJ, Rader JS, Gibb RK, Goodfellow PJ. Frequent HOXA11 and THBS2 promoter methylation, and a methylator phenotype in endometrial adenocarcinoma. *Clin Cancer Res* 2003; **9**: 2277-2287 [PMID: 12796396]

42 **Etoh T**, Kanai Y, Ushijima S, Nakagawa T, Nakanishi Y, Sasako M, Kitano S, Hirohashi S. Increased DNA methyltransferase 1 (DNMT1) protein expression correlates significantly with poorer tumor differentiation and frequent DNA hypermethylation of multiple CpG islands in gastric cancers. *Am J Pathol* 2004; **164**: 689-699 [PMID: 14742272 DOI: 10.1016/S0002-9440(10)63156-2]

43 **Liu Z**, Zhao J, Chen XF, Li W, Liu R, Lei Z, Liu X, Peng X, Xu K, Chen J, Liu H, Zhou QH, Zhang HT. CpG island methylator phenotype involving tumor suppressor genes located on chromosome 3p in non-small cell lung cancer. *Lung Cancer* 2008; **62**: 15-22 [PMID: 18358560 DOI: 10.1016/j.lungcan.2008.02.005]

44 **Yamashita K**, Dai T, Dai Y, Yamamoto F, Perucho M. Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* 2003; **4**: 121-131 [PMID: 12957287 DOI: 10.1016/S1535-6108(03)00190-9]

45 **Sandoval J**, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 2011; **6**: 692-702 [PMID: 21593595]

46 **Hinoue T**, Weisenberger DJ, Lange CP, Shen H, Byun HM, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM, Tollenaar RA, Laird PW. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* 2012; **22**: 271-282 [PMID: 21659424 DOI: 10.1101/gr.117523.110]

47 **Ogino S**, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn* 2006; **8**: 582-588 [PMID: 17065427 DOI: 10.2353/jmoldx.2006.060082]

48 **Jass JR**. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 2007; **50**: 113-130 [PMID: 17204026 DOI: 10.1111/j.1365-2559.2006.02549.x]

49 **Ogino S**, Kawasaki T, Kirkner GJ, Ohnishi M, Fuchs CS. 18q loss of heterozygosity in microsatellite stable colorectal cancer is correlated with CpG island methylator phenotype-negative (CIMP-0) and inversely with CIMP-low and CIMP-high. *BMC Cancer* 2007; **7**: 72 [PMID: 17474983 DOI: 10.1186/1471-2407-7-72]

50 **Kolbe DL**, DeLoia JA, Porter-Gill P, Strange M, Petrykowska HM, Guirguis A, Krivak TC, Brody LC, Elnitski L. Differential analysis of ovarian and endometrial cancers identifies a methylator phenotype. *PLoS One* 2012; **7**: e32941 [PMID: 22403726 DOI: 10.1371/journal.pone.0032941]

51 **Sánchez-Vega F**, Gotea V, Margolin G, Elnitski L. Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. *Epigenetics Chromatin* 2015; **8**: 14 [PMID: 25960768 DOI: 10.1186/s13072-015-0007-7]

52 **Bedard PL**, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature* 2013; **501**: 355-364 [PMID: 24048068 DOI: 10.1038/nature12627]

53 **Kanchi KL**, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, McMichael JF, Wyczalkowski MA, Larson DE, Schmidt HK, Miller CA, Fulton RS, Spellman PT, Mardis ER, Druley TE, Graubert TA, Goodfellow PJ, Raphael BJ, Wilson RK, Ding L. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun* 2014; **5**: 3156 [PMID: 24448499 DOI: 10.1038/ncomms4156]

54 **Gupta R**, Flanagan S, Li CC, Lee M, Shivalingham B, Maleki S, Wheeler HR, Buckland ME. Expanding the spectrum of IDH1 mutations in gliomas. *Mod Pathol* 2013; **26**: 619-625 [PMID: 23307057 DOI: 10.1038/modpathol.2012.210]

55 **Rodríguez-Paredes M**, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med* 2011; **17**: 330-339 [PMID: 21386836 DOI: 10.1038/nm.2305]

56 **Bernstein BE**, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006; **125**: 315-326 [PMID: 16630819 DOI: 10.1016/j.cell.2006.02.041]

57 **Ciriello G**, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 2013; **45**: 1127-1133 [PMID: 24071851 DOI: 10.1038/ng.2762]

58 **Sánchez-Vega F**, Gotea V, Petrykowska HM, Margolin G, Krivak TC, DeLoia JA, Bell DW, Elnitski L. Recurrent patterns of DNA methylation in the ZNF154, CASP8, and VHL promoters across a wide spectrum of human solid epithelial tumors and cancer cell lines. *Epigenetics* 2013; **8**: 1355-1372 [PMID: 24149212 DOI: 10.4161/epi.26701]

59 **Hoadley KA**, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V, Zhang J, Kandoth C, Akbani R, Shen H, Omberg L, Chu A, Margolin AA, Van't Veer LJ, Lopez-Bigas N, Laird PW, Raphael BJ, Ding L, Robertson AG, Byers LA, Mills GB, Weinstein JN, Van Waes C, Chen Z, Collisson EA, Benz CC, Perou CM, Stuart JM. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014; **158**: 929-944 [PMID: 25109877 DOI: 10.1016/j.cell.2014.06.049]

60 **Selaru FM**, David S, Meltzer SJ, Hamilton JP. Epigenetic events in gastrointestinal cancer. *Am J Gastroenterol* 2009; **104**: 1910-1912

[PMID: 19661933 DOI: 10.1038/ajg.2008.145]
61 **Cancer Genome Atlas Network.** Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; **487**: 330-337 [PMID: 22810696 DOI: 10.1038/nature11252]
62 **Cancer Genome Atlas Research Network**, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497**: 67-73 [PMID: 23636398 DOI: 10.1038/nature12113]
63 **Issa JP**. Colon cancer: it's CIN or CIMP. *Clin Cancer Res* 2008; **14**: 5939-5940 [PMID: 18829469 DOI: 10.1158/1078-0432.CCR-08-1596]
64 **Cancer Genome Atlas Research Network.** Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014; **513**: 202-209 [PMID: 25079317 DOI: 10.1038/nature13480]
65 **Liu JB**, Wu XM, Cai J, Zhang JY, Zhang JL, Zhou SH, Shi MX, Qiang FL. CpG island methylator phenotype and Helicobacter pylori infection associated with gastric cancer. *World J Gastroenterol* 2012; **18**: 5129-5134 [PMID: 23049225 DOI: 10.3748/wjg.v18.i36.5129]
66 **Sato F**, Meltzer SJ. CpG island hypermethylation in progression of esophageal and gastric cancer. *Cancer* 2006; **106**: 483-493 [PMID: 16362978 DOI: 10.1002/cncr.21657]
67 **Kaz AM**, Wong CJ, Luo Y, Virgin JB, Washington MK, Willis JE, Leidner RS, Chak A, Grady WM. DNA methylation profiling in Barrett's esophagus and esophageal adenocarcinoma reveals unique methylation signatures and molecular subclasses. *Epigenetics* 2011; **6**: 1403-1412 [PMID: 22139570 DOI: 10.4161/epi.6.12.18199]
68 **Eads CA**, Lord RV, Wickramasinghe K, Long TI, Kurumboor SK, Bernstein L, Peters JH, DeMeester SR, DeMeester TR, Skinner KA, Laird PW. Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res* 2001; **61**: 3410-3418 [PMID: 11309301]
69 **Galipeau PC**, Prevo LJ, Sanchez CA, Longton GM, Reid BJ. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. *J Natl Cancer Inst* 1999; **91**: 2087-2095 [PMID: 10601379]
70 **Kaz AM**, Grady WM. Epigenetic biomarkers in esophageal cancer. *Cancer Lett* 2014; **342**: 193-199 [PMID: 22406828 DOI: 10.1016/j.canlet.2012.02.036]
71 **Guo M**, Ren J, House MG, Qi Y, Brock MV, Herman JG. Accumulation of promoter methylation suggests epigenetic progression in squamous cell carcinoma of the esophagus. *Clin Cancer Res* 2006; **12**: 4515-4522 [PMID: 16899597 DOI: 10.1158/1078-0432.CCR-05-2858]
72 **Wong JJ**, Hawkins NJ, Ward RL, Hitchins MP. Methylation of the 3p22 region encompassing MLH1 is representative of the CpG island methylator phenotype in colorectal cancer. *Mod Pathol* 2011; **24**: 396-411 [PMID: 21102416 DOI: 10.1038/modpathol.2010.212]
73 **Bronner CE**, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 1994; **368**: 258-261 [PMID: 8145827 DOI: 10.1038/368258a0]
74 **Thibodeau SN**, French AJ, Roche PC, Cunningham JM, Tester DJ, Lindor NM, Moslein G, Baker SM, Liskay RM, Burgart LJ, Honchel R, Halling KC. Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes. *Cancer Res* 1996; **56**: 4836-4840 [PMID: 8895729]
75 **Fu T**, Pappou EP, Guzzetta AA, Jeschke J, Kwak R, Dave P, Hooker CM, Morgan R, Baylin SB, Iacobuzio-Donahue CA, Wolfgang CL, Ahuja N. CpG island methylator phenotype-positive tumors in the absence of MLH1 methylation constitute a distinct subset of duodenal adenocarcinomas and are associated with poor prognosis. *Clin Cancer Res* 2012; **18**: 4743-4752 [PMID: 22825585 DOI: 10.1158/1078-0432.CCR-12-0707]
76 **Fang M**, Ou J, Hutchinson L, Green MR. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Mol Cell* 2014; **55**: 904-915 [PMID: 25219500 DOI: 10.1016/j.molcel.2014.08.010]
77 **Carragher LA**, Snell KR, Giblett SM, Aldridge VS, Patel B, Cook SJ, Winton DJ, Marais R, Pritchard CA. V600EBraf induces gastrointestinal crypt senescence and promotes tumour progression through enhanced CpG methylation of p16INK4a. *EMBO Mol Med* 2010; **2**: 458-471 [PMID: 20941790 DOI: 10.1002/emmm.201000099]
78 **Serra RW**, Fang M, Park SM, Hutchinson L, Green MR. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *Elife* 2014; **3**: e02313 [PMID: 24623306 DOI: 10.7554/eLife.02313]
79 **Zhao M**, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res* 2013; **41**: D970-D976 [PMID: 23066107 DOI: 10.1093/nar/gks937]
80 **Masuda Y**, Futamura M, Kamino H, Nakamura Y, Kitamura N, Ohnishi S, Miyamoto Y, Ichikawa H, Ohta T, Ohki M, Kiyono T, Egami H, Baba H, Arakawa H. The potential role of DFNA5, a hearing impairment gene, in p53-mediated cellular response to DNA damage. *J Hum Genet* 2006; **51**: 652-664 [PMID: 16897187 DOI: 10.1007/s10038-006-0004-6]
81 **de Beeck KO**, Van Laer L, Van Camp G. DFNA5, a gene involved in hearing loss and cancer: a review. *Ann Otol Rhinol Laryngol* 2012; **121**: 197-207 [PMID: 22530481]
82 **Akino K**, Toyota M, Suzuki H, Imai T, Maruyama R, Kusano M, Nishikawa N, Watanabe Y, Sasaki Y, Abe T, Yamamoto E, Tarasawa I, Sonoda T, Mori M, Imai K, Shinomura Y, Tokino T. Identification of DFNA5 as a target of epigenetic inactivation in gastric cancer. *Cancer Sci* 2007; **98**: 88-95 [PMID: 17083569 DOI: 10.1111/j.1349-7006.2006.00351.x]
83 **Kim MS**, Chang X, Yamashita K, Nagpal JK, Baek JH, Wu G, Trink B, Ratovitski EA, Mori M, Sidransky D. Aberrant promoter methylation and tumor suppressive activity of the DFNA5 gene in colorectal carcinoma. *Oncogene* 2008; **27**: 3624-3634 [PMID: 18223688 DOI: 10.1038/sj.onc.1211021]
84 **Belshaw NJ**, Elliott GO, Foxall RJ, Dainty JR, Pal N, Coupe A, Garg D, Bradburn DM, Mathers JC, Johnson IT. Profiling CpG island field methylation in both morphologically normal and neoplastic human colonic mucosa. *Br J Cancer* 2008; **99**: 136-142 [PMID: 18542073 DOI: 10.1038/sj.bjc.6604432]
85 **Maruyama R**, Akino K, Toyota M, Suzuki H, Imai T, Ohe-Toyota M, Yamamoto E, Nojima M, Fujikane T, Sasaki Y, Yamashita T, Watanabe Y, Hiratsuka H, Hirata K, Itoh F, Imai K, Shinomura Y, Tokino T. Cytoplasmic RASSF2A is a proapoptotic mediator whose expression is epigenetically silenced in gastric cancer. *Carcinogenesis* 2008; **29**: 1312-1318 [PMID: 18310659 DOI: 10.1093/carcin/bgn060]
86 **Harada K**, Hiraoka S, Kato J, Horii J, Fujita H, Sakaguchi K, Shiratori Y. Genetic and epigenetic alterations of Ras signalling pathway in colorectal neoplasia: analysis based on tumour clinicopathological features. *Br J Cancer* 2007; **97**: 1425-1431 [PMID: 17923875 DOI: 10.1038/sj.bjc.6604014]
87 **Hughes LA**, Khalid-de Bakker CA, Smits KM, van den Brandt PA, Jonkers D, Ahuja N, Herman JG, Weijenberg MP, van Engeland M. The CpG island methylator phenotype in colorectal cancer: progress and problems. *Biochim Biophys Acta* 2012; **1825**: 77-85 [PMID: 22056543 DOI: 10.1016/j.bbcan.2011.10.005]
88 **Yagyu R**, Furukawa Y, Lin YM, Shimokawa T, Yamamura T, Nakamura Y. A novel oncoprotein RNF43 functions in an autocrine manner in colorectal cancer. *Int J Oncol* 2004; **25**: 1343-1348 [PMID: 15492824]
89 **Jiang X**, Hao HX, Growney JD, Woolfenden S, Bottiglio C, Ng N, Lu B, Hsieh MH, Bagdasarian L, Meyer R, Smith TR, Avello M, Charlat O, Xie Y, Porter JA, Pan S, Liu J, McLaughlin ME, Cong F. Inactivating mutations of RNF43 confer Wnt dependency in pancreatic ductal adenocarcinoma. *Proc Natl Acad Sci USA* 2013; **110**: 12649-12654 [PMID: 23847203 DOI: 10.1073/pnas.1307218110]
90 **Tian X**, Zhang S, Liu HM, Zhang YB, Blair CA, Mercola D, Sassone-Corsi P, Zi X. Histone lysine-specific methyltransferases and demethylases in carcinogenesis: new targets for cancer therapy and prevention. *Curr Cancer Drug Targets* 2013; **13**: 558-579 [PMID: 23713993]
91 **Lombaerts M**, Middeldorp JW, van der Weide E, Philippo K, van Wezel T, Smit VT, Cornelisse CJ, Cleton-Jansen AM. Infiltrating

leukocytes confound the detection of E-cadherin promoter methylation in tumors. *Biochem Biophys Res Commun* 2004; **319**: 697-704 [PMID: 15178462 DOI: 10.1016/j.bbrc.2004.05.041]

92 **Carter SL**, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012; **30**: 413-421 [PMID: 22544022 DOI: 10.1038/nbt.2203]

93 **Yoshihara K**, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013; **4**: 2612 [PMID: 24113773 DOI: 10.1038/ncomms3612]

94 **Egger G**, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004; **429**: 457-463 [PMID: 15164071 DOI: 10.1038/nature02625]

95 **Giannakis M**, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, Saksena G, Lawrence MS, Qian ZR, Nishihara R, Van Allen EM, Hahn WC, Gabriel SB, Lander ES, Getz G, Ogino S, Fuchs CS, Garraway LA. RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet* 2014; **46**: 1264-1266 [PMID: 25344691 DOI: 10.1038/ng.3127]

96 **Ryland GL**, Hunter SM, Doyle MA, Rowley SM, Christie M, Allan PE, Bowtell DD, Gorringe KL, Campbell IG. RNF43 is a tumour suppressor gene mutated in mucinous tumours of the ovary. *J Pathol* 2013; **229**: 469-476 [PMID: 23096461 DOI: 10.1002/path.4134]

97 **Teschendorff AE**, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013; **29**: 189-196 [PMID: 23175756 DOI: 10.1093/bioinformatics/bts680]

98 **Hothorn T**. Partykit: A Modular Toolkit for Recursive Partytioning in R. *J Mach Learn Res* 2015; **16**: 3905-3909

99 **Gao J**, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; **6**: pl1 [PMID: 23550210 DOI: 10.1126/scisignal.2004088]

100 **Carter H**, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 2013; **14** Suppl 3: S3 [PMID: 23819870 DOI: 10.1186/1471-2164-14-S3-S3]

P- Reviewer: Arora Z, Park J, Sun XT, Voutsadakis IA
S- Editor: Gong XM **L- Editor**: A **E- Editor**: Lu YJ

Published by **Baishideng Publishing Group Inc**
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com