ANNUAL REVIEWS

Annual Review of Pathology: Mechanisms of Disease

DNA Methylation Profiling: An Emerging Paradigm for Cancer Diagnosis

Antonios Papanicolau-Sengos and Kenneth Aldape

Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland 20892, USA; email: antonios.papanicolau-sengos@nih.gov, kenneth.aldape@nih.gov

Annu. Rev. Pathol. Mech. Dis. 2022. 17:295–321

First published as a Review in Advance on November 4, 2021

The *Annual Review of Pathology: Mechanisms of Disease* is online at pathol.annualreviews.org

https://doi.org/10.1146/annurev-pathol-042220-022304

Keywords

cancer, DNA methylation, methylation profiling, classification, machine learning, circulating tumor DNA

Abstract

Histomorphology has been a mainstay of cancer diagnosis in anatomic pathology for many years. DNA methylation profiling is an additional emerging tool that will serve as an adjunct to increase accuracy of pathological diagnosis. Genome-wide interrogation of DNA methylation signatures, in conjunction with machine learning methods, has allowed for the creation of clinical-grade classifiers, most prominently in central nervous system and soft tissue tumors. Tumor DNA methylation profiling has led to the identification of new entities and the consolidation of morphologically disparate cancers into biologically coherent entities, and it will progressively become mainstream in the future. In addition, DNA methylation patterns in circulating tumor DNA hold great promise for minimally invasive cancer detection and classification. Despite practical challenges that accompany any new technology, methylation profiling is here to stay and will become increasingly utilized as a cancer diagnostic tool across a range of tumor types.

INTRODUCTION

The goal of histological and molecular cancer pathology is to subdivide cancer into clinically relevant categories to guide treatment and assess prognosis. Currently, standard classification methods include light microscopy, immunohistochemistry, fluorescence in situ hybridization (FISH), DNA/RNA sequencing, and a very limited amount of DNA methylation testing. Although whole exome/genome sequencing, whole transcriptome sequencing, and whole methylome sequencing methods exist and could potentially be used to classify cancer in a routine clinical setting, these methods are technically difficult, require substantial resources, and produce a tremendous amount of data that can be challenging to analyze. Recent technological advances have made large-scale DNA methylation testing possible, permitting characterization of a portion of the cancer methylome in preanalytically, analytically, and postanalytically manageable ways.

Methylation of the cytosine component of DNA in cytosine-phosphate-guanine (CpG) dinucleotides (a 5′ cytosine followed by a 3′ guanine) is a crucial biological mechanism. CpG dinucleotides are overrepresented in gene promoters (1) and their methylation can silence the respective genes (2). In addition, CpG dinucleotides are located in other parts of the genome and are also subject to methylation (3). CpG methylation is an epigenetic feature, a phenomenon defined as "a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" (4, p. 781). De novo DNA methylation is mediated primarily by DNMT3A and DNMT3B (5) and maintained by DNMT1 through DNA replication (6, 7), which is the basis of the heritability and clinical utility of DNA methylation. DNA methylation is a highly stable molecular feature that can be reliably detected in tissue, including cancer (8–10).

Cancer has a complex methylation profile that, when compared with nonneoplastic tissues, is characterized by overall DNA hypomethylation (11) but by DNA hypermethylation at CpG islands. Promoter CpG island hypermethylation and gene body demethylation correlate with gene silencing (12–14). However, the effect of DNA methylation on cancer gene expression has substantial complexity that does not fit in a simple model (15). Patterns of DNA methylation are tissue specific (16), a phenomenon that extends to cancer. This property gives rise to tumor-specific DNA methylation profiles, which are the basis of DNA methylation as a tool for cancer classification (17).

Although sampling of tumor tissue is required for clinical-grade cancer classification, the presence of circulating cancer DNA (18), in combination with cancer-specific molecular profiles, makes blood an attractive source for cancer detection, diagnosis, and classification. The use of DNA methylation signatures in blood and other bodily fluids provides some advantages over the detection of somatically acquired cancer-specific single nucleotide variants for cancer detection or classification. A DNA methylation signature that is based on thousands of CpG sites can remain robust even when not all the features are detected. In contrast, an approach that relies on the presence of cancer DNA mutations is likely to suffer from low sensitivity due to the generally low number of somatic mutations in cancer. New highly sensitive DNA methylation interrogation technologies make a minimally invasive approach to cancer classification a viable strategy.

In this review we describe DNA methylation interrogation technologies, examine current knowledge of their use to classify cancer, and comment on their promise and limitations. The focus is not on the methylation status of specific genes but rather on the concept of DNA methylation patterns as differentiators between distinct cancer classes. In the rest of this article we refer to DNA methylation as methylation.

ANALYSIS OF METHYLATION STATUS

A common method of methylation analysis is based on bisulfite conversion of DNA, a method that converts unmethylated cytosine to uracil but leaves methylated cytosine intact (19–21).

Determination of methylation patterns from bisulfite-treated DNA can be accomplished in many ways including, but not limited to, methylation-specific polymerase chain reaction (MSP) (22), methylation-sensitive single nucleotide primer extension (Ms-SNuPE) (23), pyrosequencing (24), methylation-specific restriction endonuclease analysis (25), and whole genome bisulfite sequencing (WGBS) (26). We describe a small selection of these methods with a focus on BeadChip technology, which has been successfully used for research and clinical applications.

In MSP, primer design is exploited to determine methylation status. Whether an amplification product is made depends on if the template DNA anneals with a labeled primer designed to bind an initially methylated sequence or a labeled primer designed to bind an initially unmethylated sequence. The methylation status of the locus can be determined by the presence or absence of a polymerase chain reaction product (22).

Ms-SNuPE is an alternate technique based on designing primers that anneal one nucleotide short of the locus of interest on the template DNA. A labeled single nucleotide added in the subsequent nucleotide extension reaction allows for the determination of the methylation status. For example, if a thymine or adenine is added, it can be deduced that this locus was unmethylated, while cytosine or guanine would be added if the opposite were true (23).

The creation of arrays using the principles of bisulfite conversion along with hybridization to an array of probes allowed for large-scale interrogation of methylation. Commonly used arrays include Illumina's GoldenGate, 27K, 450K, and 850K (Illumina Methylation EPIC) BeadChip arrays, which have been iteratively designed to interrogate an increasing number of mostly CpG methylation sites. The GoldenGate interrogates 1,536 CpG sites while the 27K interrogates 27,578 CpG sites in promoter regions (27, 28). Promoters constitute 41% and 25.8% of the interrogated sites in the 450K and 850K, respectively. The rest of the sites are divided between gene body, intergenic, and 3′ untranslated regions (29, 30). The 27K and 450K use Infinium I technology, while the 850K also uses Infinium II technology (29). The formalin-fixed, paraffin-embedded (FFPE) DNA requirement adequate for classification purposes has been reported to be as low as 50 ng (31).

The signal produced by BeadChip arrays is fluorescence. Similar to MSP, Infinium I consists of beads covered by M or U probes that are, respectively, specific for methylated or unmethylated versions of a locus. A single nucleotide extension and release of light from the fluorescent molecule associated with that nucleotide is possible only if the probe and the locus on the template DNA anneal. The intensity of the fluorescence produced in association with the paired M and U probes can then be used to evaluate the methylation status (**Figure 1*a***). Infinium II does not contain U and M probes. Instead, similar to Ms-SNuPE, Infinium II relies on probes that bind one nucleotide short of the locus of interest. The color of the fluorophore attached to the nucleotide added during single nucleotide extension is used to determine if the locus is methylated or not (**Figure 1*b***).

After single nucleotide extension, the chip is scanned to measure the intensity and specific fluorescence color emitted in association with each bead. Data are stored in paired intensity data files (i.e., in IDAT format), which separately contain green and red fluorescent signals in addition to fluorescence intensity. An IDAT file pair is approximately 20 megabytes in size for the 450K array and approximately 30 megabytes for the 850K array.

The IDAT data can be processed in various ways. Most recent studies that have used BeadChip arrays have calculated beta values (methylated probe intensity divided by the sum of methylated and unmethylated probe intensities) for each locus (32). The data are then filtered to exclude extraneous or contaminating measurements, probes with low signal quality, cross-hybridizing probes, and probes known to harbor high frequencies of single-nucleotide polymorphisms. To create a classifier, probes with a high degree of variation across the samples to be classified are prioritized. For example, the top 10,000 to 30,000 most differentially methylated probes among a cohort of

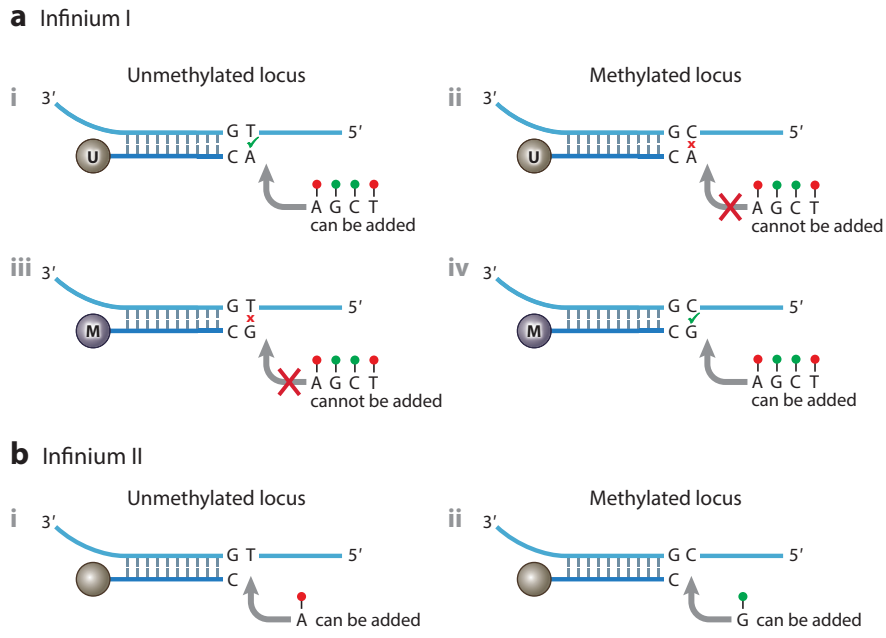

Figure 1

Comparison of Infinium I and II technologies. In Infinium I, the sequence of the DNA probe attached to U ("unmethylated") or M ("methylated") beads is reflective of the specific sequence of the interrogated DNA in its unmethylated or methylated form, respectively. A fluorescent signal is produced by the nucleotide added after the methylated/unmethylated site of interest. (*a*, *i*) In this particular example, the nucleotide at the 3′ end of the DNA probe is an adenine (A) that binds with a complementary thymine (T), indicating that the initially extracted interrogated DNA must have contained an unmethylated cytosine (C) that was converted to a uracil by bisulfite (subsequent polymerase chain reaction replaced the uracil with a thymine). Because of this 3′ match, a further nucleotide with an attached fluorophore can be added. (*a*, *ii*) An initially methylated version of the same extracted interrogated DNA would not have been usable as a template for further nucleotide additions. (*a, iii)* An initially methylated version of the same locus as the prior example would have a 3′ mismatch with the M probe because only interrogated DNA with an initially methylated cytosine (and therefore unconvertable by bisulfite) could have matched. No further 3′ nucleotides can be added. However, the same extracted initially methylated interrogated DNA can match with the 3′ guanine (G) of the M probe, making the addition of another nucleotide possible with subsequent production of a fluorescent signal (*a*, *iv*). For Infinium I, any fluorescence is considered a signal, and the specific color does not matter. (*b*) In Infinium II, there is only one type of bead (no U or M beads), and a fluorescent signal is produced by the nucleotide at the methylated/unmethylated site. As such, the specific nucleotide that is added and the corresponding fluorescent signal color do matter. In this example, the addition of an adenine suggests an unmethylated site that will fluoresce red (*b*, *i*), while the addition of a guanine suggests a methylated site that will fluoresce green (*b*, *ii*). Figure adapted with permission from Illumina, Inc.

samples are considered for further analysis. BeadChip data can also be visualized by dimension reduction methods such as t-distributed stochastic neighbor embedding (t-SNE) (33) and uniform manifold approximation and projection (UMAP) (34), which are used to generate genome-wide copy number plots (35), to determine promoter methylation status in O6-methyl-guanine-DNA methyltransferase (MGMT) (36), and to detect certain fusions such as tandem duplication in KIAA1549-BRAF (37).

Most classifiers have been generated using the random forest classifier, a machine learning method that relies on the creation of many trees that consist of randomly selected features. The

output decisions of all the trees are tallied, and the final class is defined as the most frequently voted output (38). The output is a calibrated score that reflects the confidence of a call. The cutoff for what is an acceptable score varies. Recent classifiers have defined the threshold as at least 0.9 (39, 40), although for the central nervous system (CNS) classifier a 0.84 cutoff provides the best balance between sensitivity and specificity (35, 39) and has been used in a clinical setting (35, 41) (**Figure 2**).

WGBS is the most comprehensive approach to methylation but is expensive and produces an extremely large amount of data that are challenging to analyze (26, 27). WGBS has so far not been used for cancer classification purposes.

The above methods depend on bisulfite conversion, which is inherently destructive to DNA. Affinity capture-based DNA immunoprecipitation does not need bisulfite conversion. This method immunoprecipitates, amplifies, and sequences strands of DNA that contain methylated cytosines. Due to the absence of bisulfite conversion, this method allows for DNA inputs as low as 1 ng, making it ideal for interrogation of circulating cell-free DNA (cfDNA) methylation. A major limitation of this method is that it does not have single-nucleotide resolution. Instead, it allows for the identification of at least 100 base pairs of DNA with methylation (42). Methylation-sensitive enzymatic cleavage of DNA is another bisulfite-independent method (43). These methods are currently being studied for clinical methylation applications.

TUMOR METHYLATION DETECTION IN BLOOD

Detecting cancer in circulating blood is, as a procedure, minimally invasive and therefore highly desirable in the clinical setting. Complex nonspecific clinical workups can be avoided through the use of a blood-based marker that can identify not only the presence of cancer but also the specific cancer type (44). This would allow for a simple algorithm of requesting additional confirmatory studies such as imaging and a biopsy.

A recent study tested cfDNA from female cancer patients and healthy donors (45). A pancancer panel of *APC*, *FOXA1*, and *RASSF1A* gene promoters was used to detect, but not classify, breast, colorectal, and lung cancer and achieved 72% sensitivity and 74% specificity. A cancer-specific panel was also designed. The gene promoter, corresponding cancer, and sensitivity and specificity of each promoter for the detection of that cancer by interrogating cfDNA methylation were as follows: *SCGB3A1* for breast cancer with sensitivity 16.8% and specificity 80%, *SEPT9* for colorectal cancer (CRC) with sensitivity 11.1% and specificity 98.9%, and *SOX17* for lung cancer with sensitivity 39.4% and specificity 85.1% (45). The same group performed an equivalent study on male cancer patients and healthy donors (46). A pancancer panel of *FOXA1*, *RARbeta2*, and *RASSF1A* promoters to detect, but not classify, prostate cancer, CRC, and lung cancer achieved 64% sensitivity and 70% specificity. From the set of genes interrogated in this study, only the *SEPT9* and *SOX17* promoters were found to be methylated in all three cancers. The sensitivity of each methylated gene to detect the presence of a particular cancer was quite low, ranging from 8% for the detection of CRC by the methylated *SEPT9* promoter to 29% for the detection of lung cancer and prostate cancer by the methylated *SOX17* promoter. However, specificity of methylated *SEPT9* and *SOX17* promoters for the detection of the three cancers was very high, ranging from 93% to 100%. The authors also discovered that *APC/RASSF1A* promoter methylation is an independent predictor of worse disease-specific mortality in lung cancer (46).

A probabilistic method of cancer classification using cfDNA methylation has also been described. This relies on CpG clusters with a sufficiently large difference of methylation between plasma from patients without cancer (obtained from publicly available WGBS data) and tumors in The Cancer Genome Atlas (TCGA) database (mostly 450K data from breast cancer, colon cancer,

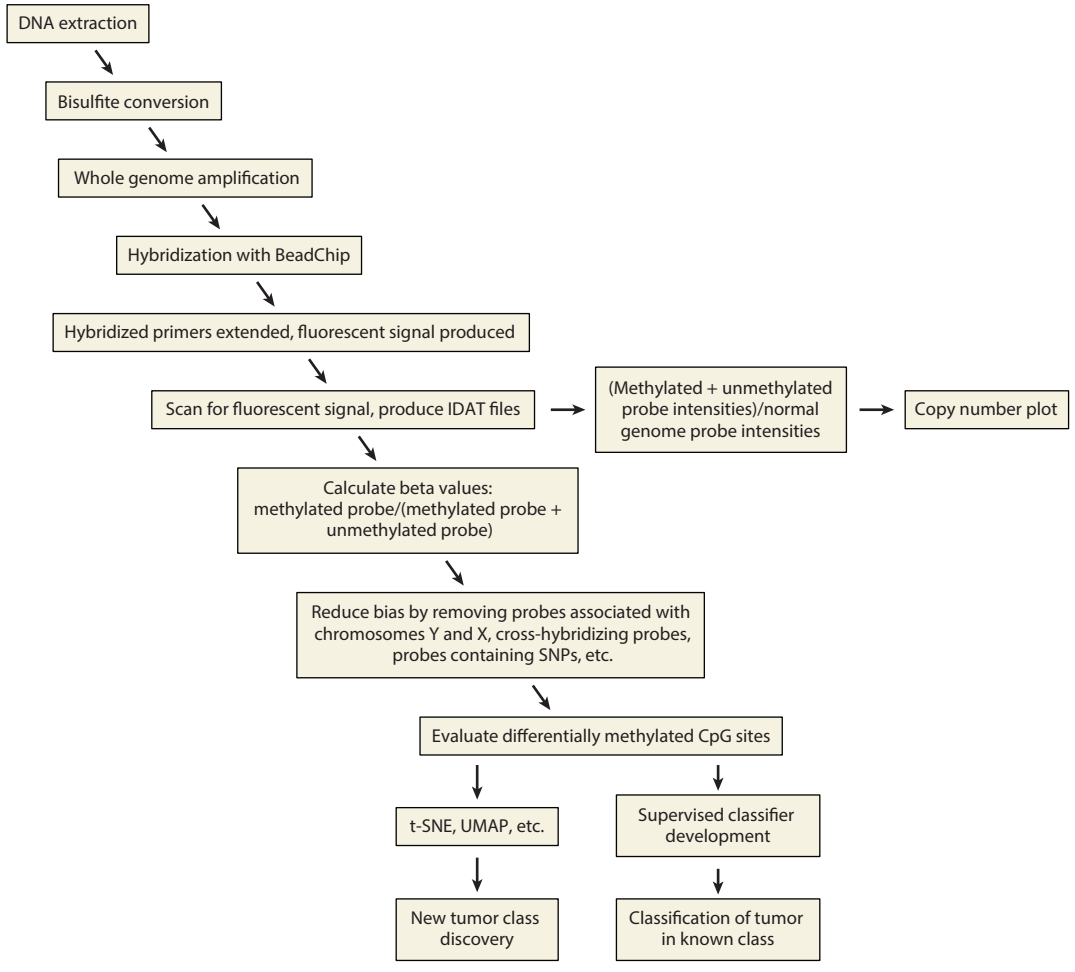

Figure 2

Tumor sample processing, BeadChip workflow, and methylation data analysis. The extracted DNA is bisulfite converted and amplified (uracils are replaced by thymines in this step). After BeadChip hybridization and primer extension, a fluorescent signal is produced and scanned by a high-resolution scanner. For a given methylation locus, the ratio between the methylated version divided by the sum of the methylated and unmethylated versions is the beta value. Probes need to be filtered on the basis of the needs of the study. For classification of cancer by methylation status, the probes that are highly informative need to be kept for further analysis. The 10,000 to 30,000 probes that display the most variable profile across the tumors included in the study are usually sufficient to capture enough of the variation between the tumors to generate distinct tumor profiles. Unsupervised methods such as t-SNE or UMAP can be used to visualize the clustering of tumors with a known diagnosis, and the positioning of a new sample can be evaluated. In the context of expanded tumor cohorts, these methods can also be used to identify new clusters that may constitute new tumor classes. A supervised classifier can be trained with specimens of known diagnoses. New specimens can then be analyzed by the classifier. The methylation data can also be used to create copy number plots. The fluorescence intensity of the methylated plus unmethylated probes for a given CpG site are added and divided by the corresponding intensities of a normal genome. This ratio gives a relative copy number for the site of interest. The inability to evaluate whole genome gain or loss is a limitation of this approach (35). Abbreviations: CpG, cytosine-phosphate-guanine; IDAT, intensity data file format; SNP, single-nucleotide polymorphism; t-SNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection.

clear cell renal cancer, papillary renal cancer, hepatocellular carcinoma, lung adenocarcinoma, and lung squamous carcinoma). The model was trained to classify cfDNA from cancer patients as originating from one of five cancer sites (breast, colon, kidney, liver, and lung cancer) or noncancer cfDNA. Critically, this model accounts for cfDNA methylation distributions that can vary on the basis of various mixtures of noncancer and cancer cfDNA. When using cancer-patient cfDNA and cfDNA from patients without cancer (from both publicly available and author-performed WGBS studies), this method's error rate of classification was 27%. This result was a substantial improvement from the error rate of classifications achieved by random forest and support vector machine (SVM) methods, which had error rates of 65% and 60%, respectively (47).

Cell-free methylated DNA immunoprecipitation and high-throughput sequencing data from blood can be used to distinguish gliomas from extracranial cancers and specimens from healthy individuals with an area under the curve (AUC) >0.9. This method has been further used to construct t-SNE methylation profiles, resulting in clusters specific to intracranial tumors [isocitrate dehydrogenase (IDH)-mutant and wild-type (WT) glioma, low-grade glial neuronal tumors, meningioma, and hemangiopericytoma] and metastases of unknown origin primary to the brain (42, 48).

A recent study investigated the possibility of a pancancer diagnosis from cfDNA using a focused bisulfite sequencing panel (44). A plasma cohort from 2,482 untreated cancer patients with more than 50 types of cancers and 4,207 individuals without cancer was separated into training and validation sets. The validation set had a sensitivity of 43.9% in all stage I–III cancer types, a sensitivity of 67.3% in the 12 common cancer types at stages I–III, and a specificity of 99.7%. Most importantly, 96% of cases with a circulating cancer signal received a prediction of site of origin, and 93% of these predictions were correct (44).

METHYLATION CLASSIFICATION IN SOLID TUMORS

Astrocytic Tumors

Methylation of the *MGMT* promoter is the first methylation marker known to predict clinical response to a chemotherapeutic agent. The *MGMT* gene codes for MGMT, which is critical for the repair of alkylated DNA. Functional MGMT can counteract the effects of alkylating agents (49), and *MGMT* promoter methylation silences the gene (50) rendering gliomas vulnerable to alkylating agents (51). Patients with primary gliomas, most prominently anaplastic astrocytoma, glioblastoma multiforme (GBM), and anaplastic oligodendroglioma, who had *MGMT* promoter methylation and were treated with the alkylating agent carmustine or temozolomide had significantly improved progression-free survival (PFS) and overall survival (OS) compared with their *MGMT* promoter unmethylated counterparts (51–53). *MGMT* promoter methylation is a good prognostic, but not predictive, marker in IDH-mutant tumors and is chemotherapy response predictive in IDH-WT high-grade gliomas (51–54).

Analysis of glioblastoma from TCGA led to an interesting and previously unsuspected insight that linked a specific somatic cancer mutation with widespread changes in methylation. In a study of 272 GBMs, 9% of cases had a methylated profile that was called a glioma CpG island methylator phenotype (G-CIMP) (55). It was found that these patients were relatively young and had a significantly improved survival rate after multivariate analysis and that their G-CIMP tumor status almost exactly matched the presence of *IDH1* mutations in codon 132. This finding became even more clinically relevant upon examination of diffuse lower-grade gliomas (LGGs), most of which have *IDH1* or *IDH2* mutations as well as concomitant G-CIMP-positive status. In an expanded cohort that included LGGs, lower grade was correlated with a higher proportion of G-CIMP status, and G-CIMP presence was an independent predictor of survival. Paired primary and recurrent

gliomas retained their methylation profiles over time, confirming methylation as a stable marker. This study confirmed that IDH mutation in gliomas is coincident with G-CIMP and that both biomarkers are associated with an improved prognosis, including in a subset of morphologically unequivocal glioblastomas (55). Subsequent work determined that *IDH1* mutation is associated with methylome remodeling, including DNA hypermethylation, causing G-CIMP (56).

Adult and pediatric GBMs were further subclassified in a study that included 210 GBMs (57). The authors used hierarchical unsupervised clustering of methylation in addition to DNA sequencing, copy number changes, and gene expression to divide GBMs into IDH-mutated, H3–3A K27-mutant, H3–3A G34-mutant, mesenchymal, receptor tyrosine kinase (RTK) I, and RTK II classes. The IDH-mutated group had the best survival, was highly methylated, and was mutually exclusive with the K27-mutant and G34-mutant groups, which were the most unmethylated. The RTK I group was characterized by an overrepresentation of *PDGFRA* amplification; the RTK II group was characterized by the adult GBM pattern of chromosome 7 gain/chromosome 10 loss, *EGFR* amplification, and *CDKN2A* loss; and the mesenchymal subset had fewer copy number changes, no distinct mutations, and a methylation profile reminiscent of nonneoplastic brain. K27-mutant cases had a mostly midline location, while the rest of the cases were usually in the cerebrum (57). Further work has emphasized the clinical outcome overlap between IDH-WT glioblastomas and lower-grade IDH-WT astrocytoma (58, 59), culminating in the proposal to create a category of "diffuse astrocytic glioma, IDH-WT, with molecular features of glioblastoma, WHO grade IV," or molecular glioblastoma (60, 61).

A study of 92 choroid plexus tumors (62) identified three methylation groups: a pediatric, mainly supratentorial, low-risk group 1; an adult, mainly infratentorial, low-risk group 2; and a pediatric, mainly supratentorial, high-risk group 3. All three groups contained choroid plexus papillomas and atypical choroid plexus papillomas (aCPPs), but only group 3 contained choroid plexus carcinomas (CPCs). As expected, CPCs were associated with a relatively short PFS and OS. All tumor-related deaths were in group 3. Although World Health Organization (WHO) grade was an independent predictor of PFS/OS and methylation clustering was not, all recurrent aCPPs were in cluster 3 (62).

Further methylation studies in various CNS tumor types have expanded findings to include the discovery of two clinical groups of diffuse leptomeningeal glioneuronal tumors (63), the identification of infratentorial IDH-mutated astrocytomas as a glioma class (64), disambiguation of papillary glioneuronal tumors (65), and prognosis subsets of primary pineal parenchymal tumors (66).

Meningioma

The methylation of meningiomas has been extensively studied for classification as well as for estimation of prognosis. An initial study showed that unsupervised clustering segregated meningiomas into two groups with distinct recurrence-free survival (67). A subsequent report (68) examined 497 meningiomas and 309 other extra-axial skull tumors that might histologically mimic meningioma variants. It was shown that methylation profiles could indeed distinguish meningioma from other types of extra-axial tumors and mimickers. Matched primary and recurrent meningioma from the same patients had a similar methylation profile, indicating stability of the methylation profile over time. Two main meningioma methylation clusters were identified and were further subclassified into six classes. These classes were characterized by incorporating mutation, copy number, and PFS profiles. This characterization resulted in three benign classes that were highly enriched for grade I meningiomas, two intermediate prognosis classes that had a mixture of grade I–III meningiomas, and one malignant class that was highly enriched for grade III meningiomas. Grade II meningiomas were scattered in all classes. Mutational and copy number

profiles were enriched in specific classes, although significant mixing of these profiles was seen. For example, the two main methylation clusters had *NF2* mutations. Comparing PFS by WHO grade versus PFS by methylation classes suggested that methylation was a better PFS predictor. The methylation classes were predictive of PFS and disease-specific survival in an independent meningioma cohort (68). To address the difficulty of predicting meningioma recurrence, the same study authors developed a methylation-based meningioma 5-year recurrence-free survival predictor that divided meningiomas into high and low risk. Through the use of one discovery and three validation cohorts from multiple institutions, the predictor was found to independently predict recurrence risk when compared with WHO grade, degree of resection, and copy number alterations burden. On the basis of these findings, a nomogram allowing for a specific 5-year risk assignment was created (69).

Recent work has identified a distinct methylation profile in clear cell meningiomas and, on the basis of clinical outcomes, provided justification to consider them as WHO grade II tumors (70).

The abovementioned studies suggest that methylation profiling may be a better predictor of outcomes than the morphology-based WHO classification. In addition, the substantial genetic heterogeneity of the methylation classes suggests that the cell precursors may be better reflected by the methylation profile than by the mutational profile.

Embryonal Central Nervous System Tumors

The medulloblastoma (MB) WNT, SHH, group 3, and group 4 classes were initially identified using expression profiling (71, 72). An array methylation study (9) recapitulated these four classes in the majority of 100 frozen MBs. To further test the capacity of methylation to classify MBs, an SVM classifier was trained using the methylation data of these 100 cases. Application on 130 FFPE MBs resulted in confident classification in 126 out of 130 cases. Similar to that of the expression-based classification, the methylation WNT group was found to have a favorable prognosis. In the non-WNT set, methylation status of the *MX11* and *IL8* genes was also found to be an independent predictor of prognosis (9). In a subsequent study (31), methylation and expression data class assignment were found to be 95.3% concordant in a set of 107 frozen tissue specimens. A 48-CpG signature was used to train an SVM classifier on the frozen tissue specimens and was tested on 169 FFPE MBs with 97.6% concordance. An additional 56 specimens with both methylation and expression data had concordant methylation and expression classes in 94.6% of specimens (31). Further studies confirmed that methylation is a robust method that can be used for MB classification (73–76).

Methylation has been used to further subclassify childhood MBs on the basis of the methylation profile of 428 primary specimens and a validation set of 276 specimens. The WNT subset remained unchanged, but the other four classes were segregated into two classes each. This new methylation-based classification was used to create four risk stratifications for MB patients: favorable, standard, high, and very high (77).

A study of 192 atypical teratoid/rhabdoid tumors (ATRTs) identified three epigenetic groups with distinct demographic, expression, methylation, and *SMARCB1* mutation characteristics. ATRT-TYR was characterized by melanosomal marker overexpression, ATRT-SHH by SHH-pathway activation, and ATRT-MYC by MYC expression. Notably, ATRT-TYR and ATRT-SHH were found to be relatively hypermethylated. These clusters were confirmed by WGBS in a subset of cases (78). An unsupervised clustering comparison of the ATRT methylation profile against the cribriform neuroepithelial tumor (CRINET), a nonrhabdoid entity that also has *SMARCB1* loss, revealed that CRINETs classify with ATRT-TYR. However, CRINETs had a significantly better OS compared with ATRT-TYR, indicating that methylation does not necessarily capture

all features related to prognosis (79). A recently described pineal entity that arises in adults, has *SMARCB1* loss, and clusters close to ATRT-MYC tumors has been designated as a "desmoplastic myxoid tumor, *SMARCB1*-mutant" (80).

Primitive neuroectodermal tumors (PNETs) of the CNS are a heterogeneous group of rare embryonal tumors with rosettes. A methylation study proved that intraocular medulloepitheliomas and CNS embryonal tumors with multilayered rosettes have distinct methylation profiles (81). In a subsequent comprehensive methylation analysis (82), unsupervised clustering and t-SNE were performed in 323 CNS-PNETs. These were compared against reference cohorts of well-characterized CNS tumors. Sixty-one percent of the CNS-PNETs clustered with a wide variety of tumor clusters including embryonal tumors with multilayered rosettes (ETMRs), *MYCN*-amplified high-grade gliomas, RTK-subgroup high-grade gliomas, IDH-mutant high-grade gliomas, MBs, and many others. On the basis of the clustering pattern and further copy number, DNA, and RNA sequencing studies, four new classes were identified: CNS NB-FOXR2, CNS EFT-CIC, CNS HGNET-MN1, and CNS HGNET-BCOR. This study indicates that the majority of what are considered CNS-PNETs are in fact a mixture of a plethora of other tumors, and that the four novel classes, in addition to ETMRs, form a molecular-based classification of what, on the basis of current knowledge, might constitute true CNS-PNETs. Importantly, this study included 50 specimens forming various clusters that consisted of a very low number of tumors each, suggesting the existence of molecular classes that are very rare and will need characterization in the future (82).

Ependymal Tumors

Ependymal tumors are mostly pediatric entities that can occur in supratentorial, posterior fossa, or spinal locations. On the basis of RNA expression, ependymomas (EPNs) of the posterior fossa were initially demonstrated to consist of two distinct groups: aggressive group A, which is found in young patients, and group B, which is found in older patients (83). Group A was subsequently found to have DNA hypermethylation (84). To further define ependymal tumor molecular classes, 500 ependymal tumors were analyzed for methylation, copy number status, and RNA fusion status. Unsupervised clustering revealed three molecular classes within each of the spinal, posterior fossa, and supratentorial compartments. A class of subependymomas was present in each compartment. The remaining subclasses included myxopapillary EPNs and EPNs with frequent 22q loss in the spine, EPN groups A and B in the posterior fossa, and supratentorial *YAP1*- and *RELA1*-translocated classes. The *RELA1*-translocated cases had prominent chromothripsis. Compared with the WHO classification at the time, this molecular classification improved ependymal tumor prognosis stratification with very poor outcomes associated with group A of the posterior fossa and RELA-translocated supratentorial EPNs. Additionally, the methylation profile of primary and recurrent tumors was similar. The molecular classes were confirmed by RNA-expression profiling in a subset of cases (85). Methylation profiling is now considered the gold standard for subtyping of ependymal neoplasms, although other methodologies such as L1CAM immunohistochemistry in supratentorial tumors (86) and H3K27 trimethylation in posterior fossa tumors (87) also play significant roles.

CNS Tumor Classifier

The abovementioned efforts culminated in the creation of a broad functionally and clinically impactful CNS tumor classifier that combined methylation profiles of a very large reference cohort with a sophisticated machine learning approach. In this work (39), a CNS neoplasm classifier was developed using the 450K array and a 2,801-specimen reference cohort to initially create

methylation class categories via the random forest classifier. This resulted in 82 distinct neoplasm classes, many of which did not fit into well-known WHO categories, and 9 control tissue classes (**Figure 3**). A validation cohort of 1,104 specimens was used to test the classifier. Investigators scored each case for the impact of the methylation classifier by comparing the standard pathology (premethylation testing) diagnosis with the final integrated (postmethylation testing) diagnosis. Eighty-eight percent of cases had a match score of a least 0.9. Of these, 86% of cases had concordant morphology and methylation class, including a subset assigned to a molecular subgroup that could not be assigned by morphology; 13% of cases received a revised diagnosis on the basis of the methylation profile (**Figure 4**); and the remaining 1% of cases could not be reconciled. The cases scoring less than 0.9 consisted of a mixture that could constitute yet-undescribed subsets of CNS tumor entities (39). The utility of the classifier is now evident and is reflected in the revised WHO classification of CNS neoplasms, in which methylation profiles are considered in many entities. Indeed, it is likely that some entities in the next WHO classification iteration will rely solely on the methylation classifier for diagnosis.

This CNS classifier has been increasingly used at a variety of institutions. In one medical center (88), 56% of adult brain tumors had a calibrated score of at least 0.84. Twenty-five percent of the diagnoses were confirmed and 48% were refined, while 25% of the patients were given a new diagnosis. In 2% of cases, the methylation score was noncontributory. Notable findings included the observation that 41% of glial tumors with a high score but without high-grade features were classified as high-grade tumors by methylation. Other clinically impactful results included reclassifications of MB to diffuse leptomeningeal glioneuronal tumor and of anaplastic ependymoma to pleomorphic xanthoastrocytoma (88). In a second medical center, (41) 66% of adult and pediatric brain tumors had a score of at least 0.9. In this high-score set, 62% of diagnoses were confirmed, 20% of diagnoses were refined, and a new diagnosis was given for 15% of the patients, with a WHO grade change in 71% of these cases. Approximately 2% of cases had a high score but the classifier output was disregarded because it was thought to be misleading. A large subset of these disregarded reclassifications was in retrospect found to be more consistent with the clinical course of the methylation class as opposed to the original diagnosis (41). The reasons for the classification variations between institutions are complex and, at a minimum, likely include the differences of specimen types usually analyzed by these institutions. Refinement and enrichment of the CNS classifier is ongoing. As an example, rosette-forming glioneuronal tumors were recently found to form a distinct methylation cluster (89).

Sarcoma

Sarcomas are diagnostically difficult neoplasms because of their rarity and the subsequent dearth of diagnostic experience by most pathologists. In addition, only approximately 20% of sarcomas have a characteristic translocation (90). This is reflected by the significant number of sarcoma methylation classifier studies in the past decade.

In the earliest effort to classify sarcomas by methylation, 80 primary sarcomas of eight histologies and two adipose tissue samples were analyzed. Unsupervised clustering resulted in four main subgroups: a synovial sarcoma cluster, a myxoid/round cell sarcoma cluster, and two clusters consisting of a mixture of sarcomas. A random forest classifier had an accuracy of only 70%. Using a partitioning around medoids algorithm, the investigators identified eight clusters, two of which (myxofibrosarcomas and undifferentiated pleomorphic sarcomas) were combined to produce seven final clusters. Using these seven clusters, a random forest model was found to have an accuracy of 82% (91).

In a rhabdomyosarcoma study (92), methylation analysis was performed on 53 cases. Unsupervised clustering produced two alveolar sarcoma and two embryonal rhabdomyosarcoma clusters.

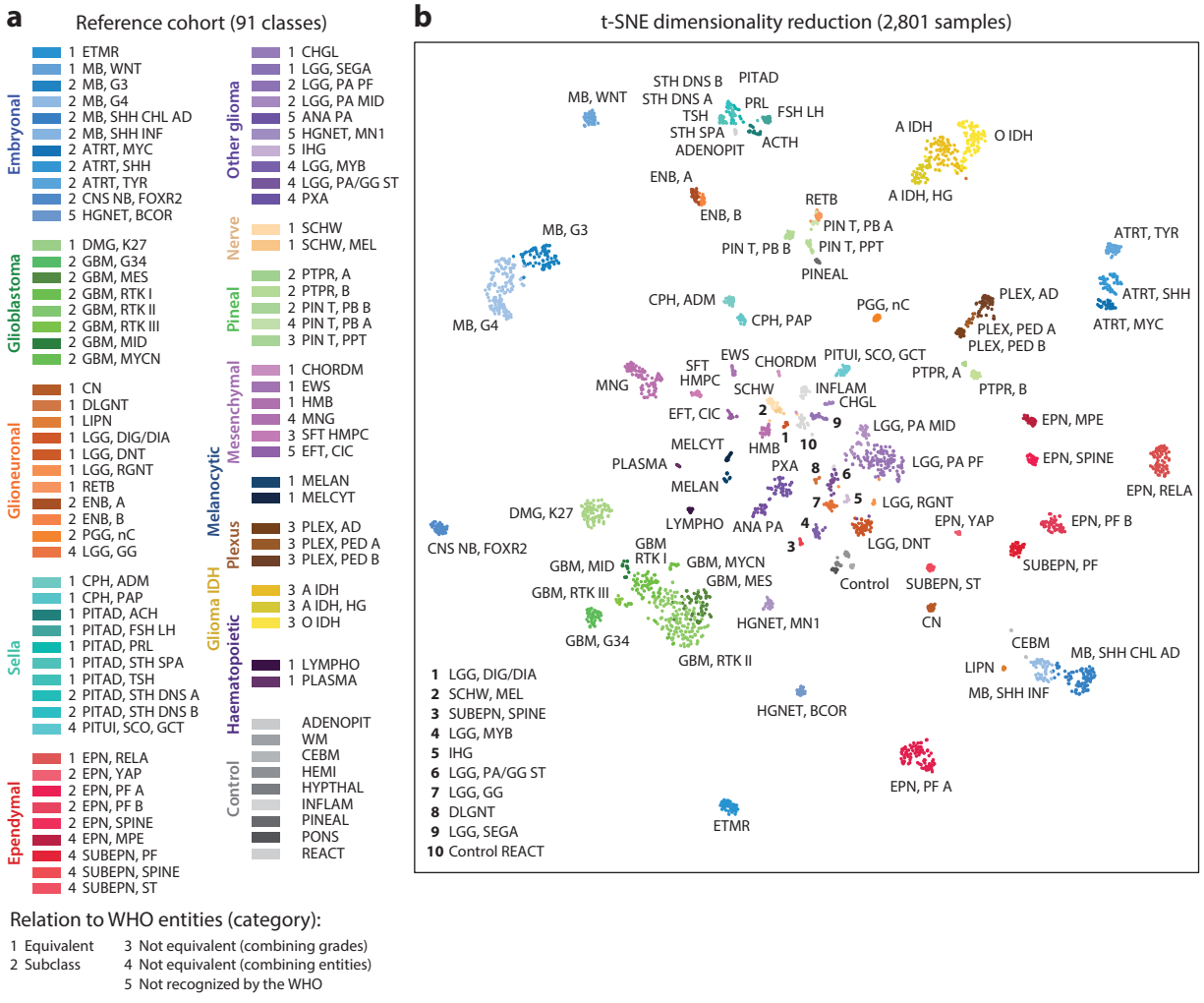

Figure 3

An annotated t-distributed stochastic neighbor embedding (t-SNE) visualization of central nervous system (CNS) tumor methylation data and World Health Organization (WHO) entity relationship with methylation classes (39). t-SNE is only one way of classifying tumor methylation data and is distinct from score-based supervised methods such as the commonly used random forest classifier. (*a*) This t-SNE visualization contains a reference set of 82 CNS tumor classes and 9 control tissues. The tumor classes have been separated by histology and molecular subtype. The relation of known WHO entities with the methylation classes varies from equivalent to not recognized by the WHO (*lower left*). (*b*) This dimensionality reduction method depends on unsupervised clustering and can cluster distinct tumor classes in a way that can be appreciated by the naked eye. Although a two-dimensional version is shown here, a three-dimensional version can also be constructed and manually manipulated to further analyze the relationship between clusters. This type of visualization can be used to identify new cancer methylation clusters with a distinct molecular profile. For such new molecular cancer entities, a score-based supervised classification approach would generate a no-match output. However, a method such as t-SNE allows for a visual identification of the cluster, further analysis of the molecular profile, and eventual inclusion into the supervised classification method so that this new cancer can be classified in the future. Note that due to space constraints, bold numbers in panel *b* are abbreviations for certain categories; they are defined in the key. Figure adapted with permission from Reference 39; copyright 2018 Springer Nature.

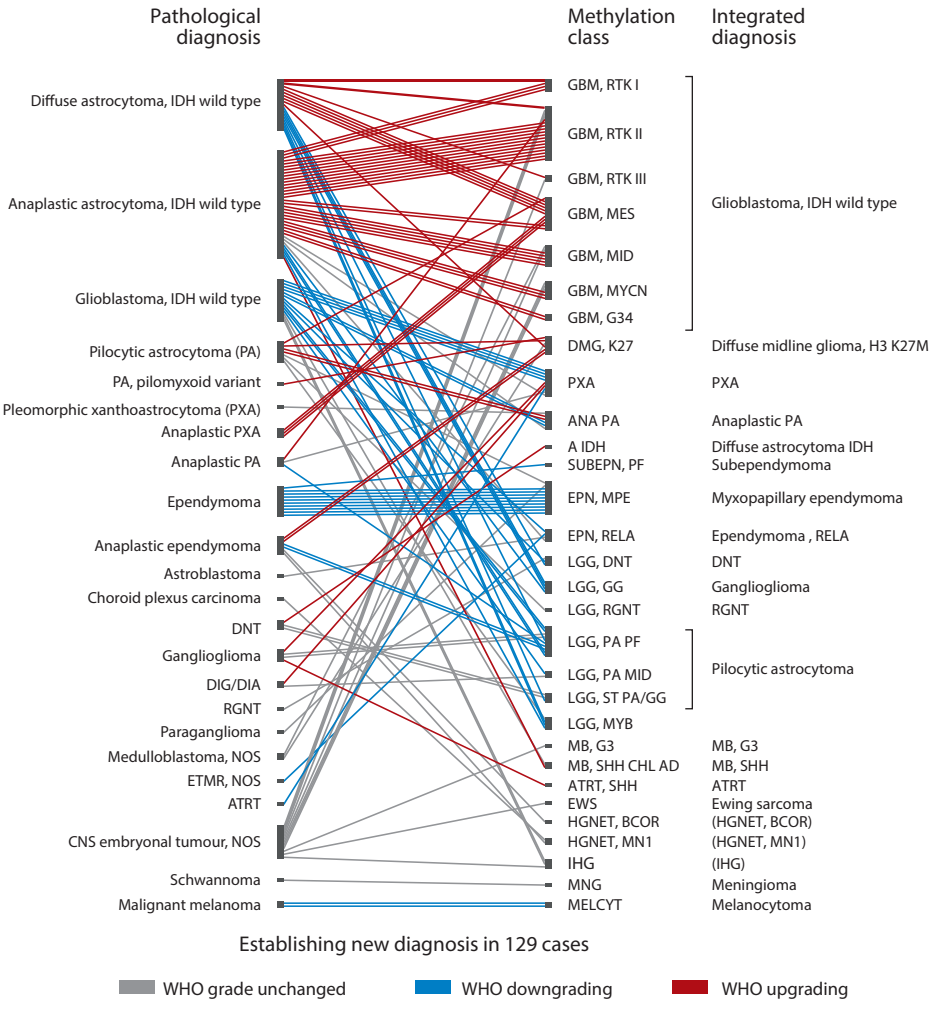

Figure 4

Subset of the central nervous system methylation classifier validation cohort. One hundred and twenty-nine of 977 cases (13%) with a high calibrated score resulted in a new diagnosis when methylation classification was considered. These revised diagnoses, including upgrading and downgrading of histological diagnoses, were in some cases highly clinically impactful. Figure adapted with permission from Reference 39; copyright 2018 Springer Nature.

These clusters correlated with distinct copy number and DNA mutation profiles. Most importantly, the two embryonal rhabdomyosarcoma clusters had a remarkably different OS.

A subsequent study investigated an unusual group of intracranial tumors (93). When compared with the methylation profile of well-defined CNS tumors and sarcomas by unsupervised clustering, a group of 22 mostly pediatric intracranial tumors formed a distinct cluster. The morphology of these 22 cases was not specific, and various diagnoses had been previously assigned. Surprisingly, 21 of 22 cases were found to have *DICER1* mutations. Germline testing was possible in five patients, two of whom had a germline *DICER1* mutation. In addition, two embryonal rhabdomyosarcomas of the uterus were found to cluster with this *DICER1*-mutated group. Findings

suggested that these tumors are part of a greater *DICER1*-mutated intra- and extracranial group of sarcomas that includes some germline *DICER1*-mutant cases. The new entity name "spindle cell sarcoma with rhabdomyosarcoma-like features, *DICER1* mutant" was proposed by the authors. This study emphasized the power of methylation to unify histologically disparate groups and identify them as a distinct molecular group.

In the first large-scale angiosarcoma methylation investigation, 36 primary and secondary (UV and radiation-induced) angiosarcomas were studied. Unsupervised clustering resulted in A and B clusters, each of which separated in two additional subclusters. Cluster A consisted almost entirely of secondary cases, had significantly more chromosomal instability, and had improved OS compared with cluster B. Cluster B contained a mixture of primary and secondary cases but contained almost all the primary cases. Cluster B1 was also found to consist of seven *MGMT* promoter methylated cases, raising the possibility that *MGMT* promoter methylation status could be used to predict response to agents such as temozolomide (94).

Methylation profiling was also applied to uterine tumors. Unsupervised hierarchical clustering and t-SNE of various epithelial and mesenchymal uterine tumors was performed. Most of the distinct diagnoses clustered in corresponding distinct clusters. Carcinosarcomas clustered with endometrial carcinomas, implying a common origin. In addition, low- and high-grade endometrial stromal sarcomas formed distinct clusters, with the high-grade cluster containing *YWHAE*- and *BCOR*-rearranged subclusters, consistent with and expanding earlier findings (95).

To classify three disparate sarcoma groups by methylation, methylation profiles of 10 Ewing sarcomas, 11 synovial sarcomas, and 15 osteosarcomas were examined. Methylation in each histology was found to be distinct upon unsupervised hierarchical clustering. The methylation data were used to train a random forest classifier, which was tested on publicly available methylation profiles of the same three histologies. Most of the validation cases were classified into their previously identified histology, including 10 out of 10 synovial sarcomas, 85 out of 86 osteosarcomas, and 14 out of 15 Ewing sarcomas. The single misclassified osteosarcoma had a relatively high expression of *EWSR1* and *ETV1* but no *EWSR1-ETV1* translocation, suggesting that this neoplasm may be related to Ewing sarcoma through a non-*EWSR1-ETV1* translocation mechanism. The authors also used the classifier on two challenging cases, one of which was confidently classified as osteosarcoma while the other could not be classified (96). A unique group of undifferentiated round cell sarcomas (URCSs) with *EWSR1-NFATc2* fusion that have a distinct methylation cluster when compared with Ewing and other sarcomas has also been discovered. This pattern suggests that URCSs with *EWSR1-NFATc2* could be an entity distinct from classic Ewing sarcomas (97).

An expansive study (98) created a reference set of 18 indisputably diagnosed sarcoma types and 10 nonneoplastic reactive soft tissues to attempt classification of 30 small blue round cell tumors (SBRCTs). The way each SBRCT visually best fit within the reference t-SNE clusters was used to assign a histology. Almost half (14 out of 30) of the SBRCTs clustered with Ewing sarcoma; six cases clustered with SBRCTs with *CIC* alteration; and the remaining cases clustered with SBRCTs with *BCOR* alteration, synovial sarcomas, malignant rhabdoid tumors, mesenchymal chondrosarcomas, and classic adamantinomas. Additional studies including immunohistochemistry, FISH, sequencing, and copy number studies confirmed that 25 out of 30 SBRCTs had been correctly classified by t-SNE.

Methylation has also been used as evidence to potentially consolidate soft tissue entities. For example, intra- and extracranial desmoplastic small round cell tumors have been found to occupy almost coincident clusters in a t-SNE, suggesting a single or two very closely related entities (99).

When considered together, these investigations strongly suggested that a comprehensive soft tissue methylation classifier similar to the one for the CNS would be possible. To achieve this, a reference cohort of 1,077 tumors consisting of 62 methylation cancer classes and

54 histological diagnoses in addition to three nonneoplastic control classes was used. Prominent findings by unsupervised analysis included the consolidation of histologies into single clusters, such as a cluster that consisted of undifferentiated sarcomas, myxofibrosarcomas, and a subset of pleomorphic liposarcomas. In addition, two methylation classes were suggestive of sarcomas that are not currently included in the WHO classification. Validation was performed on 428 sarcomas. Sixty-one percent of cases had concordant morphology and methylation class while 14% of cases were discrepant and other molecular data were used to assist in diagnosis resolution. Of these discrepant cases, 49% were revised in favor of the methylation classification, 44% could not be resolved, and 7% did not receive a revised diagnosis despite a high confidence score. Twenty-five percent of the validation set could not be confidently classified despite a generally satisfactory tissue neoplastic content, possibly indicating additional sarcoma methylation classes that were not reflected in the training set. This study is the second major effort to develop a clinically useful methylation classifier and provides strong evidence for clinical applicability (40).

Peripheral Nerve Sheath Tumors

In a study of 171 peripheral nerve sheath tumors, numerous distinct methylation profiles were detected. Most importantly, the clinically aggressive malignant peripheral nerve sheath tumors (MPNSTs) were found to have a distinct methylation profile compared with benign neoplasms such as schwannoma, nerve sheath myxoma, ganglioneuroma, and subtypes of neurofibroma. An exceptional class was the atypical neurofibroma and low-grade MPNST, which had an identical methylation profile distinct from other tumors; this finding suggests that these two tumors are closely related, a possibility that is consistent with their excellent prognosis (100). Seven tumors that were thought to be conventional sporadic MPNSTs but had no immunohistochemical loss of NF1 clustered in an unexpected fashion. Subsequent histological review, in conjunction with the clinical course that was known for some of the patients, resulted in a revised histological diagnosis of these seven tumors to three schwannomas, one solitary fibrous tumor, one biphenotypic sinonasal sarcoma, one benign tumor not further classified, and one malignant tumor not further classified (101).

Colorectal and Endometrial Adenocarcinoma

CRC serves as a prototype for methylation-based classification of cancer. In sporadic CRC, the highly methylated CpG island methylator phenotype (CIMP) is distinct from CRC with loss of heterozygosity (102) and is associated with mismatch repair deficiency, hypermutation, *BRAF* mutation, and *MLH1* promoter hypermethylation (103–106). Methylation of the *MLH1* promoter was discovered to be associated with *MLH1* silencing (107). CRC methylation classes were subsequently refined to CIMP-high with high methylation, methylated *MLH1* promoter, and *BRAF* mutation; CIMP-low with KRAS mutation enrichment; and non-CIMP (104). In addition to *BRAF* V600E mutation, *MLH1* promoter methylation is a useful marker to distinguish microsatellite instability-high (MSI-H) sporadic CRC (methylated pMLH1, *BRAF* mutant) from MSI-H Lynch syndrome–related CRC (unmethylated *MLH1* promoter, *BRAF* wild type) (108). In conjunction with sequencing codon V600 in *BRAF*, *MLH1* promoter methylation is routinely used to determine whether a microsatellite unstable, MLH1-deficient CRC adenocarcinoma is sporadic or possibly Lynch syndrome related. Additional germline testing of mismatch repair genes is required to diagnose Lynch syndrome (109). *MLH1* promoter hypermethylated, *MLH1* silenced, microsatellite unstable, hypermutated endometrial cancers—a similar population of CIMP—is a distinct subset that also needs to be worked up for the possibility of Lynch syndrome (110–112).

Kidney Neoplasms

In clear renal cell carcinoma (cRCC), there is a correlation between higher methylation and increasing stage/grade (113). Similarly, hypermethylation correlates with poor survival in cRCC, chromophobe renal cell carcinoma (chromophobe RCC), and type 1 and type 2 papillary RCC (pRCC) (114). A CIMP group with a particularly bad prognosis was identified in pRCC type 2 (115). The clinical significance of methylation-based RCC groupings was confirmed in further work, which showed that methylation status is a prognosis stratifier (116). A separate study identified a 172-CpG-site methylation-based prognosis predictor for nonmetastatic cRCC (117).

Histology-based methylation classifiers have so far been limited to the identification of genes differentially methylated in cRCC, pRCC, chromophobe carcinoma, and oncocytoma (118); methylation features distinguishing cRCC and chromophobe RCC (119); and the identification of two distinct RCC epiclusters, one of which contains cRCC, pRCC, mucinous/spindle cell carcinomas, and translocation RCC, and the other of which contains oncocytoma and chromophobe RCC (120).

Although there is currently no mature kidney cancer classifier, the abovementioned evidence suggests that methylation might be a way to identify diagnostic and prognostic molecular classes and possibly disambiguate eosinophilic renal tumors (121) and diagnoses of exclusion such as unclassified RCCs (122).

Thyroid Neoplasms

Methylation of *TIMP3*, *SLC5A8*, and *DAPK* promoters has been associated with extrathyroidal invasion, lymph node metastases, multifocality, and advanced tumor stage in papillary thyroid carcinoma (PTC), classical PTC, and tall-cell PTC. *BRAF* mutation status was significantly associated with *TIMP3*, *SLC5A8*, and *DAPK* methylation (123). Subsequent studies discovered methylation clusters of thyroid tumors, including robust clustering of PTC (124, 125), and identified a 21-methylation probe set that could distinguish poor and good prognoses in well-differentiated thyroid cancers with a sensitivity of 63% and a specificity of 92% (125). Further work developed a six-CpG classifier that could distinguish benign and malignant thyroid lesions with a 94.3% and 82.4% sensitivity and specificity, respectively, with equivalent findings in additional data sets, and 86.7% sensitivity and 89.5% specificity in fine needle thyroid aspirates (126).

Other Neoplasms

Methylation has also been used to determine whether primary lung squamous cell carcinomas (LUSCs) can be distinguished from head and neck squamous carcinomas (HNSCs) metastatic to the lung, a common dilemma in surgical pathology. A study (127) used methylation data of 528 primary HNSCs, 354 LUSCs, and 74 normal lung tissues. In t-SNE plots, normal lung formed distinct clusters while HNSCs and LUSCs formed substantially overlapping clusters. From the training set, the 2,000 most variable CpG sites were chosen, and random forest, artificial neural network, and support network machine classifiers were developed. The classifiers were tested against an independent validation data set; the neural network classifier had the best performance with 96.4% correct classification and an AUC of 0.99. Notably, the training set consisted of frozen tissue but, surprisingly, FFPE specimens were assigned to the correct class with an accuracy that was higher than that of frozen samples.

Unsupervised clustering of various other tumors has revealed methylation subgroups including CIMP clusters in urothelial bladder cancer (128) and cervical cancer (129); distinct histology profiles in testicular germ cell tumors (130) and uterine carcinosarcomas (131); distinct DNA

mutation groups in prostate cancer (132), cholangiocarcinoma (133), hepatocellular carcinoma (134), and cauda equina paraganglioma (135); and clinical outcome subsets in ovarian cancer (136) and uveal melanoma (137). Expansion of these observations has the potential to lead to further development of methylation classifiers.

The Current Methylation Classifier Status, Benefits, Caveats, and Publicly Available Data

A key feature of a cancer classifier is the ability to capture a tumor entity's biological milieu, which ideally also corresponds to clinical characteristics, natural history, and therapy response. Historically, cancer classification and diagnosis by pathologists has been achieved by visual inspection of tissue characteristics following hematoxylin and eosin staining and immunohistochemistry. These methods are based on morphology and are subject to well-described interobserver variability. Additional adjunct methods such as FISH and nucleic acid sequencing can be helpful to identify alterations specific to individual tumor entities. While helpful in many instances, DNA alterations have generally not proven to be specific enough to develop comprehensive classifiers of cancer. RNA is used for determination of specific fusions, some of which are pathognomonic for specific entities or which can be used to refine the differential diagnosis. RNA expression has been suggested as a useful tool for cancer classification, but the lability of RNA is a limiting factor (8). Advantages of methylation as a classifier include its stability and its robust measurability in FFPE tissues, the most common form of tissue processing utilized for diagnostic purposes (8, 138).

Early attempts at using methylation for cancer classification utilized a small number of methylation markers and were modestly successful in creating classifiers. The development of practical methylation arrays made it possible to interrogate thousands of methylation sites, some of which have a profile that can differentiate between different classes of cancer. Specifically, the advent of 450K has been particularly impactful because it seems to capture the minimum number of methylation sites needed to create strong classifiers. On the basis of the success of recently developed methylation classifiers, it can be stated that methylation is a biological process that reflects, to a significant degree, multiple facets of cancer including morphology, DNA mutation status, RNA expression, and, by implication, clinical behavior and therapy response.

The use of methylation as a classifier of cancer has generated new and interesting questions. In the initial description of the CNS tumor classifier, implementation of methylation profiling resulted in a new diagnosis in 12% of cases. Some of these reclassified tumors underwent an increase or a decrease in WHO grade, including clinically significant reclassifications such as from glioblastoma to pleomorphic xanthoastrocytoma and from pilocytic astrocytoma to glioblastoma (**Figure 4**) (39). We have observed such clinically significant reclassifications, which are often consistent with other molecular characteristics of the case and, critically, the known clinical behavior of the tested tumor. The existence of such cases suggests that morphology can sometimes be limiting and that the methylation profile may provide a better representation of biological behavior. It must be emphasized that amending a diagnosis can have a major impact on treatment and therefore should entail careful consideration of the specifics of the case. Furthermore, despite its strengths, methylation must not be used blindly. The morphological and standard molecular characteristics still need to be considered for each case, and the best diagnosis is nearly always one that integrates all of the relevant data. The enduring need for the standard morphology, immunohistochemistry, and FISH approach is reflected by a very small number of cases that have an unequivocal histological diagnosis but, possibly erroneously, classify in a substantially different or ambiguous methylation class (39, 40). Other issues that are occasionally encountered include specimens with a low tumor content, which can result in a methylation pattern similar to that of the contaminating nonneoplastic cells. In addition, old tissue specimens may yield suboptimal

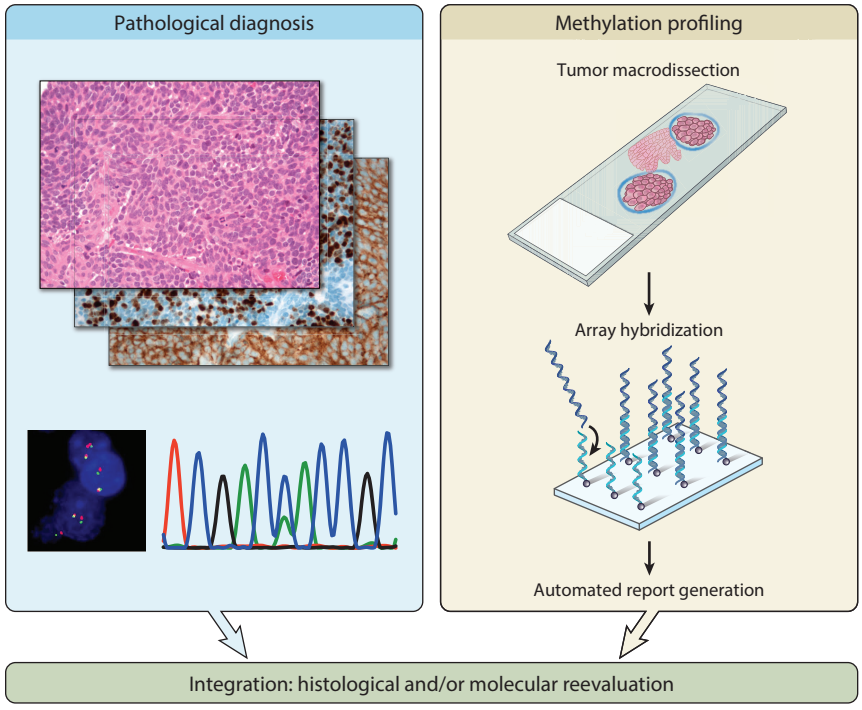

Figure 5

Integration of methylation classifiers in diagnostic workflow. Methylation can be viewed as another method of diagnosis refinement. Unlike the output of methods such as fluorescence in situ hybridization or immunohistochemistry, the output of methylation classifiers is a specific diagnosis. This can direct the pathologist toward a reevaluation of the case. Figure adapted with permission from Reference 39; copyright 2018 Springer Nature.

DNA quality that can confound accurate classification. On the basis of current capabilities and limitations, cancer methylation classifiers can be viewed as adjuncts to morphology, immunohistochemistry, and FISH studies (**Figure 5**).

The utility of methylation profiles to diagnose and classify cancers has been spearheaded by a series of studies focusing on CNS tumors (39) and soft tissue tumors (40). These studies have granularly subclassified various CNS tumor categories including MB (31), ependymoma (85), and CNS-PNET neoplasms (82). These have resulted in the segregation of tumors with similar morphology to distinct molecularly defined categories. The reverse is also true. As one example, soft tissue tumors with different histologic appearances have been shown to be molecularly identical, leading to the convergence of previously disparate entities into one (40).

Another novel aspect of methylation classifiers is the use of machine learning. Random forest is the classification machine learning method used in recent studies (39, 40), although what constitutes the best machine learning classifier method is still an open question. Recent data suggest that the elastic net method and the ridge-penalized multinomial logistic regression–calibrated SVM method with linear kernels achieve the highest AUC and minimize misclassification error (139). Additional studies and community consensus will be required to determine the best classification machine learning method.

An important aspect of classifiers is the representativeness of the training set. The methylation classification of a cancer class that was not included in the training set will likely result in no match

or a low-confidence match. The currently available studies have used an empirical minimum of eight (39) or seven (40) cases to define a methylation group. What constitutes an ideal minimum to define a molecular class remains to be seen, but there is no doubt that to achieve a complete view of the methylation classifier, a great number of tumors will have to be collected and tested. This will require interinstitutional and international collaboration with ever-expanding numbers of specimens and iteratively updated classifiers.

The most popular forms of methylation interrogation require an initial bisulfite conversion step. This chemical treatment is highly destructive to DNA and explains the substantial DNA requirement (140). Methods such as methylated DNA immunoprecipitation do not need DNA bisulfite conversion although they have different limitations including the inability to interrogate single CpG sites (42). These newer methods are not yet used in major methylation classifier studies but are likely to play a role in the development of future classifiers.

Inherent to the novelty of genome-wide methylation classifiers is the fact that current methylation interrogation technology requires instruments that most large and almost all small medical centers lack. Additional limiting factors for methylation classifiers include cost and lack of community acceptance as a standard of care in diagnostics. However, it is likely that additional classifiers will be developed for a variety of organ systems, increasing the interest and relevance of this technology, with eventual incorporation of methylation classification into diagnostic algorithms and guidelines. A possible model is a hybrid one, where only some institutions have the required machines and expertise. For next-generation DNA and RNA sequencing, such a hybrid model already exists in the form of large academic centers and private companies.

Efforts to use cfDNA methylation for classification have yielded modest results, likely because of the minute amount of tumor cfDNA and technical limitations in terms of interrogating enough methylation sites. However, new technologies are making large-scale interrogations possible and free of DNA-damaging bisulfite conversion (42). A blood-based clinical grade cancer classifier does not currently exist, although substantial inroads into the use of cfDNA methylation for cancer detection are being made (44).

TCGA and the Gene Expression Omnibus (GEO) databases contain the largest publicly accessible methylation data sets and have been used extensively to build methylation classifiers. TCGA contains a static set of GoldenGate, 27K, and 450K data for many cancers (141). Similarly, the GEO database contains cancer methylation data with new data sets uploaded directly by users as the data become available (142).

THE FUTURE OF METHYLATION AND OTHER CANCER CLASSIFIERS

The discovery of increasingly rare cancer classes and improvement of classification methods may require community acceptance of frequently updated, online-accessible classification schemes and algorithms. We anticipate the development of multiple organ-based methylation cancer classifiers and, ultimately, a pancancer classifier. To cover the entirety of the potentially discriminatory methylation sites, comprehensive approaches that will be practical in the future, such as WGBS, will provide a complete understanding of the methylome and its capacity to classify cancer. Similarly, blood-based pancancer classifiers are poised to become possible. A potential model would be to perform blood sampling from patients with symptoms suspicious for cancer. Depending on the robustness of the classifier, biopsy of the tumor itself may or may not be required. However, much work and necessary improvements in detection and standardization will be required before tumors can be reliably diagnosed using such a minimally invasive approach.

The integration of methylation profiling into the diagnostic pathology of CNS tumors has shown the potential for improved diagnostic accuracy as well as stratification of tumor subtypes

in clinically relevant ways. Among the attributes of methylation profiling, the ability to capture the substantial morphologic and genetic heterogeneity in primary CNS tumors remains its most robust and demonstrated feature. This is particularly valuable in histologically ambiguous tumor types that may harbor potentially targetable alterations. The recent implementation of the first clinical-grade methylation classifiers (39, 40) may herald a similar paradigm shift for neoplastic entities involving additional organ systems.

Although methylation is a biomarker that has great practical benefits, it is likely that as technology and data analysis methods improve, the use of additional genome-wide platforms will add further utility to cancer classification. In this regard, gene expression profiles show promise, and while they are outside the scope of this review, we believe it likely that such additional platforms will be utilized for cancer classification in the future. Finally, it is worth noting that pathologists have been central figures in the creation of these methylation classifiers, where the optimal clinical utility of methylation profiles requires integration of methylation data with established diagnostic methods with which anatomic pathologists are intimately familiar.

Development of a robust cancer classifier depends on training on a diverse and well-characterized set of neoplastic entities against which unknown samples will be compared. Experience to date suggests that methylation classifiers will be increasingly used to confirm and refine histomorphology-based diagnoses, identify clinically relevant subtypes, serve as quality control and teaching tools, and play a critical role in identifying new neoplastic entities.

SUMMARY POINTS

1. DNA methylation is a stable biomarker and is a mark of cell-of-origin in cancer.
2. Large-scale profiling has revealed complex and specific DNA methylation profiles in cancer.
3. DNA methylation in combination with machine learning has been successfully used to identify new cancer classes and to consolidate histologically disparate cancers.
4. Clinically useful central nervous system tumor and sarcoma DNA methylation classifiers are the first of their kind and are likely to be followed by many others.
5. Methods of cancer detection and classification from blood are currently emerging.
6. As DNA methylation interrogation becomes increasingly powerful and expansive, pan-cancer DNA methylation classifiers will become possible.
7. DNA methylation profiling will likely be adopted in cancer diagnostic criteria and will become a standard procedure performed at major centers.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We acknowledge Zied Abdullaev, Martha Quezado, and Drew Pratt for their helpful contributions to this manuscript.

LITERATURE CITED

1. Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *PNAS* 103:1412–17
2. Bird A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev*. 16:6–21
3. Edgar R, Tan PP, Portales-Casamar E, Pavlidis P. 2014. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenet. Chromatin* 7:28
4. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. 2009. An operational definition of epigenetics. *Genes Dev*. 23:781–83
5. Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–57
6. Bestor TH. 1992. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J*. 11:2611–17
7. Pradhan S, Bacolla A, Wells RD, Roberts RJ. 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J. Biol. Chem*. 274:33002–10
8. Paziewska A, Dabrowska M, Goryca K, Antoniewicz A, Dobruch J, et al. 2014. DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. *Br. J. Cancer* 111:781–89
9. Schwalbe EC, Williamson D, Lindsey JC, Hamilton D, Ryan SL, et al. 2013. DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathol*. 125:359–71
10. Vilahur N, Baccarelli AA, Bustamante M, Agramunt S, Byun HM, et al. 2013. Storage conditions and stability of global DNA methylation in placental tissue. *Epigenomics* 5:341–48
11. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, et al. 1983. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 11:6883–94
12. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, et al. 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst.* 92:564–69
13. Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, et al. 1995. 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor *p16/CDKN2/MTS1* in human cancers. *Nat. Med.* 1:686–92
14. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26:577–90
15. Spainhour JC, Lim HS, Yi SV, Qiu P. 2019. Correlation patterns between DNA methylation and gene expression in The Cancer Genome Atlas. *Cancer Inform*. 18:1176935119828776
16. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500:477–81
17. Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, et al. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* 24:132–38
18. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. 1977. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*. 37:646–50
19. Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*. 22:2990–97
20. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, et al. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS* 89:1827–31
21. Hayatsu H, Wataya Y, Kai K, Iida S. 1970. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* 9:2858–65
22. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *PNAS* 93:9821–26
23. Gonzalgo ML, Jones PA. 1997. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res*. 25:2529–31
24. Dupont JM, Tost J, Jammes H, Gut IG. 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal. Biochem.* 333:119–27

25. Hashimoto K, Kokubun S, Itoi E, Roach HI. 2007. Improved quantification of DNA methylation using methylation-sensitive restriction enzymes and real-time PCR. *Epigenetics* 2:86–91
26. Zhou L, Ng HK, Drautz-Moses DI, Schuster SC, Beck S, et al. 2019. Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Sci. Rep.* 9:10383
27. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, et al. 2009. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* 1:177–200
28. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, et al. 2006. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* 16:383–93
29. Moran S, Arribas C, Esteller M. 2016. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 8:389–99
30. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, et al. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6:692–702

31. First study to identify medulloblastoma molecular groups by DNA methylation.

31. Hovestadt V, Remke M, Kool M, Pietsch T, Northcott PA, et al. 2013. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* 125:913–16

32. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, et al. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* 11:587
33. Li W, Cerise JE, Yang Y, Han H. 2017. Application of t-SNE to human genetic data. *J. Bioinform. Comput. Biol.* 15:1750017
34. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, et al. 2018. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37:38–44
35. Capper D, Stichel D, Sahm F, Jones DTW, Schrimpf D, et al. 2018. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. *Acta Neuropathol.* 136:181–210
36. Braczynski AK, Capper D, Jones DTW, Schittenhelm J, Stichel D, et al. 2020. High density DNA methylation array is a reliable alternative for PCR-based analysis of the MGMT promoter methylation status in glioblastoma. *Pathol. Res. Pract.* 216:152728
37. Stichel D, Schrimpf D, Sievers P, Reinhardt A, Suwala AK, et al. 2020. Accurate calling of KIAA1549-BRAF fusions from DNA of human brain tumours using methylation array-based copy number and gene panel sequencing data. *Neuropathol. Appl. Neurobiol.* 47:406–14
38. Breiman L. 2001. Random forests. *Mach. Learn.* 45:5–32

39. CNS DNA methylation classifier description. This is the first large-scale DNA methylation cancer classifier.

39. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, et al. 2018. DNA methylation-based classification of central nervous system tumours. *Nature* 555:469–74

40. Sarcoma DNA methylation classifier description. This is the second large-scale DNA methylation cancer classifier.

40. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, et al. 2021. Sarcoma classification by DNA methylation profiling. *Nat. Commun.* 12:498

41. Priesterbach-Ackley LP, Boldt HB, Petersen JK, Bervoets N, Scheie D, et al. 2020. Brain tumour diagnostics using a DNA methylation-based classifier as a diagnostic support tool. *Neuropathol. Appl. Neurobiol.* 46:478–92
42. Shen SY, Burgener JM, Bratman SV, De Carvalho DD. 2019. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat. Protoc.* 14:2749–80
43. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, et al. 2006. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* 34:528–42
44. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, et al. 2020. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* 31:745–59
45. Nunes SP, Moreira-Barbosa C, Salta S, Palma de Sousa S, Pousa I, et al. 2018. Cell-free DNA methylation of selected genes allows for early detection of the major cancers in women. *Cancers* 10:357
46. Constancio V, Nunes SP, Moreira-Barbosa C, Freitas R, Oliveira J, et al. 2019. Early detection of the major male cancer types in blood-based liquid biopsies using a DNA methylation panel. *Clin. Epigenet.* 11:175
47. Kang S, Li Q, Chen Q, Zhou Y, Park S, et al. 2017. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* 18:53

48. Nassiri F, Chakravarthy A, Feng S, Shen SY, Nejad R, et al. 2020. Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. *Nat. Med.* 26:1044–47
49. Pegg AE, Dolan ME, Moschel RC. 1995. Structure, function, and inhibition of O^6-alkylguanine-DNA alkyltransferase. *Prog. Nucleic Acid Res. Mol. Biol.* 51:167–223
50. Watts GS, Pieper RO, Costello JF, Peng YM, Dalton WS, Futscher BW. 1997. Methylation of discrete regions of the O6-methylguanine DNA methyltransferase (MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene. *Mol. Cell. Biol.* 17:5612–19
51. Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, et al. 2000. Inactivation of the DNA-repair gene *MGMT* and the clinical response of gliomas to alkylating agents. *N. Engl. J. Med.* 343:1350–54
52. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, et al. 2004. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. *Clin. Cancer Res.* 10:1871–74
53. Paz MF, Yaya-Tur R, Rojas-Marcos I, Reynes G, Pollan M, et al. 2004. CpG island hypermethylation of the DNA repair enzyme methyltransferase predicts response to temozolomide in primary gliomas. *Clin. Cancer Res.* 10:4933–38
54. Wick W, Meisner C, Hentschel B, Platten M, Schilling A, et al. 2013. Prognostic or predictive value of *MGMT* promoter methylation in gliomas depends on *IDH1* mutation. *Neurology* 81:1515–22
55. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, et al. 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 17:510–22
56. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, et al. 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483:479–83
57. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, et al. 2012. Hotspot mutations in *H3F3A* and *IDH1* define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22:425–37
58. Hartmann C, Hentschel B, Wick W, Capper D, Felsberg J, et al. 2010. Patients with *IDH1* wild type anaplastic astrocytomas exhibit worse prognosis than *IDH1*-mutated glioblastomas, and *IDH1* mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas. *Acta Neuropathol.* 120:707–18
59. Wijnenga MMJ, Dubbink HJ, French PJ, Synhaeve NE, Dinjens WNM, et al. 2017. Molecular and clinical heterogeneity of adult diffuse low-grade IDH wild-type gliomas: Assessment of TERT promoter mutation and chromosome 7 and 10 copy number status allows superior prognostic stratification. *Acta Neuropathol.* 134:957–59
60. Brat DJ, Aldape K, Colman H, Holland EC, Louis DN, et al. 2018. cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV." *Acta Neuropathol.* 136:805–10
61. Weller M, Weber RG, Willscher E, Riehmer V, Hentschel B, et al. 2015. Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. *Acta Neuropathol.* 129:679–93
62. Thomas C, Sill M, Ruland V, Witten A, Hartung S, et al. 2016. Methylation profiling of choroid plexus tumors reveals 3 clinically distinct subgroups. *Neuro Oncol.* 18:790–96
63. Deng MY, Sill M, Chiang J, Schittenhelm J, Ebinger M, et al. 2018. Molecularly defined diffuse leptomeningeal glioneuronal tumor (DLGNT) comprises two subgroups with distinct clinical and genetic features. *Acta Neuropathol.* 136:239–53
64. Banan R, Stichel D, Bleck A, Hong B, Lehmann U, et al. 2020. Infratentorial IDH-mutant astrocytoma is a distinct subtype. *Acta Neuropathol.* 140:569–81
65. Hou Y, Pinheiro J, Sahm F, Reuss DE, Schrimpf D, et al. 2019. Papillary glioneuronal tumor (PGNT) exhibits a characteristic methylation profile and fusions involving *PRKCA*. *Acta Neuropathol.* 137:837–46
66. Pfaff E, Aichmuller C, Sill M, Stichel D, Snuderl M, et al. 2020. Molecular subgrouping of primary pineal parenchymal tumors reveals distinct subtypes correlated with clinical parameters and genetic alterations. *Acta Neuropathol.* 139:243–57

55. First study to describe G-CIMP, IDH-mutant gliomas.

67. Olar A, Wani KM, Wilson CD, Zadeh G, DeMonte F, et al. 2017. Global epigenetic profiling identifies methylation subgroups associated with recurrence-free survival in meningioma. *Acta Neuropathol*. 133:431–44
68. Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, et al. 2017. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol*. 18:682–94
69. Nassiri F, Mamatjan Y, Suppiah S, Badhiwala JH, Mansouri S, et al. 2019. DNA methylation profiling to predict recurrence risk in meningioma: development and validation of a nomogram to optimize clinical management. *Neuro Oncol*. 21:901–10
70. Sievers P, Sill M, Blume C, Tauziede-Espariat A, Schrimpf D, et al. 2021. Clear cell meningiomas are defined by a highly distinct DNA methylation profile and mutations in *SMARCE1*. *Acta Neuropathol*. 141:281–90
71. Northcott PA, Hielscher T, Dubuc A, Mack S, Shih D, et al. 2011. Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct. *Acta Neuropathol*. 122:231–40
72. Taylor MD, Northcott PA, Korshunov A, Remke M, Cho YJ, et al. 2012. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol*. 123:465–72
73. Pietsch T, Schmidt R, Remke M, Korshunov A, Hovestadt V, et al. 2014. Prognostic significance of clinical, histopathological, and molecular characteristics of medulloblastomas in the prospective HIT2000 multicenter clinical trial cohort. *Acta Neuropathol*. 128:137–49
74. Schwalbe EC, Hicks D, Rafiee G, Bashton M, Gohlke H, et al. 2017. Minimal methylation classifier (MIMIC): a novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. *Sci. Rep.* 7:13421
75. Ramaswamy V, Remke M, Bouffet E, Bailey S, Clifford SC, et al. 2016. Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol*. 131:821–31
76. Korshunov A, Sahm F, Zheludkova O, Golanov A, Stichel D, et al. 2019. DNA methylation profiling is a method of choice for molecular verification of pediatric WNT-activated medulloblastomas. *Neuro Oncol*. 21:214–21
77. Schwalbe EC, Lindsey JC, Nakjang S, Crosier S, Smith AJ, et al. 2017. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. *Lancet Oncol*. 18:958–71

78. First comprehensive description of ATRT epigenetic groups.

78. Johann PD, Erkek S, Zapatka M, Kerl K, Buchhalter I, et al. 2016. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* 29:379–93

79. Johann PD, Hovestadt V, Thomas C, Jeibmann A, Hess K, et al. 2017. Cribriform neuroepithelial tumor: molecular characterization of a SMARCB1-deficient non-rhabdoid tumor with favorable long-term outcome. *Brain Pathol*. 27:411–18
80. Thomas C, Wefers A, Bens S, Nemes K, Agaimy A, et al. 2020. Desmoplastic myxoid tumor, SMARCB1-mutant: clinical, histopathological and molecular characterization of a pineal region tumor encountered in adolescents and adults. *Acta Neuropathol*. 139:277–86
81. Korshunov A, Jakobiec FA, Eberhart CG, Hovestadt V, Capper D, et al. 2015. Comparative integrated molecular analysis of intraocular medulloepitheliomas and central nervous system embryonal tumors with multilayered rosettes confirms that they are distinct nosologic entities. *Neuropathology* 35:538–44

82. Showed that many histological CNS-PNETs have DNA methylation patterns of a variety of non-PNET tumors. Identified new molecular classes of CNS-PNETs.

82. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DTW, et al. 2016. New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* 164:1060–72

83. Witt H, Mack SC, Ryzhova M, Bender S, Sill M, et al. 2011. Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer Cell* 20:143–57
84. Mack SC, Witt H, Piro RM, Gu L, Zuyderduyn S, et al. 2014. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* 506:445–50

85. First large-scale study to describe ependymal tumor DNA methylation classes, some with a particularly poor prognosis.

85. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, et al. 2015. Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* 27:728–43

86. Gessi M, Giagnacovo M, Modena P, Elefante G, Gianno F, et al. 2019. Role of immunohistochemistry in the identification of supratentorial C11ORF95-RELA fused ependymoma in routine neuropathology. *Am. J. Surg. Pathol.* 43:56–63

87. Panwalkar P, Clark J, Ramaswamy V, Hawes D, Yang F, et al. 2017. Immunohistochemical analysis of H3K27me3 demonstrates global reduction in group-A childhood posterior fossa ependymoma and is a powerful predictor of outcome. *Acta Neuropathol*. 134:705–14
88. Jaunmuktane Z, Capper D, Jones DTW, Schrimpf D, Sill M, et al. 2019. Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre. *Acta Neuropathol. Commun.* 7:24
89. Sievers P, Appay R, Schrimpf D, Stichel D, Reuss DE, et al. 2019. Rosette-forming glioneuronal tumors share a distinct DNA methylation profile and mutations in *FGFR1*, with recurrent co-mutation of *PIK3CA* and *NF1*. *Acta Neuropathol*. 138:497–504
90. Nakano K, Takahashi S. 2018. Translocation-related sarcomas. *Int. J. Mol. Sci.* 19:3784
91. Renner M, Wolf T, Meyer H, Hartmann W, Penzel R, et al. 2013. Integrative DNA methylation and gene expression analysis in high-grade soft tissue sarcomas. *Genome Biol*. 14:r137
92. Seki M, Nishimura R, Yoshida K, Shimamura T, Shiraishi Y, et al. 2015. Integrated genetic and epigenetic analysis defines novel molecular subgroups in rhabdomyosarcoma. *Nat. Commun.* 6:7557
93. Koelsche C, Mynarek M, Schrimpf D, Bertero L, Serrano J, et al. 2018. Primary intracranial spindle cell sarcoma with rhabdomyosarcoma-like features share a highly distinct methylation profile and *DICER1* mutations. *Acta Neuropathol*. 136:327–37
94. Weidema ME, van de Geer E, Koelsche C, Desar IME, Kemmeren P, et al. 2020. DNA methylation profiling identifies distinct clusters in angiosarcomas. *Clin. Cancer Res.* 26:93–100
95. Kommoss FKF, Stichel D, Schrimpf D, Kriegsmann M, Tessier-Cloutier B, et al. 2020. DNA methylation-based profiling of uterine neoplasms: a novel tool to improve gynecologic cancer diagnostics. *J. Cancer Res. Clin. Oncol.* 146:97–104
96. Wu SP, Cooper BT, Bu F, Bowman CJ, Killian JK, et al. 2017. DNA methylation-based classifier for accurate molecular diagnosis of bone sarcomas. *JCO Precis. Oncol.* 2017:PO.17.00031
97. Koelsche C, Kriegsmann M, Kommoss FKF, Stichel D, Kriegsmann K, et al. 2019. DNA methylation profiling distinguishes Ewing-like sarcoma with *EWSR1–NFATc2* fusion from Ewing sarcoma. *J. Cancer Res. Clin. Oncol.* 145:1273–81
98. Koelsche C, Hartmann W, Schrimpf D, Stichel D, Jabar S, et al. 2018. Array-based DNA-methylation profiling in sarcomas with small blue round cell histology provides valuable diagnostic information. *Mod. Pathol.* 31:1246–56
99. Lee JC, Villanueva-Meyer JE, Ferris SP, Cham EM, Zucker J, et al. 2020. Clinicopathologic and molecular features of intracranial desmoplastic small round cell tumors. *Brain Pathol*. 30:213–25
100. Bernthal NM, Putnam A, Jones KB, Viskochil D, Randall RL. 2014. The effect of surgical margins on outcomes for low grade MPNSTs and atypical neurofibroma. *J. Surg. Oncol.* 110:813–16
101. Rohrich M, Koelsche C, Schrimpf D, Capper D, Sahm F, et al. 2016. Methylation-based classification of benign and malignant peripheral nerve sheath tumors. *Acta Neuropathol*. 131:877–87
102. Goel A, Nagasaka T, Arnold CN, Inoue T, Hamilton C, et al. 2007. The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. *Gastroenterology* 132:127–38
103. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, et al. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–37
104. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun HM, et al. 2012. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* 22:271–82
105. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. 1999. CpG island methylator phenotype in colorectal cancer. *PNAS* 96:8681–86
106. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, et al. 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat. Genet.* 38:787–93
107. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, et al. 1997. Methylation of the *hMLH1* promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* 57:808–11
108. Bettstetter M, Dechant S, Ruemmele P, Grabowski M, Keller G, et al. 2007. Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin. Cancer Res.* 13:3221–28

93. The authors used DNA methylation to identify a distinct group of *DICER1*-mutated cancers found in different anatomical sites. Some of these patients had germline *DICER1* mutations.

109. Adar T, Rodgers LH, Shannon KM, Yoshida M, Ma T, et al. 2017. A tailored approach to BRAF and MLH1 methylation testing in a universal screening program for Lynch syndrome. *Mod. Pathol.* 30:440–47
110. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. 2013. Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67–73
111. Esteller M, Levine R, Baylin SB, Ellenson LH, Herman JG. 1998. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* 17:2413–17
112. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, et al. 1999. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum. Mol. Genet.* 8:661–66
113. Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, et al. 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499:43–49
114. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, et al. 2018. The Cancer Genome Atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep.* 23:313–26
115. Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. 2016. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* 374:135–45
116. Chen W, Zhuang J, Wang PP, Jiang J, Lin C, et al. 2019. DNA methylation-based classification and identification of renal cell carcinoma prognosis-subgroups. *Cancer Cell Int.* 19:185
117. Evelonn EA, Landfors M, Haider Z, Kohn L, Ljungberg B, et al. 2019. DNA methylation associates with survival in non-metastatic clear cell renal cell carcinoma. *BMC Cancer* 19:65
118. Slater AA, Alokail M, Gentle D, Yao M, Kovacs G, et al. 2013. DNA methylation profiling distinguishes histological subtypes of renal cell carcinoma. *Epigenetics* 8:252–67
119. Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, et al. 2014. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 26:319–30
120. Malouf GG, Su X, Zhang J, Creighton CJ, Ho TH, et al. 2016. DNA methylation signature reveals cell ontogeny of renal cell carcinomas. *Clin. Cancer Res.* 22:6236–46
121. Ruiz-Cordero R, Rao P, Li L, Qi Y, Atherton D, et al. 2019. Hybrid oncocytic/chromophobe renal tumors are molecularly distinct from oncocytoma and chromophobe renal cell carcinoma. *Mod. Pathol.* 32:1698–707
122. Sirohi D, Smith SC, Agarwal N, Maughan BL. 2018. Unclassified renal cell carcinoma: diagnostic difficulties and treatment modalities. *Res. Rep. Urol.* 10:205–17
123. Hu S, Liu D, Tufano RP, Carson KA, Rosenbaum E, et al. 2006. Association of aberrant methylation of tumor suppressor genes with tumor aggressiveness and BRAF mutation in papillary thyroid cancer. *Int. J. Cancer* 119:2322–29
124. Mancikova V, Buj R, Castelblanco E, Inglada-Perez L, Diez A, et al. 2014. DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int. J. Cancer* 135:598–610
125. Bisarro dos Reis M, Barros-Filho MC, Marchi FA, Beltrami CM, Kuasne H, et al. 2017. Prognostic classifier based on genome-wide DNA methylation profiling in well-differentiated thyroid tumors. *J. Clin. Endocrinol. Metab.* 102:4089–99
126. Barros-Filho MC, Bisarro dos Reis M, Beltrami CM, de Mello JBH, Marchi FA, et al. 2019. DNA methylation-based method to differentiate malignant from benign thyroid lesions. *Thyroid* 29:1244–54
127. Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, et al. 2019. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci. Transl. Med.* 11(509):eaaw8513
128. Weinstein JN, Akbani R, Broom BM, Wang W, Verhaak RGW, et al. 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507:315–22
129. Burk RD, Chen Z, Saller C, Tarvin K, Carvalho AL, et al. 2017. Integrated genomic and molecular characterization of cervical cancer. *Nature* 543:378–84
130. Shen H, Shih J, Hollern DP, Wang L, Bowlby R, et al. 2018. Integrated molecular characterization of testicular germ cell tumors. *Cell Rep.* 23:3392–406

131. Cherniack AD, Shen H, Walter V, Stewart C, Murray BA, et al. 2017. Integrated molecular characterization of uterine carcinosarcoma. *Cancer Cell* 31:411–23
132. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, et al. 2015. The molecular taxonomy of primary prostate cancer. *Cell* 163:1011–25
133. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, et al. 2017. Integrative genomic analysis of cholangiocarcinoma identifies distinct *IDH*-mutant molecular profiles. *Cell Rep*. 19:2878–80
134. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, et al. 2017. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 169:1327–41.e23
135. Schweizer L, Thierfelder F, Thomas C, Soschinski P, Suwala A, et al. 2020. Molecular characterization of CNS paragangliomas identifies cauda equina paragangliomas as a distinct tumor entity. *Acta Neuropathol*. 140:893–906
136. Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, et al. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* 474:609–15
137. Robertson AG, Shih J, Yau C, Gibb EA, Oba J, et al. 2017. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer Cell* 32:204–20.e15
138. Wehmas LC, Hester SD, Wood CE. 2020. Direct formalin fixation induces widespread transcriptomic effects in archival tissue samples. *Sci. Rep.* 10:14497
139. Maros ME, Capper D, Jones DTW, Hovestadt V, von Deimling A, et al. 2020. Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data. *Nat. Protoc.* 15:479–512
140. Grunau C, Clark SJ, Rosenthal A. 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res*. 29:e65
141. Tomczak K, Czerwinska P, Wiznerowicz M. 2015. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19:A68–77
142. Clough E, Barrett T. 2016. The Gene Expression Omnibus database. *Methods Mol. Biol.* 1418:93–110