Advances in the profiling of DNA modifications: cytosine methylation and beyond

Nongluk Plongthongkum, Dinh H. Diep* and Kun Zhang*

Abstract | Chemical modifications of DNA have been recognized as key epigenetic mechanisms for maintenance of the cellular state and memory. Such DNA modifications include canonical 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC). Recent advances in detection and quantification of DNA modifications have enabled epigenetic variation to be connected to phenotypic consequences on an unprecedented scale. These methods may use chemical or enzymatic DNA treatment, may be targeted or non-targeted and may utilize array-based hybridization or sequencing. Key considerations in the choice of assay are cost, minimum sample input requirements, accuracy and throughput. This Review discusses the principles behind recently developed techniques, compares their respective strengths and limitations and provides general guidelines for selecting appropriate methods for specific experimental contexts.

Partially methylated domains
(PMDs). Large contiguous regions of the genome (mean length ~153 kb) that display an intermediate level of DNA methylation (average < 70%).

Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92037, USA.
**These authors contributed equally to this work.*
Correspondence to K.Z.
e-mail: kzhang@bioeng.ucsd.edu
doi:10.1038/nrg3772
Published online 27 August 2014

DNA methylation refers to the addition of methyl groups to the adenine or cytosine bases of DNA. These methyl groups can be added or removed and can remain stable throughout multiple cell divisions. Whole-genome maps of 5-methylcytosine (5mC) — the most common mark of DNA methylation — have revealed intriguing patterns in human and mouse DNA such as cell state-dependent occurrences of 5mC in contexts other than canonical CpGs and in partially methylated domains (PMDs), and conserved regions depleted of 5mCs across mouse and human species. The commonly known DNA methyltransferases (DNMTs), which are well-known for depositing methyl groups on cytosine to yield 5mC in CpG contexts, have been shown to deposit methyl groups at non-CpG sites[1]. Generation and maintenance of non-CpG methylation seems to be tightly regulated, as such modifications are enriched in specific cell types, such as pluripotent cells and neural progenitors, as well as in adolescent and adult cortex tissues[2–6]. By contrast, partially methylated domains (PMDs) have been predominantly found in non-pluripotent cells and non-cortex tissue types[2,4,7]. These PMDs have been associated with low transcription rates, lamina-associated domains and late-replicating domains. Next, different classes of methylation-depleted regions named unmethylated regions (UMRs), DNA methylation valleys (DMVs) and DNA methylation canyons (DMCs) have been defined[7–9]. These regions tend to be conserved across cell types and across mouse and human species. Both methylation valleys and canyons tend to be marked with H3K4me3 or H3K27me3 or both and each can lead to active, inactive or poised transcriptional states, respectively[7,9,10]. Strikingly, these regions cover most genes important for embryonic development[10].

In addition to DNMTs, a class of enzymes has been recently described in mammalian cells to produce epigenetic modifications such as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These newly identified 'writers' are members of the ten-eleven translocation (TET) protein family and can sequentially oxidize 5mC to form 5hmC, 5fC and 5caC. These cytosine modifications can be collectively referred to as 5mC oxidation derivatives. Some studies (reviewed in REFS 11,12) have suggested that 5mC oxidation derivatives may exist as demethylation intermediates and that their presence may be related to the development and maintenance of methylation-free regulatory regions in mammalian genomes[11,12]. The presence of 5mC at major satellites and other transposable elements has been reported to be necessary for genome stability[11] whereas depletion of 5mC in a small number of transposable elements

Ten-eleven translocation (TET). DNA-binding enzymes that have been found to be methylcytosine dioxygenases in mammals and include TET1, TET2 and TET3. They were named for the genetic variant found in the *TET1* gene sequence in acute myeloid and lymphocytic leukaemia.

Short interspersed nuclear elements (SINEs). A subtype of transposable elements reverse-transcribed from RNA molecules. They do not encode a functional reverse transcriptase protein and cover a substantial portion of primate genomes, including all Alu sequences.

Long terminal repeats (LTRs). Stretches of DNA sequences that are identical and repeat in hundreds to thousands of copies. LTRs were founds at the end of retrotransposons and act as insertion sites for viruses to insert their genetic material into the host genome.

is tissue specific and may lead to enhancer function[13]. 5hmC has been detected at short interspersed nuclear elements (SINEs) and long terminal repeats (LTRs)[14–17], whereas 5fC and 5caC have been identified at major satellite repeats[16]. These findings point to potential 5mC turnover at these regions. However, further investigation is required to determine the exact function of these DNA modifications. Proteins that preferentially bind to 5hmC, 5fC or 5caC have recently been identified[18,19] and are suggested to be the 'readers' that connect these rare DNA modifications to phenotypic consequences. The exciting discovery of these DNA modifications was only possible with the advancement of techniques for characterizing these modifications and with the development of computational approaches for interpreting increasingly large data sets.

Recent method development and optimization has been mainly focused on four aspects: improving accuracy, increasing throughput, lowering sample input and reducing costs. It has recently been shown that many human biopsy samples, such as blood, tumour or brain sections, contain a diverse population of cells. Variation in cellular composition may contribute to the differences between presumably identical tissue samples, complicating interpretation of the population-wide measurements. Thus, assessment of a limited amount of relatively homogeneous cells obtained using laser microdissection or using flow cytometry sorting is highly desirable. Epigenetic profiling of large patient cohorts is an increasingly popular strategy for understanding the genetic and environmental interactions behind many common diseases, particularly diseases related to ageing and metabolic disorders. Robust, cost-effective and scalable assays are needed for studying large cohorts with sample sizes in the thousands or tens of thousands. Technical improvements along these lines have been crucial for population profiling of disease cohorts and continue to be imperative for extending DNA methylation assays for routine diagnosis in the clinic.

In this Review, we discuss recent technical advances in the profiling of DNA modifications, including 5mC, as well as three 5mC oxidation derivatives. As the principles and challenges of DNA methylation assays have been extensively covered previously[20], we focus on more recent developments. Importantly, we highlight the features that make each of these technologies suitable for specific types of studies. Recent reviews have covered in detail the principles behind methods for detection of 5hmC,[21,22] so we discuss these methods only briefly. Although developments in computational methods for interpreting DNA methylation data are important and involve many technical considerations, these methods have been the focus of previous reviews[23,24], so are not discussed here.

Progress in mapping cytosine methylation

The development of methods for sequencing-based quantification of cytosine methylation has been particularly active, because they can provide quantification in the form of digital counts, allowing for merging of data across different sequencing runs or batches of sequencing libraries, as well as for meta-analysis of data from different studies with less batch effects. Sequencing-based methods can be subcategorized as non-targeted enrichment methods, targeted enrichment methods and whole-genome methods (TABLE 1). Whole-genome methods can determine the pattern of DNA methylation across the entire genome, minus certain repetitive

Table 1 | **Overview of quantification methods**

DNA modification	Measurement	Non-targeted enrichment	Targeted enrichment	Whole genome	Arrays
5mC	Absolute (single base)	RRBS, mRRBS, LCM-RRBS or scRRBS	Microdroplet PCR Bisulphite, Patch PCR, mTACL, BSPP, LHC-BS (pre- and post-conversion) or RSMA	WGBS, T-WGBS or PBAT	Infinium BeadChip
	Relative (peak)	MRE-seq, MeDIP-seq, MBD-seq or MethylCap-seq			CHARM or MeKL-ChIP
5mC oxidation derivatives	Absolute (single base)	RRHP	Locus-specific sequencing with TAB-seq, oxBS-seq, CAB-seq, fCAB-seq or redBS-seq	TAB-seq, oxBS-seq, CAB-seq, fCAB-seq, redBS-seq	
		Reduce representation sequencing with TAB-seq, oxBS-seq, CAB-seq, fCAB-seq or redBS-seq			
	Relative (peak)	DIP-seq, anti-CMS, hMe-Seal,fC-Seal, GLIB, JBP1, EpiMark or Aba-seq			

Aba-seq, DNA-modification-dependent restriction endonuclease *Aba*SI coupled with sequencing; anti-CMS, anti-cytosine-5-methylenesulfonate; BSPP, bisulphite padlock probe; CAB-seq, chemical modification-assisted bisulphite sequencing; CHARM, comprehensive high-throughput arrays for relative methylation; DIP-seq, DNA immunoprecipitation and shotgun sequencing; fCAB-seq, 5fC chemical modification-assisted bisulphite sequencing; fC-Seal, a 5-formylcytosine selective chemical labelling (fC-Seal) approach for the affinity purification and genome-wide profiling of 5fC; GLIB, glucosylation, periodate oxidation and biotinylation; JBP1, J-binding protein 1; LCM-RRBS, laser-capture microdissection-reduced representation bisulphite sequencing; LHC-BS (pre- and post-conversion) , liquid hybridization capture based bisulphite sequencing; MBD-seq, methyl-CpG-binding domain protein sequencing; MeDIP-seq, methylation DNA immunoprecipitation sequencing; MeKL-ChIP, methylated DNA, kinase pretreated ligation-mediated PCR amplification-chromatin immunoprecipitation; MethylCap-seq, methylation DNA capture sequencing; MRE-seq, methylation restriction enzyme sequencing; mRRBS, multiplexed reduced representation bisulphite sequencing; mTACL, methylation target capture and ligation; oxBS-seq, oxidative bisulphite sequencing; PBAT, post-bisulphite adaptor tagging; redBS-seq, reduced bisulphite sequencing; RRBS, reduced representation bisulphite sequencing; RRHP, reduced representation 5hmC profiling; RSMA, methylation-sensitive restriction enzyme-based assay; scRRBS, single-cell reduced representation bisulphite sequencing; TAB-seq, TET-assisted bisulphite sequencing; T-WGBS, transposase-based library construction; WGBS, whole-genome bisulphite sequencing.

Box 1 | **Shotgun sequencing of bisulphite converted DNA**

Bisulphite conversion is a chemical treatment of DNA that leads to the deamination of cytosine to uracil. However, methylated cytosines are resistant to deamination and remain as cytosines. Shotgun sequencing of bisulphite converted DNA (BS-seq) results in thymidine being read at unmethylated cytosine positions and cytosine being read at methylated cytosine positions. Quantification of 5-methylcytosine (5mC) at individual positions is the ratio of methylated cytosines over total cytosines called.

Bisulphite conversion can be performed on prepared shotgun sequencing libraries, but must use adaptors that do not have unmethylated cytosines that can be converted to uracils (usually termed 'methylated adaptors'). Following conversion, the shotgun library is amplified by PCR, using primers that anneal to the common adaptor sequences. PCR on converted DNA requires a polymerase that can tolerate uracil residues. Furthermore, double-stranded DNA ceases to be complementary after bisulphite conversion, so molecules amplified from the forward strand can be distinguished from molecules amplified from the reverse strand.

A recent report confirmed that 5-hydroxymethylcytosines (5hmCs) are also resistant to deamination[86]. With standard BS-seq methods, 5mC and 5hmC are indistinguishable and would be read as 5mC+5hmC. This is especially a concern in cells with high 5hmC levels, such as brain tissues. Base-resolution sequencing of 5hmC and 5mC using a method that can distinguish between the two modifications found that 4.2% of total cytosines were 5mC and 0.87% were 5hmC[5], which represents a 20% difference in quantification if 5hmC were incorrectly called as 5mC. As no other mammalian tissues have been found to have the same level of 5hmC as in the brain, BS-seq can still be used to quantify 5mC in most cell types as the overlap of 5hmC readings can be considered negligible.

regions that cannot have reads uniquely assigned to them. By contrast, non-targeted and targeted enrichment methods provide information on specific subsets of sites in the genome. Targeted approaches are more cost-effective and efficient in analysing the genomic regions of interest, thus providing similar advantages to array-based methods. Both whole-genome and non-targeted approaches have been broadly used in developmental biology and epidemiology. Owing to their comprehensive coverage, whole-genome methods represent an important resource for charting the DNA methylation landscape in various cell types and conditions. Key advances in 5mC assays have been in the reduction of sample inputs towards the single-cell level. The need to lower sample inputs has been recognized as an important area of research as it enables the assay to be applied to cell types that may be difficult to isolate. Other notable developments in mapping 5mC include methods that measure 5mC and other epigenetic marks simultaneously.

Many sequencing-based methods use bisulphite conversion and shotgun sequencing (BS-seq) (BOX 1). The errors from bisulphite conversion include failed conversion of unmethylated cytosines to methylated cytosine, which contributes to type I errors, and inappropriate conversion of 5mC to thymine, which contributes to type II errors. These errors can be estimated using control methylated and unmethylated DNA sequences and extensive studies have shown that bisulphite conversion can be optimized to be highly efficient and selective[25,26]. Although 5hmC is also resistant to conversion and gives rise to the combined 5mC and 5hmC signals from BS-seq, the abundance of 5hmC is relatively low with it being at most 20% relative to 5mC in the brain[5] and <0.1% in human embryonic stem cells (hESCs), blood and spleen[27].

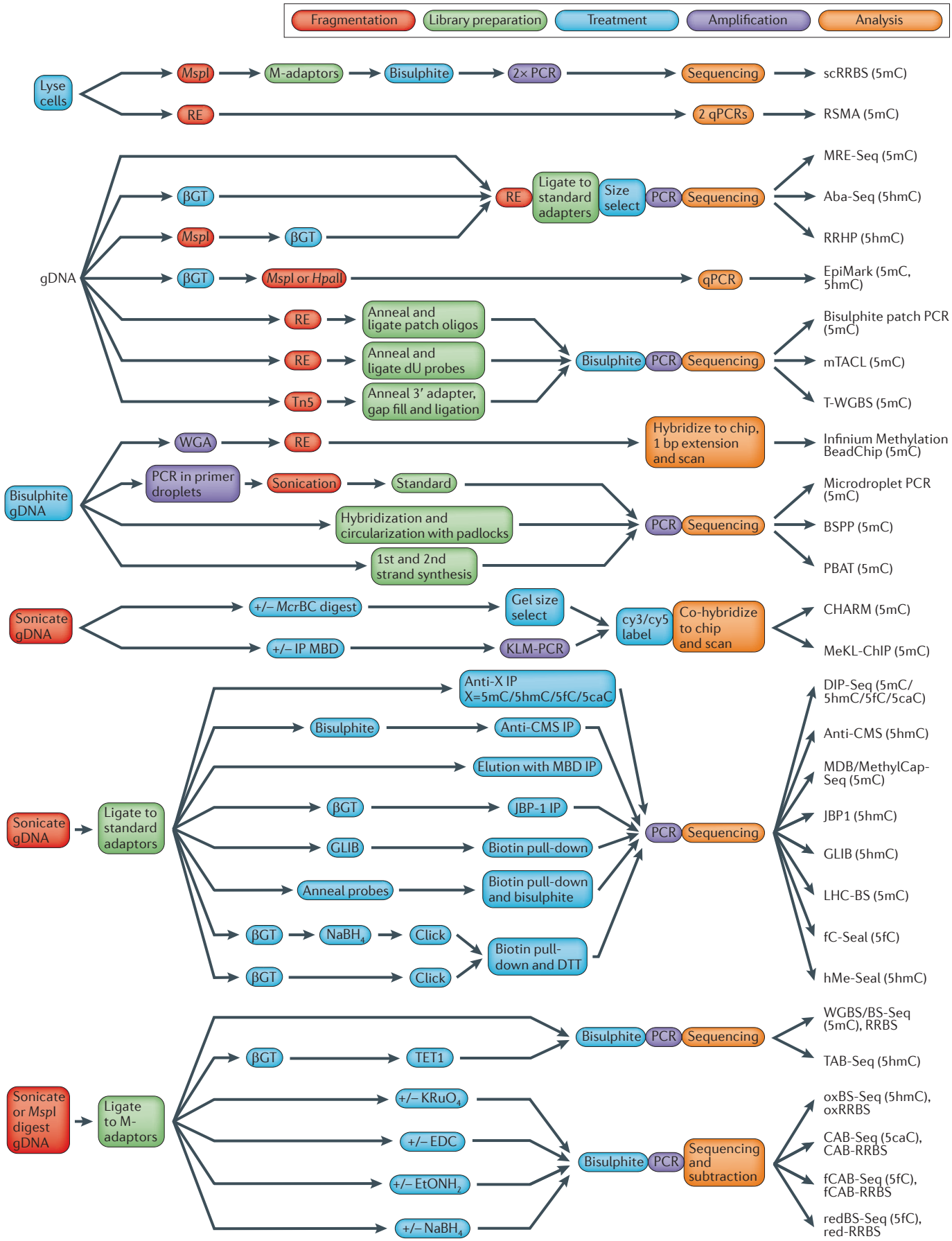
Non-targeted enrichment methods. It is known that DNA methylation occurs predominantly in CpG dinucleotides, however the occurrence of CpG in vertebrate genomes is substantially lower than the occurrence of other dinucleotides. For example, ~53% of non-overlapping 100 bp regions in the human genome (Hg 19) contain no CpG site. Selecting for CpG sites or methylated fragments before sequencing can reduce the sequencing requirements by at least tenfold. Numerous protocols have been developed based on this strategy, including reduced representation bisulphite sequencing (RRBS). RRBS takes advantage of methylation-insensitive restriction enzymes such as *Msp*I and *Ape*KI to select for CpG-rich sequences (FIG. 1). Using *Msp*I alone allows for coverage of up to 10% of CpGs[28], whereas double restriction with *Ape*KI and *Msp*I allows for coverage of up to 20% of CpGs in human cells[29]. Sample barcoding has been implemented for multiplexing RRBS (mRRBS)[28]. Additional efforts have reduced required input to ~1 ng of DNA from samples obtained using laser-capture microdissection (LCM-RRBS)[30] or single cells (scRRBS)[31] (FIG. 1). Although RRBS shares many features with whole-genome BS-seq (WGBS), coverage from RRBS is generally limited to CpG-dense regions, which leaves many CpG-sparse regions (that include some regions that overlap with functional enhancers) uncharacterized. For example, epigenetic abnormalities in low CpG-density regions specific to human induced pluripotent stem cells (hiPSCs) were consistently detected using WGBS and targeted assays but not by RRBS owing to the lack of coverage in enhancers and intronic regions[29,32].

In other assays, including methylation restriction enzyme sequencing (MRE-seq), methylation DNA immunoprecipitation sequencing (MeDIP-seq) and methyl-CpG-binding domain protein sequencing (MBD-seq), methylated DNA fragments are enriched for quantification bysequencing. Digestion by methylation-sensitive enzymes coupled with PCR amplification has traditionally been used to quantify 5mC levels at specific loci. Some methylation-sensitive restriction enzymes include *Mcr*BC, *Hpa*II, *Hin*6I and *Aci*I. Variants of non-targeted methods might also use a combination of methylation-sensitive and methylation-insensitive enzymes such as *Hpa*II and *Msp*I[33]. Recently, techniques using methylation-sensitive enzymes have been coupled with sequencing in the MRE-seq protocol[34]. MRE-seq can sample ~10% of CpGs in the human genome with only ~1.5 Gb of sequencing data (TABLE 2). Antibodies may also be used to enrich for methylated DNA. In the MeDIP-seq protocol, monoclonal antibodies specific to 5mC were used to enrich for methylated DNA fragments before sequencing. MeDIP-seq can quantify 5mC levels at a larger fraction of repeats than any other sequencing-based method[16,34]. Additionally, MeDIP can be performed on as little as 50 ng of starting DNA[35]. With ~17–18 Gb of sequencing, MeDIP-seq libraries can be saturated at 90% of total CpG coverage[36] (TABLE 2). Next, in the MBD-seq protocol, the Methyl-CpG-binding domain (MBD) of methyl-CpG-binding protein 2 (MeCP2) was used to capture methylated DNA[37,38]. This protocol has also been modified with fractionation

Type I errors
The errors that result when there are false positives or when falsely rejecting the null hypothesis.

Type II errors
The errors that result when there are false negatives or incorrect failure to reject the null hypothesis

Laser-capture microdissection (LCM)
A method for isolating specific cells or specific areas from cell, tissue or organism samples using laser cutting under microscopic visualization.



◀ Figure 1 | **Methods for quantifying 5-methylcytosine (5mC) and its oxidized derivatives.** The procedures may involve fragmentation, treatment, library preparation, amplification and analysis. Sonication, restriction enzyme (RE) digestion or Tn5 (transposase) may be used to fragment the genomic DNA (gDNA). Treatment of gDNA may involve enzymatic or chemical modifications, or selection by IP (immunoprecipitation), fragment size or probe. Bisulphite conversion is the primary treatment of gDNA that also results in fragmentation of the gDNA. Sequencing-based methods are usually performed with library preparation for sequencing, which can be done with 'standard' adaptors (standard adaptors used in sequencing library construction) or 'M-adaptors' (methylated adaptors used prior to bisulphite treatment). Other library preparation variants are also used. Amplification or limited amplification is necessary for methods using bisulphite treatment and sequencing-by-synthesis technologies and for low input. Finally, analysis comprises of sequencing, arrays scanning or quantitative PCR. Aba-seq, DNA-modification-dependent restriction endonuclease *Aba*SI coupled with sequencing; βGT, β-glucosyltransferase; BSPP, bisulphite padlock probe; CAB-RRBS, chemical modification-assisted bisulphite-representation bisulphite sequencing; CAB-seq, chemical modification-assisted bisulphite sequencing; CHARM, comprehensive high-throughput arrays for relative methylation; CMS, cytosine-5-methylsulphonate; DIP-seq, DNA immunoprecipitation and shotgun sequencing; fCAB-RRBS, Fc antigen binding-representation bisulphite sequencing; fCAB-seq, 5fC chemical modification-assisted bisulphite sequencing; fC-Seal, 5-formylcytosine selective chemical labelling; GLIB, glucosylation, periodate oxidation and biotinylation; JBP1, J-binding protein 1; LHC-BS, liquid hybridization capture-based bisulphite sequencing; MDB, methyl-DNA binding domain; MeKL-ChIP, methylated DNA kinase pretreated ligation-mediated PCR amplification chromatin immunoprecipitation; MethylCap-seq, methylation DNA capture sequencing; MRE-seq, methylation restriction enzyme sequencing; mTACL, methylation target capture and ligation; oxBS-seq, oxidative bisulphite sequencing; oxRRBS, oxidative RRBS; PBAT, post-bisulphite adaptor tagging; redBS-seq, reduced bisulphite sequencing; red-RRBS, reduced representation bisulphite sequencing; RRHP, reduced representation 5hmC profiling; RSMA, methylation-sensitive restriction enzyme-based assay; scRRBS, single cell reduced representation bisulphite sequencing; TAB-seq, TET-assisted bisulphite sequencing; T-WGBS, transposase-based library construction; WGBS/BS-seq, whole-genome bisulphite sequencing.

Ligation capture
A method for capturing restriction enzyme-digested DNA molecules via the annealing of an oligonucleotide containing complementary sequences to adaptor oligonucleotides to the DNA molecules and to the adaptor oligonucleotides. The adaptors and DNA molecules are then ligated together, allowing for PCR amplification of only the ligated products.

Bisulphite padlock probe (BSPP) capture
A method for capturing the target CpG sites of bisulphite treated genomic DNA with bisulphite padlock probes (BSPP). The two capturing arms of the BSPPs are designed to flank the region of interest. After annealing padlock probes to target regions, polymerization is preformed to fill the gap and two ends of the padlock probe are joined together to form circularized DNA. The captured regions are amplified with barcoded adaptor primers and sequenced.

elution using low-to-high salt concentrations in methylation DNA capture sequencing (MethylCap-seq). MethylCap-seq provides more uniform coverage across regions of different CpG densities and a higher total CpG coverage[39]. Typical MethylCap-seq libraries are sequenced at 2–3 Gb, although this may be far from saturation. To facilitate the transition to large cohort studies MeDIP-seq, MBD-seq and MethylCap-seq have been automated[37,39,40].

There are several technical caveats related to methylation enrichment methods (TABLE 3). For example, they do not quantify methylation at single-base resolution. The resulting data can only be interpreted as the relative abundance of 5mC in genomic windows (or peaks) of various sizes. Furthermore, in samples where copy number variations may cause bias, it is important to additionally sequence the input or negative controls to normalize the copy number differences at the genomic level. All three enrichment methods have some bias, which can be attributed to incomplete digestion of DNA, low specificity or sequence-dependent specificity of the selecting agents. Quantitative comparisons have shown that MeDIP-seq and MethylCap-seq have lower accuracy for mapping partially methylated regions compared to RRBS[41]. Additionally, hESCs assayed using MeDIP-seq showed enriched methylation in areas containing CA and CT repeats but hESCs assayed using WGBS showed a lack of methylation or partial methylation at the same regions[42]. With WGBS data currently available for many different cell types, it is possible to assess the sequence context bias genome-wide from these methylation enrichment technologies using WGBS data as the reference.

Targeted enrichment methods. Targeted methylation sequencing can be carried out using PCR amplification, ligation capture, bisulphite padlock probe (BSPP) capture or liquid hybridization. The development of these approaches relies heavily on improvements in microarray-based DNA synthesis technologies to make oligonucleotide synthesis more affordable and on more sophisticated oligonucleotide design to achieve higher specificities and sensitivities.

For PCR amplification of target loci from bisulphite converted DNA, primers need to be designed based on the bisulphite converted reference genome. PCR amplification can be difficult to multiplex for simultaneous assaying hundreds to thousands of targets. This limitation has recently been mitigated by performing multiple singleplex PCR amplification in emulsion droplets[43] (FIG. 1). The Raindance microdroplet PCR technology has allowed 3,500 loci to be targeted with 99% sensitivity and 90% specificity. Currently, the maximum number of targets for this approach is 22,000. Increased target sizes would require DNA input in the microgram range because the number of droplets scales linearly with the number of targets and each droplet requires multiple copies of the genomic template. In addition, this approach depends on a library of pre-made primer droplets, which would incur an initial investment for a customized set of targets.

There are two variants of ligation capture where enrichment of enzymatically digested DNA is achieved by annealing oligonucleotides followed by ligation with common adaptor sequences for PCR. Two notable examples of ligation capture are methylation target capture and ligation (mTACL)[44] and bisulphite Patch PCR[45]. In mTACL, a library of 'dU capture probes' is synthesized, that contains probes each composed of complementary sequences to the targeted DNA fragments flanked by sequences complementary to two common regions shared by all dU probes ('adaptors') and all thymidines in these sequences have been substituted with uracils. Common adaptors anneal to the probes next to the hybridized DNA and are then ligated to this hybridized DNA before bisulphite conversion and PCR amplification. Probes are then destroyed by a combination of enzymatic digestion and heat. Large mTACL libraries (with 19,250 probes) can be quantified through hybridization on tiling arrays, which has enabled the quantification of ~85% of targeted CpGs across 221 samples. Smaller mTACL libraries using 383 dU probes have also been prepared for pyrosequencing, but in these cases it has not been possible to determine the specificity of mTACL capture using the full probe set. In bisulphite Patch PCR, fragments of digested DNA are ligated to 'Patch oligos' that contain common adaptor sequences with an exonuclease-resistant 3′ modification. Subsequent incubation with exonucleases removes

Table 2 | **Comparison of assays**

	Assays	Sample requirement	Batch throughput limit	CpG coverage (Human and/or Mouse)	Sequenced gigabases (Human and/or Mouse)
Non-targeted enrichment	RRBS	10–100 ng	96 reactions per plate	10–20%	1–2
	mRRBS	100 ng	96 reactions per plate	10–20%	1*
	LCM-RRBS	1 ng	96 reactions per plate	~10%	1*
	scRRBS	One cell	96 reactions per plate	~10%	1*
	oxRRBS	2 µg	96 reactions per plate	10–20%	1–2
	MRE-seq	1–3 µg	96 reactions per plate	1–6%	1–2
	MeDIP-seq	50 ng–5 µg	96 reactions per plate	60–90%	2–4
	MBD-seq	1 µg	96 reactions per plate	60%	2–3
	MethylCap-seq	1 µg	96 reactions per plate	60–80%	2–3
	DIP-seq	50 ng–5 µg	96 reactions per plate	60–90%	2–4
	Anti-CMS	5 µg	96 reactions per plate	60–90%	2–4
	hMe-Seal	10 µg	96 reactions per plate	NA	2–4
	fC-Seal	50 µg	96 reactions per plate	NA	2–4
	GLIB	1–10 µg	96 reactions per plate	NA	2–4
	JBP1	1 µg	96 reactions per plate	NA	NA
	oxRRBS	100 ng–1 µg	96 reactions per plate	10–20%	1–2
	EpiMark	20 ng	96 reactions per plate	NA	NA
	Aba-seq	50 ng	96 reactions per plate	NA	>10
	RRHP	100 ng	96 reactions per plate	NA	0.2–0.5
Targeted enrichment	Microdroplet PCR	4 µg	8 samples per chip	<1%	8
	Bisulphite Patch PCR	250 ng	96 reactions per plate	<<<1%	0.005
	mTACL (analyse with sequencing)	200 ng	96 reactions per plate	1%	NA‡
	BSPP	500 ng	96 reactions per plate	2%	4
	Pre-conversion LHC	2–3 µg	96 reactions per plate	15%	8
	Post-conversion LHC	500 ng	96 reactions per plate	2%	4–7
Whole genome	WGBS	10 ng	96 reactions per plate	>90%	60–100
	T-WGBS	1–10 ng	96 reactions per plate	~90%	60*
	PBAT	125 pg–10 ng	96 reactions per plate	~90%	60*
	scBS-seq	One cell	96 reactions per plate	9–40%	0.2–5
	TAB-seq	1–3 µg	96 reactions per plate	>90%	160
	oxBS-seq	100 ng–1 µg	96 reactions per plate	>90%	320
	CAB-seq	NA	96 reactions per plate	>90%	320
	fCAB-seq	NA	96 reactions per plate	>90%	320
	redBS-seq	4 µg	96 reactions per plate	>90%	320
Arrays	Infinium BeadChip 450K	500 ng	12 samples per chip	2%	NA
	CHARM 2.0	2 µg	1 sample per chip	19%	NA
	MeKL-chip	10–20 ng	1 sample per chip	19%	NA

Aba-seq, DNA-modification-dependent restriction endonuclease AbaSI coupled with sequencing; βGT, β-glucosyltransferase; BSPP, bisulphite padlock probe; CAB-RRBS, chemical modification-assisted bisulphite- representation bisulphite sequencing; CAB-seq, chemical modification-assisted bisulphite sequencing; CHARM, comprehensive high-throughput arrays for relative methylation; CMS, cytosine-5-methylsulphonate; MeKL-chip, methylated DNA, kinase pretreated ligation-mediated PCR amplification-chromatin immunoprecipitation; DIP-seq, DNA immunoprecipitation and shotgun sequencing; fCAB-RRBS, Fc antigen binding-representation bisulphite sequencing; fCAB-seq, 5fC chemical modification-assisted bisulphite sequencing; fC-Seal, 5-formylcytosine selective chemical labelling; gDNA, genomic DNA; GLIB, glucosylation, periodate oxidation and biotinylation; IP, immunoprecipitation; JBP1, J-binding protein 1; LHC-BS, liquid hybridization capture based bisulphite sequencing; M-adaptors, sequencing library construction using methylated adaptors; MDB/MethylCap-seq, methyl-DNA binding domain/MethylCap-seq; MRE-seq, methylation restriction enzyme sequencing; mTACL, methylation target capture and ligation; NA, not applicable; oxBS-seq, oxidative bisulphite sequencing; oxRRBS, oxidative RRBS; PBAT, post-bisulphite adaptor tagging; RE, restriction enzyme; redBS-seq, reduced bisulphite sequencing; red-RRBS, reduced representation bisulphite sequencing; RRHP, reduced representation 5hmC profiling; RSMA, Methylation-sensitive restriction enzyme-based assay; scBS-seq, single cell BS-seq scRRBS, single cell reduced representation bisulphite sequencing; TAB-seq, TET-assisted bisulphite sequencing; T-WGBS, transposase-based library construction; WGBS/BS-seq, whole-genome bisulphite sequencing. *Low input samples may saturate at lower sequencing amount due to low library complexity. ‡Full mTACL libraries were not sequenced.

unwanted products and excess primers followed by bisulphite conversion, PCR amplification with barcoded primers and pyrosequencing (FIG. 1). Although this method has been applied on only 94 loci, 100% sensitivity and 90% specificity across 48 samples was achieved.

In the BSPP method[46,47], bisulphite-converted DNA is captured with a high specificity by padlock probes containing two short capture sequences that are joined through a common linker sequence (FIG. 1). Capture sequences are selected to be complementary to bisulphite converted DNA and to have minimal overlap with CG dinucleotides. A distinct feature of BSPP is that between hundreds and hundreds of thousands of probes can be pooled into a single capturing reaction. Specificities of over 97% have been demonstrated with BSPP. Consistent capture of more than 500,000 CpG sites from only 500 ng of genomic DNA has been achieved using a library of approximately 330,000 padlock probes[46] (TABLE 2). The main advantages of BSPPs are flexibility and scalability in selecting genomic targets. The availability of oligonucleotide libraries has been a major barrier in adopting BSPPs. However, oligonucleotide libraries are now increasingly accessible and affordable from multiple commercial vendors.

Liquid hybridization capture, which has been used successfully in exome capture and sequencing, was recently extended for use in targeted 5mC quantification with liquid hybridization capture-based bisulphite sequencing (LHC-BS) (FIG. 1). First, genomic DNA is prepared by fragmentation and ligation of common adaptors. The targeted fragments are then hybridized to biotinylated oligonucleotides, selected by affinity enrichment, bisulphite converted, PCR amplified and sequenced. A former limitation of LHC-BS was low library complexity and hence large amounts of input DNA were required[48,49]. To lower the input requirement, targeted fragments have been enriched from a post-amplification WGBS preparation with custom oligonucleotides that hybridize to post-conversion DNA[50]. Optimizations in bisulphite conversion and post-conversion amplification have helped to overcome this initial difficulty (J. Stamatoyannopoulos, personal communication), which has led to the development of the commercial SureSelect MethylSeq method by Agilent Technology. Another commercially available method (NimbleGen SeqCap EZ) is based on the same principle. The specificity of liquid hybridization methods may vary (between 70%–80%), heavily depending on oligonucleotide design, and is generally lower than the specificity achieved with other ligation-based target enrichment methods. However, liquid hybridization generally has more uniform target coverage compared to ligation capture and BSPP capture, which translates to better sensitivity and cost efficiency.

For projects with large sample sizes, bisulphite Patch PCR, BSPP capture and liquid capture assays can be multiplexed by tagging individual samples with unique DNA barcodes. Sample barcoding at an early step allows for pooling of tens to hundreds of samples for single-tube processing, exponentially reducing the labour and assay cost while improving the efficiency and data consistency.

Whole-genome methods. Whole-genome bisulphite sequencing (WGBS) (FIG. 1) has been considered the 'gold standard' in DNA methylation profiling, as it can provide single-base resolution with full genome coverage without the biases associated with selecting agents (TABLE 3). Recent developments in shotgun library construction methods have reduced WGBS sample input requirements to the nanogram level, equivalent to a few thousand cells. Low-input WGBS methods have been instrumental for mapping methylation in rare cell types such as primordial germ cells and oocytes[51–54]. The readouts from WGBS assays are digital counts of unmethylated and methylated cytosines found at individual genomic locations. Higher precision is typically achieved by minimizing amplification, as well as increasing library complexity and sequencing depth.

One recent development is the introduction of transposase-based library construction for WGBS (T-WGBS; also known as Tn5mC-seq)[55,56] (FIG. 1). The assay uses a Tn5 transposase derivative to fragment double-stranded DNA and to append a 5′ methylated adaptor in a single step. This step is followed by annealing of the 3′ adaptor, gap-filling and ligation. Additional processing of the shotgun library follows a BS-seq protocol (FIG. 1). This method reduces the number of steps required for sequencing library construction and also reduces the DNA input requirement to less than 50 ng. As little as 1 ng of input DNA is sufficient to generate a sequencing library, although the library will probably require substantially more amplification, resulting in reduced complexity[55]. There are some technical caveats in T-WGBS, including a potential bias introduced by the use of transposase and an increased risk of losing library complexity after many cycles of PCR amplification. Furthermore, T-WGBS was initially demonstrated on purified DNA. DNA purification can lead to excessive loss of materials when input is very limited, and direct bisulphite treatment on cells without the DNA purification step might lead to incomplete bisulphite conversion[51].

Another low-input method for whole-genome methylation sequencing is post-bisulphite adaptor tagging (PBAT)[57] (FIG. 1), which has been demonstrated with as little as 125 pg of DNA. Here, adaptor tagging precedes bisulphite treatment to reduce sample loss. Moreover, a bisulphite conversion protocol with heat denaturation has been optimized to enhance DNA fragmentation. A shotgun sequencing library is then generated from bisulphite treated single-stranded DNA using two rounds of random priming with primers containing four random nucleotides on the 3′ end. Similar to other methods using random primers, a limitation of this approach is that the fraction of aligned and usable sequencing reads is relatively low compared to standard adaptor ligation methods. In addition, the resulting libraries are difficult to quantify accurately and sequence because of their extremely small quantities. This approach has been made commercially available with the EpiGnome Methyl-seq kit (Epicentre), which can provide highly diverse and uniform coverage libraries that can be generated from as little as 50 ng input.

Liquid hybridization
A method for capturing fragmented DNA molecules via the annealing of biotinylated oligonucleotides to the DNA molecules. The binding of biotin to streptavidin beads allows for washing and removal of uncaptured DNA molecules, and subsequent elution of the captured DNA molecules.

Microdroplet PCR
Massively parallel PCR amplification of target sequences in microdroplets. The process involves the preparation of template and PCR mix in picoliter volume and primer droplets, combination of individual template and primer pair droplet, pooling the fused droplets for thermal cycling, and releasing of PCR amplicons for purification and sequencing.

Barcoded primers
Unique DNA sequences that are incorporated into adaptor sequences for tagging of different samples before sample pooling and shotgun sequencing.

Pyrosequencing
A sequencing-by-synthesis method based on the detection of phyrophosphate released upon nucleotide incorporation.

Shotgun library construction
The generation of a sequencing library involving random fragmentation of DNA and addition of adaptor sequences to both ends of DNA fragments before sequencing.

Transposase-based library construction
A procedure to generate a sequencing library using the transposase Tn5 to insert common transposon sequences to DNA. DNA segments are then amplified by annealing of primers to the transposon sequences.

Tn5 transposase
A member of the RNase superfamily of proteins that harbours retroviral integrases to catalyse random insertion of transposon DNA into target DNA.

Table 3 | **Sources of bias**

	Method	Amplification	Effect of copy number variation on quantification	Incomplete treatment (chemical or enzyme digestion)	Background signals	Batch to batch variation	Cross hybridization	Sensitivity to sequence context
5mC assays	BS-seq	Low*	None	Low	None	None	None	Medium‡ (RRBS only)
	Low input BS-seq	High§	None	Low	None	None	None	Medium (RRBS only)
	BS-based arrays	Low	None	Low	None	Medium	Medium	None
	Non-BS-seq	Low	Medium	Low	High	None	None	Medium (MeDIP/MBD)
	Low input non-BS-seq	High	Medium	None	High	None	None	Medium (MeDIP/MBD)
	Non-BS- arrays	Low	Medium	None	High	Medium	Medium	None
5mC oxidation derivatives assays	BS-based seq	Low	None	Low	None	None	None	None
	Non-BS-based seq	Low	Medium	Low	High	None	None	Medium
	RE-based 5hmC seq	Low	None	Low	None	None	None	Medium
All	Third generation sequencing	Low	None	None	None	None	None	None

5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; BS-seq, bisulphite sequencing; MBD, Methyl-CpG-binding domain; MeDIP, methylation DNA immunoprecipitation sequencing; RE-based 5hmC seq; RRBS, reduced representation bisulphite sequencing. *Low, can be ignored in most cases. ‡Medium, should be considered when performing analyses. §High, must be considered when performing analyses.

Most experimental methods for whole-genome methylation sequencing are easily scalable, as they use inexpensive reagents and are performed in tubes or plates compatible with automation. However, the amount of sequencing required is generally high, and insufficient sequencing might result in sparse coverage and random sampling errors at single CpG sites. For example, to detect a minimum methylation level of 0.10 (coverage of at least 10×) for a minimum of 99.5% of the human genome requires an average sequencing coverage of 20× or 60 Gb total sequencing (TABLE 2). The main hindrance to the implementation of these methods is the cost of sequencing and computational efforts which remain high for typical laboratories. Local binning of data on multiple CpG sites can mitigate the random sampling error at the expense of resolution.

Single-cell methods. Cytosine methylation is a dynamic modification of DNA that can be actively created by *de novo* methyltransferases (DNMT3s) or removed by the TET proteins. It has been recognized that cell-to-cell variability can exist not only among different cell types but also within a relatively homogeneous cell population. Therefore, there is a growing interest in characterizing methylation patterns in single cells to distinguish functional variability from transient and stochastic noise.

Methylation-sensitive restriction enzyme-based assay (RSMA)[58], is a robust method for quantifying 5mC at specific loci in single cells. Methylation-sensitive enzymes that can be blocked by fully methylated or hemimethylated DNA include *Hpa*II, *Hin*6I/*Hin*P1I and *Bst*UI. Recently, DNA methylation status was successfully characterized in single blastomere cells using *Bst*UI digestion coupled with PCR and sequencing, demonstrating epigenetic mosaicism in several imprinted regions owing to the loss of a methylation protecting protein, transcription intermediary factor 1-β (TRIM28)[59].

Bisulphite-based methods had previously been considered unsuitable for single-cell experiments, as the majority of input DNA is either lost or damaged after bisulphite conversion. However, recent improvements in bisulphite treatment protocols have enabled the use of this approach in single-cell analysis. Methylation levels at specific loci in single oocytes or embryos at the two-cell stage or the 16-cell stage have been quantified by combining low input bisulphite conversion with limiting dilutions, PCR and pyrosequencing[60]. Very recently, a modified RRBS protocol was successfully applied to single mouse ESCs, and in male and female pronuclei (scRRBS)[31] (FIG. 1). This was accomplished by eliminating all purification steps before bisulphite conversion, as well as by using carrier tRNA to improve the binding of DNA to the column and to reduce sample loss. Approximately 1 million unique CpG sites from single mouse ESCs were covered and close to 50% of these sites were sequenced deep enough for quantification. This approach has been used to profile the methylomes of early human embryos[61]. Single-cell BS-seq (scBS-seq) is a recent single-cell whole genome methylation profiling method that modifies the post-bisulfite adaptor tagging (PBAT) protocol. Bisulfite conversion is performed on single cells prior to five rounds of the first strand random priming to tag and

Binning
A computational technique frequently used to reduce noise by grouping sequencing reads mapped to contiguous genomic segments.

increase amount of the starting fragments, followed by second strand random priming and amplification. This approach can detect up to 48.4% of CpGs sites in mouse genome and represents the largest amount of methylation information that has been successfully obtained from a single cell to date[62]. Further improvements in this direction will probably involve reducing DNA damage and sample loss during bisulphite conversion, adapting low-bias whole-genome amplification methods to bisulphite converted DNA, and reducing the sample loss during liquid handling by simplifying the procedure or using microfluidic reactors.

Array-based methods. Array-based assays have been widely adopted owing to their low costs, ease of use and high throughput. Two examples of such assays include the comprehensive high-throughput arrays for relative methylation (CHARM)[63] (FIG. 1) and the Illumina Infinium bead chips[64]. CHARM chips do not provide single-base resolution but can be coupled to any methylation enrichment protocol. These arrays can also be synthesized to include non-CpG methylation and repetitive regions. The Illumina Infinium 450K BeadChips can quantify CpG and a very small fraction of non-CpG methylation at single-base resolution (FIG. 1). Microarrays have been coupled with MeDIP or MBD (mCIP-chip, MeDIP-chip, MeDIP-on-RepArray and MDB-chip) to specifically target promoters and repeat regions[65–67]. Promisingly, sample requirements for array-based methods have been reduced to ~10 ng of starting fragmented DNA with kinase pretreated ligation-mediated PCR amplification of 5mC enriched DNA and hybridization to custom tiling arrays (MeKL-ChIP methylated DNA, kinase pretreated ligation-mediated PCR amplification-chromatin immunoprecipitation)[68] (FIG. 1). For this approach, MBDs are used to enrich for methylated fragments, followed by end-repair, adaptor ligation and PCR amplification. DNA fragments enriched by targeted selection may also be hybridized to standard microarrays[44]. Cross hybridization remains a primary source of bias. Furthermore, the high input requirement for arrays means that some amplification must be performed when the input material is limited.

There are a number of possible improvements to array-based methylation assays. Technical improvements should be mainly focused on increasing the number of features per array, which translates to a lower cost per site. This involves substantial upfront investments and resources that are available only to large commercial providers, and it typically takes years to add new content to such arrays. Nonetheless, when a highly optimized commercial assay is widely adopted by the community, a large number of data sets are generated by various groups with a standardized protocol at a consistent quality, which allows for meta-analyses at a level well beyond a single study. Finally, assays like Illumina BeadChips tend to have lower performance on formalin-fixed paraffin-embedded tissue samples. However, this can be mitigated by 'restoring' damaged DNAs before the BeadChips assay[69].

Simultaneous quantification of epigenetic marks. Multiple epigenetic modifications operate in parallel, but in most previous studies DNA methylation, chromatin modifications and chromatin accessibility have only been correlated using data from independent experiments on different samples. It is highly desirable to obtain information on multiple or all modifications from the same sample in order to understand the combinatorial effects of these modifications on gene regulation.

The interaction between DNA methylation and histone modifications was recently characterized by integrating chromatin immunoprecipitation (ChIP) with bisulphite sequencing (ChIP-BS-seq; also known as BisChIP-seq and ChIP-BMS)[70–72] (FIG. 2). For example, several studies have used H3K27me3-specific antibodies to enrich for transcriptionally silent chromatin before bisulphite conversion and shotgun sequencing[70,71]. This approach has enabled simultaneous quantification of DNA methylation, histone modifications and transcription factor binding to investigate the interplay between the various layers of chromatin modifications.

Single chromatin molecule analysis at the nanoscale (SCAN) (FIG. 2) was developed using nanofluidic devices for the quantification of epigenetic markers on single DNA molecules[73,74]. SCAN was first developed to detect single native chromatin by monitoring fluorescent colour signatures tagged on DNA and chromatin molecules[73]. Recent improvements of this technique allow for real-time detection of multiple fluorescence-activating molecules and for sorting DNA fragments on the basis of fluorescence signatures with 98% accuracy[74]. SCAN applied to healthy tissue and cancer cells has revealed both antagonism and colocalization patterns of gene silencing marks, such as H3K9me3 and H3K27me3, and DNA methylation[75], although the SCAN readouts still remain to be connected with the sequence information of single DNA molecules.

Another approach called nucleosome occupancy and methylome sequencing (NOMe-seq) (FIG. 2) was developed to determine the genome-wide footprints of nucleosome position and DNA methylation on the same DNA molecules[76,77]. In NOMe-seq, nuclei are treated with GpC methyltransferase (*M.Cvi*PI) before bisulphite conversion and deep sequencing. The accessibility of the enzyme to GpC sites is used to map nucleosome position, as GpC sites in the genomic regions occupied by nucleosomes are protected from methylation by *M.Cvi*PI. By mapping both GpC methylation to determine nucleosome position and endogenous CpG methylation at single-base resolution, the colocalization between nucleosome position and DNA methylation status on the same DNA molecules can be determined. The advantage of assessing nucleosome position by GpC methyltransferase is that it can be used to map genome-wide localizations with high resolution. Using this technique, the different frequencies of nucleosome occupancy and DNA methylation on specific genomic regions, such as binding sites of the transcriptional repressor CCCTC-binding factor (CTCF) and promoter regions, have been identified[77].

Nucleosome
A basic unit of DNA packaging in eukaryotes that consists of section of DNA (~166 bp) wrapping around a histone core. Nucleosome structure helps to condense DNA into smaller volume. Nucleosomes are subunits of chromatin.

GpC methyltransferase (*M.Cvi*PI)
An enzyme from Chlorella virus that methylates all cytosines within the double-stranded dinucleotide recognition sequence 5′… GC…3′.

CCCTC-binding factor
(CTCF). A chromatin binding factor with highly conserved zinc finger domains that control binding to consensus DNA target sequences. CTCF regulates transcription by binding to chromatin insulators and preventing interaction between the promoter and enhancers or silencers.

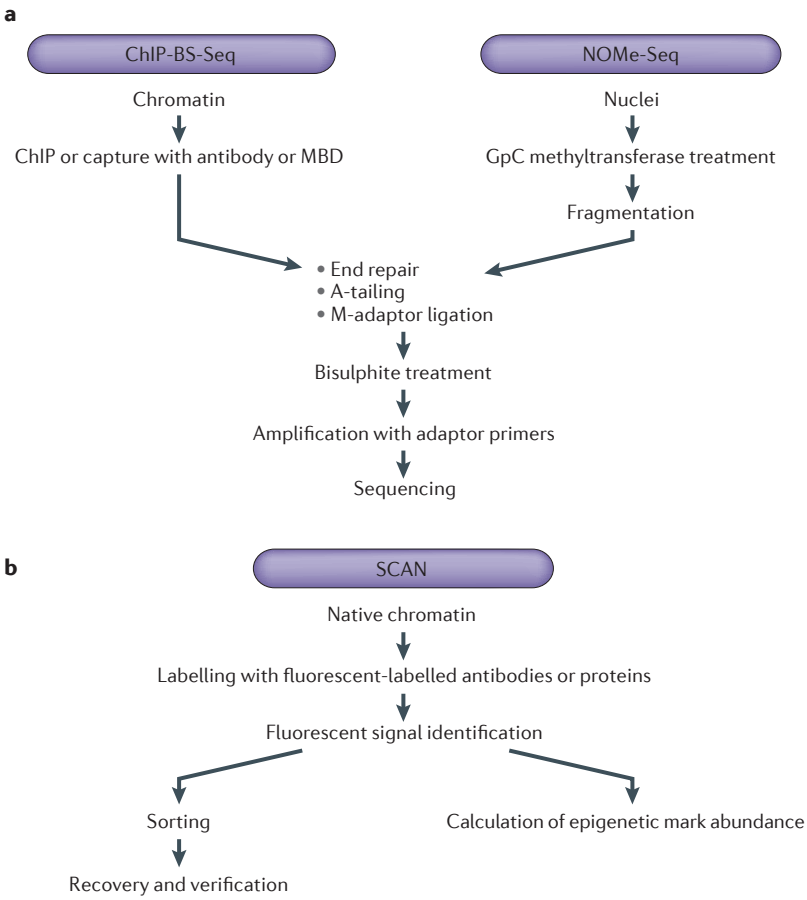


Figure 2 | **Simultaneous detection of 5mC and other epigenetic modifications.** **a** | Chromatin-immunoprecipitation bisulphite sequencing (ChIP-BS-seq; also known as BisChIP-seq and ChIP-BMS) assays identify the interaction between DNA methylation and histone marks. Chromatin is immunoprecipitated with a specific antibody or binding protein to enrich for the histone mark of interest before performing BS-seq of enriched DNA. Nucleosome occupancy and methylome sequencing (NOMe-seq) determines the position of nucleosome and 5-methylcytosine (5mC) on the same DNA. GpC methyltransferase treatment is used to map the nucleosome position. Only accessible DNA is methylated at GpC positions. Standard BS-seq is used to map the position of 5mC. **b** | For single chromatin molecule analysis at the nanoscale (SCAN), the histone marks of interest or 5mC are bound with fluorescent-labelled proteins or antibodies. The labelled DNA passes through the nanofluidic channel and is sorted according to their fluorescent labels. MBD, methyl-CpG-binding domain.

Mapping 5mC oxidation derivatives

In this section, we discuss the techniques for detecting the three recently discovered cytosine modifications: 5hmC, 5fC and 5caC (FIG. 3). Detection of these modifications in mammalian genomes is technically challenging as their abundances are extremely low. The detectable levels of 5hmC are less than 1% of total cytosine in human and mouse brain tissues[78]. Quantification of 5fC and 5caC is even more challenging as the level of these two bases is lower than 20 ppm of total cytosine[78]. Early efforts in mapping these nucleotide variants relied on non-targeted enrichment methods that utilize either specific antibodies or chemical labelling (TABLE 1). Recently, whole-genome methods have also been developed that combine enzymatic or chemical treatments with BS-seq to generate single-base resolution maps for all 5mC oxidation derivatives. Finally, progress in third-generation sequencing, such as single-molecule real-time (SMRT) sequencing and nanopore sequencing, has allowed for direct reading of 5hmC, 5fC and 5caC without DNA amplification or DNA treatment.

Third-generation sequencing
A new progression of sequencing technology that aims to improve throughput and reduce sequencing cost and time. The main goals of third-generation sequencing are to eliminate the DNA amplification step before sequencing and to enable real-time signal monitoring.

Non-targeted enrichment methods. There are three main subcategories of non-targeted enrichment methods for quantification of 5mC oxidation derivatives: antibody-based, selective chemical labelling and restriction enzyme-based methods. Similar to enrichment methods for 5mC, these non-targeted enrichment methods are substantially more cost-efficient than whole-genome methods and are currently more accessible for most laboratories.

DNA immunoprecipitation and shotgun sequencing (DIP-seq) (FIG. 1) assays for the mapping of 5mC oxidation derivatives have been performed in mouse ESCs[16,79,80], and 5hmC DIP-seq (also known as hmeDIP-seq) has been applied to hESCs[27,81]. Similar to MeDIP, such enrichment approaches might preferentially detect cytosine modifications in CpG-dense regions and simple repeats[82], however, such maps have revealed important patterns of deposition for these modifications in relation to the activity of the DNA-modifying enzymes (TET1, TET2 and TDG). DIP-seq of 5hmC, 5fC and 5caC are currently feasible using commercial antibodies with well-characterized specificities[16,17,27,79,83,84]. Bisulphite treatment of 5hmC converts it to cytosine-5-methylsulphonate (CMS), a variant that can be immunoprecipitated using CMS-specific antibodies to detect the levels of 5hmC[15,85,86]. Anti-CMS enrichment is more sensitive and specific than 5hmC enrichment due to the negative charge of CMS. This method is also less dependent on 5hmC density than other methods.

In addition to antibody-based enrichment, selective labelling of 5hmC using β-glucosyltransferase (βGT) is another enrichment strategy for mapping 5hmC in mammalian genomes. βGT has been used to transfer radiolabelled glucose from a UDP-[^{3}H]-glucose donor to 5hmC, thereby enabling quantification of the global abundance of 5hmC[87]. Similarly, in the hMe-Seal assay, βGT can be used to add azide-glucose to 5hmC, which can be subsequently tagged with biotin[88]. The presence of an azide group (N3) allows for tagging with biotin or other molecules using click chemistry[89]. Affinity enrichment for biotin-N_3-5gmC is achieved using avidin-horseradish peroxidase (HRP) and allows for accurate global quantification of 5hmC[88]. In another method, 5hmC undergo glucosylation with βGT-mediated transfer, periodate oxidation with $NaIO_4$, biotinylation with aldehyde and subsequent pull-down (GLIB; glucosylation, periodate oxidation and biotinylation)[15,90] (FIG. 1). The product of glucosylation, 5gmC, can also be directly captured using J-binding proteins (for example, J-binding protein 1 (JBP1)[91] (FIG. 1). Approaches yielding 5gmC provide greater flexibility for single-molecule quantification as they are selectively labelled with different tags, such as fluorescent dyes[92]. Recently, covalent chemical labelling strategies have been extended for the

β-glucosyltransferase (βGT). An enzyme that transfers the glucose residue of uridine diphosphate glucose (UDP-Glc) specifically to the hydroxyl group of 5-hydroxymethylcytosine to generate β-glucosyl-5hmC (5gmC).

mapping of 5fC with 5fC-selective-chemical labelling in fC-Seal (FIG. 1) where all 5hmCs are first blocked with unmodified UDG-Glc, then 5fC is reduced to 5hmC with sodium borohydride ($NaBH_4$) and labelled with biotinylated-UDG-Glc for enrichment and sequencing[93]. 5fC can also be detected using PAGE following specific labelling with a highly active amino-containing fluorescent dye[94].

Although chemical-labelling-based methods might intuitively be expected to result in lower bias related to CpG density, as well as higher specificity, accuracy and better mapping resolution, a recent comparison showed that antibody-based enrichment can yield highly specific and accurate measurements[16]. A caveat associated with chemical labelling is that abasic (apurinic–apyrimidinic) sites in the genome, which are generated during DNA

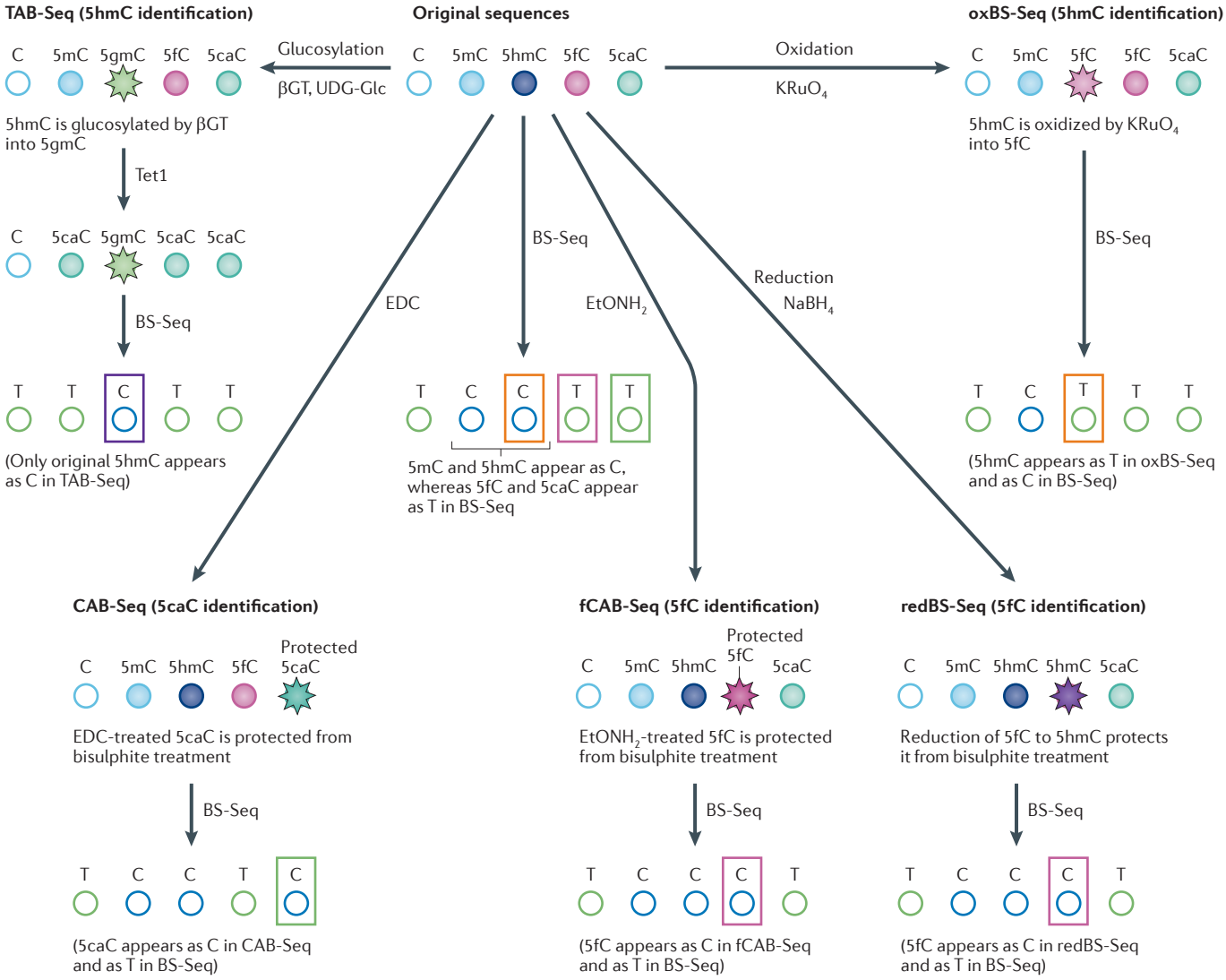


Figure 3 | **Assays for mapping 5-methylcytosine (5mC) oxidation derivatives at single-base resolution.** 5-hydroxymethylcytosine (5hmC) in single-base resolution is characterized by the two methods, oxidative bisulphite sequencing (oxBS-seq) and TET-assisted bisulphite sequencing (TAB-seq). In oxBS-seq, 5hmC is oxidized to 5-formylcytosine (5fC) by potassium perruthenate ($KRuO_4$). After bisulphite treatment and amplification, it appears as thymidine. 5hmC can be identified by subtracting thymidine from oxBS-seq from cytosine by traditional bisulphite sequencing (BS-seq). For TAB-seq, 5hmC in DNA is glucosylated to 5-glucosylmethylcytosine (5gmC) by β-glucosyltransferase (βGT). DNA is subsequently treated with ten-eleven translocation methylcytosine dioxygenase 1 (TET1) to convert all modified cytosines except 5gmC to 5-carboxyctosine (5caC). After BS-seq of TET1 treated DNA, only 5gmC from original 5hmC appears as cytosine. To identify 5caC, DNA is treated with 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC) to protect deamination of 5caC during bisulphite treatment. The protected 5caC is read out as cytosine in chemical modification-assisted bisulphite sequencing (CAB-seq) instead of thymidine in regular BS-seq. Similar to 5caC, 5fC chemically modification-assisted bisulphite sequencing (fCAB-seq) identifies 5fC by treating DNA with O-ethylhydroxylamine ($EtONH_2$) to modify 5fC and protect it from deamination during bisulphite treatment. 5fC can be reduced by sodium borohydride ($NaBH_4$) into 5hmC by the reduced bisulphite sequencing (redBS-seq) method. Both protected 5fC and 5hmC from original 5fC are read out as cytosine in fCAB-seq and redBS-seq. The signals of 5caC and 5fC are identified by subtracting cytosine from CAB-seq and fCAB-seq and/or redBS-seq, respectively, from thymidine in BS-seq. UDG-Glc, uridine diphosphate glucose.

damage and repair, can also be labelled and contribute to nonspecific background. By contrast, abasic sites have no effect on antibody enrichment as long as specific antibodies are used. Abasic site-related background noise in chemical-labelling-based data tends to have a larger effect on rare modifications, such as 5fC and 5caC. Thus, carefully selected controls or blocking reagents need to be included in chemical labelling experiments.

Similar to the detection of 5mC abundance using methylation-sensitive restriction enzymes, restriction endonuclease digestion coupled with glucosylation and PCR has been applied to map the abundance of 5hmC in the mammalian genome. *Msp*I and its isoschizomer *Hpa*II are used to map 5hmC in CpG sites within the CCGG sequence[95] as these two enzymes have different sensitivities to glucosylation of 5hmC (5-glucosylmethylcytosine or 5gmC). Treatment of DNA with βGT protects 5hmC from *Msp*I-mediated but not *Hpa*II-mediated digestion. After treatment with βGT, the DNA is digested with either *Msp*I or *Hpa*II, and amplified by locus-specific real-time PCR. The relative abundance of 5hmC at specific loci is then determined on the basis of the difference between real-time PCR readouts of *Msp*I-digested DNA and *Hpa*II-digested DNA. This method allows for the detection of 5hmC in low abundance regions and is commercially available as the EpiMark 5-hmC and 5-mC Analysis kit from New England Biosystems.

Another approach used DNA-modification-dependent restriction endonuclease *Aba*SI coupled with sequencing (Aba-seq)[96] (FIG. 1) to map genome-wide distribution of 5hmC at high resolution. *Aba*SI is used following βGT treatment to cleave dsDNA at the 3′ end of 5gmC at a fixed length of roughly 11–13 nucleotides or 9–11 nucleotides away from the 5gmC recognition site[96,97]. The resulting DNA fragments can then be converted into sequencing libraries for deep sequencing. The locations of 5hmC are determined by mapping the cleaved ends. Aba-seq enables quantification of 5hmC at almost single-base resolution and sensitive detection of low abundance sites.

Zymo Research has developed a reduced representation 5hmC profiling (RRHP) (FIG. 1) method for 5hmC profiling at single-base resolution. In RRHP, *Msp*I is used to digest genomic DNA at GC-rich regions before attaching adaptors to both ends of the digested fragments. With βGT treatment, 5hmC at adaptor junctions is protected from *Msp*I re-digestion. Only DNA fragments with intact adaptors at both ends are amplified and sequenced. The location of 5hmC at positions where adaptors are ligated can be determined by deep sequencing. Although only 5hmC positions at *Msp*I recognition sites can be mapped, RRHP is a bisulphite treatment-free method and reduces the required amount of input DNA to ~100 ng.

The main advantage of restriction enzyme-based methods in comparison to other enrichment methods is that they can use considerably less input DNA.

Whole-genome methods. Methods have been developed to discriminate individual 5mC oxidation derivatives at single-base resolution and whole-genome coverage using either enzymatic modifications or chemical treatments before bisulphite conversion and deep sequencing. Similar to conventional BS-seq, all of the following methods can be performed with locus-specific PCR after conversion instead of whole-genome sequencing.

To map 5hmC at single-base resolution, oxidative bisulphite sequencing (oxBS-seq) (FIG. 1) chemically modifies 5hmC through an oxidation reaction[98]. Potassium perruthenate ($KRuO_4$) is used to oxidize 5hmC to 5fC, which is then detected as thymidine after performing BS-seq on the treated DNA, while 5mC remains unchanged and is detected as cytosine (FIG. 3). An optimized oxBS-seq protocol has been reported to reduce oxidation-induced DNA degradation[99]. While 5mC is directly quantified, the quantification of 5hmC is obtained by subtraction of 5mC signals in oxBS-seq from 5hmC+5mC signals of BS-seq on the same samples. As the subtraction is performed on the percentages of modification resulting from two random samplings of cytosines, it is necessary to further compute the confidence interval for each quantification.

Another method for sequencing-based 5hmC detection, called Tet-assisted bisulphite sequencing (TAB-seq) (FIG. 1), involves enzymatic conversion of non-5hmC modifications[14,100]. In TAB-seq, βGT-treated DNA is further treated with the TET1 protein, which converts 5mC and 5fC to 5caC, while 5gmC generated from original 5hmC is protected (FIG. 3). BS-seq is then performed on these DNA in either whole-genome sequencing or locus-specific sequencing allowing only 5gmC to be unconverted, while the unmodified cytosine, 5mC, 5fC and 5caC are read asthymidines. The position and abundance of 5hmC in human and mouse ESCs were determined by TAB-Seq[100].

In addition to 5hmC, single-base resolution mapping of 5caC and 5fC in mammalian genomes has recently been achieved. These two bases are deaminated by bisulphite treatment and read as thymidine in sequencing. To distinguish 5caC from other bases, chemical modification-assisted bisulphite sequencing (CAB-seq)[101] (FIG. 1) has been developed. A reaction involving 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC)-based coupling selectively labels the carboxyl group in 5caC (FIG. 3). In this reaction, EDC catalyses the formation of an amide bond between the carboxyl group of 5caC and the primary amine group, protecting 5caC from deamination during bisulphite treatment. The protected 5caC modifications are unconverted in BS-seq. In addition, 5caC-containing DNA fragments can be enriched before BS-seq using selective biotin labelling through the click chemistry described above. The quantification of 5caC can be obtained by subtraction of 5hmC+5mC signals from BS-seq on the same samples.

Finally, two very recent reports described methods of 5fC quantification using chemical modification-assisted bisulphite sequencing, fCAB-seq (5fC chemically assisted bisulphite sequencing) and redBS-seq (reduced bisulphite sequencing) (FIG. 1), for whole-genome profiling at single-base resolution[93,102] (FIG. 3). In fCAB-seq, 5fC bases are chemically modified before BS-seq

Click chemistry
A nonspecific chemical reaction that combines small modular units and is used to generate or label a substance. Azide alkyne Huisgen cycloaddition, in which an azide and alkyne interact to form a 1,2,3-triazole (with 5-membered ring) is the most popular click chemistry reaction. Click chemistry has been used for selectively labelling biomolecules.

Glucosylation
The process of transferring a glucose residue from a nucleotide sugar derivative, such as from uridine diphosphate glucose (UDP-Glc) to a target molecule.

J-binding proteins
Proteins that specifically bind to the base J (β-D-glucopyranosyloxymethyl uracil), a modified form of uracil found in the DNA of a number of organisms, such as human pathogen Trypanosoma and the kinetoplastids. Base J is formed by hydroxylation of thymidine and subsequent glycosylation by glycosyltransferase enzyme.

Isoschizomer
Restriction enzymes that have the same recognition sequences and cleave at the same positions.

through the use of *O*-ethylhydroxylamine ($EtONH_2$). The modified 5fC is protected from bisulphite treatment and remains unchanged and therefore appears similar to the original 5mC and 5hmC in sequencing, while 5caC and cytosine are replaced with thymidine (FIG. 3). The quantification of 5fC is determined by subtraction of 5hmC+5mC signals from a typical readout of BS-seq on the same sample. In redBS-seq, $NaBH_4$ is used to reduce 5fC to 5hmC, followed by BS-seq (FIG. 3). Similar to previous methods, the 5fC level is obtained by subtraction of 5hmC+5mC signals from BS-seq on the same sample.

In comparison to whole-genome 5mC detection methods, the sequencing efforts required for these techniques are several-fold greater because deep sequencing is required to enable detection of variants in extremely low abundances. Particularly for oxBS-seq, CAB-seq, fCAB-seq and redBS-seq the amount of sequencing efforts is doubled, as all of these methods require subtraction of BS-seq signals or subtraction from BS-seq signals. To reduce sequencing effort, the converted DNA can be digested with *Msp*I to select for CpG-rich fragments before BS-seq library preparation[98,102].

Only a few groups have been able to apply these methods. Due to the limited data available, it is still unclear whether robust and consistent results can be achieved from the enzymatic and chemical conversions. High-quality spike-in controls must be used for these experiments to estimate both the rate of failed conversion and the rate of inappropriate conversion, which can both vary due to technical and biological variation. These methods are still not applicable to rare cell types because they require microgram levels of input DNA.

Direct sequencing of unamplified DNA. Developments in third-generation sequencing technologies have enabled the detection of DNA modifications directly without any chemical or enzymatic reactions. Pacific Biosciences has developed SMRT sequencing technology that enables direct detection of 5mC and 5mC oxidation derivatives based on nucleotide incorporation time or the kinetic signature[103–105]. A shortcoming of SMRT sequencing is that the kinetic signature differences between variants can be too close, leading to poor quantification accuracy. However, chemical labelling or enzymatic conversion can improve signal detection in some cases. Conversion of 5hmC to N_3-5gmC or to HS-N_3-5gmC using previously described methods prior to SMRT sequencing can increase quantification accuracy because the kinetic signal differences between the converted bases and 5mC are larger than the kinetic signal differences between the converted bases between 5hmC and 5mC[106]. In addition, the weak kinetic signal from 5mC can also be improved as 5caC has been found to have much stronger kinetic signals than 5mC, and oxidation of 5mC to 5caC can enhance 5mC signals for detection in bacterial or archeal genomes[105]. With an average read length of 8.5 kb, this technology enables the detection of 5mC and its oxidation derivatives on individual methylation haplotypes and in repetitive regions.

Nanopore is another emerging technology for single-molecule DNA sequencing[107–111]. Both solid-state and protein nanopores have been used to characterize 5hmC[112,113]. Similar to SMRT, nanopore technology has the advantages of long read length, rapid sequencing and no chemical modification. However, the error in base identification is still high for SMRT sequencing (18–20%)[114], and the error rate for nanopore sequencing has not yet been fully characterized. The throughput of third-generation sequencing remains to be improved by orders of magnitude before routine analysis of mammalian size genomes becomes feasible.

Conclusion

Three methods have been broadly adopted by the scientific community for 5mC quantification. Illumina Infinium 450K BeadChips have been widely applied to rapid profiling of samples in large cohorts, and will remain as the most accessible method. Further improvements to the Illumina 450K BeadChips technology will probably involve the development of newer versions of BeadChip with expanded contents. WGBS and RRBS have been used to map 5mC in many cell and tissue types in major consortium projects like ENCODE, and by individual research groups that have expertise in sequencing-based assays and related bioinformatics analyses. Improvements to WGBS and RRBS in recent years towards lower sample input have enabled many groups to map rarer cell types; RRBS has been shown to be effective in single cells. However, due to the inherent bias towards CpG dense regions, RRBS might have limited applicability in detecting methylation changes in CpG sparse regions. Targeted bisulphite sequencing methods such as hybridization capture, BSPP and microfluidic PCR are highly complementary to RRBS in that they provide the flexibility to analyse arbitrary subsets of the genome at a cost that is orders of magnitude lower than that of WGBS. With the availability of commercial products, the barrier of entry for average users has been greatly reduced. Liquid capture methods can be implemented on exome–sequencing pipelines available at many core facilities or major commercial providers. They will become strong competitors with Illumina BeadChips and RRBS. BSPP and microfluidic PCR are likely to be appealing options for many focused studies, in which tens to thousands of regions of interest in tens to thousands of samples need to be analysed quickly at minimal costs.

The discovery of other forms of cytosine modifications, such as 5hmC, 5fC and 5caC, in mammalian genomes has stimulated many efforts to quantify these modified bases in different cell types, particularly in brain tissues. The challenge for these efforts is the rarity of 5mC oxidation derivatives in mammalian genomes. Antibody-based and specific chemical or enzymatic modifications were used to enrich for and determine relative abundance in the genome with limited resolution. Application of BS-seq on chemically or enzymatically modified DNA has allowed for single-base resolution quantification of 5hmC, 5caC and 5fC. It remains unclear whether these assays will be as widely applied to large-scale projects as the 5mC

assays. This is partly due to the assay cost and technical knowledge involved. More importantly, it is still uncertain what biological insights we will gain by profiling these lower abundant marks in large numbers of human specimens.

SMRT and nanopore sequencing both allow for direct reading of 5mC and other DNA modifications on single molecules without amplification or DNA treatment. However, the full potential of these third-generation methods remains to be materialized. The throughput and accuracy of these methods needs to be substantially improved before they become serious contenders to second-generation-based methods.

In conclusion, we have recently witnessed major improvements and refinements in many DNA methylation assays, as well as the development of newer techniques for characterizing additional cytosine modifications. Commercialization of many of these methods has also attracted investigators from other fields, and has contributed to the rapid growth of epigenetics research in general. The current pace of advancement needs to continue in order to develop affordable and accurate assays to quantify 5mC and its oxidation derivatives, especially at the most phenotypically relevant sites, with the eventual goal of bringing these assays to routine use in clinical diagnostics.

1. Arand, J. *et al. In vivo* control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8**, e1002750 (2012).
2. Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genet.* **45**, 1198–1206 (2013).
3. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
4. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
5. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
6. Ziller, M. J. *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7**, e1002389 (2011).
7. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature Genet.* **46**, 17–23 (2014).
8. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
9. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
10. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
11. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* **13**, 484–492 (2012).
12. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14**, 204–220 (2013).
13. Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature Genet.* **45**, 836–841 (2013).
14. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature Protoc.* **7**, 2159–2170 (2012).
15. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
16. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013).
17. Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neurosci.* **14**, 1607–1616 (2011).
18. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
19. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
20. Laird, P. W., Principles and challenges of genome-wide DNA methylation analysis. *Nature Rev. Genet.* **11**, 191–203 (2010).
21. Shen, L. & Zhang, Y. 5-hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr. Opin. Cell Biol.* **25**, 289–296 (2013).
22. Song, C.-X., Yi, C. & He, C. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nature Biotech.* **30**, 1107–1116 (2012).
23. Bock, C. Analysing and interpreting DNA methylation data. *Nature Rev. Genet.* **13**, 705–719 (2012).
24. Krueger, F. *et al.* DNA methylome analysis using short bisulfite sequencing data. *Nature Methods* **9**, 145–151 (2012).
25. Grunau, C., Clark, S. J. & Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **29**, e65 (2001).
26. Ogino, S. *et al.* Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J. Mol. Diagn.* **8**, 209–217 (2006).
27. Nestor, C. E. *et al.* Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* **22**, 467–477 (2012).
28. Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* **13**, R92 (2012).
29. Wang, J. *et al.* Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. *BMC Genomics* **14**, 11 (2013).
30. Schillebeeckx, M. *et al.* Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced adrenocortical neoplasia in the mouse. *Nucleic Acids Res.* **41**, e116 (2013).
31. Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).
32. Ruiz, S. *et al.* Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **109**, 16196–16201 (2012).
33. Okuizumi, H. *et al.* Restriction landmark genome scanning. *Methods Mol. Biol.* **791**, 101–112 (2011).
34. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
35. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nature Protoc.* **7**, 617–636 (2012).
36. Clark, C. *et al.* A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the infinium HumanMethylation450 BeadChip® for methylome profiling. *PLoS ONE* **7**, e50233 (2012).
37. Aberg, K. A. *et al.* MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case–control samples. *Epigenomics* **4**, 605–621 (2012).
38. Lan, X. *et al.* High resolution detection and analysis of CpG dinucleotides methylation using MBD-seq technology. *PLoS ONE* **6**, e22226 (2011).
39. Brinkman, A. B. *et al.* Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* **52**, 232–236 (2010).
40. Butcher, L. M. & Beck, S. AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay. *Methods* **52**, 223–231 (2010).
41. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotech.* **28**, 1106–1114 (2010).
42. Matarese, F., Carrillo-de Santa Pau, E. & Stunnenberg, H. G. 5-hydroxymethylcytosine: a new kid on the epigenetic block? *Mol. Syst. Biol.* **7**, 562 (2011).
43. Komori, H. K. *et al.* Application of microdroplet PCR for large-scale targeted bisulfite sequencing. *Genome Res.* **21**, 1738–1745 (2011).
This paper demonstrates a fully automated method for quantification of DNA methylation on 2100 genes.
44. Nautiyal, S. *et al.* High-throughput method for analyzing methylation of CpGs in targeted genomic regions. *Proc. Natl Acad. Sci. USA* **107**, 12587–12592 (2010).
45. Varley, K. E. & Mitra, R. D. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Res.* **20**, 1279–1287 (2010).
46. Diep, D. *et al.* Library-free methylation sequencing with bisulfite padlock probes. *Nature Methods* **9**, 270–272 (2012).
This paper describes a method for high-throughput padlock probes that sequence methylated DNA.
47. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotech.* **27**, 353–360 (2009).
48. Lee, E.-J. *et al.* Targeted bisulfite sequencing by solution hybrid selection and massively parallel sequencing. *Nucleic Acids Res.* **39**, e127 (2011).
49. Wang, J. *et al.* High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing. *BMC Genomics* **12**, 597 (2011).
50. Ivanov, M. *et al.* In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. *Nucleic Acids Res.* **41**, e72 (2013).
51. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
52. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
53. Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439 (2013).
54. Kobayashi, H. & Kono, T. DNA methylation analysis of germ cells by using bisulfite-based sequencing methods. *Methods Mol. Biol.* **825**, 223–235 (2012).
55. Adey, A. & Shendure, J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res.* **22**, 1139–1143 (2012).
This study shows that whole-genome bisulphite sequencing can be performed on ~1–10 ng of genomic DNA.
56. Wang, Q. *et al.* Tagmentation-based whole-genome bisulfite sequencing. *Nature Protoc.* **8**, 2022–2032 (2013).
57. Miura, F. *et al.* Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **40**, e136 (2012).
58. Kantlehner, M. *et al.* A high-throughput DNA methylation analysis of a single cell. *Nucleic Acids Res.* **39**, e44 (2011).
59. Lorthongpanich, C. *et al.* Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* **341**, 1110–1112 (2013).
The first paper to use single-cell DNA methylation analysis to address an important biological problem.

60. El Hajj, N. *et al.* Limiting dilution bisulfite (pyro) sequencing reveals parent-specific methylation patterns in single early mouse embryos and bovine oocytes. *Epigenetics* **6**, 1176–1188 (2011).
61. Guo, H. *et al.* The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
62. Smallwood, S.A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11**, 817–820 (2014).
63. Irizarry, R. A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* **18**, 780–790 (2008).
64. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
65. Yalcin, A. *et al.* MeDIP coupled with a promoter tiling array as a platform to investigate global DNA methylation patterns in AML cells. *Leukemia Res.* **37**, 102–111 (2013).
66. Gilson, E. & Horard, B. Comprehensive DNA methylation profiling of human repetitive DNA elements using an MeDIP-on-RepArray assay. *Methods Mol. Biol.* **859**, 267–291 (2012).
67. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
68. Oliver, V. F. *et al.* A novel methyl-binding domain protein enrichment method for identifying genome-wide tissue-specific DNA methylation from nanogram DNA samples. *Epigenetics Chromatin* **6**, 1–11 (2013).
69. Dumenil, T. D. *et al.* Genome-wide DNA methylation analysis of formalin-fixed paraffin embedded colorectal cancer tissue. *Genes Chromosomes Cancer* **53**, 537–548 (2014).
70. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
Together with reference 69, these papers describe DNA precipitation followed by bisulphite treatment to create a map of DNA methylation patterns associated with chromatin modifications.
71. Statham, A. L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* **22**, 1120–1127 (2012).
72. Li, Y. & Tollefsbol, T. O. Combined chromatin immunoprecipitation and bisulfite methylation sequencing analysis. *Methods Mol. Biol.* **791**, 239–251 (2011).
73. Cipriany, B. R. *et al.* Single molecule epigenetic analysis in a nanofluidic channel. *Anal. Chem.* **82**, 2480–2487 (2010).
74. Cipriany, B. R. *et al.* Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. *Proc. Natl Acad. Sci. USA* **109**, 8477–8482 (2012).
This study discusses a nanofluidic device for sorting single methylated DNA molecules.
75. Murphy, P. J. *et al.* Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. *Proc. Natl Acad. Sci. USA* **110**, 7772–7777 (2013).
76. You, J. S. *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proc. Natl Acad. Sci. USA* **108**, 14497–14502 (2011).
77. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
This study describes NOMe-seq, a method using *M. Cvi*PI treatment and bisulphite conversion to produce a genome-wide base resolution map of nucleosome positioning and DNA methylation on the same DNA molecules.
78. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
79. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
80. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
81. Stroud, H. *et al.* 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
82. Thomson, J. P. *et al.* Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res.* **41**, e206 (2013).
83. Inoue, A. *et al.* Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676 (2011).
84. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
85. Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122–126 (2013).
86. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).
87. Szwagierczak, A. *et al.* Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
88. Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotech.* **29**, 68–72 (2011).
89. Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl Acad. Sci. USA* **104**, 16793–16797 (2007).
90. Pastor, W. A. *et al.* The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nature Protocols* **7**, 1909–1917 (2012).
91. Robertson, A. B. *et al.* Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nature Protocols* **7**, 340–350 (2012).
92. Michaeli, Y. *et al.* Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem. Commun. (Camb.)* **49**, 8599–8601 (2013).
93. Song, C.-X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
This study describes the first genome-wide base resolution map of 5fC in mESCs.
94. Hu, J. *et al.* Selective chemical labelling of 5-formylcytosine in DNA by fluorescent dyes. *Chemistry* **19**, 5836–5840 (2013).
95. Kinney, S. M. *et al.* Tissue-specific distribution and dynamic changes of 5-Hydroxymethylcytosine in mammalian genomes. *J. Biol. Chem.* **286**, 24685–24693 (2011).
96. Sun, Z. *et al.* High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep.* **3**, 567–576 (2013).
97. Shankaranarayanan, P. *et al.* Single-tube linear DNA amplification (LinDA) for robust ChIP–seq. *Nature Methods* **8**, 565–567 (2011).
98. Booth, M. J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).
99. Booth, M. J. *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protoc.* **8**, 1841–1851 (2013).
100. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
101. 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).
102. Booth, M. J. *et al.* Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nature Chem.* **6**, 435–440 (2014).
Together with reference 98 this paper describes a method for producing a whole-genome base resolution map of 5hmC (in mice and humans).
103. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
104. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**, 461–465 (2010).
Together with reference 96 this paper describes a method for producing a whole-genome base resolution map of 5hmC.
105. Clark, T. *et al.* Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* **11**, 4 (2013).
106. Song, C.-X. *et al.* Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nature Methods* **9**, 75–77 (2012).
107. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nature Biotech.* **26**, 1146–1153 (2008).
108. Purnell, R. & Schmidt, J. Measurements of DNA immobilized in the alpha-hemolysin nanopore. *Methods Mol. Biol.* **870**, 39–53 (2012).
109. Stoddart, D. *et al.* Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore.*Proc. Natl Acad. Sci. USA* **106**, 7702–7707 (2009).
110. Butler, T. Z. *et al.* Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl Acad. Sci. USA* **105**, 20647–20652 (2008).
111. Manrao, E. A. *et al.* Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS ONE* **6**, e25723 (2011).
112. Wanunu, M. *et al.* Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.* **133**, 486–492 (2011).
113. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl Acad. Sci. USA* **110**, 18904–18909 (2013).
114. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biol.* **14**, R51 (2013).

Acknowledgements
The authors gratefully acknowledge members of the Zhang laboratory for their proofreading of the manuscript. We apologize to those authors whose works were not covered here due to space constraints. This work is funded by the US National Institutes of Health grants R01GM097253 and R01AG042187. D.H.D. is supported by a UCSD-CIRM pre-doctoral fellowship.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
EU BLUEPRINT Epigenome Programme: http://www.blueprint-epigenome.eu/
NIH Eipgenomics Roadmap Programme: http://www.roadmapepigenomics.org/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF