

Genetic landscape of breast cancer and mutation tracking with circulating tumor DNA in Chinese women

Yinghao Wang^{1,*}, Lizhi Lin^{1,*}, Linhong Li^{2,*}, Jialiang Wen¹, Yili Chi¹, Rutian Hao¹, Xuanxuan Dai¹, Yizuo Chen¹, Duping Huang¹, Yili Zhou¹, Jie You¹, Zhiqiang Ye¹, Hao Chen¹, Lingli Jin¹, Danxiang Chen¹, Fan Yang¹, Erjie Xia¹, Xueyan Ma², Fengyu Guo², Yunguang Tong², Min Zheng^{3,&}, Ouchen Wang¹

¹Department of Breast Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, P.R. China

²Omigen, Inc., Hangzhou 310000, P.R. China

³Department of Breast Surgery, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, P.R. China

*Equal contribution

Correspondence to: Ouchen Wang, Min Zheng, Yunguang Tong; email: woc@wmu.edu.cn, zhengmin_wz@163.com, tony.tong@omigen.com

Keywords: breast cancer, digital molecular identifier, circulating tumor DNA, mutation landscape, Chinese women

Received: January 28, 2021

Accepted: March 6, 2021

Published: April 23, 2021

Copyright: © 2021 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Considerable efforts have been devoted to exploring the breast cancer mutational landscape to understand its genetic complexity. However, no studies have yet comprehensively elucidated the molecular characterization of breast tumors in Chinese women. This study aimed to determine the potential clinical utility of peripheral blood assessment for circulating tumor-derived DNA (ctDNA) and comprehensively characterize the female Chinese population's genetic mutational spectrum. We used Omi-Seq to create cancer profiles of 273 patients enrolled at The First Affiliated Hospital of Wenzhou Medical University. The gene landscape results indicate *PIK3CA* and *TP53* as the most frequently detected genes, followed by *ERBB2*, in Chinese breast cancer patients. The accuracy of *ERBB2* copy number variations in tissue/formalin-fixed and paraffin-embedded samples was 95% with 86% sensitivity and 99% specificity. Moreover, mutation numbers varied between different molecular cell-free DNA subtypes, with the basal-like patients harboring a higher number of variants than the luminal patients. Furthermore, ratio changes in the max ctDNA allele fraction highly correlated with clinical response measurements, including cancer relapse and metastasis. Our data demonstrate that ctDNA characterization using the Omi-Seq platform can extend the capacity of personalized clinical cancer management.

INTRODUCTION

Breast cancer is one of the most common cancers and the second leading cause of cancer-related deaths in women worldwide [1, 2]. The incidence of breast cancer in China is approximately 40/100,000, and 70,000 people have died because of breast cancer in 2015 [3]. Breast cancer is a heterogeneous disease, resulting in

complex clinical characteristics and molecular subtypes. Thanks to the development of high-throughput sequencing technology, breast cancer is composed of at least four molecular-specific diseases with different characteristics, clinical behavior, and treatment responses. These intrinsic molecular subtypes are defined as: basal-like, human epidermal growth factor receptor 2 (HER2) enrichment, and luminal A and B

subtypes [4, 5]. Because of intrinsic molecular heterogeneity, response to treatment differs between different molecular subtypes. Hence, accurately characterizing the tumor molecular subtype is necessary to guide optimal treatment decisions and determine patient prognosis [6]. Despite significant advances in treatment, metastasis remains the leading cause of breast cancer-related death; this is because of the associated tumor heterogeneity, owing to clonal evolution, limiting therapeutic efficacy and response time [7–9]. Therefore, evaluating tumor clonal heterogeneity is of significant value for breast cancer patients [10]. However, imaging examination, the main traditional diagnostic method for breast cancer, cannot determine the patient's gene mutations and tumor clonal heterogeneity and cannot facilitate the development of individualized treatment regimens. Of note, next-generation sequencing (NGS) technologies and bioinformatics tools have increased the understanding of breast cancer molecular heterogeneity, thereby providing an avenue for the identification of targeted mutations and subsequent personalized patient treatment strategies [11].

From the first detection of extracellular nucleic acids in peripheral blood in 1948 to the detection and quantification of cancer-related DNA mutations in 1994, liquid biopsies are continually evolving in the field of patient care [12]. Nucleic acids circulating in peripheral blood and originating from tumor masses are designated circulating tumor DNA (ctDNA), and their detection serves as a framework for the non-invasive detection of genomic alterations. Moreover, plasma-derived ctDNA is suitable for real-time monitoring of therapy responses because of its short lifetime (approximately 2 hours) [13]. Several studies have explored and evaluated the potential clinical value of ctDNA in cancer management, including for early detection, auxiliary evaluation of minimal residual disease, treatment monitoring, and drug resistance assessment [14–17].

Although deep sequencing platforms have been widely adopted for investigating subpopulations in complex biological samples, certain inherent limitations related to the relatively low yield of ctDNA and the errors introduced during sample preparation and sequencing interfere are their limitations [18]. Moreover, although molecular identifiers can theoretically limit the prevalence of these associated errors, certain challenges remain. First, synthesizing double-stranded adapters with randomized molecular identifiers is difficult. Second, owing to the difficulty of generating high-quality single molecular identifier (SMI) adaptors, the ligation efficiency may be compromised and require large amounts of input DNA [19]. The Omi-Seq

platform (developed by Omigen, Inc.) may address some of these challenges as it calculates digital molecular identifiers (DMIs) using a standard adapter pool with defined barcodes that can readily synthesize high-quality sequence reads.

In the present study, we used Omi-Seq to create cancer profiles of female Chinese breast cancer patients. The resulting ctDNA genomic mutational profiles were then compared between patients with early and metastatic cancer and different pathological subtypes. This study's primary aim was to define the potential clinical utility of peripheral blood ctDNA detection and analysis and characterize the genetic mutation spectrum within the female Chinese breast cancer population.

RESULTS

Patients' characteristics

A total of 273 breast cancer patients were enrolled in the study (Supplementary Table 1). All patients were women, and the median age was 52 years. The patients were categorized based on tumor subtype, with 77 luminal A (28.2%), 74 luminal B (27.1%), 67 HER2+ (24.5%), and 45 basal-like (16.5%) patients included. Some of the patients had received monotherapy with an aromatase inhibitor (AI), while others had been administered combination AI or fulvestrant and CDK4/6 inhibitor therapy. However, no patients had received prior poly ADP ribose polymerase (PARP) inhibitor therapy.

The ctDNA mutational landscape in Chinese women with breast cancer

High-frequency molecular variants were analyzed in pairs of tumor tissue and plasma. The consistency, defined as harboring at least one common gene mutation site, was evaluated between the tumor tissue and paired plasma samples. The consistency values between tissue and plasma samples in stage I, II, III, and IV breast cancers were 80% (32/40), 80% (36/45), 100% (13/13), and 100% (3/3), respectively. The top three most prevalent mutated genes were as follows: *PIK3CA* (tissue: 42.31%, plasma: 24.04%), *TP53* (tissue: 32.69%, plasma: 43.27%), and *ERBB2* (tissue: 19.23%, plasma: 14.42%). The top three variants were *PIK3CA* p.H1047R (c.3140A>G), *PIK3CA* p.E545K (c.1633G>A), and *PIK3CA* p.E542K (c.1624G>A) (Figure 1). Our findings were consistent with those reported by Andre et al [20].

Next, matched DMI-tagged cell-free DNA (cfDNA) was sequenced using the Omigen 101 genes panel kit at a targeted sequencing depth of 40,000×–50,000×. The

mutation profile was derived from 205 detected female Chinese patients' cfDNA with an overall mutation detection rate of 94.15% (193/205; Figure 2). The allele fraction (AF) detected in our cohort ranged from 0.05%–32.48%. Some of these mutations were actionable, including SNVs, such as *PIK3CA* p.H1047L, p.E545K, and p.E542K, *BRCA1/2* loss-of-function oncogenic mutation, and copy number mutations, such as *ERBB2* amplification. No mutation was detected from the panel of the remaining 12 patients. Moreover, the patients with negative cfDNA tests were primarily in stage I or II of disease, including six patients in stage I and four in stage II (Figure 2).

These data suggest that somatic mutations with frequencies as low as 0.05% could be detected in DMI-tagged cfDNA of early-stage breast cancer patients with a targeted sequencing depth of 40,000×–50,000×.

Mutation landscape across different molecular subtypes

The distribution of mutations in patients was further analyzed based on four molecular subtypes. The prevalence of changes in cancer-related genes/pathways was compatible with that reported in previous studies of treated advanced breast malignancies, including frequent oncogenic mutations in the PI3K pathway and loss-of-function mutations in the DNA damage response and tumor suppressor pathways [21, 22]. Frequently varied genes in all four molecular subtypes are presented in Figure 3. The *PIK3CA* detection rates in basal-like (38.24%, 13/34) were higher than in luminal B (31.67, 19/60), HER2+ (30.61%, 15/49), and luminal A (26.98 %, 17/63) breast cancers (Figure 3A). Meanwhile, *PTEN* mutations were more frequently detected in HER2+ and basal-like subtypes and were less common in the luminal type, with detection rates of

14.29% (7/49), 11.76% (4/34), 9.52% (6/63), and 8.33% (5/60), respectively (Figure 3A). *BRCA1/2* mutations were predominantly detected in basal-like patients (14.71%, 11.76%; Figure 3B), while *TP53* mutations were frequently detected in all four molecular subtypes (38.24%, 36.73%, 31.75%, and 25% for basal-like, HER2+, luminal A, and luminal B, respectively; Figure 3C).

cfDNA yield and ctF (highest cfDNA allele frequency) correlate with cancer stage

To investigate the correlations between cfDNA and corresponding clinical characteristics, we analyzed cfDNA yield in healthy participants and breast cancer patients with different disease stages. Considering that tumors generate a certain amount of free nucleotides into the peripheral blood, the cfDNA yield could represent the amount of released tumor DNA in the patient's body [23, 24]. We found that cfDNA was significantly associated with disease stage ($p = 0.033$, $r = 0.9$; Figure 4A). Meanwhile, ctF, defined as the highest cfDNA allele frequency in patients with multiple alterations, was found to be significantly associated with disease stage ($p = 0.032$, $r = 0.97$; Figure 4B).

Number of ctDNA mutations in different molecular subtypes and disease stages

Cancers with higher tumor mutational burdens (TMBs) are at a higher risk of being recognized by immune cells [25]. Hence, we used the total number of ctDNA non-synonymous coding mutations detected per megabase to calculate TMB and detected the highest TMB in the basal-like subtype, achieving statistical significance in comparison with the luminal subtypes (luminal A: $p = 0.0047$, luminal B: $p = 0.044$; Figure 5B). Our data also

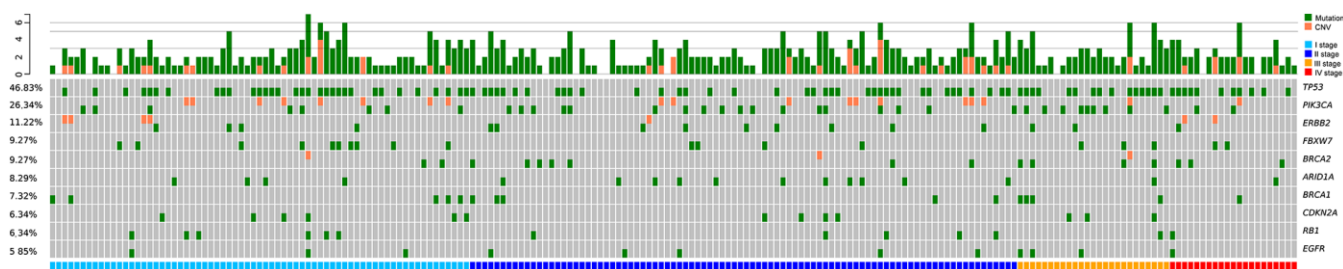


Figure 1. Low-frequency somatic mutations detected in DMI-tagged ctDNA from Chinese breast cancer patients. Mutational profiles derived from DMI-tagged ctDNA from stage I (blue), II (deep blue), III (yellow), and IV (red) breast cancers. Each column represents one patient. Different colors represent different types of mutations. Green and orange colors represent mutations and CNV, respectively. Each row represents one gene. The top bar graph denotes the number of mutations detected in each patient. The sidebar represents the proportion of patients with a mutation in a certain gene. CNV, copy number variation; ctDNA, circulating tumor-derived DNA; DMI, digital molecular identifier.

shows that the total number of ctDNA mutations positively correlates with the breast cancer stage ($p = 0.114$, $r = 0.89$; Figure 5A), although this result was not statistically significant.

NGS detection using tissue samples better reflects *ERBB2* copy number amplification than using blood samples

We next sought to explore the application of NGS detection for predicting *ERBB2* copy number

amplification in breast cancer patients. To this end, we applied a 101 gene panel for detecting *ERBB2* copy number amplification in 123 tissues and 202 blood samples. Immunohistochemistry (IHC) and Fluorescence *in situ* hybridization (FISH) were performed simultaneously. We then analyzed and compared the *ERBB2* copy number amplification results obtained by all the methods. The NGS results for tissue samples were highly consistent with IHC and FISH, with a sensitivity of 86%, a specificity of 99%, and accuracy of 95% (ratio of *ERBB2* duplication

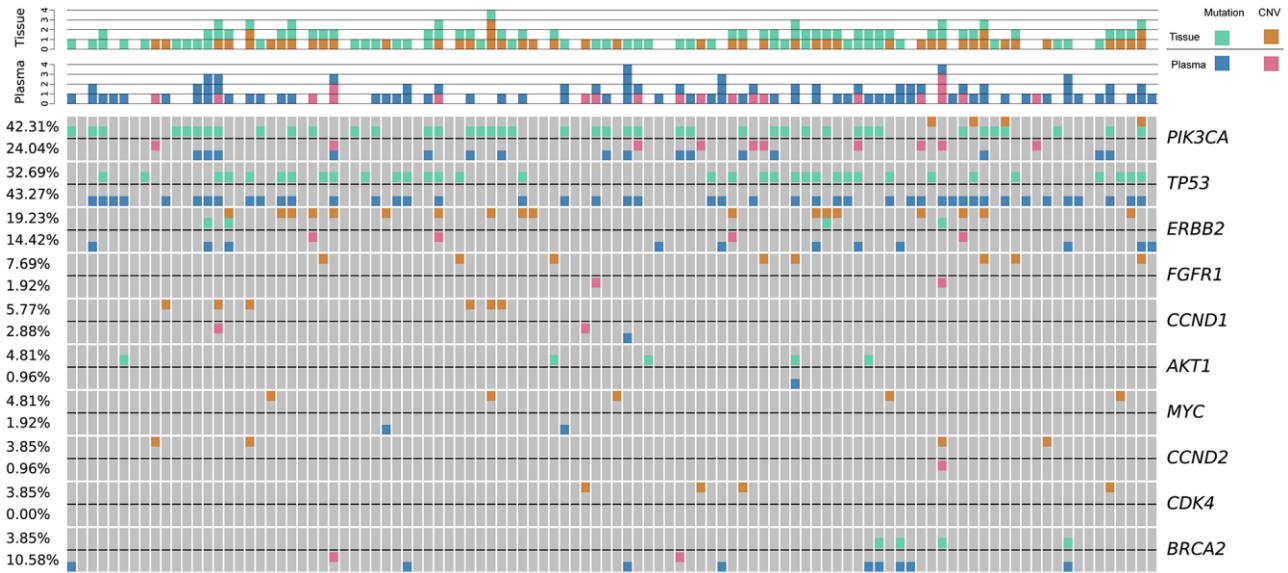


Figure 2. The genetic landscape of tissue DNA and plasma ctDNA alterations in Chinese breast cancer patients. Green and orange colors represent mutations, and CNV detected in tissue samples, respectively; blue and pink colors represent mutations, and CNV detected in cfDNA samples, respectively. Each column represents one patient. Each row represents one gene. The top bar denotes the number of mutations detected in each patient. The sidebar represents the proportion of patients with a mutation in a certain gene. CNV, copy number variation; ctDNA, circulating tumor-derived DNA; cfDNA, cell-free DNA.

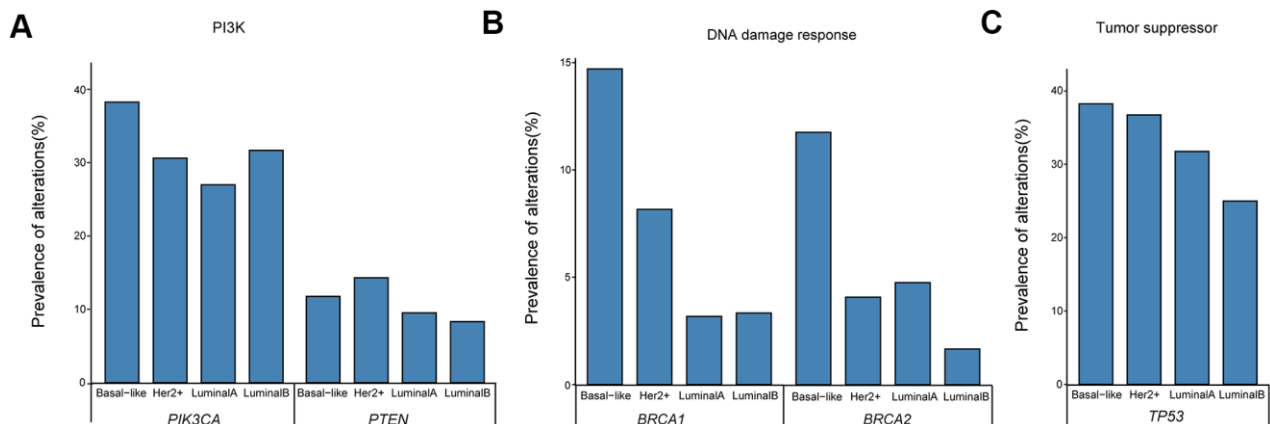


Figure 3. Prevalence of ctDNA oncogenic mutations in the (A) PI3K pathway and loss-of-function mutations in (B) DNA damage response and (C) tumor suppressor pathways. ctDNA, circulating tumor-derived DNA.

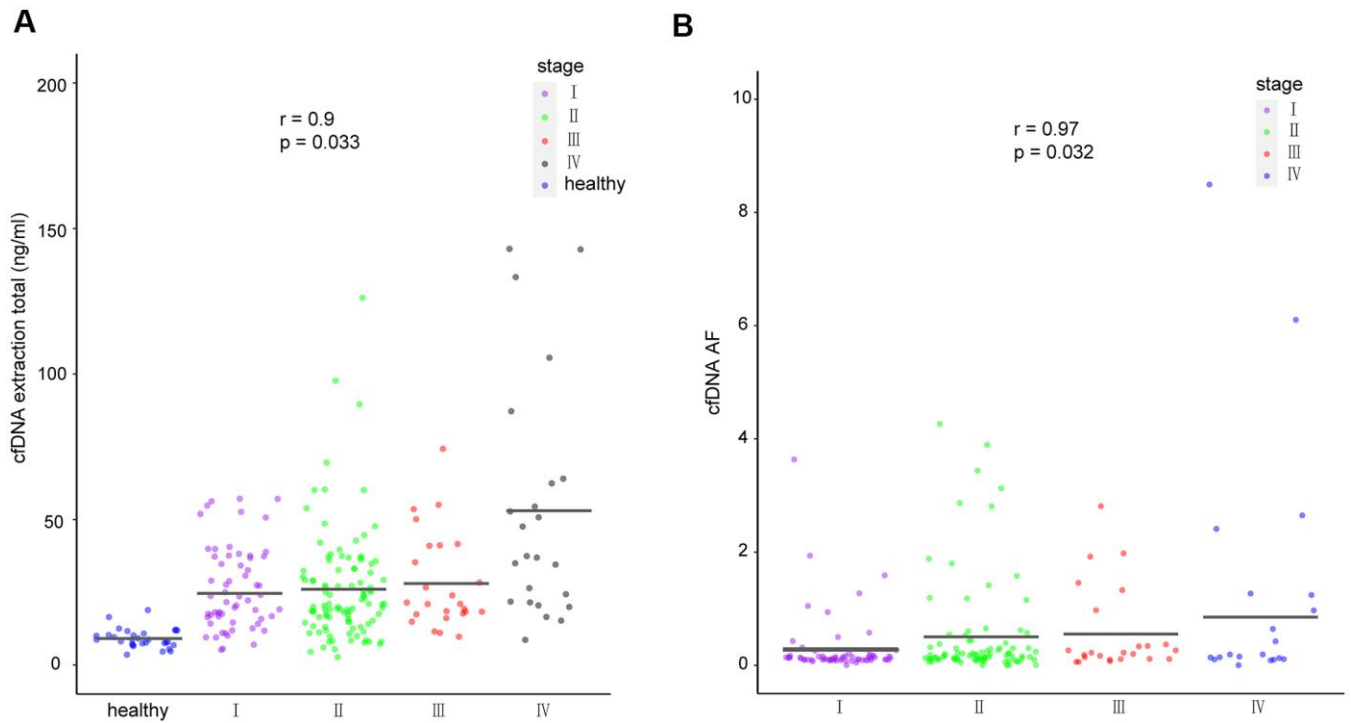


Figure 4. cfDNA in healthy individuals and breast cancer patients. (A) Amount of cfDNA extracted from all healthy individuals and cancer patients of different stages. (B) Mutant allele frequency of cfDNA detected in patients with different cancer stages. The means for each group are represented by the black lines in each column. cfDNA, cell-free DNA.

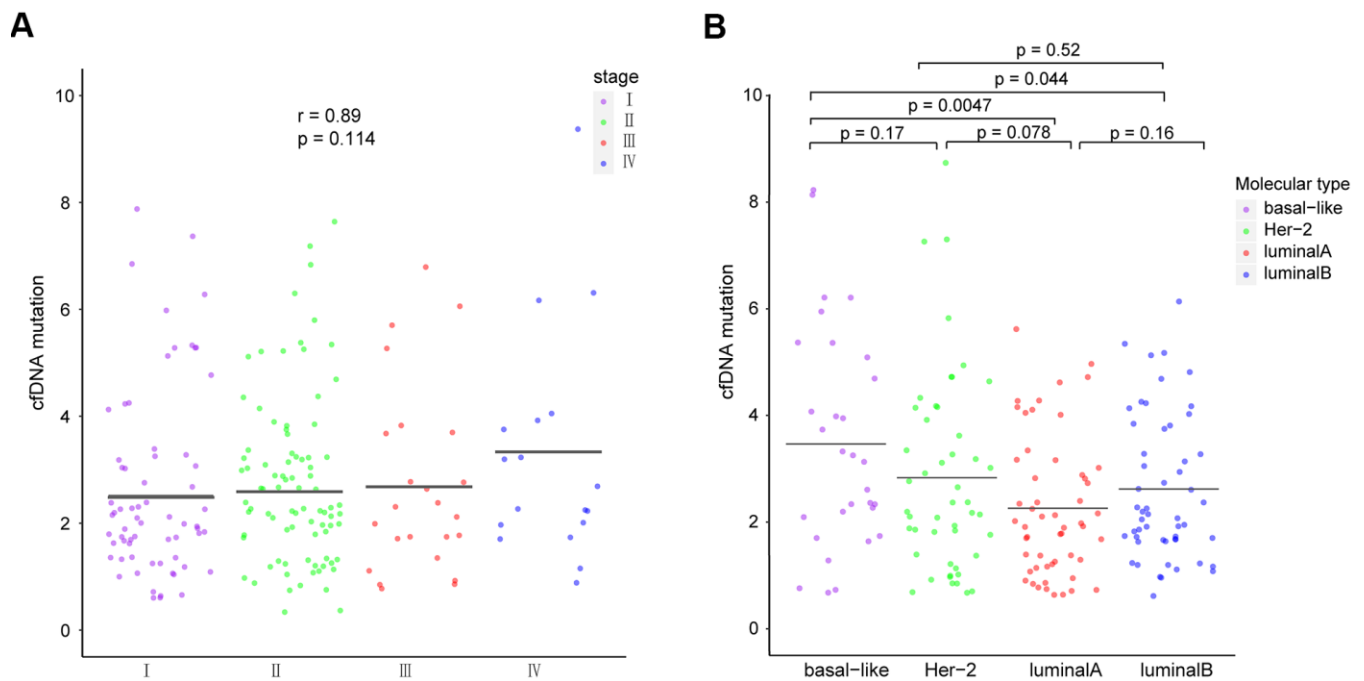


Figure 5. Comparison of ctDNA among patients with different (A) breast cancer stages and (B) molecular subtypes. The means of each group are represented by black lines in each column. ctDNA, circulating tumor-derived DNA.

region $\geq 75\%$. The duplication of *ERBB2* is identified when the proportion of amplified *ERBB2*'s exon region greater than or equal to 0.75). However, the results of NGS for blood samples were relatively inconsistent with those of IHC and FISH results, presenting a sensitivity of only 5.4%, a specificity of 97%, and accuracy of 71% (Supplementary Table 2).

For samples with an IHC test result of 2+, an additional FISH test was required to determine the final *ERBB2* expression. We analyzed the FISH and NGS results for tissue samples simultaneously to evaluate the consistency between the two detection methods in IHC 2+ samples. We obtained 100% consistency in 12 tissue samples, implying the perfect accuracy of NGS for the detection of *ERBB2* copy number amplification in IHC 2+ samples (Supplementary Table 3). Our findings, therefore, support the application of NGS for the detection of *ERBB2* copy number amplification in tissue samples, particularly when combined with NGS and IHC, which would serve to improve the accuracy of detecting *ERBB2* copy number amplification.

Dynamic monitoring and prognosis assessment

The application of ctDNA for dynamic monitoring and prognostic assessment of tumor patients was also explored. We focused on serial samples of 15 patients who underwent two or more ctDNA tests.

Eleven patients presented with decreased ctF ratio (ctF ratio: ctF2/ctF1, ctF1 refers to the highest AF in the first test, and ctF2 refers to the highest AF in the second test) (Figure 6, Supplementary Table 4), which is consistent with their imaging findings: 10 patients had significantly shrunken tumor size, and 1 patient had no significant increase in tumor size. Meanwhile, four patients had increased ctF ratios during or after treatment, 3 of whom P112, P219, and P86, had metastasis or relapse. The remaining patient required continued follow-up and observation to determine the clinical outcome.

Patient P112 had locally advanced breast cancer, which was first diagnosed in December 2015, and received chemotherapy, endocrine therapy, and radiotherapy. The disease then progressed in March 2017 with liver metastasis. The first ctDNA detection (T1) was performed in April 2018, and postoperative lumbar spine metastasis occurred during treatment. The second detection (T2) was performed four months later when the disease had continuously progressed. Based on the patient's genetic detection results, we recommended alpelisib plus fulvestrant treatment. We tracked the *PIK3CA* p.E545K and *TP53* p.E286K mutations and found that their frequency increased from 0.64% and

0.18% to 2.65% and 1.54%, respectively (Figure 7). The change in ctDNA in the patient's plasma was consistent with the clinical disease progression, resulting in a poor prognosis. A similar situation occurred in patient P219 with late-stage breast cancer, and the *TP53* p.Y220C mutation increased from 0.16% to 0.97%, with a ctF ratio of 2.310. Additionally, patient p311 was diagnosed with HER2+ breast cancer and received neoadjuvant chemotherapy and target therapy. The patient presented with decreased ctF ratio and tumor size during mutation tracking and reached complete clinical response during surgery.

To further understand the molecular characteristics and clonal evolution in our study samples, clonal analysis was performed for several patients using mutation tracking data. Certain clones were undetectable following treatment in patients P080, P105, P268, and P391. Moreover, ctF values decreased after receiving treatment in these patients, suggesting a decreased tumor mutational burden or disappearance of specific clonal clusters resulting in a relatively better prognosis (Figure 8). However, new emerging clones were observed following treatment in patient P086.

Patient P086 was diagnosed with triple-negative breast cancer (TNBC) in April 2018 and received the first ctDNA detection (T1) before standard neoadjuvant chemotherapy with no significant change observed in the tumor during this period. The patient then discontinued our treatment schedule for four months and came back with a larger mass. The second ctDNA detection was performed in February 2019 (T2), after which an additional four rounds of chemotherapy were performed. No significant change in the tumor was observed during the treatment period. The third ctDNA detection and surgery were performed in May 2019. The newly emerged mutation at T2 was *ERBB2* p.R678Q, which was also detected at T3 with an increase in mutation frequency from 2.81% to 3.9%. The ctF value then increased, while the clone analysis demonstrated an increase in clones. These analyses indicated that the patient exhibited a tendency for disease progression, consistent with the patient's clinical manifestations. Studies have shown that liquid biopsy is more informative than traditional methods, such as the CA-125 test and imaging, capable of predicting tumor recurrence 7–10 months earlier [26].

Moreover, previous reports indicate that *ERBB2* p.R678Q likely represents an oncogenic mutation [27]. Herein, the cytological study results showed that anti-HER antibody drugs, trastuzumab and pertuzumab, as well as small molecule kinase inhibitors, lenatinib, and lapatinib, elicited inhibitory effects on tumor cells harboring *ERBB2* p.R678Q mutations [28]. Although

patient P086 tested negative for HER2 amplification using IHC and FISH, an oncogenic missense mutation in the HER2 subtype was detected using ctDNA, suggesting that HER2-targeted therapy may be an effective follow-up treatment option.

Considering that patient P086 was relatively insensitive to standard neoadjuvant chemotherapy regimen, we next sought to analyze the HRD (homologous recombination deficiency) status of several TNBC patients, including patient P086, to explore optional individualized treatments for TNBC patients with different HRD statuses. The HRD score predicts TNBC or *BRCA1/2*-mutated patient response to DNA damaging neoadjuvant chemotherapy. Patients with an HRD score ≥ 42 were defined as HR deficient and tended to be more sensitive to platinum-containing regimens [29]. Meanwhile, sequencing data analysis revealed that the HRD-sum of patient P086 was remarkably high (Supplementary Table 5), which might be related to the patient's relatively poor prognosis. Among the five TNBC patients, two were HR deficient, accounting for 40% of the tested patients; the result was consistent with that reported in previous literature (42%) [30]. HRD-positive patients might not benefit from conventional chemotherapy, neoadjuvant chemotherapy, or targeted therapy; however, PARP inhibitors may represent an effective treatment option for these patients [31].

DISCUSSION

Several studies have recently explored the breast cancer mutational landscape to better understand its genetic complexity [21, 32, 33]. Nevertheless, to the best of our knowledge, ours is the first study to fully elucidate the comprehensive molecular characterization of breast tumors in Chinese women. Specifically, we identified the ctDNA profiles of different pathological subtypes, evaluated the incidence of clinically operable changes,

compared the blood and tissue NGS of paired sequenced patients, and evaluated the similarity of blood and imaging evaluations for disease monitoring.

Hence the key findings of our study are as follows: First, ctDNA evaluation effectively captured the landscape of breast cancer genetic alterations, including SNVs, indels, and CNVs. Second, across the entire cohort, genetic alterations were detectable in the blood of >90% of the evaluated patients using the 101 gene panel. Third, the amount of cfDNA extracted from the blood of breast cancer patients and the detected ctF positively correlated with the disease stage. Fourth, differences in the frequency, number, and distribution of gene mutations in the different molecular subtypes of breast cancer were demonstrated.

In our cohort, we found that the *PIK3CA* gene exhibits a higher mutational frequency in TNBC patients, while the *PTEN* gene exhibits higher mutations in HER2+ patients. We also confirmed that TNBC patients exhibited a higher mutation frequency and higher TMB than luminal breast cancer patients, indicating their possible sensitive response to appropriate targeted therapies. Theoretically, tumors with high mutations would generate more proteins that could be recognized by T cells; thus, immunotherapy drugs, such as immune checkpoint inhibitors, may elicit a strong response against these tumors [34]. Therefore, we speculated that the proportion of TNBC patients who benefited from immunotherapy might be higher for the Chinese breast cancer population.

Comparing the test results of paired blood and tissue samples, we found a relatively high consistency in all stages, particularly within patients with late-stage disease. Compared with tissue sampling, ctDNA detection has greater potential for accurately assessing spatial tumor heterogeneity, reflecting the multiple metastatic sites, while also representing a more convenient method for continuous sampling.

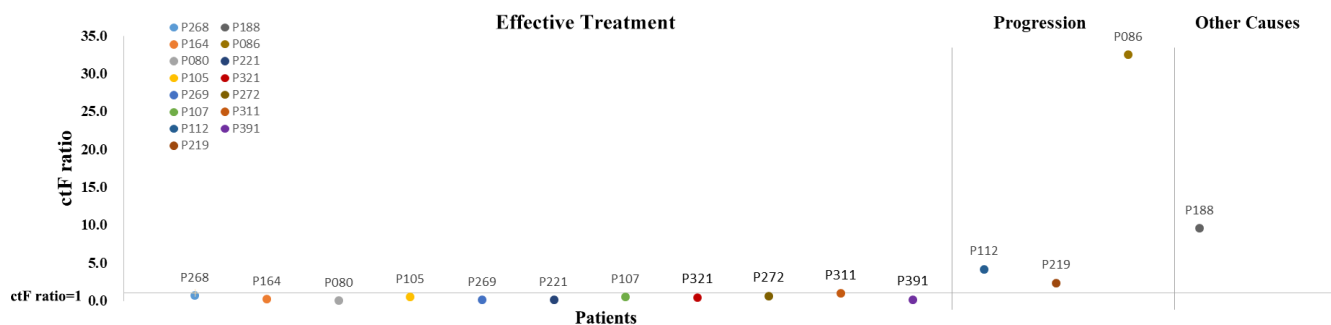


Figure 6. Scatter diagram for ctF ratio of 15 patients with two or more cfDNA tests. The horizontal line on the abscissa indicates that the ctF ratio is 1. Each dot represents one patient. cfDNA, cell-free DNA; ctF, cfDNA allele frequency.

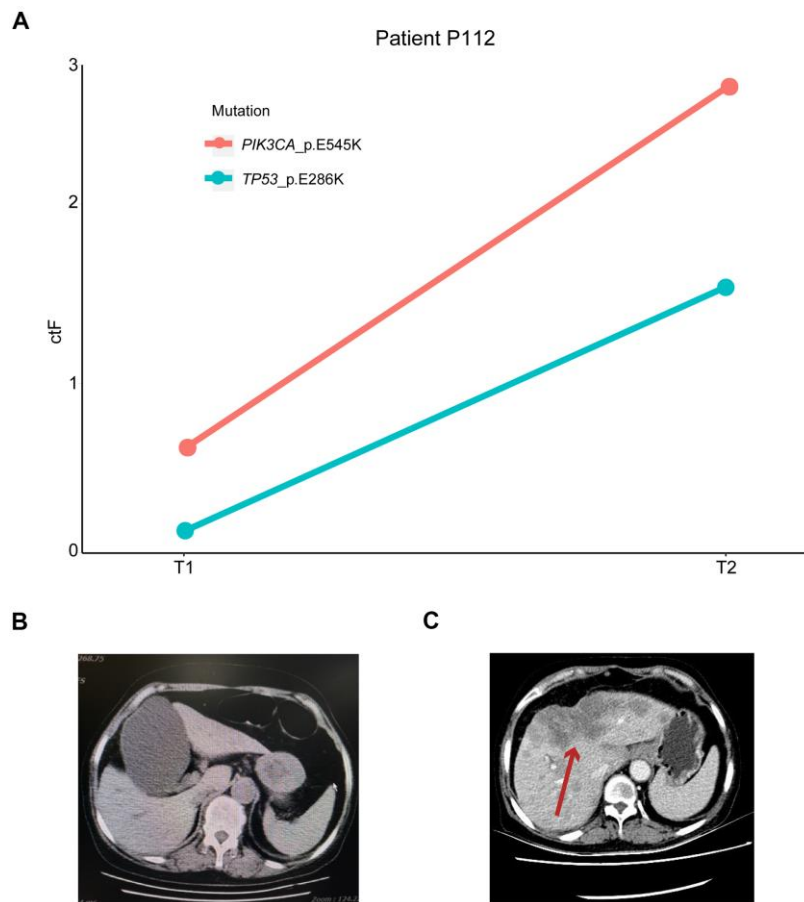


Figure 7. Mutation tracking in metastasis and recurrence patients. (A) ctDNA mutation tracking in patient P112. Computed tomography scan of P112 patient on (B) T1 and (C) T2 follow-ups. The red arrow indicates the patient's metastatic foci. ctDNA, circulating tumor-derived DNA.

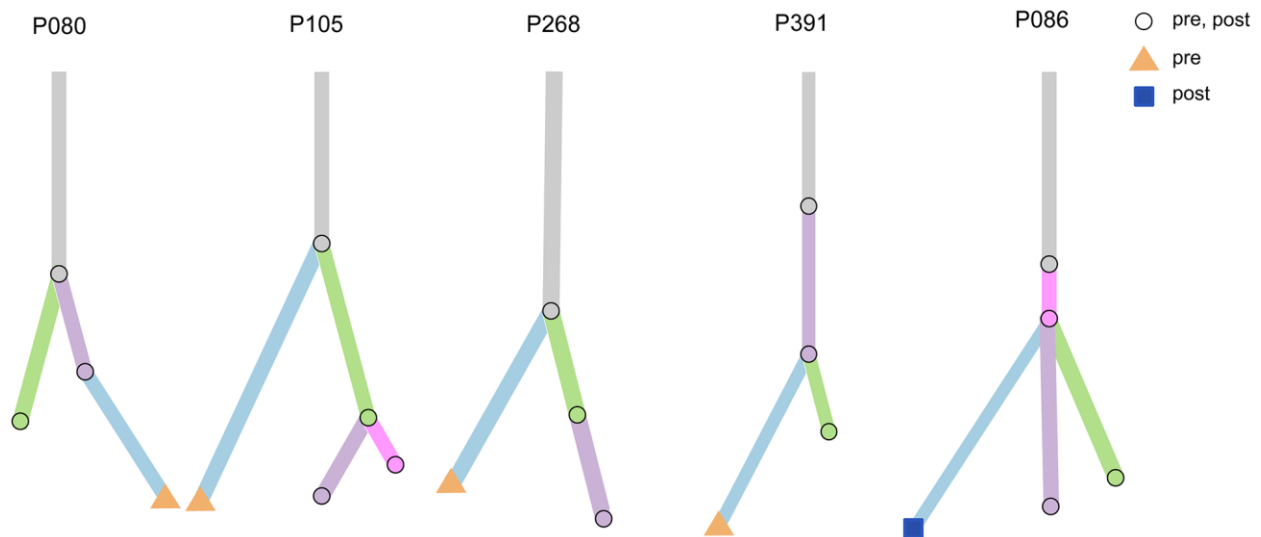


Figure 8. Clonal analysis of multiple testing in five patients. Five examples of evolutionary trees. The circles represent mutations present in both pre-treatment and post-treatment; the triangles represent mutations present only in pre-treatment, and squares represent mutations present only in post-treatment.

ERBB2 copy number amplification is an essential mutation in breast cancer. Several *ERBB2* mutation-targeting therapies have been proven to be effective for breast cancer patients [35]. Hence, the Food and Drug Administration has approved lapatinib, neratinib, trastuzumab, as well as other small molecule inhibitors and antibody-drug conjugate drugs for the treatment of HER2+ (overexpression/amplification) breast cancer patients. Therefore, accurate detection of *ERBB2* gene amplification is crucial to ensure the effective treatment of breast cancer patients. In our study, the NGS *ERBB2* copy number amplification results for tissue samples were highly consistent with those of IHC and FISH but with poor consistency for blood samples. The low consistency observed in blood samples may be due to high intratumor heterogeneity, which would cause variable amplification of the *ERBB2* copy number in different tumor tissue sites. Additionally, the proportion of examined tumor cells in the tumor tissue samples was higher than 20%; however, only low amounts of DNA were released by the tumor cells into the blood. Hence, the existing NGS analysis platform has certain limitations associated with the analysis of copy number amplification from tumor blood samples.

The present study has several limitations: First, paired sequencing data were only available for a subset of patients. Second, baseline blood collection of all patients was not performed at the same time. Most of our patients have already received multiple treatments; their different treatment options may affect the patient's mutation spectrum. Third, the prognostic value of frequently altered genes, such as *TP53*, *PIK3CA*, and *ERBB2*, requires further confirmation as our cohort's overall survival data was not mature. Finally, our study was conducted in a single center with a limited number of patients. It is necessary to verify our results through a multi-center cooperative study.

In our study, patients with decreased ctF ratios appeared more likely to reach complete clinical response or present with higher Miller Payne grades than those with increased ctF ratios after neoadjuvant chemotherapy. Therefore, we postulate that the decreased ctF ratio could reflect the tumor degradation rate and the effectiveness of minimal residual disease clearance. However, further studies with additional statistical analyses are needed to confirm this hypothesis.

MATERIALS AND METHODS

Patients and samples

Thirty healthy and 273 breast cancer patients were enrolled at The First Affiliated Hospital of Wenzhou Medical University. All patient-derived specimens and

patients' information were collected and archived under protocols approved by the institutional review board of the First Affiliated Hospital of Wenzhou Medical University. Informed consent was obtained from each patient for the use of their blood and tissue samples. All patients underwent complete tumor staging according to the seventh edition tumor, node, and metastasis (TNM) criteria of breast cancer [36]. A 10-mL sample of whole peripheral blood was collected at initial diagnosis in cfDNA blood collection tubes (Omigen, Hangzhou, China) and was processed for plasma and white blood cells (WBC) within 72 h, resulting in the collection of approximately 4 mL of plasma. The first blood sample from each patient was collected before tissue biopsy/surgery, and some patients had multiple time points of follow-up sampling after therapy. Tumor tissues were collected and either analyzed fresh or were formalin-fixed and paraffin-embedded (FFPE). *Her2* status was tested using IHC. Samples assigned a value of 2+ were further confirmed using FISH. From April 2018 to September 2019, there were 205 cfDNA and 131 tissue/FFPE samples; 104 samples were paired with tissue/FFPE and cfDNA samples.

DNA extraction

Genomic DNA was extracted from the tumor tissue, FFPE, or WBC using an Omigen extraction kit (Omigen, Inc., Hangzhou, China), followed by fragmentation using an Omigen fragmentation kit (Omigen, Inc., Hangzhou, China). cfDNA was extracted from plasma using an OMInano isolation kit (Omigen, Inc., Hangzhou, China). DNA yield was quantified with Qubit HS DNA (Life Technologies, Carlsbad, CA, USA). All the procedures were performed according to the manufacturer's instructions.

Library construction

A DNA library was prepared using an Omigen lib prep kit (Omigen, Inc., Hangzhou, China). Briefly, DNA was end prepared, then ligated with barcoded molecular adaptors, followed by polymerase chain reaction (PCR) amplification to add indexes. Library yield was quantified with Qubit, and the size was measured using a 2100 bioanalyzer (Agilent, Santa Clara, CA, USA).

Hybrid capture

To ensure minimal coverage, libraries for each pool were prepared according to the Omigen libraries pooling guideline (China patent: 201910515833.4). Libraries were pooled and target enriched using the Omigen 101 genes panel (Supplementary Table 6) and Omigen hybrid capture kit (Omigen, Inc., Hangzhou, China). The captured pool was then amplified using

PCR, and quality control (QC) was performed using a 2100 bioanalyzer. The enriched pool was sequenced using Illumina X10 (Illumina, Inc., USA) with paired-end reads and average sequencing depth of 8,000×–13,000× and 40,000×–50,000× for tissue DNA and DMI-tagged cfDNA samples, respectively.

Sequence data processing and analysis

Somatic single-nucleotide variant (SNV), fusion, and copy number variation (CNV) were analyzed and annotated using the Omigen 101 genes panel (Supplementary Table 6) and Omigen bioinformatics program pipeline including QC, alignment, and variant annotation, followed by filtering. Thirty cfDNA samples from healthy participants were used as background controls.

Exome FASTQ sequences were mapped to the human assembly NCBI build 37 (hg19) using the Bowtie2 alignment software [37]. SAMtools (v0.1.19) was used to convert SAM files to compressed BAM files and sort the BAM files using chromosomal coordinates [38]. PCR duplicates were marked using Picard (v1.131) and subsequently sorted. Marked BAM files were realigned using a Genome Analysis Toolkit (GATK-3.6) at intervals with indel mismatches. GATK was used for performing local realignment for indel calls, and VarDict was used for detecting SNVs [39, 40]. Germline variants were removed by filtering out variants in normal blood samples. Gene annotations and function prediction scores were performed using Annovar, dbSNP build135, 1000Genomes, Polyphen, Avsift, and COSMIC [41–44]. Non-synonymous somatic mutations were identified in the bulk exome sequencing data of five TNBC patients with two or three longitudinal samples. Bulk tumor genomic copy number profiles were estimated from the pair-end exome sequencing depth using the CNV kit [45]. Mutation clusters with only one mutation were excluded from further analysis. The PyClone cluster frequencies were calculated as the mean variant allele frequencies of mutations within each cluster and were visualized using the R package ‘clonevol’ [46, 47].

HRD score calculation

A portion of sporadic TNBCs and *BRCA1/2*-mutated tumors have DNA repair defects and are sensitive to DNA damage treatment [31]. Based on the loss of heterozygosity (LOH), telomere allelic imbalance (TAI), and large-scale state transition (LST), three independent DNA-based genomic instability measurements were developed. The HRD score represents the arithmetic sum of LOH (number of LOH regions >15 Mb but less than the length of a whole

chromosome) + TAI (regions of allelic imbalance that extend to the subtelomere but do not cross the centromere) + LST (breakpoints between regions of imbalance >10 Mb after filtering out regions <3 Mb). HRD scores range between 0 and 100. Tumors with HRD scores ≥ 42 or *BRCA1/2* mutations were defined as having an HR deficient status. Tumors with an HRD score of <42 that also lacked the *BRCA1/2* mutation were defined as HR non-deficient. The threshold of 42 was selected based on the fifth percentile of HRD scores in tumors with known *BRCA1/2* mutations or methylation status [48, 49]. This process was achieved using a scarHRD software based on whole-exome sequencing data [50].

Statistical analyses

A *t*-test was performed to compare ctDNA yield, mutation number, and the highest AF between different disease stages and different molecular subtypes. Moreover, their correlation was tested using Pearson’s correlation coefficient. All P values were two-sided, and values <0.05 were considered statistically significant. All statistical analyses were performed using the R language, version 3.6.0.

Abbreviations

ctDNA: circulating tumor-derived DNA; HER2: human epidermal growth factor receptor 2; NGS: next-generation sequencing; SMI: single molecular identifier; DMIs: digital molecular identifiers; AI: aromatase inhibitor; PARP: poly ADP ribose polymerase; cfDNA: cell-free DNA; AF: allele fraction; ctF: the highest cfDNA allele frequency; TMBs: tumor mutational burdens; IHC: Immunohistochemistry; FISH: Fluorescence *in situ* hybridization; TNBC: triple-negative breast cancer; HRD: homologous recombination deficiency; TNM: tumor, node, and metastasis; WBC: white blood cells; FFPE: formalin-fixed and paraffin-embedded; PCR: polymerase chain reaction; QC: quality control; SNV: single-nucleotide variant; CNV: copy number variation; LOH: loss of heterozygosity; TAI: telomere allelic imbalance; LST: large-scale state transition.

AUTHOR CONTRIBUTIONS

Conceptualization, L.L., Y.W., L.L., and J.W.; methodology, L.L., Y.W., L.L., J.W., Y.C., and X.D.; software, R.H., D.H., Y.Z., X.M., F.G. and Y.C.; validation, L.L., Y.W., L.L., J.Y., H.C., and Z.Y.; formal analysis, H.C., L.J. D.C., F.Y. and E.X.; investigation, L.L., Y.W., L.L., D.C., F.Y., E.X. and Y.Z.; resources, Y.Z., Y.C., R.H. and X.D.; data curation, L.L., Y.W., L.L., X.D., and Y.C.; writing—original draft preparation,

L.L., Y.W., L.L.; writing—review and editing, L.L., Y.W., L.L.; visualization, H.C., L.J., X.M., F.G., D.C., F.Y. and E.X.; supervision, Y.T., M.Z., and O.W.; project administration, Y.T., M.Z., and O.W.; funding acquisition, M.Z., and O.W. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

The authors thank all the patients who participated in this study and their families and the technical support from Omigen, Inc.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This study was funded by the Foundation of Wenzhou Municipal Science and Technology Bureau, China (No. Y20170030, Y20190209), the Wenzhou Science and Technology Project (Y20190209; Y20190204), and the Natural Science Foundation of Zhejiang Province (LGF18H160031).

Editorial Note

[&]This corresponding author has a verified history of publications using the personal email address for correspondence.

REFERENCES

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016; 66:115–32. <https://doi.org/10.3322/caac.21338> PMID:[26808342](https://pubmed.ncbi.nlm.nih.gov/26808342/)
2. Torre LA, Sauer AM, Chen MS Jr, Kagawa-Singer M, Jemal A, Siegel RL. Cancer statistics for Asian Americans, Native Hawaiians, and Pacific Islanders, 2016: converging incidence in males and females. *CA Cancer J Clin.* 2016; 66:182–202. <https://doi.org/10.3322/caac.21335> PMID:[26766789](https://pubmed.ncbi.nlm.nih.gov/26766789/)
3. Zheng RS, Sun KX, Zhang SW, Zeng HM, Zou XN, Chen R, Gu XY, Wei WW, He J. [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi.* 2019; 41:19–28. <https://doi.org/10.3760/cma.j.issn.0253-3766.2019.01.005> PMID:[30678413](https://pubmed.ncbi.nlm.nih.gov/30678413/)
4. Ellsworth RE, Blackburn HL, Shriver CD, Soon-Shiong P, Ellsworth DL. Molecular heterogeneity in breast

cancer: state of the science and implications for patient care. *Semin Cell Dev Biol.* 2017; 64:65–72.

<https://doi.org/10.1016/j.semcdb.2016.08.025> PMID:[27569190](https://pubmed.ncbi.nlm.nih.gov/27569190/)

5. Ma D, Jiang YZ, Liu XY, Liu YR, Shao ZM. Clinical and molecular relevance of mutant-allele tumor heterogeneity in breast cancer. *Breast Cancer Res Treat.* 2017; 162:39–48. <https://doi.org/10.1007/s10549-017-4113-z> PMID:[28093659](https://pubmed.ncbi.nlm.nih.gov/28093659/)
6. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ, Albain KS, André F, Bergh J, Bonnefoi H, Bretel-Morales D, Burstein H, et al, and Panel members. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol.* 2013; 24:2206–23. <https://doi.org/10.1093/annonc/mdt303> PMID:[23917950](https://pubmed.ncbi.nlm.nih.gov/23917950/)
7. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. *N Engl J Med.* 2013; 368:842–51. <https://doi.org/10.1056/NEJMra1204892> PMID:[23445095](https://pubmed.ncbi.nlm.nih.gov/23445095/)
8. Klaassen Z, Shay R, Moses KA, Terris MK. Words of wisdom. Re: volume outcomes of cystectomy—is it the surgeon or the setting? *Eur Urol.* 2014; 65:846. <https://doi.org/10.1016/j.eururo.2013.12.032> PMID:[24559904](https://pubmed.ncbi.nlm.nih.gov/24559904/)
9. Beca F, Polyak K. Intratumor heterogeneity in breast cancer. *Adv Exp Med Biol.* 2016; 882:169–89. https://doi.org/10.1007/978-3-319-22909-6_7 PMID:[26987535](https://pubmed.ncbi.nlm.nih.gov/26987535/)
10. Arnedos M, Vicier C, Loi S, Lefebvre C, Michiels S, Bonnefoi H, Andre F. Precision medicine for metastatic breast cancer—limitations and solutions. *Nat Rev Clin Oncol.* 2015; 12:693–704. <https://doi.org/10.1038/nrclinonc.2015.123> PMID:[26196250](https://pubmed.ncbi.nlm.nih.gov/26196250/)
11. Desmedt C, Voet T, Sotiriou C, Campbell PJ. Next-generation sequencing in breast cancer: first take home messages. *Curr Opin Oncol.* 2012; 24:597–604. <https://doi.org/10.1097/CCO.0b013e328359554e> PMID:[23014189](https://pubmed.ncbi.nlm.nih.gov/23014189/)
12. Friel AM, Corcoran C, Crown J, O'Driscoll L. Relevance of circulating tumor cells, extracellular nucleic acids, and exosomes in breast cancer. *Breast Cancer Res Treat.* 2010; 123:613–25. <https://doi.org/10.1007/s10549-010-0980-2> PMID:[20549336](https://pubmed.ncbi.nlm.nih.gov/20549336/)

13. Hojbjerg JA, Madsen AT, Schmidt HH, Sorensen SF, Stougaard M, Meldgaard P, Sorensen BS. Intra-individual variation of circulating tumour DNA in lung cancer patients. *Mol Oncol*. 2019; 13:2098–106. <https://doi.org/10.1002/1878-0261.12546> PMID:[31306545](https://pubmed.ncbi.nlm.nih.gov/31306545/)
14. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, Marafioti T, Kirkizlar E, Watkins TB, et al, and TRACERx consortium, and PEACE consortium. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017; 545:446–51. <https://doi.org/10.1038/nature22364> PMID:[28445469](https://pubmed.ncbi.nlm.nih.gov/28445469/)
15. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, Silliman N, Tacey M, Wong HL, Christie M, Kosmider S, Skinner I, Wong R, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med*. 2016; 8:346ra92. <https://doi.org/10.1126/scitranslmed.aaf6219> PMID:[27384348](https://pubmed.ncbi.nlm.nih.gov/27384348/)
16. Pantel K, Alix-Panabières C. Liquid biopsy and minimal residual disease - latest advances and implications for cure. *Nat Rev Clin Oncol*. 2019; 16:409–24. <https://doi.org/10.1038/s41571-019-0187-3> PMID:[30796368](https://pubmed.ncbi.nlm.nih.gov/30796368/)
17. Wang W, Song Z, Zhang Y. A Comparison of ddPCR and ARMS for detecting EGFR T790M status in ctDNA from advanced NSCLC patients with acquired EGFR-TKI resistance. *Cancer Med*. 2017; 6:154–62. <https://doi.org/10.1002/cam4.978> PMID:[28000387](https://pubmed.ncbi.nlm.nih.gov/28000387/)
18. Iwahashi N, Sakai K, Noguchi T, Yahata T, Matsukawa H, Toujima S, Nishio K, Ino K. Liquid biopsy-based comprehensive gene mutation profiling for gynecological cancer using CAncer Personalized Profiling by deep Sequencing. *Sci Rep*. 2019; 9:10426. <https://doi.org/10.1038/s41598-019-47030-w> PMID:[31320709](https://pubmed.ncbi.nlm.nih.gov/31320709/)
19. Kou R, Lam H, Duan H, Ye L, Jongkam N, Chen W, Zhang S, Li S. Benefits and challenges with applying unique molecular identifiers in next generation sequencing to detect low frequency mutations. *PLoS One*. 2016; 11:e0146638. <https://doi.org/10.1371/journal.pone.0146638> PMID:[26752634](https://pubmed.ncbi.nlm.nih.gov/26752634/)
20. André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, Iwata H, Conte P, Mayer IA, Kaufman B, Yamashita T, Lu YS, Inoue K, et al, and SOLAR-1 Study Group. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. *N Engl J Med*. 2019; 380:1929–40. <https://doi.org/10.1056/NEJMoa1813904> PMID:[31091374](https://pubmed.ncbi.nlm.nih.gov/31091374/)
21. Tao Z, Li T, Feng Z, Liu C, Shao Y, Zhu M, Gong C, Wang B, Cao J, Wang L, Du Y, Lizaso A, Li B, et al. Characterizations of cancer gene mutations in Chinese metastatic breast cancer patients. *Front Oncol*. 2020; 10:1023. <https://doi.org/10.3389/fonc.2020.01023> PMID:[32695676](https://pubmed.ncbi.nlm.nih.gov/32695676/)
22. Juvekar A, Burgal LN, Hu H, Lunsford EP, Ibrahim YH, Balmañà J, Rajendran A, Papa A, Spencer K, Lyssiottis CA, Nardella C, Pandolfi PP, Baselga J, et al. Combining a PI3K inhibitor with a PARP inhibitor provides an effective therapy for BRCA1-related breast cancer. *Cancer Discov*. 2012; 2:1048–63. <https://doi.org/10.1158/2159-8290.CD-11-0336> PMID:[22915751](https://pubmed.ncbi.nlm.nih.gov/22915751/)
23. Xue X, Teare MD, Hoken I, Zhu YM, Woll PJ. Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. *Clin Chim Acta*. 2009; 404:100–04. <https://doi.org/10.1016/j.cca.2009.02.018> PMID:[19281804](https://pubmed.ncbi.nlm.nih.gov/19281804/)
24. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, Anagnostou V, Fiksel J, Cristiano S, Papp E, Speir S, Reinert T, Orntoft MW, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med*. 2017; 9:eaan2415. <https://doi.org/10.1126/scitranslmed.aan2415> PMID:[28814544](https://pubmed.ncbi.nlm.nih.gov/28814544/)
25. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019; 51:202–06. <https://doi.org/10.1038/s41588-018-0312-8> PMID:[30643254](https://pubmed.ncbi.nlm.nih.gov/30643254/)
26. Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, et al. Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. *PLoS Med*. 2016; 13:e1002198. <https://doi.org/10.1371/journal.pmed.1002198> PMID:[27997533](https://pubmed.ncbi.nlm.nih.gov/27997533/)
27. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, Monsey J, Goel N, Aronson AB, Li S, Ma CX, Ding L, Mardis ER, Ellis MJ. Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov*. 2013; 3:224–37. <https://doi.org/10.1158/2159-8290.CD-12-0349> PMID:[23220880](https://pubmed.ncbi.nlm.nih.gov/23220880/)

28. Pahuja KB, Nguyen TT, Jaiswal BS, Prabhaskar K, Thaker TM, Senger K, Chaudhuri S, Kljavin NM, Antony A, Phalke S, Kumar P, Mravic M, Stawiski EW, et al. Actionable activating oncogenic ERBB2/HER2 transmembrane and juxtamembrane domain mutations. *Cancer Cell*. 2018; 34:792–806.e5. <https://doi.org/10.1016/j.ccell.2018.09.010> PMID:30449325
29. Telli ML, Hellyer J, Audeh W, Jensen KC, Bose S, Timms KM, Gutin A, Abkevich V, Peterson RN, Neff C, Hughes E, Sangale Z, Jones J, et al. Homologous recombination deficiency (HRD) status predicts response to standard neoadjuvant chemotherapy in patients with triple-negative or BRCA1/2 mutation-associated breast cancer. *Breast Cancer Res Treat*. 2018; 168:625–30. <https://doi.org/10.1007/s10549-017-4624-7> PMID:29275435
30. Bertucci F, Ng CK, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, Soria JC, Dien AT, Adnani Y, Kamal M, Garnier S, Meurice G, Jimenez M, et al. Genomic characterization of metastatic breast cancers. *Nature*. 2019; 569:560–64. <https://doi.org/10.1038/s41586-019-1056-z> PMID:31118521
31. Chopra N, Tovey H, Pearson A, Cutts R, Toms C, Proszek P, Hubank M, Dowsett M, Dodson A, Daley F, Kriplani D, Gevensleben H, Davies HR, et al. Homologous recombination DNA repair deficiency and PARP inhibition activity in primary triple negative breast cancer. *Nat Commun*. 2020; 11:2662. <https://doi.org/10.1038/s41467-020-16142-7> PMID:32471999
32. Davis AA, Jacob S, Gerratana L, Shah AN, Wehbe F, Katam N, Zhang Q, Flaum L, Siziopikou KP, Platanius LC, Gradishar WJ, Behdad A, Cristofanilli M. Landscape of circulating tumour DNA in metastatic breast cancer. *EBioMedicine*. 2020; 58:102914. <https://doi.org/10.1016/j.ebiom.2020.102914> PMID:32707446
33. Zhang X, Zhao W, Wei W, You Z, Ou X, Sun M, Yin Y, Tang X, Zhao Z, Hu C, Liu F, Deng J, Mao L, et al. Parallel analyses of somatic mutations in plasma circulating tumor DNA (ctDNA) and matched tumor tissues in early-stage breast cancer. *Clin Cancer Res*. 2019; 25:6546–53. <https://doi.org/10.1158/1078-0432.CCR-18-4055> PMID:31350313
34. Klempner SJ, Fabrizio D, Bane S, Reinhart M, Peoples T, Ali SM, Sokol ES, Frampton G, Schrock AB, Anhorn R, Reddy P. Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence. *Oncologist*. 2020; 25:e147–59. <https://doi.org/10.1634/theoncologist.2019-0244> PMID:31578273
35. Oh DY, Bang YJ. HER2-targeted therapies - a role beyond breast cancer. *Nat Rev Clin Oncol*. 2020; 17:33–48. <https://doi.org/10.1038/s41571-019-0268-3> PMID:31548601
36. Edge SB, Compton CC. The American joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010; 17:1471–74. <https://doi.org/10.1245/s10434-010-0985-4> PMID:20180029
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–59. <https://doi.org/10.1038/nmeth.1923> PMID:22388286
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–79. <https://doi.org/10.1093/bioinformatics/btp352> PMID:19505943
39. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–303. <https://doi.org/10.1101/gr.107524.110> PMID:20644199
40. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, Johnson J, Dougherty B, Barrett JC, Dry JR. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016; 44:e108. <https://doi.org/10.1093/nar/gkw227> PMID:27060149
41. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013; Chapter 7:Unit7.20. <https://doi.org/10.1002/0471142905.hg0720s76> PMID:23315928
42. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, and 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015; 526:68–74. <https://doi.org/10.1038/nature15393> PMID:26432245

43. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 2017; 45:D777–83.
<https://doi.org/10.1093/nar/gkw1121>
PMID:27899578
44. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc.* 2015; 10:1556–66.
<https://doi.org/10.1038/nprot.2015.105>
PMID:26379229
45. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol.* 2016; 12:e1004873.
<https://doi.org/10.1371/journal.pcbi.1004873>
PMID:27100738
46. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A, Shah SP. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods.* 2014; 11:396–98.
<https://doi.org/10.1038/nmeth.2883> PMID:24633410
47. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, Maher CA. ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol.* 2017; 28:3076–82.
<https://doi.org/10.1093/annonc/mdx517>
PMID:28950321
48. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, Tran TV, Williams D, Iliev D, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer.* 2012; 107:1776–82.
<https://doi.org/10.1038/bjc.2012.451>
PMID:23047548
49. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, Olson JE, Godwin AK, Pankratz VS, Olswold C, Slettedahl S, Hallberg E, Guidugli L, et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol.* 2015; 33:304–11.
<https://doi.org/10.1200/JCO.2014.57.1414>
PMID:25452441
50. Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, Birkbak NJ, Eklund AC, Syed A, Szallasi Z. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer.* 2018; 4:16.
<https://doi.org/10.1038/s41523-018-0066-6>
PMID:29978035

SUPPLEMENTARY MATERIALS

Supplementary Tables

Supplementary Table 1. Basic clinical, demographic, and pathologic characteristics of the breast cancer patient cohort (Number = 273).

Characteristic	Number	Percent (%)
Age (years)		
≤ 60	202	73.99
> 60	71	26.01
Molecular subtype		
Unknown	10	3.66
Luminal A	77	28.21
Luminal B	74	27.11
HER2-positive	67	24.54
Basal-like	45	16.48
TNM ¹ stage		
0	7	2.56
I	94	34.43
II	111	40.66
III	31	11.36
IV	30	10.99

¹Tumor, node, and metastasis according to the 7th edition.

Supplementary Table 2. Next-generation sequencing detection for the prediction of *ERBB2* copy number amplification in breast cancer patients.

Type	TP ³	FP ⁴	TN ⁵	FN ⁶	Sensitivity	Specificity	Accuracy
TS1	30	1	87	5	0.86	0.99	0.95
CF2	3	5	141	53	0.054	0.97	0.71

¹Tissue; ²cfDNA (circulating tumor-derived DNA); ³true positive; ⁴false positive; ⁵true negative; ⁶false negative.

Supplementary Table 3. Consistency between FISH1 and NGS3 methods for the detection of ERBB2 CNV4 in IHC2 2+ samples.

Sample ID	IHC	FISH	NGS
WY00105	2+	+	+
WY00287	2+	-	-
WY00289	2+	+	+
WY00313	2+	-	-
WY00314	2+	-	-
WY00325	2+	-	-
WY00371	2+	-	-
WY00378	2+	-	-
WY00381	2+	-	-
WY00417	2+	-	-
WY00427	2+	+	+
WY00448	2+	+	+

Supplementary Table 4. Patient information from multiple ctDNA tests.

ID	Age	Stage	Molecular type	Treatment	During treatment	Tumor size after treatment (cm)	ctF ratio	Note
P268	69	III A	luminal B	Epirubicin,Cyclophosphamide,Doxorubicin,Docetaxel,Trastuzumab	/	3.4	0.615	
P164	51	III C	luminal B	Cyclophosphamide,Epirubicin,Surgery	/	0.4	0.182	
P080	61	IV	Her2+	Docetaxel,Trastuzumab,Surgery	3.1	0.5	0.04	
P105	47	III C	Her2+	Cyclophosphamide,Doxorubicin,Docetaxel,Trastuzumab,Surgery	/	0.5	0.423	
P269	50	I	luminal B	Cyclophosphamide,Epirubicin,Docetaxel,surgery	1.8	1.3	0.106	
P221	63	III C	luminal B	Doxorubicin,Cyclophosphamide,Docetaxel	/	2	0.075	
P107	46	II B	Her2+	Cyclophosphamide,Epirubicin,Docetaxel,Trastuzumab,Surgery	/	2.5	0.444	
P321	35	II B	luminal B	Cyclophosphamide,Epirubicin	2.3	2.1	0.359	
P272	45	III A	luminal B	Epirubicin,Cyclophosphamide	/	0.8	0.588	
P311	45	III C	TNBC	Cyclophosphamide,Doxorubicin,Docetaxel,Carboplatin	/	1.6	0.899	
P391	35	II A	TNBC	Cyclophosphamide,Epirubicin,Docetaxel,Carboplatin	/	2.7	0.101	
P112	69	IV	luminal A	Epirubicin,Cyclophosphamide,Gemcitabine,Cisplatin,Nedaplatin,Docetaxel,Zoledronic acid,Letrozole,Surgery	/	4	4.141	Metastasis
P219	39	IV	luminal B	Cyclophosphamide,Epirubicin,Toremifene,Carboplatin,Docetaxel,Surgery	/	4.2	2.310	relapse
P086	44	II B	TNBC	Carboplatin,Cyclophosphamide, Epirubicin,Docetaxel,Surgery	2.6	2.6	32.500	Resistant
P188	55	II B	luminal B	Doxorubicin,Cyclophosphamide,Docetaxel	/	0.9	9.545	

Supplementary Table 5. HRDa score of five TNBC patients.

	HRD-LOH^b	HRD-LST^c	HRD-TAI^d	HRD-sum
WY00086	16	24	34	74
WY00249	2	4	8	14
WY00254	14	22	13	49
WY00330	15	14	12	41
WY00086	16	24	34	74
WY00385	10	11	5	26

^aHomologous recombination deficiency; ^bLoss of heterozygosity; ^cTelomere allelic imbalance; ^dLarge-scale state transition.

Supplementary Table 6. Omigen 101 genes.

AKT1	CDKN2A	GNAS	NF1	RICTOR
AKT2	CRLF2	HNF1A	NFE2L2	RIT1
ALDOA	CTNNB1	HRAS	NOTCH1	RMRP
ALK	DDR2	IDH1	NOTCH2	RNF43
APC	DNMT3A	IDH2	NRAS	ROS1
AR	EGFR	JAK2	NTRK1	RUNX1
ARAF	ERBB2	KDR	NTRK3	SF3B1
ARID1A	ERCC2	KEAP1	PALB2	SMAD4
ATM	ESR1	KIT	PDGFRA	SMO
BCOR	ETV6	KRAS	PDGFRB	SRSF2
BLM	EZH2	LEPROTL1	PIK3CA	STAG2
BRAF	FBXW7	MAP2K1	PIK3R1	STK11
BRCA1	FGFR1	MAP2K2	PMS2	TBC1D12
BRCA2	FGFR2	MAPK1	PTEN	TERT
BRIP1	FGFR3	MET	PTPN11	TET2
CCND1	FLT3	MSH2	RAF1	TP53
CCND3	FOXA1	MTOR	RB1	TSC1
CDH1	GATA3	MYC	RET	TSC2
CDK4	GNA11	NBN	RHEB	U2AF1
CDK6	GNAQ	NEAT1	RHOA	VHL
				ZNF143