TAPS: The development of a direct and base-resolution sequencing method for DNA methylation

Yibin Liu[1, 2]†, Chun-Xiao Song[3, 4†]

[1]College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, China

[2]Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan 430072, China

[3]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK

[4]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK

[†]Corresponding author. Email: chunxiao.song@ludwig.ox.ac.uk; liuyibin@whu.edu.cn

A more efficient DNA methylation sequencing method is highly sought-after for biomedical research and clinical diagnostics. In this in-focus article, we discuss how we developed TET-assisted pyridine borane sequencing (TAPS) as a new bisulfite-free and base-resolution sequencing method for DNA methylation[1].

5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) are the most important epigenetic modifications in human DNA. For decades, bisulfite sequencing (BS-Seq) has been the gold standard for 5mC and 5hmC sequencing. It converts unmodified cytosine to uracil while 5mC and 5hmC remain intact. However, the harsh bisulfite conversion substantially degrades DNA, and the indirect detection greatly reduces the sequence complexity. These technical obstacles of BS-Seq limit its applications. The methyl group at the C5 position of the cytosine base is chemically stable, which limits the direct conversion of 5mC by chemical

reactions. However, 5mC can be oxidized to 5hmC, 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) sequentially by ten-eleven translocation (TET) enzymes. This biochemical function of TET enzymes has been applied in numerous sequencing methods for cytosine modifications including TET-assisted bisulfite sequencing (TAB-Seq), which provides single-base resolution detection of 5hmC[2]. In TAB-Seq, 5mC is oxidised to 5caC by the TET enzyme while 5hmC is protected by β-glucosyltransferase (βGT) as β-glucosyl-5-hydroxymethylcytosine (5gmC). Bisulfite treatment converts all unmodified cytosines and 5caC (derived from 5mC) to uracil or 5-carboxyluracil, which are read as T during PCR amplification. The original 5hmC bases, on the other hand, remain protected as 5gmC, which are then read as C during PCR and distinguished at base-resolution by sequencing.

Inspired by TAB-Seq, we sought to develop a new method to sequence 5mC and 5hmC directly at base-resolution. Instead of focusing on the chemically stable 5mC, we envisioned that we could exploit its more reactive oxidized derivative 5caC. The new method would consist of two steps: (1) oxidation of 5mC and 5hmC to 5caC by the TET enzyme; (2) a new reaction that converts 5caC into another base. To overcome the drawbacks of bisulfite treatment, this new reaction should be (1) specific to 5caC, causing a base sequence change that enables direct read-out of 5caC without affecting the unmodified cytosine and the rest of genome; (2) mild to minimise DNA damage. We firstly screened reported organic reactions of 4-aminonicotinic acid, which is structurally similar to 5caC. However, these reactions failed on our 5caC-containing model DNA oligo. This may have been due to the low reactivity of the amino group on cytosine base or the incompatibility of DNA with these organic reactions. While attempting cyclization reactions on the model DNA, we added 2-methylpyridine borane, a commonly used reducing reagent, aiming to capture possible reaction intermediates. We were expecting an increased molecular weight (MW) after the proposed reaction. Instead, we observed a reduction

of 41 Da in MW. Initially, we thought the reaction had failed so we repeated it several times. We repeatedly observed a single and clean peak on mass spectrometry with the same reduced mass indicating that an efficient reaction had occurred. We then applied the reaction on longer 5caC-containing model DNA, and confirmed the base sequence had changed by a restriction enzyme digestion assay.

Our next task was to figure out the identity of the reaction product. If it was a decarboxylation product, the MW loss would be 44 Da. Adding two hydrogen atoms results in dihydrocytosine, which brings the MW loss to 42 Da. However, these two products would still be read as C and not change the sequence. Alternatively, if the product is uracil, the MW loss would be 43 Da. Adding two hydrogen atoms results in dihydrouracil (DHU) and brings the MW loss to 41 Da. Based on these analyses, we applied the reaction on 2'-deoxycytidine-5-carboxylic acid and confirmed the product was indeed 2'-deoxy-5,6-dihydrouridine by ^{1}H NMR, ^{13}C NMR, and high resolution mass spectrometry. To the best of our knowledge, this is a previously unknown reductive decarboxylation and deamination reaction. DHU is a natural base modification found in transfer ribonucleic acid. As an uracil derivative with a C5-C6 saturated bond, DHU can be recognized by both DNA and RNA polymerases as T.

The borane reduction chemistry was the 5caC conversion reaction we were looking for. By combining TET oxidation and the borane reduction reaction, it could induce a C-to-T transition of 5mC and 5hmC. We named the new method TAPS. We further optimized TAPS for next-generation sequencing and achieved high sensitivity and specificity with mouse TET1 and pyridine borane. TAPS has two key advantages over BS-Seq: (1) High sequencing quality and mapping rate. TAPS directly converts 5mC and 5hmC while unmodified cytosines remain unchanged. By contrast, BS-seq converts nearly all cytosine to thymine which leads to lower

sequence complexity and mapping rate; (2) Less DNA damage. TAPS utilizes mild enzymatic and chemical reactions at 37°C, in comparison to the high temperature used by BS-seq, which causes severe DNA damage. Consequently, TAPS lowers the sequencing cost by half and the resulting data is much faster to process computationally compared to BS-seq. In addition, TAPS preserves DNA fragments over 10 kilobases long, which allows accurate long range methylation sequencing with third-generation sequencing technologies such as Nanopore and SMRT sequencing[3].

Furthermore, we developed derived methods including TAPS with βGT blocking (TAPSβ) for 5mC-specific sequencing, chemical-assisted pyridine borane sequencing (CAPS) for 5hmC-specific sequencing, and pyridine borane sequencing for 5fC and 5caC sequencing[4]. Together these methods provide a comprehensive toolkit for direct and quantitative sequencing of all four cytosine epigenetic modifications. The versatile borane reduction chemistry-based methods have demonstrated advantages over bisulfite-based or other bisulfite-free methods (Table 1), and could lead to wider epigenetic applications in academic research and clinical diagnostics[5].

Bisulfite?	Bisulfite-based			Bisulfite-free					
Direct detection?	Indirect detection					Direct detection			
Quantitative?	Yes					No	Yes		
Methods	BS-Seq	TAB-Seq	oxBS	ACE-seq	EM-seq	hmC-CATCH	**TAPS**	**TAPSβ**	**CAPS**
C **5mC** **5hmC**	T C C	T T C	T C T	T T C	T C C	C C T	**C** **T** **T**	**C** **T** **C**	**C** **C** **T**
Detect	5mC+ 5hmC	5hmC	5mC	5hmC	5mC+ 5hmC	5hmC	5mC+ 5hmC	5mC	5hmC

Table 1. Comparison of TAPS-based methods and other base-resolution sequencing methods for 5mC and/or 5hmC. BS-Seq: bisulfite sequencing. TAB-Seq: TET-assisted bisulfite sequencing. oxBS: oxidative bisulfite sequencing. ACE-seq: APOBEC-coupled epigenetic sequencing. EM-seq: Enzymatic Methyl-seq. hmC-CATCH: chemical-assisted C-to-T conversion of 5hmC sequencing. TAPS: TET-assisted pyridine borane sequencing. TAPSβ: TAPS with βGT blocking. CAPS: chemical-assisted pyridine borane sequencing.

Competing interests

C.-X.S. is a consultant to Exact Sciences Innovation.

Acknowledgments

We thank Felix Jackson for editing the manuscript.

References

(1) Liu, Y., Siejka-Zielinska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Bockler, B. and Song, C. X. (2019) Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol. 37*, 424-429.

(2) Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J. H., Jin, P., Ren, B. and He, C. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell 149*, 1368-1380.

(3) Liu, Y., Cheng, J., Siejka-Zielinska, P., Weldon, C., Roberts, H., Lopopolo, M., Magri, A., D'Arienzo, V., Harris, J. M., McKeating, J. A. and Song, C. X. (2020) Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biol. 21*, 54.

(4) Liu, Y., Hu, Z., Cheng, J., Siejka-Zielinska, P., Chen, J., Inoue, M., Ahmed, A. A. and Song, C. X. (2021) Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nat. Commun. 12*, 618.

(5) Siejka-Zielinska, P., Cheng, J., Jackson, F., Liu, Y., Soonawalla, Z., Reddy, S., Silva, M., Puta, L., McCain, M. V., Culver, E. L., Bekkali, N., Schuster-Bockler, B., Palamara, P. F., Mann, D., Reeves, H., Barnes, E., Sivakumar, S. and Song, C. X. (2021) Cell-free DNA TAPS provides multimodal information for early cancer detection. *Sci. Adv. 7*, eabh0534.