

ORIGINAL ARTICLE

Whole-genome circulating tumor DNA methylation landscape reveals sensitive biomarkers of breast cancer

Luo Hai¹  | Lingyu Li¹ | Zongzhi Liu^{1,3,4}  | Zhongsheng Tong^{2,*} | Yingli Sun^{1,3,4,*}

¹Central Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

²Department of Breast Oncology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

³University of Chinese Academy of Sciences, Beijing, China

⁴CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

*Correspondence

Yingli Sun, Central Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, China.
Email: scaroll99@hotmail.com

Zhongsheng Tong, Department of Breast Oncology, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China.
Email: tongzhongsheng@tjmuch.com

Luo Hai, Lingyu Li, and Zongzhi Liu contributed equally to this work.

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2019YFC1315701; National Natural Science Foundation of China, Grant/Award Number: 22005343; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academic of Medical Sciences and Peking Union Medical College, Shenzhen Research Project, Grant/Award Number: SZ2020ZD004; Shenzhen Science and Technology Program, Grant/Award Number: KCXFZ20201221173008022; Sanming

Abstract

The changes in circulating tumor DNA (ctDNA) methylation are believed to be early events in breast cancer initiation, which makes them suitable as promising biomarkers for early diagnosis. However, applying ctDNA in breast cancer early diagnosis remains highly challenging due to the contamination of background DNA from blood and low DNA methylation signals. Here, we report an improved way to extract ctDNA, reduce background contamination, and build a whole-genome bisulfite sequencing (WGBS) library from different stages of breast cancer. We first compared the DNA methylation data of 74 breast cancer patients with those of seven normal controls to screen candidate methylation CpG site biomarkers for breast cancer diagnosis. The obtained 26 candidate ctDNA methylation biomarkers produced high accuracy in breast cancer patients (area under the curve [AUC] = 0.889; sensitivity: 100%; specificity: 75%). Furthermore, we revealed potential ctDNA methylated CpG sites for detecting early-stage breast cancer (AUC = 0.783; sensitivity: 93.44%; specificity: 50%). In addition, different subtypes of breast cancer could be well distinguished by the ctDNA methylome, which was obtained through our improved ctDNA-WGBS method. Overall, we identified high specificity and sensitivity breast cancer-specific methylation CpG site biomarkers, and they will be expected to have the potential to be translated to clinical practice.

KEYWORDS

biomarkers, breast cancer, circulating tumor DNA, early detection

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *MedComm* published by Sichuan International Medical Exchange & Promotion Association (SCIMEA) and John Wiley & Sons Australia, Ltd.

Project of Medicine in Shenzhen,
Grant/Award Numbers: SZSM201812062,
SZSM201612097; Shenzhen Key Medical
Discipline Construction Fund,
Grant/Award Number: SZXK075

1 | INTRODUCTION

Breast cancer is the most common cancer with the second-highest cancer-related morbidity in women worldwide.¹ Early diagnosis followed by timely treatment can significantly improve the survival of breast cancer patients. To date, clinical methods for breast cancer detection include medical imaging detection (e.g., ultrasonic testing, X-ray imaging, computed tomography, magnetic resonance imaging, positron emission tomography-computed tomography),^{2–6} serum antigen protein markers (e.g., cancer antigen 15-3, carcinoembryonic antigen, and cancer antigen 125),⁷ and tissue biopsy.⁸ These methods have their advantages but also have inevitable disadvantages. For example, medical imaging clearly shows the morphology and location of tumor tissues. However, imaging detection may cause harm to patients when using contrast agents and high-energy rays, and medical imaging detection usually lags behind tumor progression.⁹ Serum antigen protein markers have broad applications. Nevertheless, those serum biomarkers may underestimate tumor heterogeneity, resulting in higher misdiagnosis rates for breast cancer.¹⁰ At present, tissue biopsy is the “gold standard” method. However, biopsy is expensive and invasive and has an unignorable false-negative rate in small tumors.¹¹ Additionally, punctures in tumors might also lead to postoperative bleeding, pain, and infection.¹² Meanwhile, single-site biopsy cannot reveal intratumor and multimetastatic genetic and phenotypic heterogeneity in breast cancer processing, resulting in the loss of potential therapeutic targets and opportunities.¹³ Overall, conventional biomarkers have been applied for a long time in clinical diagnosis and have the value of continued application. Nevertheless, there is still an urgent need to explore novel biomarkers with noninvasive sampling, high specificity and sensitivity for early breast cancer diagnosis and classification.

Tumor tissue can release one- or two-strand DNA fragments named circulating tumor DNA (ctDNA) into the body fluid, which offers a new test approach.¹⁴ There are many apparent advantages of ctDNA as a surrogate for current clinical methods: (i) Blood collection with ctDNA is noninvasive. (ii) The 2-hour half-life of ctDNA in plasma allows it to be used for real-time and dynamic monitoring of tumor progression.¹⁵ (iii) ctDNA can reflect relapse after surgery ahead of medical imaging for several months.¹⁶

(iv) ctDNA detection shrinks the bias caused by intratumoral genetic heterogeneity. Because ctDNA has the above advantages, many attempts are ongoing to use ctDNA as an early diagnostic biomarker of breast cancer.^{17–19}

DNA methylation has been reported to play essential roles in tumor occurrence and development.^{20–22} A previous study demonstrated that the accuracy of DNA methylation in predicting the risk of breast cancer is better than that of copy number variants by tissue.²³ In addition to genomic variations, an increasing number of studies have reported the potential application of ctDNA methylation as a biomarker of various cancers, including breast cancer,²⁴ hepatocellular cancer,²⁵ lung cancer,²⁶ and colon cancer.²⁷ However, routine whole-genome bisulfite sequencing (WGBS) is difficult to detect, limited by the low concentration of ctDNA in blood and background contamination.²⁸ Here, we improved the ctDNA extraction method, and only fragments of approximately 160–180 bp were applied in the subsequent ctDNA WGBS library construction. We also increased the sequencing depths to obtain better signals for data analysis. With the above efforts, we could obtain the whole-genome DNA methylation landscape of breast cancer and reveal biomarkers for breast cancer, including early-stage breast cancer. In addition, different subtypes of breast cancer could be distinguished by the ctDNA methylome. Our work showed that the ctDNA methylome could be a promising noninvasive method to detect breast cancer.

2 | RESULTS

2.1 | An improved method was developed to obtain ctDNA with less background contamination and a high-quality WGBS library for sequencing

We developed an improved method to build a WGBS library in trace quantities of ctDNA. We utilized the selective binding properties of the silica membrane to extract ctDNA from plasma using QIAamp circulating nucleic acid kits (Figure 1A). First, the plasma samples were lysed in an optimized buffer and adjusted to binding conditions. Then, the samples were loaded directly onto a spin column. In this step, ctDNA was bound to the silica membrane, and contaminants were completely removed in wash steps.

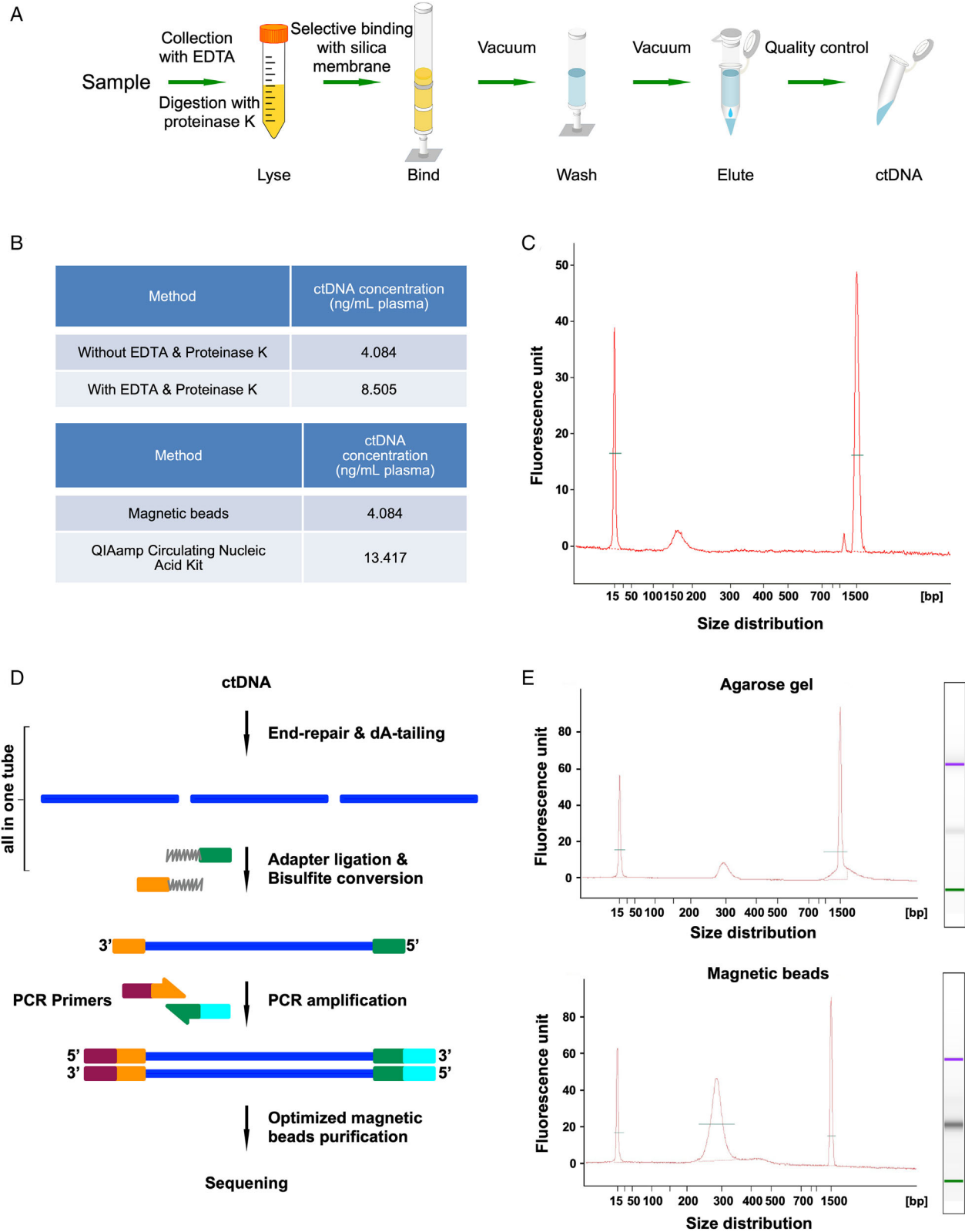


FIGURE 1 Optimization of circulating tumor DNA (ctDNA) extraction and whole-genome bisulfite sequencing (WGBS) library preparation. (A) The workflow shows the processing of ctDNA extraction from plasma. (B) The comparison of ctDNA extraction method: with EDTA and proteinase K versus without EDTA and proteinase K; magnetic beads method versus QIAamp circulating nucleic acid kit method. (C) Extracted plasma ctDNA has sharp peaks of approximately 160–180 bp, indicating that the obtained DNA fragments are high-quality ctDNA. (D) The workflow shows the process of ctDNA methylation library construction. (E) The head-to-head comparison showed that the magnetic bead method significantly increased the recovery ratio of the resulting library compared to the agarose gel method

Finally, pure ctDNA was eluted in small volumes of a low-salt buffer for downstream applications. Meanwhile, we optimized the extraction method of ctDNA and found that higher amounts of ctDNA could be extracted by adding EDTA and proteinase K (Figure 1B). The explanation for the improvement in ctDNA extraction performance might be that the addition of EDTA could inhibit the degradation of ctDNA by nucleases in blood. In addition, proteinase K could improve ctDNA yield by releasing protein-bound ctDNA. Moreover, the ctDNA amount obtained was higher using the QIAamp circulating nucleic acid kit-based protocol than the conventional magnetic-bead-based protocol. The quality control analysis demonstrated that the DNA fragments have sharp peaks of approximately 160–180 bp (Figure 1C), indicating that our method could obtain high-quality ctDNA.

As shown in Figure 1D, purified ctDNA was ligated with an adapter and converted using bisulfate with an “all-in-one tube” operation. Specifically, the process of end repair, dA-tailing, adapter ligation, and bisulfite conversion was performed in the same tube, considering fewer ctDNA requirements and material losses. Then, the fragments were completed by PCR amplification, followed by purification. We optimized the purification method of the ctDNA library by comparing the agarose gel extraction method with the magnetic bead method. The head-to-head comparison showed that the magnetic bead method significantly increased the recovery ratio of the resulting library (Figure 1E). Bisulfite treatment of ctDNA led to the conversion of unmodified cytosines to uracil while keeping 5-methylcytosine unchanged, which could be mapped at a single base resolution after PCR and sequencing. With the above efforts, our ctDNA-WGBS method could be applied for whole-genome base-level resolution detection of 5-methylcytosine in ctDNA from breast cancer patients. Eventually, we could obtain the whole-genome DNA methylation landscape of breast cancer and reveal potential biomarkers.

2.2 | ctDNA methylation biomarkers enabled sensitive detection of breast cancer

To validate the efficacy of the method, we collected breast cancer samples and healthy controls (Table S1). We first analyzed the genome-wide methylation patterns to compare the differences in the global methylation level and distribution of ctDNA from breast cancer patients and healthy controls. The results showed that the overall methylation level had no significant differences (Figure 2A). By dividing the methylation beta-level (0–1) into 10 intervals, the results revealed that the breast cancer patients and healthy controls showed similar distribution patterns (Figure 2B).

Then, we tried to compare the ctDNA methylation data of breast cancer patients with healthy controls to screen candidate CpG sites. Feature selection was performed by least absolute shrinkage and selection operator (LASSO)-penalized logistic regression, resulting in 26 ctDNA methylation CpG sites as potential biomarkers for breast cancer diagnosis. The hierarchical clustering results suggested a precise classification of healthy individuals ($n = 7$) and patients ($n = 74$) in the training dataset using the methylation levels of these CpG sites (Figure 3A). The receiver operating characteristic (ROC) curve demonstrated that the sensitivity, specificity, and area under the curve (AUC) were 100%, 75%, and 0.889, respectively (Figure 3B). Meanwhile, the box plot shows the position of 26 potential biomarkers (Figure 3C).

2.3 | ctDNA methylation biomarkers enabled the sensitive diagnosis of early-stage breast cancer

With these data in hand, we then decided to evaluate the early diagnostic ability of ctDNA methylation CpG sites. Specifically, we recruited seven early-stage breast cancer patients from the Tianjin Medical University Cancer Institute and Hospital (TMUCIH), and the healthy control groups enrolled age-matched Chinese women who were confirmed by physical examination. The heatmap results suggested a precise classification of healthy persons and early breast cancer patients (Figure 4A). Then, feature selection was performed by a random forest algorithm, resulting in 10 CpG sites as potential biomarkers for the early diagnosis of breast cancer (Figure 4B). The principal component analysis (PCA) data showed that healthy individuals and early-stage breast cancer patients were precisely classified (Figure 4C). Furthermore, we validated the identified biomarkers for early diagnosis using independent cohorts named test set 2, which comprised 12 Chinese female patients with early-stage breast cancer recruited from TMUCIH between May 2018 and October 2018. The ROC curve demonstrated that the AUC was 0.783, and the sensitivity and specificity were 93.44% and 50%, respectively (Figure 4D). Collectively, these results show that the 10 ctDNA methylation biomarkers obtained could sensitively distinguish early-stage breast cancer.

2.4 | ctDNA methylation biomarkers distinguished different subtypes of breast cancer

The heterogeneity of breast cancer suggests that the tumor may consist of phenotypically different cancer cell

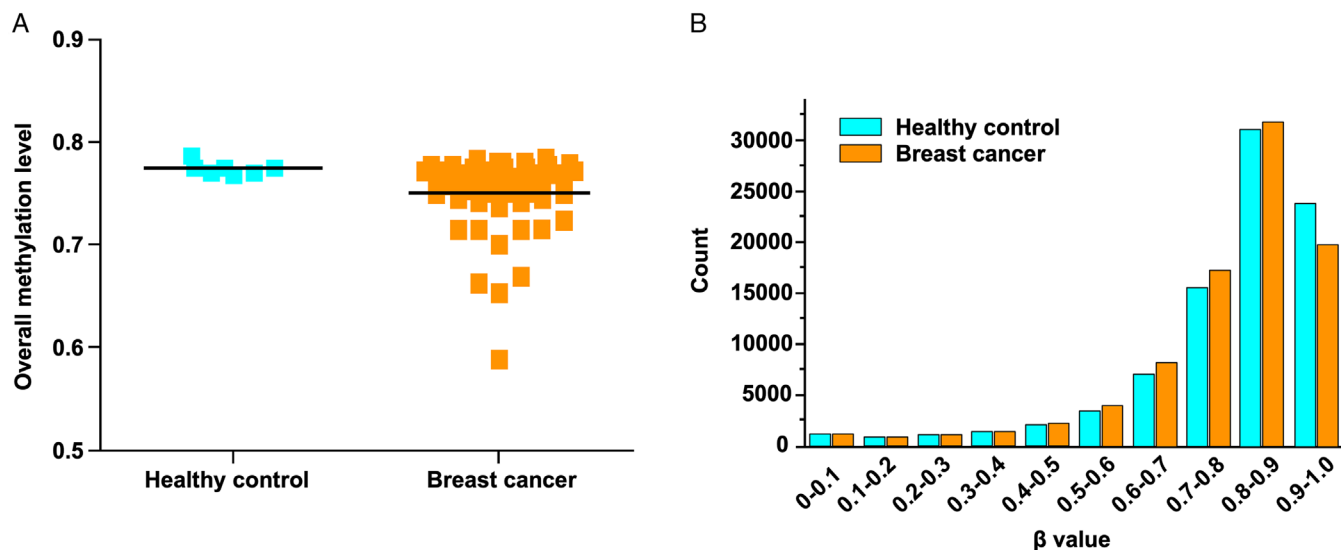


FIGURE 2 The vertical scatter plot and histogram show the whole-genome DNA methylation level (A) and distribution (B) of circulating tumor DNA (ctDNA) from healthy controls and breast cancer patients

populations with diverse properties and expressions of various functional molecules. According to hormone and epidermal growth factor receptor expression status, breast cancer can be divided into several molecular subtypes: triple-negative breast cancer (TNBC), luminal A (LA), luminal B (LB), human epidermal growth factor receptor 2 (HER2) enriched, normal-like, and so forth.

Encouraged by sensitive and specific early-stage breast cancer detection, we constructed three breast cancer datasets to validate the subtyping ability of differentially methylated CpG sites, including TNBC patients and non-TNBC controls, LA patients and non-LA controls, and LB patients and non-LB controls. For each dataset, we selected particular differentially methylated CpG sites for analysis. First, the heatmap plot showed that the obtained differentially methylated CpG sites could be used to divide the TNBC patients ($n = 15$) and non-TNBC controls ($n = 30$) into two groups (Figure 5A). Then, the corresponding CpG sites performed well for subtyping LA patients ($n = 3$) and non-LA controls ($n = 22$) (Figure 5B). Finally, the methylation level of CpG sites in ctDNA could also efficiently discriminate between LB patients ($n = 11$) and non-LB controls ($n = 14$) (Figure 5C). Collectively, these results suggested that the differentially methylated CpG sites we identified could be used to subtype breast cancer by heatmap analysis.

Based on the encouraging outcomes, we performed PCA to further assess the capability of the as-obtained ctDNA differentially methylated CpG sites to classify the subtypes of breast cancer. As shown in Figure 5D, the results indicated a precise classification of TNBC patients and non-TNBC controls. In addition, two distinct clusters of LA patients versus non-LA controls were sep-

arated (Figure 5E). Meanwhile, the PCA demonstrated that the LB patients were clustered together, and non-LB controls were grouped at the top of the coordinate plot (Figure 5F).

To further validate our obtained biomarkers, we conducted external validation of the identified CpG sites using independent The Cancer Genome Atlas (TCGA) dataset. The t -distributed stochastic neighbor embedding analysis indicated that TNBC ($n = 15$), LA ($n = 12$), HER2-positive (HER2⁺) LB ($n = 10$), HER2-negative (HER2⁻) LB ($n = 26$), and HER2 ($n = 9$) patients could be well distinguished (Figure 5G). Although further systematic evaluations are still needed, these preliminary results indicate that our study provides an unprecedented method to develop novel biomarkers complementary to traditional diagnostics for classifying various subtypes of breast cancer.

3 | DISCUSSION

Breast cancer is the leading cause of death for women worldwide. X-ray mammography has been used to screen breast cancer in many countries and has been evaluated as the best method thus far. However, the reports of overdiagnoses of the method are increasing.²⁹ In addition, the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) results provide negligible molecular information. Hence, it could not distinguish the precise details of the disease, for example, benign or malignant, early stage or other stages, TNBC or other subtypes, and so forth. More importantly, it is not very applicable in Asia because the structure of Asian women is quite different from women in Western countries. Therefore, it is

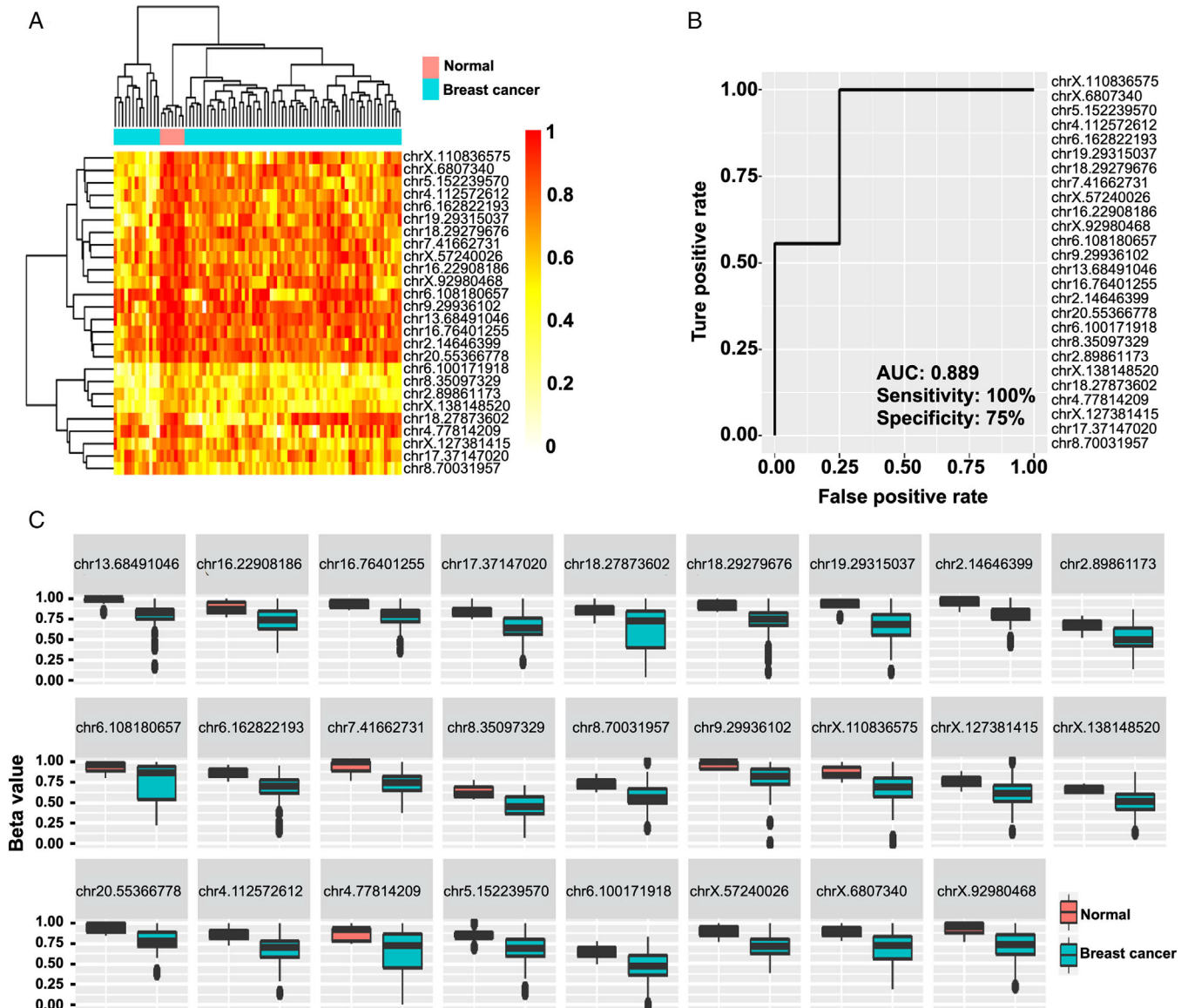


FIGURE 3 Methylation CpG sites selected for breast cancer diagnosis. (A) Heatmap of the DNA methylation levels of the methylated CpG sites in normal controls and breast cancer patients (mean difference >0.15 , $p < 0.05$, standard deviation <0.1). (B) Receiver operating characteristic curve of a predictive model comprising 26 potential biomarkers. (C) The box plot shows the position of 26 candidate circulating tumor DNA (ctDNA) methylation CpG site biomarkers

critical to develop a noninvasive method for breast cancer diagnosis.

Among all noninvasive diagnostic methods, ctDNA methylation detection is the most promising method. Epigenetic abnormalities often lead to cancer, and many researchers have reported that 5-methylcytosine is very important in detecting cancer.³⁰ Furthermore, DNA methylation provides several more advantages than copy number variants and somatic mutation.³¹ ctDNA has many advantages and is widely studied.³² However, applying ctDNA in breast cancer early diagnosis remains highly challenging due to the contamination of background DNA from blood and low DNA methylation signals. To over-

come the above obstacles, some researchers mixed several samples into one sample to obtain ctDNA levels to build the library.³³ Unfortunately, those studies lost many messages and could not acquire individual differences.

We developed an improved ctDNA-WGBS method to evenly profile whole-genome methylation patterns from trace quantities of ctDNA in only 200 μ l of plasma compared with the standard 5–20 ml of plasma: (i) Streck tubes were used to collect blood. The preservatives contained in Streck tubes stabilize nucleated blood cells, prevent the release of cellular genomic DNA, inhibit nuclease-mediated degradation of ctDNA, and help improve the stability of ctDNA. (ii) A two-step centrifugation was used:

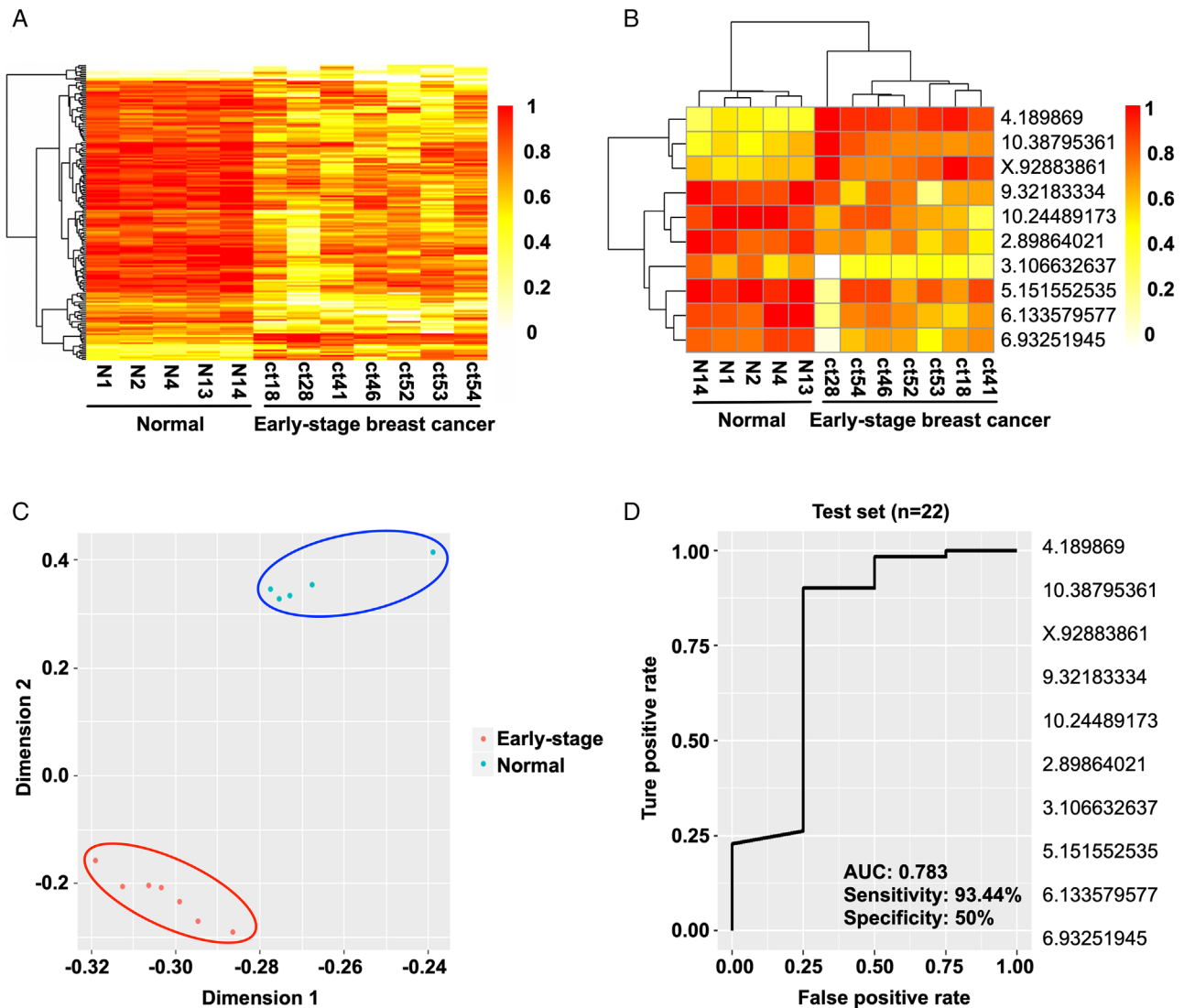


FIGURE 4 Methylation CpG sites in circulating tumor DNA (ctDNA) fragments selected for the early diagnosis of breast cancer. (A) Heatmap derived from cluster analysis of CpG sites (mean difference >0.15 , $p < 0.05$, standard deviation <0.1). (B) Heatmap of the top 10 CpG sites between normal controls and early-stage breast cancer samples. (C) Principal component analysis shows that these CpG sites can distinguish between normal controls and early-stage breast cancer patients. (D) Receiver operating characteristic curve suggested the classification of normal controls and early-stage breast cancer patients using the same 10 differentially methylated CpG sites that could be biomarkers

EDTA was added at the first centrifugation, and proteinase K was added at the second centrifugation. EDTA could inhibit the degradation of ctDNA by nucleases in blood, and proteinase K could improve ctDNA yield by releasing protein-bound ctDNA. (iii) We compared the ctDNA separation method and found that the amount of ctDNA obtained was higher using the silica membrane-based protocol than the conventional magnetic-bead-based protocol. (iv) During the extraction process, quality control needs to be performed to check whether the size of the enriched DNA fragments is appropriate. Only fragments of approximately 160–180 bp were applied in the subsequent ctDNA WGBS library construction. (v) To

achieve less ctDNA requirement and reduce material loss, purified ctDNA was ligated with an adapter and converted using bisulfate with an “all-in-one tube” operation. That is, the process of end repair, dA-tailing, adapter ligation, and bisulfite conversion was performed in the same tube. (vi) We increased the sequencing depths to obtain better signals for data analysis. With the above efforts, we could reduce background contamination and build a WGBS library from trace quantities of ctDNA.

Our analysis focused on detecting and classifying breast cancer from ctDNA to identify potential biomarkers at the whole-genome scale. The results showed that ctDNA methylation is a potential biomarker for the early

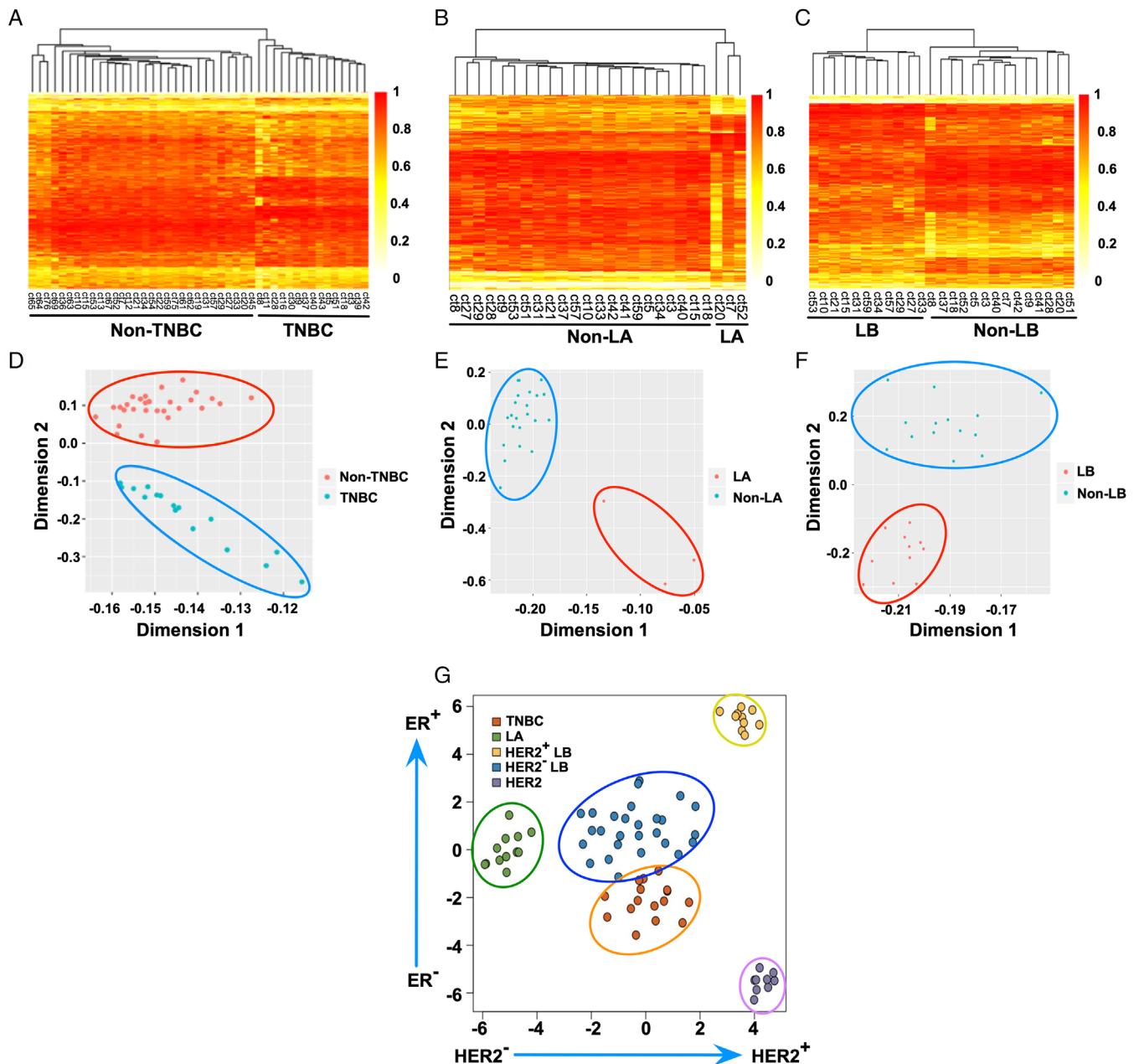


FIGURE 5 Methylation CpG sites in circulating tumor DNA (ctDNA) fragments selected for subtyping of breast cancer. (A) Heatmap of the DNA methylation levels of methylated CpG sites in triple-negative breast cancer (TNBC) and non-TNBC. (B) Heatmap of the DNA methylation levels of methylated sites in luminal A (LA) and non-LA. (C) Heatmap of the DNA methylation levels of methylated CpG sites in luminal B (LB) and non-LB. (D) Principal component analysis of TNBC and non-TNBC. (E) Principal component analysis of LA and non-LA. (F) Principal component analysis of LB and non-LB. (G) *t*-Distributed stochastic neighbor embedding plot suggested the classification of TNBC, LA, human epidermal growth factor receptor 2 (HER2)⁺ LB, HER2⁻ LB, and HER2 patients using independent The Cancer Genome Atlas dataset data

detection and molecular subtyping of breast cancer. Our work showed that different extraction methods could significantly influence the quality of ctDNA and are critical for identifying biomarkers with the ctDNA methylome. It is also very important to improve library construction to obtain enough peak-enriched ctDNA for sequencing. In addition, increasing the sequencing depth is critical for

signal amplification. With all these efforts, we could significantly improve the sensitivity and specificity of breast cancer prediction with the ctDNA methylome compared to recently reported work.³⁴

According to industry research institutions, the cancer screening and early detection market is estimated to be more than 162 billion RMB yuan in China.³⁵ We

developed a noninvasive liquid biopsy technology with high breast cancer detection sensitivity while significantly reducing blood sample usage. The technology can significantly reduce the amount of blood used for testing, and the sample consumption is only one-tenth of that of commercial kits, which can improve patient compliance. Moreover, the noninvasive liquid biopsy strategy combined with conventional biomarkers might reduce the frequency of unnecessary biopsies in the clinic. The protocol was effective at developing novel sensitive and specific biomarkers for the diagnosis, early detection, and molecular subtyping of breast cancer. We anticipate that the improved ctDNA-WGBS method and diagnostic biomarkers of breast cancer have substantial clinical translation potential.

Several limitations also need to be acknowledged in this study. First, the blood samples from breast cancer patients and healthy controls were relatively small. Furthermore, the lack of multicenter patient cohorts for validation during the screening of breast cancer ctDNA methylation biomarkers may lead to some potential bias. We will focus on investigating the optimal ctDNA methylation biomarkers with large-scale ctDNA samples from multicenter patient cohorts in future work.

4 | MATERIALS AND METHODS

4.1 | Sample collection

We recruited two cohorts of Chinese breast cancer patients from the Tianjin Medical University Cancer Institute and Hospital (TMUCIH): (i) Test set 1 consisted of early-stage ($n = 7$) and advanced-stage ($n = 44$) breast cancer female patients of Chinese descent from the TMUCIH between December 2016 and December 2017. In sum, we enrolled 51 Chinese female patients with early/advanced breast cancer (50.87 ± 11.05 as the mean age \pm standard deviation [SD]) in test set 1. (ii) Test set 2 comprised Chinese female patients with early-stage ($n = 12$) and advanced-stage ($n = 11$) breast cancer recruited from the TMUCIH between May 2018 and October 2018. For test set 2, 23 Chinese female patients with early/advanced breast cancer (53.47 ± 8.88 as the mean age \pm SD) were recruited. Collectively, there were 74 female patients with breast cancer enrolled in this study.

The healthy control groups enrolled age-matched Chinese women without cancer from the Beijing Institute of Genomics (BIG), confirmed by physical examination, ultrasound scans of breast, and mammographic screening. Informed consent or a waiver of consent was obtained from each participant. In total, seven healthy women were recruited into healthy control groups.

4.2 | Phenotypic evaluation

The diagnosis of breast cancer was histologically confirmed. We conducted phenotypic analyses of age, menopausal status, stage, pathology features, and molecular subtype in the two cohorts. Information on estrogen receptor (ER) and progesterone receptor (PR) status was assessed by immunohistochemistry (IHC) and extracted from pathology reports. Specifically, the paraffin samples were cut into $4\text{-}\mu\text{e}$ paraffin samples were cut into hology features, and molecular subtype in the two cohorts. Information on estrogen receptor (E mol/L citrate buffer (pH = 6.0) for 10 min for antigen retrieval. After cooling, the slides were washed with PBS (pH = 7.4) and incubated with 0.3% H_2O_2 solution for 10 min to block endogenous peroxidase activity. Primary polyclonal antibodies against ER, PR, and HER2 were applied to the samples and incubated overnight at 4al r subtype in the two cohorts. Informath PBS, and the antibody reaction was visualized using a fresh substrate solution containing diaminobenzidine. Images were captured by a Leica DC500 camera (Solms, Germany) on a microscope equipped with Leica DMRA2 fluorescent optics under 10 orts. Innification, and the expression level was defined as the mean density of ER-, PR-, and HER2-specific staining/area. Tumors were classified as ER/PR-negative if IHC staining of tumor cell nuclei was less than 1% reactivity. HER2 was assessed through IHC or fluorescence in situ hybridization (FISH) if the IHC results were borderline. The clinical grouping of molecular subtypes was defined by the status of hormone receptor and HER2 according to the St. Gallen 2017 criteria.³⁶

4.3 | Circulating tumor DNA extraction

Whole blood of healthy and cancer samples was collected using ctDNA blood collection tubes (Strect, USA). Plasma was then separated by centrifugation at $1900 \times g$ for 10 min and $18,000 \times g$ for 10 min at room temperature with EDTA and protinease K. ctDNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit (55114) or magnetic beads method according to the manufacturer's instructions. The ctDNA concentration was determined using a Qubit dsDNA HS Assay Kit (Life Technologies, Q32851). The extracted ctDNA was stored at -80°C for further library construction.

4.4 | ctDNA methylation library construction

As shown in Figure 1D, we used the DNA to build the methylation library. We mixed 20 ng ctDNA and 1/1000–5/

1000 ng unmethylated lambda DNA (D1521, Promega). Unmethylated lambda DNA was sheared by a Covaris S220 instrument (Life Technologies, Thermo Fisher Scientific), and the size of fragments was approximately 200 bp. First, DNA fragments were synthesized via end-repair and A-tailing reactions. Second, NEBNext Ultra Ligation Module (E7445S/L, NEB) reagent was used to ligate the methylation adaptor. Then, we used the EZ DNA Methylation-Gold Kit (D5005, Zymo) to bisulfite-converted DNA according to the manufacturer's recommendations. Next, DNA was amplified by PCR, which was carried out using primers. The primers came from Illumina next-generation sequencing. PCR was performed to amplify the fragments. DNA gel extraction and magnetic bead extraction were performed to compare the different influences of library quality. Finally, we selected effective captures (~290 bp) as the resulting library using Agencourt AMPure XP beads (Beckman Coulter). All binding and washing procedures were performed at room temperature. The purity of the libraries was analyzed by Qubit dsDNA HS assay (Thermo Fisher). A total of 100 G data were obtained for each sample.

4.5 | Quality control and mapping of WGBS data

The raw sequencing reads (fastq format) were trimmed to remove sequencing adapters, amplification primers, and low-quality bases in read ends using trim_galore (version 0.4.2) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). After quality control, the reads were mapped to the human reference genome (hg19) using Bismark (version 0.14.2) with default parameters.³⁷ Next, the PCR duplicates were also removed using Bismark (version 0.14.2) with default parameters.³⁷ The bisulfite conversion rate was calculated by the spike-in of totally unmethylated lambda DNA. Only the mapped and duplicate-removed reads (bam format) were used for the subsequent bioinformatics analysis.

4.6 | Data processing and analysis

We sequenced DNA methylation libraries by using X-ten systems (Illumina). The total reads were assessed by FASTQC (version 0.11.8). The raw sequencing reads were then cleaned by Trim Galore (version 0.4.5) to clip sequencing adapters and low-quality reads. Subsequently, the trimmed reads were aligned against the human reference genome using the Bismark (version 0.14.3) tool. Finally, differential methylation analysis of the DNA methylation sequencing data was performed by the limma package. For each dataset, the top CpG sites with *p*-value

<0.05 and absolute mean methylation difference >0.15 were selected. Sites were annotated by PeakAnnotator (version 1.4).

ACKNOWLEDGMENTS

We thank Professor Nan Wu from Peking Union Medical College Hospital. We are also grateful to our laboratory colleagues for their assistance with the experiments and in the preparation of the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

ETHIC STATEMENT

Our study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (approved protocol no. 20210205). All individuals were adequately informed and signed an informed consent form before participating in the study.

AUTHOR CONTRIBUTIONS

Luo Hai participated in data analysis, original draft writing, and review & editing. Ling yu Li and Zongzhi Liu participated in sample collection, data analysis, and paper writing. Zhongsheng Tong and Yingli Sun designed the experiment and revised the article. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The code and data that support the findings of this study are openly available in GitHub at <https://github.com/zhq921/cfWGBS-bioinfo-pip>. The raw sequencing data reported in this manuscript are publicly available at the Genome Sequence Archive under accession number CRA001142 (<https://ngdc.cncb.ac.cn/gsa/browse/CRA001142>). All the data and materials are available from the corresponding author upon reasonable request.

ORCID

Luo Hai  <https://orcid.org/0000-0001-8084-3742>

Zongzhi Liu  <https://orcid.org/0000-0002-1109-5573>

REFERENCES

1. Cao W, Chen HD, Yu YW, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J*. 2021;134(7):783-791.
2. Singh BK, Verma K, Panigrahi L, et al. Integrating radiologist feedback with computer aided diagnostic systems for breast cancer risk prediction in ultrasonic images: An experimental investigation in machine learning paradigm. *Expert Syst Appl*. 2017;90(30):209-223. <https://doi.org/10.1016/j.eswa.2017.08.020>
3. Nieves LM, Hsu JC, Lau KC, et al. Silver telluride nanoparticles as biocompatible and enhanced contrast agents for X-ray imaging: an in vivo breast cancer screening study. *Nanoscale*. 2021;13(1):163-174.

4. Jafari SH, Saadatpour Z, Salmaninejad A, et al. Breast cancer diagnosis: imaging techniques and biochemical markers. *J Cell Physiol.* 2018;233(7):5200-5213.
5. Kuhl CK. Abbreviated magnetic resonance imaging (MRI) for breast cancer screening: rationale, concept, and transfer to clinical practice. *Annu Rev Med.* 2019;70:501-519.
6. He M, Liu C, Shi Q, et al. The predictive value of early changes in 18F-fluoroestradiol positron emission tomography/computed tomography during fulvestrant 500 mg therapy in patients with estrogen receptor-positive metastatic breast cancer. *Oncologist.* 2020;25(11):927-936.
7. Loke SY, Lee ASG. The future of blood-based biomarkers for the early detection of breast cancer. *Eur J Cancer.* 2018;92:54-68.
8. Park HL, Hong J. Vacuum-assisted breast biopsy for breast cancer. *Gland Surg.* 2014;3(2):120-127.
9. Nguyen T, Levy G, Poncelet E, et al. Overview of digital breast tomosynthesis: clinical cases, benefits and disadvantages. *Diagn Interv Imaging.* 2015;96(9):843-859.
10. Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value. *Clin Chem.* 2006;52(3):345-351.
11. Buono G, Gerratana L, Bulfoni M, et al. Circulating tumor DNA analysis in breast cancer: is it ready for prime-time. *Cancer Treat Rev.* 2019;73:73-83.
12. Marrugo-Ramírez J, Mir M, Samitier J. Blood-based cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy. *Int J Mol Sci.* 2018;19(10):2877.
13. He Z, Chen Z, Tan M, et al. A review on methods for diagnosis of breast cancer cells and tissues. *Cell Prolif.* 2020;53(7):e12822.
14. Dawson SJ, Tsui DWY, Murtaza M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *New Engl J Med.* 2013;368(13):1199-1209.
15. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med.* 2008;14(9):985-990.
16. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature.* 2017;545(7655):446-451.
17. Rohanizadegan M. Analysis of circulating tumor DNA in breast cancer as a diagnostic and prognostic biomarker. *Cancer Genet.* 2018;228:159-168.
18. Shang M, Chang C, Pei Y, et al. Potential management of circulating tumor DNA as a biomarker in triple-negative breast cancer. *J Cancer.* 2018;9(24):4627-4634.
19. Fiala C, Diamandis EP. Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. *BMC Med.* 2018;16(1):166.
20. Sina AAI, Carrascosa LG, Liang Z, et al. Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker. *Nat Commun.* 2018;9:4915.
21. Guo S, Diep D, Plongthongkum N, et al. 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):635-642.
22. Zhang H, Dong P, Guo S, et al. Hypomethylation in HBV integration regions aids non-invasive surveillance to hepatocellular carcinoma by low-pass genome-wide bisulfite sequencing. *BMC Med.* 2020;18:200.
23. Gao Y, Widschwendter M, Teschendorff AE. DNA methylation patterns in normal tissue correlate more strongly with breast cancer status than copy-number variants. *EBioMedicine.* 2018;31:243-252.
24. Tzanikou E, Lianidou E. The potential of ctDNA analysis in breast cancer. *Crit Rev Clin Lab Sci.* 2020;57(1):54-72.
25. Xu R, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater.* 2017;16(11):1155-1161.
26. Liang W, Zhao Y, Huang W, et al. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics.* 2019;9(7):2056-2070.
27. Mitchell SM, Ho T, Brown GS, et al. Evaluation of methylation biomarkers for detection of circulating tumor DNA and application to colorectal cancer. *Genes.* 2016;7(12):125.
28. Luo H, Wei W, Ye Z, et al. Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends Mol Med.* 2021;27:482-500.
29. Welch HG, Prorok PC, O'Malley AJ, et al. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N Engl J Med.* 2016;375(15):1438-1447.
30. Toh TB, Lim JJ, Chow EKH. Epigenetics of hepatocellular carcinoma. *Clin Transl Med.* 2019;8(1):13.
31. Gao Y, Widschwendter M, Teschendorff AE. DNA methylation patterns in normal tissue correlate more strongly with breast cancer status than copy-number variants. *EBioMedicine.* 2018;31:243-252.
32. Jia N, Sun Z, Gao X, et al. Serial monitoring of circulating tumor DNA in patients with metastatic colorectal cancer to predict the therapeutic response. *Front Genet.* 2019;10:470.
33. Legendre C, Gooden GC, Johnson K, et al. Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. *Clin Epigenetics.* 2015;7(1):100.
34. Liu MC, Oxnard GR, Klein EA, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol.* 2020;31(6):745-759.
35. Cao M, Li H, Sun D, et al. Cancer burden of major cancers in China: a need for sustainable actions. *Cancer Commun.* 2020;40(5):205-210.
36. Curigliano G, Burstein HJ, Winer EP, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen international expert consensus conference on the primary therapy of early breast cancer 2017. *Ann Oncol.* 2017;28(8):1700-1712.
37. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571-1572.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hai L, Li L, Liu Z, Tong Z, Sun Y. Whole-genome circulating tumor DNA methylation landscape reveals sensitive biomarkers of breast cancer. *MedComm.* 2022;3:e134.
<https://doi.org/10.1002/mco2.134>