Implication of targeted next generation sequencing in circulating tumor DNA from breast cancer patients

A Thesis Submitted to
the Department of Cancer Biomedical Science
in Partial Fulfillment of the Requirements
for the Master’s Degree of Science

Min Jeong Kwon

January 2019

NATIONAL CANCER CENTER
GRADUATE SCHOOL OF CANCER SCIENCE AND POLICY

Professor Sun-Young Kong
We hereby approve the M.S. thesis of
Min Jeong Kwon

January 2019

In Hae Park
Chairman of Thesis Committee

Seung-Tae Lee
Thesis Committee Member

Sun-Young Kong
Thesis Committee Member

NATIONAL CANCER CENTER
GRADUATE SCHOOL OF CANCER SCIENCE AND POLICY
ABSTRACT

Implication of targeted next generation sequencing in circulating tumor DNA from breast cancer patients

Circulating tumor DNA (ctDNA) is a non-invasive biomarker for early detection and prognosis in cancer cells that are released into the blood. Next generation sequencing (NGS) has been used to detect various mutations in patients and to study patient treatment. However, more research is needed in comparison with clinical studies. Therefore, we aim to identify whether the genetic characteristics of circulating tumor DNA (ctDNA) in breast cancer patients reflect the tumor DNA genome profile and its clinical implications.

In this study, 84 breast cancer patients were enrolled and NGS in ctDNA using QIAseq Targeted DNA panels and Oncomine Breast panel was performed. Whole-exome sequencing (WES) was performed using tumor DNA from 49 patients and was compared with ctDNA NGS results. Then, we validated and monitored using Droplet Digital PCR (ddPCR). The fractional abundance was analyzed by QuantaSoft software (Biorad, USA). Variants and copy number alterations (CNA) were detected in 74 of 84
patients including 73 variants and the most common variants were \textit{KMT2C} (33%), \textit{TP53} (16%) and \textit{MUC16} (8%). We compared NGS of ctDNA and tumor DNA in 49 patients. At least one mutation was found in 41 patients and the concordance rate was 81%. Kaplan-Meier (KM) curve of the survival rate according to mutation detection. Revealed that the patients with more than two variants showed poorer overall survival (OS) than patients with less than two variants in the Oncomine panel \((P = 0.0282, \text{Hazard ratio (HR)} = 4.9925, 95\% \text{ confidence interval (CI): } 1.19-20.99)\). We also monitored somatic variants of ctDNA in 10 patients. Changes in the allele frequency of ctDNA reflected therapeutic response and values of tumor marker.

This study confirmed that the genetic profile of ctDNA in breast cancer patients reflects the tumor DNA genome profile. We evaluated the clinical utility of ctDNA by comparing it with clinical results.
# Contents

1. Introduction ........................................................................................... 1  
   1.1 Breast cancer ...................................................................................... 1  
   1.2 Circulating tumor DNA (ctDNA) ...................................................... 2  
   1.3 Next-generation sequencing (NGS) ................................................... 4  
   1.4 Droplet Digital PCR (ddPCR) ........................................................... 5  

2. Purposes of this study ............................................................................ 6  

3. Materials and Methods ......................................................................... 7  
   3.1 Study design ....................................................................................... 7  
   3.2 Sample collection and blood processing ............................................ 8  
   3.3 DNA extraction and quantitation ....................................................... 9  
   3.4 QIAseq Targeted DNA panels .......................................................... 9  
   3.5 Oncomine Breast ctDNA Assay .......................................................... 11  
   3.6 ddPCR .............................................................................................. 13  
   3.7 Detection limit of ddPCR ................................................................. 14  
   3.8 Statistical analysis ............................................................................ 15  

   vi
4. Result ............................................................................................................... 16

4.1 Characteristics of breast cancer patients .............................................. 16

4.2 NGS in ctDNA and tumor DNA ............................................................ 17

4.3 Comparison of two panel assay results ............................................. 24

4.4 Concordance between ctDNA and tumor DNA ................................. 26

4.5 Association between ctDNA NGS and ddPCR ............................... 30

4.6 Survival according to variants detection ......................................... 31

4.7 Identification of the molecular mechanism and clinical significance  
............................................................................................................................................ 36

4.8 Limit of detection (LOD) and limit of quantification (LOQ) in  
ddPCR ......................................................................................................................... 40

4.9 Longitudinal monitoring of somatic variants in ctDNA .................. 41

5. Discussion ..................................................................................................... 47

6. Bibliography ............................................................................................... 51

7. Acknowledgments ...................................................................................... 62
List of Tables

Table 1. List of genes covered by QIAseq Targeted Human Breast Cancer Panels........................................................................................................... 10

Table 2. List of genes covered by Oncomine Breast cfDNA Assay v2 .... 12

Table 3. The types of genes and positive and negative DNA used for LOD evaluation ................................................................................................. 15

Table 4. Characteristics of breast cancer patient with tissue and ctDNA NGS.................................................................................................................. 17

Table 5. Concordance of mutation between QIAseq Targeted DNA and and Oncomine Breast cfDNA................................................................. 24

Table 6. Concordance of mutation between QIAseq Targeted DNA and and Oncomine Breast cfDNA................................................................. 37
List of Figures

Figure 1. Cancer incidence and mortality in Korea, 2018 ......................... 2

Figure 2. Circulating tumor DNA (ctDNA) in blood .............................. 3

Figure 3. Principle of droplet digital PCR ............................................. 6

Figure 4. Experiment scheme ............................................................... 7

Figure 5. Analysis pipeline of QIAseq Targeted Human Breast Cancer
    Panels .......................................................................................... 11

Figure 6. Analysis pipeline of Oncomine Breast cfDNA Assay v2 .......... 13

Figure 7. Number of patients with variants and CNA in ctDNA of breast
    cancer patients by NGS ................................................................ 18

Figure 8. The proportion of variants detected in NGS using ctDNA in
    breast cancer patients ..................................................................... 18

Figure 9. The frequencies of detected gene mutations and copy number
    alterations in ctDNA panel in breast cancer patients ...................... 19

Figure 10. Detection of variants in ctDNA using QIAseq Targeted DNA
    panels in 53 breast cancer patients ............................................... 21
Figure 11. Variants and CNA results of 38 patients with Oncomine breast ctDNA panel

Figure 12. Results of detected variants and CNA in WES using Tumor DNA

Figure 13. Comparison of allele frequency of Oncomine and QIAseq cfDNA panel

Figure 14. Results of detected variants in ctDNA or tumor DNA

Figure 15. Total of 28 genes detected in patients per gene in NGS results using tumor DNA or ctDNA

Figure 16. Comparison of variants in ctDNA and tumor DNA

Figure 17. Association between NGS and ddPCR for detection of mutation in ctDNA

Figure 18. KM curve according to the presence of ctDNA variants

Figure 19. KM curve according to the number of variants detected in ctDNA panel

Figure 20. KM curve of PFS according to number of mutations per stage in QIAseq Targeted and Oncomine Breast cfDNA panel
Figure 21. KM curve of OS according to number of mutations per stage in QIAseq Targeted and Oncomine Breast cfDNA panel .......... 35

Figure 22. The position of detected mutation and clinical significance for each gene .......................................................................................................................... 39

Figure 23. Detection limit of ddPCR ................................................................. 41

Figure 24. Study subjects of monitored ....................................................... 42

Figure 25. Serial monitoring of somatic variants in ctDNA ................. 46
1. Introduction

1.1 Breast cancer

In women, breast cancer is the most commonly diagnosed cancer in the world and is the cause of major cancer deaths (Bray et al. 2018). The incidence of breast cancer patients in Korea has increased from 1993 to 2018, with the highest incidence in women and the fifth highest mortality rate (Figure 1) (Jung, Won, Kong, and Lee 2018). The number of cases of breast cancer is increasing. The 5-year relative survival rate for breast cancer in 2006-2010 was over 90% due to the development of chemotherapy, surgery and radiotherapy (Jung, Won, Kong, Lee, et al. 2018). However, it is necessary to monitor treatment response and disease progression because tumor recurrence or progression may occur (Yang et al. 2018).

Blood markers such as Cancer antigen 15-3 (CA 15-3), cancer fetal antigen (CEA), human epidermal growth factor receptor 2 (HER2), and MUC1 are the most common tumor markers used in the diagnosis and treatment of breast cancer (Tan et al. 2018, Berghuis et al. 2017). Therefore, biomarkers for the treatment of personalized chemotherapeutic agents are important.
1.2 Circulating tumor DNA

Circulating tumor DNA (ctDNA), a nucleic acid isolated from blood cells, is known to circulate from cancer cells into blood. ctDNA in the blood of cancer patients is secreted into the blood during various cell changes such as apoptosis, necrosis, and secretion (Figure 2) (Diaz and Bardelli 2014). Therefore, the proliferation of cancer cells and the size of cancer tissues can be important factors of ctDNA (Schwarzenbach, Hoon, and Pantel 2011). The presence of ctDNA is associated with radiological outcomes and disease activity (Lipson et al. 2014). Detecting ctDNA in plasma or serum can act as a liquid biopsy without unnecessary toxicity during treatment of the disease (Nicolini, Ferrari, and Duffy 2018).
Figure 2. Circulating tumor DNA (ctDNA) in blood (Diaz and Bardelli 2014).

Previous studies on ctDNA have been reported in a variety of carcinomas such as non-small cell lung cancer, breast cancer, pancreatic cancer, thyroid cancer, colon cancer, and ctDNA has been shown to reflect response to molecular target therapy (Yang et al. 2018). In other studies, ctDNA has been shown to be of value as a biomarker for prediction of prognosis in cancers such as cervical cancer, colorectal cancer, pancreatic cancer, and melanoma (Rapisuwon, Vietsch, and Wellstein 2016). In breast cancer, ctDNA has been shown to be useful as a potential marker and biomarker for early detection (Cheng et al. 2018) and may be indicative of a therapeutic response (Madhavan et al. 2014). Thus, ctDNA can be an effective tool for early diagnosis of the natural course of disease or prediction of efficacy in breast cancer patients because blood samples can be collected repeatedly.
1.3 Next-generation sequencing

High sensitivity is required in detection of ctDNA due to the low presence of ctDNA in plasma and high noise ratio (Lupini et al. 2018, Mehrotra et al. 2018). Polymerase chain reaction (PCR)-based assays or next-generation sequencing (NGS)-based assays are performed to detect ctDNA (Zill et al. 2015). NGS is a powerful tool in molecular screening programs for the diagnosis and management of diseases because it can detect somatic mutations below 5% and ctDNA mutations in small amounts (Page et al. 2017, Risberg et al. 2018). NGS can detect multiple genetic variants and is useful for screening patients for targeted therapy (Iwama et al. 2017).

In other study, drug targets were selected using NGS for breast cancer, pancreatic cancer, lung cancer, and thyroid cancer. A total of 70 genes were used in 102 patients with non-small cell lung cancer (NSCLC) to identify the targetable driver and resistance mutation of the patient (Thompson et al. 2016). In one study, a comparison of mutations using tumors and ctDNA in breast cancer patients showed high concordance and indicated the need for tissue and blood-based NGS (Chae et al. 2017). Therefore, NGS is useful in the detection of various mutations in patients and in patients’ treatment studies. Although there are many studies on the progression and diagnosis of disease using NGS in ctDNA of metastatic breast cancer, there is a need for further study on the analysis compared with clinical studies.
1.4 Droplet Digital PCR

Droplet Digital PCR (ddPCR) is a third-generation PCR capable of quantitative and qualitative analysis of genes (Hadfield et al. 2011). A total of 20 μl of PCR product is divided into 20,000 droplets and amplified. Each droplet is calculated as a positive droplet (1) and a negative droplet (0) depending on the amplification. You can use QuantaSoft to check the value per copy of a sample using fast statistics (Figure 3) (Pinheiro et al. 2012).

Low levels of mutation can be detected in ctDNA due to the development of low ddPCR and NGS. NGS can detect multiple genes, but errors can occur during the sequencing process. ddPCR is simple to work with and can detect and quantify 0.01% mutations (Demuth et al. 2018). These advantages make them suitable for monitoring mutations during target therapy. In one study, mutations of ESR1 gene were detected after Aromatase Inhibitor (AI) treatment in 43 cases of ER-positive breast cancer with skin or liver metastasis. ddPCR in ctDNA was used to monitor ESR 1 mutations in patients receiving endocrine therapy (Wang et al. 2016).
Figure 3. Principle of droplet digital PCR

2. Purpose of this study

The aim of this study was to evaluate the correlation between ctDNA and tumor DNA through NGS in breast cancer patients and to confirm whether ctDNA can be used as a therapeutic marker.

Previous studies have shown higher concordance of > 90% in NGS results using ctDNA and tumor DNA (Chae et al. 2017). However, there are not many studies to compare with clinical data.
In this study, we hypothesized that genetic profile of Circulating tumor DNA (ctDNA) in breast cancer patients reflects the tumor DNA genome profile and it would have clinical implication.

1) To confirm association of genome profile of ctDNA and tumor DNA in breast cancer patients.
2) To evaluate methods of NGS and ddPCR for ctDNA genome profile.
3) To evaluate the association of NGS results of ctDNA for clinical relevance.

3. Materials and Methods and materials

3.1 Study design

Figure 4. Experiment scheme
The study recruited a total of 84 breast cancer patients with informed consent from the Institutional Review Board at the National Cancer Center from 2016 to 2018. Tissue and blood samples were collected from the patients. gDNA and ctDNA were extracted from tissue and plasma. Then, whole exome sequencing and targeted panel were performed on the tissue gDNA and ctDNA respectively. And then the variants and monitoring by ddPCR were validated. We also analyzed the association of clinical data with ctDNA results (Figure 4).

### 3.2 Sample collection and blood processing

Tumor biopsy at the time of diagnosis or tumor at the time of surgery was collected. And blood was collected before surgery or after chemotherapy. 12 ml of blood samples were collected from tubes containing baseline and EDTA, and 5 ml of blood samples were collected from patients who were able to follow up. Blood samples were processed within 2 hours after the collection. Centrifugation at 3000 rpm for 10 min at 4 °C was performed to extract ctDNA from plasma. Plasma was carefully centrifuged at 16,000 g for 10 min at 4 °C to remove the residual supernatant and any remaining contaminants including cells. 1 ml of separated plasma was aliquoted in a 1.5 ml tube and immediately stored in a deep freezer at -80 °C. Whole blood was also stored in a deep freezer at -80 °C after obtaining 1 mL aliquots for validation of germline mutation.
3.3 DNA extraction and quantitation

cT-DNA from plasma were extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Germany) from 2 mL of plasma according to the manufacturer’s instructions. cT-DNA samples were quantified using the Qubit dsDNA HS (High Sensitivity) Assay Kit (Life Technologies, Carlsbad, CA, USA). Genomic DNA was extracted from 1 ml of whole blood with a QIAamp DNA Mini Kit (Qiagen, Germany). Tumor DNA was isolated from formalin-fixed paraffin embedded (FFPE) tissue and fresh tissue using the Qiagen AllPrep DNA/RNA FFPE kit.

3.4 QIAseq Targeted DNA panels

20 ng (range 7.2-39.6) of cT-DNA was used to prepare the library using QIAseq Targeted DNA panels for breast cancer (DHS-001Z, QIAGEN) according to the manufacturer’s instructions. The QIAseq Targeted DNA panels for breast cancer include exonic regions of genes plus 10 bases to cover the intron and exon junction coding for 93 genes (Table 1). The library preparation and targeted enrichment preparations were performed using the Human Breast Cancer Panel and QIAseq 12-Index I. The procedure followed the QIAseq ™ Targeted DNA Panel Handbook. Next-generation sequencing was performed on Illumina Next Seq 500 High output using digital sequencing by single primer extension and molecular barcodes. Variants analysis was performed using the online software QIAseq Target DNA panel analysis according to the GATK pipeline. Somatic variants
were identified in COSMIC (Catalogs of Somatic Mutations in Cancer) or ClinVar or dbSNP. IGV browsers have confirmed the errors that occurs at the end of reading of the position (Figure 5) (Robinson et al. 2011, Robinson et al. 2017, Meacham et al. 2011). The average mean read depth was 21,991X and mean molecular barcode depth was 665X. The analysis used values of allele frequency of 1-10% and included the group with high and moderate annotation impact.

Table 1. List of genes covered by QIAseq Targeted Human Breast Cancer Panels

<table>
<thead>
<tr>
<th>QIAseq Targeted Human Breast Cancer Panels</th>
<th>ACVR1B</th>
<th>BMPR1A</th>
<th>CDKN2A</th>
<th>ERCC4</th>
<th>HERC1</th>
<th>MEN1</th>
<th>NEK2</th>
<th>PPM1L</th>
<th>SMAD4</th>
<th>XRCC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>BRCA1</td>
<td>CHEK2</td>
<td>ESR1</td>
<td>HOXB13</td>
<td>MLH1</td>
<td>NF1</td>
<td>PTEN</td>
<td>SMARCA4</td>
<td>ZBED4</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>BRCA2</td>
<td>CSMD1</td>
<td>EXOC2</td>
<td>IRAK4</td>
<td>MRE11A</td>
<td>PALB2</td>
<td>PTGFR</td>
<td>STK11</td>
<td>FGFR1</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>BRIP1</td>
<td>CTNNB1</td>
<td>EXT2</td>
<td>ITCH</td>
<td>MSH2</td>
<td>PALLD</td>
<td>RAD50</td>
<td>SYNE1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>CASP8</td>
<td>DIRAS3</td>
<td>FAM175A</td>
<td>KMT2C</td>
<td>MSH6</td>
<td>PBRM1</td>
<td>RAD51</td>
<td>TGFB1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td>CBFB</td>
<td>EGFR</td>
<td>FANCC</td>
<td>KRAS</td>
<td>MUC16</td>
<td>PCGF2</td>
<td>RAD51</td>
<td>TP53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXIN2</td>
<td>CCND1</td>
<td>EP300</td>
<td>FBXO32</td>
<td>MAP2K4</td>
<td>MUTYH</td>
<td>PIK3CA</td>
<td>RAD51D</td>
<td>TRAF5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAP1</td>
<td>CDH1</td>
<td>EPCAM</td>
<td>FGFR2</td>
<td>MAP3K1</td>
<td>MYC</td>
<td>PIK3R1</td>
<td>RB1</td>
<td>VHL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARD1</td>
<td>CDK4</td>
<td>ERBB2</td>
<td>GATA3</td>
<td>MDM2</td>
<td>N8N</td>
<td>PMS1</td>
<td>RET</td>
<td>WEE1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLM</td>
<td>CDK6</td>
<td>ERBB3</td>
<td>GEN1</td>
<td>MED12</td>
<td>NCOR1</td>
<td>PMS2</td>
<td>SEPT9</td>
<td>XRCC2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10
3.5 Oncomine Breast ctDNA Assay

The ctDNA sample was amplified using the Oncomine Breast cfDNA Assay v2, consisting of 10 nucleotide single nucleotide variations with mutations in the cancer and 3 copy number genes (CCND1, ERBB2, FGFR1) (Table 2). Oncomine™ Breast cfDNA Assay v2 contains 9 genes (~52 hotspots) composed of hotspots and TP53 gene with full length cover (~80% coverage). This library preparation was performed using the Oncomine™ Breast cfDNA Research Assay according to the Oncomine™ Cell-Free Research Assay User's Guide. Briefly, target was amplified by PCR using 2-40 ng of ctDNA. The target amplicons were
purified and the target amplicons were amplified with barcoded primers. The barcode libraries purified and size selected the barcoded library. The resulting libraries were quantified using the Ion Library TaqMan® Quantitation Kit (Thermo Fisher, USA). The prepared libraries were then sequenced on an Ion S5 XL Sequencer using an Ion 530™ kit and Ion 540™ kit (Thermo Fisher Scientific, USA). Oncomine TagSeq Breast V2 Liquid Biopsy Workflow 2. 5.6 version was used. The mutation was called by Ion Reporter visualization. Somatic variants were identified in COSMIC, ClinVar, dbSNP, ExAC, 5000 Exomes and the errors were confirmed using IGV browsers. Molecular coverage detected at least 2 or more. Somatic variants were identified using Sanger sequencing for allele frequencies of mutation ≥ 30 % (Figure 6). Limit of detection was calculated by dividing the molecular amplicon coverage by the minimum variant family minus 0.5. All samples in this study had more than 15,000,000 final library Ion Sphere Particles (ISPs), median read coverage of 39,704X (range 24,453~70,527), median molecular coverage of 2,333X (range 594-6,592).

Table 2. List of genes covered by Oncomine Breast cfDNA Assay v2

<table>
<thead>
<tr>
<th>AKT1</th>
<th>EGFR</th>
<th>ERBB2*</th>
<th>ERBB3</th>
<th>ESR1</th>
<th>KRAS</th>
<th>PIK3CA</th>
<th>TP53</th>
<th>FBXW7</th>
<th>SF3B1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCND1*</td>
<td>FGFR1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: 3 copy number genes
3.6 Droplet Digital PCR

ddPCR was performed on ctDNA and gDNA obtained from plasma and tissues for mutation detection. Each probe assays was obtained from Bio-rad (Hercules, CA, USA): AKT1 p.E17K: dHsaCP2000031 and WT: dHsaCP2000002, ERBB2 p.V842I: dHsaIS2504516 and WT: dHsaIS2504516 and dHsaIS2504517, KRAS p.G12D: dHsaCP2000001 and WT: DHsacP2000011 and WT: dHsaCP2000077, PIK3CA p.H1047R: dHsaCP2000077 and WT: dHsaCP200007 and WT: dHsaCP2000011 and WT: dHsaCP2000012, KRAS p.g12V: dHsaCP2000005 and WT: dHsaCP2000006, PIK3CA p.E542K: dHsaCP2000073 and WT:
dHsaCP2000074, dHsaCP2000078, TP53 p.R175H: dHsaCP2000105 and WT: dHsaCP2000106, TP53 p.R196 *: dHsaCP2000121 and WT: dHsaCP2000122, TP53 p.Y220C: dHsaCP2500536 and WT: dHsaCP2500537, p.R306 *: dHsaCP2500552 and WT: dHsaCP2500553. And ESR1 p.D538G, p.E380Q, p.Y537N, and p.Y537S were ordered from Life Technologies using primers and probes. The reaction was carried out in 20 μL of reaction volume according to the manufacturer’s recommendations, which was performed using the extracted DNA (9 μL), 2 × droplet PCR supermix for probe (10 μL), 20X mutation and reference primer / probe mix (FAM / HEX) (1 μL). The emulsion was made using a QX200 droplet generator (Bio-rad). The resulting droplets were transferred to a 96-well PCR plate and PCR was performed. The PCR conditions were 10 minutes at 95 °C, 30 seconds at 94 °C for 40 cycles, and 98 °C for 1 minute at 55 °C for 10 minutes. The PCR products were then incubated at 98 °C for 10 minutes. The sample plate was read in a QX100 droplet reader. Results were calculated by counting the number of positive and negative droplets using Bio-rad QuantaSoft software (Biorad-Laboratories, Pleasanton, CA, USA) and calculating the copy number of the target gene.

3.7 Detection limit of ddPCR

To assess the limit of detection (LOD) using ddPCR, wild type and mutant DNA of each gene was serially diluted to 50, 10, 1, 0.5, 0.25, 0.1, 0.05 and 0.01%. Types of genes and DNA were described in Table 3. The detection limit was confirmed
by ddPCR with serially diluted DNA. All samples were tested three times in duplicate. In table 3, the types of genes and positive and negative DNA used for LOD evaluation are shown.

**Table 3. The types of genes and positive and negative DNA used for LOD evaluation.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein change</th>
<th>Positive cell line or DNA</th>
<th>Negative cell line or DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>p.E17K</td>
<td>50% Mutant FFPE DNA (Horizon discovery, cat: HD167)</td>
<td>100% Matched Wild type FFPE DNA (Horizon discovery, cat: HD172)</td>
</tr>
<tr>
<td>ESR1</td>
<td>p.D538G</td>
<td>synthetic DNA (Life Technologies)</td>
<td>synthetic DNA (Life Technologies)</td>
</tr>
<tr>
<td>KRAS</td>
<td>p.G12S</td>
<td>50% Mutant DNA (Horizon discovery, cat: HD288)</td>
<td>100% Matched Wild type DNA (Horizon discovery, cat: HD710)</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>p.E545K</td>
<td>MCF-7</td>
<td>MDA-MB-231</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>p.H1047R</td>
<td>T47D</td>
<td>MDA-MB-231</td>
</tr>
<tr>
<td>TP53</td>
<td>p.R175H</td>
<td>CCFR-CEM cell line</td>
<td>MCF10A</td>
</tr>
<tr>
<td>TP53</td>
<td>p.R306*</td>
<td>MOLT-4</td>
<td>MCF10A</td>
</tr>
</tbody>
</table>

**3.8 Statistical analysis**

All analyses were performed with GraphPad Prism 5.0 software (GraphPad Software Inc., La Jolla, CA, USA). To compare the correlation of the allele frequency values between the ctDNA panel and the ddPCR, we used a two-tailed test using the Pearson correlation coefficient. The number of median variants detected in the panel and the survival curve of the patient were calculated using Kaplan-Meier (KM) curves. The log-rank test was used to compare the survival
curves. Progression free survival (PFS) was measured from the day of diagnosis to the day of progression or death. Overall survival (OS) was calculated from the day of diagnosis to the day of last follow-up or death. The effects of the number of median variants detected in the panel on OS or PFS were presented as hazard ratio (HR) and 95% confidence interval (CI).

4. Result

4.1 Characteristics of breast cancer patients

A total of 84 patients were enrolled in this study. The characteristics of the study subjects are shown in Table 4. The median age of the study subjects was 47 (range 23-74). The immunohistochemistry (IHC) subtype at the time of diagnosis was triple-negative breast cancer (TNBC), hormone receptor (HR)+ / human epidermal growth factor receptor 2 (HER2) -, HR- / HER2 +, HR + / HER2 + and the numbers were 45 (53.6%), 32 (38.1%), 5 (6.0%) and 2 (2.4%) respectively. NGS of ctDNA was performed on 84 patients. We performed tumor NGS in 49 patients to compare with ctDNA results. Thirty-three conserved FFPE and 16 fresh tissue
samples were obtained from biobank. The median interval between tissue and blood sample collection was 6 months. 22 (44.1%) were less than 3 months and 27 (55.1%) were more than 3 months. Of the 84 patients who performed NGS using ctDNA, 74 had mutations detected.

Table 4. Characteristics of breast cancer patient with tissue and ctDNA NGS

<table>
<thead>
<tr>
<th></th>
<th>All patients (N = 84)</th>
<th>Patients with Matched Tissue NGS (N = 49)</th>
<th>Mutation Detected in ctDNA (N = 84)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>84</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>Median age (range)</td>
<td>47 (23-74)</td>
<td>47 (28-68)</td>
<td>47 (23-74)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>31</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>III</td>
<td>22</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>IV</td>
<td>28</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Subtype</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNBCa</td>
<td>45</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>HR+/HER2 c</td>
<td>32</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>HR-/HER2+</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>HR+/HER2+</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Interval between tissue and blood sample collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median month (range)</td>
<td></td>
<td>6 (0-73)</td>
<td></td>
</tr>
<tr>
<td>≤3 months</td>
<td>22</td>
<td>44.9</td>
<td></td>
</tr>
<tr>
<td>&gt; 3 months</td>
<td>27</td>
<td>55.1</td>
<td></td>
</tr>
</tbody>
</table>

*TNBC: triple-negative breast cancer, HR: hormone receptor, HER2: human epidermal growth factor receptor 2

4.2 NGS in ctDNA and tumor DNA

NGS were performed using ctDNA in 84 breast cancer patients. The QIAseq Targeted DNA panels were performed in 53 patients and the Oncomine Breast
cfDNA panels in 38 patients. Mutation and copy number alterations (CNA) were detected in 74 of 84 patients including 73 variants, 8 CNAs, and 7 mutations and CNAs were detected with sequencing (Figure 7). And 10 patients had no mutations. Of the 317 mutations detected in 84 genes, 308 variants and 9 CNAs were detected. Mutations were most commonly identified in KMT2C (33%), TP53 (16%) and MUC16 (8%). Other genes were PIK3CA (5%), ESR1 (5%) and ERBB2 (2%) (Figure 8, 9).

Figure 7. Number of patients with variants and CNA in ctDNA of breast cancer patients by NGS. 73 mutations, 8 CNAs, and 7 mutations and CNAs were detected.

Figure 8. The proportion of variants detected in NGS using ctDNA in breast cancer patients.
Figure 9. The frequencies of detected gene mutations and copy number alterations in ctDNA panel in breast cancer patients.
Variants were detected in 50 patients with QIAseq Targeted DNA panels and 31 patients with Oncomine Breast cfDNA panels. Variants were identified in 33/93 (35%) in QIAseq Targeted DNA and 10/12 (83%) in Oncomine Breast cfDNA panels of the genes sequenced. The median number of mutated genes per patient in the QIAseq Targeted DNA panels was 2.0 (SD ±2.2, range: 0-10) and 1.0 (SD ±1.2, range: 0-5) in the Oncomine Breast cfDNA panel (Figure 10,11). The median number of total variants per patients in the QIAseq and Oncomine panels was 4.0 (SD ±3.4, range: 0-20) and 2.0 (SD ±2.8, range: 0-11), respectively.

Whole-exome sequencing (WES) was performed in 49 breast tumor tissues, 33 of which were archival formalin-fixed paraffin-embedded (FFPE) and 16 from fresh tissue. 44 (90%) patients had one or more mutations and 1 patient had CNA. 50 genes and 232 mutations were identified. The median number of mutated genes per patients was 3.0 (SD ±3.5, range: 0-14) and the median number of total mutations per patients in the WES was 4.0 (SD ±4.6, range: 0-22) (Figure 12).
Figure 10. Detection of variants in ctDNA using QIAseq Targeted DNA panels in 53 breast cancer patients. Variants were found in 50 of 53 breast cancer patients and median number of variants per patient was 4.0 (SD ± 3.4, range: 0-20).
Figure 11. Variants and CNA results of 38 patients with Oncomine breast ctDNA panel. Variants and CNAs were found in 31 of 38 breast cancer patients and the median number of variants per patient was 2.0 (SD ± 2.8, range: 0-11).
Among 49 patients with breast cancer, 44 patients had variants and 1 patient had CNA. And the median number of variants per patients was 4.0 (SD ± 4.6, range: 0-22).

Figure 12. Results of detected variants and CNA in WES using Tumor DNA.
4.3 Comparison of two panel assay results

We studied ctDNA with two panels, QIAseq Targeted DNA panels and Oncomine Breast cfDNA panels. Seven patients sequenced using both panels. Eight mutations and three CNA genes are common in both panels. At least one mutation was identified in all 7 patients. Six genes and 18 mutations were detected. A total of 13 mutations were commonly detected and concordance was 96% (Table 5). The range of variant allele frequency was 0.53-68.06 and both panels showed similar percentages (Figure 13). QIAseq Targeted DNA panels showed a slightly higher allele frequency. Four mutations were found on the Oncomine panel but not on the Qiagen panel and one mutation was detected only on the Qiagen panel. In one patient, amplification was confirmed in the CCND1 gene by the Oncomine panel only.

<table>
<thead>
<tr>
<th>QIAseq Targeted DNA panels</th>
<th>Oncomine Breast cfDNA panel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>Positive</td>
<td>13</td>
</tr>
<tr>
<td>Negative</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5. Concordance of mutation between QIAseq Targeted DNA and and Oncomine Breast cfDNA
Figure 13. Comparison of allele frequency of Oncomine and QIAseq cfDNA panel. The frequency of variant alleles ranged from 0.53 to 68.06. Four mutations were found only in the Oncomine cfDNA panel and one mutation was found only in the QIAseq cfDNA panel.
4.4 Concordance between ctDNA and tumor DNA

Mutations were compared in 49 patients who performed NGS using tumor DNA and ctDNA. The most frequently detected genes in ctDNA were \textit{KMT2C}, \textit{TP53}, and \textit{MUC16}, and the Tumor DNA were \textit{TP53}, \textit{KMT2C}, and \textit{MUC16}. The genes most frequently found in ctDNA or tumor DNA were \textit{TP53}, \textit{KMT2C}, and \textit{MUC16} (Figure 14). The average number of genes per patient was 3.0 (SD ±1.8, range: 1-7). A total of 28 genes were detected in tumor DNA or ctDNA. There were a total of 9 genes mutated in both ctDNA panel and WES and the top 4 mutated genes were \textit{TP53}, \textit{MUC16}, \textit{PIK3CA}, and BRCA1 in both tumor DNA and ctDNA (Figure 15).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure14}
\caption{Results of detected variants in ctDNA or tumor DNA. The most commonly detected genes in 49 patients who performed NGS using ctDNA and tumor DNA were \textit{TP53}, \textit{KMT2C} and \textit{MUC16}.}
\end{figure}
We compared 7 genes that were covered in QIAseq Targeted DNA panels and Oncomine Breast ctDNA panels and WES analysis. Variants were detected in 41 (84%) patients, 23 (47%) patients had more than one mutation in ctDNA, and 33 (67%) patients had more than one sample in tumor DNA. 53 and 50 mutations were found in ctDNA and tumor DNA respectively, and 5 variants were detected by both. In Figure 16 A, red color defined that variants were detected at the same position in plasma and tissue. An orange color indicates that one variant has concordant variant and at least one discordant variant in the same gene. Purple means that there are different variants in the same gene, blue in the tissues only, and green in variants detected only in ctDNA. As a result, the concordance rate of tumor DNA with ctDNA was 81%. And the concordance on positives was defined as the detection of mutations in both ctDNA and tumor DNA at the same gene location and it was 7.1%. The median time interval between tissue and blood collection was 6 months (range: 0-73 months). We compared the results of <10, 10-30, and >30 months to compare the differences in tissue and blood collection time with mutation correlations. Positive concordance of tissue and blood according to collection intervals of <10, 10-30, and >30 months were 9.3%, 6.3% and 0%, respectively (Figure 16B).
Figure 15. Total of 28 genes detected in patients per gene in NGS results using tumor DNA or ctDNA. The top three mutant genes were KMT2C, TP53, and MUC16 in tumor DNA or ctDNA.
Figure 16. Comparison of variants in ctDNA and tumor DNA.  

A. The concordance of variants detection between matched tumor and ctDNA was 81.0%. However, the rate of variants detected at the same position in both ctDNA and tumor DNA was 7.1%.  

B. Positive concordance in groups with differences in tissue and blood sampling periods of less than 10 months was higher than in groups of 10-30 months (9.3%).  

C. There were five mutations detected in both ctDNA and tumor DNA at the same gene location.

*Concordance = 81%
4.5 Association between ctDNA NGS and ddPCR

To validate NGS results using ctDNA, we compared the frequency of ctDNA mutations with ddPCR. The frequency of 56 mutations in the hotspot mutations of AKT1, ESR1, ERBB2, PIK3CA, KRAS, and TP53, which are well known as therapeutic markers of cancer, were compared. The results of NGS and ddPCR showed high association ($R^2 = 0.9731$, $p < 0.0001$) (Figure 17).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein change</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>p.E17K</td>
<td>4</td>
</tr>
<tr>
<td>ERBB2</td>
<td>p.V842I</td>
<td>1</td>
</tr>
<tr>
<td>ESR1</td>
<td>p.D538G</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>p.Y537N</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>p.Y537S</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>p.G12D</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p.G12S</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p.G12V</td>
<td>1</td>
</tr>
<tr>
<td>KRAS</td>
<td>p.E542K</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p.E545K</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>p.H1047R</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>p.R175H</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>p.R248Q</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>p.R248W</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>p.R306*</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>p.Y220C</td>
<td>8</td>
</tr>
</tbody>
</table>

**Total Number** 56

Figure 17. Association between NGS and ddPCR for detection of mutation in ctDNA.
4.6 Survival according to variants detection

This study identified the association between presence of detected mutation and survival using KM curve (Figure 18). ctDNA detection and non-detection were not associated with PFS in breast cancer patients \((P = 0.7420)\). The OS of the patients with detected mutations in ctDNA showed a tendency to be shorter than that of the patients without detected mutation. However, it was not statistically significant \((P = 0.4875)\).

We confirmed whether the number of variants detected was associated with survival (Figure 19). We have identified PFS and OS according to the number of mutations detected using QIAseq Targeted panel and Oncomine Breast cfDNA panels, respectively. Survival was analyzed according to the median number of mutations detected in each panel. The QIAseq Targeted panel was performed on ctDNA in 53 breast cancer patients and the median of the detected mutations was 4. Oncomine Breast cfDNA panels were performed in 38 patients, with a median of 2 detected mutations. PFS according to the number of mutations in the Qiaseq panel was not significantly different. Patients with \(\leq 4\) variants in ctDNA had a shorter OS than patients with \(>4\) variants but were not statistically significant \((P = 0.5325)\). Similarly, the PFS according to the number of mutations in the Oncomine panel was not significantly different. Patients with more than two variants showed poorer OS than patients with less than two variants in the Oncomine panel \((P = 0.0282, \text{HR} = 4.9925, 95\% \text{ CI}: 1.19-20.99)\). The median overall survival periods
of patients with >2 mutations and patients with ≤2 mutations were 27 and 47 months, respectively.

Next, we analyzed the survival rate according to the number of mutations in the stage group. Using the QIAseq Targeted and Oncomine Breast cfDNA panels, we identified PFS and OS according to stage group of 1, 2 and 1, 2, 3. Then, we analyzed survival with the cut point median of the number of variants detected. In the Qiaseq and Oncomine panels of the PFS, there were no significant differences according to the number of variants in the stage group (Figure 20). In the oncomine panel, stage 1, 2 ($P = 0.0215$, HR = 38.39, 95% CI: 1.7-859.7) and 1, 2, 3 groups ($P = 0.0308$, HR = 12.90, 95% CI: 1.3-131.4) showed poorer OS in patients with variant number of > 2 than ≤ 2 (Figure 21).

![Figure 18. KM curve according to the presence of ctDNA variants.](image)

PFS and OS did not show statistically significant difference between ctDNA detected and not detected groups.
Figure 19. KM curve according to the number of variants detected in ctDNA panel. (A) Patients with >4 variants in ctDNA had a better OS than patients with ≤4 variants in QIAseq Targeted panel but were not statistically significant ($P = 0.5325$). (B) OS of patients with >2 variants were statistically significantly worse than patients with ≤2 variants in Oncomine Breast cfDNA panel ($P = 0.0282$).
Figure 20. KM curve of PFS according to number of mutations per stage in QIAseq Targeted and Oncomine Breast cfDNA panel. Analysis was performed on the QIAseq (A) and Oncomine panels (C) after dividing into stage 1, 2 and 3, 4 group. Figures B and D were divided into stage 1, 2, 3 group and stage 4 group in each panel. Stage groups did not showed association with mutations in PFS.
Figure 21. KM curve of OS according to number of mutations per stage in QIAsseq Targeted and Oncomine Breast cfDNA panel. A, C. QIAsseq Targeted panel showed no statistically significant differences according to the number of variants in the stage group. B, D. In the Oncomine Breast cfDNA panel, stage 1, 2 ($P = 0.0215$) and stage 1, 2, 3 groups ($P = 0.0308$) showed poorer OS in group with less than 2 mutations.
4.7 Identification of the molecular mechanism and clinical significance

We compared molecular mechanisms for 10 genes, breast cancer-related studies, and clinical significance. Previous reports and molecular mechanisms for 10 genes in breast cancer are summarized in table 6. Studies related to breast cancer such as mutations and deletions of genes that have been associated with survival or drug sensitivity have been conducted. Among them, KMT2C and MUC16 have been studied for the expression of mRNA, but studies on DNA have been lacking. It is also found in many cancers, but its role in tumorigenesis is uncertain (Pereira et al. 2016b). Therefore, clinical significance was confirmed for 8 genes except for KMT2C and MUC16 genes.

For each gene, the domain and the position of the detected mutation were indicated (Figure 22). Circles indicate the number of patients with mutations detected in ctDNA. Red is pathogenic, orange is likely pathogenic, green is benign and gray means unknown. Among them, pathogenic clinical significance was shown in TP53, PIK3CA, ERBB2, ATK1 and ESR1 genes. Therefore, we decided to monitor the changes in the frequency of ctDNA allele in four genes in patients who were able to follow up.
Table 6. Concordance of mutation between QIAseq Targeted DNA and Oncomine Breast cfDNA

<table>
<thead>
<tr>
<th>Gene</th>
<th>Research Description</th>
<th>Molecular mechanism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KMT2C</strong></td>
<td>One study reported a co-regulator of ER and a prognostic marker for ER+ breast cancer. In contrast, in another study, KMT2C showed high mutation but survival was not significant. KMT2C has a frequency of about 8% in the tumors of breast cancer patients. KMT2C has been studied for the expression of mRNA and it was significantly associated with short survival when KMT2C was deleted. However, DNA studies rarely progressed and were not included in the panel's gene using ctDNA.</td>
<td>Chromatin modification</td>
<td>(Gala et al. 2018, Liu et al. 2015) (Wang et al. 2011) (Manso et al. 2016)</td>
</tr>
<tr>
<td><strong>TP53</strong></td>
<td>As a tumor suppressor gene, it affects cell proliferation, apoptosis, and DNA repair. Mutations in TP53 lead to uncontrolled reproduction of damaged cells. It is found in almost all types of cancer and is also implicated in hereditary cancers such as Li-Fraumeni syndrome.</td>
<td>DNA damage control</td>
<td>(Yin et al. 2002) (Marcel et al. 2010) (Olivier et al. 2006)</td>
</tr>
<tr>
<td><strong>MUC16</strong></td>
<td>CA-125, also known as MUC16, is a tumor antigen used for ovarian cancer monitoring.</td>
<td>-</td>
<td>(Pereira et al. 2016a) Genecards Onim</td>
</tr>
<tr>
<td><strong>PIK3CA</strong></td>
<td>The most frequent marker of ER-positive and HER2-positive breast cancer is the utility of PIK3CA mutation as an early marker of treatment response in advanced breast cancer.</td>
<td>PI3K pathway</td>
<td>(O'Leary et al. 2018) (Nicolini, Ferrari, and Duffy 2018) (Kodahl et al. 2018)</td>
</tr>
<tr>
<td><strong>BRCA1</strong></td>
<td><em>BRCA1</em> is a gene with clinical significance that identifies hereditary cancers in breast cancer patients. <em>BRCA1</em> mutations also indicate sensitivity to the PARP inhibitor Olaparib drug.</td>
<td>cell cycle checkpoints, DNA damage response</td>
<td>(Duffy et al. 2015) (Slavin et al. 2018) (Berghuis et al. 2017, Hu et al. 2018) (Deng 2006)</td>
</tr>
<tr>
<td><strong>NF1</strong></td>
<td>NF1 is found in a variety of cancers and is associated with cancer risk and drug resistance. NF1 mutations and deficiencies in breast cancer patients were associated with estrogen receptor signaling and decreased survival.</td>
<td>Ras/MAPK signalling,</td>
<td>(Dischinger et al. 2018) (Usitalo et al. 2017)</td>
</tr>
<tr>
<td><strong>AKT1</strong></td>
<td>AKT1 is associated with the PI3K signaling pathway and inhibits invasion and metastasis by inhibiting the migration of breast cancer cells.</td>
<td>PI3K pathway</td>
<td>(Ooms et al. 2015)</td>
</tr>
<tr>
<td>Gene</td>
<td>Research Description</td>
<td>Molecular mechanism</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>ERBB2</td>
<td>ERBB2 receptor overexpression induces tumor growth by constitutively activating MAPK and PI3K / AKT signaling pathways and is associated with a poor prognosis</td>
<td>Growth factor receptors</td>
<td>(Contino et al. 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Nicolini, Ferrari, and Duffy 2018)</td>
</tr>
<tr>
<td>CDH1</td>
<td>CDH1 is associated with developing diffuse gastric and lobular breast cancer. Deletion of CDH1 activates T-regulatory (Treg) cell signaling and target immune checkpoint pathways as well as the formation of pseudotumors of invasive lobular carcinoma with immune cell invasion.</td>
<td>Cell cycle</td>
<td>(An et al. 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Corso et al. 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(van der Post et al. 2015)</td>
</tr>
<tr>
<td>ESR1</td>
<td>The ER (ESR1) gene is associated with resistance to endocrine therapy.</td>
<td>Transcriptional regulation</td>
<td>(Du et al. 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Schiavon et al. 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Wang et al. 2016)</td>
</tr>
</tbody>
</table>
Figure 22. The position of detected mutation and clinical significance for each gene.
4.8 Limit of detection (LOD) and limit of quantification (LOQ) in ddPCR

To evaluate the detection limit of ddPCR probe, mutant DNA was serially diluted with wild type DNA at 50, 10, 1, 0.5, 0.25, 0.1, 0.05 and 0.01%. The limit of quantification (LOQ) was measured to evaluate a reliable quantitative value with precision and accuracy (Dobnik et al. 2016). LOQ was determined as a sample with a coefficient of variation (CV) of less than 25% for all amplicons (George et al. 2013, Floren et al. 2015, Dobnik et al. 2016). The LOD and LOQ of all samples were tested three times in duplicate. Based on the results, we determined that the LOD of AKT1 p.E17K, ESR1 p.D538G, KRAS p.G12S, PIK3CA p.E545K, p.H1047R, TP53 p.R175H, p.R306 * was 0.25, 1.00, 0.50, 1.00, 0.10, 0.50 and 0.50%, respectively (Figure 23).
4.9 Longitudinal monitoring of somatic variants in ctDNA

Variants were detected in tumor DNA or ctDNA in 81 patients. We performed serial monitoring of somatic variants in ctDNA in 9 patients. Thirty-two patients were excluded from monitoring due to death, patient refusal or unavailable blood sample. After designing the probe and establishing the ddPCR condition, continued
monitoring in 40 patients was necessary (Figure 24). Monitoring was performed by ddPCR for allele frequencies for somatic variants in ctDNA of \textit{PIK3CA} (p.E545K, p.H1047R) and \textit{TP53} (p.R175H, p.Y220C, p.R306 *) genes.

Baseline was defined as the day when tumor tissue was obtained. NGS was performed on the day the blood was first collected and somatic mutation was confirmed. Figure 22 shows the change in allele frequency (%) of mutations in ctDNA using ddPCR at the time of blood collection. Tumor marker values of CA15-3 (U / mL), Serum Her-2 (ng / mL) and CEA (ng / mL) were also shown. Nine patients were classified as having increased, decreased or not detected variant allele frequency (VAF).

**Figure 24. Detection limit of ddPCR.** Changes in allele frequencies of somatic variants in ctDNA were monitored by ddPCR in 9 of 81 patients. \textit{PIK3CA} and \textit{TP53} genes were monitored.
Patient 64 (TNBC) was diagnosed with breast cancer in September 2016 and PIK3CA p.H1047R (VAF 10%) was detected in tumor gDNA in February 2017. However, the variant was not detected in NGS and ddPCR results in ctDNA 2 months later. In the second blood sample 13 months later, variant was not detected in ctDNA and the CA15-3 concentration was normal. She was treated with Paclitaxel, but the disease progressed. She received treatment with the combination of Vinorelbine and Cisplatin but did not respond. And variants of PIK3CA p.H1047R (VAF 6.6%) were detected in ctDNA. Other drugs such as Vinorelbin / Cisplatin, Exemestane / Everolimus and Irinotecan / Cisplatin did not reflect response to treatment. The variant AF (VAF 11%) and CA15-3 concentration elevated with disease progression (Figure 25A).

Patient 37 (estrogen receptor (ER) negative, progesterone receptor (PR) negative, HER2 positive) was diagnosed with breast cancer in June 2016 and tumor tissue was obtained in August 2017. She was treated with herceptin and capecitabine. And radiation therapy (RT) was administered at 1 month. No mutations were found in the NGS results of the tumor tissues. At 6 and 12 months, no mutations were found in ctDNA. At 21 months, a mutation of PIK3CA p.E545K was found (VAF 0.16) but was very low. The patient is currently being followed up without recurrence of the disease (Figure 25B).

Patient 74 (ER positive, PR positive, HER2 negative) was diagnosed with breast cancer in October 2013. In November 2016, TP53 p.R175H mutation (VAF 37.96%) was found in NGS using tumor tissue. She received therapy of fulvestrant
combined with palbociclib but the tumor metastasized to liver. She received Paclitaxel treatment but had intolerable adverse reaction. At 13 months, variants of $TP53$ p.Y220C (VAF 0.08%), p.R175H (VAF 2.43%) and p.R306* (VAF 0.61%) were detected in the ctDNA using NGS (Figure 25C).

Patients 46 (ER positive, PR positive, HER2 negative) was diagnosed in July 2014 and variants were not detected in tumor tissues in February 2018. One month later, $TP53$ p.R306* (VAF 0.8%) was detected in ctDNA and tumor markers such as CA15-3 and CEA were detected as high. At 6 months, the VAF of mutations in ctDNA with CA15-3 and CEA decreased to 0.25%. However, the disease has progressed and is being treated with eribulin (Figure 25D).

Patients 82 (TNBC) was diagnosed in November 2016 and variants were not detected in tumor tissues in May 2018. $TP53$ p.R306* (VAF 0.12%) variant was detected in the patient's ctDNA at 1 month. Variant was not detected at 6 months after modified radical mastectomy (MRM) at 5 months. The patient underwent RT at 7 months and continues to follow (Figure 25E).

Patient 42 (TNBC) was diagnosed in March 2017. In August, MRM was performed to obtain tumor tissue and blood was collected. The $TP53$ p.Y220C variant was detected in blood ctDNA using NGS (Qiagen panel VAF 14.36%, Oncomine panel VAF 9.25%) and ddPCR (VAF 4.40). One month later, the patient was treated with capecitabine and underwent RT. In the second collection, ctDNA VAF decreased to 0.50% (Figure 25F).

Patient 45 (ER positive, PR positive, HER2 negative) was diagnosed in
September 2017 and *PIK3CA* p.E545K (VAF 11.3%) was detected in tumor gDNA in February 2018. A low allele frequency variant (VAF 0.10%) was found in ctDNA but less than LOD. RT was performed at 1 month and variant was not detected at 8 months (Figure 25G).

Patient 63 (TNBC) was diagnosed in July 2014 and *PIK3CA* p.H1047R (VAF 26%) was detected in the tumor tissue. The first blood collection was obtained 25 months after tumor tissue and variation was not detected in ctDNA (Figure 25H).

Patient 67 (TNBC) was diagnosed in July 2014 and *PIK3CA* p.E545K (VAF 8%) was detected in the tumor tissue. Tumor tissues and blood sampling intervals were 39 months, and variant was not detected in ctDNA during monitoring (Figure 25I).
Figure 25. Serial monitoring of somatic variants in ctDNA. Clinical results in 7 patients were consistent with the changes in detected variants (A-G), and 2 patients without variants did not show clinical relevance (H-I).
5. Discussion

In this study, we demonstrated that ctDNA reflects the genomic profile of matched tumor DNA in breast cancer patients using NGS. We identified changes of the allele frequency in ctDNA according to therapeutic response and values of tumor marker.

First, we performed NGS using ctDNA panel in breast cancer patients. As a result, variants of *KMT2C, TP53, MUC16, AR* and *PIK3CA* genes were most frequently detected. Previous reports have shown that *TP53, PIK3CA, ERBB2* and *KRAS* variants were most commonly identified in ctDNA panel from breast cancer patients (Frenel et al. 2015, Chae et al. 2017, Maxwell et al. 2017). When comparing with other studies, the *TP53* and *PIK3CA* genes were the most commonly detected.

Previous studies have compared tumor DNA and ctDNA results with NGS, showing 50-88% concordance (Adalsteinsson et al. 2017, Thompson et al. 2016, Chen et al. 2016, Beije et al. 2016). In our study, the concordance rate for detection of variants between matched tumor and ctDNA was 81.0%. However, the rate of variants detected at the same position in both ctDNA and tumor DNA was 7.1%. Similar to our study, a study by Chae et al. showed high concordance of 94.2% and concordance on positives of 10.8% (Chae et al. 2017).

The reason for low concordance rate of variants between tumor and ctDNA is the wide interval between tumor and blood sample collection (median: 6 months, range: 0-73 months). Four out of five mutations detected at the same position were
0-1 months interval between tumor and blood collection. Positive concordance in groups with differences in tissue and blood sampling periods of less than 10 months was higher than in groups of 10-30 months. In another study, a group with a time interval difference of ≤ 2 weeks (100%) than a group of > 6 months (60%) showed higher concordance in 50 lung cancer patients (Thompson et al. 2016). As a result, reducing the time interval between samples is important in studies to confirm the concordance.

The sequencing depth, gene frequency threshold and gene coverage position also may be the reason for the low concordance between matched tumor and ctDNA. In our study, the LOD of WES using tumor DNA was 5% and the depth was 200X. The LOD of the QIAseq targeted DNA panels and Oncomine Breast cfDNA using ctDNA was 1% and 0.05-0.35%, and the mean read depth was 21,991X and 39,704X, respectively. Since ctDNA panels have lower LOD than WES using tumor DNA, more variants can be detected. Several previous studies have compared hotspot genes position and showed high concordance between ctDNA and tumor DNA. A comparison of exon 19 deletion and L858R EGFR mutation in NSCLC patients showed a high agreement of 80-98% (Zhu et al. 2015). In the prostate cancer study, there were 89% agreement between the 9 genes including the driver genes AR, BRCA2, and ATM (Wyatt et al. 2017). In breast cancer, tissue and ctDNA showed high agreement with PIK3CA mutation and ERBB2 amplifications (Liang et al. 2016). In our study, QIAseq Targeted DNA panel includes the position of a broader range of genes as well as hotspot. Therefore, our
study has lower concordance than other studies.

Our study evaluated the clinical relevance of ctDNA to NGS results. Patients with more than two variants showed poor survival in OS than patients with less than two mutations in the Oncomine panel ($P = 0.0282, \text{HR} = 4.9925, 95\% \text{ CI: 1.19-20.99}$). Also, patients with two or more mutations had worse OS in stages 1, 2 and 1, 2, 3 group. Many studies have shown that detected variants are associated with survival of cancer patients. In other study, the mutant negative group had a better OS than the positive group in solid tumor ($\text{HR} = 0.26, 95\% \text{ CI 0.1409-0.9520, } P <0.04$) (Mehrotra et al. 2018). In other studies, TNBC and early breast cancer patients with variants in ctDNA had significantly shorter disease-free survival (DFS) than patients without variants (Garcia-Murillas et al. 2015, Chen et al. 2017).

To confirm the usefulness of the ctDNA, we monitored the changes of VAF according to therapeutic response and values of tumor marker. In our study, \textit{PIK3CA} and \textit{TP53} genes were monitored. The variant was simultaneously increased in patients with PD, and the variant was decreased in patients who underwent RT. In addition, the change in mutant allele frequency was consistent with the tumor marker measurements of breast cancer. However, several mutations were detected at a value below the ddPCR LOD. It was also monitored in the ctDNA of a small number of patients. Therefore, we need to perform more patient monitoring.

In summary, we have confirmed that the genetic profile of ctDNA in breast cancer patients reflects the tumor DNA genome profile. More mutations in ctDNA have
been associated with poor OS. We evaluated the utility of ctDNA by identifying changes in VAF according to the therapeutic response and the value of tumor markers. In further studies, we need to monitor a larger number of patients to assess whether ctDNA results can be helpful in clinical decisions. In addition, we need to find out abnormal variants as prognostic markers in ctDNA from breast cancer patients and monitor therapeutic responses by specific target drugs for variants.
BIBLIOGRAPHY


10.1038/s41598-018-22312-x.


ACKNOWLEDGEMENT

Many people have provided valuable time and generous support. I sincerely thank Prof. Sun-Young Kong for advice and guidance. With her help, I was able to broaden my knowledge and complete my thesis. Also, I sincerely appreciate all my thesis committee members, Prof. In Hae Park and Seung-Tae Lee for their valuable comments.

I want to thank Junnam Lee, Eun Hae Cho of Greencross genome, and Charny Park of National Cancer Center for their excellent support for NGS analysis. I am grateful to Han Na Yang, who provided technical support, and Kyoung Hee An, who helped to collect the specimen and clinical data.

I would like to express my gratitude to my colleagues Min Kyeong Kim, Jin-Sun Ryu for their care and great help for my thesis.

Finally, I would like to thank my parents for their generous support. My parents helped me mentally in completing my thesis.