

Cite this: *Chem. Sci.*, 2017, 8, 7443

## Fluorogenic labeling and single-base resolution analysis of 5-formylcytosine in DNA†

Chaoxing Liu,<sup>‡a</sup> Yafen Wang,<sup>‡a</sup> Wei Yang,<sup>a</sup> Fan Wu,<sup>a</sup> Weiwu Zeng,<sup>a</sup> Zonggui Chen,<sup>b</sup> Jinguo Huang,<sup>a</sup> Guangrong Zou,<sup>a</sup> Xiong Zhang,<sup>a</sup> Shaoru Wang,<sup>a</sup> Xiaocheng Weng,<sup>a</sup> Zhiguo Wu,<sup>b</sup> Yu Zhou<sup>b</sup> and Xiang Zhou<sup>id</sup>\*<sup>a</sup>

5-Formylcytosine (5fC), which plays an important role in epigenetic functions, has received widespread attention in many related fields. Here, we demonstrate a new design for both the fluorogenic switch-on detection and single-base resolution analysis of 5fC through selectively reacting a reagent with 5fC to yield an intramolecular cyclization nucleobase. The generated product, bearing a similar benzothiazole-iminocoumarin scaffold, is highly fluorescent and enables us to qualitatively and quantitatively detect 5fC moieties in  $\gamma$ -irradiated calf thymus DNA. Additionally, losing the exocyclic 4-amino group in 5fC causes the incorporation of dATP through base pairing with the generated nucleobase during polymerase extension, which helped us to analyze the 5fC sites in both single- and double-stranded oligonucleotides. Our Sanger and Illumina sequencing results show great potential in single-base resolution analysis of 5fC. It is hopeful that a similar design may be used for more detection targets.

Received 23rd August 2017  
Accepted 4th September 2017

DOI: 10.1039/c7sc03685j

rsc.li/chemical-science

### Introduction

5-Formylcytosine (5fC) is present at a level of 0.02% to 0.002% that of cytosine in mouse embryonic stem cells (mESCs).<sup>1,2</sup> It also can be found in many cells and tissues,<sup>3</sup> such as Hela cells,<sup>4</sup> HEK293T cells,<sup>5</sup> hepatocellular carcinoma tissues<sup>6</sup> and colorectal carcinoma tissues.<sup>7</sup> It is a transient intermediate in the demethylation of 5-methylcytosine (5mC) by ten-eleven translocation (TET) family enzymes,<sup>8,9</sup> and it is stable in various genomes,<sup>10</sup> with mounting evidence to suggest that it plays a vital role in epigenetic functions.<sup>11</sup> It is associated with gene regulation,<sup>12</sup> alterations in DNA structures,<sup>13,14</sup> cell differentiation<sup>15,16</sup> and some illnesses.<sup>17</sup> Research on 5fC has accelerated the complete understanding of genetic and epigenetic regulation. 5fC can be effectively fluorogenically labeled by hydrazine,<sup>18</sup> amine,<sup>19</sup> amidoxyl<sup>20</sup> and indole derivatives.<sup>21</sup> However, 5-formyluracil (5fU), the modified thymine (T) counterpart of 5fC, is more active,<sup>22–24</sup> which causes many fluorescent reagents to prefer to react with 5fU.<sup>25–27</sup> As far as we know, there is no report of reagents that can fluorogenically switch on 5fC but that do not disturb the fluorescence detection of 5fU and abasic sites

(AP). The difference between 5fC and 5fU is minor, but the reactivity of the aldehyde present in 5fU is more chemoselective, making it a big challenge to design a reagent that can only fluorogenically switch on 5fC.

As for the sequencing of 5fC, the recent discovery of cytosine (C) modifications in genomic DNA has attracted widespread attention and has energized the field of epigenetics.<sup>28</sup> Balasubramanian and co-workers made a breakthrough in the genome-wide map of 5fC in mESCs using the commercial probe O-(biotinylcarbazoylmethyl)hydroxylamine.<sup>29</sup> After that, they created a reduced bisulfite sequencing method (redBS-seq) based on the selective chemical reduction of 5fC to 5-hydroxymethylcytosine (5hmC) followed by bisulfite treatment which can detect 5fC in DNA at single-base resolution.<sup>30</sup> He and co-workers presented a pair of methods, fCAB-seq and fC-Seal, which employ 5fC-selective chemical manipulation to enable base-resolution analysis and its affinity enrichment.<sup>31</sup> Their studies were noteworthy, systematic and timely. Yi *et al.* creatively reported the fC-CET method (bisulfite-free, selective chemical labeling of 5fC and subsequent C-to-T transition during PCR) for the single-base analysis of 5fC that avoided the DNA degradation of many bisulfite-based methods.<sup>32</sup> Subsequently, the modified method (CLEVER-seq) was used in single cell sequencing of 5fC.<sup>33</sup> It is a remarkable and significant finding that reagents undergoing base modification to result in a C-to-T transition during PCR could have wider applications in epigenome sequencing.<sup>32–34</sup> However, reported reagents have been sparse. Thus, the synthesis of reagents that could be used in bisulfite-free, single-base analysis of 5fC is highly needed.

<sup>a</sup>College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, The Institute for Advanced Studies, Hubei Province Key Laboratory of Allergy and Immunology, Wuhan University, Wuhan, Hubei, 430072, P. R. China. E-mail: xzhou@whu.edu.cn; Fax: +86-27-68756663; Tel: +86-27-68756663

<sup>b</sup>College of Life Science, Wuhan University, Wuhan, Hubei, 430072, P. R. China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc03685j

‡ These authors contributed equally.



Herein, we provide new insight into meeting the demands of both fluorogenic labeling and single-base resolution analysis of 5fC in DNA. In Fig. 1, 2-(5-chlorobenzothiazol-2-yl)acetonitrile (CBAN) was chosen to selectively react with 5fC to generate a 5-formyl-2'-deoxyctidine-CBAN adduct (CB-C), which is similar to the synthesis that uses 2-benzothiazoleacetonitrile and 2-aminobenzaldehyde to form 2-amino-3-(1,3-benzothiazol-2-yl)-quinoline.<sup>35</sup> The direct condensation of aminobenzaldehydes and cyano reagents is very effective and highly selective. Other syntheses that do not involve an aminobenzaldehyde structure cannot proceed through intramolecular cyclization with cyano reagents (Scheme S1†).<sup>33,35,36</sup> Among naturally modified nucleobases, only 5fC contains the same structure. Furthermore, the benzothiazole-aminocoumarin fluorophore possesses excellent photophysical properties, such as a high quantum yield and good photostability.<sup>37–40</sup> The generated nucleoside CB-C contains a similar scaffold that may be highly fluorescent. Meanwhile, it loses its exocyclic 4-amino group, which is a competent proton donor in 5fC, thus failing to base pair with guanine (G). This leads to C-to-T transition during polymerase extension and base pairing with adenine (A), which can further lead to C–T conversion signals after PCR amplification and sequencing.<sup>32,33</sup> Thus, we hypothesize that this design can be used in both highly selective fluorescence switch-on detection and single-base resolution analysis of 5fC.

## Results and discussion

### Quantitative and qualitative detection of 5fC

Firstly, to verify the feasibility of our design, 5-formyl-2'-deoxycytidine was reacted with CBAN in a methanol piperidine solution to cyclize *via* the addition of an exocyclic 4-amino group to the cyano group to give the corresponding nucleotide derivative CB-C (Fig. S12 and S13†). The absorbance and fluorescence emission properties of CB-C were acquired in various buffer solutions. The absorbance was detected at 389 nm, and the fluorescence emission maximum was 430 nm (Fig. S8†). Next, to demonstrate the highly efficient fluorescence-based switch-on detection of 5fC, we treated CBAN with canonical deoxynucleosides and their modifications. Only a dramatic fluorescence enhancement ( $\lambda_{\text{ex}}$ : 389 nm,  $\lambda_{\text{em}}$ : 430 nm) for 5-formyl-2'-deoxycytidine can be observed compared to other deoxynucleosides, such as those containing the 5fU moiety, through direct fluorescence readout without further purification (Fig. 2a). We also reacted 5-formyl-2'-deoxyuridine with

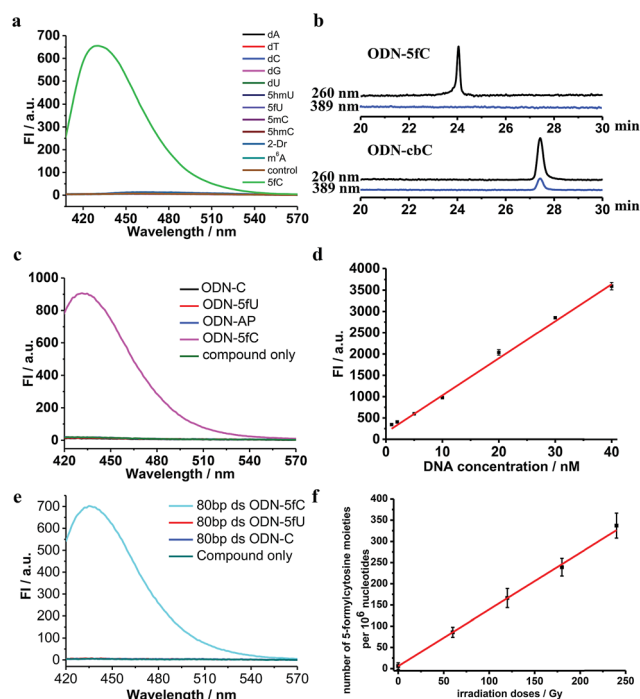


Fig. 2 Quantitative and qualitative detection of 5fC. (a) Fluorescence emission spectra ( $\lambda_{\text{ex}}$ : 389 nm,  $\lambda_{\text{em}}$ : 430 nm) of different nucleosides (100 nM) after incubation with CBAN (100 nM) in MES buffer (100 mM, pH 6.0). (b) RP-HPLC trace at  $\lambda = 260$  nm (black) and 389 nm (blue) of ODN-5fC and ODN-cbC, which were labeled by a reaction with CBAN under optimized conditions. (c) Fluorescence emission spectra ( $\lambda_{\text{ex}}$ : 389 nm,  $\lambda_{\text{em}}$ : 430 nm) of ODNs after reaction with CBAN. (d) Correlation of the fluorescence intensity (at 430 nm) of ODN-5fC after fluorogenic labeling by CBAN with DNA concentration. (e) Fluorescence emission spectra ( $\lambda_{\text{ex}}$ : 389 nm,  $\lambda_{\text{em}}$ : 430 nm) of various dsDNAs that contain 5fU, 5fC or only canonical nucleobases after treatment with CBAN. (f) Quantification of 5fC in  $\gamma$ -irradiated calf thymus DNA (50  $\mu\text{g}$ ) at different irradiation doses (0–240 Gy) ( $\lambda_{\text{ex}}$ : 389 nm,  $\lambda_{\text{em}}$ : 430 nm).

CBAN to generate a 5-formyl-2'-deoxyuridine-CBAN adduct (CB-U) without cyclization (Fig. S14†). The following studies revealed that CB-U is not fluorescent and does not affect the qualitative and quantitative fluorescence detection of 5fC. This may be because the decyclization scaffold contributes to the lack of fluorescence.<sup>39,40</sup> Then, ODN-5fC containing one 5fC site was reacted with CBAN in MES buffer (pH 6.0) at 60 °C for 10 h. Complete conversion to the new product ODN-cbC was recorded using RP-HPLC (monitored at 260 nm and 389 nm) (Fig. 2b). The integrity of the labeled DNAs was confirmed by MALDI-TOF MS (Fig. S1†). The enzymatically digested mononucleosides were analyzed through LC-MS to ensure the reaction of 5fC to yield CB-C (Fig. S4†). In control experiments, we also incubated other ODNs (the 5fC site was replaced by C, 5mC, 5hmC, 5hmU, 5fU and AP) with CBAN under the same conditions. Only 5fU could be labeled with CBAN in the same manner as 5fC, and the other ODNs showed no reactions according to the results of the RP-HPLC and denaturing polyacrylamide gel electrophoresis (PAGE) analyses (Fig. S2 and S3†). Furthermore, the fluorescence intensities of different single-stranded (ss)

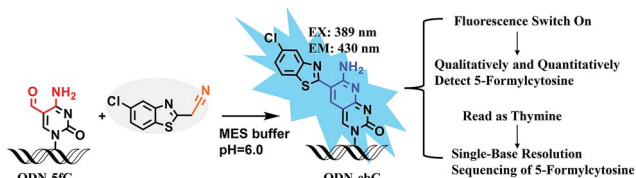


Fig. 1 Strategies of both fluorogenic labeling and single-base resolution analysis of 5fC in DNA.



ODNs after incubation with CBAN under the same conditions demonstrated that CBAN is an excellent reagent for the highly selective fluorogenic tagging of 5fC in DNA, and there was negligible fluorescence observed for the control ODNs, indicating that AP, which also contains aldehydes, did not perturb the fluorescence detection of 5fC (Fig. 2c). A linear correlation between the concentration of ODN-5fC and fluorescence intensity ranging from 0 nM to 40 nM was also observed (Fig. 2d). All of these results indicate that this is a highly selective and fluorescence-based switch-on method for the qualitative and quantitative detection of 5fC. Since most biological samples bearing 5fC are in double-stranded (ds) forms, it is very important to determine whether CBAN can effectively label 5fC in dsDNAs. Therefore, we used a series of 80-bp dsDNAs (containing 5fC or 5fU sites or just canonical nucleosides) as a model test. The fluorescence intensities proved that the reagent can also fluorogenically label 5fC in dsDNA compared to that of other controls, hinting at more possibilities for CBAN in future applications (Fig. 2e).

### Quantification of 5fC in $\gamma$ -irradiated calf thymus DNA at different irradiation doses

Recently, Wagner *et al.* reported that 5fC can be generated from the oxidation of 5mC by being exposed to ionizing radiation in an oxygenated aqueous solution.<sup>41</sup> With all of these encouraging findings, we commenced applying our method to complex biological samples ( $\gamma$ -irradiated calf thymus DNA) to qualitatively and quantitatively detect 5fC formation. First, 5-methyl-2'-deoxycytidine and ODN-5mC (pre-dissolved in an oxygenated aqueous solution) were subjected to <sup>60</sup>Co irradiation (17.4 Gy min<sup>-1</sup>, 60 min) at room temperature. After that, the LC-MS data showed the exact generation of 5fC in both  $\gamma$ -irradiated 5-methyl-2'-deoxycytidine and enzymatically digested  $\gamma$ -irradiated ODN-5mC solutions (Fig. S5 and S6†). Next, we treated CBAN with 5-methyl-2'-deoxycytidine,  $\gamma$ -irradiated 5-methyl-2'-deoxycytidine, ODN-5mC,  $\gamma$ -irradiated ODN-5mC, ODN-C and  $\gamma$ -irradiated ODN-C solutions. We found that only solutions containing a  $\gamma$ -irradiated 5mC moiety were highly fluorescent (Fig. S9†). Then, we prepared  $\gamma$ -irradiated calf thymus DNA (0.5 mg mL<sup>-1</sup>, prior to being bubbled with oxygen for 1 h) at various doses of <sup>60</sup>Co irradiation (0–240 Gy at 1.84 Gy min<sup>-1</sup>). After incubation with CBAN, the fluorescence spectra were measured using excitation at 389 nm and emission at 430 nm. We obtained 1.30 5fC moieties per 10<sup>6</sup> nucleotides per Gy (1.84 Gy min<sup>-1</sup>, <sup>60</sup>Co irradiation) upon evaluation of the slope of the graph shown in Fig. 2f. Comparatively, quantification can also be performed using quantitative mass spectrometry. So we used reported LC-MS/MS analysis methods to detect enzymatically digested  $\gamma$ -irradiated calf thymus DNA<sup>42</sup> and found 1.27 5fC moieties per 10<sup>6</sup> nucleotides per Gy (Fig. S7†). Our results are in reasonable agreement with that of Wagner *et al.*, who reported a formation rate of 0.45 fC moieties per 10<sup>6</sup> nucleotides per Gy (1.2 Gy min<sup>-1</sup>).<sup>41</sup> The slightly higher result might be due to more dissolved oxygen, a higher dose per minute or different batches of biological samples. It is worth noting that the modified nucleoside was also found to be induced by a hydroxyl radical

produced from a Fenton-type reaction.<sup>43</sup> These reasons all might contribute to a higher content of 5fC in the assays. The quantification of 5fC in  $\gamma$ -irradiated calf thymus DNA at different irradiation doses suggests that this is a highly efficient fluorescence-based switch-on detection method of 5fC in DNA. In addition, quantification using quantitative mass spectrometry verified the feasibility of our method. There is no denying that the changeable levels of natural nucleobase modifications in living cells can be obtained using MS-based isotope tracing,<sup>44</sup> and our fluorescence-based switch-on detection method lacks such advantages. Even if the directly competing method (MS with isotope standards) bears the properties of warmer conditions, lower detection limit and higher sensitivity,<sup>5,7,45</sup> our fluorescence-based switch-on detection method does not require expensive MS instruments or professional operation and has many properties such as faster detection, easier operation, visualization technology for qualitative detection and so on which can also be effective supplements in some situations.

### Primer extension assay with KF exo-DNA polymerase

On the other hand, to determine whether such a cyclic adduct could enable the C-to-T transition of 5fC during PCR amplification, similar primer extension assays were prepared.<sup>32</sup> ODN2-5fC and ODN2-C (where the 5fC sites were replaced by C) were reacted with CBAN. After primer extension with Klenow fragment exo-(KF exo-) DNA polymerase, the mixture solution was subjected to denaturing PAGE analysis, revealing the incorporation of dATP into the primers of the ODN2-5fC template (after reaction with CBAN) rather than dGTP when compared to ODN2-C (Fig. 3a). To compare the incorporation efficiency of dATP and dGTP, various concentrations of dATP and dGTP were used in the primer extension reaction. The denaturing PAGE analysis demonstrated that dGTP showed a similar increasing tendency to dATP, but with an obviously slower incorporation rate (Fig. 3b). All of these results showed that the cyclization

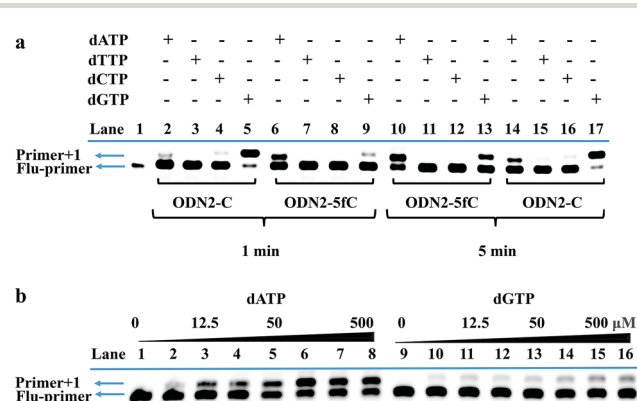


Fig. 3 Primer extension assay with KF exo-DNA polymerase. (a) Incorporation of dATP, dTTP, dCTP and dGTP opposite the original 5fC site of ODN2-5fC (after being labeled with CBAN) for 1 and 5 min. ODN2-C in which the 5fC site was replaced by C was used as a control. (b) Incorporation of dATP or dGTP into the corresponding primers of ODN2-5fC (after reaction with CBAN) with various concentrations of dATP or dGTP (0–500  $\mu$ M).





nucleoside generated from 5fC effectively pairs with A, which encouraged us to do more tests.

### Single-base resolution analysis of 5fC

Finally, an 80-mer oligonucleotide bearing two 5fC sites (80-SS-fC) was labeled with CBAN, PCR amplified and then subjected to Sanger sequencing. The Sanger sequencing results of 80-SS-fC before and after the reaction with CBAN showed exact effective C-to-T signals (Fig. S10†). It is also important to ensure that this reagent can be applied to dsDNAs. The above 80 bp-dsODN-fC was used as a model test. Samples containing 5fC were reacted with CBAN to generate cyclization nucleosides (Fig. 4a, A strands). After the chemicals were removed, excessive primers (primer 1 and 2) and dNTP were added at 95 °C for denaturation for 8 minutes. Then, the samples were immediately put on ice for 5 minutes. Next, KF exo-DNA polymerase and buffer were added at 37 °C for 1 cycle of primer extension to achieve dATP incorporation through pairing with CB-C in the complementary strands (B strands). Then, we obtained a high T signal in the original 5fC sites from the sequencing of the amplification of the B strands. To our delight, the Sanger sequencing results of the 80 bp-dsODN-fC before and after being treated with CBAN revealed an effective C-to-T conversion in the application to the dsDNAs (Fig. 4b). We also investigated multiple 5fC-containing dsDNA segments (four sites) using an Illumina sequencing system. The DNA was incubated with CBAN and then treated using the above protocol. After purification, it was subjected to Illumina library preparation using the ThruPLEX® DNA-seq Kit (Rubicon genomics). An obvious difference in the conversion ratios between the CBAN-treated and untreated samples suggested that 5fC can be detected at base resolution when using the Illumina sequencing method (Fig. S11†). In comparison, peroxotungstate can be used to oxidize 5hmC to thT to undergo base pairing with A and therefore be applied in the analysis of 5hmC at base-resolution.<sup>34</sup> However, it can't be directly applied to the dsDNAs besides with treatment under thermal cycle conditions. A 1,3-indandione derivative<sup>32</sup> and malononitrile<sup>33</sup> were also employed in the analysis of 5fC at base-resolution. However, the reaction times of the 1,3-indandione derivative (24 h) and malononitrile

(20 h) were longer than CBAN (10 h). 2,3,3-trimethylindole derivatives can react with 5fC to produce a fluorescent nucleobase in 3 M NaCl solution.<sup>21</sup> However, the yield and selectivity of the reagents restricted their applications because it can also fluorogenically switch-on 5fU and the condensation was not completed even when the best optimal labeling conditions for 5fC were chosen. Even though similar reagents were shown before, bearing the similar property of converting the coding potential of 5fC from G to A, the reagent we have presented here is also novel as a bifunctional molecule for the simultaneous selective and fluorogenical switch-on detection and analysis of 5fC in a single-base resolution.

## Conclusions

In conclusion, herein we have described a conceptually new type of reagent for use in fluorogenic switch-on detection and base-resolution analysis of 5fC through a selective reaction with 5fC to generate an intramolecular cyclization nucleobase. As far as we know, this is the first reagent that can fluorogenically switch on 5fC while leaving 5fU and AP undisturbed. Its excellent fluorescence properties enabled us to detect the 5fC moieties of  $\gamma$ -irradiated calf thymus DNA. Additionally, loss of the exocyclic 4-amino group, which is a competent proton donor in 5fC, allows the nucleobase to undergo base pairing with A and leads to C-to-T conversion during polymerase extension. We also revealed the possibility of single-base resolution analysis of 5fC for both single- and double-stranded oligonucleotides by employing this reagent. The Sanger and Illumina sequencing results showed the great potential of this complex in the single-base resolution analysis of 5fC. A similar design could be further applied to both qualitative and quantitative detection and single-base resolution analysis of natural C modifications in more cell and tissue studies in the near future.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We thank Professor Chengqi Yi (School of Life Sciences, Peking University) for the analysis of 5fC, technical assistance and for providing DNA samples at the beginning of this project. We also thank Dr Haifang Li (Analysis Center, Tsinghua University) for the DNA MALDI-TOF test instructions. This work was supported by the National Natural Science Foundation of China (21432008, 81373256 and 21572172).

## Notes and references

- 1 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.
- 2 T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeyer and T. Carell, *Angew. Chem., Int. Ed.*, 2011, **50**, 7008–7012.

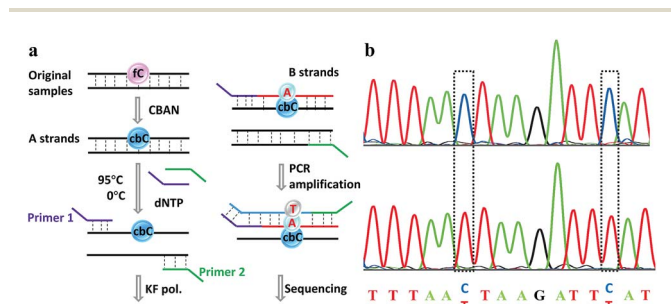


Fig. 4 Single-base resolution analysis of 5fC. (a) Illustration of a strategy for samples containing 5fC moieties. (b) Sanger sequencing analysis before and after CBAN treatment of the oligonucleotides containing two sites of 5fC. The original 5fC sites are surrounded by dotted lines.



- 3 M. Y. Liu, J. E. DeNizio, E. K. Schutsky and R. M. Kohli, *Curr. Opin. Chem. Biol.*, 2016, **33**, 67–73.
- 4 W. Huang, M. Lan, C. Qi, S. Zheng, S. Wei, B. Yuan and Y. Feng, *Chem. Sci.*, 2016, **7**, 5495–5502.
- 5 S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song and Y. Wang, *Nucleic Acids Res.*, 2013, **41**, 6421–6429.
- 6 H. Y. Zhang, J. Xiong, B. L. Qi, Y. Q. Feng and B. F. Yuan, *Chem. Commun.*, 2016, **52**, 737–740.
- 7 D. Gackowski, M. Starczak, E. Zarakowska, M. Modrzejewska, A. Szpila, Z. Banaszkiwicz and R. Olinski, *Anal. Chem.*, 2016, **88**, 12128–12136.
- 8 A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee and A. C. Drohat, *J. Am. Chem. Soc.*, 2013, **135**, 15813–15822.
- 9 L. Zhang, W. Chen, L. M. Iyer, J. Hu, G. Wang, Y. Fu, M. Yu, Q. Dai, L. Aravind and C. He, *J. Am. Chem. Soc.*, 2014, **136**, 4801–4804.
- 10 M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell and S. Balasubramanian, *Nat. Chem. Biol.*, 2015, **11**, 555–557.
- 11 M. Su, A. Kirchner, S. Stazzoni, M. Muller, M. Wagner, A. Schroder and T. Carell, *Angew. Chem., Int. Ed.*, 2016, **55**, 11797–11800.
- 12 Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He and A. Tokmakoff, *ACS Chem. Biol.*, 2016, **11**, 470–477.
- 13 E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi and S. Balasubramanian, *Nat. Struct. Mol. Biol.*, 2015, **22**, 44–49.
- 14 S. Wang, Y. Long, J. Wang, Y. Ge, P. Guo, Y. Liu, T. Tian and X. Zhou, *J. Am. Chem. Soc.*, 2014, **136**, 56–59.
- 15 M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Muller, A. Giese, H. A. Kretzschmar and T. Carell, *Angew. Chem., Int. Ed.*, 2015, **54**, 12511–12514.
- 16 L. Wang, J. Zhang, J. Duan, X. Gao, W. Zhu, X. Lu, L. Yang, J. Zhang, G. Li, W. Ci, W. Li, Q. Zhou, N. Aluru, F. Tang, C. He, X. Huang and J. Liu, *Cell*, 2014, **157**, 979–991.
- 17 Y. Chen, T. Hong, S. Wang, J. Mo, T. Tian and X. Zhou, *Chem. Soc. Rev.*, 2017, **46**, 2844–2872.
- 18 L. Xu, Y. C. Chen, J. Chong, A. Fin, L. S. McCoy, J. Xu, C. Zhang and D. Wang, *Angew. Chem., Int. Ed.*, 2014, **53**, 11223–11227.
- 19 J. Hu, X. Xing, X. Xu, F. Wu, P. Guo, S. Yan, Z. Xu, J. Xu, X. Weng and X. Zhou, *Chem.–Eur. J.*, 2013, **19**, 5836–5840.
- 20 P. Guo, S. Yan, J. Hu, X. Xing, C. Wang, X. Xu, X. Qiu, W. Ma, C. Lu, X. Weng and X. Zhou, *Org. Lett.*, 2013, **15**, 3266–3269.
- 21 B. Samanta, J. Seikowski and C. Höbartner, *Angew. Chem., Int. Ed.*, 2016, **55**, 1912–1916.
- 22 R. E. Hardisty, F. Kawasaki, A. B. Sahakyan and S. Balasubramanian, *J. Am. Chem. Soc.*, 2015, **137**, 9270–9272.
- 23 Y. H. Jang, L. C. Sowers, T. Çağın and W. A. Goddard, *J. Phys. Chem. A*, 2001, **105**, 274–280.
- 24 M. Habgood, S. L. Price, G. Portalone and S. Irrera, *J. Chem. Theory Comput.*, 2011, **7**, 2685–2688.
- 25 W. Hirose, K. Sato and A. Matsuda, *Angew. Chem., Int. Ed.*, 2010, **49**, 8392–8394.
- 26 C. Liu, Y. Chen, Y. Wang, F. Wu, X. Zhang, W. Yang, J. Wang, Y. Chen, Z. He, G. Zou, S. Wang and X. Zhou, *Nano Res.*, 2017, **10**, 2449–2458.
- 27 C. Liu, Y. Wang, X. Zhang, F. Wu, W. Yang, G. Zou, Q. Yao, J. Wang, Y. Chen, S. Wang and X. Zhou, *Chem. Sci.*, 2017, **8**, 4505–4510.
- 28 X. Wu and Y. Zhang, *Nat. Rev. Genet.*, 2017, **18**, 517–534.
- 29 E. A. Raiber, D. Beraldi, G. Ficuz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik and S. Balasubramanian, *Genome Biol.*, 2012, **13**, R69.
- 30 M. J. Booth, G. Marsico, M. Bachman, D. Beraldi and S. Balasubramanian, *Nat. Chem.*, 2014, **6**, 435–440.
- 31 C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin and C. He, *Cell*, 2013, **153**, 678–691.
- 32 B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He and C. Yi, *Nat. Methods*, 2015, **12**, 1047–1050.
- 33 C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang and C. Yi, *Cell Stem Cell*, 2017, **20**, 720–731.
- 34 G. Hayashi, K. Koyama, H. Shiota, A. Kamio, T. Umeda, G. Nagae, H. Aburatani and A. Okamoto, *J. Am. Chem. Soc.*, 2016, **138**, 14178–14181.
- 35 A. Y. Hassan, *Phosphorus, Sulfur, and Silicon and the Related Elements*, 2009, vol. 184, pp. 2856–2869.
- 36 K. Wang, E. Herdtweck and A. Dömling, *ACS Comb. Sci.*, 2012, **14**, 316–322.
- 37 K. Komatsu, Y. Urano, H. Kojima and T. Nagano, *J. Am. Chem. Soc.*, 2007, **129**, 13447–13454.
- 38 K. G. Reddie, W. H. Humphries, C. P. Bain, C. K. Payne, M. L. Kemp and N. Murthy, *Org. Lett.*, 2012, **14**, 680–683.
- 39 T. Kim, H. Kim, Y. Choi and Y. Kim, *Chem. Commun.*, 2011, **47**, 9825–9827.
- 40 C. Wang, S. Yang, M. Yi, C. Liu, Y. Wang, J. Li, Y. Li and R. Yang, *ACS Appl. Mater. Interfaces*, 2014, **6**, 9768–9775.
- 41 G. S. Madugundu, J. Cadet and J. R. Wagner, *Nucleic Acids Res.*, 2014, **42**, 7450–7460.
- 42 Y. Tang, J. Xiong, H. Jiang, S. Zheng, Y. Feng and B. Yuan, *Anal. Chem.*, 2014, **86**, 7764–7772.
- 43 H. Cao and Y. Wang, *Nucleic Acids Res.*, 2007, **35**, 4833–4844.
- 44 T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller and T. Carell, *Nat. Chem. Biol.*, 2014, **10**, 574–581.
- 45 S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schroder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld and T. Carell, *J. Am. Chem. Soc.*, 2013, **135**, 14593–14599.

