

Altered cfDNA Fragmentation Profile in Hypomethylated Regions as Diagnostic Marker in Breast Cancer

Jun Wang (✉ wangjun07@szu.edu.cn)

Shenzhen University <https://orcid.org/0000-0002-5470-4631>

Ming Yang

Shenzhen University

Mingyang Su

Shenzhen University

Lirong Shu

Shenzhen University

Hongxian Wang

Huazhong University of Science and Technology Union Shenzhen Hospital

Xiaoqian Wu

Huazhong University of Science and Technology Union Shenzhen Hospital

Guocheng Zhong

Shenzhen University

Zhixiong Tang

Shenzhen University

Peng Chen

Shenzhen University

Shasha Zhang

Shenzhen University

Yun Wang

Shenzhen University

Li Yu

Shenzhen University

Deming Gou

Shenzhen University <https://orcid.org/0000-0002-3312-841X>

Research Article

Keywords: cfDNA, methylation, fragment size, breast cancer diagnosis

Posted Date: May 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-490423/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Breast cancer, the most common malignancy in women, has been proved to have both altered plasma cell-free DNA (cfDNA) methylation and fragmentation profiles, nevertheless, simultaneously detecting both of them for breast cancer diagnosis has never been reported. Moreover, although fragmentation pattern of cfDNA is determined by nuclease digestion of chromatin, structure of which may be affected by DNA methylation, whether cfDNA methylation and fragmentation are biologically related or not still remains unclear.

Methods: Improved cfMeDIP-seq were utilized to characterize both cfDNA methylation and fragmentation profiles in 25 plasma samples from both healthy individuals and patients with breast cancer. The feasibility of using cfDNA fragmentation profile in hypo- and hyper- methylated regions as diagnostic markers for breast cancer was evaluated.

Results: Mean size of cfDNA fragments ranging from 100 to 220 base pairs (bp) was found to increase from 170.06 (Input libraries) to 173.04 (IP libraries) bp in healthy individuals, which was not observed in patients with breast cancer (170.51 to 170.71 bp). Furthermore, mean size of cfDNA fragments mapped to hypomethylated regions decreased more in patients with breast cancer (4.60 bp, 172.33 bp in hypermethylated regions to 167.73 bp in hypomethylated regions) than healthy individuals (2.87 bp, 174.54 bp in hypermethylated regions to 171.67 bp in hypomethylated regions). The feasibility of using abnormality of short cfDNA fragments ratio in hypomethylated genomic regions for diagnosis of breast cancer in validation cohort was evaluated. 7 out of 11 patients were detected as having breast cancer (63.6% sensitivity), whereas no healthy individuals were mis-detected (100% specificity).

Conclusion: We identified enriched short cfDNA fragments after 5mC-immunoprecipitation (IP) in patients with breast cancer, and demonstrated the enriched short cfDNA fragments might originate from hypomethylated genomic regions. Furthermore, we proved the feasibility of using differentially methylated regions (DMRs)-dependent cfDNA fragmentation profile for breast cancer diagnosis.

Background

Breast cancer is the leading diagnosed malignancy in women worldwide with more than 2 million new cases and 0.6 million deaths each year [1]. The most effective way to achieve better survival for patients with breast cancer is early diagnosis [2]. Cell-free DNA (cfDNA) in blood is emerging as an important non-invasive biomarker for early diagnosis and monitoring the progression of cancer [3–7]. Current research on cfDNA-based cancer detection approaches mainly focus on the differences of methylation or fragment size between cancer- and noncancer-derived cfDNA as they are considered to appear at an early phase of carcinogenesis [3, 8–11].

Global hypomethylation and hypermethylation of tumor suppressor genes have been demonstrated to present in breast cancer [12, 13], and altered methylation of specific genes in cfDNA could serve as biomarkers for early diagnosis have also been widely reported [14–17]. Cancer-derived cfDNA fragments

were proved to be shorter than noncancer-derived cfDNA fragments recently, which led to the aberrant size distribution of cfDNA fragments in patients with cancer [3, 11, 18, 19]. Furthermore, genome-wide cfDNA fragmentation profiling was reported to achieve 70% detecting sensitivity with 95% specificity as biomarker for breast cancer diagnosis [3].

These studies suggested that both altered methylation and aberrant fragmentation were present in cancer-derived cfDNA. Conceptually, approaches detecting these changes simultaneously can better differentiate the origin of cfDNA, and thus improve cancer detection efficacy. However, whether methylation and fragmentation of cfDNA are biologically related or happened independently have not been reported, and need to be investigated.

Since cfDNA is originated from the nucleases digestion of chromatin during multiple cellular processes including apoptosis, necrosis and active cellular secretion [20], fragmentation pattern of cfDNA should be closely related to the accessibility of chromatin, which may be affected by epigenetic modification, nucleosome position and location of transcription machinery [19–23]. Therefore, we hypothesized that methylation profile of cfDNA, which had implications for chromatin remodeling, should be related to fragmentation profile of cfDNA.

In this study, we used the improved cfMeDIP-seq approach to investigate whether the differential methylation of cfDNA in patients with breast cancer was related to cfDNA fragmentation profile or not (Fig. 1). And we further evaluated the possibility of detecting both methylation and fragmentation of cfDNA for better efficacy of breast cancer diagnosis.

Methods

Sample collection and cfDNA extraction

Blood samples from patients with breast cancer in discovery cohort were obtained at the time of treatment in Shenzhen University General Hospital. Blood samples from patients with breast cancer in validation cohort were obtained at the time of diagnosis, before tumor resection or therapy from Huazhong University of Science and Technology Union Shenzhen Hospital. Blood samples from healthy individuals in discovery cohort and validation cohort were obtained at the time of routine screening from Shenzhen University General Hospital and The Third People's Hospital of Shenzhen respectively. This study was approved by the Institutional Review Board of Shenzhen University General Hospital and Huazhong University of Science and Technology Union Shenzhen Hospital according to established ethical guidelines as outlined in the Declaration of Helsinki. All patients signed an informed consent document approved by the Institutional Review Board before entering any study. Clinical characteristics for all participants in this study were listed in Table S1.

All blood samples for participants in this study were collected in tubes containing EDTA as anticoagulant, and processed immediately for plasma isolation. In general, whole blood were first centrifuged at 1000g for 10min at 4 °C for plasma and cellular components separation, and followed by centrifugation at

16000g for 10min at 4 °C for further purifying plasma. The purified plasma was then stored at -80 °C. cfDNA was extracted from plasma using MiniMax™ High Efficiency Cell-Free DNA Isolation Kit (Apostle, A17622-250) according to manufacturer's instructions. The concentration and quality of cfDNA were assessed by the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Q32854) and Bioanalyzer 2100 (Agilent Technologies).

cfMeDIP-seq library construction and sequencing

cfDNA was used for cfMeDIP-seq library preparation with the method described previously with the following modifications [6, 24].

(1) ~ 10 to 20ng cfDNA was ligated with a pool of eight unique paired end Illumina adapters with 8-bp molecular barcodes instead of the NEBNext adapters (NEBNext Multiplex Oligos for Illumina kit, New England BioLabs) (table S6), and the ligation was conducted by using KAPA Hyper Prep kit (KAPA biosystems, KK8504) according to manufacturer's instructions.

(2) The 5-mC monoclonal antibody (Diagenode, C02010021) immunoprecipitated cfDNA and input cfDNA were amplified using Kapa HiFi Hotstart Mastermix (KAPA biosystems, KK8504) and oligos listed in table S6.

(3) The multiplexed libraries were subjected for BioAnalyzer analysis before sequencing on Illumina Novaseq platform at HaploX (Shenzhen, China) with 2 × 150-bp paired-end (PE) reads.

(4) Input and IP libraries were sequenced at 0.5 and 5 respectively.

The specificity of the immunoprecipitation reaction and fold enrichment ratio in sequencing IP libraries were evaluated using the MagMeDIP kit (Diagenode, C02010021) according to the manufacturer's instructions.

Data processing and analysis

Raw reads of cfMeDIP-seq Input and IP libraries were processed according to the following steps. (1) Each reads were labelled with the molecular barcode identified in the leading 8-bp sequences of R1 and R2 reads with 1 mismatch allowed, and then the molecular barcode sequence was removed from raw reads. (2) Illumina sequencing adapter and low quality sequence were removed with cutadapter (version 2.10) and trimmomatic (version 0.39) respectively. (3) Pair reads with insert size less than 20 bp were also eliminated for further analysis. (4) The filtered reads were aligned against the human reference genome (version hg19) using BWA (version 0.7.17-r1188). (5) Only properly paired and unique mapped read pairs with a MAPQ score above 13 were kept, and PCR duplicates defined as having the same genomic start, end and molecular barcode were removed as well. The remaining mapped read pairs in SAM files were converted to BAM format using SAMtools (version 1.7) for further analysis.

cfDNA fragment size analysis

To calculate fragment size of cfDNA, the bam file obtained above was first processed by R package GenomicAlignments (version 1.24.0), and then a Granges object was generated for calculating the fragment size of each cfDNA molecule by R package GenomicRanges (version 1.40.0). Density plot was generated for illustrating the size distribution of cfDNA fragment through R package ggplot2 (version 3.3.2). Short cfDNA fragments were defined as having lengths between 100 bp and 150 bp and long fragments as having lengths between 151 bp and 220 bp according to previous study [3]. Short fragments ratio was calculated as the counts of short cfDNA fragments mapped to the investigated regions or genomic windows dividing by the counts of long cfDNA fragments mapped to the same regions or windows in sequencing libraries. Input-adjusted short fragments ratio was calculated through dividing the short fragments ratio in investigated regions or genomic windows by the short fragments ratio in whole human reference genome (version hg19) of corresponding Input library. Genome-wide cfDNA fragmentation profiles in Input and IP libraries for participants in discovery cohort were calculated without GC adjustment according to the methods reported in previous study [3].

Identification of differentially methylated regions (DMRs)

For each sample from participants, we computed cfDNA fragment counts per 10-kb non-overlapping windows across human reference genome (version hg19), filtered out windows with the mean counts less than 10, and R package DESeq2 (version 1.28.1) with default parameters was used for calling DMRs at $p_{adj} < 0.05$. Hypermethylated and hypomethylated regions were defined as the genomic windows that have $\log_2\text{FoldChange} > 1$ and $\log_2\text{FoldChange} < -1$ in patients with breast cancer compared with healthy individuals, and then illustrated in volcano and heatmap by ggplot2 (version 3.3.2) and pheatmap (version 1.0.12) R packages. Density plot was generated through R package ggplot2 (version 3.3.2) to show fragment size distribution of the cfDNA mapped to hypermethylated and hypomethylated regions. Differentially methylated 10-kb windows were selected as DMRs according to the following criteria. (1) the selected genomic windows should have at least 20 unduplicated cfDNA fragments for all samples including patients with breast cancer and healthy individuals; (2) the selected genomic windows should have input-adjusted short fragments ratio of less than 10 for any samples investigated. For samples of lung cancer from another study [25], same data processing and analysis were used without deduplication step, and DMRs were called at $p\text{-value} < 0.05$ and $|\log_2\text{FoldChange}| > 1$.

Diagnostic model for breast cancer detection

To distinguish patients with breast cancer from healthy individuals using fragmentation profiles in DMRs, we calculated the median input-adjusted short fragments ratio in each differentially hypomethylated 10-kb windows of healthy individuals ($n = 8$) as a baseline profile. We then evaluated the Pearson correlation of the fragmentation profile in each participants from validation cohort to the baseline profile. Cut-offs threshold was determined as the correlation value that can classify healthy individuals and patients with breast cancer at maximum specificity and sensitivity. Receiver operating characteristic (ROC) curve was used to evaluate the classifiers for predicting breast cancer through the R package pROC (version 1.16.2).

Results

Altered cfDNA fragmentation profile upon 5mC-immunoprecipitation (IP)

As it has been reported that cancer-derived cfDNA fragments may have altered methylation and smaller size [6, 11], we decided to focus on cfDNA fragments ranging from 100 to 220 base pairs (bp) to investigate whether the release of cancer-derived cfDNA was related to DNA methylation or not. In a preliminary analysis in discovery cohort, cfDNA extracted from plasma of 3 healthy individuals (H1, H2 and H3) and 3 breast carcinoma patients (P1, P2 and P3) in recovery period with low tumor burden were used for cfMeDIP-seq library construction with some modifications (fig. S1, A-E, and table S1). Both Input and IP libraries were sequenced for pair-end reads with around $0.5 \times$ and $5 \times$ coverage respectively (table S2). Interestingly, we observed a decrease of short cfDNA fragments (100–150 bp) density and short fragments ratio (defined as the ratio of short cfDNA fragments to long cfDNA fragments (151–220 bp)) in IP libraries compared with it in corresponding Input libraries for healthy individuals (Fig. 2, A-C and G), whereas these phenomena were not seen in patients with breast cancer (Fig. 2, D-F and H). Furthermore, mean cfDNA fragments size was found to increase from 170.06 (Input libraries) to 173.04 (IP libraries) bp in healthy individuals, which was not observed in cancer patients (170.51 to 170.71 bp) as well (fig. S2, A and B). To examine differences between healthy individuals and cancer patients, percentage change of short fragments ratio from IP library to corresponding Input library was calculated, we found that patients with breast cancer had significant smaller changes compared with healthy individuals (fig. S2, C-E and table S3).

To find out the short fragments ratio variation across human genome, genome-wide cfDNA fragmentation profiles in both Input (Fig. 2I, upper panel) and IP (Fig. 2I, middle panel) libraries were shown in 5-Mb windows for participants in discovery cohort according to the method described previously [3], changes of cfDNA fragmentation profile (IP - Input) due to 5mC-IP were calculated through subtracting the short fragments ratio in Input libraries from the short fragments ratio in IP libraries in each 5-Mb genomic window (Fig. 2I, lower panel). Smaller changes of short fragments ratio between IP library and Input library were observed in almost all genomic windows across human genome for patients with breast cancer.

Overall, these results suggested that more cancer-derived short cfDNA fragments were enriched during 5mC-IP reaction than noncancer-derived short cfDNA fragments. Therefore, we hypothesized that the enrichment of short cfDNA fragments in cancer patients might be due to the differences in methylation profiles.

Relationship Between Methylation And Fragment Size In Cfdna

To examine origins of the enriched short cfDNA fragments in patients with breast cancer, we first identified 2,211 differentially methylated regions (DMRs) between cfDNA of patients and healthy individuals (1,241 hypermethylated, 970 hypomethylated in patients at $p_{adj} < 0.05$ and $|\log_2\text{FoldChange}| > 1$ with each region represented 10kb genomic window) (Fig. 3, A and B, and table S4). We further evaluated DMRs-dependent cfDNA fragmentation pattern in IP libraries, it was found that cfDNA released from hypomethylated regions had higher short fragments ratio than hypermethylated regions in both patients and healthy individuals (Fig. 3C). Analysis of percentage change for short fragments ratio in hypomethylated regions compared with hypermethylated regions showed patients with breast cancer had increased short fragments ratio in hypomethylated regions compared with healthy individuals (Fig. 3D), which indicated that enriched cancer-derived short cfDNA fragments might be mainly released from hypomethylated regions.

In accordance with increased short fragments ratio in hypomethylated regions, size distribution of cfDNA fragments mapped to hypomethylated regions was found to shift to the direction of smaller size compared with cfDNA fragments mapped to hypermethylated regions, and this shift was to a greater extent in patients with breast cancer (Fig. 4, A-F). Moreover, mean size of cfDNA fragments mapped to hypomethylated regions decreased more in patients with breast cancer (4.60 bp, 172.33 bp in hypermethylated regions to 167.73 bp in hypomethylated regions) than healthy individuals (2.87 bp, 174.54 bp in hypermethylated regions to 171.67 bp in hypomethylated regions). Collectively, these findings again demonstrated that in contrast to healthy individuals, patients with breast cancer had enriched short cfDNA fragments during 5mC-IP reaction, which might mainly originated from hypomethylated genomic regions.

To further confirm the origin of short cfDNA fragments, size of cfDNA fragments in patients with lung cancer from another study were also investigated [25]. As expected, patients with lung cancer had higher percentage change of short fragments ratio in hypomethylated regions compared with it in hypermethylated regions (fig. S3, A and B).

Breast Cancer Diagnostic Accuracy In Validation Cohort

To verify whether the findings obtained from discovery cohort could be applied for diagnosis of breast cancer, we performed cfMeDIP-seq for cfDNA extracted from 11 patients with breast cancer (P4 - P14) and 8 healthy individuals (H4 - H11) in validation cohort (Table S1). All patients with breast cancer had not undergone previous treatment and were confirmed through biopsy. Similarly, increased short cfDNA fragments density in IP libraries of patients with breast cancer was observed (Fig. S6, A and B, and fig. S7). Within the identified 731 DMRs, higher percentage change of short fragments ratio as well as greater shift of size distribution of cfDNA fragments in hypomethylated regions compared with hypermethylated regions were also found for patients with breast cancer (Fig. S8, A-D, fig. S9, and table S5).

Subsequently, we assessed whether DMRs-dependent cfDNA fragmentation profile could differentiate cancer patients from healthy individuals in validation cohort. It was found that abnormal input-adjusted short fragments ratio in specific hypomethylated genomic windows were present for most of the patients with breast cancer, whereas it remained consistent in healthy individuals (fig. S10 and fig. S11).

We then developed an approach called ‘correlation assessment of DMRs-dependent cfDNA fragmentation profile’ to evaluate the abnormality of short fragments ratio in 72 frequently altered hypomethylated genomic windows with at least 20 unduplicated cfDNA fragments for all samples and input-adjusted short fragments ratio of no more than 10 for any samples were identified within each window. Correlation analysis of input-adjusted short fragments ratio from each participant to the median input-adjusted short fragments ratio of healthy individuals in the 72 hypomethylated windows was performed. It was found that healthy individuals had higher correlation with an average of 0.83, whereas patients with breast cancer had lower correlation with an average of 0.68 (Fig. 6A). If using the correlation value as classifier for detecting patients as being healthy or having cancer, at a threshold of 0.72, we detected 7 out of 11 patients as having breast cancer (63.6% sensitivity), whereas no healthy individuals were mis-detected (100% specificity) (Table 1). Receiver operator characteristic analysis for the detection of patients with cancer had an area under the curve (AUC) value of 0.909 (95% confidence interval, 0.771–1.000) (Fig. 6B). Taken together, DMRs-dependent cfDNA fragmentation profiling could distinguish patients with breast cancer and healthy individuals.

Table 1
Effect of cut-offs threshold in detecting breast cancer in validation cohort

Cut-offs (correlation to Healthy median)	Sensitivity (%)	Specificity (%)
0.82	100.0%	75.0%
0.80	90.9%	75.0%
0.76	72.7%	75.0%
0.72	63.6%	100.0%

Discussion

Genome-wide DNA methylation alteration has been demonstrated to present in neoplastic tissue and lead to the changes of chromatin structure [26, 27], which is the direct source for releasing cfDNA into plasma. However, it is still unknown to what extent DNA methylation may affect the initiation and release of cfDNA. In this study, we not only proved that cfDNA fragment size was associated with cfDNA

methylation, but also suggested that DMRs-dependent cfDNA fragmentation profile might provide an alternative approach for breast cancer diagnosis with high efficacy.

Although the fact that size of cancer-derived cfDNA is smaller compared to noncancer-derived cfDNA has been unravelled recently [11], the cause of this shortening remains unclear. Differences in nucleosome wrapping and action mode of nuclease during apoptosis are considered to determine the size of cfDNA fragments in plasma [28]. As nucleosome compaction and rigidity decrease upon DNA demethylation [29, 30], hypomethylated regions in genome should be more vulnerable to nuclease digestion during apoptosis in theory. In accordance with this hypothesis, our results showed that cfDNA originated from hypomethylated regions in patients with breast cancer tend to have significant smaller size compared with healthy individuals, which might be the result of excessive digestion of the wrapped DNA in nucleosome by nuclease with decreasing nucleosome compaction in hypomethylated regions (Fig. 7). Furthermore, global DNA hypomethylation presented in white blood cells of patients with breast cancer could lead to the genome instability and opening of chromatin [31, 32], and thus aggravated the nuclease digestion (Fig. 7). Despite the variation of cfDNA fragmentation profile in hypomethylated regions in patients with breast cancer, it was relatively consistent in healthy individuals. We identified that short fragments ratio of cfDNA mapped to both hypermethylated regions and hypomethylated regions had less changes in healthy individuals, and we supposed this phenomenon was an indicator of genome integrity and stability.

Hypomethylation in promoter regions of oncogenes was found to occur in breast carcinomas [33, 34], therefore aberrant short cfDNA fragments might partially originated from certain oncogene. Indeed, previous studies suggested that short cfDNA fragments harbor footprints of transcription factors [19]. In this study, cfDNA mapped to TRAF3IP3, PTPRN2 and GALNT9 gene locus in hypomethylated regions were found to have significant increased short fragments ratio in patients with breast cancer. Increased expression of these three genes during tumor growth has been reported previously, which might indicate demethylation in their promoter regions [35–38], and thus lead to the release of short cfDNA fragments due to excessive digestion by nucleases. In addition, most of genomic windows in hypomethylated regions that have altered short fragments ratio in patients with breast cancer were found to colocalized with histone modification marker H3K27ac (data not shown), therefore DNA modification might also work in conjunction with histone modification for determining the fragmentation pattern of cfDNA in breast cancer.

This study showed that high sensitivity and specificity detection of early-stage breast cancer could be achieved by characterizing the fragmentation profile of cfDNA in DMRs. As genome-wide fragmentation profiles varied slightly for participants in validation cohort, differentiating patients with breast cancer from healthy individuals became difficult under this circumstance. Besides, although various DMRs were identified, cancer-related DMRs rather than individual variations-related DMRs still required further discrimination. Nevertheless, through detecting DMRs-dependent cfDNA fragmentation profile, we could not only precisely focus on the genomic regions that might lead to aberrant cfDNA release, but also help to evaluate diagnostic value of each DMR. If the variation of DMRs-dependent cfDNA fragmentation in

patients with breast cancer could be confirmed in a larger population-based cohort study in future, it could be utilized as a companion approach to routine diagnostic method for detecting patients with breast cancer.

Aberrant epigenetic modifications including changes in DNA methylation, histone modifications and chromatin remodeling are considered to occur at very early stage in neoplastic development and cancer initiation [39–42], hypomethylated intergenic and intronic regions have further been demonstrated to appear early in the transition from normal to neoplastic cells [26, 43, 44]. Release of short cfDNA fragments in hypomethylated regions thus should also occur at early stage during cancer development, which can guarantee early and real-time monitoring of breast cancer development through DMRs-dependent cfDNA fragmentation profile.

Chromatin remodeling involves the assembly of nucleosomes and regulation of DNA accessibility, which may differ depending on the tissue investigated. Through calculating cfDNA short fragments ratio in DMRs, it is possible to remind us the original chromatin structure profile and thus help to inform tissue of origin, which has been demonstrated to be feasible in another study [45]. For example, the altered cfDNA fragmentation profile in TRAF3IP3, PTPRN2 and GALNT9 gene locus together with their upregulated expression could remind us the chromatin changes due to the development of breast carcinomas. In future, DMRs-dependent cfDNA fragmentation profile should be further characterized together with chromatin changes in multiple cancer types to validate the results obtained in this study.

Conclusions

To summarize, through analyzing cfDNA methylation and fragment size simultaneously, this study reveal that the short cfDNA fragments were possibly originated from hypomethylated DNA regions in patients with breast cancer, and demonstrated the feasibility of using a DMRs-dependent cfDNA fragmentation profiling method for detecting breast cancer. Several limitations should also be taken into consideration. The population in this study was relatively small, so as to eliminate misleading, cfDNA samples in discovery cohort were from patients that in recovery period and undergoing treatment, whereas cfDNA samples in validation cohort were from patients at the time of diagnosis. In searching for differentiated methylation profile between patients with breast cancer and healthy individuals, we defined genomic windows with 10kb in length that have $p_{adj} < 0.05$ and $|\log_2\text{FoldChange}| > 1$ as DMRs, which might not be the most appropriate selecting threshold. With more participants as well as more patients with multiple cancer types investigated, identification of DMRs for calculating cfDNA fragment ratio still need further calculation.

Declarations

Acknowledgments

n/a

Authors' contributions

D.G. designed the research. J.W. improved the cfMeDIP-seq library preparation and high-throughput sequencing method. Z.T., S.Z., M.S. conducted cfMeDIP-seq libraries generation. G.Z., H.W., X.W. collected plasma samples. P.C., M.Y. performed comprehensive bioinformatics analyses. J.W. and D.G. wrote the manuscript, and all authors participated in discussion, data interpretation, and manuscript editing.

Authors' information

n/a

Funding

This work was supported by National Natural Science Foundation of China (91739109, 81970053, 81570046, 81870045 and 81700054); Guangdong Provincial Key Laboratory of Regional Immunity and Diseases (2019B030301009); Shenzhen Municipal Basic Research Program (JCYJ20190808123219295 and JCYJ20170818144127727); Interdisciplinary Innovation Team Project of Shenzhen University (843-00000325), Science and Technology Project of Shenzhen Nanshan District (Health Care, 2018012), and the start-up funds from Shenzhen University (to J.W.).

Availability of data and materials

NGS data will be deposited in NCBI's Sequence Read Archive (SRA) and are available through SRA accession number no. upon acceptance.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Shenzhen University General Hospital and Huazhong University of Science and Technology Union Shenzhen Hospital according to established ethical guidelines as outlined in the Declaration of Helsinki. All patients signed an informed consent document approved by the Institutional Review Board before entering any study.

Consent for publication

n/a

Competing interests:

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA

Cancer J Clin 68(6):394–424

2. Welch HG, Prorok PC, O'Malley AJ, Kramer BS (2016) Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N Engl J Med* 375(15):1438–1447
3. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, Jensen S, Medina JE, Hruban C, White JR et al (2019) Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* 570:385–389
4. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, Grimes B, Krysan K, Yu M, Wang W et al (2017) CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol* 18(1):53
5. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K (2017) Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet* 49(4):635–642
6. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T et al (2018) Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* 563(7732):579–583
7. Hulbert A, Jusue-Torres I, Stark A, Chen C, Rodgers K, Lee B, Griffin C, Yang A, Huang P, Wrangle J (2017) Early detection of lung cancer using DNA promoter hypermethylation in plasma and sputum. *Clin Cancer Res* 23(8):1998–2005
8. Moss J, Zick A, Grinshpun A, Carmon E, Maoz M, Ochana BL, Abraham O, Arieli O, Germansky L, Meir K et al (2020) Circulating breast-derived DNA allows universal detection and monitoring of localized breast cancer. *Ann Oncol* 31(3):395–403
9. Xu Z, Sandler DP, Taylor JA (2020) Blood DNA methylation and breast cancer: A prospective case-cohort analysis in the sister study. *J Natl Cancer Inst* 112(1):87–94
10. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Liu MC, Oxnard GR, Klein EA, Smith D, Richards D et al (2020) Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* 31(6):745–759
11. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, Mair R, Goranova T, Marass F, Heider K (2018) Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 10(466):eaat4921
12. Tanas AS, Sigin VO, Kalinkin AI, Litviakov NV, Slonimskaya EM, Ibragimova MK, Ignatova EO, Simonova OA, Kuznetsova EB, Kekeeva TV et al (2019) Genome-wide methylotyping resolves breast cancer epigenetic heterogeneity and suggests novel therapeutic perspectives. *Epigenomics* 11(6):605–617
13. Karsli-Ceppioglu S, Dagdemir A, Judes G, Ngollo M, Penault-Llorca F, Pajon A, Bignon YJ, Bernard-Gallon D (2014) Epigenetic mechanisms of breast cancer: an update of the current knowledge. *Epigenomics* 6(6):651–664
14. Uehiro N, Sato F, Pu F, Tanaka S, Kawashima M, Kawaguchi K, Sugimoto M, Saji S, Toi M (2016) Circulating cell-free DNA-based epigenetic assay can detect early breast cancer. *Breast Cancer Res* 18(1):129

15. Bacolod MD, Huang J, Giardina SF, Feinberg PB, Mirza AH, Swistel A, Soper SA, Barany F (2020) Prediction of blood-based biomarkers and subsequent design of bisulfite PCR-LDR-qPCR assay for breast cancer detection. *BMC Cancer* 20(1):85
16. Salta S, Fontes-Sousa SPN, Lopes M, Freitas P, Caldas M, Antunes M, Castro L, Antunes F, P, Palma de Sousa S et al: **A DNA methylation-based test for breast cancer detection in circulating cell-free DNA.** *J Clin Med* 2018, **7**(11):420
17. Gharibian A, Hashemi S, Hadi M, Abdolghafoorian H, Ardekani A (2015) Serum/plasma DNA methylation pattern and early detection of breast cancer. *Clin Cancer Investig J* 4:120–128
18. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, Gligorich KM, Rostomily RC, Bronner MP, Shendure J (2016) Fragment length of circulating tumor DNA. *PLoS Genet* 12(7):e1006162
19. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J (2016) Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 164(1–2):57–68
20. Chiu RWK, Heitzer E, Lo YMD, Mouliere F, Tsui DWY (2020) Cell-free DNA fragmentomics: the new “Omics” on the block. *Clin Chem* 66(12):1480–1484
21. Han DSC, Ni M, Chan RWY, Chan VWH, Lui KO, Chiu RWK, Lo YMD (2020) The biology of cell-free DNA fragmentation and the roles of DNASE1, DNASE1L3, and DFFB. *Am J Hum Genet* 106(2):202–214
22. Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ et al (2016) MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun* 7:11485
23. Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V (2015) Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genom* 16(13):S1
24. Shen SY, Burgener JM, Bratman SV, De Carvalho DD (2019) Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat Protoc* 14(10):2749–2780
25. Xu W, Lu J, Zhao Q, Wu J, Sun J, Han B, Zhao X, Kang Y (2019) Genome-wide plasma cell-free DNA methylation profiling identifies potential biomarkers for lung cancer. *Dis Markers* 2019:4108474
26. Sheaffer KL, Elliott EN, Kaestner KH (2016) DNA hypomethylation contributes to genomic instability and intestinal cancer initiation. *Cancer Prev Res (Phila)* 9(7):534–546
27. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA et al (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 44(1):40–46
28. Bronkhorst AJ, Ungerer V, Holdenrieder S (2019) The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif* 17:100087
29. Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee T-H (2010) DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc* 132(6):1782–1783

30. Collings CK, Waddell PJ, Anderson JN (2013) Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res* 41(5):2918–2931
31. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM (2011) DNA methylation in white blood cells: association with risk factors in epidemiologic studies. *Epigenetics* 6(7):828–837
32. Delgado-Cruzata L, Wu HC, Perrin M, Liao Y, Kappil MA, Ferris JS, Flom JD, Yazici H, Santella RM, Terry MB (2012) Global DNA methylation levels in white blood cell DNA from sisters discordant for breast cancer from the New York site of the Breast Cancer Family Registry. *Epigenetics* 7(8):868–874
33. Gupta A, Godwin AK, Vanderveer L, Lu A, Liu J (2003) Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma. *Cancer Res* 63(3):664–673
34. Singh P, Yang M, Dai H, Yu D, Huang Q, Tan W, Kernstine KH, Lin D, Shen B (2008) Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. *Mol Cancer Res* 6(11):1710–1717
35. Nasarre P, Bonilla IV, Metcalf JS, Hilliard EG, Klauber-DeMore N (2018) TRAF3-interacting protein 3, a new oncotarget, promotes tumor growth in melanoma. *Melanoma Res* 28(3):185–194
36. Sengelaub CA, Navrazhina K, Ross JB, Halberg N, Tavazoie SF (2016) PTPRN2 and PLCβ1 promote metastatic breast cancer cell migration through PI(4,5)P2-dependent actin remodeling. *EMBO J* 35(1):62–76
37. Pangen RP, Channathodiyil P, Huen DS, Eagles LW, Johal BK, Pasha D, Hadjistephanou N, Nevell O, Davies CL, Adewumi AI et al (2015) The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. *Clin Epigenetic* 7(1):57
38. Berois N, Gattolliat C-H, Barrios E, Capandeguy L, Douc-Rasy S, Valteau-Couanet D, Bénard J, Osinaga E (2013) GALNT9 gene expression is a prognostic marker in neuroblastoma patients. *Clin Chem* 59(1):225–233
39. Werner RJ, Kelly AD, Issa JJ (2017) Epigenetics and precision oncology. *Cancer J* 23(5):262–269
40. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L (2018) Cancer epigenetics: Moving forward. *PLoS Genet* 14(6):e1007362
41. van der Pol Y, Mouliere F (2019) Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA. *Cancer Cell* 36(4):350–368
42. Kanwal R, Gupta S (2012) Epigenetic modifications in cancer. *Clin Genet* 81(4):303–311
43. Goetz SE, Vogelstein B, Hamilton SR, Feinberg AP (1985) Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* 228(4696):187–190
44. Feinberg AP, Gehrke CW, Kuo KC, Ehrlich M (1988) Reduced genomic 5-methylcytosine content in human colonic neoplasia. *Cancer Res* 48(5):1159–1161
45. Sun K, Jiang P, Cheng SH, Cheng THT, Wong J, Wong VWS, Ng SSM, Ma BBY, Leung TY, Chan SL et al (2019) Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Res* 29(3):418–427

Figures

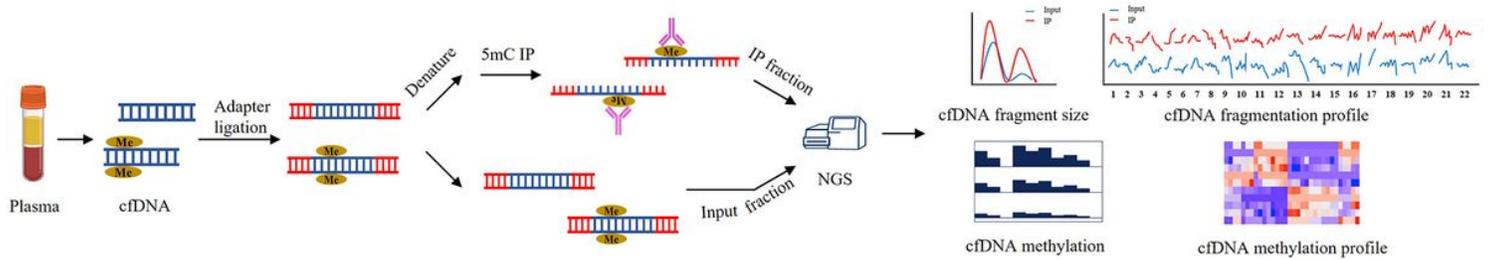


Figure 1

Schematic representation of the improved cfMeDIP-seq approach used in this study. Plasma was collected from patients with breast cancer and healthy individuals. cfDNA was extracted and processed into adapter ligation and 5mC-immunoprecipitation (IP) for sequencing library construction. cfDNA methylation and fragmentation profile were identified through analyzing the NGS data.

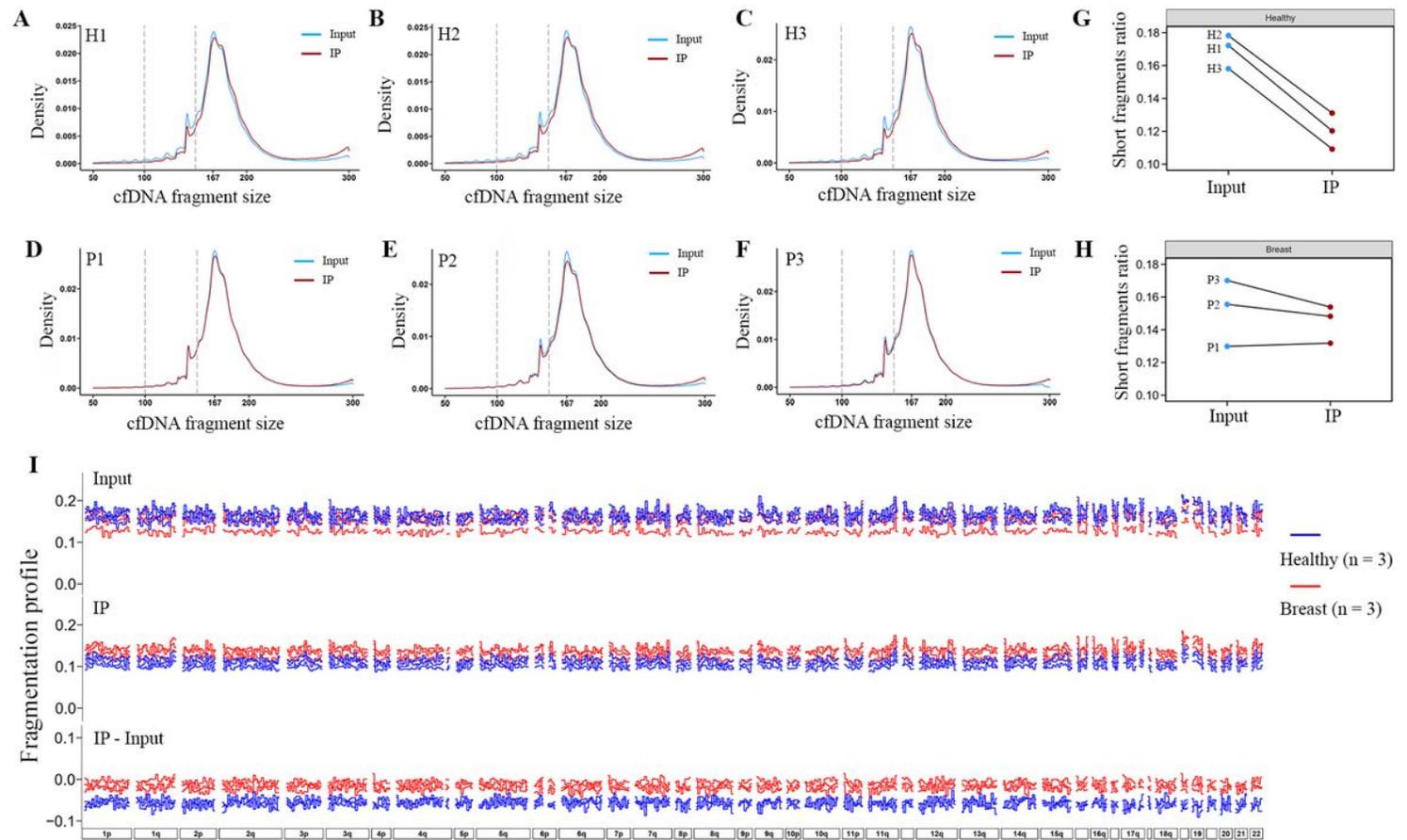


Figure 2

Altered fragmentation profiles of methylated cfDNA in patients with breast cancer. (A-F) Distribution of cfDNA fragment size in Input library (blue line) and IP library (red line) were shown for healthy individuals (H1, H2 and H3) (A - C), and patients with breast cancer (P1, P2 and P3) (D - F). The vertical dashed line indicated cfDNA fragment size at 100 bp and 150 bp. (G and H) Evaluation of short cfDNA fragments

ratio changes (defined as the ratio of short cfDNA fragments (100 – 150 bp) to long cfDNA fragments (151 - 220 bp)) in IP libraries compared with corresponding Input libraries were shown for healthy individuals (G) and patients with breast cancer (H) in discovery cohort. (I) Genome-wide cfDNA fragmentation profiles (the ratio of short cfDNA fragments (100 - 150 bp) to long cfDNA fragments (151 – 220 bp)) in Input (upper panel) and IP (middle panel) libraries were shown in 5-Mb windows for patients with breast cancer (red, n = 3) and healthy individuals (blue, n = 3), changes of cfDNA fragmentation profile (IP - Input, lower panel) were calculated through subtracting the short fragments ratio in Input libraries from the short fragments ratio in IP libraries and shown in each 5-Mb window. Healthy, healthy individuals; Breast, patients with breast cancer.

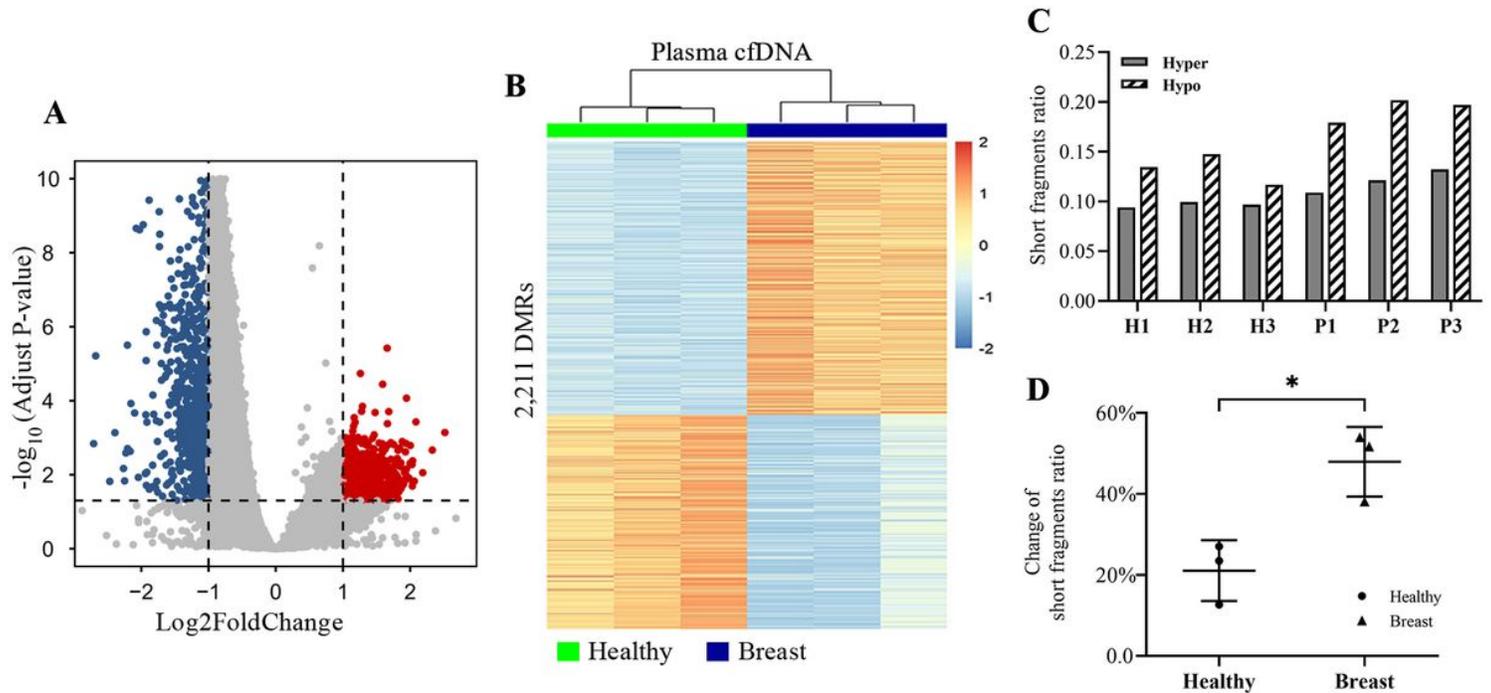


Figure 3

Short cfDNA fragments ratio among DMRs in discovery cohort. (A) Volcano plot of DMRs from patients with breast cancer (n = 3) versus healthy individuals (n = 3). Significantly hypermethylated windows were highlighted in red dots with $p_{adj} < 0.05$, $\log_2\text{Foldchange} > 1$, significantly hypomethylated windows were highlighted in blue dots with $p_{adj} < 0.05$ and $\log_2\text{Foldchange} < -1$. (B) Heatmap of the 2,211 DMRs identified in plasma cfDNA from patients with breast cancer and healthy individuals. (C) Short fragments ratio of cfDNA in hypermethylated and hypomethylated regions in patients with breast cancer and healthy individuals. (D) Percentage change of short fragments ratio in hypomethylated regions compared with it in hypermethylated regions for patients with breast cancer and healthy individuals. Hyper, hypermethylated genomic region; Hypo, hypomethylated genomic regions; Healthy, healthy individuals; Breast, patients with breast cancer; * represents P value < 0.05.

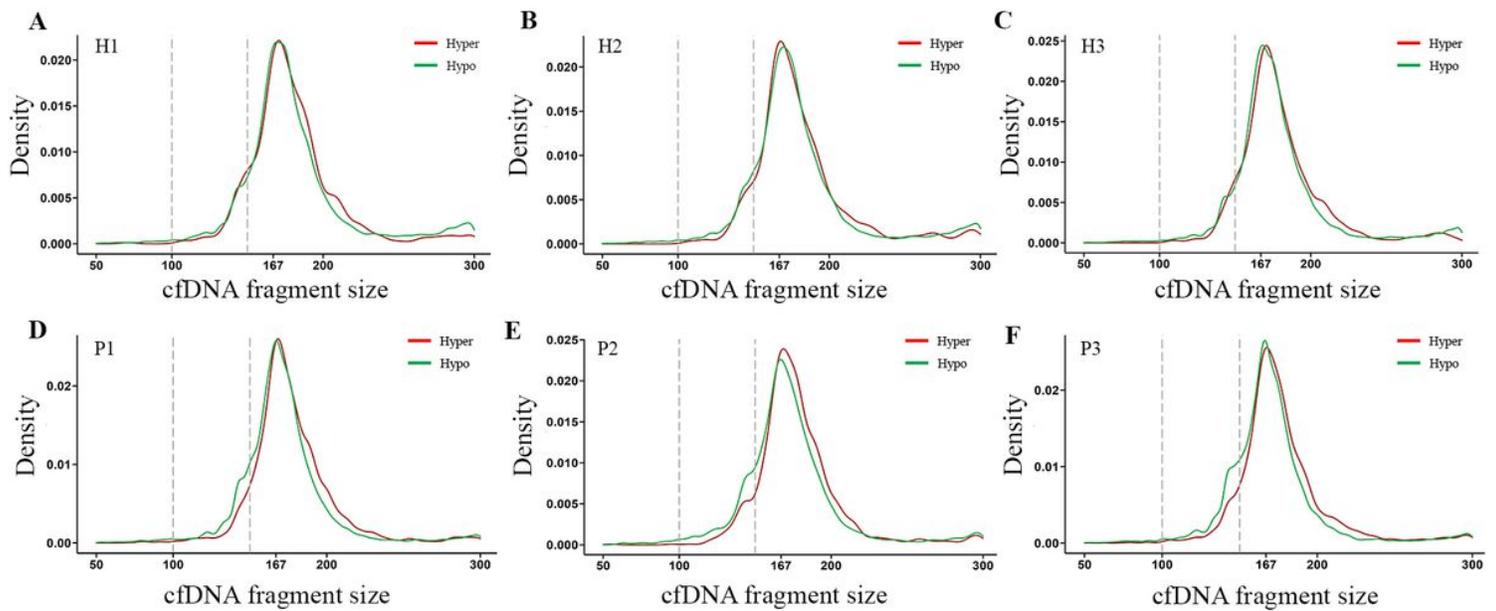


Figure 4

Altered cfDNA fragmentation patterns in hypomethylated regions in patients with breast cancer. (A-C) Distribution of cfDNA fragment size was shown for healthy individuals H1, H2 and H3 in hypermethylated genomic regions (red) and hypomethylated genomic regions (green). (D-F) Distribution of cfDNA fragment size was shown for patients with breast cancer P1, P2 and P3 in hypermethylated genomic regions (red) and hypomethylated genomic regions (green). The vertical dashed line indicated cfDNA fragment size at 100 bp and 150 bp.

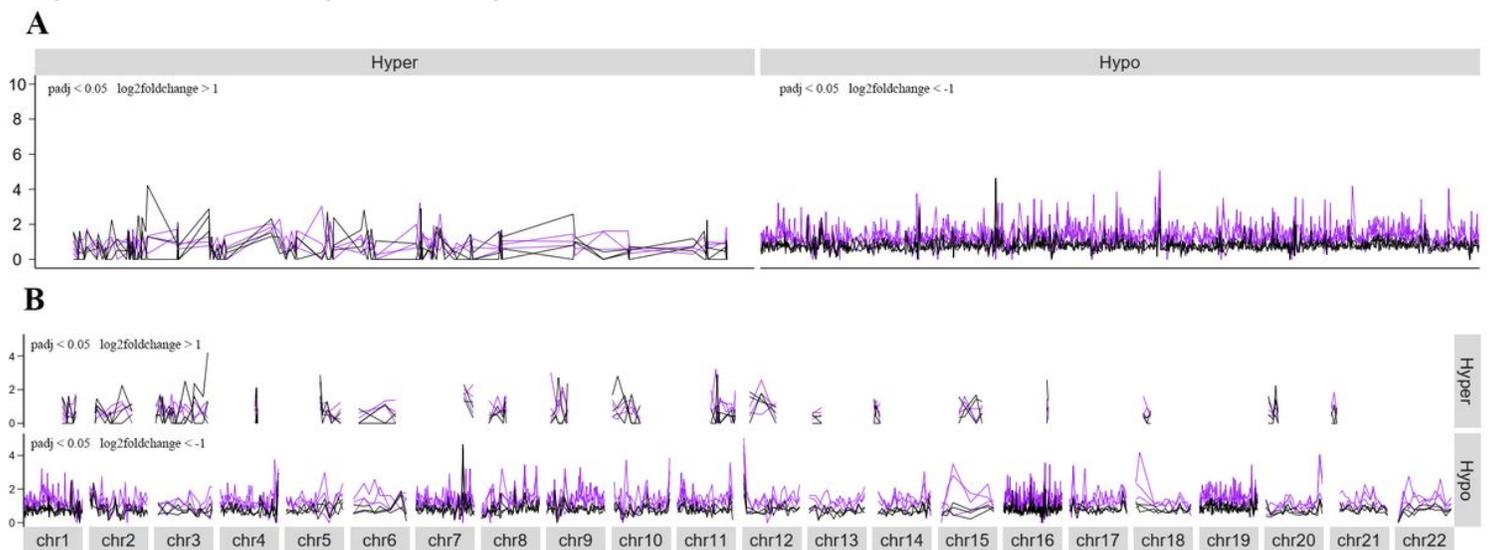


Figure 5

DMRs-dependent cfDNA fragmentation profiles. (A) Input-adjusted short fragments ratio were shown with 10-kb windows in hypermethylated and hypomethylated regions for both patients with breast cancer (purple) and healthy individuals (black). (B) Distribution of the cfDNA fragmentation profile mentioned above was shown across human genome. The input-adjusted short fragments ratio in each 10-kb

window was calculated by dividing short fragments ratio in each 10-kb window by the short fragments ratio in corresponding input libraries. Differentially methylated 10-kb windows were selected for representation according to the following criteria: (1) hypermethylated 10-kb windows have $p_{adj} < 0.05$ and $\log_2\text{FoldChange} > 1$; (2) hypomethylated 10-kb windows have $p_{adj} < 0.05$ and $\log_2\text{FoldChange} < -1$; (3) the selected windows should have at least 20 deduplicated cfDNA fragments for all samples including patients with breast cancer and healthy individuals; (4) the selected windows should have input-adjusted short fragments ratio of less than 10 for any samples analyzed. Hyper, hypermethylated genomic resiong; Hypo, hypomethylated genomic regions.

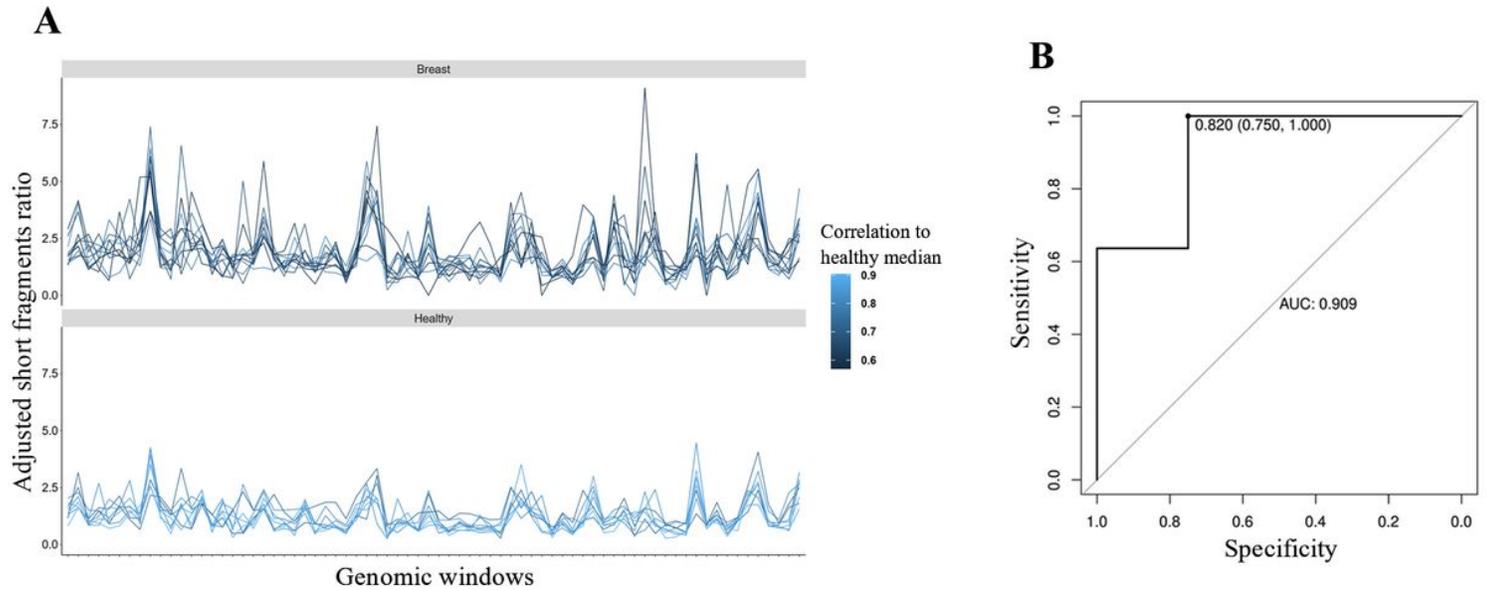


Figure 6

Detection of breast cancer using DMRs-dependent cfDNA fragmentation profile. (A) Input-adjusted short fragments ratio was depicted for hypomethylated genomic windows, individual profile was colored according to their Pearson correlation to the healthy median in each genomic window. (B) Receiver operator characteristics for breast cancer detection using correlation assessment of DMRs-dependent cfDNA fragmentation profile. AUC=0.909; 95% CI (0.771-1.000). Healthy, healthy individuals; Breast, patients with breast cancer.

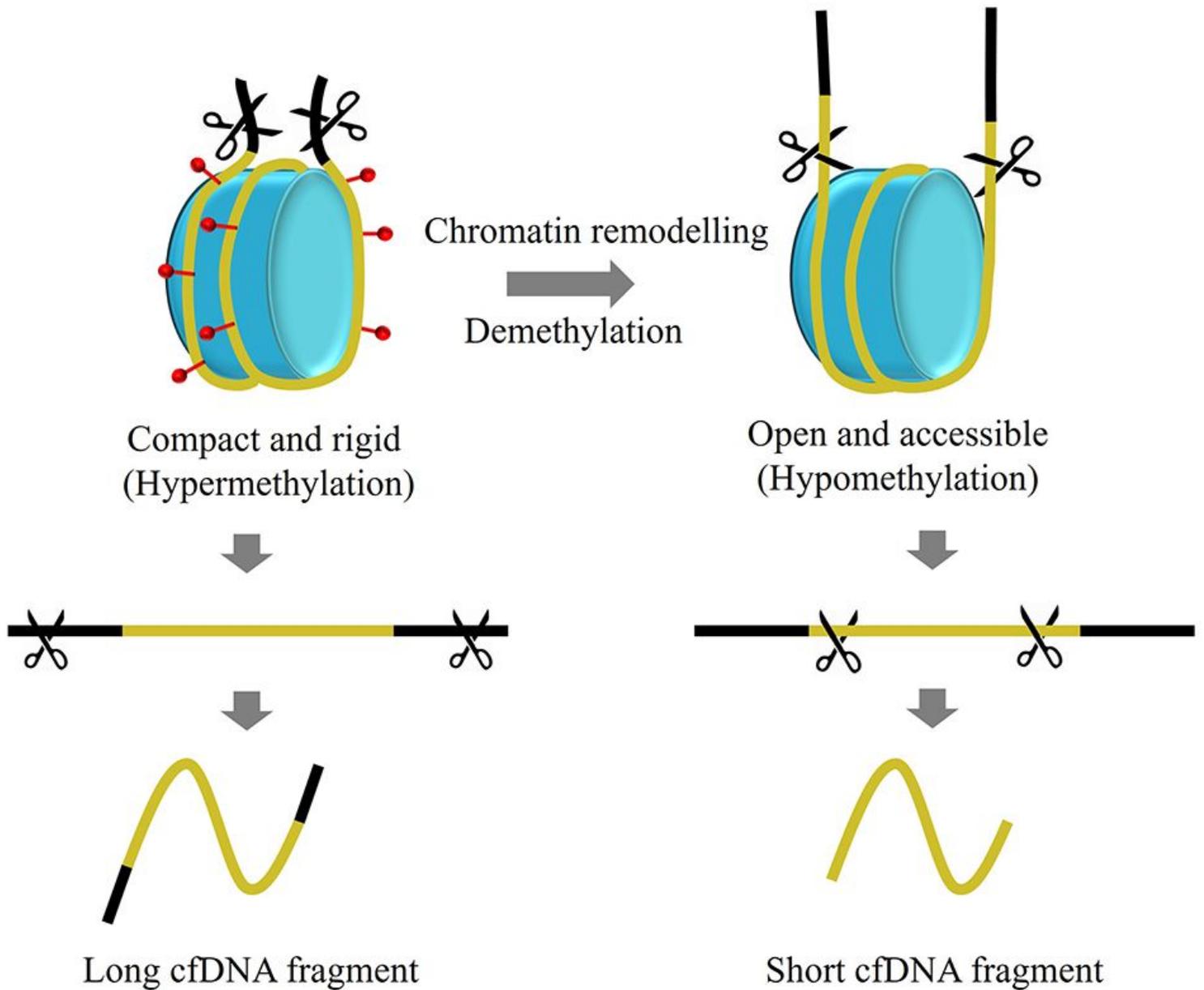


Figure 7

Illustration of the potential relationship between cfDNA methylation and fragment size. Nucleosomes with wrapped DNA (yellow line) exist as compact and rigid structure under normal physiological conditions, cfDNA is released through nuclease digestion (scissors) of the linker sequences (black line) and thus produce long cfDNA fragments. During cancer development, chromatin remodeling and DNA demethylation lead to global hypomethylation, and result open and accessible wrapped DNA in nucleosomes, therefore nuclease digestion during apoptosis produce short cfDNA fragments.