

Quantitation and mapping of the epigenetic marker 5-hydroxymethylcytosine

Ying Qing^{1,†}, Zhiqi Tian^{1,2,†}, Ying Bi³, Yongyao Wang^{1,2}, Jiangang Long², Chun-Xiao

Song^{3,4,*}, Jiajie Diao^{1,*}

¹Department of Cancer Biology, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA. ²Center for Mitochondrial Biology and Medicine, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology Xi'an Jiaotong University, Xi'an 710049, China. ³Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁴Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

*Correspondence should be addressed to C.S. (email: chunxiao.song@ludwig.ox.ac.uk) or J.D. (email: jiajie.diao@uc.edu).

†These authors contributed equally to this work.

Summary: We here review primary methods used in quantifying and mapping 5-hydroxymethylcytosine (5hmC), including global quantification, restriction enzyme-based detection, and methods involving DNA-enrichment strategies and the genome-wide sequencing of 5hmC. As discovered in the mammalian genome in 2009, 5hmC, oxidized from 5-methylcytosine (5mC) by ten-eleven translocation (TET) dioxygenases, is increasingly being recognized as a biomarker in biological processes from development to pathogenesis, as its various detection methods have shown. We focus in particular on an ultrasensitive single-molecule imaging technique that can detect and quantify 5hmC from trace samples and thus offer information regarding the distance-based relationship between 5hmC and 5mC when used in combination with fluorescence resonance energy transfer.

Keywords: Epigenetic modification, single molecule imaging

Abbreviations: FRET: Förster Resonance Energy Transfer or Fluorescence Resonance Energy Transfer; 5hmC: 5-Hydroxymethylcytosine; 5mC: 5-methylcytosine; TET: ten-eleven translocation (TET)

Introduction

Epigenetic mechanisms play a pivotal role in many physiological and pathological processes, including tumorigenesis [1,2]. Among major DNA modifications in the mammalian genome along with 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC) was recently recognized as the sixth base and could play multiple roles in transcriptional regulation other than mediating the DNA demethylation pathway [3,4].

Discovered in mammalian genomic DNA in 2009, 5hmC is produced from 5mC via a hydroxylation reaction catalyzed by ten-eleven translocation (TET) dioxygenases [5,6]. 5hmC can be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TET, both of which will be removed and replaced with unmodified cytosine by base excision repair to complete the DNA demethylation pathway. Highly tissue-specific and stable, 5hmC occurs at substantial levels in specific tissues, most notably in the brain [7-9]. Recent studies have established that in addition to mediating an oxidative demethylation process, 5hmC is an important epigenetic marker that could regulate chromatin structure and transcription, as well as playing pivotal roles in tumorigenesis, embryogenesis, the nervous system, and various other pathological conditions [4,10-12]. The loss or low levels of 5hmC have been observed in various solid tumors such as those in lung [13], kidney [14], and colorectal cancer [15], as well as melanomas [16,17] and brain tumors [18]. Moreover, the balance between 5hmC and 5mC in distinct genic contexts, including those of gene promoters and gene bodies, is likely critical to transcriptional fine-tuning [4]. Accordingly, 5hmC, like 5mC, is increasingly recognized as a biomarker in disease diagnostics [19].

Knowledge of the distribution of 5hmC and its interaction with 5mC is necessary in

order to fully elucidate how 5hmC acts in its biological role. Many techniques have been used to detect 5hmC, first by global quantification by Tahiliani *et al.* [5] and Kriaucionis *et al.* [6]. In those studies, DNA was isolated from different types of cells, enzymatically digested to a single nucleoside, and analyzed via one- or two-dimensional thin-layer chromatography (TLC) assay to identify 5hmC as a new component in mammalian genomic DNA. As an alternative method, T4 bacteriophage glucosyltransferase (β -GT) can load a glucose moiety from uridine diphosphoglucose (UDP-Glc) onto the hydroxyl group of the 5hmC base [20]. With the help of that enzyme, Szwagierczak *et al.* [21] conducted the quantitative detection of 5hmC with a liquid scintillation analyzer. Song *et al.* [22] further developed the method by employing the chemically modified UDP-Glc to attach an azide-modified glucose onto the 5hmC base, which allowed them to attach a biotin to the azide group and capture the 5hmC-containing DNA using streptavidin-coated magnetic beads for sequencing analysis. Among other techniques, base-resolution sequencing for 5hmC has also been developed [23,24].

However, all of the aforementioned techniques have failed to identify the interaction between 5hmC and 5mC in the same DNA molecule. However, that information is critical, especially for studying the physiological and pathological roles of 5hmC, given three observations. First, the intragenic gene body ratio of 5hmC to 5mC is a better predictor for gene expression than either marker on its own [25]. Second, the amount of 5hmC varies significantly in different cell types [7]. Third, 5hmC is oxidized from 5mC [5], which implies that the interplay between 5hmC and 5mC can be crucial in gene regulation [5,6].

Using β -GT and a selective chemical labeling strategy, specific fluorophores can be

attached to 5hmC and 5mC residues simultaneously. Optical imaging and analysis of fluorescent-labeled 5hmC and 5mC with an ultrasensitive single-molecule microscope provides a promising method of reveal 5hmC and 5mC sites simultaneously in the same DNA molecule [26]. Here, we reviewed the primary methods for detecting 5hmC, with particular focus on the novel ultrasensitive single-molecule imaging technique, which can offer information regarding the distance-based relationship and interplay between 5hmC and 5mC.

Global Quantification of 5hmC

Qualitative and quantitative detection of 5hmC by TLC or high-pressure liquid chromatography with mass spectrometry

Tahiliani *et al.* [5] detected 5hmC as a new component in mammalian DNA while studying the function of TET proteins, the mammalian homologs of the trypanosome proteins JBP1 and JBP2 [27,28] that have been proposed to oxidize the 5-methyl group of thymine. By using one-dimensional TLC, they found that TET1 can catalyze the conversion of 5mC to another unidentified species in cells overexpressing TET1 and used mass spectrometry (MS) to identify the novel nucleotide as 5hmC. To further confirm that 5hmC is a physiological constituent in the mammalian genome, they also verified the existence and performed the quantification of 5hmC in mouse embryonic stem cells by using TLC assay. Kriaucionis *et al.* [6] employed a similar two-dimensional TLC assay to discover a novel nucleotide in the genomic DNA isolated from Purkinje neurons and, using high-pressure liquid chromatography (HPLC) and MS analyses, further confirmed the newfound nucleotide to be 5hmC. Those groundbreaking findings offered a new avenue for studying DNA

demethylation pathways; however, due to the limited sensitivity of the techniques, quantification accuracy has remained poor.

HPLC combined with MS (HPLC-MS) is more effective than TLC, and can simultaneously detect multiple components, including 5hmC [29]. Furthermore, by using synthetic stable isotope-labeled reference compounds, the technique can be converted into a quantitative method. Globisch *et al.* [7] and Munzel *et al.* [30] applied that very technique to investigate the distribution of 5hmC in different tissue types and found that the amount of 5hmC was high in tissues of the central nervous system, average in heart and kidney tissues, and low in liver and testis tissues. Although further liquid chromatography–MS- or MS-based quantitative detection of 5hmC has been performed by several research teams [15,31-36], HPLC-MS remains by far the most accurate method for global 5hmC quantification, given its greater sensitivity and selectivity in quantifying the total mass of 5hmC. However, the method requires expensive equipment, considerable expert knowledge, and ample DNA, and is limited by its inability to provide sequence information.

Detection of global 5hmC by specific antibodies

At present, several commercially available 5hmC-specific antibodies can be used for the immune-detection of global 5hmC. For example, when used in the classical dot-blot technique, such antibodies provide a rapid, inexpensive approach for performing the semi-quantification of 5hmC in tissues or cells [8,37]. Among its advantages, immunohistochemistry and immunofluorescence with 5hmC-specific antibodies allows the visualization of the distribution of 5hmC in tissues [7,37-39]. Although the techniques are easy to conduct, their quantitative accuracy is relatively poor and relies heavily on the

specificity and sensitivity of the antibody involved. To partly solve those problems, an alternative approach treats the DNA with bisulfite to convert 5hmC into cytosine 5-methylenesulfonate (CMS), which can be used to cultivate an antibody that allows the indirect detection and quantification of 5hmC [38,40,41]. In comparison with anti-5hmC immune detection, that method is more specific and produces less background, as well as being less dependent on 5hmC density. than.

Labeling 5hmC with glucose by β -GT

β -GT is a glucosyltransferase from T-4 bacteriophages capable of attaching a glucose onto a 5hmC base in DNA. Szwagierczak *et al.* used that same method to load a radioactive labeled UDP-[^3H] glucose onto 5hmC and quantify its amount in different mouse tissues with a liquid scintillation analyzer [21]. Among its benefits, β -GT does not have sequence or density bias, which provides an advantage over antibodies.

Labeling 5hmC with fluorescent reporters

A recent study has described a spectroscopic method for the quantitative detection of global 5hmC in genomic DNA [42] that uses β -GT to conduct the enzymatic glucosylation of 5hmC to facilitate a subsequent step of glucose oxidation. With the subsequent formation of aldehyde moieties, an oxime ligation covalently links the moieties to a fluorescent reporter. The level of 5hmC in the labeled sample is in direct proportion to the fluorescence intensity, and can thus be measured.

Detection of 5hmC in specific DNA sequences

To determine the presence of 5hmC in certain restriction enzyme recognition sites, a restriction enzyme-based method known as combined glycosylation restriction analysis

(CGRA) can distinguish 5hmC and 5mC [43-47]. Restriction enzymes such as *MspI* and *GlaI* cleave only sites containing 5hmC or 5mC, not those with glucosylated 5hmC. Because CGRA analysis uses β -GT to attach a glucose to 5hmC in order to convert it into glucosylated 5hmC, 5hmC becomes protected from cleavage by glucosylation. Afterward, qPCR can be used to determine the amount of DNA templates cut by restriction enzymes before and after treatment with β -GT, thereby allowing the estimation of approximate absolute 5hmC levels in the target sequence by comparing copy numbers in both samples. Such a technique can furnish information on locus-specific hydroxymethylation, yet can only be applied to specific DNA sequences containing certain restriction enzyme recognition sites.

Methods Involving DNA-Enrichment Strategies

Methods involving DNA enrichment and isolation strategies often employ sequencing techniques to detect 5hmC after the enrichment and isolation of DNA with 5hmC. To date, several enrichment techniques for DNA with 5hmC have been applied, including enrichment by 5hmC-specific antibodies, by chemical labeling, and by restriction enzymes.

Enrichment of DNA with 5hmC by specific antibodies

Hydroxymethylated DNA immunoprecipitation (hMeDIP) is a method that uses a 5hmC-specific antibody to form a DNA-antibody complex, thereby allowing DNA with 5hmC to be collected and isolated for downstream analysis by, for example, high-throughput sequencing [8,48-50]. Comparing the immunoprecipitated DNA to an untreated control sample by qPCR or high-throughput sequencing can achieve the relative enrichment of a specific locus in the tested DNA. Such a method has been used to study the distribution of

5hmC in the genomes of mice, as well as human embryonic stem cells and neural progenitor cells [48,49]. Similarly, 5hmC can be modified to CMS by treatment with bisulfite and immunoprecipitated by a CMS-specific antibody for subsequent analysis, thereby offering more sensitive and specific enrichment [51,52]. However, the disadvantages of hMeDIP include its dependence on 5hmC density, its inability to give single-base resolution, and its off-target affinity of the antibody for simple repeats and CA dinucleotides [51,53-55].

Enrichment of DNA with 5hmC with the glucosylation of 5hmC

As mentioned earlier, T4 bacteriophage β -GT can be used to load a glucose moiety from UDP-Glc onto 5hmC. By extension, the hydroxymethyl-selective chemical labeling (hMeSeal) method was skillfully developed by Song *et al.*, who transferred the engineered 6-N₃-glucose onto the hydroxyl group of 5hmC to bio-orthogonally label it in genomic DNA [22]. The presence of the azide group allows a biotin or any other tag to be attached by Huisgen cycloaddition (i.e., click) chemistry [56,57]. They used the biotin tag to capture DNA with 5hmC for sensitive detection, affinity enrichment, and sequencing in order to reveal the genomic locations of 5hmC. The major advantage of the method is that it represents a more efficient way to capture or enrich DNA with 5hmC, because covalent labeling along with biotin-based affinity purification offers far more selective and comprehensive capture. Furthermore, other tags such as specific fluorophores can be readily installed to achieve the optical imaging of 5hmC. More recently, the method was modified by using an engineered Tn5 transposase-based library construction strategy to incorporate sequencing adaptors to the DNA fragments via a single transposase-catalyzed DNA tagmentation step [58]. That modified approach, termed nano-hmC-Seal, allows the profiling

of 5hmC in genomic DNA isolated from only about 1,000 cells.

From *Trypanosoma*, the J-binding protein 1 (JBP1) can specifically recognize and bind to 5hmC once attached with a glucose by β -GT [59]. Robertson *et al.* developed a method to modify JBP1 to affix it to magnetic beads and subsequently pulled down 5hmC-containing DNA by JBP1-coated magnetic beads for downstream applications, including qPCR and DNA sequencing [60]. The method provides a more straightforward way to enrich DNA with 5hmC; however, its enrichment efficiency is less than that of other glucosylation-based methods. Cui *et al.* improved the method by using a recombinant JBP1 protein to achieve greater efficiency [61]. Nevertheless, the method provides only a semi-quantitative detection of 5hmC at a resolution of about 50 bp. As a comparative analysis of three affinity-based 5hmC enrichment methods indicated, 5hmC profiles produced by JBP1 5hmC pull-down show poor overlap with hydroxymethyl-DNA immunoprecipitation (hMeDIP) and hMeSeal, meaning that JBP1 5hmC pull-down is assumably a poor reporter of 5hmC profiles [62].

A combined enzymatic–chemical technique, termed glucosylation, periodate oxidation, and biotinylation, has also been applied in detecting 5hmC [51,63,64]. It relies on β -GT to add glucose to 5hmC, after which glycols in the glucose are oxidized by sodium periodate to produce aldehydes, thereby allowing the biotins to attach to them in aqueous solution under mild reaction conditions. Ultimately, biotinylated DNA fragments with 5hmC are precipitated with streptavidin-conjugated beads and eluted for subsequent high-throughput sequencing. The background precipitation of DNA produced by that technique is greater than that of other precipitation methods such as the anti-CMS method.

Restriction Enzyme-Based Sequencing of 5hmC

Currently, several methods of detecting 5hmC incorporate restriction enzymes. Reduced representation 5-hydroxymethylcytosine profiling (RRHP), an approach developed by Petterson *et al.*, combines enzymatic digestion with high-throughput sequencing by using β -GT [65]. Genomic DNA is first digested by restriction enzyme *MspI*, after which the ends of digested fragments are ligated with adapters, one of which reconstitutes the *MspI* site. The adapter-ligated fragments are thereafter glucosylated with β -GT so that fragments with glucosylated 5hmC at the junction resist a second *MspI* digestion, while fragments presenting unmodified C or 5mC cleave. Only fragments with adapters at both ends after the second *MspI* digestion are amplified and sequenced. That assay can be used when the 5hmC density is low, and it can achieve single-base resolution of 5hmC positions.

Other restriction enzyme-based methods include DNA-modification-dependent restriction endonuclease *AbaS* I coupled with sequencing (Aba-seq) and Pvu-Seal-seq [66,67]. *AbaS* I recognizes glucosylated 5hmC specifically and cleaves at 11-13 bp downstream of the recognition site. Sun *et al.* exploited *AbaS* I to digest genomic DNA, and ligated the digested DNA with biotinylated P1 adapters for capture by streptavidin beads, followed by the ligation of other P2 adapters to the captured DNA for library construction [66]. 5hmC profiles are determined by the high-throughput sequencing of the library. Mooijman *et al.* further developed Aba-seq by including a cell-specific barcode in the adapters ligated to the *AbaS* I digested DNA to realize single-cell 5hmC sequencing [68]. However, the cleavage efficiency of *AbaS* I depends on a symmetrical pattern of recognition sites; cleavage efficiency

decreases only when one of two cytosines around the cleavage site is glucosylated 5hmC. *PvuRtsI* I, which belongs to the same family as *AbaS* I, can specifically recognize 5hmC and does not require a symmetrical pattern of the glucosylated cytosines, meaning that it can theoretically detect every 5hmC in the genome, despite its poor specificity compared with *AbaS* I. To overcome that limitation, Sun *et al.* combined a Seal-based enrichment technique involving 5hmC with *PvuRtsI* I to develop a new method dubbed Pvu-Seal-seq [67]. In that modified Pvu-Seal-seq method, DNA was first digested by *PvuRtsI* I, followed by the attachment of a biotin to 5hmC and the enrichment of the DNA fragments using magnetic beads. The enriched DNA fragments were then analyzed by high-throughput sequencing to detect 5hmC.

Whole Genome Sequencing

The genome-wide base-resolution sequencing of 5hmC is primarily realized by two strategies: one based on the whole-genome bisulfite sequencing technique, the other using the third-generation sequencing technique. The conventional bisulfite sequencing technique fails to distinguish 5hmC and 5mC, which both read as cytosine (C) when sequenced after bisulfite treatment, whereas C and 5caC bases are observed as thymine (T). TET-assisted bisulfite sequencing (TAB-seq), a modified genome-wide bisulfite sequencing developed by Yu *et al.*, involves the glucosylation of 5hmC by β -GT to protect it from oxidation by TET1, followed by the oxidation of 5mC to 5-carboxylcytosine (5caC) by excessive TET1 [23,69]. In the subsequent bisulfite treatment, both unmodified cytosine and 5caC are converted to uracil and read as T, leaving glucosylated 5hmC unaffected and hence read as cytosine C.

Therefore, the whole-genome mapping of 5hmC can be achieved by TAB-seq. Another bisulfite sequencing-based method known as oxidative bisulfite sequencing (oxBS-seq) adopts a similar idea for base-resolution mapping of 5hmC [24,70]. In that method, 5hmC is selectively oxidized to 5fC by potassium perruthenate (K₂ReO₇) and subsequently converted to uracil by bisulfite treatment and observed as T. Meanwhile, 5mC remains the same and is read as C, which makes it possible to discriminate 5hmC from 5mC. Therefore, 5hmC levels can be deduced by the subtraction of 5mC signals measured by oxBS-seq from 5hmC+5mC signals measured by conventional bisulfite sequencing.

The invention and application of the third-generation sequencing technique makes it possible to detect DNA modification directly without any chemical or enzymatic modifications [71-74]. In single-molecule real-time (SMRT) sequencing, four distinguishable nucleotides labeled with fluorophores are incorporated into new complementary DNA strands by DNA polymerases. By observing and analyzing the arrival time and duration of the corresponding fluorescence signals, information concerning polymerase dynamics can be obtained, which allows the direct detection of various DNA modifications, including 5hmC. However, a major drawback of SMRT sequencing is the dismal accuracy of its sequencing data, which translates to an error rate of approximately 15% [75]. Another new single-molecule detection technique is nanopore sequencing [72,73]. By electrophoretically driving single strands of DNA through a nanopore, which invariably has a highly confined space, the single-strand DNA can be detected and analyzed in single-nucleotide steps, which can distinguish cytosine, 5mC, and 5hmC. However, the error rate of nanopore sequencing remains unclear, hence preventing its wide application.

Single-Molecule Optical Imaging Technique

Optical mapping techniques selectively label epigenetic modifications on DNA molecules with fluorescent reporter molecules and can offer epigenetic information of an individual DNA molecule of up to 1 Mbp long [76]. By extending the DNA molecules in a linear configuration, epigenetic information patterns along the DNA molecules can be directly visualized with the help of a fluorescence microscope. Michaeli *et al.* reported a method using that optical mapping technique to specifically label 5hmC with a fluorophore along the genomic DNA molecule [77]. They used β -GT to load a glucose modified with an azide group onto 5hmC, after which the azide covalently bound to the 5hmC site further reacted with a fluorescently labeled alkyne via a click chemistry, Huisgen cycloaddition reaction to allow a fluorophore to attach to the 5hmC site. Using fluorescence imaging, the single-molecule detection and quantification of global 5hmC in genomic DNA is thus possible.

Other than the genome-wide mapping of 5hmC, another piece of information critical for the study of 5hmC is its interplay with 5mC. However, none of the aforementioned methods can provide such information. In response, we have developed a versatile ultrasensitive single-molecule imaging technique that we used to study the colocalization of 5hmC and 5mC in mouse genomic DNA [26]. With the combination of single-molecule fluorescence resonance energy transfer (smFRET), the technique can be used to measure the proximity of 5hmC and 5mC in the same DNA molecule. It is a highly modular technique, suitable for imaging both single and multiple modifications. Given its ultrahigh sensitivity, it

requires ultralow input (50 pg or less DNA), and can be applied to small and sensitive samples.

Technological Process of the Method

To perform single-molecule fluorescent assays, the first step is to label DNA epigenetic modifications with specific fluorophores (Fig. 1). Genomic DNA sample was first extracted and enzymatically digested to 50-200 bp fragments, whose average length was analyzed by agarose gel electrophoresis. Terminal transferase (TdT) was used along with modified dCTP to end-label the DNA fragments with biotin and Cy3. Once the end-labeled DNA was purified and eluted in H₂O, DNA molecules labeled with biotin could be captured on the microscope slides and their total amount counted through the Cy3 tag. To perform a genome-wide count of 5hmC, the 5hmC base in the biotin and Cy3 end-labeled DNA were tagged with an azide-modified glucose by β -GT. Via the Huisgen cycloaddition reaction, 5hmC was ultimately labeled with Cy5, after which a passivated quartz slide coated with polyethylene glycol to eliminate nonspecific DNA binding was assembled and coated with neutravidin. The dye-labeled DNA conjugated with biotin was immobilized on the slide via the specific interaction between biotin and neutravidin. With the help of a single-molecule total internal reflection fluorescence (TIRF) microscope, the total number of DNA molecules could be identified by counting the green fluorophores (Cy3) and the amount of 5hmC containing molecules obtained by counting the red ones (Cy5). Apart from counting the fluorophores by direct emission, multiple fluorophores were detected by photo-bleaching in individual DNA molecules. By taking the fluorophore counts, average length of the DNA

fragments, and the multiple fluorophore correction into account, the ultimate 5hmC levels could be obtained.

For the simultaneous optical imaging of 5hmC and 5mC, a dual-labeling method was adopted to label 5hmC and 5mC at once (Fig. 2). Genomic DNA sample was extracted and digested as described above, after which DNA fragments were end-labeled with biotin only, followed by the labeling of 5hmC with Cy5 according to the abovementioned method and the blocking of unlabeled 5hmC. Next, 5mC was labeled with Cy3 by a one-pot procedure using β -GT and TET1 [78]. 5mC was oxidized to 5hmC by TET1, thereby allowing Cy3 to attach to it via an azide-modified glucose. By counting the red (Cy5) and green (Cy3) fluorophores, the amount of 5hmC and 5mC, respectively, could be obtained. As estimated by the dual-labeled DNA absorption spectrum, the labeling efficiencies of both modifications were about 60%, which is acceptable compared with the other 5hmC labeling assay that yielded 70-85% efficiency [79].

Using the dual-labeling method described above, this single-molecule imaging technique provides an opportunity to measure the proximity of 5hmC and 5mC in the same DNA molecule with smFRET. Because 5mC modifications far exceed 5hmC in genomic DNA, we switched the fluorescent labels so that 5hmC was labeled by Cy3 and 5mC by Cy5 to ensure that acceptors (5mC-Cy5) outnumbered donors (5hmC-Cy3). After synthetic DNA was constructed, on which 5hmC and 5mC were separated by defined distances (22, 9, and 1 bp) for smFRET measurements, the low FRET (~ 0.1) state was detected when 5hmC and 5mC were separated by 22 bp and middle-FRET (~ 0.6) and high-FRET (~ 0.82) states observed when 5hmC and 5mC were separated by 9 and 1 bp. A synthetic DNA with adjacent

and opposing hemihydroxymethylated/hemimethylated CpG sites (5hmC/5mCpGs) was constructed, in which a high-FRET (~ 0.78) state was observed. The technique was next applied to mouse cerebellum DNA, which revealed a distinct high-FRET peak (~ 0.78) with no middle-FRET peaks in between. To discern whether the high-FRET state was caused by 5hmC and 5mC's adjacency on the same strand of DNA molecule or by 5hmC and 5mC on the opposing hemihydroxymethylated/hemimethylated CpG site, the denaturation of the DNA molecule was performed prior to the smFRET experiment. Ultimately, the FRET signal was retained after denaturation when 5hmC and 5mC were on the same strand of the synthetic DNA, whereas it disappeared when the synthetic DNA consisted of 5hmC and 5mC on two strands. The FRET signal of the mouse cerebellum DNA disappeared by denaturing, thereby indicating that the high-FRET state was caused by adjacent and opposing CpG sites.

Verification of the methodology

To verify the efficiency and measure the background of the stepwise labeling technique for 5hmC detection, we constructed a synthetic DNA-bearing biotin and fluorophore as a positive control and a synthetic DNA-bearing biotin incubated with fluorophore without enzymes as a negative one. Comparing the labeled sample DNA with the positive and negative controls confirmed that the stepwise labeling technique was highly efficient and exhibits minimal background. Moreover, 5hmC quantification results measured by the technique were approximately the same as those measured by previous HPLC-MS methods in bulk samples. The same verification process was conducted for the 5mC labeling technique, which confirmed labeling to be highly efficient.

To verify that smFRET signals were not from fully hydroxymethylated CpG sites

(5hmC/5hmCpGs) or fully methylated CpG sites (5mC/5mCpGs), we also constructed synthetic DNAs containing such sites and confirmed that no smFRET signals could be observed in the DNAs. By using mouse cerebellum genomic DNA, we further verified that the high-FRET events were detectable only in 5hmC and 5mC dual-labeled samples, not in single-labeled samples with only donors or acceptors.

Potential application of the method

To apply the above-described assay, studies have indicated that gene-environment interactions converge at the metabolic–epigenome–genome axis to regulate gene expression and phenotypic outcomes via altered epigenetic regulation. The mitochondrion, as an intracellular organelle with its own genome, is a central player of the metabolic–epigenome–genome axis capable of interacting with epigenetic mechanisms in different ways [80].

Since the mid-1970s, it has been known that mammalian mitochondria can methylate their own genome. Recent advances strongly suggest that mitochondrial DNA is also subject to epigenetic modifications, particularly by the 5-methylcytosine and 5-hydroxymethylcytosine markers [81]. In 2011, a mitochondrially targeted DNMT1 transcript variant was discovered that uses an upstream alternative translation start site that prompts the inclusion of a mitochondrial targeting sequence [44]. That finding renews interest in mammalian mtDNA methylation. Human mtDNA, at less than 1% of total cellular DNA, comprises 16,569 bp with 435 CpG sites and 4,747 cytosine residues at non-CpG sites. Both cytosines at CpG sites and cytosines at non-CpG sites can be methylated and hydroxymethylated, although with different frequencies. In particular, subsequent studies concluded that 5mC occurred predominantly at the dinucleotide sequence CpG. Mounting

evidence supports the association of mitochondrial DNA methylation with various disease conditions, environmental exposure, drug treatment, and aging, among other conditions [82,83]. Moreover, it is likely that future advances in research on mitochondrial epigenetics will reveal the mitochondrial epigenome as a useful biomarker and putative therapeutic target. However, current studies have been limited by the use of global or targeted methods of measuring DNA methylation. Accurate, large-scale mapping of mtDNA 5mC and 5hmC patterns obtained from blood samples is not yet available. Because of its ultrasensitivity, that single-molecule imaging technique promises to become a more accurate method of profiling 5mC and 5hmC in small genomes such as mtDNA.

Conclusion

We have reviewed major methods applied in investigating the epigenetic modification of 5hmC, with particular focus on the newly emerging technique of single-molecule imaging in combination with smFRET. To our knowledge this is the only measurement technology that can provide information about the colocalization states of 5hmC and 5mC sites. Moreover, that novel single-molecule imaging technique has an ultrahigh sensitivity for which no more than 50 pg of DNA is required for each measurement. That is orders of magnitude less than the quantity of DNA required by conventional quantification techniques such as HPLC-MS or other fluorescence-based methods. As a highly modular method, the future development and application of the technique could illuminate the study of various epigenetic modifications.

Acknowledgments

We thank the 973-program for support (2015CB856304). Authors have no conflict of interest to declare.

Figure Captions

Fig. 1. Technological process for single-molecule fluorescent assays of 5hmC. DNA fragments are end-labeled with biotin and Cy3 by TdT, after which the 5hmC bases are tagged with azide-modified glucoses by β -GT, which are ultimately labeled with Cy5 via the Huisgen cycloaddition reaction. The labeled DNA is captured on the microscope slide coated with neutravidin via the interaction of biotin and neutravidin, and imaged with single-molecule TIRF microscopy.

Fig. 2. Technological process for smFRET analysis between dual-labeled 5mC and 5hmC. DNA fragments are end-labeled with biotin by TdT, while 5hmC bases are labeled with Cy3 via the attachment of azide-modified glucoses by β -GT. The 5mC bases are oxidized to 5hmC by TET1 and tagged with azide-modified glucoses by β -GT, which allows Cy5 to attach to them. The labeled DNA is immobilized on the microscope slide and imaged with single-molecule TIRF microscopy to measure the distance between 5hmC and 5mC via smFRET analysis.

References

- 1 **Seton-Rogers S.** 2015. Tumour evolution: Genetic and epigenetic co-dependency. *Nat Rev Cancer* **15**: 637-.
- 2 **Baylin SB, Jones PA.** 2011. A decade of exploring the cancer epigenome — biological and translational implications. *Nat Rev Cancer* **11**: 726-34.
- 3 **Munzel M, Globisch D, Carell T.** 2011. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew Chem Int Ed Engl* **50**: 6460-8.
- 4 **Moen EL, Mariani CJ, Zullo H, Jeff-Eke M, et al.** 2015. New themes in the biological functions of 5-methylcytosine and 5-hydroxymethylcytosine. *Immunol Rev* **263**: 36-49.
- 5 **Tahiliani M, Koh KP, Shen Y, Pastor WA, et al.** 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-5.
- 6 **Kriaucionis S, Heintz N.** 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**: 929-30.
- 7 **Globisch D, Munzel M, Muller M, Michalakis S, et al.** 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**: e15367.
- 8 **Nestor CE, Ottaviano R, Reddington J, Sproul D, et al.** 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res* **22**: 467-77.
- 9 **Gross JA, Lefebvre F, Lutz P-E, Bacot F, et al.** 2016. Variations in 5-methylcytosine and 5-hydroxymethylcytosine among human brain, blood, and saliva using oxBS and the Infinium MethylationEPIC array. *Biology Methods and Protocols* **1**.
- 10 **Shen L, Song CX, He C, Zhang Y.** 2014. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu Rev Biochem* **83**: 585-614.
- 11 **Huang Y, Rao A.** 2014. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet* **30**: 464-74.
- 12 **Pastor WA, Aravind L, Rao A.** 2013. TETonic shift: biological roles of TET proteins

- in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**: 341-56.
- 13 **Liao YF, Gu J, Wu YB, Long X, et al.** 2016. Low level of 5-Hydroxymethylcytosine predicts poor prognosis in non-small cell lung cancer. *Oncol Lett* **11**: 3753-60.
 - 14 **Chen K, Zhang J, Guo Z, Ma Q, et al.** 2016. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res* **26**: 103-18.
 - 15 **Zhang L-T, Zhang L-J, Zhang J-J, Ye X-X, et al.** 2013. Quantification of the sixth DNA base 5-hydroxymethylcytosine in colorectal cancer tissue and C-26 cell line. *Bioanalysis* **5**: 839-45.
 - 16 **Lian CG, Xu YF, Ceol C, Wu FZ, et al.** 2012. Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell* **150**: 1135-46.
 - 17 **Micevic G, Theodosakis N, Taube JM, Bosenberg MW, et al.** 2016. Attenuation of genome-wide 5-methylcytosine level is an epigenetic feature of cutaneous malignant melanomas. *Melanoma Research* **Publish Ahead of Print**.
 - 18 **Jin SG, Jiang Y, Qiu RX, Rauch TA, et al.** 2011. 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res* **71**: 7360-5.
 - 19 **Vasanthakumar A, Godley LA.** 2015. 5-hydroxymethylcytosine in cancer: significance in diagnosis and therapy. *Cancer Genet* **208**: 167-77.
 - 20 **Song C-X, Sun Y, Dai Q, Lu X-Y, et al.** 2011. Detection of 5-Hydroxymethylcytosine in DNA by Transferring a Keto-Glucose by Using T4 Phage β -Glucosyltransferase. *ChemBioChem* **12**: 1682-5.
 - 21 **Szwagierczak A, Bultmann S, Schmidt CS, Spada F, et al.** 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**: e181.
 - 22 **Song CX, Szulwach KE, Fu Y, Dai Q, et al.** 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**: 68-72.
 - 23 **Yu M, Hon GC, Szulwach KE, Song CX, et al.** 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**: 1368-80.
 - 24 **Booth MJ, Branco MR, Ficz G, Oxley D, et al.** 2012. Quantitative sequencing of

- 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**: 934-7.
- 25 **Mellén M, Ayata P, Dewell S, Kriaucionis S, et al.** 2012. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151**: 1417-30.
 - 26 **Song CX, Diao JJ, Brunger AT, Quake SR.** 2016. Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *P Natl Acad Sci USA* **113**: 4338-43.
 - 27 **Liu S, Ji DB, Cliffe L, Sabatini R, et al.** 2014. Quantitative Mass Spectrometry-Based Analysis of beta-D-Glucosyl-5-Hydroxymethyluracil in Genomic DNA of *Trypanosoma brucei*. *J Am Soc Mass Spectr* **25**: 1763-70.
 - 28 **Cliffe LJ, Hirsch G, Wang J, Ekanayake D, et al.** 2012. JBP1 and JBP2 Proteins Are Fe²⁺/2-Oxoglutarate-dependent Dioxygenases Regulating Hydroxylation of Thymidine Residues in Trypanosome DNA. *J Biol Chem* **287**: 19886-95.
 - 29 **He YF, Li BZ, Li Z, Liu P, et al.** 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-7.
 - 30 **Munzel M, Globisch D, Bruckl T, Wagner M, et al.** 2010. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* **49**: 5375-7.
 - 31 **Feng J, Zhou Y, Campbell SL, Le T, et al.** 2010. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* **13**: 423-30.
 - 32 **Le T, Kim KP, Fan G, Faull KF.** 2011. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal Biochem* **412**: 203-9.
 - 33 **Du C, Kurabe N, Matsushima Y, Suzuki M, et al.** 2015. Robust quantitative assessments of cytosine modifications and changes in the expressions of related enzymes in gastric cancer. *Gastric Cancer* **18**: 516-25.
 - 34 **Liu S, Wang J, Su Y, Guerrero C, et al.** 2013. Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA.

Nucleic Acids Research **41**: 6421-9.

- 35 **Amouroux R, Nashun B, Shirane K, Nakagawa S, et al.** 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol* **18**: 225-33.
- 36 **Tsuji M, Matsunaga H, Jinno D, Tsukamoto H, et al.** 2014. A validated quantitative liquid chromatography–tandem quadrupole mass spectrometry method for monitoring isotopologues to evaluate global modified cytosine ratios in genomic DNA. *Journal of Chromatography B* **953–954**: 38-47.
- 37 **Kudo Y, Tateishi K, Yamamoto K, Yamamoto S, et al.** 2012. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Science* **103**: 670-6.
- 38 **Leoni C, Montagner S, Deho L, D'Antuono R, et al.** 2015. Reduced DNA methylation and hydroxymethylation in patients with systemic mastocytosis. *Eur J Haematol* **95**: 566-75.
- 39 **Yang H, Liu Y, Bai F, Zhang JY, et al.** 2013. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**: 663-9.
- 40 **Ko M, Huang Y, Jankowska AM, Pape UJ, et al.** 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**: 839-43.
- 41 **Huang Y, Pastor WA, Zepeda-Martínez JA, Rao A.** 2012. The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protocols* **7**: 1897-908.
- 42 **Tamar S, Ori G, Uri H, Yael M, et al.** 2016. Simple and cost-effective fluorescent labeling of 5-hydroxymethylcytosine. *Methods and Applications in Fluorescence* **4**: 044003.
- 43 **Davis T, Vaisvila R.** 2011. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *J Vis Exp*.
- 44 **Shock LS, Thakkar PV, Peterson EJ, Moran RG, et al.** 2011. DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. *Proc Natl Acad Sci U S A* **108**: 3630-5.
- 45 **Song C-X, Yu M, Dai Q, He C.** 2011. Detection of 5-hydroxymethylcytosine in a combined glycosylation restriction analysis (CGRA) using restriction enzyme TaqαI.

Bioorganic & Medicinal Chemistry Letters **21**: 5075-7.

- 46 **Sun X, Chung TH, Tan D, Kim A.** 2016. Practical guidelines and consideration of using RRHP for 5hmC detection.
- 47 **Kinney SM, Chin HG, Vaisvila R, Bitinaite J, et al.** 2011. Tissue-specific Distribution and Dynamic Changes of 5-Hydroxymethylcytosine in Mammalian Genomes. *J Biol Chem* **286**: 24685-93.
- 48 **Ficz G, Branco MR, Seisenberger S, Santos F, et al.** 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**: 398-402.
- 49 **Stroud H, Feng SH, Kinney SM, Pradhan S, et al.** 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* **12**.
- 50 **Tan L, Xiong L, Xu W, Wu F, et al.** 2013. Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Research*.
- 51 **Pastor WA, Pape UJ, Huang Y, Henderson HR, et al.** 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394-7.
- 52 **Nestor CE, Reddington JP, Benson M, Meehan RR.** 2014. Investigating 5-Hydroxymethylcytosine (5hmC): The State of the Art. In Stockert JC, Espada J, Blázquez-Castro A, eds; *Functional Analysis of DNA and Chromatin*. Totowa, NJ: Humana Press. p 243-58.
- 53 **Thomson JP, Hunter JM, Nestor CE, Dunican DS, et al.** 2013. Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Research* **41**: e206.
- 54 **Amouroux R, McEwen KR, Hajkova P.** 2014. Current technological advances in mapping new DNA modifications. *Drug Discovery Today: Disease Models* **12**: 15-26.
- 55 **Hack LM, Dick ALW, Provençal N.** 2016. Epigenetic mechanisms involved in the effects of stress exposure: focus on 5-hydroxymethylcytosine. *Environmental Epigenetics* **2**.
- 56 **Sletten EM, Bertozzi CR.** 2009. Bioorthogonal chemistry: fishing for selectivity in a

- sea of functionality. *Angew Chem Int Ed Engl* **48**: 6974-98.
- 57 **Speers AE, Cravatt BF.** 2004. Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol* **11**: 535-46.
- 58 **Han D, Lu X, Shih AH, Nie J, et al.** 2016. A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Molecular Cell* **63**: 711-9.
- 59 **Robertson AB, Dahl JA, Vagbo CB, Tripathi P, et al.** 2011. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research* **39**.
- 60 **Robertson AB, Dahl JA, Ougland R, Klungland A.** 2012. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nature Protocols* **7**: 340-50.
- 61 **Cui L, Chung TH, Tan D, Sun X, et al.** 2014. JBP1-seq: a fast and efficient method for genome-wide profiling of 5hmC. *Genomics* **104**: 368-75.
- 62 **Thomson JP, Hunter JM, Nestor CE, Dunican DS, et al.** 2013. Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res* **41**.
- 63 **Huang Y, Pastor WA, Zepeda-Martinez JA, Rao A.** 2012. The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protoc* **7**: 1897-908.
- 64 **Pastor WA, Huang Y, Henderson HR, Agarwal S, et al.** 2012. The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protoc* **7**: 1909-17.
- 65 **Petterson A, Chung TH, Tan D, Sun XG, et al.** 2014. RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biol* **15**.
- 66 **Sun Z, Terragni J, Borgaro JG, Liu Y, et al.** 2013. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep* **3**: 567-76.
- 67 **Sun Z, Dai N, Borgaro JG, Quimby A, et al.** 2015. A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol Cell* **57**: 750-61.
- 68 **Mooijman D, Dey SS, Boisset J-C, Crosetto N, et al.** 2016. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage

- p>reconstruction.
- Nature Biotechnology*
- 34**
- : 852-6.
- 69 **Yu M, Hon GC, Szulwach KE, Song CX, et al.** 2012. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc* **7**: 2159-70.
 - 70 **Booth MJ, Ost TW, Beraldi D, Bell NM, et al.** 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* **8**: 1841-51.
 - 71 **Flusberg BA, Webster DR, Lee JH, Travers KJ, et al.** 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**: 461-5.
 - 72 **Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, et al.** 2013. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A* **110**: 18904-9.
 - 73 **Wanunu M, Cohen-Karni D, Johnson RR, Fields L, et al.** 2011. Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J Am Chem Soc* **133**: 486-92.
 - 74 **Eid J, Fehr A, Gray J, Luong K, et al.** 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133-8.
 - 75 **Matsuda T, Matsuda S, Yamada M.** 2015. Mutation assay using single-molecule real-time (SMRT(TM)) sequencing technology. *Genes Environ* **37**: 15.
 - 76 **Levy-Sakin M, Ebenstein Y.** 2013. Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. *Curr Opin Biotechnol* **24**: 690-8.
 - 77 **Michaeli Y, Shahal T, Torchinsky D, Grunwald A, et al.** 2013. Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem Commun (Camb)* **49**: 8599-601.
 - 78 **Zhang L, Szulwach KE, Hon GC, Song C-X, et al.** 2013. Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. *Nature Communications* **4**: 1517.
 - 79 **Shahal T, Green O, Hananel U, Michaeli Y, et al.** 2016. Simple and cost-effective fluorescent labeling of 5-hydroxymethylcytosine. *Methods and Applications in Fluorescence* **4**: 044003.
 - 80 **Chinnery PF, Elliott HR, Hudson G, Samuels DC, et al.** 2012. Epigenetics,

- epidemiology and mitochondrial DNA diseases. *Int J Epidemiol* **41**: 177-87.
- 81 **Iacobazzi V, Castegna A, Infantino V, Andria G.** 2013. Mitochondrial DNA
methylation as a next-generation biomarker and diagnostic tool. *Mol Genet Metab* **110**:
25-34.
- 82 **Byun HM, Baccarelli AA.** 2014. Environmental exposure and mitochondrial
epigenetics: study design and analytical challenges. *Hum Genet* **133**: 247-57.
- 83 **van der Wijst MG, Rots MG.** 2015. Mitochondrial epigenetics: an overlooked layer
of regulation? *Trends Genet* **31**: 353-6.

Figure 1

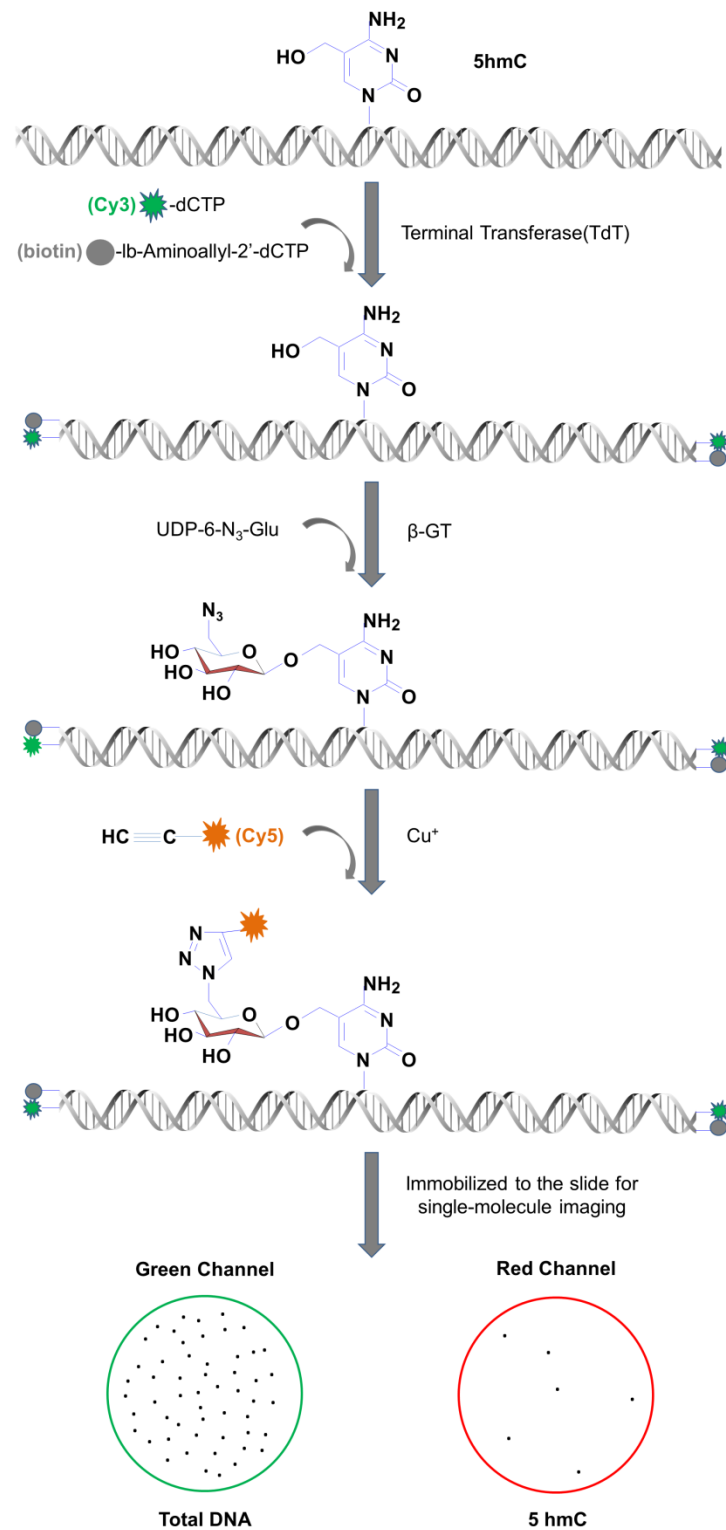


Figure 2

