Characterizing DNA methylation alterations from The Cancer Genome Atlas

Daniel J. Weisenberger


The Cancer Genome Atlas (TCGA) Research Network is an ambitious multi-institutional consortium effort aimed at characterizing sequence, copy number, gene (mRNA) expression, microRNA expression, and DNA methylation alterations in 30 cancer types. TCGA data have become an extraordinary resource for basic, translational, and clinical researchers and have the potential to shape cancer diagnostic and treatment strategies. DNA methylation changes are integral to all aspects of cancer genomics and have been shown to have important associations with gene expression, sequence, and copy number changes. This Review highlights the knowledge gained from DNA methylation alterations in human cancers from TCGA.

Find the latest version:

http://jci.me/69740-pdf
Characterizing DNA methylation alterations from The Cancer Genome Atlas

Daniel J. Weisenberger

USC Epigenome Center, USC Norris Comprehensive Cancer Center, Department of Biochemistry and Molecular Biology, University of Southern California, Keck School of Medicine, Los Angeles, California, USA.

The Cancer Genome Atlas (TCGA) Research Network is an ambitious multi-institutional consortium effort aimed at characterizing sequence, copy number, gene (mRNA) expression, microRNA expression, and DNA methylation alterations in 30 cancer types. TCGA data have become an extraordinary resource for basic, translational, and clinical researchers and have the potential to shape cancer diagnostic and treatment strategies. DNA methylation changes are integral to all aspects of cancer genomics and have been shown to have important associations with gene expression, sequence, and copy number changes. This Review highlights the knowledge gained from DNA methylation alterations in human cancers from TCGA.

The Cancer Genome Atlas overview

The Cancer Genome Atlas (TCGA) Research Network is a multi-institutional consortium aimed at performing comprehensive molecular profiling of 10,000 primary tumors spanning 30 cancer types (Table 1). Molecular profiling includes whole-exome sequencing for mutation detection, RNA sequencing for gene expression profiling, microRNA sequencing for microRNA expression profiling, single nucleotide polymorphism arrays for determining somatic copy number alterations (SCNAs), and Illumina Infinium BeadArrays for DNA methylation profiling (1–3). The novelty of TCGA stems from large sample numbers, centralized pathological review, selection of samples with a high fraction of tumor nuclei, genomic characterizations using nucleic acids isolated from the same tissue samples, and data integration for high-level pathway interpretations. This Review summarizes the TCGA DNA methylation findings to date, focusing on the molecular features of CpG island methylator phenotypes (CIMPs) shared among cancer types, the relationship between DNA methylation profiles and chromatin modifiers, and epigenetic silencing of key driver genes of carcinogenesis.

Overview of cancer epigenetics

The processes of DNA methylation, chromatin modifications, nucleosome positioning, and transcription factor binding shape the epigenome and are implicated in cellular differentiation, imprinting, and X chromosome inactivation (reviewed in refs. 4, 5). In mammalian cells, DNA methylation almost exclusively occurs at the C-5 position of cytosines in the sequence context of 5′-CpG-3′ (reviewed in ref. 6). Most CpGs are methylated but are located in low-density CpG regions. However, there are regions of the genome termed “CpG islands” that contain higher CpG and G+C content, are typically unmethylated in normal somatic tissues, and are frequently located in the promoter/5′ regions of genes (7).

Human cancers exhibit DNA hypomethylation in repetitive elements, low-density CpG regions (8–15), and lamin-associated domains (16–19). This reduction in DNA methylation occurs concomitant with locus-specific DNA hypermethylation in CpG islands (reviewed in ref. 20) and CpG island shores (21). DNA hypermethylation in gene promoter regions is a frequent event in every human cancer and can inversely correlate with gene expression (reviewed in refs. 4, 5, 22). Epigenetic silencing of tumor suppressor genes has been demonstrated in several cancer types, including the cyclin-dependent kinase inhibitor 2A (CDKN2A, also referred to as p16INK4A) and secreted frizzled-related protein genes in colorectal and lung cancers (23, 24), the breast cancer 1, early onset (BRCA1) gene in breast and ovarian cancers (25), and the von Hippel–Lindau tumor suppressor E3, ubiquitin protein ligase gene in kidney cancers (24).

There is also interest in understanding coordinated cancer-associated DNA methylation events across the genome. For example, polycomb repressive complex (PRC) gene targets in ES cells also exhibit cancer-specific DNA hypermethylation (26–28). PRCs repress expression of transcription factors and drivers of development and differentiation in ES cells (29). This enrichment of DNA hypermethylation at ES PRC targets suggests that crosstalk between PRC and DNA methylation machineries may occur early in tumorigenesis (28).

CIMPs in human cancers

In 1999, Issa and colleagues first described a distinct subset of human colorectal cancers with extensive DNA hypermethylation of a subset of CpG islands that remained unmethylated in the remaining colorectal tumors (30) and are therefore distinguished from general cancer-specific DNA methylation for a specific tumor type (Figure 1). These tumors were classified as positive for a CIM. TCGA Research Network and others have identified CIMPs in breast, colorectal, and endometrial tumors as well as in glioblastomas and acute myeloid leukemias, but not in serious ovarian, lung squamous, or kidney renal cell cancers. TCGA has unveiled similar and unique characteristics between CIMPs of different tumor types that have potential implications for the development of novel cancer diagnostics and therapeutic agents.

Colonorectal CIMP is tightly associated with the BRAFV600E mutation.

Colonorectal CIMP has been identified and characterized using candidate and genome-scale approaches in numerous reports after Issa first identified CIMP in colorectal cancer (30). Colonorectal CIMP tightly associates with mutation of the V-raf murine sarcoma viral oncogene homolog B (BRAF) gene correlating with
the V600 amino acid (BRAF\textsuperscript{V600E}), DNA methylation of the mutL homolog 1 (MLH1) promoter region, microsatellite instability (MSI), location in the proximal colonic region, and female gender (31–35). Ogino et al. first described a CIMP-low (CIMP-L) subgroup as having an attenuated CIMP phenotype, and showed an association with mutations in the kirsten rat sarcoma viral oncogene homolog (KRAS) gene rather than BRAF (36). Similarly, Shen et al. identified the CIMP2 subgroup as displaying CIMP-associated DNA methylation, with enrichment in KRAS mutations (32). KRAS mutations are also enriched in non-CIMP tumors, but are strikingly absent in CIMP-high (CIMP-H) tumors (34–37). TCGA confirmed CIMP-H, CIMP-L, and two non-CIMP subgroups of colorectal cancer (34, 35). CIMP-H was present in approximately 15% of colorectal tumors, the majority of which also showed elevated mutation rates (hypermutated) and showed few SCNAS in contrast to the majority of colorectal tumors, which are non-CIMP, non-hypermutated, and microsatellite stable but which show substantial SCNAS (34).

The molecular mechanisms that explain the tight correlation between CIMP-H and the BRAF\textsuperscript{V600E} mutation are not well understood. TCGA did not identify driver events, such as specific mutations or SCNAS in a \textit{trans}-acting factor in CIMP-H and BRAF\textsuperscript{V600E} colorectal tumors. While both the BRAF\textsuperscript{V600E} mutation and DNA methylation changes are thought to occur early in colorectal tumorigenesis, it is unclear whether CIMP or BRAF\textsuperscript{V600E} mutation is the initiating event. Hinoue and colleagues did not observe CIMP after introducing exogenous BRAF\textsuperscript{V600E} into a wild-type BRAF, non-CIMP colorectal cancer cell line (38). However, Hinoue et al. did identify CIMP-specific epigenetic silencing of IGF-binding protein 7 (IGFBP7), which mediates BRAF\textsuperscript{V600E}-induced apoptosis and cellular senescence. As BRAF\textsuperscript{V600E} has been implicated in oncogene-induced senescence in melanomas and colorectal cancers (39), CIMP-specific silencing of IGFBP7 may create a favorable context for the generation of the BRAF\textsuperscript{V600E} mutation in CIMP-positive tumors (38). However, additional experiments are needed to determine the molecular mechanism linking CIMP and BRAF\textsuperscript{V600E} in colorectal cancer.

**Isocitrate dehydrogenase mutations and DNA hypermethylation in glioma and AML.** Glioblastoma multiforme (GBM) was the first cancer selected by TCGA Research Network for molecular characterization (40). The majority of GBM tumors are considered primary or de novo, and approximately 5% are secondary GBMs which progress from lower-grade tumors. Parsons and colleagues used a comprehensive sequencing strategy and identified specific heterozygous somatic point mutations in the isocitrate dehydrogenase 1 (IDH1)
non-G-CIMP patients had shorter survival outcomes. Interestingly, even though the M6 tumors were proneural, patients with this tumor type did not show survival advantages, which suggests that the aberrant molecular features of G-CIMP may be important in conferring the survival advantage in G-CIMP patients.

Because of the expanded GBM tumor collection and exome sequencing depth, TCGA identified amplifications of MYC (v-myc avian myelocytomatosis viral oncogene homolog) in G-CIMP tumors. MYC is a transcription factor that is frequently altered in cancer and is involved in cell cycle progression, transformation, and apoptosis (reviewed in ref. 47). In addition, TCGA also identified ATRX (α-thalassemia/mental retardation syndrome X-linked) somatic mutations in G-CIMP tumors, which are highly correlated with IDH1 mutations. ATRX belongs to the SWI/SNF family of chromatin remodelers and functions as an ATPase and helicase that facilitates the substitution of variant histone H3.3 into chromatin at telomeres (48).

ATRX mutations are predominant in the alternative lengthening of telomeres (ALT), a process by which telomere length is maintained independent of telomerase in cancer cells (48, 49). Interestingly, TCGA identified telomerase (TERT) promoter mutations in 21 of 25 GBM tumors sequenced, and these mutations correlated with increased TERT gene expression. Notably, all four tumors without TERT mutations did not display elevated TERT gene expression, but instead contained ATRX alterations. Therefore, it is possible that GBM tumors maintain telomere length by either TERT mutations to reactivate TERT gene expression or via ATRX mutations in ALT.

The link between the IDH1R132H mutation and G-CIMP has generated tremendous attention from basic and translational scientists. Recent studies have shown that introducing exogenous IDH1R132H into immortalized cell lines with endogenous IDH1WT was sufficient to drive G-CIMP–based DNA methylation events and increased occupancy of histone H3 lysine 9 dimethylation (H3K9me2), histone H3 lysine 27 trimethylation (H3K27me3), and H3K36me3, which correlate with methylated DNA regions in the cancer genome (50, 51).

These epigenomic changes may occur as a result of the function of the mutant IDH1 proteins. IDH1WT functions as a dimer to catalyze the reduction of nicotinamide adenine dinucleotide phosphate (NADP+) to NADPH by converting of isocitrate to α-ketoglutarate (α-KG) (reviewed in ref. 52). However, the IDH mutant catalyzes the conversion of α-KG to D-2-hydroxyglutarate (2-HG) (53–55), resulting in elevated 2-HG levels (53, 56, 57).

IDH1 mutant-mutant (IDH1MUT-MUT) homodimers or mutant–WT (IDH1MUTWT) heterodimers have been identified in vitro (58). However, recent in vitro experiments have shown that the presence of the IDH1WT protein is associated with increased 2-HG levels (54), and that IDH1MUTWT dimers show more enzymatic activity toward α-KG than IDH1MUT-MUT alone (59), suggesting that both forms may be required for 2-HG production.

2-HG inhibits the TET family of enzymes and Junonj-C domain containing histone lysine demethylases, which normally utilize α-KG as a co-substrate (60, 61). Thus, the production of 2-HG by mutant IDH1-containing enzymes effectively inhibits TET activity (Figure 2A). TETs catalyze the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and ultimately to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (62, 63). 5fC and 5caC are substrates for the thymine DNA glycosylase-mediated base excision repair pathway that ultimately results in replacement with an unmethylated cytosine, effectively demethylating the locus (64–67) (Figure 2B).
lesions after the initial DNA strand-nicking step in the repair pathway. The nicks accumulate and are thought to promote an apoptotic response that results in cell death.

O-6-methylguanine methyltransferase (MGMT) removes methyl groups from the O-6 position of guanines, thereby rendering TMZ ineffective. Promoter DNA hypermethylation–based silencing of MGMT sensitizes GBM tumors to TMZ, and as a result has been used as a diagnostic barometer for selecting TMZ as a treatment option for GBM patients. TCGA recently identified MGMT DNA methylation in nearly 50% of GBM patients, and MGMT DNA methylation was more prevalent in G-CIMP tumors than in non–G-CIMP tumors (44). MGMT DNA methylation correlated with patient response in GBMs belonging to the classical gene expression subgroup only, and not in the proneural, neural, or mesenchymal groups.

Breast cancer CIMP. Breast cancer is a complex and heterogeneous disease, and breast tumor subgroups have been proposed based on the expression status of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2; also known as ERBB2). TCGA identified five DNA methylation subgroups of breast tumors (78), with one subgroup exhibiting a CIMP-like DNA methylation signature (B-CIMP), as previously described by Fang and colleagues (79). B-CIMP tumors were positive for ER, PR, and HER2 expression and were enriched for the luminal B gene expression subgroup as well as epigenetic silencing of genes in the Wnt-signaling pathway (78), as has also been described for colorectal tumors (34).

Endometrial carcinoma CIMP. Endometrial cancers are classified into two groups: the type I endometrioid tumors that are hormone receptor positive with good prognosis and the type II serous tumors that mostly occur in older women and correlate with poor outcome. TCGA identified four DNA methylation subgroups of endometrial tumors (80), with one subgroup displaying a CIMP-like (E-CIMP) DNA hypermethylation profile. E-CIMP was previously identified by Whitcomb and colleagues (81). Similar to colorectal CIMP tumors, the E-CIMP tumors were hypermethylated, MSI positive due to MLH1 promoter DNA hypermethylation, and did not contain TP53 somatic mutations or extensive SCNAS (34, 80). However, E-CIMP tumors did not harbor BRAFV600E or IDH1 mutations, as described in colorectal and glioma CIMP tumors, respectively, pointing to an alternative mechanism of CIMP-specific DNA methylation in endometrial tumors.

Similarities and differences among CIMPs of individual human cancers

TCGA has confirmed the presence of CIMP in colorectal, breast, and endometrial cancers, while providing the first comprehensive view of G-CIMP. However, the molecular events that result in CIMP-specific DNA methylation events for each tumor type remain unclear and suggest that universal presentation of CIMP is not apparent across tumor types. For instance, colorectal CIMP is associated with BRAFV600E mutation, but G-CIMP is tightly linked to IDH1 mutations. However, B-CIMP and E-CIMP tumors are not associated with these gene mutations. Moreover, colorectal and endometrial CIMPs are associated with MSI via epigenetic silencing of MLH1, but this was not identified in CIMPs from other tumor types. Therefore, CIMP-specific DNA methylation targets may be largely non-overlapping with no consensus CIMP DNA methylation signature across tumor types, suggesting that CIMPs may be manifested by several molecular pathways and diverse sets of genomic alterations. However, formal analyses to determine the extent of common and unique CIMP-spe-
cific DNA hypermethylation signatures between tumor types are still needed and are currently in progress by TCGA researchers.

Normal-like DNA methylation subgroups of human cancers
Unsupervised analyses of TCGA breast and endometrial cancers each identified subgroups of tumors with normal-like DNA methylation profiles. In breast cancers, this subgroup showed enrichment with the basal-like (triple-negative) gene expression group, which in which ER, PR, and HER2 were not expressed. In endometrial tumors, the normal-like DNA methylation subgroup was enriched in the serous-type tumors. Both subgroups displayed TP53 somatic mutations and extensive SCNAS and overall were similar to serous ovarian tumors (78, 80, 82). These findings indicate that these subtypes may share a common mechanism of epigenomic changes in tumorigenesis separate from their CIMP counterparts. CIMP-positive tumors generally present with genomic stability and the absence of TP53 mutations, with the exception of G-CIMP tumors, which show enrichment for TP53 mutations.

CDKN2A inactivation in lung squamous cell carcinoma
TCGA Research Network has recently reported the integrated genomic characterization findings of lung squamous cell carcinoma (LUSC) (83). Overall, LUSCs displayed increased rates of DNA sequence alterations compared with other tumor types. Among the genes that were significantly mutated, CDKN2A alterations are of interest. The CDKN2A locus encodes for two well-characterized cell cycle–regulating proteins, p16<sub>INK4A</sub> and p14<sub>ARF</sub>, that are encoded by alternate splicing of an overlapping set of exons in the locus. p16<sub>INK4A</sub> inactivates cyclin-dependent kinases that phosphorylate the retinoblastoma protein, resulting in a G1 phase arrest, while p14<sub>ARF</sub> stabilizes p53 by inducing MDM2 degradation and blocking its function (84). TCGA showed that CDKN2A inactivation occurs in over 70% of LUSCs by a mutually exclusive combination of epigenetic silencing, inactivating mutations and deletions, with each contributing to CDKN2A gene expression alterations, suggesting that multiple routes of CDKN2A inactivation are important to LUSC tumorigenesis.

BRCA1 epigenetic silencing in serous ovarian cancers
Integrative analyses of DNA methylation and gene expression data identified 168 epigenetically silenced genes in TCGA serous ovarian tumors (82). These genes display promoter DNA hypermethylation together with reduced gene expression. Among those is BRCA1, an important driver of carcinogenesis that encodes for a protein involved in repairing DNA double-strand breaks. Interestingly, TCGA demonstrated that BRCA1 was inactivated by mutation and epigenetic silencing in a mutually exclusive manner in ovarian tumors (82). Although neither BRCA1 mutation nor epigenetic silencing correlated with prognosis (85), BRCA1 epigenetic silencing was more abundant than somatic mutations in ovarian tumors, and patients with ovarian cancer with BRCA1 epigenetic silencing were younger than those with BRCA1 mutation inactivation, suggesting that epigenetic alterations of BRCA1 are important early events in ovarian tumorigenesis.

Chromatin modifier gene mutations affect the cancer methylome
Somatic mutations of chromatin-modifier genes have been reported in several human cancers (reviewed in ref. 5). While the effects of these mutations on the cancer methylome have not been completely explored, TCGA AML tumors displayed substantial levels of mutations in chromatin-modifier genes, with 30% of AML tumors harboring somatic mutations in chromatin modifier genes and 44% of tumors with mutations in DNA methylation–related genes. AML tumors with fusion events involving the myeloid/lymphoid or mixed-lineage leukemia (MLL) gene family or concordant mutations in nucleophosmin (NPM1), fms-related tyrosine kinase (FLT3), and DNA (cytosine-5-)–methyltransferase 3 alpha (DNMT3A) displayed DNA hypomethylation (75). The MLL gene family encodes histone lysine methyltransferases, and DNMT3A codes for a de novo DNA methyltransferase. These findings point to a strong link between cancer genetics and epigenetics that will be better understood by forthcoming mechanistic studies.

TCGA also identified somatic mutations of SET domain containing 2 (SETD2) in 11.5% of KIRC tumors (86). SETD2 is a H3K36 methyltransferase, and H3K36me3 marks are associated with DNA methylation at actively transcribed gene regions (87). Recent work from Chantalat and colleagues has also identified H3K36me3 marks in facultative and constitutive heterochromatin (88). TCGA demonstrated that KIRC tumors with SETD2 mutations displayed DNA hypomethylation at non-promoter regions that are marked by H3K36me3 and DNA hypermethylation at non-H3K36me3 occupied regions in normal kidney cells, suggesting a crosstalk between DNA methylation and H3K36 methylation in shaping the epigenome.

Future directions
Several interesting applications utilizing TCGA DNA methylation data are possible. For instance, TCGA data can be analyzed to delineate between driver and passenger DNA methylation events in carcinogenesis. Also of interest is determining an understanding of DNA methylation changes across several cancer types, not only in relation to CIMP, but also in identifying specific cancer-associated DNA methylation alterations for potential diagnostic purposes. Such DNA methylation–based biomarkers of disease can be used to identify primary tumors as belonging to specific subgroups and may have utility in personalized medicine. In addition, DNA methylation biomarkers can be used as tools for early detection of cancer in patient blood and a mechanism to track response to therapy.

TCGA has been and will continue to be a widely used discovery and validation resource for other genomics-based projects. TCGA will also continue to inspire the development of novel informatics approaches for integrative, high-level summaries of a wide range of molecular data types. With the tremendous volume of genomic data generated by TCGA efforts, functional experiments are needed to fully characterize cancer genomes. Currently, TCGA Research Network is completing genomic characterizations across several cancers to identify epigenomic signatures that will provide a powerful view into the molecular events that shape human carcinogenesis.

Acknowledgments
I thank Peter A. Jones and Peter W. Laird for helpful discussions and constructive comments.

Address correspondence to: Daniel J. Weisenberger, Department of Biochemistry and Molecular Biology, University of Southern California, 1450 Biggy Street, NRT G511, Los Angeles, California 90089-9601, USA. Phone: 323.442.7786; Fax: 323.442.7880; E-mail: weisenbe@usc.edu.

The Journal of Clinical Investigation http://www.jci.org Volume 124 Number 1 January 2014 21
46. Figueroa ME, et al. Leukemic IDH1 and IDH2


