

## ORIGINAL ARTICLE

# 5-Hydroxymethylcytosine profiling from genomic and cell-free DNA for colorectal cancers patients

Pingting Gao<sup>1</sup>  | Shengli Lin<sup>1</sup> | Mingyan Cai<sup>1</sup> | Yan Zhu<sup>1</sup> | Yanqun Song<sup>2,3</sup> | Yi Sui<sup>2,3</sup> | Jing Lin<sup>2,3</sup> | Jiaxiuyu Liu<sup>2,3</sup> | Xingyu Lu<sup>2,3</sup> | Yunshi Zhong<sup>1</sup> | Yuehong Cui<sup>4</sup> | Pinghong Zhou<sup>1</sup>

<sup>1</sup>Zhongshan Hospital, Endoscopy Center and Endoscopy Research Institute, Fudan University, Shanghai, China

<sup>2</sup>Shanghai Epican Genetech Co. Ltd., Zhangjiang Hi-Tech Park, Shanghai, China

<sup>3</sup>Shanghai Epican Biotech Co.Ltd., Shanghai, China

<sup>4</sup>Medical Oncology Department, Zhongshan Hospital, Fudan University, Shanghai, China

## Correspondence

Yuehong Cui, Medical Oncology Department, Zhongshan Hospital, Fudan University, Shanghai, China.

Email: cui.yuehong@zs-hospital.sh.cn and

Pinghong Zhou, Zhongshan Hospital, Endoscopy Center and Endoscopy Research Institute, Fudan University, Shanghai, China. Email: zhou.pinghong@zs-hospital.sh.cn

## Funding information

This study was supported financially by the following grants: Grant from the National Natural Science Foundation of China (81670483) and Study on Prevention and Control of Major Chronic Non-Infectious Diseases, National Key Research and Development Plan of China (No. 2017YFC1308902).

## Abstract

5-Hydroxymethylcytosine (5hmC) is a DNA modification that is generated by the oxidation of 5-methylcytosine (5mC) in a reaction catalyzed by the ten-eleven translocation (TET) family enzymes. It tends to mark gene activation and affects a spectrum of developmental and disease-related biological processes. In this manuscript, we present a 5hmC selective chemical labelling technology (hmC-Seal) to capture and sequence 5hmC-containing DNA fragments with low input. We tested 10 tumour/adjacent colon cancer tissues and 10 tumour/healthy plasma samples. Furthermore, we tested if this methodology could generate the 5hmC differential genes among cancer patients, healthy controls and precancerous adenoma patients from plasma. Robust cancer-specific epigenetic signatures were identified for colon cancers. The results show that 5hmC is mainly distributed in gene active regions. The results also indicate the potential application of 5hmC change signals in early stage of colon cancer, even show potential in the diagnosis of precancerous adenoma. We demonstrated the robustness of the 5hmC-Seal method in tissue and cell-free DNA (cfDNA) as potential biomarkers. Moreover, this study provides the potential value and feasibility of 5hmC-Seal approach on colorectal cancer (CRC) early detection. We believe this strategy could be an effective liquid biopsy-based diagnosis and a potential prognosis method for colon cancer using cfDNA.

## KEYWORDS

5-hydroxymethylcytosine, colon cancer, hmC-Seal, precancerous adenoma

## 1 | INTRODUCTION

5-methylcytosine (5mC) is one of the most important epigenetic modification that impacts global gene expression in mammals. Genome-wide 5mC pattern is extensively reshaped during cell

differentiation and mammalian cell development, as well as the development and treatment of the human diseases such as cancer.<sup>1,2</sup> In contrast to methylation, demethylation has been observed in mammalian genome for decades<sup>3-6</sup> including passive demethylation and active demethylation. Active demethylation is mediated by the

Pingting Gao, Shengli Lin and Mingyan Cai contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

ten-eleven translocation (TET) family of dioxygenases that oxidize the 5mC modification to 5-hydroxymethylcytosine (5hmC),<sup>7,8</sup> and further to 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)<sup>9-11</sup> in a step-wise manner. 5-Hydroxymethylcytosine is not only the 'intermediate' during active demethylation pathways but also acts as a stable DNA mark that plays crucial epigenetic roles.<sup>12-18</sup> Recently developed genome-wide sequencing methods of 5hmC in various mammalian cells and tissues associate the distribution as a marker for active gene expression.<sup>19-25</sup> 5-Hydroxymethylcytosine is enriched in enhancers, gene-bodies and promoters, and the fold changes in 5hmC level in 5hmC sequencing maps correlate with changes in gene expression levels in RNA-Seq.<sup>25,26</sup>

Cell-free DNA (cfDNA) originating from different tissues into the circulating blood has been studied for a long time and has shown significant roles in clinical diagnosis,<sup>27</sup> boosting the rapid development of the liquid biopsy-field. cfDNA-based biomarkers and detection tools offer substantial advantages over the intrinsic methods. The minimally invasive blood test has revolutionary potentials in clinics, having higher patient compliance, is clinically convenient, cost-efficient and enables dynamic monitoring.<sup>28</sup> Tumour-related somatic mutations in cfDNA have been well studied shown to be consistent with the mutations found in tumour tissue, which has been applied in dynamic monitoring of the drug treatment. However, mutation frequency is low and hard to provide the information on tissue of origin, which hampers the application of mutation detection methods into a universal diagnostic or prognostic method. Sensing hypermethylated 5mC region has been shown as an effect way to detect tumour biomarkers from plasma.<sup>29-31</sup> 5-Hydroxymethylcytosine could serve as a parallel or more valuable biomarker for human diseases because it represents active gene expression changes compared to the gene silencing effect in the hypermethylation region. Once the 5hmC patterns can be sensitively and robustly detected, disease-specific biomarkers could be identified.

Next-generation sequencing is an advanced platform for detecting cytosine modification patterns because of its ability in capturing complex information. A selective chemical labelling-based technology platform named 5hmC-Seal has been applied to map 5hmC using low-input DNA. Here, we showed that 5hmC-Seal technology is robust for 5hmC profiling in low-input DNA including cell-free DNA. We detected differentially expressed 5hmC regions in both tissue gDNA and cfDNA in colon cancer patients. This technology showed high potential in real world clinics.

## 2 | METHODS

### 2.1 | Study design and sample preparation

A total of five colorectal cancer patients and three precancerous adenoma patients above 20 years were diagnosed in Zhongshan Hospital at Fudan University, China from July 2017 to September 2017. All specimens were collected from patients who were newly diagnosed, and were about to undergo surgery, as well as received

no neoadjuvant therapy pre-operation. The control plasma samples were collected from healthy individuals who visited the clinic for medical examination. This study was approved by the Ethical Committee of Medical Research, Shanghai Zhongshan Hospital of Fudan University, and written informed consents were obtained from all patients before the surgery.

Paired cancer tissues and para-carcinoma tissues from five patients were stored at  $-80^{\circ}\text{C}$  after surgical removal. The gDNA was isolated using the Quick-gDNA MicroPrep (Zymoresearch, California, USA) kits according to the manufacturer's protocol. Ten plasma samples were collected from colon cancer patients and healthy individuals. The cfDNA was isolated by the QIAamp Circulating Nucleic Acid Kit (Qiagen, Santa Clarita, CA) according to the manufacturer's protocol.

### 2.2 | Spike-in probe

In this study, two similar spike-in probes with unique sequences named 5hmC spike-in and no5hmC spike-in were designed. 5hmC spike-in: 5'-CTGTCATGGTGACAAAGGCATCC\*GGCAGAAATGCCACACAGCCTCTT TAACCAGCACGCCAACCGCCTCTGCTTCGGCCCTGGTCCGCAGCTGACAAGGTCTTCATAATAGAGAAATCCTG-3', C\*-5hmC modifications. no5hmC spike-in: 5'-CTGTCATGGTGACAAAGGCATCGCAGCGAAATGCCACACAGCCTCTT TAACCAGCACGCCAACCGCCTCTGCTTCGGCCCTGGTCCGCAGCTGACAAGGTCTTCATAATAGAGAAATCCTG-3'.

### 2.3 | 5hmC-Seal-seq library preparation and sequencing

The 5hmC library preparation and sequencing were conducted as described previously.<sup>32</sup> Briefly, we applied the T4 bacteriophage  $\beta$ -glucosyltransferase to transfer an engineered glucose moiety containing an azide group onto the hydroxyl group of 5hmC. Then, 5hmC-containing DNA fragments were labelled by chemical modification with biotin on the azide group for further affinity enrichment. PCR amplification was utilized to amplify the captured DNA fragments, followed by the purification of the PCR products using AMPure XP beads according to the manufacturer's instructions. Afterwards, the sequencing was performed on the Illumina NextSeq 500 platform.

For robustness validation, a standard 5hmC-containing gDNA (500 ng, Catalog# D5018, Zymo Research) isolated from human brain and spleen tissue was tested by 5hmC-Seal approach. This set is an ideal control for detection and quantification methods against 5mC and 5hmC as both the modified cytosines are present at physiologically relevant levels and loci.

### 2.4 | Sequencing data processing

Illumina reads were poste-processed and mapped to the human hg19 assembly using the bowtie program with default parameters. We used samtools<sup>33</sup> to generate bigwig files, and deepTools<sup>34</sup> was adopted to plot the line chart of signal distribution. Model-based

analysis of ChIP-seq (MACS)<sup>35</sup> was used to identify the 5hmC-enriched regions (peaks) in each sample (the *q*value cut-off to call significant regions is 0.05). We used FeatureCount<sup>36</sup> to determine feature counts on gene body for further study. Euclidean distance based on *rlog*-transformed 5hmC signals was used to evaluate the difference between samples, the Bioconductor DESeq2 package<sup>37</sup> was applied to detect the genotype-specific genes, and R package 'pheatmap' (<https://cran.r-project.org/web/packages/pheatmap/index.html>)<sup>38</sup> was used to visualize the distance in a heatmap figure. Gene Ontology (GO) (<http://www.geneontology.org/>) term analyses was performed by R package 'clusterProfiler' (<http://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html>)<sup>39</sup> and visualized as functional gene network by R package 'pathview'.<sup>40</sup>

### 3 | RESULTS

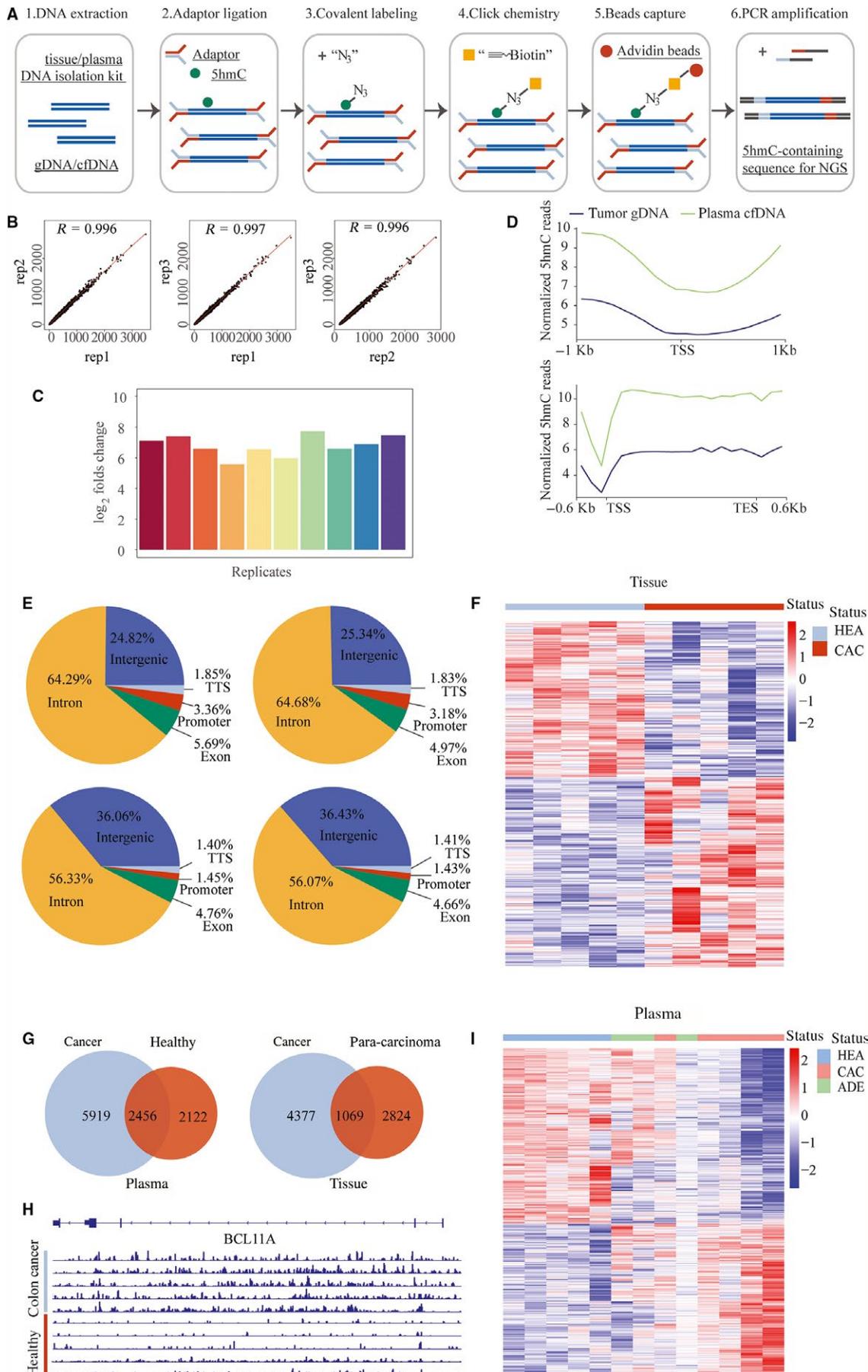
First, we tested the robustness of 5hmC-Seal approach. Although the core chemistry of this technology has been developed for years, a systematic study on the robustness of this technology as a clinical kit is still lacking. We used a commercially available standard 5hmC-containing gDNA (Zymo research) isolated from human brain and spleen tissue to ensure the consistency of the sample. The workflow of 5hmC-Seal profiling method in this study was shown in Figure 1A. As published before, the labelling, capture and washing steps were further optimized for capturing 5hmC-containing DNA fragments from low-input DNA.<sup>32,41</sup> To verify the reproducibility of the technology, we repeated the library construction step of a 10-ng standard DNA sample for 10 times continuously for 10 days with three different technicians, and the correlation between the 10 replicates was analysed via Pearson correlation. To our delight, the 10 5hmC profiling maps were highly correlated to each other ( $R > 0.99$ , three representative correlation analysis results shown in Figure 1B). This test was expanded to three independent batches of reagents with a continuous 20-day test and the similar high quality (data not shown) was maintained. To further study the reliability of the global 5hmC pattern, we visualized the 5hmC signals on the genome browser view, as shown in Figure 2; the 5hmC signals were distributed consistently on the genome among different replicates. Next, we designed two

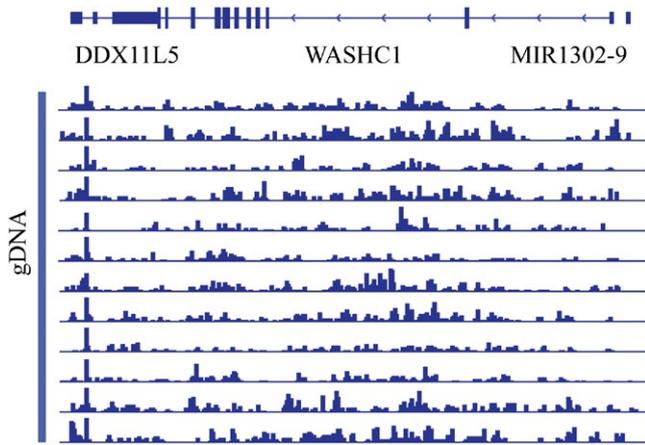
spike-in probes with only limited differences near the oligo ends to function as different sequences during analysis, one contains the 5hmC modification in the middle and one contains normal cytosine. With this design, the sequence content between the two spike-ins is almost identical, thus this could be the ideal model for testing the 5hmC capture efficiency during the 5hmC-Seal assay.

Next, we applied this method in paired colorectal cancer samples, pairs of carcinoma tissues and adjacent tissues from five colorectal cancer patients, as well as plasma samples from colorectal cancer patients ( $n = 5$ ) and health controls ( $n = 5$ ). In general, the average all reads in 20 samples was 30 million bp. The all reads in the five cancer plasma samples were higher, which ranged from 34 million to 47 million bp with a mean value of 41 million bp. We revealed that 5hmC was enriched within the gene body. In addition, the genomic enrichment pattern of 5hmC was consistently observed in both tissue gDNA and blood cfDNA (as shown in Figure 1D). Interestingly, the gene body enrichment of the blood cfDNA is much higher than the enrichment pattern from tissue gDNA, indicating an increased 5hmC level in the gene body of cfDNA compared to the tissue. Furthermore, we calculated the distribution of 5hmC sequencing peaks in various genetic region categories including intergenic, intron, exon (5'untranslated region (UTR), 3'UTR, coding-exons), Transcriptional start site (TSS) and promoter (-2 kb to +0.5 kb relative to TSS) regions to analyse the 5hmC distribution along the genomic features. 5-Hydroxymethylcytosine peaks were enriched within the gene body and kept similar distribution patterns between tumour samples and healthy controls, no matter in gDNA or cfDNA (Figure 1E). We observed different 5hmC peak distribution in the intron and intergenic region between cfDNA sample and tissue sample. Higher peak percentage has been found in intron in tissue samples, while blood samples have higher percentage in intergenic regions.

To address the technology potentials in clinical diagnosis and prognosis, we tested if there are specific 5hmC profile patterns between the tumour samples and para-carcinoma tissues. As shown above, the 5hmC reads were mainly distributed in the gene body, therefore, define that the differential 5hmC region on the gene body could be a robust method to cost-effectively collect enough reads in statistics, which is important for a practical clinical purpose. We

**FIGURE 1** 5-Hydroxymethylcytosine (5hmC)-Seal profiling in colorectal cancer samples. A, Workflow of 5hmC-Seal profiling from gDNA and cell-free DNA (cfDNA) is shown. Purified fragmented gDNA and cfDNA is ligated with sequencing adaptors. 5-Hydroxymethylcytosine-containing fragments are selectively labelled with a biotin group in a two-step reaction. The biotin-labelled fragments are captured on the streptavidin beads, followed by PCR amplification and next-generation sequencing (NGS). B, Scatterplots showing correlation between 5hmC-Seal replicates with Pearson correlation (*r*) displayed. From left to right: correlation between replicate 1 and replicate 2, between replicate 1 and replicate 3, between replicate 2 and replicate 3. C, Column diagram showed the log<sub>2</sub> enrichment fold of 10 replicates. D, Normalized 5hmC reads from gDNA and cfDNA matched in TSS and gene body regions. E, Peak distribution of 5hmC in four groups of samples. Left to right and top to bottom, tumour tissue samples; para-carcinoma tissue; plasma of tumor patients; plasma from healthy individuals. F, Heatmap shows clustering of gDNA samples from tumour tissue (CAC) and para-carcinoma tissues (HEA). Genes and samples were clustered by Euclidean distance using centred *rlog*-transformed expression counts. G, Venn diagram showing the overlap of differential 5hmC peaks in gDNA and cfDNA samples. H, Genome browser views of 5hmC signals detected in BCL11A gene region from libraries generated with gDNA from tissue colon cancer patients and para-carcinoma tissues. I, The heat map shows clustering of cfDNA samples from plasma of colon cancer patients (CAC), precancerous adenoma (ADE) and healthy individuals (HEA). Genes and samples were clustered by Euclidean distance using centred *rlog*-transformed expression counts

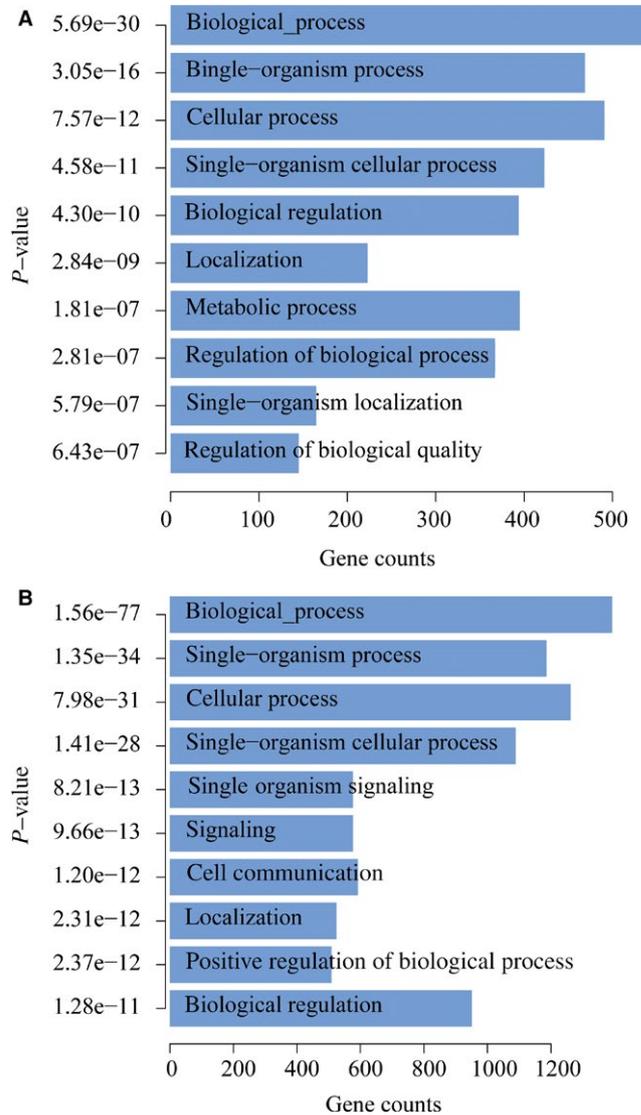




**FIGURE 2** Representative genome browser views of 5-hydroxymethylcytosine (5hmC) signals. Genome browser views of 5hmC signals detected in indicated region from libraries generated from standard 5hmC gDNA replicates showing reliable 5hmC signal distribution

calculated and normalized the 5hmC reads for each gene body from each sample. Differential genes with remarkable 5hmC level difference were identified and listed in Table S1. Then, we applied an unsupervised hierarchical clustering of those differentially modified 5hmC loci, and the samples from colorectal cancer patients were distinctly separated from healthy individuals (Figure 1F). We used Venn diagram to show the overlap of differential 5hmC peaks between cancer samples and healthy control samples. To gain insight into the dynamics of the 5hmC changes, we quantified the number of peaks that were gained or lost in each group. Comparison of plasma from cancer patients and healthy individuals revealed a substantially higher number of peaks gained in cancer (5919) than that which is lost in cancer (2122). Similarly, comparison between cancer tissue and para-carcinoma tissues indicated a higher number of peaks gained in the cancer tissue (4377) than lost in the cancer tissue (2824). The overlap peaks between cancer plasma and healthy plasma was 2456, and 1069 peaks coexist in cancer tissues and para-carcinoma tissues (Figure 1G). When we visualized the 5hmC signal in loci, the 5hmC peaks are strongly increased in the cancer tissue across the gene body of the representative genes (Figures 1H & 2), indicating that this 5hmC signal change could be a long-distance event. Gene Ontology enrichment analysis indicated that these 5hmC-change-related genes may be associated with important biological pathways during cancer development (as shown in Figure 3A).

Finally, we tested if this methodology could generate the 5hmC differential genes among cancer patients, healthy controls and pre-cancerous adenoma patients from plasma, which is the key to evaluate the potential methods in non-invasive clinic diagnosis. We compared five plasma samples from colon cancer patients, three plasma samples from adenoma patients and five plasma samples from healthy individuals. We first identified the differential 5hmC genes ( $P < 0.05$ ) and unsupervised hierarchical clustering with these genes was applied to all the samples. As shown in Figure 1I, the differential 5hmC genes could clearly cluster the samples from cancer patients

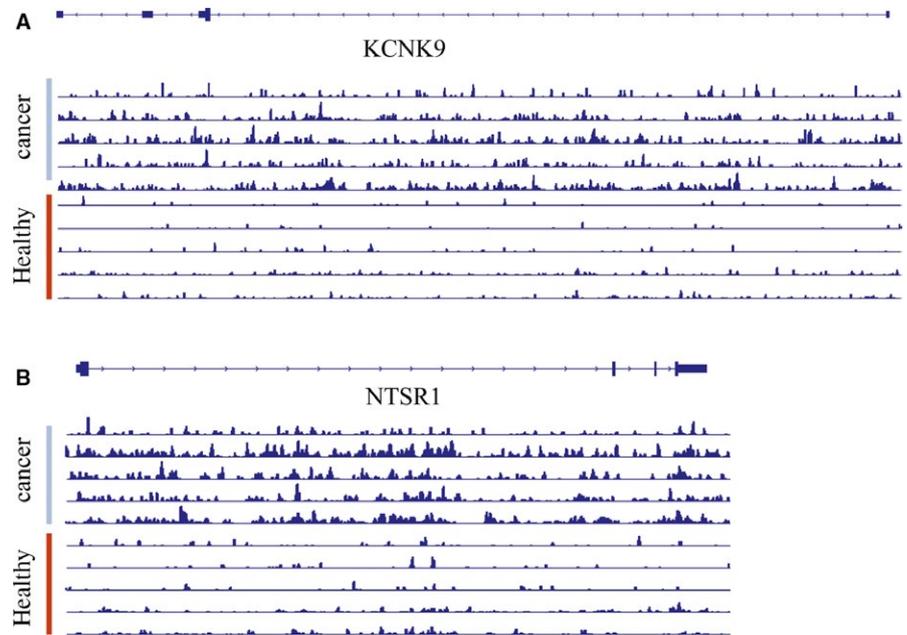


**FIGURE 3** Gene Ontology analysis shows the top categories of 5-hydroxymethylcytosine (5hmC) signal differential genes. A, Top categories of 5hmC signal differential genes identified from cancer tissues and para-carcinoma tissues. B, Top categories of 5hmC signal differential genes identified from plasma of cancer patients and plasma of healthy person

and healthy individuals into two classes, meanwhile, the three adenoma samples hold intermediate 5hmC signal patterns between the healthy and colorectal cancer (CRC) individuals. The results indicate the potential application of 5hmC change signals in colon cancer early, even in the pre-cancerous adenoma. To better understand the differential genes found in the blood, GO analysis was adopted, the results revealed that the differential 5hmC genes from blood samples exhibited similar pathways with that from tissue samples (Figure 3B).

## 4 | DISCUSSION

In this study, spike-in probes with only limited difference were adopted to verify the reliability of the 5hmC-Seal method for



**FIGURE 4** Representative genome browser views of 5-hydroxymethylcytosine (5hmC) signals from cancer tissues and para-carcinoma tissues. Genome browser views of 5hmC signals detected in TWIK-related acid-sensitive K<sup>+</sup> channel 3 (KCNK9) (A) and Neurotensin Receptor 1 Gene (NTSR1). B, gene regions of libraries generated from gDNA of cancer tissues (cancer) and para-carcinoma (healthy) tissues, showing reliable 5hmC signal distribution

5hmC profiling. As a result, the sequence content between the two spike-ins is almost identical, thus this could be the ideal model for testing the 5hmC capture efficiency during the 5hmC-Seal assay. The average log<sub>2</sub> fold enrichment of the spike-in probes in 10 replicates was around seven (Figure 1C), which is consistent with the published data,<sup>21</sup> indicating the high 5hmC capture affinity and the robustness of 5hmC-Seal profiling technology in such low-input DNA.

Next, we applied this method in pairs of carcinoma tissues and adjacent tissues from five colorectal cancer patients, as well as plasma samples from colorectal cancer patients and health controls. We revealed that 5hmC was enriched within the gene body, while it was under-represented in regions near TSS, which is consistent with the previous studies.<sup>23</sup>

Decreased global 5hmC levels in various cancer tissues were reported in previous studies,<sup>15,42</sup> however, 5hmC gained peaks have been discovered as well.<sup>32,43</sup> Our data again indicated that the decreasing global 5hmC level may not represent the 5hmC change in genome in diseases, but rather a total level change in background. When we visualized the 5hmC signal in loci, the 5hmC peaks are strongly increased in the cancer tissue across the gene body of the representative genes (Figures 1H & 4), indicating this 5hmC signal change could be a long-distance event.

In summary, we established a continuous quality control assay for testing the repeatability and stability of 5hmC-Seal technology, and further proved that this 5hmC profiling method is robust in mapping low-input DNA in a cost-effective manner. 5-Hydroxymethylcytosine occurs in gene bodies, indicating that the genomic locations of 5hmC is associated with actively expressed genes. Utilizing a robust and highly efficient profiling-based approach to map 5hmC samples from patients with cancer, we were able to identify differential 5hmC peaks and genes that

can distinguish tumour tissues from the adjacent normal tissues. Our study showed that we can also identify differential 5hmC signals in plasma from colon cancer patients and healthy controls. On account of CRC are mostly sporadic and develop from removable precancerous lesions (adenomas) and curable early stage cancer, thus screening for CRC has high potential could reduce morbidity and mortality.<sup>44</sup> This study provides the potential value and feasibility of 5hmC-Seal approach on CRC early detection. Moreover, the results indicated that 5hmC profiling of cfDNA from liquid biopsies could serve as parallel or more valuable markers for non-invasive diagnosis and prognosis of various diseases.

## 5 | CONCLUSIONS

We demonstrated the robustness of the 5hmC-Seal method in tissue and cfDNA as potential biomarkers. Moreover, this study provides the potential value and feasibility of 5hmC-Seal approach on CRC early detection. We believe this strategy could be an effective liquid biopsy-based diagnosis and potentially serve as a prognosis method for colon cancer using cfDNA.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethical Committee of Medical Research, Shanghai Zhongshan Hospital of Fudan University, and written informed consents were obtained from all patients before the surgery.

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIAL

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## ACKNOWLEDGEMENT

Not applicable.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## AUTHORS' CONTRIBUTIONS

PT Gao, XY Lu, SL Lin, MY Cai, YH Cui, PH Zhou designed the research and collected the samples from patients. PT Gao, Y Zhu, SL Lin and YQ Song and performed the experiments. Y Sui, J Lin, XYJ Liu and XY Lu analysed the data. PT Gao, YH Cui, XY Lu, YS Zhong and PH Zhou were the major contributors in writing the manuscript. All authors read and approved the final manuscript.

## ORCID

Pingting Gao  <https://orcid.org/0000-0001-8208-3134>

## REFERENCES

1. Taby R, Issa JP. Cancer epigenetics. *CA Cancer J Clin*. 2010;60(6):376-392.
2. Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol*. 2013;14(11):215.
3. Guo F, Li X, Liang D, et al. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell*. 2014;15(4):447-458.
4. Suzuki H, Gabrielson E, Chen W, et al. A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat Genet*. 2002;31(2):141-149.
5. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature*. 2000;403(6769):501-502.
6. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol*. 2010;11(9):607-620.
7. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009;324(5929):929-930.
8. Tahiliani M, Koh KP, Shen YH, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009;324(5929):930-935.
9. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011;333(6047):1300-1303.
10. He YF, Li BZ, Li Z, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011;333(6047):1303-1307.
11. Pfaffeneder T, Hackner B, Truss M, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl*. 2011;50(31):7008-7012.
12. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*. 2011;13(1):7-13.
13. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem*. 2014;6(12):1049-1055.
14. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 2010;466(7310):1129-1133.
15. Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*. 2010;468(7325):839-843.
16. Ficz G, Branco MR, Seisenberger S, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 2011;473(7347):398-U589.
17. Guo JU, Su YJ, Zhong C, Ming G-L, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. 2011;145(3):423-434.
18. Koh KP, Yabuuchi A, Rao S, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*. 2011;8(2):200-213.
19. Mooijman D, Dey SS, Boisset JC, et al. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat Biotechnol*. 2016;34:852-856.
20. Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*. 2011;473(7347):394-397.
21. Song CX, Szulwach KE, Fu Y, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol*. 2011;29(1):68-72.
22. Wu H, D'Alessio AC, Ito S, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*. 2011;473(7347):389-393.
23. Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012;149(6):1368-1380.
24. Booth MJ, Branco MR, Ficz G, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. 2012;336(6083):934-937.
25. Han D, Lu X, Shih AH, et al. A highly sensitive and robust method for genomewide 5hmC profiling of rare cell populations. *Mol Cell*. 2016;63:711-719.
26. Konstandin N, Bultmann S, Szwagierczak A, et al. Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia*. 2011;25(10):1649-1652.
27. Lo YM, Tein MS, Lau TK, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet*. 1998;62(4):768-775.
28. Alix-Panabieres C, Pantel K. Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. *Cancer Discov*. 2016;6(5):479-491.
29. Xu R-H, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*. 2017;16:1155-1161.
30. Guo S, Diep D, Plongthongkum N, Fung H-L, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*. 2017;49(4):635.
31. Warren JD, Wei X, Bunker AM, et al. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med*. 2011;9(1):133.
32. Li W, Zhang X, Lu X, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res*. 2017;27(10):1243-1257.

33. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
34. Ramírez F, Dündar F, Diehl S, et al. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-W191.
35. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012;7(9):1728-1740.
36. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
38. Kolde R. Pheatmap: pretty heatmaps. 2012; R package version 0.7.4.
39. Yu G, Wang LG, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
40. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. 2013;29(14):1830-1831.
41. Song CX, Yin S, Ma L, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res*. 2017;27:1231-1242.
42. Evans B, Griner E; Reproducibility Project: Cancer Biology. Registered report: oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Elife*. 2015;4:e07420.
43. Han D, Lu X, Shih AH, et al. A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol Cell*. 2016;63(4):711-719.
44. Souverijn JH. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;371(2):187-188.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Gao P, Lin S, Cai M, et al. 5-Hydroxymethylcytosine profiling from genomic and cell-free DNA for colorectal cancers patients. *J Cell Mol Med*. 2019;23:3530-3537. <https://doi.org/10.1111/jcmm.14252>