EPIGENOMICS
2026, VOL. 18, NO. 1, 101–115
https://doi.org/10.1080/17501911.2025.2586452

Taylor & Francis
Taylor & Francis Group

PERSPECTIVE

5mC and 5hmC methylation sequencing: the power of 6-base sequencing in a multiomic era

Robert Crawford [a], Tom Charlesworth [a], Stephen Riffle [b], Kurt Yardley[a] and Robert Osborne [a]

[a]biomodal Ltd, Chesterford Research Park, Saffron Walden, UK; [b]Independent Consultant, Denver, CO, USA

ABSTRACT

The integration of genetic and epigenetic information is essential for a comprehensive understanding of genome function and regulation. Traditional sequencing methods often fall short in capturing both genetic variants and epigenetic modifications such as 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) simultaneously. Recent advances in 6-base sequencing have enabled the simultaneous, base-resolution detection of canonical bases and key cytosine modifications in a single workflow. This review explores the biological significance of 5mC and 5hmC, discusses current methods to achieve 6-base sequencing, and highlights recent applications in academic and clinical settings.

PLAIN LANGUAGE SUMMARY

This review describes the importance of studying DNA methylation, specifically two markers of DNA methylation known as 5mC and 5hmC. DNA methylation is something that all human cells do and it helps them control when and where genes are active. In this review, we give special attention to a new class of technology that helps researchers detect both 5mC and 5hmC. In recent years, studies have shown that 5mC and 5hmC form patterns across human DNA, and that these patterns help direct cell behavior in ways that are different from one another. That direction is important for human development and may also play a role in disease. Researchers are particularly interested in these markers because several studies have shown that cancer cells often have unusual patterns of 5mC and 5hmC. Until recently, studying 5mC and 5hmC has been difficult because researchers have lacked technology that is able to differentiate between the two markers. What they did have allowed them to study DNA methylation broadly, but these studies lacked the detail needed to separate 5mC patterns from 5hmC patterns. Because of this, our understanding of DNA methylation is limited and lacking in consistency. Here, we describe a new type of sequencing technology – known as 6-base sequencing – that can help researchers reliably study both 5mC and 5hmC. So far, these studies have already revealed new aspects of human genetics and disease.

ARTICLE HISTORY
Received 25 August 2025
Accepted 3 November 2025

KEYWORDS
Genetics; epigenetics; DNA methylation; 5mC; 5hmC; modC; 6-base sequencing; WGBS

1. Introduction

Epigenetics was first defined in 1942 by the British developmental biologist Conrad Hal Waddington to describe the complex of unknown processes that mediate the flow of information from DNA to phenotype [1]. Nearly a century of study ensued, revealing a wide range of epigenetic processes that all share the common trait of altering the accessibility of nucleic acid chains without mutating their underlying sequence.

Among the most studied epigenetic processes is cytosine methylation and hydroxymethylation. These modifications affect cytosine's interactions with DNA-binding proteins, including transcriptional effectors, and are instrumental in the dynamic regulation of transcriptional activity. Consequently, cytosine modification plays a critical role in guiding cell fate decisions, preserving genomic stability, and stabilizing cellular phenotypes. Mounting evidence suggests that disruption of cytosine methylation and hydroxymethylation patterns may be an early step in the genesis and progression of multiple pathological conditions, including colorectal cancer, Alzheimer's disease, and others [2–5].

Studying the relationship between epigenetic aberrations and genetic variation is essential to understanding complex biological processes and disease mechanisms. However, doing so has historically been challenging. Among the many hurdles researchers face are basic technological limitations that make it difficult to comprehensively study cytosine modifications. Conventional sequencing technologies typically require separate workflows for genetic and epigenetic analyses which limit the scope of epigenetic studies due to increased sample input requirements, cost, and analytical complexity.

The emergence of 6-base sequencing technologies represents a recent and significant advancement in the field of epigenomics. By enabling the simultaneous detection of the four canonical DNA bases (A, T, C, G) along with 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), these technologies offer a more comprehensive view of the genome and the factors that dictate gene-phenotype relationships.

CONTACT Robert Crawford robert.crawford@biomodal.com biomodal Ltd, Chesterford Research Park, The Trinity Building, Little Chesterford, Saffron Walden CB10 1XL, UK

Article highlights

- 5mC and 5hmC are stable epigenetic marks with distinct influences over gene expression.
- 5mC is generally associated with gene repression; whereas 5hmC is indicative of gene reawakening.
- Changes in 5hmC patterns appear to predict 5mC dynamics and may prove to be valuable clinical biomarkers
- Traditional methylation sequencing conflates 5mC and 5hmC patterns, reporting instead on total modified cytosine (modC).
- 6-base sequencing technologies enable simultaneous interrogation of 5mC and 5hmC patterns
- By resolving 5mC from 5hmC, 6-base sequencing is exposing new dynamics in gene regulation, enabling sensitive detection of circulating tumor DNA (ctDNA), and more.

In this review, we delve into the epigenetic influence of 5mC and 5hmC, focusing on recent findings connecting these modifications to disease processes and the potential of 6-base sequencing technologies to deepen our understanding of genetics and health.

2. Cytosine modifications and their functional roles

2.1. The epigenetic alphabet: beyond A, T, C, and G

Though 5-methylcytosine (5mC) was first observed in 1904 by American chemists Treat Johnson and Robert Coghill of Yale, the modified base wasn't recognized as a member of the core genetic alphabet (A, T, C, G) until 1950 [6–8]. This 'fifth base' drew attention when Chargaff reported that DNA contained equal amounts of adenine and thymine, yet cytosine levels remained slightly – and consistently – below those of guanine [9]. Gerard Wyatt, at the University of Cambridge, showed that, to balance C and G levels, Chargaff needed to account for the large fraction of 5mC in the genome [8–11].

5mC is formed when a methyl group (donated by S-adenosylmethionine) is enzymatically transferred to the 5′ carbon of cytosine's pyrimidine ring by for example human DNA methyltransferases DNMT1, DNMT3A and DNMT3B [12]. Though not exclusive, 5mC is often formed where cytosine precedes guanine residues in sequence (CpG dinucleotides), and is particularly enriched within heterochromatic (closed) DNA [13]. Once formed, 5mC can be highly stable, with myriad cellular mechanisms acting to preserve its patterns through cell division [14].

However, 5mC is not a permanent epigenetic mark. Demethylation may occur through both passive and active processes. The former takes place during cell division when, through various means, methyltransferase enzymes are prevented from acting, thereby leaving newly incorporated cytosine bases unmodified [15].

In contrast, active demethylation is an enzymatic process that occurs in both dividing and non-dividing cells. Ten-Eleven Translocase (TET) enzymes mediate the multistep process, beginning with the oxidation of 5mC to form the stable intermediate 5-hydroxymethylcytosine (5hmC). After subsequent oxidation steps (forming the short-lived intermediates 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC)), base excision and repair machinery replace the modified base with an unmodified cytosine (Figure 1) [15]. In recent years, the stability and unique biological influence of the 5hmC intermediate – as well as its prominence in neuronal tissues – has led it to be viewed as the 6th base in life's genetic alphabet [16,17].

Both 5mC and 5hmC appear to be critical to genomic function, each playing an active role across a wide range of processes, from gene regulation to chromatin remodeling, genomic stability and ultimately cell fate determination [18–22].

2.2. The physiochemical consequences of cytosine modification

The biological significance of 5mC and 5hmC is both recognized and enigmatic. To date, there is no known somatic vertebrate cell type that is viable in the absence of DNA methylation machinery, suggesting a core requirement for cytosine modification [23]. Yet, the precise mechanics of how 5mC and 5hmC exert influence over cell viability is not fully understood.

At a high level, these markers influence cell behavior by modulating DNA's accessibility to transcriptional machinery. While the addition of a methyl or hydroxymethyl group to

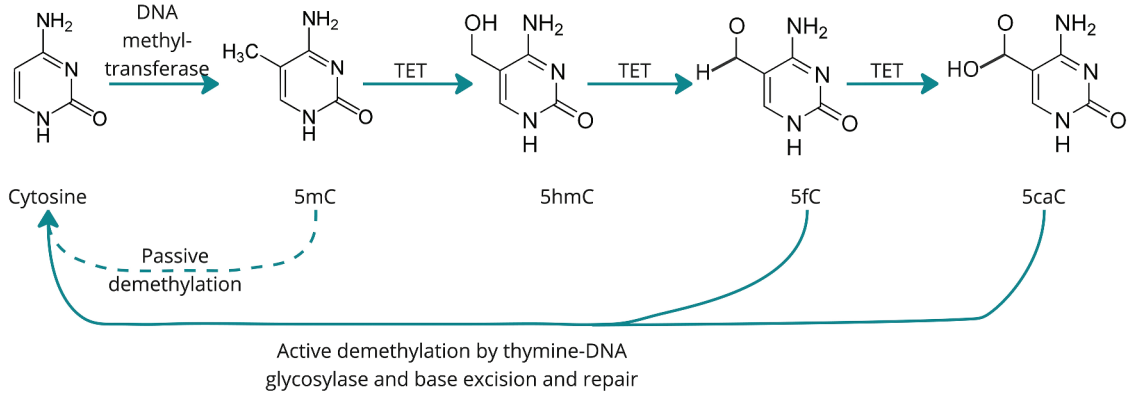

Figure 1. Cytosine modification cycle. Schematic depicting the methylation of cytosine by DNA methyl transferase proteins, followed by passive (diffusion based) demethylation or active demethylation through oxidation by Ten-eleven translocase (TET) enzymes. Oxidation of 5mC gives rise to 5hmC. Base excision repair (BER) proteins, such as thymine DNA glycosylase (TDG) act to remove subsequent oxidation products (5-fC and 5-caC), replacing them with unmodified cytosine.

cytosine does not alter its base pairing, these moieties occupy space in the major and minor grooves of the DNA helix. Here, they can have both steric and physicochemical effects on the DNA superstructure that ultimately affect protein-binding dynamics [23,24].

The magnitude and direction of any cytosine modification's effect appears highly context dependent, influenced by the bases surrounding the modified cytosine, chromatin structure, and interactions with the protein's DNA binding domain. As a result, cytosine modifications can have simultaneously opposing effects, such that 5mC in a specific locus both repels certain transcription factors while attracting others [23–26].

It stands to reason that 5mC and 5hmC, possessing different steric and chemical properties, affect protein binding in distinct ways. Indeed, modeling of transcription factor-binding preferences shows modification-specific effects. For example, the CCAAT/enhancer-binding protein beta (C/EBPβ) transcription factor has demonstrated a strong preference for binding core recognition sequences containing 5mC but not 5hmC [25,27]. Such an effect has been observed for many transcription factors, including cAMP response element-binding protein (CREB1), upstream stimulatory factor 1 (USF1), and others [25,28–30]. Therefore, these related modifications exhibit distinct effects on DNA–protein interactions and may play unique roles in genomic function. It should be noted, too, that cytosine modifications may have effects on DNA accessibility that are independent of transcription factor binding [31].

Despite the complexity of predicting any single modification's effect on genomic activity, generalities have been observed. Hypermethylation (concentrated formation of 5mC) is traditionally associated with transcriptional repression. Enrichment of 5mC in gene promoters, for example, has long been associated with repression of nearby transcriptional start sites and reduced gene expression. Such an effect may be mediated through a combination of decreased transcriptional activator binding and increased recruitment of repressors. In contrast, transformation of 5mC into 5hmC is typically associated with emerging transcriptional activity from associated genes, particularly when enriched in gene bodies [15,32–34].

The distinct effects of 5mC and 5hmC on transcriptional effector recruitment provide cells with a means for honing gene expression, and, by extension, cell identity. This is particularly relevant during embryonic development as stem cells rapidly transition across cell states en route to terminal differentiation.

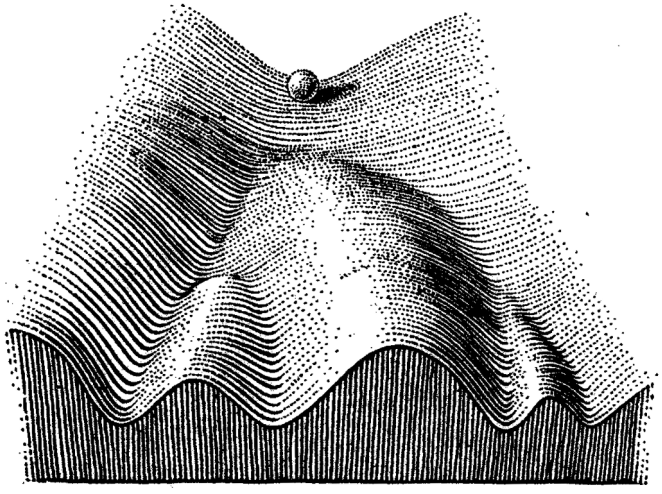

Figure 2. Waddington's landscape. Cartoon depicting a ball poised above a landscape of bifurcating valleys, representing the semi-stochastic guiding of cellular differentiation. From: the strategy of the genes: a discussion of some aspects of theoretical biology, C. H. Waddington, 1957, fig. 4 in chapter 2, p. 29. This edition published in 2014 by Routledge, an imprint of the Taylor & Francis group. © 1957 George Allen & Unwin Ltd. Reproduced by permission of Taylor & Francis group [35].

2.3. Relevance of 5mC and 5hmC in development

Waddington defined the epigenome as a complex of processes that helped to guide organismal development, specifically the derivation of differentiated cell types from a single genome [1]. This was visually depicted as a ball (representing a stem cell) rolling down a hill that progressively bifurcates into distinct valleys (differentiated cell types) (Figure 2). In this metaphorical image, unknown epigenetic factors make up the valley walls that guide the ball's progression and prevent its lateral movement (transdifferentiation) [35].

Mapping current knowledge onto Waddington's landscape, we can say that cytosine modifications (5mC, 5hmC) do function like the valley walls, guiding and maintaining cell identity by regulating transcription factor binding and gene expression [31,36].

Following embryo fertilization, the zygote must bring two distinct and unequal genomes into functional unity. Cytosine methylation is a critical tool in this process. During a period of significant epigenetic change, both parent genomes are demethylated and subsequently remethylated, enabling cells to hypermethylate (silence) redundant genes, establish necessary gene expression profiles, and begin to both divide and differentiate [37]. During this period, cytosine methylation levels show a logarithmic rate of change as gene expression requirements fluctuate across developmental time and space. When nearing terminal differentiation, the rate of change decreases as methylation levels stabilize into cell and tissue-specific patterns [38,39].

As predicted by Waddington, these patterns play an important role in defining and maintaining cell identity [40]. However, the influence that cytosine modifications have on the genome goes well beyond the honing of gene expression. These markers help to maintain genomic stability, drive genetic mutation, and may be essential to the development of common diseases. But, to understand the breadth of 5mC and 5hmC influence, we must first be able to detect them.

2.4. Methylation detection technologies

Faithfully recording 5mC and 5hmC levels across the genome is essential to advancing our understanding of human genomics and may prove valuable for clinical decision making. Yet, simply determining cytosine's modification status, both locally and globally, is a non-trivial task. Several methods have been developed to differentiate 5mC from unmodified cytosine with varying degrees of sensitivity, specificity, and scalability [41,42]. Many of these methods require deep technical and bioinformatic expertise, and few are capable of distinguishing 5mC from

Table 1. Methylation sequencing technology.

Modality	Method	Resolution	Input Mass (ng)	Summary	References
modC	LC-MS/MS	Global	1–1000	Highly sensitive and accurate for global quantification	[43]
(5mC + 5hmC)	MeDIP-seq	Regional	10–−1000	Antibody-based enrichment; biased toward CpG-dense regions	[44]
	MBD-seq		500–1000	Uses MBD proteins for enrichment; misses low-density methylation	[45]
	EPIC array		500–1000	Probe-based hybridization; limited coverage	[46]
	WGBS	Base-resolution	200–5000	Historical gold standard for 5mC detection; cannot distinguish 5hmC	[47]
	EM-seq		10–200	Enzymatic, bisulfite-free; detects both 5mC and 5hmC	[48]
	TAPS		10–200	Detects modC (5mC + 5hmC)	[49]
5mC only	TAPS-β	Base-resolution	0.5–250	Detects 5mC specifically using βGT protection	[50]
	DM-Seq		10	Enzymatic, bisulfite-free	[51]
	PacBio 5-base HIFI		5000	Simultaneous detection of genetic and epigenetic information	[52]
5hmC only	hMeDIP-seq	Regional	500–5000	Antibody-based enrichment; biased toward high-density 5hmC regions	[53]
	MSA Array		50–250	Probe-based hybridization; limited coverage	[54]
	oxBS-seq	Base-resolution	1000 – 5000	Oxidizes 5hmC to 5fC; indirect inference of 5hmC	[55]
	TAB-seq		500–2000	Detects 5hmC specifically using glycosylation protection	[56]
6-base	evoC	Base-resolution	5–80	Simultaneous detection of genetic and epigenetic information	[57,58]
	ONT		100 -1200	Simultaneous detection of genetic and epigenetic information	[59,60]

A non-exhaustive list of available techniques and technologies for detecting cytosine modifications. Note that input mass ranges are based on published quantities and should be interpreted as both approximate and highly dependent on context (quality of DNA source, desired coverage, etc.).

Abbreviations: modC = total modified cytosine; 5mC = 5-methylcytosine; 5hmC = 5-hydroxy methylcytosine; 5fC = 5-formylcytosine; 6-base = sequencing that resolves 5mC, 5hmC, and the four canonical DNA bases (A, T, C, G). LC-MS/MS = liquid chromatography mass spec; MeDIP-seq = Methylated DNA Immunoprecipitation sequencing; MBD-seq = Methyl-CpG binding domain sequencing; EPIC array = sequencing method based on bisulfite conversion and use of the Infinium MethylationEPIC array technology; WGBS = whole genome bisulfite sequencing; EM-seq = enzymatic methylation sequencing; TAPS = TET-assisted pyridine borane sequencing; TAPS-β = TET-assisted pyridine borane sequencing beta; DM-seq = direct enzymatic methylation sequencing; hMeDIP-seq = hydroxymethylated DNA immunoprecipitation sequencing; MSA Array = bisulfite conversion based methylation sequencing microarray; oxBS-seq = oxidative bisulfite sequencing; TAB-seq = Tet-assisted bisulfite sequencing; evoC = duet evoC enzymatic conversion-based sequencing; ONT = Oxford Nanopore Technologies long-read sequencing; βGT = beta-glucosyltransferase.

5hmC (Table 1). Because of this, progress in the field has been marred by inconsistencies and technical limitations [31,33,42,61].

Fortunately, recent technological advances – such as the advent of 6-base sequencing methods – may bring relief to the field, greatly improving our ability to study these consequential epigenetic markers.

2.5. What are you detecting, really?

Several methods for detecting 5mC have been developed, ranging from high-performance liquid chromatography and mass spectrometry to bisulfite sequencing [41]. Generally, these methods can be categorized based on the resolution of their readout and, consequently, their application in global, local, or base-level analyses (Table 1). Here, we primarily focus our discussion on high-resolution technologies that provide base-level readouts. For a comprehensive overview of current methylation detection technologies and their applications, we refer the reader to some influential reviews on the topic [41,62].

Currently, the most common approaches to methylation detection rely on a bisulfite conversion step followed by either whole genome sequencing (WGBS) or fluorescent detection using microarray technology. In both cases, DNA exposed to bisulfite will undergo a conversion wherein unmodified cytosine bases are deaminated to uracil. Subsequent amplification via polymerase chain reaction converts uracil into thymine residues (representing a C to T conversion, denoted C > T). In this reaction, 5mC and 5hmC are protected from conversion and, following PCR, will be preserved as unmodified cytosine.

Following conversion, DNA may be interrogated by either next generation sequencing (as in the case of WGBS) or by probe hybridization (as in microarray technology). For WGBS, unmodified cytosine bases can be differentiated from modified cytosines by comparing sequencing results from untreated and bisulfite treated samples. In the latter, a gain of thymine (C > T), indicates an unmodified base, while the preservation of cytosine (C > C) denotes the presence of either 5mC or 5hmC. For array-based technologies, modified bases are detected through direct probe binding of converted sequences.

Array-based methylation detection has been widely used for targeted analyses, enabling researchers to conserve resources while achieving base-level data for pre-defined loci. However, these technologies often require a high input DNA mass, and their accuracy may be reduced by probe cross-hybridization, single nucleotide polymorphisms, and probe-specific dye biases [63]. Additionally, arrays are inherently limited in scope. Therefore researchers seeking a genome-wide view of methylation opt for WGBS.

While WGBS has been widely used for many years, several limitations have driven the field to look for alternatives. Bisulfite conversion is a damaging process that greatly reduces the amount of available material for sequencing [64,65]. Consequently, methods rooted in bisulfite conversion often require high DNA input mass, which can be problematic for applications with relatively little source DNA, such as circulating tumor DNA [66,67]. The development of post-bisulfate library preparation methods may help to mitigate this issue and reduce mass requirements [66].

Additionally, bisulfite conversion reduces genome complexity, with the bulk of its resulting sequencing data containing just three nucleotides (A,T,G). This greatly complicates sequencing analysis, slows the experimental process, and increases false-positive matches during alignment [68]. Reduced genomic complexity also presents unique challenges to researchers who hope to use cost-saving targeted sequencing methods. Further, in relying on C > T conversion, this approach obscures bona fide C > T polymorphisms, which represent the most common mutations in the human genome and in cancer [69,70].

Like most methods of methylation detection (Table 1), microarrays and WGBS are unable to differentiate between 5mC and 5hmC – both will be conflated into a single total modified cytosine (modC) readout. Such a conflation means that bisulfite sequencing is not a true measure of DNA methylation. In reality, it can only provide an assessment of cytosine modification levels, with no indication of which modification type is present. Failure to distinguish between these two biologically distinct epigenetic markers can mask potential gene-phenotype links [33].

To overcome these issues, multiple single-base conversion methods have been developed to enable differentiation of 5hmC from 5mC, including oxidative bisulfite sequencing, TET-assisted pyridine borane sequencing-beta (TAPS-β), Tet-assisted bisulfite sequencing (TAB-seq) and APOBEC-coupled epigenetic sequencing [71–74]. Though valuable, these methods often require separate sample processing and sequencing workflows, which may increase sample input requirements, experimental costs, and time requirements. Additionally, generating phased epigenetic and genetic variation with these methods is not often viable. As with any conversion-based method, the reliability of these methods heavily depends on conversion rate efficiency, which can be quite low [34,75]. Low conversion efficiency can lead to both false positives and false negatives.

There is a significant need for technology that enables the simultaneous and unequivocal sequencing of the four canonical bases (A,T,G,C) alongside both modified cytosine bases (5mC and 5hmC).

2.6. 6-base sequencing technologies

True 6-base sequencing technologies are defined as those that can simultaneously identify all 6-bases (A, T, C, G, 5mC, and 5hmC) in a single workflow. At present, there are only two such platforms: biomodal's duet evoC and Oxford Nanopore Technologies (ONT). It should be noted that technologies like Pacific Bio's 5-Base HiFi sequencing enable the detection of the four canonical bases as well as 5mC. While this technology has gained in popularity, it does not yet allow for 5hmC detection and is thus not discussed in depth here. We refer interested readers to some recent literature on the technology [76].

2.6.1. Oxford Nanopore technologies (ONT) for 5mC and 5hmC detection

Oxford Nanopore Technologies (ONT) offers a unique approach to epigenetic sequencing by directly detecting DNA modifications during sequencing, without the need for chemical conversion or amplification. ONT platforms, such as the MinION and PromethION, measure changes in ionic current as single DNA strands pass through a nanopore. The resulting shifts in current are base-specific and can be uniquely affected by base modifications, allowing for the detection of 5mC and, with more recent models and algorithms, 5hmC. Machine learning models (such as nanopolish, DeepSignal, and Megalodon) support signal shift interpretation.

In a recent study, it was demonstrated that a single ONT run may reliably distinguish 5mC from 5hmC at CpG sites when models are trained on known standards [77]. This same study demonstrated the technology's ability to provide duplex sequencing data, thereby improving base-calling accuracy and enabling detection of asymmetric methylation.

The primary advantage of ONT's technology is that it directly reads native DNA, preserving the integrity of both genetic and epigenetic data. The lack of chemical and enzymatic conversion steps during library preparation means that nanopore sequencing preserves the integrity of the original DNA sequence and can be relatively quick. Further, and in contrast to biomodal's duet evoC, ONT's long-read sequencing provides long-range context that can be helpful when studying allele-specific methylation [78].

Nonetheless, ONT has several drawbacks that must be considered. The raw read accuracy for nucleotide sequences has greatly improved in recent years, but error rates remain relatively high, posing potential challenges for variant calling in repetitive or low-complexity regions [77–80].

Diminished accuracy is partly related to the complexity of inferring base identity, which requires subjective interpretation of event data. Event data is the accumulation of statistics describing the nanopore's current variance as DNA passes through the polarized channel. This data is influenced by multiple bases at a time, such that each pulse of recorded event data measures the effect of roughly 6 bases at a time (depending on the length of the nanopore sensing region). Not only can information be lost as raw data is translated into event data for analysis, but base-calling can vary depending on the software employed [81,82]. This is further complicated by the emerging nature of 6-base sequencing – in order to infer 5mC or 5hmC's presence, base-calling models must be well-trained and leverage ground-truth for both modifications [77].

Due to these limitations, certain sequence contexts are prone to false-positive calls with nanopore technology. In particular, GC-rich and repetitive regions – both of which are commonly home to modified cytosine bases – are known to elevate the false-positive rate of modified-base calls [77].

Practical limitations include the need for bioinformatics expertise and the technology's need for large DNA input with a mass between > 100ng and 1 µg [83]. Additionally, fragmented DNA such as that of FFPE samples or cell-free DNA (cfDNA) significantly limits achievable yield. Such requirements limit the technology's application where available DNA input material is low, as is often the case when working with liquid biopsies, for example [84].

2.6.2. Biomodal's duet multiomics solution evoC

Biomodal's duet evoC platform leverages a single, enzymatic-conversion workflow that enables simultaneous detection of A, T, C, G, 5mC, and 5hmC (Figure 3). It uses a two-strand system wherein the original DNA strand is linked (via hairpin loop) to a synthetic complementary strand. Subsequent enzymatic treatments (including methyl transfer, glycosylation, TET oxidation, and deamination) convert cytosine modifications into distinguishable base-pairing patterns, which are decoded using a 16-state two-base code [85]. Herein, base-calling is informed by a two-base code, which improves accuracy and reduces bias.

In a recent comparative study [85], duet evoC demonstrated 98.55% sensitivity and 99.95% specificity for 5mC detection. The technology was shown to have high concordance with comparable WGBS and EM-seq workflows, while

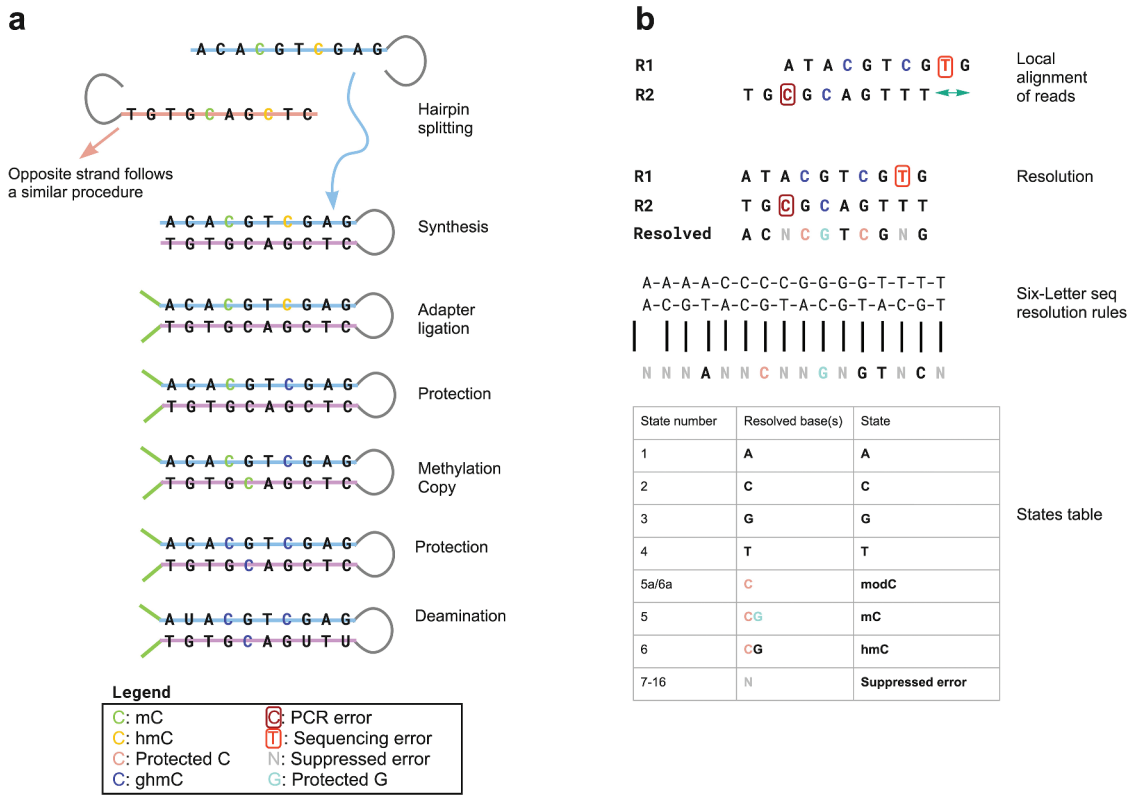

Figure 3. 6-base sequencing with duet evoC. a. Schematic of six-base epigenetic sequencing laboratory workflow. Hairpins are ligated to double-stranded dna and the strands are separated. The 3′–5′ strand is omitted for clarity, but follows a similar procedure to the 5′–3′ strand. An additional copy strand is synthesized using Klenow exo-polymerase and short sequencing adapters are ligated. 5hmCs are protected by glycosylation by beta-glucosyltransferase (BGT) and a methyl-copy step copies only the 5mCs from the original to the copy strand using DNMT5. All modCs are protected through oxidation by TET2 and glycosylation by beta-glucosyltransferase (BGT). Treatment by APOBEC3A and UvrD helicase is used to simultaneously open up and deaminate the hairpin. Unprotected Cs are deaminated from C to U (read as t). b. Overview of the resolution rules and states under the six- decoding model. A protected C on the original strand, signifying modC, is denoted by pink in the diagram and table; a G opposite a protected C on the copy strand is denoted by light blue. The 5mC is denoted by a protected (pink) C followed by a protected (blue) G and 5hmC is denoted by a protected (pink) C followed by an unprotected (black) G. Figure reprinted from fullgrabe et al. In accordance with its creative commons attribution 4.0 international license: http://creativecommons.Org/licenses/by/4.0/ [82].

requiring lower DNA input. In addition, phased analyses are improved through detection of genetic and epigenetic variants on the same DNA molecule.

Thus far, duet evoC has been shown to have several advantages for 6-base sequencing. Firstly, the technology uses an all-enzymatic workflow that avoids harsh, DNA-denaturing bisulfite treatments. This means the resulting data has higher library complexity, more uniform coverage, and larger fragment recovery – all of which translate to more reliable and comprehensive sequencing. Secondly, duet evoC can be used on fragmented or damaged DNA samples, meaning it can be leveraged to analyze a wide range of sample types (including formalin-fixed, paraffin-embedded tissues and cfDNA) [86]. TKTK

Additionally, duet evoC is compatible with illumina short-read sequencing platforms, enabling its integration into existing sequencing pipelines. It is also notable that, unlike ONT, duet evoC has an ultra low DNA input requirement (~5 ng), making it well suited for cfDNA analyses [4,57,58]. Lastly, duet evoC includes a built in error suppression modality that removes PCR and sequencing errors before alignment.

As with any technology, duet evoC is not without its drawbacks. It's a multistep assay and its current compatibility is confined to short-read sequencing technology, limiting the phasing of genetic and epigenetic variants across longer genomic regions. Additionally, to resolve the 5th and 6th bases requires collapsing paired end reads to a pseudo single end format, which increases sequencing requirements. However, the impact of this will be mitigated as the cost of sequencing on short-read systems continues to drop [87].

Further, duet evoC's reliance on base-conversion means that its accuracy is heavily dependent on conversion efficiency. Lack of efficiency can lead to false positive and false negative interpretations. At present, the high conversion efficiency of biomodal's technology mitigates this risk [88].

3. The epigenetic mechanisms of disease

Over the past century, many thousands of genetic associations have been identified linking disease phenotypes to underlying genetic mutations [88]. Yet, this extensive list only describes associations for a third of known human diseases, very few of which can be wholly attributed to discrete mutations [89]. It is increasingly clear that disease is driven by the combination of genetic and epigenetic factors not by either alone.

Therefore to better understand, diagnose, and treat human disease, it is necessary to study the influence of both genetic and epigenetic variation. Much of what we know about the influence of cytosine modifications is born from traditional technologies like

WGBS. Yet, progress has been complicated by statistical noise and contradictory findings due, in part, to the conflation of 5mC and 5hmC by many technologies. With the advent of 6-base sequencing, however, new insights are being gained. Below, we briefly review how cytosine modification may contribute to disease pathology, highlighting along the way where 6-base sequencing is helping to advance the field.

3.1. Mutational hotspots

Cytosine modification has been shown to contribute to disease pathology through multiple non-exclusive mechanisms. One such mechanism concerns the physical nature of 5mC, which closely resembles the chemical structure of thymine. Cytosine frequently undergoes spontaneous hydrolytic deamination, which results in the formation of uracil (C > U). Base excision repair (BER) machinery can readily detect the discordant base pairing (U:G) and take action to restore cytosine. However, when 5mC undergoes deamination, it results in thymine. Attempts to resolve this mispairing (T:G) can, at minimum, result in an epigenetic mutation as the deaminated base is replaced with an unmodified cytosine. Alternatively, if BER machinery fails to recognize the mispairing before DNA replication, the deaminated 5mC will be paired with an adenine on the newly synthesized strand, leading to a stable C > T mutation [69,90,91]. Indeed, C > T mutations are among the most common in the human genome, accumulating in an age-dependent manner and frequently implicated in disease genesis [69,91].

Therefore, loci with 5mC represent mutational hotspots from which diseases like cancer, neurodegeneration, and other heritable diseases can arise [69,91,92].

The extent to which 5hmC deamination occurs, much less its contribution to stable genomic mutation remains unclear. However, studies have shown that deamination of 5hmC can produce 5-hydroxymethylated uracil (5hmU). The resulting 5hmU:G mispairing may be restored to cytosine through multiple repair mechanisms, including long-patch BER similar to 5mC. If the latter occurs, it may inadvertently lead to the loss of cytosine modification in neighboring residues [16,93]. The aberrant loss of 5mC and 5hmC in the region would then constitute an epigenetic mutation.

3.2. Genomic instability

Beginning early in embryonic development, 5mC (alongside multiple other epigenetic factors) plays a critical role in establishing and repressing heterochromatic regions of the genome. Heterochromatin is essential to chromosomal segregation, telomere formation, and genomic stability [18–20]. Among its many roles, heterochromatin helps to repress the expression of repetitive and transposable elements. When expressed, these elements may reinsert into the genome at random loci, potentially disrupting gene expression, prompting chromosomal rearrangements, and increasing DNA damage. The collective effect of abnormal transposable element expression is increased genomic instability and increased likelihood of disease development [94–96].

In helping to repress transposable elements contained within heterochromatin, 5mC is essential to genomic stability. In one recent study, Sepulveda et al. [22] used a combination of WGBS, ONT, and duet evoC 6-base sequencing to show that dysregulation of TET proteins prompted global hypomethylation, including notable loss of 5mC in several heterochromatic loci. Hypomethylation in these regions correlated with both a corresponding increase in 5hmC levels and the expression of transposable elements – specifically endogenous retroviruses (ERVs), intracisternal A-type particles (IAP) retrotransposons, and a family of long interspersed elements (L1) – contained therein [22]. Such a finding is significant because it indicates that 5mC has a repressive influence over transposable elements that are frequently expressed in cancers, Alzheimer's disease, and other such conditions [95,96].

Additionally, Sepulveda et al.'s [22] use of 6-base sequencing to distinguish 5mC from 5hmC enabled them to observe the dynamics of heterochromatin demethylation in finer temporal detail, noting that 5hmC levels spike rapidly following TET dysregulation. Therefore, detection of 5hmC in heterochromatic regions may be an early indicator of approaching genomic instability.

3.3. Surrogate mutations

A potential consequence of genetic mutation is abnormal gene expression. In the development of cancer, for example, gain of function mutations in oncogenic transcription factors, loss of function mutations in tumor suppressor genes, or regulatory element mutations can drive malignant transformation. C > T mutations occurring in the Telomerase Reverse Transcriptase (TERT) promoter are among the most common mutations observed across cancer types [92,97,98]. These mutations enable recruitment of ETS transcription factors to the TERT promoter, leading to increased promoter activity and a greater proliferative potential in affected cells.

However, such mutations are not required to inspire pathological gene expression. Epigenetic changes – including DNA methylation, disrupted patterns of histone modifications, and chromatin remodeling – can occur independently of genetic mutations and are capable of both initiating and sustaining cancerous growth [99]. By affecting mitotically heritable changes in gene expression, deposition of 5mC and 5hmC in both gene bodies and regulatory elements can function as surrogate mutations [100,101].

Here again, TERT is a good example. Multiple reports suggest that, in the absence of promoter mutations, increased expression of the TERT enzyme is achieved in malignant cells by altering promoter cytosine methylation status [102]. Though not well understood, it's hypothesized that excessive formation of 5mC in the TERT promoter may repel repressive transcription factors (such as WT1) without displacing activating transcription factors, leading to a net increase in gene expression.

Of note, a recent study suggests that the co-occurrence of genetic mutation and loss of 5hmC in the TERT promoter is associated with aggressive cancer phenotypes in thyroid cancers, indicating a potentially synergistic relationship between genomic mutation and cytosine modification [103]. Such

results also highlight the potential prognostic value of measuring 5hmC.

While cytosine modification can function as a surrogate mutation, it's notable that 5mC and 5hmC formation is dynamic and reversible. Indeed, hypermethylation of the BRCA1 promoter has been observed in triple negative breast cancers and can indicate tumor-cell vulnerability to platinum-based agents. Yet, unlike loss of function mutations in BRCA1, promoter methylation can be reversed following treatment, giving rise to treatment resistant subpopulations [104].

In these examples, we see that cytosine modifications at the gene level may not only play a critical role in tumor transformation by emulating tumor-driver mutations, but their dynamic nature may enable tumor cells to respond and adapt to environmental triggers. These illustrative examples are just a few of many, and we encourage those interested in 5mC and 5hmC's influence over tumorigenic gene expression to read some of the listed reviews on this subject [99,105–107].

Differential patterns of 5mC and 5hmC are associated with nonmalignant disease progression as well. Over the past decade, several studies have reported distinctive methylation patterns that are associated with Alzheimer's disease and some of its defining characteristics [2,3,108]. Though less studied, 5hmC shows a 10-fold enrichment in brain tissue and is expected to play a prominent role in neurodegenerative disease [109,110]. In support of this, postmortem examination of Alzheimer brain tissue from 1079 patients has demonstrated substantial reductions in 5hmC proximal to disease-associated genes, with a notable absence of detectable genetic mutations [111].

Similar results have been reported for Parkinson's disease, amyotrophic lateral sclerosis (ALS), fragile X-associated tremor/ataxia syndrome (FXTAS), Friedreich ataxia (FRDA), and Huntington's disease (HD) [112–114].

3.4. Cultivating disease progenitor cells

Disruptions in 5mC and 5hmC patterns on both a local and global scale may be a fundamental process to both the formation and maintenance of pathological cell identities. A strong example of this can be found in the cancer epigenome.

Though substantial differences exist between cancer types, the transformation and evolution of tumor cells is believed to follow a similar progression, whereby initiating mutations increase the likelihood of further mutation and clonal expansion. Each generation of clones accumulate a new complement of mutations that eventually give rise to a heterogeneous tumor cell population [115].

As made clear above, mutations that initiate tumorigenesis and contribute to intratumoral heterogeneity do not need to be exclusively genetic. Abnormal distribution of 5mC and 5hmC is hypothesized to be a foundational step toward malignant transformation [116]. Feinberg et al. have proposed a model of cancer development that begins with the alteration of methylation patterns, which consequently elicit stem-like properties in cells and leave them vulnerable to the effects of tumor driver mutations [116].

This epigenetic progenitor model is supported by several lines of evidence. Indeed, while the methylation status of individual genes may vary, hypomethylation at the genomic level is ubiquitous across cancer types and has been observed at every stage of the tumor lifecycle, from pre-malignant to advanced stages [105,116,117]. The ubiquity of hypomethylation (and local hypermethylation events) supports the notion that early epigenetic changes in cancer progenitor cells are maintained during clonal expansion.

Additionally, the prevalence of putative tumor driver mutations in nonmalignant tissues suggests that the influence of these powerful mutations is dependent on additional factors, one of which may be epigenetic context [118]. Cytosine modifications have already been demonstrated to regulate the influence of tumorigenic mutations, as highlighted in prostate cancer. Among the most heritable forms of cancer, genetic studies have identified roughly 160 risk-associated loci for prostate cancer. Though thousands of single nucleotide polymorphisms (SNPs) can be found in these loci, they collectively explain only 28% of the disease's total familial risk [119–122].

In a recent study, Ahmed et al. found that one of these risk loci harbors a cis enhancer for the MYC oncogene [119]. Efforts to link SNPs within this enhancer to MYC activity had thus far produced mixed results. The authors show that enhancer activity is dependent on the methylation status of a nearby CTCF binding site. Hypermethylation of this preferred CTCF binding site drives the protein to seek out lower affinity alternatives, including a secondary binding site located near the MYC enhancer locus. Binding to this secondary site alters DNA geometry, enabling enhancer activity and subsequent MYC expression.

In other words, neither hypermethylation of the CTCF binding site nor mutation of the MYC enhancer are consequential on their own. It is only when combined that an oncogenic effect is observed.

Similar results were found by Galasso et al. in chronic lymphocytic leukemia (CLL) [123], wherein the effect of a SNP (rs1001179) in the catalase gene was found to be modulated by the concomitant demethylation of neighboring CpG dinucleotides. The combination of local hypomethylation and genetic mutation of the CAT gene not only led to its malignant expression, but may prove to be a valuable tool for predicting disease aggressiveness.

Feinberg's epigenetic progenitor model [116] is further supported by the epigenetic link to clonal hematopoiesis (CH). CH is a pre-malignant condition in which a single hematopoietic stem cell acquires a somatic mutation and expands to produce a disproportionately large fraction of circulating blood cells. Although initially viewed as benign, recent findings indicate that CH is associated with an increased risk of hematologic cancer and cardiovascular disease, as well as overall mortality for these conditions [124].

Notably, the mutations that drive clonal expansion frequently occur in genes responsible for DNA methylation maintenance, including DNMT3A and TET2. Loss-of-function mutations in these genes can disrupt global 5mC and 5hmC patterns, resulting in a mosaic of epigenetic states that permit greater plasticity and potential activation of oncogenic programs [124,125].

Recent findings using 6-base sequencing support the notion that repatterning of cytosine modifications is an

early step in disease development. Puddu et al. analyzed cfDNA and showed that combining 5mC and 5hmC signals significantly improves early detection of colorectal cancer (CRC), compared to treating them as a single signal [4]. They observed that many regions with increased 5hmC in stage I CRC later showed decreased 5mC in stage IV, suggesting progressive demethylation during tumor evolution. These findings indicate that 5hmC enrichment marks early disease stages, while loss of 5mC becomes more pronounced in advanced stages, highlighting the complementary value of both modifications for detection and monitoring.

Notably, the most sensitive method for detecting circulating tumor DNA (ctDNA) across stages combined the individual analyses of 5mC and 5hmC patterns, such that 5mC + 5hmC achieved an area under the curve of 0.95, outcompeting 5mC alone (0.7), 5hmC alone (0.54), and modC (0.66). This is significant because the Centers for Medicare & Medicaid Services guidelines suggest that blood based biomarker testing for CRC should have a minimum of 74% sensitivity at 90% specificity [126]. Puddu et al.'s 5mC + 5hmC model exceeds this, showing an 85% sensitivity at 95% specificity [4].

These findings may at first appear contradictory: why would the combination of 5mC and 5hmC measurements be different from modC? The answer is related to the finer resolution gained by observing individual marker changes, specifically where gain of 5hmC co-occurs with the loss of 5mC. In analyses that use modC, these anti-correlated loci may appear to cancel out, obscuring the changes that are occurring on a base-level. By analyzing 5hmC independently from 5mC, these changes can be recorded as distinct patterns. Subsequently combining the analysis of these patterns enables ctDNA to be defined by more complex, distinctive patterns, ultimately enabling more sensitive ctDNA detection.

Together, these findings strongly suggest that global rearrangement of cytosine modifications is an early step in disease development, one that may increase cellular plasticity and create a permissive genomic environment wherein disease-driving mutations can influence cellular behavior. Such a model for disease development is also hypothesized to apply to nonmalignant conditions where alterations in cytosine modifications are similarly observed throughout disease development [116].

4. Embracing the 6-base genome

With the recent development of 6-base sequencing technologies, it is likely that we'll see a significant expansion in our understanding of epigenetic disease mechanisms over the next decade. Already, this technology is driving advancements in academic and clinical settings.

4.1. Fundamental research

Despite decades of progress, many questions remain about the fundamental workings of epigenetic gene regulation. How, for instance, do asymmetric cytosine modifications affect cell behavior; when are these modifications most influential; and why is 5hmC so enriched in neuronal cell types? Answering these questions will not be easy, but technologies like 6-base sequencing can help by exposing the dynamics of gene regulation in high resolution.

Enhancer regulation, for example, has proven to be a complex process involving both chromatin modifications and DNA methylation. Acetylation of histone 3 at lysine 27 (H3K27Ac) proximal to enhancers increases DNA accessibility, transcription factor binding, and the expression of gene networks under the enhancer's influence [127,128]. Demethylation of enhancer sequences would be expected to have a similar effect and correlate closely with histone acetylation dynamics. However, results thus far have been discordant, with some studies reporting that enhancer demethylation may lag behind histone acetylation (and gene expression changes) by as much as 48 hours [31,129,130]. Results from Guerin et al. suggest that this discordance may be explained by the conflation of 5mC and 5hmC [31].

Using a combination of bisulfite and 6-base sequencing, Guerin et al. [31] detailed the time course of enhancer epigenetic changes during stem cell differentiation. Using bisulfite sequencing, they found that histone acetylation, transcription factor binding, and gene expression all preceded the loss of modified cytosine (modC), suggesting that cytosine modifications may be decoupled from gene expression dynamics.

However, with 6-base sequencing, Guerin et al. [31] reported that conflating 5hmC and 5mC obscured key temporal dynamics of DNA demethylation. Their observations show that 5hmC levels begin to spike at enhancers *prior to* histone modification and transcription factor binding, priming the location for increased activity providing sufficient signal such that machine learning algorithms could faithfully predict future chromatin accessibility based on 5hmC patterns.

To achieve full demethylation took several days, accounting for the apparent lag observed with bisulfite sequencing. In conflating cytosine modifications, bisulfite sequencing saw the balanced spike in 5hmC and loss of 5mC as a net neutral, leading to the misleading readout of little to no change in cytosine modification levels.

The gain in resolution achieved through 6-base sequencing enabled Guerin et al. [31] to reveal new insights into the temporal process of enhancer activation and demonstrate the value of 5hmC in neuronal progenitor cell differentiation. These results reflect a broader trend wherein 6-base sequencing exposes nuanced demethylation dynamics, often showing 5hmC levels spiking before modified cytosine levels dip [4,22].

Further nuance into enhancer regulation was recently provided by Hardwick et al., who used a form of 6-base sequencing to assess the symmetry of cytosine modifications as they appear on the DNA double helix [131]. The vast majority of modified cytosines occur in CpG dinucleotides. Each CpG loci in the double helix has two cytosine residues that can be independently modified. Importantly, some DNA-binding proteins have shown a preference for either dual or hemimethylated dinucleotide sites [132–134].

Through 6-base sequencing, Hardwick et al. found that 5hmC was a predominantly (98%) asymmetrical modification, occurring on only one of two CpG cytosines [131]. Further,

they show that asymmetrical hydroxymethylation was most prominent among primed enhancers (as opposed to active enhancers). These results indicated that the symmetry (or lack thereof) of 5hmC formation may play an important role in defining and maintaining enhancer states.

4.2. 5mC and 5hmC as clinical biomarkers

In addition to supporting basic research on the mechanics of epigenetic influence, 6-base sequencing is being used for clinical applications as well, particularly in the identification and development of potential biomarkers.

4.2.1. Early cancer detection

Cytosine modification is already the subject of significant focus as a blood-based biomarker for early cancer detection [135]. Epigenetically reprogrammed tumor cells may release DNA into the bloodstream through various means. This ctDNA may carry diagnostically valuable mutations and epigenetic patterns [136]. However, distinguishing ctDNA from the far more abundant milieu of cfDNA that originates from healthy tissue is non-trivial. To be identifiable, ctDNA fragments must carry a mutational pattern that is unique to the malignant genome. The more detailed this pattern is, the more reliable its identification will be.

At present, ctDNA is most often defined by genetic mutations but this approach faces significant limitations for early-stage cancer detection. The burden of nonsynonymous mutations in known cancer genes does not differ substantially between early- and late-stage tumors [137]. Moreover, small tumors shed fewer DNA fragments, making mutation-bearing ctDNA exceedingly rare [138]. In contrast, aberrant DNA methylation arises early in tumorigenesis and affects thousands of CpG sites across the genome. By exploiting these widespread and tissue-specific epigenetic alterations, methylation-based assays can aggregate signals across multiple CpGs within individual fragments. This strategy enhances sensitivity at low tumor fractions and improves specificity by mitigating sequencing noise.

This was emphasized by Klein et al. [139] who showed that ctDNA detection based on cytosine modification patterns (measured with WGBS) could improve the positive predictive value of ctDNA detection across several cancer types. But while overall sensitivity was 90.1% for stage IV cancers (at 99.5% specificity), sensitivity at stages I-III was only 40.7%. As with genetic mutation, the shift in modC may be too subtle in early-stage cancers and limit reliable ctDNA detection. However, as discussed above, distinct methylation patterns can be obscured when 5mC and 5hmC are conflated. Therefore 6-base sequencing may help to further improve ctDNA recognition and early cancer detection [4,5].

Recent evidence also suggests that considering patterns from both genetic and epigenetic data can also improve detection sensitivities. Wang et al. examined the benefits of using multiomic data in the detection of ctDNA from hepatocellular carcinoma (HCC) patients [140]. Identifying genetic mutations and DNA methylation patterns (modC) improved detection sensitivity by 50% compared to mutations alone, and 33% compared to methylation alone.

Collectively, these studies suggest that the inclusion of cytosine modifications in cfDNA analysis can support sensitive ctDNA detection and, specifically, that 6-base sequencing can be particularly beneficial by exposing higher resolution epigenetic patterns.

4.2.2. Tracing tissue of origin

Recent studies, including Klein et al., have demonstrated the feasibility of using ctDNA methylation signatures to classify tumors by tissue of origin [117,139]. This capability is particularly valuable in the context of multi-cancer early detection (MCED) and cancers of unknown primary (CUP) [139,141]. Despite these advances, certain tissues remain challenging to distinguish due to overlapping methylation profiles. For example, tumors originating from the upper gastrointestinal tract and pancreas often exhibit similar methylation signatures, complicating precise classification. Moreover, the resolution of current classifiers is limited by the depth and breadth of reference methylation datasets, as well as by the technical constraints of ctDNA analysis. The integration of genome-wide maps of both 5mC and 5hmC could hold promise for improving tissue-of-origin inference.

4.2.3. Tumor characterization and therapeutic guidance

Several recent studies have demonstrated the value of studying 5hmC and 5mC as distinct biomarkers for primary tumor characterization [100,101,142–144]. In keeping with the trend of 5hmC forecasting early epigenomic changes, Guler et al., found that anti-PD1 therapeutics elicited an early change in cfDNA 5hmC levels among patients with non-small cell lung cancer [142]. This gain of 5hmC helped to predict which patients were likely to show a diminished response to treatment. Further examination showed that 5hmC patterns prior to treatment could similarly be used to estimate treatment response. These results suggest that the incorporation of 5hmC profiling in ctDNA analysis can help guide therapeutic decision making and potentially forecast disease recurrence.

5. Conclusion and future perspectives

Cytosine modification is essential to human life, so much so that 5mC and 5hmC are now recognized as the 5th and 6th DNA bases [11,17]. Together, these epigenetic markers guide stem cells down Waddington's landscape, enabling stem cells with identical genomes to differentiate and give rise to distinct cell types. It is increasingly clear that these same epigenetic markers can guide cells toward a pathological identity.

In the above sections, we describe how 5mC and 5hmC are likely to exert influence, physically altering interactions between DNA and the proteins that bind it. Importantly, 5mC and 5hmC have distinct effects on protein binding and should therefore be considered independently when studying both developmental and pathological processes.

Both 5mC and 5hmC appear to have important roles to play in disease pathology. Abnormal patterns of modC are well documented in disease contexts, with extensive evidence showing global hypomethylation as a hallmark of the cancer genome [145]. Hypomethylation may lead to the expression of oncogenic gene networks, increased genomic instability, and

increased cellular plasticity. Similarly, the local hypermethylation of tumor suppressor genes may facilitate transformation and affect therapeutic response. Notably, 5mC and 5hmC exist alongside genetic mutations and may affect the functional consequences of these variants (and vice versa).

The vast majority of studies exploring the impact of cytosine modifications have relied on WGBS and related technologies. Though valuable, these technologies are unable to differentiate between 5mC and 5hmC, providing instead a readout of the total cytosine modification (modC) level. Consequently, most studies reporting on DNA methylation have not studied DNA methylation specifically. In conflating 5mC and 5hmC, these technologies obscure the dynamic loss of 5mC and gain of 5hmC that may occur as genes reawaken from a silenced state. The result of this is lower resolution sequencing, contradictory results, and decreased sensitivity in biomarker detection.

To date, only two platforms are capable of faithfully reporting on the four canonical bases (A, T, C, G) as well as both 5mC and 5hmC (so-called 6-base sequencing).

ONT's long-read sequencing technology enables rapid 6-base sequencing and provides long-range context that can be valuable for phased analyses. However, accurately inferring base identity from subtle shifts in nanopore current is a nontrivial task. This is particularly challenging when attempting to call bases like 5hmC, for which ground truth training data may be limited. Additionally, the technology requires a high mass of input DNA, and suffers from yield limitations with fragmented DNA (e.g., cfDNA) which may limit its application in biomarker detection.

In contrast, biomodal's duet evoC platform provides a direct readout of base identity, enabled by a multistep enzymatic conversion workflow. The process requires relatively little input DNA (as low as 5ng) and its use of enzymatic conversion preserves sample integrity. This means that duet evoC is ideal when sequencing fragmented or damaged samples, such as those found in formalin-fixed, paraffin-embedded tissues and liquid biopsies.

6-base sequencing has already shed important light on the dynamics of epigenetic gene regulation. In multiple circumstances, 5hmC has been shown to be an early indicator of gene expression change, preceding detectable loss of 5mC on both a local and global scale [4,22,31]. This information may be valuable for early disease detection, predicting therapeutic response, and deepening our understanding of basic genomic processes.

In the near future, we're likely to see an expanding application of 6-base sequencing, particularly in the context of neurodegenerative disease. 5hmC is known to be enriched in neuronal tissue and has already been implicated in the pathophysiology of conditions like Alzheimer's and Parkinson's disease [146,147]. 6-base sequencing may help to resolve the epigenetic changes that occur early in disease and potentially reveal valuable diagnostic biomarkers.

From a broader perspective, 6-base sequencing is still in its early days, but it is already reshaping our understanding about what can be asked about genome regulation. In the coming years, its integration with single-cell and spatial methods should illuminate how 5mC and 5hmC patterns vary within tissue architectures, across cell lineages, and in response to environmental and pathological pressures. In parallel, coupling genetic and epigenetic data in allele-aware analyses will sharpen our understanding of how sequence variation and cytosine modification act together to drive phenotypic change.

Given these advances, researchers are likely to transition away from antiquated technologies that measure only modC in favor of more detailed readouts. If pursued with methodological rigor and thoughtful experimental design, these advances will not only deepen our mechanistic understanding of gene regulation but also accelerate the discovery of clinically actionable biomarkers, moving us closer to connecting the topology of Waddington's landscape to real-world trajectories of health and disease.

6. Future perspective

In the coming years, 6-base sequencing technologies are likely to become standard in DNA methylation studies, with particular promise in the study of cancer and neurodegenerative disease. Additionally, as researchers embrace multiomics, we will likely see an expanding integration of 6-base sequencing with other omics and single-cell technologies. We expect that the increased adoption of 6-base sequencing technologies (and their continued evolution) will lead to myriad discoveries that advance our understanding of both human genetics and disease development; while also giving rise to clinically valuable biomarkers. We are undoubtedly entering an exciting era for epigenetic research.

Author contributions

Robert Crawford: Conceptualization, Writing – original drafting, reviewing & editing. Tom Charlesworth: Writing – original drafting, reviewing & editing. Stephen Riffle: Writing – original drafting, reviewing & editing, Kurt Yardley – Writing – original drafting, reviewing & editing, Robert Osborne – Writing – original drafting, reviewing & editing.

Disclosure statement

Kurt Yardley, Robert Crawford, Robert Osborne and Tom Charlesworth are employees of biomodal. Stephen Riffle is an independent consultant who receives funding from biomodal.

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Funding

This paper was not funded.

ORCID

Robert Crawford http://orcid.org/0009-0003-2487-4626

Tom Charlesworth http://orcid.org/0000-0001-9233-7679
Stephen Riffle http://orcid.org/0000-0002-5947-242X
Robert Osborne http://orcid.org/0000-0002-1914-1239

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Waddington CH. An introduction to modern genetics. 1st ed. London (UK): Routledge; 1950.
2. Fetahu IS, Ma D, Rabidou K, et al. Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease. Sci Adv. 2019;5(8). doi: 10.1126/sciadv.aaw2880
3. Shireby G, Dempster EL, Policicchio S, et al. Dna methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. Nat Commun. 2022;13(1):5620. doi: 10.1038/s41467-022-33394-7
4. Puddu F, Johansson A, Modat A, et al. 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. bioRxiv. 2024 [cited 2025 Aug 20]. doi: 10.1101/2024.10.30.621123
   •• **This article demonstrates how 6-base sequencing can be used to improve early cancer detection.**
5. Nj W, Rashid M, Yu S, et al. Hydroxymethylation profile of cell-free DNA is a biomarker for early colorectal cancer. Sci Rep. 2022;12(1). doi: 10.1038/s41598-022-20975-1
6. Johnson TB, Wheeler HL. 5-methylcytosine. Am Chem J. 1904;31:591–601.
7. Johnson TB, Coghill RROP. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. J Am Chem Soc. 1925;47(11):2838–2844. doi: 10.1021/ja01688a030
8. Wyatt GR. Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem J. 1951;48(5):581–584. doi: 10.1042/bj0480581
9. Chargaff E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia. 1950;6(6):201–209. doi: 10.1007/BF02173653
10. Wyatt GR, Cohen SS. A new pyrimidine base from bacteriophage nucleic acids. Nature. 1952;170(4338):1072–1073. doi: 10.1038/1701072a0
11. Tompkins JD. Discovering DNA methylation, the history and future of the writing on DNA. J Hist Biol. 2022;55(4):865–887. doi: 10.1007/s10739-022-09691-8
12. Lanata CM, Chung SA, Criswell LA. Dna methylation 101: what is important to know about DNA methylation and its role in SLE risk and disease heterogeneity. Lupus Sci Med. 2018;5(1):e000285. doi: 10.1136/lupus-2018-000285
13. Jang HS, Shin WJ, Lee EJ, et al. CpG and non-CpG methylation in epigenetic gene regulation and brain function. Genes (Basel). 2017;8(6):148. doi: 10.3390/genes8060148
14. Petryk N, Bultmann S, Bartke T, et al. Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucl Acids Res. 2021;49(6):3020–3032. doi: 10.1093/nar/gkaa1154
15. Moore L, Le T, Fan G. Dna methylation and its basic function. Neuropsychopharmacol. 2013;38(1):23–38. doi: 10.1038/npp.2012.112
16. Grin I, Ishchenko AA. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. Nucl Acids Res. 2016;44(8):3713–3727. doi: 10.109.3/nar/gkw059
17. Kraus TF, Guibourt V, Kretzschmar HA. 5-hydroxymethylcytosine, the "sixth base", during brain development and ageing. J Neural Transm (Vienna). 2015;122(7):1035–1043. doi: 10.1007/s00702-014-1346-4
18. Pappalardo XG, Barra V. Losing DNA methylation at repetitive elements and breaking bad. Epigenet Chromatin. 2021;14(25). doi: 10.1186/s13072-021-00400-z
19. Toubiana S, Velasco G, Chityat A, et al. Subtelomeric methylation distinguishes between subtypes of immunodeficiency, centromeric instability and facial anomalies syndrome. Hum Mol Genet. 2018;27(20):3568–3581. doi: 10.1093/hmg/ddy265
20. Bourc'his D, Bestor T. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96–99. doi: 10.1038/nature02886
21. Honig F, Murrell A. Cell identity and 5-hydroxymethylcytosine. Epigenet Chromatin. 2025;18(36). doi: 10.1186/s13072-025-00601-w
22. Sepulveda H, Li X, Arteaga-Vazquez LJ, et al. Ogt prevents DNA demethylation and suppresses the expression of transposable elements in heterochromatin by restraining TET activity genome-wide. Nat Struct Mol Biol. 2025;32(7):1282–1296. doi: 10.1038/s41594-025-01505-9
23. Kribelbauer JF, Lu KJ, Rohs R, et al. Towards a mechanistic understanding of DNA methylation readout by transcription factors. J Mol Biol. 2019;432(6):1801–1815. doi: 10.1016/j.jmb.2019.10.021
   • **The authors provide a detailed description of how cytosine modifications (5mC and 5hmC) affect protein-DNA interactions.**
24. Li S, Peng Y, Panchenko AR. Dna methylation: precise modulation of chromatin structure and dynamics. Curr Opin Struct Biol. 2022;75(102430):102430. doi: 10.1016/j.sbi.2022.102430
25. Viner C, Ishak CA, Johnson J. Modeling methyl-sensitive transcription factor motifs with an expanded epigenetic alphabet. Genome Biol. 2024;25(1):11. doi: 10.1186/s13059-023-03070-0
26. de Mendoza A, Nguyen TV, Ford E, et al. Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. Genome Biol. 2022;23(163). doi: 10.1186/s13059-022-02728-5
27. Sayeed SK, Zhao J, Sathyanarayana BK, et al. C/EBPβ (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. Biochem Biophys Acta. 2015;1849(6):583–589. doi: 10.1016/j.bbagrm.2015.03.002
28. Khund-Sayeed S, He X, Holzberg T, et al. 5-hydroxymethylcytosine in E-Box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. Integr Biol (Camb). 2016;8(9):936–945. doi: 10.1039/c6ib00079g
29. Khund-Sayeed S, He X, Tillo D, et al. 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) enhance the DNA binding of CREB1 to the C/EBP half-site tetranucleotide GCAA. Biochemistry. 2016;55(49):6940–6948. doi: 10.1021/acs.biochem.6b00796
30. Liu L, Zhang Y, Liu M, et al. Structural insights into the specific recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL effectors. J Mol Biol. 2020;432(4):1035–1047. doi: 10.1016/j.jmb.2019.11.023
31. Guerin LN, Scott TJ, Yap JA, et al. Temporally discordant chromatin accessibility and DNA demethylation define short- and long-term enhancer regulation during cell fate specification. Cell Rep. 2025;44(5):115680. doi: 10.1016/j.celrep.2025.115680
   •• **This paper highlights the negative effect of conflating 5mC and 5hmC by resolving temporal dynamics of gene regulation with 6-base sequencing.**
32. Madvedeva YA, Khamis AM, Kulakovskiy IV, et al. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014;15(1):119. doi: 10.1186/1471-2164-15-119
33. Li L, Gao Y, Wu Q, et al. New guidelines for DNA methylome studies regarding 5-hydroxymethylcytosine for understanding transcriptional regulation. Genome Res. 2019;29(4):543–553. doi: 10.1101/gr.240036.118
   • **This paper provides guidelines for the evaluation of 5mC and 5hmC, detailing the added benefit of including 5hmC as a separate marker from 5mC.**
34. Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012;149(6):1368–1380. doi: 10.1016/j.cell.2012.04.027
35. Waddington CH. The strategy of the genes. 1st ed. London (UK): Routledge; 1957.
36. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128(4):128(4)635–638. doi: 10.1016/j.cell.2007.02.006
37. Menezo Y, Clement P, Clement A, et al. Methylation: an ineluctable biochemical and physiological process essential to the transmission of life. Int J Mol Sci. 2020;21(23):9311. doi: 10.3390/ijms21239311

38. Horvath S. Dna methylation age of human tissues and cell types. Genome Biol. 2013;14(10):3156. doi: 10.1186/gb-2013-14-10-r115
39. He B, Zhang C, Zhang X, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. Nat Commun. 2021;12(1). doi: 10.1038/s41467-021-24425-w
40. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019;20(10):590–607. doi: 10.1038/s41580-019-0159-6
41. Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. Biology (Basel). 2016;5(1):3. doi: 10.3390/biology5010003
42. He B, Yao H, Yi C. Advances in the joint profiling technologies of 5mC and 5hmC. RSC Chem Biol. 2024;5(6):500–507. doi: 10.1039/D4CB00034J
43. Dai N, Corrêa IR. Liquid chromatography–mass spectrometry analysis of cytosine modifications. Methods In Mol Biol. 2021:2198. doi: 10.1007/978-1-0716-0876-0_6
44. Dna Methylation Workshop - Medip [Internet]. Diagenode Inc; 2022 [cited 2025 Aug 20]. Available from: https://diagenode.co.jp/wp-content/uploads/2022/01/002-2022_01_DNA_Methylation_Workshop.pdf
45. Lan X, Adams C, Landers M, et al. High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. PLOS ONE. 2011;6(7):e22226. doi: 10.1371/journal.pone.0022226
46. DNA Input Guidelines for Infinium MethylationEPIC Arrays [Internet]. Illumina; 2024 [cited 2025 Aug 20]. Available from: https://knowledge.illumina.com/microarray/general/microarray-general-reference_material-list/000002679
47. Whole Genome Bisulfite Sequencing [Internet]. Novogene; 2025 [cited 2025 Aug 20]. Available from: https://www.novogene.com/us-en/wp-content/uploads/sites/4/2025/07/22-Detailed-Materials-for-WGBS-202506.pdf
48. NEBNext® Enzymatic Methyl-seq (EM-seq™) [Internet]. New England Biolabs. [cited 2025 Aug 20]. Available from: https://www.neb.com/en-us/-/media/nebus/files/application-notes/technote_nebnext_enzymatic_methyl-seq.pdf
49. Liu Y, Siejka-Zielińska P, Velikova G, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol. 2019;37(4):424–429. doi: 10.1038/s41587-019-0041-2
50. Cheng J, Siejka-Zielinska P, Liu Y, et al. Endonuclease enrichment taps for cost-effective genome-wide base-resolution DNA methylation detection. Nucleic Acids Res. 2021;49(13):e76–e76. doi: 10.1093/nar/gkab291
51. Wang T, Fowler JM, Liu L, et al. Direct enzymatic sequencing of 5-methylcytosine at single-base resolution. Nat Chem Biol. 2023;19(8):1004–1012. doi: 10.1038/s41589-023-01318-1
52. Cheung WA, Johnson AF, Rowell WJ, et al. Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort. Nat Commun. 2023;14(1):3090. doi: 10.1038/s41467-023-38782-1
53. CD Genomics. Medip-seq/hmedip-seq (DNA methylation & hydroxymethylation profiling) [Internet]. [cited 2025 Aug 20]. Available from: https://www.cd-genomics.com/medip-sequencing.html
54. Infinium™ Methylation Screening Array [Internet]. Illumina; 2024 [cited 2025 Aug 20]. Available from: https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-methylation-screening-array-data-sheet-m-gl-01893/infinium-methylation-screening-array-data-sheet-m-gl-01893.pdf
55. oxBS-seq [Internet]. Cd genomics. [cited 2025 Aug 20]. Available from: https://www.cd-genomics.com/oxbs-seq.html
56. Skvortsova K, Bogdanovic O. Tab-seq and ACE-seq data processing for genome-wide DNA hydroxymethylation profiling. Methods In Mol Biol. 2021:2272. doi: 10.1007/978-1-0716-1294-1_9
57. duet multiomics solution +modC [Internet]. Biomodal. 2025 [cited 2025 Aug 20]. Available from: https://biomodal.com/products/duet-modc/
58. duet multiomics solution evoC [Internet]. Biomodal. 2025 [cited 2025 Aug 20]. Available from: https://biomodal.com/products/duet-evoc/
59. Liu Y, Rosikiewicz W, Pan Z, et al. Dna methylation-calling tools for Oxford nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol. 2021;22(1):295. doi: 10.1186/s13059-021-02510-z
60. Chemistry Technical Document [Internet]. Oxford Nanopore Technologies; 2024 [cited 2025 Aug 20]. Available from: https://nanoporetech.com/document/chemistry-technical-document
61. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucl Acids Res. 2010;38(11):e125–e125. doi: 10.1093/nar/gkq223
62. Erlitzki N, Kohli RM. An overview of global, local, and base-resolution methods for the detection of 5-hydroxymethylcytosine in genomic DNA. Methods Mol Biol. 2024;2842:325–352. doi: 10.1007/978-1-0716-4051-7_17
63. Neefs I, Ibrahim J, Peeters M, et al. The technology landscape for detection of DNA methylation in cancer liquid biopsies. Epigenetics. 2025;20(1):2453273. doi: 10.1080/15592294.2025.2453273
64. Olova N, Krueger F, Andrews S, et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol. 2018;19(33). doi: 10.1186/s13059-018-1408-2
65. Tanaka K, Okamoto A. Degradation of DNA by bisulfite treatment. Bioorg Med Chem Lett. 2007;17(7):1912–1915. doi: 10.1016/j.bmcl.2007.01.040
66. Han Y, Zheleznyakova GY, Marincevic-Zuniga Y, et al. Comparison of EM-seq and PBAT methylome library methods for low-input DNA. Epigenetics. 2021;17(10):1195–1204. doi: 10.1080/15592294.2021.1997406
67. Morrison J, Koeman JM, Johnson BK, et al. Evaluation of whole-genome DNA methylation sequencing library preparation protocols. Epigenet Chromatin. 2021;14(28). doi: 10.1186/s13072-021-00401-y
68. Xi Y, Li W. Bsmap: whole genome bisulfite sequence mapping program. BMC Bioinf. 2009;10(232). doi: 10.1186/1471-2105-10-232
69. Cagan A, Baez-Ortega A, Brzozowska N, et al. Somatic mutation rates scale with lifespan across mammals. Nature. 2022;604(7906):517–524. doi: 10.1038/s41586-022-04618-z
70. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94–101. doi: 10.1038/s41586-020-1943-3
71. Booth MJ, Ost TWB, Beraldi D, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841–1851. doi: 10.1038/nprot.2013.115
72. Liu Y, Hu Z, Cheng J, et al. Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. Nat Commun. 2021;12(1):618. doi: 10.1038/s41467-021-20920-2
73. Yu M, Hon GC, Szulwach KE, et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc. 2012;7(12):2159–2170. doi: 10.1038/nprot.2012.137
74. Library prep + barcoding kits [Internet]. Pacific Bioscience. 2025 [cited 2025 Oct 15]. Available from: https://www.pacb.com/products-and-services/consumables/library-prep-and-barcoding-kits/
75. Dahl JA, Robertson AB, Klungland A. Inventors; Oslo Universitetssykehus HF, assignee. Methods and kits for detection of methylation status. US20140363815A1. 2012.
76. Giehr P, Kyriakopoulos C, Lepikhov K, et al. Two are better than one: hPoxBS - hairpin oxidative bisulfite sequencing. Nucl Acids Res. 2018;46(15):e88–e88. doi: 10.1093/nar/gky422
77. Halliwell DO, Honig F, Bagby S, et al. Double and single stranded detection of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore sequencing. Commun Biol. 2025;8(243). doi: 10.1038/s42003-025-07681-0
    - **The authors provide a detailed overview of ONT sequencing and its potential for 6-base sequencing.**
78. Sigurpalsdottir BD, Stefansson OA, Holley G, et al. A comparison of methods for detecting DNA methylation from long-read

sequencing of human genomes. Genome Biol. 2024;25(69). doi: 10.1186/s13059-024-03207-9
79. Product Specifications [Internet]. Oxford Nanopore Technologies. 2025 [cited 2025 Aug 20]. Available from: https://nanoporetech.com/products/specifications
80. Nanopore sequencing accuracy Internet]. Oxford Nanopore Technologies. 2025 [cited 2025 Aug 20]. Available from: https://nanoporetech.com/platform/accuracy
81. Xie S, Leung AWS, Zheng Z, et al. Applications and potentials of nanopore sequencing in the (epi)genome and (epi)transcriptome era. Innovation (Camb). 2021;2(4):100153. doi: 10.1016/j.xinn.2021.100153
82. Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–1365. doi: 10.1038/s41587-021-01108-x
83. Input DNA/RNA QC [Internet]. Oxford Nanopore Technologies; 2024 [cited 2025 Aug 20]. Available from: https://nanoporetech.com/document/input-dna-rna-qc
84. Cell free DNA (cfDNA) information repository [Internet]. Oxford Nanopore Technologies. 2025 [cited 2025 Aug 20]. Available from: https://nanoporetech.com/document/cell-free-dna-cfdna-info-sheet
85. Füllgrabe J, Gosal WS, Creed P, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 2023;41(10):1457–1464. doi: 10.1038/s41587-022-01652-0
•• **The 6-base sequencing method from biomodoal.**
86. biomodal. Genetic and epigenetic study of formalin-damaged (FFPE) DNA with 6-base sequencing [Internet]. 2024 [cited 2025 Aug 20]. Available from: https://biomodal.com/poster/genetic-and-epigenetic-study-of-formalindamaged-ffpe-dna-with-6-base-sequencing/
87. DNA Sequencing Costs: Data [Internet]. National Human Genome Research Institute; 2023 [cited 2025 Aug 20]. Available from: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
88. Carss K, Halldorsson BV, Hou L, et al. Whole-genome sequencing of 490,640 UK Biobank participants. Nature. 2025;645(8081):692–701. doi: 10.1038/s41586-025-09272-9
89. Claussnitzer M, Cho JH, Collins R, et al. A brief history of human disease genetics. Nature. 2020;577(7789):179–189. doi: 10.1038/s41586-019-1879-7
90. Silveira AB, Houy A, Ganier O, et al. Base-excision repair pathway shapes 5-methylcytosine deamination signatures in pan-cancer genomes. Nat Commun. 2024;15(9864). doi: 10.1038/s41467-024-54223-z
91. Mort M, Ivanov D, Cooper DN, et al. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat. 2008;29(8):1037–1047. doi: 10.1002/humu.20763
92. Rheinbay E, Nielsen MM, Abascal F, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020;578(7793):102–111. doi: 10.1038/s41586-020-1965-x
93. Rusmntratip V, Sowers LC. An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. Proc Natl Acad Sci USA. 2000;97(26):14183–14187. doi: 10.1073/pnas.97.26.14183
94. Bhat A, Ghatage T, Bhan S, et al. Role of transposable elements in genome stability: implications for health and disease. J Mol Sci. 2022;23(14):7802. doi: 10.3390/ijms23147802
95. Dumitrache LC, McKinnon PJ. Out of line: transposons, genome integrity, and neurodegeneration. Neuron. 2022;110(20):3217–3219. doi: 10.1016/j.neuron.2022.09.026
96. Jang HS, Shah NM, Du AY, et al. Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genet. 2019;51(4):611–617. doi: 10.1038/s41588-019-0373-3
97. Bell RJA, Rube HT, Xavier-Magalhaes A, et al. Understanding TERT promoter mutations: a common path to immortality. Mol Cancer Res. 2016;14(4):315–323. doi: 10.1158/1541-7786.MCR-16-0003
98. Elliott K, Larsson E. Non-coding driver mutations in human cancer. Nat Rev Cancer. 2021;21(8):500–509. doi: 10.1038/s41568-021-00371-z
99. Baylin SB, Jones PA. Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol. 2016;8(9):a019505. doi: 10.1101/cshperspect.a019505
100. Sjostrom M, Zhao SG, Levy S, et al. The 5-hydroxymethylcytosine landscape of prostate cancer. Cancer Res. 2022;82(21):3888–3902. doi: 10.1158/0008-5472.CAN-22-1123
101. Wu SL, Yang L, Huang C, et al. Genome-wide characterization of dynamic DNA 5-hydroxymethylcytosine and TET2-related DNA demethylation during breast tumorigenesis. Clin Epigenet. 2024;16(125). doi: 10.1186/s13148-024-01726-7
102. Lee DD, Komosa M, Nunes NM, et al. Dna methylation of the TERT promoter and its impact on human cancer. Curr Opin Genet Dev. 2020;60:17–24. doi: 10.1016/j.gde.2020.02.003
103. Oishi N, Vuong HG, Mochizuki K, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of thyroid carcinomas with TERT promoter mutations. Endocr Pathol. 2020;31:359–366. doi: 10.1007/s12022-020-09652-z
104. Menghi F, Banda K, Kumar P, et al. Genomic and epigenomic BRCA alterations predict adaptive resistance and response to platinum-based therapy in patients with triple negative breast and ovarian carcinomas. Sci Transl Med. 2022;14(652):eabn1926. doi: 10.1126/scitranslmed.abn1926
105. Feinberg A, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004;4(2):143–153. doi: 10.1038/nrc1279
106. Guvenc C, Neckebroeck F, Antoranz A, et al. Bona fide tumor suppressor genes hypermethylated in melanoma: a narrative review. Int J Mol Sci. 2021;22(19):10674. doi: 10.3390/ijms221910674
107. Koch A, Joosten SC, Feng Z, et al. Analysis of DNA methylation in cancer: location revisited. Nat Rev Clin Oncol. 2018;15(7):459–466. doi: 10.1038/s41571-018-0004-4
108. Smith RG, Pishva E, Shireby G, et al. A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. Nat Commun. 2021;12(1):3517. doi: 10.1038/s41467-021-23243-4
109. Kraucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–930. doi: 10.1126/science.1169786
110. Szulwach K, Li X, Li Y, et al. 5-hmc–mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci. 2011;14(12):1607–1616. doi: 10.1038/nn.2959
111. Zhao J, Gu T, Gao C, et al. Brain 5-hydroxymethylcytosine alterations are associated with Alzheimer's disease neuropathology. Nat Commun. 2025;16(2842). doi: 10.1038/s41467-025-58159-w
112. Chuang YH, Lu AT, Paul KC. Longitudinal epigenome-wide methylation study of cognitive decline and motor progression in Parkinson's disease. J Parkinsons Dis. 2019;9(2):389–400. doi: 10.3233/JPD-181549
113. Al-Mahdawi S, Virmouni SA, Pook MA. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. Front Neurosci. 2014;8:8. doi: 10.3389/fnins.2014.00397
114. Wang J, Tang J, Lai M, et al. 5-hydroxymethylcytosine and disease. Mutat Res Rev Mutat Res. 2014;762:167–175. doi: 10.1016/j.mrrev.2014.09.003
115. Nowell PC. The clonal evolution of tumor cell populations: acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. Science. 1976;194(4260):23–28. doi: 10.1126/science.959840
116. Feinberg A, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nat Rev Genet. 2006;7(1):21–33. doi: 10.1038/nrg1748
• **The authors provide a compelling and detailed argument for the epigenetic progenitor hypothesis.**
117. Zhang F, Zhang X, Zhang H, et al. Pan-precancer and cancer DNA methylation profiles revealed significant tissue specificity of interrupted biological processes in tumorigenesis. Epigenetics. 2023;18(1):2231222. doi: 10.1080/15592294.2023.2231222
118. Kato S, Lippman SM, Flaherty KT, et al. The conundrum of genetic "drivers" in benign conditions. J Natl Cancer Inst. 2016;108(8):djw036. doi: 10.1093/jnci/djw036

119. Ahmed M, Soares F, Xia JH, et al. Crispri screens reveal a DNA methylation-mediated 3D genome dependent causal mechanism in prostate cancer. Nat Commun. 2021;12(1781). doi: 10.1038/s41467-021-21867-0
120. Hazelett DJ, Rhie SK, Gaddis M, et al. Comprehensive functional annotation of 77 prostate cancer risk loci. PLoS Genet. 2014;10(1):e1004102. doi: 10.1371/journal.pgen.1004102
121. Dadaev T, Saunders EJ, Newcombe PJ, et al. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. Nat Commun. 2018;9(1):2256. doi: 10.1038/s41467-018-04109-8
122. Schumacher FR, Olama AAA, Bendt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50(7):928–936. doi: 10.1038/s41588-018-0142-8
123. Galasso M, Pozza ED, Chignola R. The rs1001179 SNP and CpG methylation regulate catalase expression in chronic lymphocytic leukemia. Cell Mol Life Sci. 2022;79(10):521. doi: 10.1007/s00018-022-04540-7
124. Jaiswal S, Ebert B. Clonal hematopoiesis in human aging and disease. Science. 2019;366(6465):eaan4673. doi: 10.1126/science.aan4673
125. Kirmani S, Huan T, Van Amburg JC, et al. Epigenome-wide DNA methylation association study of CHIP provides insight into perturbed gene regulation. Nat Commun. 2025;16(4678). doi: 10.1038/s41467-025-59333-w
126. Centers for Medicare & Medicaid Services. Screening for colorectal cancer - blood-based biomarker tests [Internet]. 2021 [cited 2025 Aug 20]. Available from: https://www.cms.gov/medicare-coverage-database/view/ncacal-decision-memo.aspx
127. Creyghton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS. 2010;107(50):21931–21936. doi: 10.1073/pnas.1016071107
128. Pradeepa MM. Causal role of histone acetylations in enhancer function. Transcription. 2016;8(1):40–47. doi: 10.1080/21541264.2016.1253529
129. Barnett KR, Decato BE, Scott TJ, et al. Atac-Me captures prolonged DNA methylation of dynamic chromatin accessibility loci during cell fate transitions. Mol Cell. 2020;77(6):1350–1364.e6. doi: 10.1016/j.molcel.2020.01.004
130. Schlesinger F, Smith AD, Gingeras TR, et al. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. Genome Res. 2013;23(10):1601–1614. doi: 10.1101/gr.157271.113
131. Hardwick JS, Dhir S, Kirchner A, et al. Scotch-seq reveals that 5-hydroxymethylcytosine encodes regulatory information across DNA strands. PNAS. 2025;122(31). doi: 10.1073/pnas.251220412
132. Skene PJ, Illingworth RS, Webb S, et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell. 2010;37(4):457–468. doi: 10.1016/j.molcel.2010.01.030
133. Nishiyama A WI, Shikimachi R. CdcA7 is an evolutionarily conserved hemimethylated DNA sensor in eukaryotes. Sci Adv. 2024;10(34). doi: 10.1126/sciadv.adp5753
134. Shinkai A, Hashimoto H, Shimura C, et al. The C-terminal 4CXXC-type zinc finger domain of CDCA7 recognizes hemimethylated DNA and modulates activities of chromatin remodeling enzyme HELLS. Nucl Acids Res. 2024;52(17):10194–10219. doi: 10.1093/nar/gkae677
135. Ibrahim J, Peeters M, Camp GV, et al. Methylation biomarkers for early cancer detection and diagnosis: current and future perspectives. Eur J Cancer. 2023;178:91–113. doi: 10.1016/j.ejca.2022.10.015
136. Spector BL, Harrell L, Sante D, et al. The methylome and cell-free DNA: current applications in medicine and pediatric disease. Pediatr Res. 2023;94:89–95. doi: 10.1038/s41390-022-02448-3
137. Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017 Nov 16;171(5):1029–1041.e21. doi: 10.1016/j.cell.2017.09.042
138. Fiala C, Diamandis EP. Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. BMC Med. 2018;16(166). doi: 10.1186/s12916-018-1157-9
139. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32(9):1167–1177. doi: 10.1016/j.annonc.2021.05.806
140. Wang P, Song Q, Ren J, et al. Simultaneous analysis of mutations and methylations in circulating cell-free DNA for hepatocellular carcinoma detection. Science. 2022;14(672). doi: 10.1126/scitranslmed.abp8704
    - **This paper highlights the value of a multi-omic (DNA + epigenetics) approach to cancer detection.**
141. Conway AM, Pearce SP, Clipson A, et al. A cfDNA methylation-based tissue-of-origin classifier for cancers of unknown primary. Nat Commun. 2024;15(1):3292. doi: 10.1038/s41467-024-47195-7
142. Guler GD, Nig Y, Coruh C, et al. Plasma cell-free DNA hydroxymethylation profiling reveals anti-PD-1 treatment response and resistance biology in non-small cell lung cancer. J Immunother Cancer. 2024;12(1):e008028. doi: 10.1136/jitc-2023-008028
    - **This is a good example of how analysis of 5hmC can provide clinically valuable information.**
143. Lee MK, Brown MS, Wilkins OM, et al. Distinct cytosine modification profiles define epithelial-to-mesenchymal cell-state transitions. Epigenomics. 2022;14(9):519–535. doi: 10.2217/epi-2022-0023
144. Murcott B, Honig F, Halliwell DO, et al. Colorectal cancer progression to metastasis is associated with dynamic genome-wide biphasic 5-hydroxymethylcytosine accumulation. BMC Biol. 2025;23(100). doi: 10.1186/s12915-025-02205-y
145. Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. Cell. 2021;37(11):1012–1027. doi: 10.1016/j.tig.2021.05.002
146. Smith AR, Smith RG, Pishva E, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenet. 2019;11(1):52. doi: 10.1186/s13148-019-0636-y
147. Wang JY, Song YD, Zhang DL. Detection of early Parkinson's disease using 5-hydroxymethylcytosine episignatures in blood DNA. Age Ageing. 2025;54(6):afaf147. doi: 10.1093/ageing/afaf147