Genomic insights into cancer-associated aberrant CpG island hypermethylation

Duncan Sproul and Richard R. Meehan

Advance Access publication date 21 January 2013

Abstract
Carcinogenesis is thought to occur through a combination of mutational and epimutational events that disrupt key pathways regulating cellular growth and division. The DNA methylomes of cancer cells can exhibit two striking differences from normal cells; a global reduction of DNA methylation levels and the aberrant hypermethylation of some sequences, particularly CpG islands (CGIs). This aberrant hypermethylation is often invoked as a mechanism causing the transcriptional inactivation of tumour suppressor genes that directly drives the carcinogenic process. Here, we review our current understanding of this phenomenon, focusing on how global analysis of cancer methylomes indicates that most affected CGI genes are already silenced prior to aberrant hypermethylation during cancer development. We also discuss how genome-scale analyses of both normal and cancer cells have refined our understanding of the elusive mechanism(s) that may underpin aberrant CGI hypermethylation.

Keywords: epigenetics; epigenomics; cancer; DNA methylation; CpG islands

INTRODUCTION
It is widely accepted that carcinogenesis necessitates multiple genetic alterations that either drive cellular division or remove checkpoints regulating this process in normal cells. These same disruptions could potentially also be caused by epimutations. Epigenetic events have been strictly defined as heritable changes in gene function that are not explained by changes in DNA sequence but also more recently as ‘the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states’ [1]. Here, we refer to epimutations as heritable, abnormal alterations in the state of chromosomal regions.

Strong support for the possibility that epimutations play a significant role in cancer comes from the recent discovery that epigenetic regulators are recurrently mutated in cancer genomes and observations that the levels and distributions of epigenetic marks are altered in cancer [2]. Particular attention and research effort has focused on the hypothesis that aberrant silencing of genes by DNA hypermethylation is a key epimutational mechanism driving carcinogenesis. Importantly, a clear mechanism has been described for the inheritance of DNA methylation patterns across cellular generations [1] and, therefore, abnormal DNA methylation states fit the definition of epimutations as heritable alterations.

In normal mammalian somatic genomes, DNA methylation mainly occurs at cytosines in a CpG dinucleotide context [3]. Around 70% of CpGs in mammalian genomes are methylated but DNA methylation is bimodally distributed and is generally absent from short stretches of CpG-rich sequence known as CpG islands (CGIs) [4] which frequently correspond to the promoters of genes [5] (Figure 1A).
The remainder of the genome is relatively depleted in CpGs due to the inherent mutability of methylcytosine (mC) which is prone to spontaneous deamination [6]. Such deamination can also cause cancer-associated mutations, including many TP53 (p53) mutations [7]. Methylated cytosines in the genome are recognised by methyl CpG binding proteins (MBDs) which are hypothesised to play an important role in reading this epigenetic mark [8]. During embryogenesis, the mammalian genome undergoes a series of epigenetic reprogramming events including a wave of global demethylation followed by establishment of the bimodal pattern by the de novo methyltransferases DNMT3A and DNMT3B [9]. The maintenance DNA methyltransferase, DNMT1, ensures that this distribution is stably inherited during the remainder of development and differences in the DNA methylation profiles between adult cell types have been shown to be relatively small, particularly at CGIs which are generally maintained in a hypomethylated state regardless of gene expression status [10, 11]. Regions bordering CGIs in the human genome, termed CGI shores, have been suggested to be more variably methylated between different normal cell types [12, 13]. At present, however, it is unclear whether they represent a distinct functional genomic compartment. Hydroxymethylation of cytosines (hmC) has also been recently rediscovered in mammalian cells [14] and significant levels of this modification are present in the bodies of active genes in some somatic tissues [15].

In cancer cells, genomic levels of DNA methylation are frequently reduced compared with their normal counterparts [16]. The underlying cause(s) of this reduction is unknown, but the loss can be localised to particular types of repetitive elements [17] or chromosomal domains [18, 19] (Figure 1B). Global levels of hydroxymethylation have also recently been shown to be reduced in cancer cells [15, 20–22]. Contrasting with this overall trend, many CGIs undergo cancer-associated aberrant hypermethylation [18] (Figure 1B). Hypermethylation of CGI promoters is tightly linked with transcriptional repression of the affected gene [23] and many promoters initially shown to be aberrantly hypermethylated in cancer correspond to known tumour suppressor genes (TSGs) [24]. Aberrant CGI hypermethylation has, therefore, been viewed as an epimutation causing the silencing of TSGs (Figure 2A) [25]. Thus, a strong hypothesis is that the aberrant hypermethylation of CGIs can drive carcinogenesis and cancer progression and that epimutation events might outnumber mutations in cancer [26].

Here, we review our current understanding of aberrant CGI hypermethylation as a paradigm of a potential epimutation in cancer. In particular, we focus on how recent data from the study of cancer methylomes demonstrates that the majority of genes that are aberrantly hypermethylated in cancer are in fact already repressed in preneoplastic cells. This parallels the view that most CGI hypermethylation in
normal development occurs subsequent to silencing by other means. Based on these findings, we present alternative hypotheses as to the impact of aberrant CGI hypermethylation on the growth and development of cancers. Finally, we describe how the genome-scale analysis of cancer has advanced our understanding of the potential molecular mechanisms behind this epigenetic reprogramming.

EVIDENCE FOR INACTIVATION OF TUMOUR SUPPRESSOR GENES BY ABERRANT HYPERMETHYLATION

Evidence that aberrant CGI hypermethylation might act as an epimutation directly driving carcinogenesis is based primarily on studies of individual candidate genes. The aberrant hypermethylation of genes such as RB1 [27–30], MLH1 [31, 32] and BRCA1 [33], whose mutation is associated with inherited cancer predisposition [34], can be regarded as particularly significant. Three important pieces of evidence support the view that aberrant hypermethylation in cancer causes their silencing.

First, hypermethylation of TSGs has been observed alongside inherited germline mutations [35–37]. This suggests promoter hypermethylation can directly substitute for genetic loss of heterozygosity (LOH) as the second hit that completely disables TSG activity. The evidence is particularly strong in the case of CDKN2A (p16/ARF) where the presence of a mutation in the first exon in a colon cancer cell line facilitated the direct demonstration that hypermethylation occurs only on the wild-type allele [36]. In the majority of cases, however, such analyses have not or cannot be conducted and it remains to be demonstrated whether hypermethylation frequently occurs in an allele-specific fashion. Recent studies of BRCA1 have failed to observe instances of LOH through hypermethylation, suggesting that it might be a very rare event [38, 39].

Second, the tissue specificity of TSG hypermethylation in sporadic cancer overlaps with the tissue-specific predispositions caused by inherited mutations in these same genes. Inherited MLH1 mutations predispose to colorectal cancer and MLH1 hypermethylation is largely limited to colorectal tumours [40]. Similarly, BRCA1 mutations predispose specifically to breast and ovarian cancer and hypermethylation is limited to cancer of these tissues [40]. Within particular tissues, the phenotypes of cancers which have hypermethylated particular TSGs can overlap with the specific phenotypes of cases associated with inherited mutations in the same gene. For example, RB1 mutated and hypermethylated retinoblastomas phenocopy each other [37], colorectal tumours with either mutated or hypermethylated MLH1 have microsatellite instability (MSI) [32] and VHL hypermethylation occurs in renal cancers of the clear cell histological subtype as do VHL mutations [41]. Rare cases of inherited methylation of MLH1 and MSH2 also confer a predisposition to developing MSI colorectal cancer as do mutations in these genes, although these apparent epimutations are tightly associated with genetic variants [42, 43]. The overlap between the phenotype of mutated and hypermethylated cancers is less clear in other cases. Carriers of BRCA1 mutations develop particular types of breast cancer, specifically estrogen receptor negative (ER—ve) tumours that are often classified as belonging to one of a few special histological types [44]. BRCA1 hypermethylation was reported to be more frequent in medullary carcinomas which are ER—ve and often occur in BRCA1 carriers, but is also observed in mucinous carcinomas which are ER+ve and often occur in BRCA1 carriers. BRCA1 hypermethylation is not specific to ER—ve breast cancers [47] and that the gene expression profiles of BRCA1 mutated and methylated cancers differ [48]. BRCA1 hypermethylated serous ovarian adenocarcinomas also displayed a clinical course that was
more similar to BRCA1 wild type than BRCA1 mutant tumours [49]. It should be noted, however, that inherited and somatic mutations in important TSGs can be associated with different phenotypes. For example, inherited mutations in CDKN2A predispose to melanoma and pancreatic tumours [34] but somatic CDKN2A mutations occur in a variety of other cancer types including non-small cell lung cancer [50].

The final strong piece of evidence underpinning the causative role of aberrant DNA hypermethylation in silencing tumour suppressor genes in cancer is that they can be reactivated when methylation is removed from their promoters. This is most often achieved by treatment with the drug 5′-aza-2′-deoxycytidine (5-Aza) which causes the degradation of DNMT1, the maintenance methyltransfrase [51]. Treatment of cancer cell lines with 5-Aza causes the reactivation of hypermethylated VHL [41], MLH1 [32, 52] and BRCA1 [53]. Also, genetic knockout or RNA-mediated knockdown of DNMT results in activation of previously hypermethylated CDKN2A in a colorectal cancer cell line [54, 55]. These experiments, however, do not examine the temporal sequence of events causing gene silencing. Treatment of female mammalian cells with 5-Aza causes the activation of genes on the inactive X chromosome (Xi) [56], but silencing of genes on the Xi precedes hypermethylation of gene promoters [57, 58] and can occur in the absence of DNMTs [59, 60]. These observations demonstrate that gene activation can occur upon ablation of promoter hypermethylation even when the hypermethylation was not the initial and causative silencing event.

MOST ABERRANTLY HYPERMETHYLATED CGI GENES ARE REPRRESSED PRIOR TO HYPERMETHYLATION

The evidence for aberrant CGI hypermethylation as a direct silencer of TSGs is mostly correlative leading many to question its direct role in carcinogenesis [61–64]. Although cancer-associated hypermethylation of a gene’s promoter has been invoked to suggest it might possess tumour suppressor activity [65], many aberrantly hypermethylated genes are unlikely candidates as TSGs, for example, CALCA (Calcitonin), the first gene reported to become hypermethylated in cancer [66]. Early unbiased studies of cancer methylomes made it clear that large numbers of genes could be hypermethylated in a single specimen [67]. The most dramatic cases are cancers with CGI hypermethylator phenotypes (CIMP), first described in colorectal tumours [68] and more recently in cancers arising in other tissues [69–72]. The frequency of aberrant hypermethylation has been used to suggest that epimutations might be more significant than mutations in carcinogenesis [26], but this is difficult to reconcile with evidence suggesting that relatively few mutations are likely to be required for carcinogenesis [34, 73].

In order for aberrant hypermethylation to directly drive cancer by silencing genes, the affected genes must be expressed prior to hypermethylation. Transcriptionally repressed genes have been known to undergo hypermethylation in tissue culture for many years [74]. Recent integrated analyses of cancer methylomes together with gene expression data demonstrate that transcriptionally repressed genes are in fact also the predominant target of cancer-associated aberrant hypermethylation (Figure 2B). By analysing the methylation profiles of cancers derived from seven different tissue types, we have shown that genes which are repressed in a lineage-specific fashion in normal tissues become hypermethylated in cancers derived from that tissue, whereas housekeeping or expressed lineage-specific genes are resistant to hypermethylation [40, 75]. A study of colon cancer found that 93% of the genes hypermethylated in CIMP tumours had unaltered expression in tumours compared with normal tissue [76]. This suggests they are already repressed in the normal colon because CGI hypermethylation is tightly associated with transcriptional repression. One explanation for the low correlation frequently observed between gene expression changes and promoter hypermethylation in cancer methylomes studies is that the majority of affected genes are normally repressed in the tissue studied [71, 77–81]. A comparison of an osteosarcoma cell line to cultured mesenchymal stem cells and osteoblasts also found that the majority of aberrantly hypermethylated genes in the osteosarcoma cell line were repressed in the normal cells analysed [82]. It is possible that hypermethylation prone genes are expressed to a low level rather than repressed in normal tissue [18] but background levels of hybridisation to probes make it difficult to draw this conclusion from microarray expression data. A recent and comprehensive analysis of the normal expression of RUNX3, which is frequently hypermethylated in gastric cancers, conclusively demonstrated that it is in fact never expressed in the cells that give rise to
these tumours and supports the hypothesis that hyper-
methylation prone genes are fully repressed rather than
expressed to a low level [65].

These recent findings from the study of cancer methylomes draw parallels with our understanding of CGI hypermethylation during normal development which, most evidence suggests, occurs at genes already repressed through other mechanisms [83, 84]. As noted above, the hypermethylation of CGIs on the Xi in female cells occurs after genes have already been silenced [57, 58]. Mice deficient for the de novo methyltransferases initiate X-inactivation normally [60] and genetic deletion of Dnmt1 does not preclude X-inactivation but instead results in sporadic reactivation of the Xi later in development [59]. Repression of the paternal copy of the imprinted Meg3 (Gtl2) promoter in mouse development occurs prior to its hypermethylation [85]. The CGI promoter of the pluripotency associated transcription factor Oct-3/4 is also silenced before becoming hypermethylated during differentiation and functional studies implicate hypermethylation in stabilising its silencing [86, 87]. It is repressed gene promoters that become hypermethylated during the in vitro differentiation of mouse ES cells [88] and silencing of a transgene in chicken erythroid cells precedes its hypermethylation [89]. A stabilising role for CGI hypermethylation is also supported by observations that DNA methylation represents a bar-
errier to the reprogramming of somatic cell types to induced pluripotent stem cells (iPS cells) [90] and that its ablation increases reprogramming efficiency [91]. Taken together, findings from the study of cancer methylomes put cancer-associated aberrant CGI hypermethylation in a similar framework to CGI hypermethylation in normal development, as a largely secondary event (Figure 2B).

DO EPIGENETIC EVENTS IN CANCER FOLLOW A DRIVER AND PASSENGER MODEL?

Although the analysis of cancer methylomes demonstrates that the vast majority of genes whose CGI promoters become hypermethylated in cancer are repressed prior to hypermethylation, the possibility remains that occasionally active genes become hypermethylated and repressed (Figure 2). Such genes have been termed epigenetic drivers [92, 93] and the bulk of hypermethylated genes, which are repressed in normal untransformed tissue, are hence termed passengers, a nomenclature adopted from studies of the mutational structure of cancer genomes [94]. The integrative analyses described above do not exclude the possibility that a small proportion of hypermethylated genes might be active in the tissue of origin [40, 76] and one study has suggested that a significant proportion of genes hypermethylated in an osteosarcoma cell line are active in normal cells [82]. The estimation of the exact proportion of normally active genes that are aberrantly hyper-
methylated in cancer is, however, complicated by a number of issues. First, normal tissues consist of a heterogenous mix of cell types and cancers often originate from rare cell populations, and therefore, the bulk expression profile of a normal tissue might not be representative of a cancer’s cell of origin [95]. Second, in these analyses, one measurement of expression level is generally used for the whole gene. The existence of alternative promoters, how-
ever, means that hypermethylation of an apparently active gene may actually occur at an inactive pro-
moter, as has been described for APC in gastric cancer [96]. Clarification of this situation will require the analysis of rigorously purified cell populations using techniques that measure promoter transcrip-
tional activity, such as CAGE [97], and carefully constructed bioinformatic analysis pipelines which focus on promoters rather than whole genes.

The strong phenotype of cancer predisposition genes which are also hypermethylated [32, 33] demonstrates that they must be expressed in the affected normal tissues. A key question in these cases is whether their aberrant hypermethylation is also sec-
ondary? The potential primary repressive event would have to occur abnormally rather than as part of normal development, as is the case for the majority of aberrantly hypermethylated genes. MLH1 is frequently hypermethylated in colorectal tumours with a CIMP phenotype [76] supporting the possibility that MLH1 becomes hypermethylated through the same mechanism affecting passenger genes. BRCA1 is repressed in sporadic breast cancer in the absence of hypermethylation [47] suggesting that it too could be repressed prior to hypermethylation. Silencing of BRCA1 in cell lines can be initiated by the transcription factors SNAI1 and SNAI2 (Snail and Slug, respectively) in partnership with the histone lysine demethylase LSD1 [98]. An interesting example is CDKN2A which is frequently hypermethylated in a variety of tumour types [99]. CDKN2A is normally only expressed in cells following replicative or
oncogenic stress and it remains silent in most normal cells [100]. This suggests the possibility that the gene could become hypermethylated prior to transformation. CDKN2A is normally repressed by polycomb repressive complexes (PRCs) [101, 102] which have been implicated as part of the mechanism associated with aberrant hypermethylation (see below). Furthermore, after reactivation following genetic knockout of DNMT1 and DNMT3B in a colon cancer cell line, CDKN2A is silenced before DNA hypermethylation is re-established [103] and CDNK2A silencing also occurs prior to hypermethylation in colonies of human mammary epithelial cells (HMECs) that escape replicative arrest in culture [104].

Taken together, these observations suggest that although some hypermethylated genes might be expressed in normal tissues, their hypermethylation could follow aberrant down-regulation by other means. Another possibility is that separate mechanisms result in the hypermethylation of active and repressed genes. A recent study demonstrated that the hypermethylation of initially active genes on the Xi requires additional factors compared with initially silent genes supporting the idea that alternative mechanisms might underpin the hypermethylation of different gene types [105]. Non-coding genetic variants may also cause aberrant CGI hypermethylation, as exemplified by cases of inherited allelic hypermethylation of MLH1 and MSH2 which are tightly associated with sequence variants [42, 43]. The case of CDNK2A suggests that prior expression in normal tissue is not a perquisite for the hypermethylation of a gene playing a role in promoting carcinogenesis. Also, even in cases of driver genes where hypermethylation was not the initiating silencing event, its role in maintaining silencing might be important for the continued growth of the cancer. Potential epigenetic driver genes have been identified by screening for genes whose promoters remain aberrantly hypermethylated after genetic ablation of DNA methyltransferase activity in a hypomorphic colorectal cancer cell line [106]. The identified genes were not classically known to be TSGs and their expression in normal colon was not demonstrated but their enforced expression via a strong CMV promoter in wild-type colon cancer cell lines inhibited their growth suggesting that maintenance of hypermethylation at these genes was important for the fitness of the cell line.

While the potential that the hypermethylation of some genes is subject to positive selection because it facilitates carcinogenesis fits current data from methylome studies, a number of questions remain. One prediction of the driver and passenger model is that tumours with methylator phenotypes would demonstrate more aggressive clinical behaviour because they would be statistically more likely to have hypermethylated more TSGs or drivers. Methylator phenotypes in colorectal tumours, breast tumours and glioblastomas, however, all coincide with better clinical prognoses [69, 71, 107]. It also remains to be explained why the mutational and hypermethylation landscapes are so different if the selection of rare driver events underpins their development. Putative passengers, that is to say the majority of aberrantly hypermethylated genes, are frequently and reproducibly hypermethylated in cancer whereas the strongest candidates for driver genes, those that confer cancer predisposition when mutated, are hypermethylated much more rarely [40]. This contrasts with the mutational landscape of cancer where individual driver genes are observed to be more frequently mutated in clinical samples than passenger genes as a result of selection [108].

ALTERNATIVE HYPOTHESES AS TO THE ROLE OF ABBERRANT CGI HYPERMETHYLATION IN CANCER

Rather than playing an evolutionary neutral passenger role and representing a surrogate of general epigenetic dysregulation, the widespread hypermethylation of normally repressed genes in cancer could have other impacts on carcinogenesis and progression. As discussed above, hypermethylation of CGIs in normal development results in stable gene silencing and prevents ectopic activation [83]. It is conceivable that abundant promoter hypermethylation could influence the epigenetic plasticity of cancer cells with two possible outcomes.

Many groups have noted that genes repressed by PRCs in embryonic stem (ES) cells are frequently hypermethylated in cancer [109–111]. Other studies have reported overlaps in the gene expression profiles of ES cells and aggressive cancers [112] which, at least in part, correspond to repression of ES cell PRC targets in cancers [113]. Many of the genes targeted by PRCs in ES cells are transcription factors whose expression is key to lineage commitment during differentiation [114]. Thus, it has been
proposed that the frequent hypermethylation of ES cell PRC targets in cancer might block differentiation and maintain cancers in a stem-cell-like state [111] (Figure 3A). In support of this hypothesis, genes which are expressed late in murine-lung differentiation are reported to be frequently hypermethylated in non-small cell lung cancer [115]. IDH1R132H mutations correlate with a hypermethylator phenotype in acute myeloid leukaemia (AML) [70] and knock-in of this mutant into mouse haematopoietic stem cells results in general hypermethylation and an apparent block to differentiation [116]. These mutations, however, result in the production of an oncometabolite, 2-hydroxyglutarate, which affects the function of numerous cellular enzymes [117, 118] and the differentiation block observed in this study is not necessarily a result of CGI hypermethylation. The hypothesis that widespread hypermethylation maintains cancers in a aggressive stem-cell-like state is also inconsistent with the better clinical prognoses associated with cancer hypermethylator phenotypes [69, 71, 107].

Instead of maintaining a stem-cell-like state, a restriction of epigenetic plasticity induced by widespread hypermethylation could act as a check on cancer progression (Figure 3B). The dissemination of cancer cells from the site of origin and their survival in metastatic niches requires the activation of gene expression programs [119]. Resistance to therapy can occur as a result of secondary activating mutations [120, 121] and potentially through epigenetic gene activation. The stable gene repression associated with hypermethylation might provide a barrier to these events resulting in an inhibition of cancer progression. This hypothesis is more consistent with the better prognosis associated with hypermethylation in cancer [69, 71, 107].

One hypothesis is that aberrant CGI hypermethylation in cancer results from the over-expression or increased activity of DNMTs. Such increases were initially reported [124, 125] but are likely to be attributed to the regulation of DNMTs during the cell cycle [126, 127] and an increased number of cycling cells in cancer. A recent analysis reports that hypermethylation at some genes correlates with increased DNTM3B levels in colorectal tumours [128]. Experimental manipulation of DNMT levels in the Apcmin/+ mouse model of colorectal cancer demonstrate that higher DNMT levels promote carcinogenesis [129–131]. Dnmt3b over-expression in this model is also associated with promoter hypermethylation of the murine homologues of genes hypermethylated in human colorectal tumours [132]. On the other hand, DNMT3A mutations occur in AML and other haematological cancer genomes [133, 134] and these mutations have been shown to reduce DNMT3A enzymatic activity.

Cancers are, however, very heterogenous with respect to their genetic and epigenetic profile and the environment in which they grow; the two hypotheses outlined above may, therefore, reflect differing roles of widespread aberrant hypermethylation in different cancer types or subtypes. They both may even play a part in the same cancer at different stages of progression. For example, a hypermethylation-mediated block on differentiation may promote the initial growth of a tumour but later this restricted epigenetic landscape might prevent its metastasis. Furthermore, neither of these hypotheses is incompatible with the occasional hypermethylation of driver genes outlined above. Differentiating these possibilities requires an understanding of the mechanism responsible for aberrant CGI hypermethylation.

**THE MECHANISM(S) OF ABERRANT CGI HYPERMETHYLATION IN CANCER**

Although the mechanism(s) responsible for aberrant promoter hypermethylation in cancer remains elusive, potential hypotheses have emerged from genome-scale analyses of both normal and cancerous cells. Two main types of mechanisms have been proposed; active processes mediated by targeting of specific factors to CGIs or passive mechanisms resulting from a loss of protection against de novo methylation.

One hypothesis is that aberrant CGI hypermethylation in cancer results from the over-expression or increased activity of DNMTs. Such increases were initially reported [124, 125] but are likely to be attributed to the regulation of DNMTs during the cell cycle [126, 127] and an increased number of cycling cells in cancer. A recent analysis reports that hypermethylation at some genes correlates with increased DNTM3B levels in colorectal tumours [128]. Experimental manipulation of DNMT levels in the Apcmin/+ mouse model of colorectal cancer demonstrate that higher DNMT levels promote carcinogenesis [129–131]. Dnmt3b over-expression in this model is also associated with promoter hypermethylation of the murine homologues of genes hypermethylated in human colorectal tumours [132]. On the other hand, DNMT3A mutations occur in AML and other haematological cancer genomes [133, 134] and these mutations have been shown to reduce DNMT3A enzymatic activity.
They have not, however, been found to correlate with variations in CGI hypermethylation patterns [133, 136]. DNMT3B mutations in cancer have not currently been found but aberrant splicing of DNMT3B frequently gives rise to cancer-specific isoforms of the enzyme [137]. These isoforms lack a methyltransferase domain but they could potentially function similarly to DNMT3L, acting as cofactors to bring canonical DNMT3A and 3B to new locations and stimulating their activity [138, 139]. Although variations in DNMT activity might affect carcinogenesis, it remains to be demonstrated whether this

Figure 3: Models as to the consequences of widespread CGI hypermethylation in cancer. (A) Hypermethylation as a block to differentiation. This model predicts that key genes required for normal cellular differentiation become hypermethylated in cancer. The result is a block to their activation and thus normal differentiation processes within the cancer producing a more aggressive, stem-cell like phenotype. (B) Hypermethylation as a block to progression. Widespread hypermethylation of repressed CGI promoters might prevent their stochastic activation in individual cancer cells. If the activation of such genes facilitated survival in changing conditions, such as during metastasis to distant organs or treatment, widespread hypermethylation might restrict the potential for epigenetic adaptation and thus result in block to progression.
occurs as a result of increased CGI hypermethylation. It is also unclear as to whether MBDs play any role in the process of aberrant CGI hypermethylation. The MBDs Mbd2 and Kaiso appear have contributory roles in intestinal tumorigenesis as their deletion in mice result in reduced tumour numbers in Apca

 occurs in cancer. The analysis of cancer methylomes has also linked dysfunction in DNA demethylation pathways to aberrant CGI hypermethylation. DNA demethylation is proposed to be initiated by the conversion of mC to hmC by the ten–eleven translocation (TET) family of enzymes [158, 159]. TET1 is bound to CGIs in mouse ES cells [160, 161] leading to the proposal that it maintains the fidelity of DNA methylation patterns in cells by maintaining CGIs in a hypomethylated state [162]. As noted above, reduction in global hmC is frequent in cancer [15, 20–22] and disruptions to TET enzyme function have been linked to CGI hypermethylation. Mutations abrogating TET2 enzymatic activity are frequent in AML [163] and correlate with a hypermethylator phenotype [70]. The oncometabolite 2-hydroxyglutarate is also produced in AMLs with IDH1 or 2 mutations and one of its effects is the inhibition of TET enzyme activity [118]. IDH mutations in AML correlate with a similar hypermethylator phenotype to TET2 mutations [70] and CIMP in glioblastoma is also associated with IDH1 mutations [71]. Although one study suggested Tet1 knockdown in mouse ES cells led to CGI hypermethylation [161], the same observation was not made in a similar independent study [160]. As TET1 is also bound to the vast majority of CGIs in mouse ES cells, it is unclear why TET dysfunction in cancer might result in the preferential hypermethylation of repressed CGIs.

Epigenomic analyses of normal cells point towards DNA sequence determining genome methylation patterns through sequence-specific transcription factors (TFs) rather than vice versa [164–166]. DNMT3A and 3B have been shown to interact with normal TFs [167] and abnormal versions generated as a result of gene fusions [168]. Although a number of studies have published sequence motifs associated with genes that become hypermethylated in cancer, these have neither been reproduced nor demonstrated to correspond to the binding sites of particular TFs [78, 169]. As TFs are found at both active and repressed genes, it is unclear why their recruitment of DNMT would specifically result in the hypermethylation of repressed genes. A reproducible association does occur between general TF motifs and hypermethylation-resistant promoters [18, 170] which is consistent with their housekeeping gene status [40]. In addition to
suggesting that TF binding protects CGIs from hypermethylation, this correlation could be explained if transcriptional activity itself conferred protection from hypermethylation. One model for the generation of bimodal hypermethylation patterns in mammalian cells postulates that CGIs are protected from de novo methylation by transcription during the developmental window in which the genome is remethylated [63]. An analysis comparing normal and cancer cell lines also showed that the presence of stalled or active RNA polymerase in normal cells predicts resistance to aberrant hypermethylation in cancer cells [171]. Variations in DNA sequence that alter promoter activity also correlate with the predisposition of promoters to hypermethylation. An extra SP1 site in the RIL gene confers resistance to hypermethylation [172] and sequence variants in the MLH1 promoter that reduce promoter activity are associated with soma-wide mosaic hypermethylation [42]. It is unclear, however, if low-level transcription confers resistance to CGI hypermethylation or whether the relationship is quantitative with increasing transcription levels resulting in a lower frequency of hypermethylation but not entirely excluding it. Models of protection based on the hypothesis that transcription protects CGIs must also consider the presence of stalled RNA polymerase at PRC-marked genes [173, 174].

One of the main results to emerge from the systematic analysis of cancer methylomes is the finding that cancer-associated hypermethylation does not occur at random but affects a distinct set of genes. As noted above, several groups have documented the overlap between PRC-marked genes in ES cells and hypermethylation in cancer and this has been reproduced extensively in genome profiling studies [76, 109–111, 175]. Rather than being repressed, PRC-occupied CGIs in ES cells are proposed to exist in a poised state [176] which is resolved to either full activation or repression as differentiation proceeds [177]. A subset of PRC-marked genes in ES cells are, therefore, expected to be occupied by PRCs in differentiated cell types. The association between PRC-marked CGIs in ES cells and hypermethylation in cancer might, therefore, reflect preferential hypermethylation of those repressed CGIs associated with PRCs in adult cells during the transformation process rather than a mechanistic connection between ES cell and cancer epigenetic state. The promoters of hypermethylation-prone genes are also relatively depleted of retrotransposons compared with hypermethylation-resistant promoters [178]. This could result from evolutionary selection against retrotransposon integration near tissue-specific genes because it might disrupt essential interactions with distal regulatory elements [40]. Overall, the characteristics of hypermethylation-prone genes are consistent with the possibility that repressed, lineage-specific genes are predominantly subject to cancer-associated hypermethylation [40].

While the observation that aberrant hypermethylation affects a specific set of genes has been presented as evidence for a targeted rather than stochastic process of hypermethylation [179], it does not exclude the possibility that stochastic hypermethylation occurs with a susceptible set of genes. One recent study has suggested that normal cell line promoter methylation patterns evolve through a stochastic process [180].

Taken together, these studies suggest that many factors could be involved in the reprogramming of repressed CGIs to a hypermethylated state in cancer. It is clear that a number of different factors found at active CGI promoters are capable of maintaining them in a hypomethylated state including H3K4me3, H2A.Z, TFs, TET enzymes and active RNA transcription (Figure 4A). Repressed CGIs are also occupied by a number of factors that could potentially perform the same function including TETs, PRCs and RNA polymerase (Figure 4B). Furthermore, some of these factors are shared with active CGIs, so it is uncertain why a defect in any one factor would result in aberrant CGI hypermethylation (Figure 4B). The potential role of active recruitment of DNMTs in this picture is also unclear and the relative importance of stochastic and targeted processes in the evolution of cancer CGI methylomes remains to be determined.

SUMMARY AND CONCLUSIONS

The recent characterisation of cancer methylomes has demonstrated, contrary to the prevailing view, that the hypermethylation of CGI promoters in cancer parallels CGI hypermethylation during normal development and is secondary to silencing by other means. For most aberrantly hypermethylated promoters, this silencing occurs as a result of normal development and subsequent methylation represents an epigenetic reprogramming event. Potential driver genes that are expressed in normal cells could conceivably be directly silenced by
The overwhelming tendency for hypermethylation to occur as a secondary event, however, strongly suggests that such drivers are subject to primary aberrant silencing in cancer through other means. Hypermethylation of isolated individual repressed genes may also contribute a growth advantage to cancer by preventing their activation at later stages of carcinogenesis or progression. We should also consider whether the widespread hypermethylation of CGIs in cancer has other impacts on the growth of cancers, in particular by blocking differentiation or restricting epigenetic plasticity and adaptive potential.

The analysis of cancer genomes and methylomes have helped refine our definition of the type of gene affected by aberrant promoter hypermethylation and generated new hypotheses as to the molecular defect underpinning this epigenetic reprogramming but many questions remain to be answered. Dissection of this mechanism is also likely to lead to new insights regarding the biology of CGIs, the most abundant promoter type in our genome [63]. The reinterpretation of cancer-associated CGI hypermethylation that has occurred as a result of the advent of genome-scale datasets should also be considered as potential epimutations associated with other diseases are identified [181]. Finally, cancer epigenomes are potentially a rich source of biomarkers and specific epigenetic defects may be exploitable therapeutic targets [182]. We have not discussed these avenues of research here, but our new more global understanding of cancer-associated aberrant hypermethylation.

**Figure 4:** Potential mechanisms of cancer-associated aberrant hypermethylation. CGIs bound are occupied by multiple factors that potentially play roles in aberrant CGI hypermethylation. (A) Nucleosomes incorporating the histone variant H2A.Z and with H3K4me3 marks, TFs, active transcription and TET enzymes may all help maintain the hypomethylated state of active CGIs in normal cells. TET enzyme function is most likely compromised in at least some cancers but this still leaves other protective factors at active CGIs. (B) Similarly, a number of factors are found at inactive CGIs that could potentially play a role in maintaining their normal hypomethylated state including TET enzymes, PRCs and stalled RNA polymerase. PRCs are thought to be lost, most likely along with stalled RNA polymerase, when CGIs become aberrantly hypermethylated, but it is unclear if this plays a mechanistic role in this cancer-associated epigenetic reprogramming. The role of DNMT recruitment in the process is also unclear. White lollipops - unmethylated CpGs, black lollipops - methylated CpGs.
CGI hypermethylation should be used to guide these efforts.

As this review was going to press, de-repression of CXCR4, which has a CGI promoter, was shown to facilitate metastasis in a renal cancer cell line [183]. This further supports our proposal that widespread aberrant hypermethylation of CGI promoters in cancer could inhibit progression by blocking gene activation (Figure 3).

Key Points
- CpG islands (CGIs) frequently become aberrantly hypermethylated in cancers.
- Analysis of cancer methylomes has shown that aberrant CGI hypermethylation occurs primarily at genes that are already silent in the host tissue and is therefore not generally linked to transcriptional silencing of tumour suppressor genes.
- The predominant view is that CGI hypermethylation in normal development is also secondary to prior silencing through other mechanisms.
- Several hypotheses now exist as to the impact of CGI hypermethylation on carcinogenesis and progression. The occasional hypermethylation of rare driver genes might directly promote carcinogenesis. Widespread CGI hypermethylation could also result in more aggressive cancers by blocking cellular differentiation or act as a protective mechanism hindering progression by preventing epigenetic adaptation to changing conditions.
- The mechanism underpinning aberrant CGI island hypermethylation remains elusive but genome-scale studies have refined our hypotheses and demonstrated that a distinct gene set is affected.

Acknowledgements
We apologise to colleagues whose work could not be cited due to space limitations. We thank Wendy Bickmore (MRC Human Genetics Unit, Edinburgh, UK) for encouragement in writing this review and Peter Adams (CRUK Beatson Institute, Glasgow, UK), Robert Illingworth (MRC Human Genetics Unit) and Ragnhild Eskeland (Department of Molecular Biosciences, University of Oslo, Norway) for invaluable comments on the manuscript.

FUNDING
Work in R.M.’s lab is supported by the Medical Research Council, the Biotechnology and Biological Sciences Research Council and by the Innovative Medicine Initiative Joint Undertaking (IMI JU) under grant agreement number 115 001 (MARCAR project). D.S. is supported by the Medical Research Council, Breakthrough Breast Cancer and Breast Cancer Campaign.

References


64. Fearon ER. BRCA1 and E-cadherin promoter hypermethylation and gene inactivation in cancer-associated aberrant CpG island hypermethylation.


37. Robertson KD, Keyomarsi K, Gonzales FA, et al. Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the...


169. Takeshima H, Yamashita S, Shimazu T, et al. The presence of RNA polymerase II, active or stalled, predicts epigenetic


