Detection Methods of Cytosine and Thymine Modifications in DNA

Mark Berney[1] and Joanna F. McGouran[1,*]

[1]School of Chemistry, Trinity College Dublin, University of Dublin, Dublin, Ireland
*To whom correspondence should be addressed. Tel: +35318961422 Email: jmcgoura@tcd.ie

Abstract

Methylation of cytosine at the 5-position is a common epigenetic marker in mammalian DNA, and plays an important role in regulating gene expression. Oxidised derivatives of 5-methylcytosine have recently been discovered. As well as being intermediates in an active demethylation pathway, some of these oxidised derivatives may function as epigenetic markers in their own right. Oxidised derivatives of thymine are also known as products of DNA damage. There is evidence however that one such derivative, 5-hydroxymethyluracil, may play an epigenetic role. There is a pressing need to learn more about these modifications, due to the role epigenetic markers play in development, and diseases such as cancer. This emerging area of research requires methods for detecting cytosine and thymine modifications in DNA with a high degree of accuracy and sequence specificity. This review will introduce the biochemistry of cytosine and thymine modifications, and discuss new and established detection methods which have been developed to overcome the high degree of difficulty associated with studying these modifications in DNA.

Introduction: Cytosine and Thymine Modifications

DNA codes for all the proteins necessary for life, but the DNA sequence is not the sole determinant of all the phenotypic traits of cells and organisms. Transcription of DNA is tightly regulated by epigenetic modifications, which play an important role in controlling when and where specific genes are expressed.[1,2] Epigenetic modifications regulate important processes such as cellular differentiation, and are also implicated in various diseases including cancer, autism spectrum disorder, and other developmental diseases such as Rett syndrome.[3–5] Epigenetic control of gene expression is often mediated through covalent modification of DNA itself, or of the histone proteins which pack DNA in the nucleus. These modifications do not result in changes to the DNA sequence, but can affect the binding of certain proteins to the DNA, including transcription factors and proteins which regulate chromatin structure.[6–8]

In mammals and many other eukaryotes, the most common epigenetic covalent modification of DNA is methylation of cytosine (C) at the 5-position. Such modifications can be inherited, but can also be enzymatically introduced and removed in response to stimuli.[9] 5-Methylcytosine (mC) is introduced by the methylation of cytosine residues by DNA-methyltransferases (DNMTs) with an S-adenosylmethionine cofactor. 5-Methylcytosine is most often found in the context of 5'-cytosine-phosphate-guanine-3' (CpG) dinucleotides. Regions containing a high density of CpG sequences are found in about 72% of promoters in the human genome.[10] Methylation of these regions causes

transcriptional inactivation of the associated gene.[9] One of the roles which epigenetics plays in cancer is the silencing of tumour suppressor genes by the methylation of CpG sites in their promoters.[11] DNMT inhibitors are thus used as anti-cancer drugs.[12] Cytosine methylation may also play a role in regulating chromatin folding.[13]

It was previously assumed that demethylation of mC occurred only by passive dilution, the replication of cells without maintenance of methylation patterns. However the recent discovery of oxidised forms of mC has revealed a new active demethylation pathway, summarised in Figure 1. 5-Methylcytosine can be converted to 5-hydroxymethycytosine (hmC) by the ten-eleven translocation dioxygenase (TET) family of enzymes, which includes TET1, TET2, and TET3. A dioxygen molecule is transferred to α-ketoglutarate and mC *via* reactive Fe(III)/Fe(IV) intermediates to give succinate and hmC.[14] 5-Hydroxymethylcytosine can be further oxidised by TET enzymes to 5-formylcytosine (fC) and 5-carboxylcytosine (caC).[15] both of which can be excised by thymine-DNA-glycosylase (TDG) to give an abasic site.[16] Unmodified cytosine can then be restored by base excision repair (BER). It has also recently been shown by Carell and co-workers that fC and potentially also caC can be directly converted to C through deformylation or decarboxylation.[17,18] The AID/APOBEC family of enzymes can deaminate mC to T, resulting in mismatched T:G base pairs which are then repaired by BER.[19] There is also evidence for AID/APOBEC induced deamination of hmC *in vivo,*[20,21] although a study using purified AID/APOBECs found that they show low activity on mC compared to C, and caused no detectable deamination of hmC *in vitro.*[19]

There is evidence that oxidised mC derivatives serve not just as intermediates in this demethylation pathway, but also as epigenetic markers in their own right.[22] Modifications at the cytosine 5-position protrude into the major groove of DNA and are thus available for interaction with DNA-binding molecules.[2] Differences in their hydrogen-bonding properties, as well as steric differences, allow differentiation between them. NMR studies have shown that the formyl group in fC and the carboxyl group in caC are held rigidly in the plane of the cytosine ring due to an intramolecular hydrogen bond with the amino group at the 4-position.[23] On the other hand the hydroxy group in hmC has been found in two different conformations in crystal structures[23] and thus appears to rotate freely. This exaggerates what would otherwise be a subtle difference in the steric properties of hmC and fC, allowing for easier discrimination between these modifications by DNA-binding proteins.

Some hmC modifications, particularly those at promoters and at poised and active enhancers, appear to be stable and not subject to further oxidation by TET enzymes.[24] 5-Hydroxymethylcytosine is also enriched in exons and near transcriptional start sites.[25] 5-Hydroxmethylcytosine has been found to be especially abundant in the brain, particularly in Purkinje neurons.[26] Embryonic stem cells also exhibit elevated hmC levels. Certain cancer cell types on the other hand have lower than normal genomic hmC content.[26] All this suggests that hmC plays a role in gene regulation and that cell-type specific regulation of TET dioxygenases may be an important mediator of epigenetic control.[22]

5-Formylcytosine appears to be a semi-permanent marker in some cases as well. For example it was observed in one study that TDG removed only 50% of fC at specific genomic sites.[27] 5-Formylcytosine

is known to preferentially occur at poised enhancers in mouse embryonic stem cells,[28] and several proteins have been identified, including transcription factors and chromatin regulators, which show a preference for binding to regions containing fC.[29] Lysine residues in histone proteins have been shown to form imine crosslinkages with fC,residues in DNA, which may play a role in nucleosome positioning.[30,31] The activity of RNA polymerase II has been shown to be significantly affected by the presence of fC or caC in template DNA.[32] 5-Formylcytosine is also found in mitochondrial tRNA molecules, where it modulates codon-anticodon interactions.[33]

WT1, a zinc-finger transcription factor protein, binds most strongly to caC, due to electrostatic and hydrogen bonding interactions of the negatively charged carboxylate of caC.[34] MAX, another transcription factor involved in multiple myeloma, has a greater affinity to its binding site when C or caC is present compared to mC, hmC, or fC[35] This suggests an epigenetic role for caC.

TET enzymes have also been shown to convert mC to hmC, fC, and caC in RNA,[36–40] suggesting a role for these modifications in regulating translation. The presence of hmC in RNA has indeed been shown to facilitate translation.[40]

5-Hydroxymethyluracil (hmU) and 5-formyluracil (fU) are oxidised derivatives of thymine that occur in DNA. These thymine modifications are known to result from DNA damage by reactive oxygen species (ROS), and both are associated with mutagenesis.[41,42] Furthermore, it was discovered by Carell and co-workers in 2014 that TET enzymes oxidise thymine to produce hmU (Figure 2) in mouse embryonic stem cells, and that the hmU bases influence the binding of some chromatin remodelling proteins and transcription factors.[43] This suggests that hmU also functions as an epigenetic marker.

Cytosine and thymine derivatives play an important biological role, not only in demethylation pathways, but are not yet fully understood. There is therefore a pressing need for further study of these modifications. Such research requires the development of robust methods which can detect pyrimidine modifications in genomic DNA with a high degree of accuracy and sensitivity. Levels of cytosine modifications in DNA depend on a number of factors, such as age, tissue type, and disease state.[26,44,45] As a representative example of the relative abundances of these modifications, in mouse embryonic stem cells levels of hmC, fC, and caC were found to be 0.13%, 0.002%, and 0.0003% of all cytosines respectively.[15] 5-Methylcytosine is more abundant, present at a level of >4% of all cytosines in most healthy human tissues.[46] Levels of hmU and fU in mouse embryonic stem cells were found to be 5% and 22% relative to fC levels respectively.[43]

Chemical Detection Methods

A large number of detection methods for cytosine modifications in DNA have been developed. Genome wide levels of all cytosine modifications can be determined using mass spectrometry based methods.[18,44,45,47–49] Chemical labelling of cytosine modifications to introduce easily ionisable groups has been used to improve the sensitivity of their detection by mass spectrometry.[41-46] Methods employing high performance liquid chromatography (HPLC),[56] or high performance capillary electrophoresis[57] have also been used to determine genome wide levels of cytosine modifications. A

method based on thin-layer chromatography was used to detect fC and caC in DNA for the first time in 2011.[15]

In addition to studying genome wide levels of cytosine modifications, it is also desirable to develop methods to study these modifications at specific loci. Some early chemical methods for the sequence specific detection of mC were based on the Maxam-Gilbert sequencing method, and have previously been reviewed by Balasubramanian and coworkers.[58] Ongoing developments include two different *N*-halogeno-*N*-sodiobenzenesulfonamide reagents have been used for the detection of mC in Maxam-Gilbert sequencing (Figure 3).[59] Treatment of DNA with *N*-sodio-*N*-chloro-*p*-nitrobenzenesulfonamide (**1**) in conjunction with I_2 causes selective iodination of C residues, while treatment of DNA with *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide (**2**) causes bromination of both C and mC bases. Treatment of the DNA samples with hot piperidine then causes strand cleavage at the halogenated sites. After gel electrophoresis, comparison of the results obtained after treatment with *N*-sodio-*N*-chloro-*p*-nitrobenzenesulfonamide (**1**) and I_2 with the results obtained after treatment with *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide (**2**) allows mC loci to be identified. *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide (**2**) can also react with hmC. Enzymatic glucosylation of hmC prevents this reaction, allowing hmC residues to be distinguished using this method. These reagents can also be used in conjunction with bisulfite treatment of DNA, which converts C residues to U but does not affect mC. During PCR amplification C is replicated as U and mC is replicated as C. *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide (**2**) can then selectively react with mC.[60] In another procedure taking advantage of the fact that after bisulfite treatment and PCR, G will be incorporated opposite mC sites, while A is incorporated opposite C sites, treatment with K_2WO_4/H_2O_2 can be used to induce selective strand cleavage at G for the identification of mC loci.[60] Strand cleavage at fC can be induced by treating DNA with hot piperidine. This can also be used to detect hmC by first oxidising hmC to fC.[61] While Maxam-Gilbert sequencing methods require gel electrophoresis which is time consuming and labour intensive, they provide robust sequence specific detection of mC and hmC.

The most widely used strategy for the sequence specific detection of mC is the chemical derivatisation of DNA using sodium bisulfite in a process known as bisulfite sequencing (BS-seq), first reported in a seminal work by Frommer *et al*. in 1992.[62,63] Treatment of DNA with sodium bisulfite leads to deamination of C to U but not of mC to T. Treatment with bisulfite therefore translates epigenetic information into a change in the sequence of canonical nucleobases, which can be detected using PCR followed by DNA sequencing, in which C will be read as U and mC will be read as C. While BS-seq is a robust method for the detection of mC, it cannot distinguish oxidised mC derivatives. Cytosine-5-methylsulfonate (CMS) is formed on treatment of hmC with bisulfite, and since CMS shows the same base-pairing selectivity for guanine, it reads as C.[64] Bisulfite converts fC and caC to U, after deformylation or decarboxylation, and therefore neither of these are distinguishable from unmodified C by this method (Table 1, Entry 1).[22]

This limitation can be overcome by the use of additional chemical derivatisation steps before bisulfite treatment. Oxidation of hmC to fC with potassium perruthenate (oxBS-seq) (Table 1, Entry 2), or reduction of fC to hmC with sodium borohydride (redBS-seq) (Table 1, Entry 3), followed by bisulfite

sequencing, allows fC and hmC levels to be determined by comparing oxidised/reduced samples with untreated ones.[65] Bis(acetoxy)iodobenzene enclosed in sodium dodecyl sulfate micelles, and 2-hydroxy-2-azaadamantane have been used as alternative reagents for the oxidation of hmC to fC to avoid the problem of DNA degradation by potassium perruthenate.[66] Cu(II) perchlorate, TEMPO and bipyridine also selectively oxidise hmC to fC.[61]

Another derivatisation step that has been employed to detect hmC is the glucosylation of this residue using β-glucosyltransferase. Subsequent treatment of the DNA with TET1 oxidises all 5-modified cytosine residues to caC, except hmC which is now protected. In bisulfite sequencing hmC is then the only base that reads as C (Table 1, Entry 4). This is known as TET-assisted bisulfite sequencing (TAB-seq).[67,68] 5-Formylcytosine can also be detected using a chemically assisted bisulfite sequencing method (fCAB-seq), in which fC is first protected from deamination by the formation of an oxime with ethylhydroxylamine. In bisulfite sequencing fC then reads as C, as do mC and hmC (Table 1, Entry 5). The location of mC/hmC bases can be determined using conventional bisulfite sequencing, and the location of fC bases can then be inferred by comparison.[28] Another variation of chemically assisted bisulfite sequencing is caCAB-seq, which can detect caC using an amide bond forming reaction with a xylene-based primary amine. Inclusion of an azide group in the primary amine allows for attachment of a biotin tag for affinity enrichment of DNA fragments containing caC. The biotin tag can subsequently be removed by cleavage of a disulfide bond present in the linker. Formation of the amide bond protects caC from decarboxylation and deamination upon treatment with bisulfite, meaning that it reads as C in bisulfite sequencing (Table 1, Entry 6).[69,70] Methylase-assisted bisulfite sequencing (MAB-seq) is a method in which an enzyme is employed to convert all unmodified C residues to mC. Upon treatment with sodium bisulfite, fC and caC are then the only bases that are converted to U, and can thus be identified, although they cannot be distinguished from each other (Table 1, Entry 7).[71,72] By treating the methylated DNA with $NaBH_4$ to convert fC residues to hmC prior to bisulfite sequencing, caC can be selectively detected as it is then the only base that reads as U (Table 1, Entry 8). This is known as caMAB-seq.[72]

		Readout for:				
Entry No.	Method	C	mC	hmC	fC	caC
1	BS-seq	T	C	C	T	T
2	oxBS-seq	T	C	T	T	T
3	redBS-seq	T	C	C	C	T
4	TAB-seq	T	T	C	T	T
5	fCAB-seq	T	C	C	C	T
6	caCAB-seq	T	C	C	T	C
7	MAB-seq	C	C	C	T	T
8	caMAB-seq	C	C	C	C	T

Table 1 - Bisulfite sequencing and its modifications

Bisulfite sequencing is currently considered the gold standard for detecting epigenetic cytosine modifications. Through additional chemical modification steps robust sequence specific detection of mC, hmC, fC, and caC can be achieved. There are number of drawbacks however. About 95% of the DNA is destroyed on treatment with bisulfite, so a large sample of DNA is often needed,[73] although there have been several reports of bisulfite sequencing analysis of DNA from single cells.[74–77] Incomplete conversion of C to U can lead to errors,[78] and the use of several chemical manipulations of the sample increases the chances for bias or contamination.[79] PCR amplification can also be problematic due to the reduction in sequence complexity on conversion of C to U.[80] Bisulfite sequencing is also labour intensive, since several steps are required. A number of modified procedures combining bisulfite treatment with, for example, restriction enzymes, modified PCR assays, embedding the DNA sample in agarose beads, or adaptation of the GoldenGate genotyping assay have been reported.[81–86] The development of bisulfite-free detection methods for cytosine modifications is still of significant interest however.

Chemical Methods Beyond Bisulfite Sequencing

Detection of mC

Beyond bisulfite sequencing, a number of novel chemical derivatisation strategies based on selective oxidation have been developed to detect mC. Treatment of DNA containing mC with osmium tetroxide causes oxidation of mC to give an osmate complex, while unmodified C does not react.[87] A bipyridine ligand modified with a linker attached to a fluorescent or electrochemically active group can coordinate to the osmate and allow detection of mC.[88] The exact loci of mC residues can be determined by treatment with hot piperidine, which causes strand cleavage at oxidised mC bases but not at C, followed by polyacrylamide gel electrophoresis.

Detection of mC by selective oxidation has also been achieved using V_2O_5, or $NaIO_4$ with LiBr.[89] Enzymatic oxidation of mC to hmC by TET, and labelling with an azide containing glucose derivative followed by biotin, has also been used. Existing hmC residues must be blocked by enzymatic glucosylation prior to TET oxidation of mC.[90] Direct electrochemical oxidation of mC is also a useful detection method.[91,92] A number of other methods based on electrochemical detection,[93–96] and FRET based methods for the detection of mC have also been reported.[97–99]

Derivatisation of DNA with *O*-allylhydroxylamine has been used by Carell and co-workers to detect mC.[100] *O*-allylhydroxylamine forms an adduct with both C and mC, but the adducts have different conformations due to steric clash between the allyl group of *O*-allylhydroxylamine and the methyl group of mC. The two adducts therefore show different base pairing selectivities, as the adduct formed from mC base pairs with G, but the adduct formed from C can base pair with either G or A. Cytosine and mC can thus be distinguished using a pyrosequencing method.

Detection of hmC

β-Glucosyltransferase has been used outside the context of bisulfite sequencing for detection of hmC. Glucosylation has been used in combination with restriction enzymes in several assays.[101–103] Genome wide levels of hmC can be determined by labelling with radiolabelled glucose.[104] Alternatively, after glucosylation of hmC, the DNA may be treated with sodium periodate to oxidise the glucose, and the resulting aldehydes then allow for attachment of biotin tags *via* the formation of oxime linkages. Biotin tags allow for detection of DNA fragments containing hmC using streptavidin in a pull-down assay.[25] This method is known as GLIB (glucosylation, periodate oxidation, biotinylation). Alternatively a glucose moiety containing an azide group can be enzymatically attached to hmC, allowing for the introduction of selective reaction with an alkyne-bearing biotin group. This method, known as hMe-Seal (hmC-selective chemical labelling) was developed by Song *et al*.[105,106] Nano-hmC-Seal is an optimised version of this procedure in which DNA is first fragmented and ligated with sequencing adaptors in a single step. 5-Hydroxymethylcytosine residues are then enzymatically labelled with an azide containing glucose derivative, followed by biotin. Enrichment and sequencing of the fragments allows detection of hmC-containing regions in scant samples of DNA.[107] Non-enzymatic biotinylation of hmC has been achieved using an alkyl sulfinate reagent in a reaction similar to the production of cytosine 5-methylsulfonate from hmC upon treatment with bisulfite.[108]

Glucosylated hmC has been detected both at a genomic level and sequence specifically using boronic acid moieties. Microspheres functionalised with phenylboronic acid, upon reaction with glucosylated hmC, form boronate esters. This reaction causes an increase in fluorescence intensity which can be used to quantify hmC levels in DNA.[109] Enzymatic glucosylation of hmC followed by derivatisation of the glucose moieties by reaction with a boronic acid has also been shown to inhibit DNA replication by Taq DNA polymerase. This has led to the development of a PCR assay for sequence specific detection of hmC by Jiang and co-workers.[110]

Glucosylation of hmC in conjunction with a boronic acid has been used to develop a sensitive electrochemical biosensor for the detection of this cytosine modification in target sequences. Probe oligonucleotides were first immobilised on the surface of an electrode and complementary strands of the DNA sample under analysis hybridise to the probes. β-Glucosyltransferase is then used to glucosylate any hmC residues present. The electrode is incubated with a solution containing 1,4-phenyldiboronic acid, which binds to the glucose moieties. The 1,4-phenyldiboronic acid then immobilises alkaline phosphatase, which catalyses the hydrolysis of p-nitrophenyl phospate to *p*-nitrophenol. The production of *p*-nitrophenol can be detected as it is electrochemically active.[111] Another biosensor for hmC that also utilises alkaline phosphatase has also been developed. In this method DNA strands were first immobilised on a magnetic bead. M.HhaI DNA-methyltransferase was then used to derivatise hmC with cysteamine. This allowed for the attachment of biotin, which was recognised by avidin-conjugated alkaline phosphatase. The bound alkaline phosphatase catalysed the production of ascorbic acid from 2-phosphoascorbic acid trisodium salt, and the ascorbic acid was detected photoelectrochemically.[112] Labelling of hmC with cysteamine by M.HhaI DNA-methyltransferase has also been employed in the development of a biosensor for hmC utilising horseradish peroxidase.[113]

Glucosylated hmC has been derivatised with ferroceneboronic acid *via* formation of a boronic ester by Wang and co-workers.[114] The ferrocene-labelled DNA acts as a quencher of the electrogenerated chemiluminescence produced by $Ru(bpy)_3^{2+}$ immobilised on the surface of an electrode. This allowed hmC levels in DNA to be quantified through measurement of the decrease in intensity of the luminescence. A similar method reported by Zhou and co-workers allows quantification of hmC in DNA through the switching on of electrogenerated chemiluminescence.[115] $Ru(phen)_3^{2+}$ complexes were immobilised on graphene oxide, and the carboxyl groups in the $Ru(phen)_3^{2+}$/graphene oxide composites were reacted with 3-aminophenylboronic acid, forming amide bonds. DNA fragments under analysis were hybridised with complementary strands immobilised on the surface of an electrode. Upon enzymatic glucosylation of any hmC residues present, and subsequent formation of a boronic ester linkage to the $Ru(phen)_3^{2+}$/graphene oxide composites, electrogenerated chemiluminescence resulted. Electrogenerated chemiluminescence has also been used in an immunosensor for hmC.[116]

$[Ru(NH_3)_6]^{3+}$ has also been used for the detection of mC and hmC. DNA strands are first immobilised on the surface of an electrode. MspJI endonuclease is used to cleave strands containing mC and hmC. $[Ru(NH_3)_6]^{3+}$ binds to the anionic phosphate groups in the DNA strands and is detected electrochemically. The amount of $[Ru(NH_3)_6]^{3+}$ present depends on the length of the DNA strands, which indicates whether or not they have been cleaved by the MspJI endonuclease. Glucosylation of hmC residues protects them from recognition by the endonuclease, making the assay specific for mC.[117]

Okamoto and co-workers have developed a chemical method for the detection of hmC in which peroxotungstate, is used to selectively oxidise and deaminate hmC to trihydroxylated thymine (thT).[118] In a primer extension assay, A rather than G is then incorporated opposite thT, allowing the position of hmC residues to be determined.

Fluorescence resonance energy transfer (FRET) can be used to detect hmC and fC. 5-Hydroxymethylcytosine is oxidised to fC using $KRuO_4$, followed by labelling with hydroxylamine-BODIPY. The labelled DNA is then captured on cationic conjugated polymers (CCPs) *via* an electrostatic interaction with the negatively charged phosphodiester backbone. The CCPs have excellent light harvesting properties and can transfer energy to the BODIPY group *via* FRET. The use of FRET improves the signal to noise ratio compared to simple direct excitation of the BODIPY fluorophore, allowing for more sensitive detection.[119] CCPs have also been used in conjunction with bisulfite treatment in a FRET based detection method for mC.[99] Information about the distance between mC and hmC residues can be obtained using a technique based on FRET. Enzymatic labelling of hmC with an azide-containing glucose derivative allows for attachment of an alkyne-bearing Cy3 fluorophore to hmC residues. Subsequent treatment of the DNA sample in one pot with TET1, β-glucosyltransferase, and an azide-containing glucose derivative, then allows for labelling of mC residues with an alkyne-bearing Cy5 fluorophore. The presence of a FRET signal then indicates that an mC and a hmC residue are close together in the DNA sample. Denaturing of the DNA causes disappearance of the FRET signal if the mC and hmC residues are on complementary strands, but not if they are close together on the same strand.[120]

Detection of fC

The hMe-Seal method reported by Song *et al.*[105] for enrichment of DNA fragments containing hmC has been extended to the detection of fC. After protection of hmC by β-glucosyltransferase, fC can be selectively reduced to hmC by sodium borohydride, and the newly created hmC residues then enzymatically labelled with an azide containing glucose moiety, and alkyne-bearing biotin tags attached. This method for the detection of fC is known as 5-formylcytosine selective chemical labelling, or fC-Seal.[28] Direct labelling of fC with biotin *via* an oxime or hydrazone linkage has also been used.[61,121]

Selective labelling of fC has been achieved using a Friedlander type reaction with an azide containing derivative of 1,3-indandione (**3**) (FIgure 4).[122] An alkyne-bearing biotin group is then attached, allowing for affinity enrichment of fragments of DNA containing fC using streptavidin-coated beads. Fragments of DNA containing fC are subsequently released from the beads by cleavage of the linkers attaching the biotin groups. In PCR followed by DNA sequencing of these fragments, labelled fC residues read as T, allowing them to be identified by comparison with unlabelled samples. This method does not require the use of sodium bisulfite and thus avoids the problem of degradation of the DNA sample, and is useful for the analysis of bulk samples of DNA. The poor solubility of 1,3-indandione derivatives in water, and the need for purification steps to remove excess 1,3-indandione derivatives before PCR, make analysis of the genome of a single cell unfeasible using this method however. These issues have recently been overcome by the use of malonitrile (**4**) in place of 1,3-indandione derivatives for labelling of fC in a method known as CLEVER-seq (Figure 4).[123]

Wang *et al.* have recently reported a novel method for fC detection in which 2-(adamantyl)ethoxyamine (**5**) is used to label fC with an adamantane moiety *via* formation of an oxime linkage (Figure 4).[124] The adamantane can then be recognised by a macrocycle known as CB7 through a host-guest binding interaction. The bulky CB7-adamantane complex acts as a roadblock to enzymes that read DNA, such as a restriction endonuclease or DNA polymerase. This allows fC loci to be determined using a primer extension assay. The potential for adapting this intriguing method to quickly measure global levels of fC by covalently linking a fluorophore to the CB7 macrocycle has yet to be explored.

Another promising chemical derivatisation strategy for the detection of fC both at the genomic level and sequence specifically is the labelling of DNA with a trimethylindole derivative (**6**). These react with fC to produce hemicyanine-like chromophores (Figure 4).[125] Quantitative measurement of fC levels can then be achieved by measurement of the intensity of the fluorescence emission of the sample. Site specific detection of fC can also be achieved using a primer extension assay, since the hemicyanine-modified nucleobases act as a roadblock to Klenow DNA polymerase, which can usually bypass fC.

2-(5-Chlorobenzo[d]thiazol-2-yl) acetonitrile (CBAN) (**7**) can be used to label fC, (Figure 4). Importantly, this allows for detection that is selective for fC over the structurally similar fU modification. CBAN can also react with fU, but does not form a cyclised fluorescent product as is the case upon reaction with fC. The fluorescence of the CBAN-labelled fC residues allows for quantification of fC levels in DNA. Also, since labelling with CBAN removes the hydrogen bond donating exocyclic amino group of fC, the

base pairing properties are altered. Therefore, after labelling of a DNA sample with CBAN, followed by PCR amplification and DNA sequencing, fC residues read as T. Comparison of labelled and unlabelled samples therefore allows sequence-specific detection of fC.[126] In a further development of this strategy, a reagent named azi-BP (**8**) was used to label fC (Figure 4). The azide group in azi-BP allowed for installation of a biotin moiety for enrichment of fC-containing DNA fragements. The large azi-BP-biotin label acts as a roadblock to a DNA polymerase, allowing fC to be detected with sequence specificity using a qPCR assay. Additionally, labelled fC residues base pair with A rather than G, allowing them to be identified using Sanger sequencing.[127]

Xu *et al.* demonstrated that fC can be labelled with a 2-hydrazinyl-N-(pyren-1-yl)acetamide fluorophore (**9**) *via* formation of a hydrazone linkage (Figure 4). This allows for the determination of fC levels in a DNA sample through measurement of fluorescence intensity. Furthermore, the authors showed that when two fC residues are present in a symmetric CpG site in dsDNA, the two adjacent pyrene groups form an excimer, which leads to a shift in the emission wavelength and increased intensity of the fluorescence. This allows the levels of isolated fC residues, and fC in symmetric CpG sites to be determined.[128]

Detection of caC

While methods such caMAB-seq, the use of antibodies, or nanopore sequencing have been developed for the detection of caC, to the best of our knowledge there has thus far been no report of a bisulfite-free chemical detection method for this modification.

Exploiting Protein-DNA Interactions

The main alternative to these chemical derivatisation strategies is to use DNA-binding proteins that can recognise epigenetic cytosine bases. An early method for detecting mC was based on the use of restriction endonucleases that include CpG dinucleotides in their recognition sequences, and do not cleave DNA that is methylated at cleavage/recognition sites.[129] Recent advances have involved the combination of restriction enzymes with enzymatic glucosylation of hmC for the detection of this modification.[130–132] A limitation of restriction enzymes is that they only cleave DNA in specific sequence contexts. This specificity is useful however in applications however when only the methylation status of particular loci is of interest.

Another fruitful strategy for the detection of mC is the use of antibodies in a methylated DNA immunoprecipitation assay (MeDIP). DNA is first fragmented, typically by sonication, and then denatured. The resulting single strand fragments which contain mC are bound by monoclonal mC antibodies. The bound fragments can then be separated using immunoprecipitation protocols and analysed. Antibodies specific for each modified form of cytosine have been used in this kind of assay,[22] as have antibodies specific for cytosine 5-methylsulfonate.[25] Immunoprecipitation assays provide a straightforward method for analysis of cytosine modifications, but they are not quantitative, and the resolution is dependent on the size of the DNA fragments, as any number of mC residues in the fragment will lead to a positive signal. Antibodies have also been used in electrochemical immunosensors to detect mC [133] and hmC.[134]

Methyl-CpG-binding domains (MBDs) of MeCP2 proteins can also be used to bind mC in the context of CpG dinucleotides, and can be tethered to green fluorescent protein to allow for detection. Furthermore, the tethering of a zinc finger to the green fluorescent protein can allow targeting of a specific DNA sequence by the MBD.[135] A zinc finger fused with luciferase has also been used in combination with an MBD for sensitive detection of mC.[136] Alternatively, MBDs can be used to precipitate densely methylated DNA fragments which can then be sequenced.[137] J-binding protein 1, found in trypanosomes, can recognise glucosylated hmC.[138] In an interesting development, an artificial, fluorophore-labelled, phosphopeptide has been created by rational design to bind selectively to mC.[139]

Transcription activator-like effectors (TALEs), a type of protein found in Xanthomonas bacteria, have been used to recognise cytosine derivatives with programmable sequence specificity.[140] Both the *C*-terminal and the *N*-terminal regions are involved in binding to DNA, but the sequence specificity of TALEs is derived from the DNA-binding domain, which consists of repeat units 33-35 amino acids in length. The 12th and 13th amino acids in each repeat unit are found in a loop between two α-helices and constitute the repeat variable diresidue (RVD). The four naturally occurring RVDs recognise the four canonical nucleobases through hydrogen-bonding interactions. TALEs have been engineered which contain additional mutant RVDs that recognise modified cytosine nucleobases. So far mC and hmC have been selectively detected in this way.[141–144] Also, a mutant TALE which recognises all cytosine nucleobases except caC has been developed.[145]

The Klimašauskas and Weinhold laboratories have developed a strategy for the analysis of the methylation status of CpG sites in which a DNA-methyltransferase is used to covalently label DNA.[146–148] Cytosine-5 methyltransferase SssI is an enzyme that has been engineered to work with synthetic S-adenosylmethionine analogues for the labelling of C with functional groups other than a methyl group. This has been used to install amine or azide modifications at the 5-position of cytosine. The DNA sample is first fragmented by sonication. Unmethylated and hemimethylated CpG sites are then enzymatically labelled, while methylated CpG sites are unaffected. The azide or amino groups on the labelled cytosine residues allow for the attachment of biotin tags. The biotin groups were attached *via* linkers that contained a cleavable S-S bond to allow for detachment of DNA fragments after enrichment using streptavidin beads. DNA fragments were then amplified by PCR and sequenced or analysed on a microarray. This allowed for the identification of unmethylated CpG sites.[149] Tethered oligonucleotide primed sequencing (TOP-seq) is a further development of this strategy.[150] Unmethylated and hemimethylated CpG sites are labelled with an azide functional group using Cytosine-5 methyltransferase SssI. A double stranded DNA oligonucleotide containing an alkyne group is then attached using a copper catalysed alkyne-azide cycloaddition. This tethered oligonucleotide can then act as a primer for a DNA polymerase, *via* a mechanism which is not fully understood, to produce DNA strands that include unmethylated and hemimethylated CpG sites and their adjacent regions.

Another DNA-methyltransferase which specifically methylates hemimethylated CpG sites, but not unmodified or hydroxymethylated CpG sites, has been used to distinguish between C, mC, and hmC. DNA is first fragmented by digestion with restriction enzymes. The fragments are ligated with hairpin-shaped adaptors and then treated with a DNA polymerase which extends the DNA from the adaptor,

resulting in self-complementary hairpin-shaped DNA fragments. DNMT1 is then used to methylate all hemimethylated CpG sites in the hairpin-duplexes, but does not affect unmethylated or hydroxymethylatd CpG sites. The sample is then treated with bisulfite and denatured, resulting in a long DNA strand. PCR and sequencing then allows C, mC, and hmC residues in CpG sites to be identified by comparison of the two ends of the self complementary sequence. C residues will read as U at both ends, while mC residues read as C at both ends, and hmC residues read as C at one end and U at the other end.[151] A similar method for the detection of hmC without the use of hairpin-shaped adaptors has also been reported.[152]

DNA polymerase enzymes can be used in single-molecule real time (SMRT) sequencing of DNA to detect cytosine modifications. The DNA polymerase is immobilised in a zero-mode waveguide (ZMW), which is a zeptolitre-volume cylindrical cavity.[153] The ZMW allows optical observation to be limited to a very small volume, so that the incorporation of a single nucleotide by the DNA polymerase can be observed. The DNA strand under analysis is used as a template for the synthesis of a complementary strand using nucleosides labelled with fluorophores attached to their terminal phosphate. Incorporation of each nucleotide can be observed as a fluorescent pulse which ends when the fluorophore is cleaved by the DNA polymerase and diffuses out of the ZMW. The wavelength of the fluorescence serves to identify the nucleobase. The duration of the fluorescent pulse (pulse width) and the time interval between successive pulses (interpulse duration) can be used to characterise the kinetics of the nucleotide incorporation by the DNA polymerase, which varies when cytosine modifications are present.[154] 5-Formylcytosine and caC show strong kinetic effects, while mC and hmC have more subtle effects. Detection of mC can be improved by first oxidising mC residues to caC using TET1.[155] Naegleria TET-like oxygenase has also been used for this purpose.[156] SMRT sequencing has also been used in combination with bisulfite treatment for detection of mC.[157] Labelling of hmC using the hMe-Seal method enhances the effect on the interpulse duration, improving detection of hmC.[158] Enzymatic diglucosylation of hmC has also been used.[159] The addition of such labelling steps, whilst increasing detection sensitivity, also introduces a small extra source of error however, as for example the oxidation of mC with TET1 proceeds with only 97% conversion.[68] Circular consensus sequencing, whereby the same DNA template is read multiple times by the polymerase, can be used to increase the accuracy of SMRT methods.[160] It is speculated that further improvements could be made by mutating the DNA polymerase. Also, since the kinetic effect of the cytosine modifications is spread over several nucleotides, and depends on the sequence context[154] improved algorithms for deconvoluting the data, particularly when there are two mC residues close together, could improve the accuracy of the technique.

Another method which shows excellent promise for detecting cytosine modifications is nanopore sequencing. Proteins form nanopores in a barrier separating two compartments filled with electrolyte. The nanopores allow ions to flow through them when an electric potential is applied. DNA can also migrate through the nanopores, and in doing so modulates the ionic current in a way that depends on the structure of the nucleotides present in the nanopore. Monitoring of the ionic current over time therefore allows sequencing of the DNA strand as it moves through the pore. Controlling the kinetics of

the DNA translocation through the pore improves the accuracy of the sequencing. This has been achieved by the use of phi29 DNA polymerase, which acts as a cap on the pore and slowly threads DNA through.[161] A nanopore sequencing method using a mutant form of the MspA porin protein found in Mycobacterium smegmatis has been shown to allow discrimination between C, mC, hmC, fC, and caC.[162] Pores formed from α-hemolysin have also been used to detect cytosine modifications.[163] Chemical labelling of mC and hmC has been used to improve the detectability of these modifications using α-hemolysin pores.[164,165] Aerolysin pores have also been used, and have the advantage that they are stable under harsh conditions and can be used in serum, which is desirable for diagnostic applications.[166] Also, since aerolysin pores are narrower than MspA and α-hemolysin pores, and they contain positively charged amino acids in their lumen which interact with DNA, capping with a DNA polymerase is not required to slow down translocation. Thus far C and mC have been distinguished using aerolysin pores. Commercially available nanopore sequencing instruments have been shown to be capable of detecting mC with 95% accuracy.[167,168] Improvements in the accuracy of nanopore sequencing methods may lead to their broad application in the field of epigenetics research. Such progress may result from engineering of the pore forming proteins, and improvements in the statistical models used to analyse the data. Beyond the use of protein-based nanopores, pores consisting of carbon nanotubes embedded in a lipid bilayer have been used to detect hmC, although this first required chemical modification of hmC.[169] Solid state nanopores have also been used.[170,171]

Oligonucleotide Probes

Recently oligonucleotide probes have begun to be explored as a strategy for the detection of cytosine modifications. They offer an advantage over protein based probes such as TALEs since they are more cost-effective, as they can be readily prepared using solid-phase synthesis. A large number of modified phosphoramidites have been developed to extend this method to the synthesis of oligonucleotides containing non-native functional groups.[172–175] Oligonucleotide probes also offer easily tuneable selectivity using probe sequences which are complementary to target sites.

DNA templated photoligations have been used to detect 5-methylcytosine. A probe strand containing a terminal 5-vinyl-2'-deoxyuridine with a hydrophobic group, undergoes a [2+2] cycloaddition with the carbon-carbon double bond of mC upon irradiation. Measurement of the fluorescence emission of the product allows easy detection. Reaction with mC is much more efficient than reaction with unmodified C due to a favourable hydrophobic interaction, illustrated by the arrow in Figure 5, between the methyl group of mC and the various hydrophobic moieties which have been tested in the 5-vinyl-2'-deoxyuridine residue.[176–178] Yamayoshi *et al.* have reported oligonucleotide probes containing a psoralen group which undergo a photocrosslinking with mC in preference to C in complementary strands, which can be observed by denaturing PAGE. Interestingly, this method was also successfully used to detect mC in dsDNA. The psoralen group can also undergo photoreaction with thymine residues adjacent to the target site, leading to off-target crosslinking. Further research is underway to improve the sequence specificity of this assay.[179] Oligonucleotides modified with a 3-cyanovinylcarbazole nucleoside also selectively photocrosslink to mC.[180]

A DNA templated light-activated reaction that can detect mC through selective oxidation has been developed by Nishimoto and co-workers.[181,182] An oligonucleotide probe tethered to a sensitizing 2-methyl-1,4-naphthoquinone chromophore causes one-electron oxidation of mC in a complementary strand upon irradiation. Treatment with piperidine leads to selective oxidative strand cleavage at mC. Oligonucleotides in which mC is replaced with one of the four canonical nucleobases are much less susceptible to strand cleavage.

A number of oligonucleotide probes based on fluorescent detection have been developed for the detection of cytosine modifications. Notably, Tucker and co-workers reported DNA probes containing an anthracene fluorophore which can discriminate between all four canonical nucleobases as well as mC through changes in the intensity of the fluorescence upon formation of a duplex with the strand under analysis.[183] This method was subsequently extended to the detection of hmC.[184] The efficiency of the detection was found to be dependent on the length of the alkyl linker attaching the anthracene to the probe oligonucleotide, raising the possibility that the probes could be further optimised for the detection of fC and caC, which has yet to be investigated. Recently oligonucleotide probes labelled with a 6-carboxyfluorescein or 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein fluorophore have also been used to detect mC.[185] In this case an increase in fluorescence intensity is observed when the probes are hybridised to a complementary strand containing mC rather than C. The authors propose that this is due to the positioning of the fluorophores in a more hydrophobic environment when mC is present. The use of hybrid locked nucleic acid (LNA)/DNA probes seems to enhance this effect, as they form more compact, less hydrated duplexes with the complementary strand. Another oligonucleotide probe containing a fluorescein moiety at the 5'-end and a dabsyl quencher group at the 3'-end has been used to distinguish between C and mC. Two peptide nucleic acid oligomers are first used to displace one DNA strand from each end, allowing the oligonucleotide probe to anneal to the region under analysis. A restriction enzyme is then used, leading to strong fluorescence when C is present as the probe is cleaved, separating the fluorescein from the quencher. When mC is present however the restriction enzyme cannot digest the sample and only weak fluorescence is observed.[186] In a related strategy, fluorescently labelled oligonucleotides can be used to detect mC in dsDNA by means of a strand exchange reaction, which does not occur in the absence of mC. The products of the strand exchange reaction can visualised using gel electrophoresis.[187]

Beyond the use of fluorescence detection, an electrochemically active ferrocene acetic acid group has been conjugated to the 3'-end of a probe oligonucleotide and used in combination with a restriction enzyme to detect mC.[188] A quartz crystal microbalance has also been used to observe the hybridisation of DNA to probe oligonucleotides immobilised on a surface. The DNA sample is first treated with a restriction endonuclease that cleaves all unmethylated recognition sites. Subsequently during PCR, only DNA fragments containing mC are amplified, and then hybridise to the probe oligonucleotides.[189]

Modified nucleotides have been used to discriminate between C and mC in primer extension assays. Cytosine modifications do not interfere with normal Watson-Crick base pairing, but O^6-modified 2'-deoxyguanosine derivatives are incorporated opposite C or mC with different efficiencies.[190] Modified nucleotides have also been incorporated into triplex forming oligonucleotides to detect modified

cytosines. It has been observed by Brown and co-workers that triplex forming oligonucleotides containing a synthetic *N*-methylpyrrolocytosine base show significantly lower triplex melting temperatures when bound to strands which have mC rather than C in a CpA sequence.[191] When hmC, fC or caC is present the melting temperature is also slightly different from that observed with mC.

Okamoto and co-workers have developed oligonucleotide probes that take advantage of the selective oxidation of mC by osmium tetroxide. A modified adenine residue linked to a bipyridine ligand can be incorporated into an oligonucleotide probe. After hybridisation of the probe to the DNA strand under analysis, the melting temperature of the duplex is much greater if the bipyridine ligand is coordinated to an osmate complex. These are known as interstrand crosslink formed by osmium and nucleic acid (ICON) probes, and allow mC to be detected in a sequence specific manner.[192] The formation of a mismatched base pair between the modified adenine residue and mC disrupts the pi-stacking of the duplex, and facilitates oxidation by osmium tetroxide. The formation of the interstrand crosslink upon coordination of the bipyridine ligand to the osmate complex blocks PCR, and so can also be detected using PCR based methods.[193] It has been shown that ICON probes also undergo crosslinking when hmC is present in place of mC.[194] The ICON probes used in these methods have also been immobilised on a microarray.[195] As an alternative to crosslinking bipyridine ligands, a nucleoside modified with a 6-dimethylamino-2-acylnaphthalene fluorophore shows significantly reduced fluorescence upon complexation of osmium to an mC residue base-paired to it.[196] Fluorimetric detection has also been used in combination with crosslinking by bipyridine ligands.[197] Methylation-specific fluorescence *in situ* hybridisation (MeFISH) uses ICON probes for *in vivo* visualisation of cytosine modifications. Fluorescence *in situ* hybridisation (FISH) is first performed using ICON probes labelled with a fluorophore. The sample is then treated with osmium tetroxide, and probes that don't crosslink are removed by denaturing. Comparison of the FISH and MeFISH images allows methylation status to be deduced.[198]

The utility of crosslinking oligonucleotide probes for the detection of fC has been demonstrated for the first time by Carell and co-workers.[27] An oligonucleotide probe modified with a hydroxylamine moiety forms an oxime linkage with fC in a complementary strand. A reporter strand hybridises to a complementary region in the DNA adjacent to the crosslinked probe. The probe and reporter strand are ligated using Ampligase. Two different primers, one specific for the ligated probe-reporter strand, and one specific for the genomic DNA are added, and droplet digital PCR is performed, allowing fC levels at a specific locus to be quantified with a high degree of sensitivity.

Detection of Thymine Modifications

The development of detection methods for hmU and fU in DNA is of significant importance to understand the biological effects of these modifications. In particular there is a need to develop detection methods that are selective for hmU *vs* hmC, and especially fU *vs* fC, as these modifications are present at similar levels in mammalian DNA.[43] Indeed, a potential shortcoming in many reports of detection methods for hmC and fC is that the specificity of the detection is demonstrated using

oligonucleotides containing other C modifications as negative controls, but negative controls consisting of oligonucleotides containing hmU and fU are not used.

There have been several reports in recent years of detection methods for hmU and fU. *O*-phenylenediamine derivatives can be used to label fU *via* formation of a benzimidazole linkage. This shows selectivity for fU over fC. Balasubramanian and co-workers employed a biotinylated *o*-phenylenediamine to tag fU residues in DNA.[199] DNA fragments containing fU can then be enriched using streptavidin coated magnetic beads, and amplified by PCR. This method has been extended to the detection of hmU by first oxidising hmU to fU with $KRuO_4$. In a further development of this strategy, detection of fU has been achieved using an *o*-phenylenediamine derivative (**10**) covalently linked to a naphthalimide fluorophore and a biotin tag (Figure 6). Fluorescence of the naphthalimide moiety is quenched by photoinduced electron transfer from the *o*-phenylenediamine group. This quenching no longer occurs upon formation of a benzimidazole linkage with fU. The biotin tag allows for enrichment of DNA fragments containing fU. This labelling strategy also allowed for detection of fU using a primer-extension assay, and could be used for *in vivo* imaging of fU in HeLa cells.[200] In manner similar to the benzimidazole labelling of fU with *o*-phenylenediamine derivatives, Hirose *et al.* have shown that 2-amino-4,5-dimethoxythiophenol (**11**) can be used for labelling of fU in DNA *via* formation of a fluorescent benzothiazol-2-yl group (Figure 6).[201,202]

Another naphthalimide derivative (**12**) has also been used to selectively fluorogenically label fU through formation of a hydrazone linkage (Figure 6). An azide moiety incorporated into the naphthalimide reagent allows for an alkyne-bearing biotin tag to be introduced for enrichment of DNA fragments containing fU. The naphthalimide group also acts as a roadblock to a DNA polymerase in a primer extension assay.[203] Furthermore, by first blocking fU by reaction with 4-nitro-*o*-phenylenediamine to form a non-fluorescent product, and altering the reaction conditions, this method has been adapted for the detection of fC.[204]

5-Formyluracil can be labelled with high selectivity using 4-hydrazinyl-7-nitrobenz-[2,1,3-d]-oxadiazole (**13**) (NBDH) (Figure 6), also *via* formation of a hydrazone linkage.[205] The resulting adducts are easily detectable due to their fluorescence. 5-Formyluracil has also been detected with sequence specificity using a primer extension assay after labelling with NBDH.

The labelling of fC with trimethylindole derivatives to produce hemicyanine-like chromophores is also effective for the detection of fU. 5-Formyluracil residues can be distinguished from fC residues by the difference in emission wavelength of the chromophores that result upon labelling.[125]

The use of β-glucosyltransferase to label hmU residues with an azide-containing glucose moiety has been demonstrated by Yu *et al.*[206] A biotin tag can then be attached using click chemistry. Labelling of hmC is avoided by first treating samples with recombinant TET1, which oxidises hmC (as well as mC and fC) to caC, but does not oxidise hmU. Labelling by β-glucosyltransferase is selective for hmU residues in mismatched hmU:G sites, which are formed by deamination of hmC and are removed *in*

vivo by base excision repair. 5-Hydroxymethyluracil residues formed by oxidation of thymine on the other hand exist in hmU:A sites. The use of this technique in conjunction with other methods therefore indicates the origin of hmU modifications.

Very recently, a method for the detection of hmU at single-base resolution has been reported. First, hmU is oxidised to fU using $KRuO_4$. Under mildly basic conditions, fU ionises, due to the presence of the electron withdrawing formyl group. Ionised fU residues can then base pair with G rather than A, allowing them to be detected by a polymerase-dependent single extension followed by PCR.[207]

Conclusion

The study of cytosine modifications is a rapidly expanding area, in which current detection methods for these modifications, such as bisulfite sequencing, have found broad applicability in the analysis of genomic DNA, while the potential of other detection methods has been demonstrated only in synthetic oligonucleotide models, as summarised in Table 2. The development of bisulfite-free chemical detection methods for cytosine modifications promises greater progress in the future. In particular the emergence of strategies based on chemical labelling which blocks the action of a DNA polymerase and allows for the determination of the exact loci of epigenetic modifications is an exciting development. Other novel technologies such as nanopore sequencing and SMRT sequencing show significant promise, as improvements in their accuracy are ongoing. Procedures in which cytosine modifications are detected by fluorescent labelling also show promise as convenient detection methods, although they do not have the single-base resolution associated with, for example, bisulfite sequencing and nanopore sequencing.

The use of oligonucleotide probes for detecting cytosine modifications is an emerging area which offers potential for the development of probes to detect epigenetic modifications at specific sites without the need for DNA sequencing. Oligonucleotide probes can potentially be immobilised on a microarray for the convenient high throughput screening of DNA samples. Indeed, DNA microarrays designed to detect C to U conversion in bisulfite treated DNA samples have already been commercialised.[208] Oligonucleotide probes could also be used *in vivo* in a manner similar to established fluorescent *in situ* hybridisation (FISH) techniques as has already been demonstrated with ICON probes.[198]

Oxidised thymine derivatives are now also detectable and distinguishable from their cytosine analogues using current methods, and an improved understanding of these modifications is likely to follow.

Further advances in the detection of cytosine and thymine modifications will facilitate research on their role in development and disease, potentially leading to new therapies. Analysis of DNA is also important in diagnostics and forensics, and sequence specific probes for epigenetic markers may find applications in these areas as well.

Funding

This work was supported by the Irish Research Council (GOIPG/2017/1453) and the Wellcome Trust TCD institutional strategic support fund. The authors declare no conflict of interest.

629

630

631

	Reagents/techniques demonstrated in oligonucleotide models. Single base resolution techniques shown in **bold**	Reagents/techniques successfully applied in genomic DNA Single base resolution techniques shown in **bold**
Chemical Detection Methods	**Maxam Gilbert type sequencing using N-halogeno-N-sodiobenzenesulfonamide reagents**[59] (mC, hmC) **Cleavage of DNA at fC sites by piperidine**[61]	**Use of N-sodio-N-bromo-m nitrobenzenesulfonamide in conjunction with bisulfite treatment**[60] (mC) **K_2WO_4/H_2O_2**[60] (mC) **BS-seq**[62,63] **oxBS-seq**[65] **redBS-seq**[65] **TAB-seq**[67,68] **fCAB-seq**[28] **caCAB-seq** (in conjunction with enrichment of DNA fragments containing caC by immunoprecipitation methods)[69,70] **MAB-seq**[71,72] **caMAB-seq**[72]
Chemical Methods Beyond Bisulfite Sequencing: Detection of mC	***O*-Allylhydroxylamine**[100] **Oxidation with OsO_4**[87,88] **Oxidation with V_2O_5 or $NaIO_4$ with LiBr**[89]	Oxidation to hmC by TET followed by labelling with an azide-bearing gluocse moeity[90]
Chemical Methods Beyond Bisulfite Sequencing: Detection of hmC	Biotinylation using an alkyl sulfinate reagent[108] Electrochemical biosensing methods[111–116] $[Ru(NH_3)_6]^{3+}$ in electrochemical sensing[117]	Labelling with radiolabelled glucose[104] GLIB[25] hMe-Seal[105,106] Nano-hmC-Seal[107] Glucosylation followed by reaction with phenylboronic acid microspheres[109] Glucosylation, reaction with boronic acid, and PCR assay[110] **Peroxotungstate**[118] Cationic conjugated polymers in FRET based detection[119] Labelling of mC and hmC for FRET assay[120]
Chemical Methods Beyond Bisulfite Sequencing: Detection of fC	Biotinylation *via* a hydrazone linkage[61] **2-(Adamantyl)ethoxyamine**[124] **Trimethylindole derivative**[125] 2-Hydrazinyl-N-(pyren-1-yl)acetamide[128]	fC-Seal[28] Biotinylation *via* an oxime linkage[121] **1,3-Indandione derivative**[122] **CLEVER-seq**[123] **CBAN** (single-base resolution detection demonstrated only in oligonucleotide models)[126] azi-BP[127]

Exploiting DNA-Protein Interactions	MBDs tethered to green fluorescent protein and zinc finger[135] (mC) **Detection of caC using TALEs**[145] **Nanopore sequencing methods**[162–166,169–171]	Zinc finger fused with luciferase in combination with an MBD[136] (mC) Precipitation of DNA fragments using an MBD[137] (mC) J-binding protein 1[138] (hmC) Artificial phosphopeptide[139] (mC) **Detection of mC and hmC using TALEs**[141–144] Use of DNMT to install a cleavable biotin[149] (unmodified C) **TOP-seq**[150] (unmodified C) **Simultaneous detection of mC and hmC using DNMT1**[151] **Detection of hmC using DNMT1**[152] **Oxidation of mC to caC followed by SMRT sequencing**[155,156] **SMRT bisulfite sequencing**[157] (mC) **hMe-Seal followed by SMRT sequencing**[158] (hmC) **Diglucosylation of hmC followed by SMRT sequencing**[159]
Oligonucleotide Probes	**5-Vinyl-2'-deoxyuridine with a hydrophobic group**[176–178] (mC) **Psoralen**[179] (mC) **3-Cyanovinylcarbazole**[180] (mC) **2-Methyl-1,4-naphthoquinone**[181,182] (mC) **Anthracene**[183,184] (mC, hmC) **Fluorescein and a dabsyl quencher**[186] (mC) **Strand exchange reactions**[187] (mC) Ferrocene acetic acid group[188] (mC) **Triplex forming oligonucleotides**[191] (mC)	**Fluorescein derivatives**[185] (mC) Quartz crystal microbalance[189] (mC) ***O*^{6}-modified 2'-deoxyguanosine derivatives**[190] (mC) **ICON probes**[192] (mC/hmC, only demonstrated in oligonucleotides for detection of hmC) **Crosslinking oligonucleotides for detection of fC**[27]
Detection of Thymine Modifications	Biotinylated *o*-phenylenediamine[199] (fU, or hmU after oxidation) ***o*-Phenylenediamine derivative covalently linked to a naphthalimide fluorophore and a biotin tag** (Used for *in vivo* imaging of fU in HeLa cells)[200] **Labelling with an azide-bearing naphthalimide group through a hydrazone linkage**[203] (fU) **Trimethylindole derivative**[125] (fU) **Labelling of hmU with an azide-containing glucose moeity**[206]	2-Amino-4,5-dimethoxythiophenol[201,202] (fU) **NBDH**[205] (fU) (single-base resolution detection demonstrated only in oligonucleotide models) **Oxidation of hmU to fU followed by single extension and PCR**[207]

Table 2 - **Summary of detection methods for cytosine and thymine modifications.**

References

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Gen.* **33,** 245–254 (2003).
2. Raiber, E.-A., Hardisty, R., van Delft, P. & Balasubramanian, S. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1,** 69 (2017).
3. Kinnaird, A., Zhao, S., Wellen, K. E. & Michelakis, E. D. Metabolic control of epigenetics in cancer. *Nat. Rev. Cancer* **16,** 694–707 (2016).
4. Jakovcevski, M. & Akbarian, S. Epigenetic mechanisms in neurodevelopmental and neurodegenerative disease. *Nat. Med.* **18,** 1194–1204 (2012).
5. Xie, P., Zang, L.-Q., Li, X.-K. & Shu, Q. An epigenetic view of developmental diseases: new targets, new therapies. *World J. Pediatr.* **12,** 291–297 (2016).
6. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464,** 1082–1086 (2010).
7. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17,** 487–500 (2016).
8. Cook, P. C. *et al.* A dominant role for the methyl-CpG-binding protein Mbd2 in controlling Th2 induction by dendritic cells. *Nat. Commun.* **6,** 6920 (2015).
9. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11,** 204–220 (2010).
10. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1412–1417 (2006).
11. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21,** 5427–5440 (2002).
12. Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **12,** 447–464 (2013).
13. Yoo, J., Kim, H., Aksimentiev, A. & Ha, T. Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation. *Nat. Commun.* **7,** 11045 (2016).
14. Hu, L. *et al.* Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155,** 1545–1555 (2013).
15. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
16. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).
17. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nat. Chem. Biol.* **14,** 72–78 (2018).
18. Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chemie Int. Ed.* **51,** 6516–6520 (2012).
19. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8,** 751–758 (2012).
20. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146,** 67–79 (2011).
21. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).
22. Kubik, G. & Summerer, D. Deciphering epigenetic cytosine modifications by direct molecular recognition. *ACS Chem. Biol.* **10,** 1580–1589 (2015).
23. Szulik, M. W. *et al.* Differential stabilities and sequence-dependent base pair opening dynamics of watson–crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-

carboxylcytosine. *Biochemistry* **54,** 1294–1305 (2015).

24. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14,** 341–356 (2013).

25. Pastor, W., Pape, U., Huang, Y. & Henderson, H. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473,** 394–397 (2011).

26. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

27. Su, M. *et al.* 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chemie Int. Ed.* **55,** 11797–11800 (2016).

28. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153,** 678–691 (2013).

29. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** R119 (2013).

30. Li, F. *et al.* 5-Formylcytosine yields DNA–protein cross-links in nucleosome core particles. *J. Am. Chem. Soc.* **139,** 10617–10620 (2017).

31. Ji, S., Shao, H., Han, Q., Seiler, C. L. & Tretyakova, N. Y. Reversible DNA–protein cross-linking at epigenetic DNA marks. *Angew. Chemie Int. Ed.* **56,** 14130–14134 (2017).

32. Kellinger, M. W. *et al.* 5-Formyl- and 5-carboxyl-cytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19,** 831–833 (2012).

33. Lusic, H. *et al.* Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res.* **36,** 6548–6557 (2008).

34. Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G. & Zhang, X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev* **28,** 2304–2313 (2014).

35. Wang, D. *et al.* MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **45,** 2396–2407 (2017).

36. Basanta-Sanchez, M. *et al.* TET1-mediated oxidation of 5-formylcytosine (5fC) to 5-carboxycytosine (5caC) in RNA. *ChemBioChem* **18,** 72–76 (2017).

37. Fu, L. *et al.* Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136,** 11582–11585 (2014).

38. M., H. S. *et al.* Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem* **16,** 752–755 (2015).

39. Zhang, H.-Y., Xiong, J., Qi, B.-L., Feng, Y.-Q. & Yuan, B.-F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. *Chem. Commun.* **52,** 737–740 (2016).

40. Delatte, B. *et al.* Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351,** 282–285 (2016).

41. Djuric, Z. *et al.* Levels of 5-hydroxymethyl-2′-deoxyuridine in DNA from blood of women scheduled for breast biopsy. *Cancer Epidemiol. Biomarkers Prev.* **10,** 147–149 (2001).

42. Klungland, A. *et al.* 5-Formyluracil and its nucleoside derivatives confer toxicity and mutagenicity to mammalian cells by interfering with normal RNA and DNA metabolism. *Toxicol. Lett.* **119,** 71–78 (2001).

43. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10,** 574–581 (2014).

44. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5,** e15367 (2010).

45. Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chemie Int. Ed.* **54,** 12511–12514 (2015).

46. Gama-Sosa, M. A. *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11,** 6883–6894 (1983).

47. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chemie Int. Ed.* **50,** 7008–7012 (2011).

48. Friso, S., Choi, S.-W., Dolnikowski, G. G. & Selhub, J. A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Anal. Chem.* **74,** 4526–4531 (2002).

49. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chemie Int. Ed.* **49,** 5375–5377 (2010).

50. Tang, Y., Zheng, S.-J., Qi, C.-B., Feng, Y.-Q. & Yuan, B.-F. Sensitive and simultaneous determination of 5-methylcytosine and its oxidation products in genomic DNA by chemical derivatization coupled with liquid chromatography-tandem mass spectrometry analysis. *Anal. Chem.* **87,** 3445–3452 (2015).

51. Zeng, H. *et al.* Formation and determination of endogenous methylated nucleotides in mammals by chemical labeling coupled with mass spectrometry analysis. *Anal. Chem.* **89,** 4153–4160 (2017).

52. Xiong, J. *et al.* Heavy metals induce decline of derivatives of 5-methycytosine in both DNA and RNA of stem cells. *ACS Chem. Biol.* **12,** 1636–1643 (2017).

53. Yuan, B.-F. & Feng, Y.-Q. Recent advances in the analysis of 5-methylcytosine and its oxidation products. *TrAC Trends Anal. Chem.* **54,** 24–35 (2014).

54. Tang, Y. *et al.* Determination of oxidation products of 5-methylcytosine in plants by chemical derivatization coupled with liquid chromatography/tandem mass spectrometry analysis. *Anal. Chem.* **86,** 7764–7772 (2014).

55. Jiang, H.-P. *et al.* Determination of formylated DNA and RNA by chemical labeling combined with mass spectrometry analysis. *Anal. Chim. Acta* **981,** 1–10 (2017).

56. Tian, T., Shaoru, W., Jianguo, W. U. & Xiang, Z. Review: Advances in methodology of DNA methylation assay. *Sci. China Chem.* **54,** 1233–1243 (2011).

57. F., F. M., Roberto, R. & Jesus, C. M. Rapid quantification of DNA methylation by high performance capillary electrophoresis. *Electrophoresis* **21,** 2990–2994 (2000).

58. Booth, M. J., Raiber, E. & Balasubramanian, S. Chemical methods for decoding cytosine modifications in DNA. *Chem. Rev.* 2240−2254 (2015).

59. Wang, T. *et al.* Application of N-halogeno-N-sodiobenzenesulfonamide reagents to the selective detection of 5-methylcytosine in DNA sequences. *J. Am. Chem. Soc.* **135,** 1240–1243 (2013).

60. Wang, Y. *et al.* Highly selective detection of 5-methylcytosine in genomic DNA based on asymmetric PCR and specific DNA damaging reagents. *Anal. Chem.* **88,** 3348–3353 (2016).

61. Matsushita, T., Moriyama, Y., Nagae, G., Aburatani, H. & Okamoto, A. DNA-friendly Cu(II)/TEMPO-catalyzed 5-hydroxymethylcytosine-specific oxidation. *Chem. Commun.* **53,** 5756–5759 (2017).

62. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1827–1831 (1992).

63. Li, Y. & Tollefsbol, T. O. DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol. Biol.* **791,** 11–21 (2011).

64. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **5,** e8888 (2010).

65. Booth, M. J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **6,** 435–440 (2014).

66. Fukuzawa, S., Takahashi, S., Tachibana, K., Tajima, S. & Suetake, I. Simple and accurate single base resolution analysis of 5-hydroxymethylcytosine by catalytic oxidative bisulfite sequencing using micelle incarcerated oxidants. *Bioorganic Med. Chem.* **24,** 4254–4262 (2016).

67. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **7,** 2159–2170 (2012).

68. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149,** 1368–1380 (2012).

69. Lu, X. *et al.* Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* **135,** 9315–9317 (2013).

70. Lu, X. *et al.* Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Res.* **25,** 386–389 (2015).

71. Neri, F. *et al.* Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.* **10,** 674–683 (2015).

72. Wu, H., Wu, X., Shen, L. & Zhang, Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32,** 1231–1240 (2014).

73. Grunau, C., Clark, S. J. & Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **29,** e65 (2001).

74. Gravina, S., Dong, X., Yu, B. & Vijg, J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* **17,** 150 (2016).

75. Farlik, M. *et al.* Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. *Cell Rep.* **10,** 1386–1397 (2015).

76. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11,** 817–820 (2014).

77. Clark, S. J. *et al.* Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* **12,** 534–547 (2017).

78. Genereux, D. P., Johnson, W. C., Burden, A. F., Stöger, R. & Laird, C. D. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res.* **36,** e150 (2008).

79. Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L. & Frommer, M. DNA methylation: Bisulphite modification and analysis. *Nat. Protoc.* **1,** 2353–2364 (2006).

80. Tusnády, G. E., Simon, I., Váradi, A. & Arányi, T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res.* **33,** e9 (2005).

81. Xiong, Z. & Laird, P. W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25,** 2532–2534 (1997).

82. Herman, J. G., Graff, J. R., Myöhänen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 9821–9826 (1996).

83. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* **16,** 383–393 (2006).

84. Olek, A., Oswald, J. & Walter, J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24,** 5064–5066 (1996).

85. Eads, C. A. *et al.* MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* **28,** e32 (2000).

86. Gonzalgo, M. L. & Jones, P. A. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* **25,** 2529–2531 (1997).

87. Okamoto, A., Tainaka, K. & Kamei, T. Sequence-selective osmium oxidation of DNA: efficient distinction between 5-methylcytosine and cytosine. *Org. Biomol. Chem.* **4,** 1638–1640 (2006).

88. Tanaka, K., Tainaka, K., Kamei, T. & Okamoto, A. Direct labeling of 5-methylcytosine and its applications. *J. Am. Chem. Soc.* **129,** 5612–5620 (2007).

89. Bareyt, S. & Carell, T. Selective detection of 5-methylcytosine sites in DNA. *Angew. Chemie Int. Ed.* **47,** 181–184 (2008).

90. Zhang, L. *et al.* Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. *Nat. Commun.* **4,** 1517 (2013).

91. Kato, D. *et al.* A nanocarbon film electrode as a platform for exploring DNA methylation. *J. Am. Chem. Soc.* **130,** 3716–3717 (2008).

92. Wang, P., Mai, Z., Dai, Z. & Zou, X. Investigation of DNA methylation by direct electrocatalytic oxidation. *Chem. Commun.* **46,** 7781–7783 (2010).

93. Wang, P., Han, P., Dong, L. & Miao, X. Direct potential resolution and simultaneous detection of cytosine and 5-methylcytosine based on the construction of polypyrrole functionalized graphene nanowall interface. *Electrochem. Commun.* **61,** 36–39 (2015).

94. Wang, P., Chen, H., Tian, J., Dai, Z. & Zou, X. Electrochemical evaluation of DNA methylation level based on the stoichiometric relationship between purine and pyrimidine bases. *Biosens. Bioelectron.* **45,** 34–39 (2013).

95. Yamada, H., Tanabe, K. & Nishimoto, S. Photocurrent response after enzymatic treatment of DNA duplexes immobilized on gold electrodes: electrochemical discrimination of 5-methylcytosine modification in DNA. *Org. Biomol. Chem.* **6,** 272–277 (2008).

96. Wang, P., Wu, H., Dai, Z. & Zou, X. Picomolar level profiling of the methylation status of the p53 tumor suppressor gene by a label-free electrochemical biosensor. *Chem. Commun.* **48,** 10754–10756 (2012).

97. Wu, M., Wang, X., Wang, K. & Guo, Z. Sequence-specific detection of cytosine methylation in DNA via the FRET mechanism between upconversion nanoparticles and gold nanorods. *Chem. Commun.* **52,** 8377–8380 (2016).

98. Wang, Z.-Y., Wang, L.-J., Zhang, Q., Tang, B. & Zhang, C.-Y. Single quantum dot-based nanosensor for sensitive detection of 5-methylcytosine at both CpG and non-CpG sites. *Chem. Sci.* **9,** 1330–1338 (2018).

99. Feng, F., Wang, H., Han, L. & Wang, S. Fluorescent conjugated polyelectrolyte as an indicator for convenient detection of DNA methylation. *J. Am. Chem. Soc.* **130,** 11338–11343 (2008).

100. Münzel, M., Lercher, L., Müller, M. & Carell, T. Chemical discrimination between dC and 5MedC via their hydroxylamine adducts. *Nucleic Acids Res.* **38,** e192 (2010).

101. Szwagierczak, A. *et al.* Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5–hydroxymethylcytosine. *Nucleic Acids Res.* **39,** 5149–5156 (2011).

102. Wang, H. *et al.* Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.* **39,** 9294–9305 (2011).

103. Song, C.-X., Yu, M., Dai, Q. & He, C. Detection of 5-hydroxymethylcytosine in a combined glycosylation restriction analysis (CGRA) using restriction enzyme TaqαI. *Bioorg. Med. Chem. Lett.* **21,** 5075–5077 (2011).

104. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38,** e181 (2010).

105. Song, C.-X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotech* **29,** 68–72 (2011).

106. Song, C.-X. *et al.* 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* **27,** 1231–1242 (2017).

107. Han, D. *et al.* A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol. Cell* **63,** 711–719 (2016).

108. Wu, Q., Amrutkar, S. M. & Shao, F. Sulfinate based selective labeling of 5-hydroxymethylcytosine: application to biotin pull down assay. *Bioconjug. Chem.* **29,** 245–249 (2018).

109. Chen, H. Y. *et al.* Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA using boric acid-functionalized nano-microsphere fluorescent probes. *Biosens. Bioelectron.* **91,** 328–333 (2017).

110. Zhao, C. *et al.* Boronic acid-mediated polymerase chain reaction for gene- and fragment-specific detection of 5-hydroxymethylcytosine. *Nucleic Acids Res.* **42,** e81 (2014).

111. Zhou, Y. *et al.* Electrochemical biosensor for detection of DNA hydroxymethylation based on glycosylation and alkaline phosphatase catalytic signal amplification. *Electrochim. Acta* **174,** 647–652 (2015).

112. Yang, Z., Shi, Y., Liao, W., Yin, H. & Ai, S. A novel signal-on photoelectrochemical biosensor for detection of 5-hydroxymethylcytosine based on in situ electron donor producing strategy and all wavelengths of light irradiation. *Sensors Actuators B Chem.* **223,** 621–625 (2016).

113. Jiang, W., Lu, Y., Wang, H., Wang, M. & Yin, H. Amperometric biosensor for 5-hydroxymethylcytosine based on enzymatic catalysis and using spherical poly(acrylic acid) brushes. *Microchim. Acta* **184,** 3789–3796 (2017).

114. Zhang, Y., Li, Y., Wei, Y., Sun, H. & Wang, H. A sensitive signal-off electrogenerated chemiluminescence biosensing method for the discrimination of DNA hydroxymethylation based on glycosylation modification and signal quenching from ferroceneboronic acid. *Talanta* **170,** 546–551 (2017).

115. Wei, Y. *et al.* Electrogenerated chemiluminescence biosensing method for highly sensitive detection of DNA hydroxymethylation: combining glycosylation with $Ru(phen)_3^{2+}$-assembled graphene oxide. *J. Electroanal. Chem.* **795,** 123–129 (2017).

116. Jiang, W., Wu, L., Duan, J., Yin, H. & Ai, S. Ultrasensitive electrochemiluminescence immunosensor for 5-hydroxymethylcytosine detection based on Fe_3O_4@SiO_2 nanoparticles and PAMAM dendrimers. *Biosens. Bioelectron.* **99,** 660–666 (2018).

117. Yang, Y. *et al.* Electrochemical signal-amplified detection of 5-methylcytosine and 5-hydroxymethylcytosine in DNA using glucose modification coupled with restriction endonucleases. *Analyst* **143,** 2051–2056 (2018).

118. Hayashi, G. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine by one-pot bisulfite-free chemical onversion with peroxotungstate. *J. Am. Chem. Soc.* **138,** 14178–14181 (2016).

119. Hong, T. *et al.* Fluorescent strategy based on cationic conjugated polymer fluorescence resonance energy transfer for the quantification of 5-(hydroxymethyl)cytosine in genomic DNA. *Anal. Chem.* **85,** 10797–10802 (2013).

120. Song, C.-X., Diao, J., Brunger, A. T. & Quake, S. R. Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 4338–4343 (2016).

121. Raiber, E.-A. *et al.* Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* **13,** R69 (2012).

122. Xia, B. *et al.* Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods* **12,** 1047–1050 (2015).

123. Zhu, C. *et al.* Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20,** 720–731.e5 (2017).

124. Wang, S.-R. *et al.* Cucurbit[7]uril-driven host–guest chemistry for reversible intervention of 5-formylcytosine-targeted biochemical reactions. *J. Am. Chem. Soc.* **139,** 16903–16912 (2017).

125. Samanta, B., Seikowski, J. & Höbartner, C. Fluorogenic labeling of 5-formylpyrimidine nucleotides in DNA and RNA. *Angew. Chemie Int. Ed.* **55,** 1912–1916 (2016).

126. Liu, C. *et al.* Fluorogenic labeling and single-base resolution analysis of 5-formylcytosine in DNA. *Chem. Sci.* **8,** 7443–7447 (2017).

127. Wang, Y. *et al.* Gene specific-loci quantitative and single-base resolution analysis of 5-formylcytosine by compound-mediated polymerase chain reaction. *Chem. Sci.* **9,** 3723–3728 (2018).

128. Xu, L. *et al.* Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation. *Angew. Chemie Int. Ed.* **53,** 11223–11227 (2014).

129. Kaput, J. & Sneider, T. W. Methylation of somatic vs germ cell DNAs analyzed by restriction

endonuclease digestions. *Nucleic Acids Res.* **7,** 2303–2322 (1979).

130. Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N. & Van Oudenaarden, A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* **34,** 852–856 (2016).

131. Sun, Z. *et al.* A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol. Cell* **57,** 750–761 (2015).

132. Sérandour, A. A. *et al.* Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets. *Genome Biol.* **17,** 56–67 (2016).

133. Gao, F. *et al.* Highly efficient electrochemical sensing platform for sensitive detection DNA methylation, and methyltransferase activity based on Ag NPs decorated carbon nanocubes. *Biosens. Bioelectron.* **99,** 201–208 (2018).

134. Yang, Z. *et al.* A novel electrochemical immunosensor for the quantitative detection of 5-hydroxymethylcytosine in genomic DNA of breast cancer tissue. *Chem. Commun.* **51,** 14671–14673 (2015).

135. Stains, C. I., Furman, J. L., Segal, D. J. & Ghosh, I. Site-specific detection of DNA methylation utilizing mCpG-SEER. *J. Am. Chem. Soc.* **128,** 9761–9765 (2006).

136. Hiraoka, D. *et al.* Development of a method to measure DNA methylation levels by using methyl CpG-binding protein and luciferase-fused zinc finger protein. *Anal. Chem.* **84,** 8259–8264 (2012).

137. Serre, D., Lee, B. H. & Ting, A. H. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* **38,** 391–399 (2010).

138. Robertson, A. B. *et al.* A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **39,** e55 (2011).

139. Nomura, A. & Okamoto, A. Phosphopeptides designed for 5-methylcytosine recognition. *Biochemistry* **50,** 3376–3385 (2011).

140. Kubik, G. & Summerer, D. TALEored epigenetics: a DNA-binding scaffold for programmable epigenome editing and analysis. *ChemBioChem* **17,** 975–980 (2016).

141. Zhang, Y. *et al.* Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition. *Nat. Commun.* **8,** 901 (2017).

142. Rathi, P., Witte, A. & Summerer, D. Engineering DNA backbone interactions results in TALE scaffolds with enhanced 5-methylcytosine selectivity. *Sci. Rep.* **7,** 15067 (2017).

143. Kubik, G., Batke, S. & Summerer, D. Programmable sensors of 5-hydroxymethylcytosine. *J. Am. Chem. Soc.* **137,** 2–5 (2015).

144. Grzegorz, K., Schmidt, M. J., Penner, J. E. & Daniel, S. Programmable and highly resolved in vitro detection of 5-methylcytosine by TALEs. *Angew. Chemie Int. Ed.* **53,** 6002–6006 (2014).

145. Maurer, S., Giess, M., Koch, O. & Summerer, D. Interrogating key positions of size-reduced TALE repeats reveals a programmable sensor of 5-carboxylcytosine. *ACS Chem. Biol.* **11,** 3294–3299 (2016).

146. Dalhoff, C., Lukinavičius, G., Klimašauskas, S. & Weinhold, E. Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.* **2,** 31–32 (2006).

147. Dalhoff, C., Lukinavičius, G., Klimašauskas, S. & Weinhold, E. Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protoc.* **1,** 1879–1886 (2006).

148. Lukinavičius, G. *et al.* Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *J. Am. Chem. Soc.* **129,** 2758–2759 (2007).

149. Kriukiene, E. *et al.* DNA unmethylome profiling by covalent capture of CpG sites. *Nat. Commun.* **4,** 2190 (2013).

150. Staševskij, Z., Gibas, P., Gordevičius, J., Kriukienė, E. & Klimašauskas, S. Tethered oligonucleotide-primed sequencing, TOP-Seq: a high-resolution economical approach for DNA

epigenome profiling. *Mol. Cell* **65,** 554–564.e6 (2017).

151. Kawasaki, Y. *et al.* A novel method for the simultaneous identification of methylcytosine and hydroxymethylcytosine at a single base resolution. *Nucleic Acids Res.* **45,** e24 (2016).

152. Takahashi, S., Suetake, I., Engelhardt, J. & Tajima, S. A novel method to analyze 5-hydroxymethylcytosine in CpG sequences using maintenance DNA methyltransferase, DNMT1. *FEBS Open Bio* **5,** 741–747 (2015).

153. Larkin, J., Henley, R. Y., Jadhav, V., Korlach, J. & Wanunu, M. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nat. Nanotechnol.* **12,** 1169–1175 (2017).

154. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7,** 461–465 (2010).

155. Clark, T. A. *et al.* Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* **11,** 4–13 (2013).

156. Pais, J. E. *et al.* Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 4316–4321 (2015).

157. Yang, Y. *et al.* Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genomics* **16,** 350–359 (2015).

158. Song, C.-X. *et al.* Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat. Methods* **9,** 75–77 (2011).

159. Chavez, L. *et al.* Simultaneous sequencing of oxidized methylcytosines produced by TET/JBP dioxygenases in Coprinopsis cinerea. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E5149–E5158 (2014).

160. Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **38,** e159 (2010).

161. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18904–18909 (2013).

162. Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc.* **136,** 16582–16587 (2014).

163. Johnson, R. P., Fleming, A. M., Perera, R. T., Burrows, C. J. & White, H. S. Dynamics of a DNA mismatch site held in confinement discriminate epigenetic modifications of cytosine. *J. Am. Chem. Soc.* **139,** 2750–2756 (2017).

164. Zeng, T. *et al.* Detection of 5-methylcytosine and 5-hydroxymethylcytosine in DNA via host-guest interactions inside α-hemolysin nanopores. *Chem. Sci.* **6,** 5628–5634 (2015).

165. Li, W., Gong, L. & Bayley, H. Single-molecule detection of 5-hydroxymethylcytosine in DNA through chemical modification and nanopore analysis. *Angew. Chemie Int. Ed.* **52,** 4350–4355 (2013).

166. Yu, J., Cao, C. & Long, Y.-T. Selective and sensitive detection of methylcytosine by aerolysin nanopore under serum condition. *Anal. Chem.* **89,** 11685–11689 (2017).

167. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14,** 407–410 (2017).

168. Rand, A. C. *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14,** 411–413 (2017).

169. Liu, L., Yang, C., Zhao, K., Li, J. & Wu, H. C. Ultrashort single-walled carbon nanotubes in a lipid bilayer as a new nanopore sensor. *Nat. Commun.* **4,** 2989 (2013).

170. Wanunu, M. *et al.* Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.* **133,** 486–492 (2011).

171. Shim, J. *et al.* Detection and quantification of methylation in DNA using solid-state nanopores. *Sci. Rep.* **3,** 1389 (2013).

172. Beaucage, S. L. & Iyer, R. P. The functionalization of oligonucleotides via phosphoramidite

derivatives. *Tetrahedron* **49,** 1925–1963 (1993).

173. Jl, Y., Bannwarth, W. & Luu, B. Application of the phosphoramidite-phosphite triester approach for the synthesis of combinations between oxygenated sterols and nucleoside analogues linked by phosphodiester bonds. *Tetrahedron* **46,** 487–502 (1990).

174. Pon, R. T. A long chain biotin phosphoramidite reagent for the automated synthesis of 5′-biotinylated oligonucleotides. *Tetrahedron Lett.* **32,** 1715–1718 (1991).

175. Stetsenko, D. A. & Gait, M. J. New phosphoramidite reagents for the synthesis of oligonucleotides containing a cysteine residue useful in peptide conjugation. *Nucleosides. Nucleotides Nucleic Acids* **19,** 1751–1764 (2000).

176. Ogino, M., Taya, Y. & Fujimoto, K. Detection of methylcytosine by DNA photoligation via hydrophobic interaction of the alkyl group. *Org. Biomol. Chem.* **7,** 3163–3167 (2009).

177. Ogino, M., Taya, Y. & Fujimoto, K. Highly selective detection of 5-methylcytosine using photochemical ligation. *Chem. Commun.* **0,** 5996–5998 (2008).

178. Sakamoto, T., Ami, T. & Fujimoto, K. 5-Methylcytosine selective photoligation using photoresponsive oligonucleotides containing various 5-vinyl-2′-deoxyuridines having an aromatic group. *Chem. Lett.* **41,** 47–49 (2011).

179. Yamayoshi, A., Matsuyama, Y., Kushida, M., Kobori, A. & Murakami, A. Novel photodynamic effect of a psoralen-conjugated oligonucleotide for the discrimination of the methylation of cytosine in DNA. *Photochem. Photobiol.* **90,** 716–722 (2014).

180. Fujimo, K., Konishi-Hiratsuka, K. & Sakamoto, T. Quick, selective and reversible photocrosslinking reaction between 5-methylcytosine and 3-cyanovinylcarbazole in DNA double strand. *Int. J. Mol. Sci.* **14,** 5765–5774 (2013).

181. Tanabe, K., Yamada, H. & Nishimoto, S. I. One-electron photooxidation and site-selective strand cleavage at 5-methylcytosine in DNA by sensitization with 2-methyl-1,4-naphthoquinone- tethered oligonucleotides. *J. Am. Chem. Soc.* **129,** 8034–8040 (2007).

182. Yamada, H., Tanabe, K., Ito, T. & Nishimoto, S. The pH effect on the naphthoquinone-photosensitized oxidation of 5-methylcytosine. *Chem. Eur. J.* **14,** 10453–10461 (2008).

183. Duprey, J.-L. H. A. *et al.* Detection of DNA base variation and cytosine methylation at a single nucleotide site using a highly sensitive fluorescent probe. *Chem. Commun.* **47,** 6629–6631 (2011).

184. Duprey, J.-L. H. A. *et al.* Single site discrimination of cytosine, 5-methylcytosine, and 5-hydroxymethylcytosine in target DNA using anthracene-tagged fluorescent probes. *ACS Chem. Biol.* **11,** 717–721 (2016).

185. Taskova, M., Barducci, M. C. & Astakhova, K. Environmentally sensitive molecular probes reveal mutations and epigenetic 5-methyl cytosine in human oncogenes. *Org. Biomol. Chem.* **15,** 5680–5684 (2017).

186. Okamoto, A., Tanabe, K. & Saito, I. Site-specific discrimination of cytosine and 5-methylcytosine in duplex DNA by peptide nucleic acids. *J. Am. Chem. Soc.* **124,** 10262–10263 (2002).

187. Xu, C. *et al.* Detecting 5-methylcytosine using an enzyme-free DNA strand exchange reaction without pretreatment under physiological conditions. *Chem. Commun.* **52,** 6833–6836 (2016).

188. Liu, S., Wu, P., Li, W., Zhang, H. & Cai, C. An electrochemical approach for detection of DNA methylation and assay of the methyltransferase activity. *Chem. Commun.* **47,** 2844–2846 (2011).

189. Wang, J., Zhu, Z. & Ma, H. Label-free real-time detection of DNA methylation based on quartz crystal microbalance measurement. *Anal. Chem.* **85,** 2096–2101 (2013).

190. von Watzdorf, J., Leitner, K. & Marx, A. Modified nucleotides for discrimination between cytosine and the epigenetic marker 5-methylcytosine. *Angew. Chemie Int. Ed.* **55,** 3229–3232 (2016).

191. Johannsen, M. W., Gerrard, S. R., Melvin, T. & Brown, T. Triplex-mediated analysis of cytosine methylation at CpA sites in DNA. *Chem. Commun.* **50,** 551–553 (2014).

192. Tainaka, K., Tanaka, K. & Okamoto, A. Development of bipyridine-modified nucleobase for methylcytosine-selective crosslink reaction. *Nucleic Acids Symp. Ser.* **50,** 129–130 (2006).

193. Tanaka, K., Tainaka, K., Umemoto, T., Nomura, A. & Okamoto, A. An osmium-DNA interstrand complex: application to facile DNA methylation analysis. *J. Am. Chem. Soc.* **129,** 14511–14517 (2007).

194. Sugizaki, K., Ikeda, S., Yanagisawa, H. & Okamoto, A. Facile synthesis of hydroxymethylcytosine-containing oligonucleotides and their reactivity upon osmium oxidation. *Org. Biomol. Chem.* **9,** 4176–4181 (2011).

195. Sugizaki, K., Umemoto, T. & Okamoto, A. On-chip DNA methylation analysis using osmium complexation. *J. Nucleic Acids* **2011,** 480570 (2011).

196. Tanaka, K., Tainaka, K. & Okamoto, A. Fluorescence quenching by methylcytosine-metal complexation. *Nucleic Acids Symp. Ser.* **2006,** 139–140 (2006).

197. Sugizaki, K., Nakamura, A., Yanagisawa, H. & Okamoto, A. Ligand-incorporation site in 5-methylcytosine-detection probe modulating the site of osmium complexation with the target DNA. *Chem. Biodivers.* **9,** 2000–2007 (2012).

198. Li, Y. *et al.* Sequence-specific microscopic visualization of DNA methylation status at satellite repeats in individual cell nuclei and chromosomes. *Nucleic Acids Res.* **41,** e186 (2013).

199. Hardisty, R. E., Kawasaki, F., Sahakyan, A. B. & Balasubramanian, S. Selective chemical labeling of natural T modifications in DNA. *J. Am. Chem. Soc.* **137,** 9270–9272 (2015).

200. Liu, C. *et al.* Enrichment and fluorogenic labelling of 5-formyluracil in DNA. *Chem. Sci.* **8,** 4505–4510 (2017).

201. Hirose, W., Sato, K. & Matsuda, A. Selective detection of 5-formyl-2′-deoxyuridine, an oxidative lesion of thymidine, in DNA by a fluorogenic reagent. *Angew. Chemie Int. Ed.* **49,** 8392–8394 (2010).

202. Hirose, W., Sato, K. & Matsuda, A. Fluorescence properties of 5-(5,6-dimethoxybenzothiazol-2-yl)-2′- deoxyuridine (dbtU) and oligodeoxyribonucleotides containing dbtU. *European J. Org. Chem.* **2011,** 6206–6217 (2011).

203. Wang, Y. *et al.* Naphthalimide derivatives as multifunctional molecules for detecting 5-formylpyrimidine by both PAGE analysis and dot-blot assays. *Chem. Commun.* **54,** 1497–1500 (2018).

204. Guo, P. *et al.* Synthesis and spectroscopic properties of fluorescent 5-benzimidazolyl-2′-deoxyuridines 5-fdU probes obtained from o-phenylenediamine derivatives. *Org. Biomol. Chem.* **11,** 1610–1613 (2013).

205. Liu, C. *et al.* A highly efficient fluorescence-based switch-on detection method of 5-formyluracil in DNA. *Nano Res.* **10,** 2449–2458 (2017).

206. Yu, M., Song, C. X. & He, C. Detection of mismatched 5-hydroxymethyluracil in DNA by selective chemical labeling. *Methods* **72,** 16–20 (2015).

207. Kawasaki, F. *et al.* Sequencing 5-hydroxymethyluracil at single-base resolution. *Angew. Chemie Int. Ed.* (2018). doi:10.1002/anie.201804046

208. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8,** 389–399 (2015).

Figure Captions

Figure 1 - **Active demethylation of cytosine.** Cytosine (C) residues in DNA can be methylated by DNA-methyltransferases (DNMTs). The resulting 5-methylcytosine (mC) residues are susceptible to oxidation by the ten-eleven translocation dioxygenase (TET) family of enzymes to produce

5-hydroxymethylcytosine (hmC), and subsequently 5-formylcytosine (fC) and 5-carboxylcytosine (caC). The epigenetic role of mC is well established. There is strong evidence that hmC functions as an epigenetic marker as well, and fC and caC also appear to influence the binding of certain transcription factors and chromatin regulators, and the activity of RNA polymerase II. The oxidised derivatives of mC are also intermediates in an active demethylation pathway, as fC and caC can be excised by thymine-DNA-glycosylase (TDG) to produce an abasic site, where unmodified C residues can be restored. A recent study also found evidence that unmodified C also results from deformylation of fC, and potentially also decarboxylation of caC, as shown by the dotted arrows.

Figure 2 - **Oxidation of thymine.** Oxidation of thymine (T) residues in DNA produces 5-hydroxymethyluracil (hmU), which can be further oxidised to 5-formyluracil (fU). These oxidised thymine derivatives result from the action of reactive oxygen species (ROS). More recently it has also been shown that hmU is produced by ten-eleven translocation dioxygenase (TET) enzymes, and there is evidence that hmU may play an epigenetic role. The structural similarity between hmU and 5-hydroxymethylcytosine (hmC), and between fU and 5-formylcytosine (fC), presents a challenge in the development of detection methods capable of discriminating between these modifications.

Figure 3 - ***N*-halogeno-*N*-sodiobenzenesulfonamide reagents**. Maxam-Gilbert sequencing is a method of DNA sequencing which employs chemically induced strand cleavage at particular nucleobase sites followed by gel electrophoretic analysis. This strategy has been extended to the detection of 5-methylcytosine (mC), an epigenetic modification of cytosine (C) found in DNA. Treatment of a DNA sample with *N*-sodio-*N*-chloro-*p*-nitrobenzenesulfonamide (**1**) and I_2 followed by piperidine causes strand cleavage at C loci. Treatment with *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide (**2**) followed by piperidine causes strand cleavage at both C and mC loci. After gel electrophoresis, comparison of the results obtained after treatment with **1** with those obtained after treatment with **2** allows mC loci to be identified.

Figure 4- **Selective chemical labelling methods for detection of 5-formylcytosine (fC) in DNA**. Labelling of fC with a pyrene fluorophore (**9**) through formation of a hydrazone linkage allows for determination of fC levels. An excimer is formed when two labelled fC residues are adjacent to each other, allowing the relative position of two fC residues to be deduced. Reaction of fC with a trimethylindole derivative (**6**), the reagent azi-BP (**8**), or 2-(adamantyl)ethoxyamine (**5**) followed by complexation by a CB7 macrocycle, results in adducts which act as a roadblock to DNA-polymerases, allowing for sequence specific detection of fC in primer extension assays. Reaction with azi-BP (**8**), CBAN (**7**), malonitrile (**4**), or a 1,3-indandione derivative (**3**) alters the base pairing properties of fC, since the hydrogen bond donating exocyclic amino group is converted to a hydrogen bond acceptor. This allows for the determination of fC loci through DNA sequencing. The inclusion of an azide moiety in **3** and **8** enables the introduction of an alkyne-bearing biotin tag for the enrichment of DNA fragments containing fC.

Figure 5 - **Oligonucleotide probes for detection of epigenetic markers.** (Adapted from reference 176 with permission from the Royal Society of Chemistry.) Epigenetic modifications can be detected in DNA using oligonucleotide probes, which have the advantage of easily tuneable sequence specificity. A representative example of such a method, illustrated above, employs an oligonucleotide containing a 5-vinyl-2'-deoxyuridine residue bearing a hydrophobic moiety, which undergoes a favourable hydrophobic interaction with the methyl group of mC, as shown by the arrow. This facilitates a [2+2]

cycloaddition upon irradiation to form an interstrand crosslink. Since the interstrand crosslink is not formed when unmodified cytosine is present in place of mC this allows mC residues to be detected using probe oligonucleotides immobilised on a microarray.

Figure 6 - **Chemical labelling methods for detection of 5-formyluracil (fU).** Detection of fU in DNA can be achieved through labelling with *o*-phenylenediamine derivatives such as **10**, which react with fU to form benzimidazole moieties. Importantly, this reaction is selective for fU over the structurally similar modification 5-formylcytosine (fC), which is present at similar levels in DNA. The *o*-phenylenediamine group in **10** acts as a quencher of the naphthalimide fluorophore. Fluorescence is enhanced when the benzimidazole linkage is formed. 5-Formyluracil residues labelled with **10** block the action of DNA-polymerases, enabling sequence specific detection of fU in a primer extension assay. Similarly, reaction of fU with 2-amino-4,5-dimethoxythiophenol (**11**) leads to formation of a fluorescent benzothiazol-2-yl group, allowing fU levels in a DNA sample to be determined through measurement of the fluorescence intensity. 5-Formyluracil can also be selectively labelled by naphthalimide derivative **12**, or 4-hydrazinyl-7-nitrobenz-[2,1,3-d]-oxadiazole (NBDH) (**13**) through formation of hydrazone linkages. The resulting adducts are fluorescent, allowing fU levels in a DNA sample to be quantified. 5-Formyluracil residues labelled with either **12** or **13** also act as roadblocks to DNA-polymerases. The azide moiety in **12** enables enrichment of DNA fragments containing fU *via* reaction with an alkyne-bearing biotin derivative.

Table 1 - **Bisulfite sequencing and its modifications.** Bisulfite sequencing (BS-seq) allows for the detection of 5-methylcytosine (mC) in DNA. After treatment with sodium bisulfite cytosine (C) residues are converted to uracil (U), while mC is unaffected, and can therefore be distinguished in DNA sequencing. 5-Hydroxymethylcytosine (hmC) reads as C in BS-seq, while 5-formylcytosine (fC) and 5-carboxylcytosine (caC) read as U. Oxidative bisulfite sequencing (oxBS-seq) and reductive bisulfite sequencing (redBS-seq) are modified procedures in which hmC is first converted to fC or *vice versa*. TET-assisted bisulfite sequencing (TAB-seq) involves protection of hmC, followed by enzymatic oxidation of all other modified cytosine derivatives to caC prior to bisulfite treatment. In chemically assisted bisulfite sequencing (fCAB-seq) fC is protected from bisulfite-induced deamination with ethylhydroxylamine. Formation of an amide is used to protect caC prior to bisulfite treatment in a modification of chemically assisted bisulfite sequencing (caCAB-seq). In methylase assisted bisulfite sequencing (MAB-seq) all unmodified C residues are enzymatically converted to mC prior to bisulfite treatment. A variant of this method known as caMAB-seq involves enzymatic methylation followed by reduction of fC to hmC prior to bisulfite treatment, allowing caC to be selectively detected.

Table 2 - **Summary of detection methods for cytosine and thymine modifications.** A large number of detection methods for cytosine and thymine modifications have been developed. Many of these detection methods have been utilised in the study of genomic DNA, while the applicability of others has been demonstrated only in synthetic oligonucleotide models. While some detection methods can detect modified nucleobases with single base resolution, methods for quantifying the levels of modifications in whole genomes or genome fragments have also proven valuable. These aspects of the detection methods discussed in this review are summarised in Table 2. Detection methods which offer single base resolution are highlighted in bold.