ORIGINAL ARTICLE

Cancer Science WILEY

Short somatic alterations at the site of copy number variation in breast cancer

Fumi Murakami[1,2,3] | Yumi Tsuboi[1] | Yuka Takahashi[2] | Yoshiya Horimoto[2] |
Kaoru Mogushi[3] | Takeshi Ito[1] | Mitsuru Emi[4] | Daisuke Matsubara[1,5] (ID) |
Tatsuhiro Shibata[6] (ID) | Mitsue Saito[2] | Yoshinori Murakami[1] (ID)

[1]Division of Molecular Pathology, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

[2]Department of Breast Oncology, Juntendo University, Tokyo, Japan

[3]Juntendo, University Graduate School of Medicine, Tokyo, Japan

[4]University of Hawaii Cancer Center, Honolulu, HI

[5]Department of Pathology, Jichi, Medical University, Shimotsuke, Japan

[6]Laboratory of Molecular Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Correspondence
Yoshinori Murakami, Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
Email: ymurakam@ims.u-tokyo.ac.jp

Abstract

Copy number variation (CNV) is a polymorphism in the human genome involving DNA fragments larger than 1 kb. Copy number variation sites provide hotspots of somatic alterations in cancers. Herein, we examined somatic alterations at sites of CNV in DNA from 20 invasive breast cancers using a Comparative Genomic Hybridization array specifically designed to detect the genome-wide CNV status of approximately 412 000 sites. Somatic copy number alterations (CNAs) were detected in 39.9% of the CNV probes examined. The most frequently altered regions were gains of 1q21-22 (90%), 8q21-24 (85%), 1q44 (85%), and 3q11 (85%) or losses of 16q22-24 (80%). Gene ontology analyses of genes within the CNA fragments revealed that cascades related to transcription and RNA metabolism correlated significantly with human epidermal growth factor receptor 2 positivity and menopausal status. Thirteen of 20 tumors showed CNAs in more than 35% of sites examined and a high prevalence of CNAs correlated significantly with estrogen receptor (ER) negativity, higher nuclear grade (NG), and higher Ki-67 labeling index. Finally, when CNA fragments were categorized according to their size, CNAs smaller than 10 kb correlated significantly with ER positivity and lower NG, whereas CNAs exceeding 10 Mb correlated with higher NG, ER negativity, and a higher Ki-67 labeling index. Most of these findings were confirmed or supported by quantitative PCR of representative DNA fragments in 72 additional breast cancers. These results suggest that most CNAs are caused by gain or loss of large chromosomal fragments and correlate with NG and several malignant features, whereas solitary CNAs of less than 10 kb could be involved in ER-positive breast carcinogenesis.

KEYWORDS
breast cancer, chromosomal aneuploidy, copy number alteration, copy number variation, genomic instability

---

- -

1 | INTRODUCTION

Genomic instability is one the forces driving cancer development and progression.[1] Microsatellite instability in specific types of cancer triggered by loss of function of DNA mismatch repair enzymes causes deletions or insertions of mononucleotide to tetranucleotide repeats in the human genome, some of which inactivate tumor suppressor genes, thereby endowing cancer cells with malignant features. Chromosomal instability, triggered mainly by dysfunction of the chromosomal segregation apparatus in mitosis, is more frequently observed than other genomic instabilities in human solid cancers. It causes aneuploidy and large chromosomal alterations in the human genome, some of which activate groups of oncogenes, inactivate groups of tumor suppressor genes, and lead to imbalanced expression of a large number of genes, and thereby promote malignancy and/or confer individual features of cancer cells, including drug resistance.[2]

Copy number variation is a relatively recently identified polymorphism in the human genome involving DNA fragments larger than 1 kb.[3] Copy number variations are present in the genomes of healthy individuals as polymorphisms. It has been shown that some CNVs are associated with susceptibilities to various human diseases.[4] In cancer, CNVs provide hotspots for somatic alterations, called CNAs. Furthermore, chromosomal aneuploidy or gain/loss of large chromosomal fragments in cancer cells causes alterations of copy number at the sites of CNVs, which are also categorized as CNAs. Copy number alterations at specific chromosomal regions showing association with the development and progression of certain tumors often include oncogenes or tumor suppressor genes.[5-8]

Breast cancer is one of the malignancies associated with high morbidity in women in Japan as well as many other developed countries.[9] Although many therapeutic approaches to breast cancer have been developed and the prognosis has been greatly improved, personalized treatments and precision medicine based on the specific features of individual tumors are required to obtain better outcomes. For this purpose, several techniques, such as expression profiling of multiple genes, have been developed for predicting outcomes or the risk of recurrence.[10,11] However, these tests are used only in limited clinical situations due to their laborious procedures and high cost.

Previous studies reported that several CNVs are associated with breast cancer risk, including copy gains of *APOBEC3B* on chromosomal region 22q13.1, *GSTM1* on 1p13.3, and *RAD51C* on 17q22.[12,13] In breast cancer, gain and loss of chromosomal region were analyzed using conventional a CGH array and several hotspots of somatic alterations were reported.[14,15] However, somatic CNAs at the sites of CNV have not been adequately analyzed in breast cancer. In the present study, we examined CNAs in 20 cases with invasive carcinoma of the breast using array CGH specifically designed to detect changes at CNV sites and evaluated the clinicopathological significance of the CNAs. The findings obtained by CNV array analysis were confirmed by Q-PCR analysis of the representative DNA fragments using 72 cases of independent invasive breast cancer. These results would provide a basic information for considering possible prognostic markers of invasive breast cancer.

2 | MATERIALS AND METHODS

2.1 | Tissue samples

We investigated 20 fresh frozen and 72 FFPE primary tumors of invasive ductal carcinoma from breast cancer patients who had undergone surgery at the Department of Breast and Endocrine Surgery, Juntendo University (Tokyo, Japan) from 2006 to 2016. We excluded patients given systemic chemotherapy prior to surgery. Informed consent was obtained from all study participants prior to surgery, in accordance with the ethics board requirements of Juntendo University and the University of Tokyo. The clinical characteristics of the 20 and 72 cases are summarized in Table S1. The median age at diagnosis was 64 (range, 36-82) years for cases in the CNV array analysis and 61 (range, 33-89) years for cases in the Q-PCR analysis. Tissue samples were intraoperatively excised from the centers of surgical tumor specimens, and then immediately frozen and stored at −80°C. All samples were diagnosed and classified according to the WHO grading system and the General Rules for Clinical and Pathological Recording of Breast Cancer established by the Japanese Breast Cancer Society.[16] Tumor content in the 92 breast cancers was determined by histopathologic analysis carried out by author DM.

2.2 | DNA isolation

Genomic DNA from breast cancer tissue and PBL was extracted from tumor samples using an AllPrep DNA/RNA/Protein Mini Kit (Qiagen). Normal DNA was obtained from peripheral blood cells of the corresponding patient using a QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's protocol. Genomic DNA from FFPE specimens of breast cancer was undertaken using a QIAamp DNA FFPE Tissue Kit (Qiagen).

2.3 | Array-based comparative genomic hybridization

Genomic DNA extracted from tissue samples or peripheral blood was analyzed using a SurePrint G3 Human CNV Microarray Kit, 2 × 400 K (Agilent Technologies) in accordance with the suggestion by the CNV Laboratory, DNA Chip Research Institute (Yokohama, Japan).

2.4 | Immunohistochemistry

The FFPE specimens from the 92 patients were examined. Tissue sections were deparaffinized and hydrated through graded alcohols and xylene. Antigen retrieval was carried out with citrate buffer at pH 6.0 in an autoclave at 121°C for 10 minutes. Endogenous peroxidase was blocked with 3% hydrogen peroxide for 10 minutes.

After rinsing and blocking with 5% normal donkey serum, the sections were incubated overnight at 4°C with primary Ab. The sections were washed and incubated for 2 hours at 4°C with the Dako EnVision + Dual Link System-HRP (Dako). Diaminobenzidine (Dako EnVision kit/HRP [DAB]) was used for detection of protein. The sections were finally counterstained with hematoxylin. On IHC, ER status and PgR status were assessed semiquantitatively and reported as positive when more than 1% of the nuclei of cancer cells showed staining. Positive HER2 status was determined if strong staining of the complete membrane in more than 10% of tumor cells was observed. Details of Abs are as follows: ER, rabbit monoclonal, clone SP1 (Ventana); PgR, rabbit monoclonal, clone 1E2 (Ventana); and HER2, rabbit monoclonal, clone 4B5 (Ventana). For Ki-67 staining, mAb (MIB-1; Dako) (1:400) was used and the cells positive for nuclear Ki-67 were counted in at least 500 cancer cells in one hotspot on each sample.

2.5 | Pathological grading

Histopathological examination was carried out using H&E staining of each tumor tissue by author DM. Nuclear grade of cancer cells was evaluated on the basis of a combination of nuclear atypia and mitotic counts. According to the intensity of atypia, NG was classified into three categories, NG1, 2, and 3. This grading system has been shown to correlate with the outcomes of Japanese breast cancer patients.[17] Histological grade was evaluated using a combination of the degree of architectural atypia, nuclear atypia, and the number of mitotic figures. This grading system has three categories, HG1, 2, and 3.

2.6 | Array CGH analysis

The array CGH data were analyzed with R statistical computing software (version 3.0.2; http://www.r-project.org), and the hidden Markov model was applied, with the segmental maximum posteriori approach.[18] We started by excluding the chromosome Y data and calculated the copy number ratio of cancer DNA to normal DNA. Copy number ratios were then categorized into six states with initial mean ratios of 0.50, 0.75, 1.00, 1.25, 1.50, and 2.00 with a standard deviation of 0.1. Two states, those with copy number ratios of 0.5 and 0.75, were defined as copy number loss, while those of 1.25, 1.50, and 2.00 states were defined as copy number gain. Sample no. 10 showed extremely rare copy number loss, and the estimates of two copy number states (initial values of 0.75 and 1.00) approached 1.00 (0.9783 and 0.9998, respectively). We defined these two states as representing a normal copy number for this sample. Correlations with clinicopathological features were calculated with Fisher's exact test or the Wilcoxon rank sum test. Gene ontology analysis was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID; http://david.abcc.ncifcrf.gov/).

2.7 | Quantitative real-time PCR

Quantitative PCR was undertaken using the CFX Connect Real-Time PCR Detection System (Bio-Rad). All reactions were set up using PowerUp SYBR Green Master Mix (Applied Biosystems) for total volumes of 10 μL which contained 50 ng of genomic DNA and 300 nmol/L of each specific forward and reverse primer. Thermal cycling was carried out as follows: 120 seconds at 95°C for initial denaturation, followed by 40 cycles with 15 seconds at 95°C, and 60 seconds at 60°C. *GAPDH* was used as the reference gene for normalizing gene copy.

2.8 | Next generation sequencing

The yield and quality of genomic DNA were determined using Picogreen (Invitrogen) and further visually inspected by agarose gel electrophoresis. Targeted sequencing of genes of interest was undertaken after capture with a custom SureSelect capture reagent designed using the SureDesign tool (Agilent Technologies). Target-enriched libraries were sequenced on the Illumina HiSeq-2500 sequencing platform as described previously.[19]

2.9 | Statistical analysis

All statistical analyses were carried out with programming language R. Differences with a *P* value of less than .05 were considered to be statistically significant. In the GO analysis, Benjamini-Hochberg FDRs less than 30% were considered to be statistically significant.

3 | RESULTS

3.1 | Somatic CNAs detected in 20 breast cancers

To examine genome-wide CNAs in 20 breast cancers we used the CGH-array specifically designed to detect changes at the sites of known CNV covering the human genome with more than 412 000 probes. As CNVs are highly polymorphic between individuals, we analyzed CNAs in DNA from cancer and PBL of the same patients in each case. The average ratio of tumor content in 20 breast cancers was 80.2% with a maximum rate of 91.5% and minimum rate of 70.7%, which was used to adjust the copy number in each tumor when necessary. Somatic CNAs were detected in 39.9% of the probes examined (Figure S1). The most frequently altered regions were chromosomal fragments 1q21-22 (90%) and 8q21-24 (85%). In addition, gains at 1q44 (85%) and 3q11 (85%), as well as losses at 16q22-24 (80%) and 17p13 (75%), were frequently detected, as previously reported by investigators using conventional CGH analyses. When we focused on three CNVs, *APOBEC3B*, *GSTM1*, and *RAD51C*, whose copy gain in germline DNA was shown to be associated with increased risk of breast cancer,[12,13] loss rather

than a gain of copy number was observed in 55%, 35%, and 40% of tumors, respectively.

To validate the results of CNV array analysis, we examined DNA sequences of the specific genes located within the regions showing gain/loss of chromosomal fragments by next generation sequencing, and then copy number was estimated by comparing the numbers of sequencing read with those from PBL DNA. These analyses showed an increased amount of DNA at the *ASH1L* gene on 1q22 in 85% and at the *MYC* on 8q24 in 75% of tumors, whereas there was a decreased amount of DNA at the *CDH1* on 16q22 in 70% and at the *TP53* on 17p13 in 80% of tumors. These results are comparable to those by the present CNV array analysis in the incidence of copy gain of 90% at 1q21-22 and 85% at 8q21-24 and the incidence of copy loss of 80% at 16q22-24 and 75% at 17p13 (Table S2), suggesting that CNV array analysis is sufficiently quantitative for further studies. Then, to validate the results in a larger number of breast cancer specimens, we examined CNAs of the four DNA fragments described above using 70 additional breast cancer DNA samples from FFPE specimens. As shown in Table S3, copy gains of 1q22 and 8p24 were detected in 90% and 69%, respectively, whereas copy losses of 16q22 and 17p13 were detected in 79% and 76%, respectively, providing highly comparable results with those in 20 breast cancers. These results strongly suggest that CNAs of chromosomal fragments detected in high prevalence by CNV array analysis are common features observed in breast cancer.

By CNV array analysis, 13 of 20 tumors showed a high prevalence of CNAs, exceeding 35%, wherein the prevalence of CNAs is defined by the number of probes showing CNAs out of the total number of probes examined. Tumor no. 3 with a triple negative subtype and no. 6 with HER2-positive subtype showed the highest prevalence of CNAs, 60.3% and 60.2%, respectively. In contrast, only two tumors (no. 10 and no. 20) had a CNA prevalence of less than 15% (Figure 1); both tumors showed an ER-positive, PgR-positive, and HER2-negative state with the Ki-67 labeling index being less than 20%, suggesting a relatively low-grade malignancy. Representative

histopathological features of tumors with the highest and the lowest rate of aberrant CNAs are shown in Figure 2. Quantitative PCR analysis of 18 DNA fragments in 72 additional breast cancers showed that the incidence of CNAs varies from 10% to 56%, and that 44 of 72 (61%) tumors showed an incidence higher than 35% (Table S4). The tumor with the highest CNA incidence (56%) had a HER2-positive subtype, whereas the tumor with lowest CNA incidence (10%) was ER-positive, PgR-positive, and HER2-negative with the Ki-67 labeling index of less than 20%, again supporting the findings of the CNV array analysis of 20 breast cancers.

3.2 | Pathological significance of CNAs in breast cancer

To evaluate possible correlations between CNAs and clinicopathological features of breast cancer, we categorized the 20 tumors by 11 clinicopathological parameters. These include patient age at diagnosis of breast cancer, pre- or postmenopausal status, pT, lymph node metastasis, nuclear grade, histological grade, lymphovascular invasion, status of ER, PgR, and HER2, and Ki-67 labeling index (Table 1). First, we examined correlations of clinicopathological features with CNA fragments. Then we focused on genes located within the DNA fragments showing CNAs and examined the correlations of each of these genes with clinicopathological features. Copy number alterations of 161 genes showed significant correlations ($P < .001$) with higher NG of tumors (NG1 or NG2 in 3 tumors vs NG3 in 17 tumors) (Table 1). Copy number alterations of 52, 35, and 35 genes also showed highly significant correlations ($P < .001$) with ER negativity (positive in 14 tumors vs negative in 6 tumors), PgR-negativity (positive in 11 tumors vs negative in 9 tumors), and higher Ki-67 labeling indexes (less than 20% in 12 tumors vs 20% or higher in 8 tumors), respectively. However, fewer than 10 genes showed highly significant correlations with age, pT, lymph node metastasis, histological grade, lymphovascular invasion, or HER2 positivity.
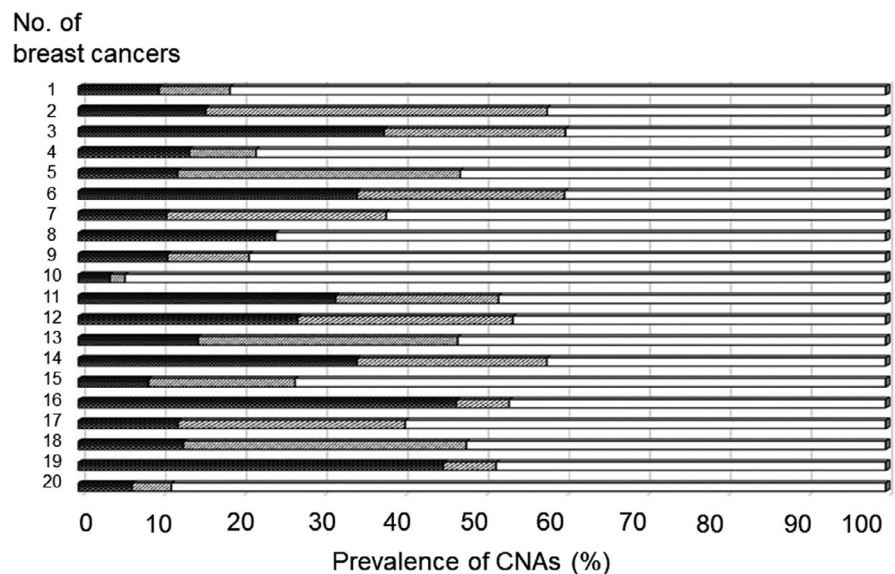
FIGURE 1 Schematic representation of a prevalence of copy number alterations (CNAs) in each breast cancer tumor detected by copy number variation array analysis in 20 breast cancers. Prevalence of CNA varies from 6.6% (no. 10) to 60.3% (no. 3). Perpendicular numbers indicate IDs of breast cancer patients. Black, hatched, and white bars indicate the prevalence of copy gain, copy loss, and no alteration, respectively
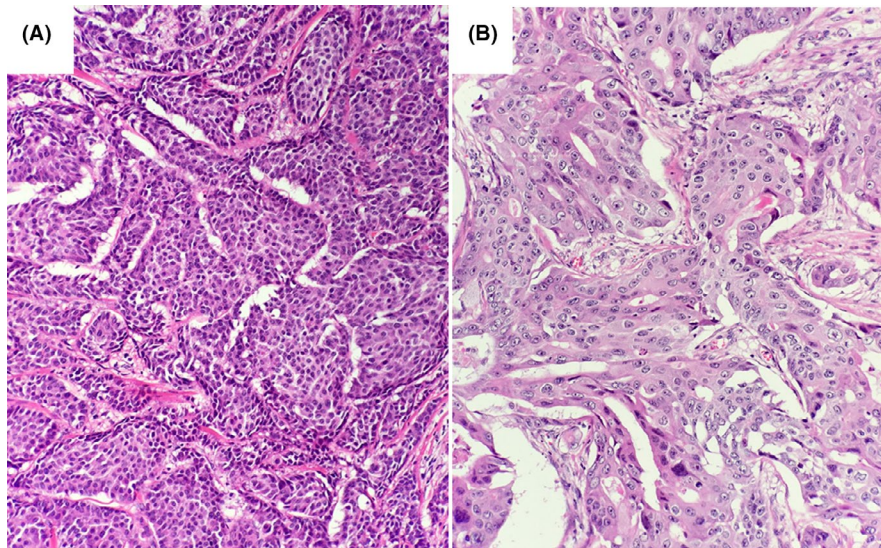
FIGURE 2 Representative histopathological features of breast cancer with lower and higher copy number alterations (CNAs). A, Breast tumor no. 10 is a low-grade carcinoma (nuclear grade 1, histological grade II) with CNAs of 6.6% showing an estrogen receptor-positive, progesterone receptor-positive, and human epidermal growth factor receptor 2-negative state (Ki-67 labeling index <20%). B, Breast tumor no. 3 is a high-grade carcinoma (nuclear grade 3, histological grade III) with CNAs of 60.3% showing a triple negative subtype with high proliferative activity (Ki-67 labeling index >20%)

TABLE 1 Number of genes within the copy number alteration fragments showing significant correlations with clinicopathological features in breast cancer patients

Clinicopathological feature		No. of cases in each group	No. of genes		
			$P < .05$	$P < .01$	$P < .001$
Age (y)	<50 vs ≥50	7:13	1463	244	3
Menopause	Pre vs post	7:13	592	417	17
pT[a]	I vs II	16:4	287	117	0
Lymph node metastasis	0 vs ≥1	9:11	882	212	6
Nuclear grade	1 vs 2 or 3	18:2	38	0	0
	1 or 2 vs 3	3:17	1726	512	161
Histological grade	I vs II or III	17:3	513	55	1
	I or II vs III	6:14	1233	262	20
Lymphovascular invasion	+ vs −	7:13	630	78	0
Estrogen receptor	+ vs −	14:6	1829	527	52
Progesterone receptor	+ vs −	11:9	2005	590	35
HER2	0, 1+, 2+ vs 3+	17:3	1677	327	8
Ki-67 labeling index	<20% vs ≥20%	12:8	1040	355	35

P values were calculated with Fisher's exact test.

HER2, human epidermal growth factor receptor 2.

[a]Defined by the TNM pathological classification.

We then undertook GO analyses of the genes showing highly significant correlations with each clinicopathological feature ($P < .01$). The GO terms that are annotated to the list of genes showing statistically significant associations with each clinicopathological feature in Table 1 ($P < .01$) were examined (Tables S5 and S6). A GO term, cell activation, was annotated with statistical significance ($P < .001$ and FDR < 30%) in six genes, *GP6C*, *LYN*, *BAX*, *PRKDC*, *TPD52*, and *KIR3DL1*, showing copy number gain in patients younger than 50 years. Similarly, three GO terms, transcription, DNA-dependent regulation of transcription, and regulation of RNA metabolic process, are annotated to 14-15 genes showing copy number gain in tumors

with HER2 positivity, whereas four GO terms, transcription, regulation of transcription, DNA-dependent regulation of transcription, and regulation of RNA metabolic process, are annotated in 48-64 genes showing copy number loss in postmenopausal tumors.

3.3 | Pathological features of tumors with high CNA prevalence

We next analyzed the clinicopathological features of tumors showing a high prevalence of CNAs exceeding 35%. High CNA prevalence

was observed mainly in tumors with higher NG, ER negativity, and a Ki-67 labeling index exceeding 20% (Table 2), as reported in previous studies using conventional array-CGH analyses. However, menopause, pT stage, lymph node metastasis, HG, lymphovascular invasion, as well as HER2 status and the *TP53* mutation, did not show significant correlation with aberrant CNVs. Quantitative PCR analysis of 18 DNA fragments in 72 additional breast cancer tumors confirmed most of the results above, except for the *TP53* mutation, because the incidence of the *TP53* mutation was significantly higher in CNA-positive tumors than in CNA-negative tumors ($P = .006$) (Table S7).

To extract the clinicopathological features that are specifically found in the present study using a CNV-oriented CGH array, we next categorized the CNAs on the basis of the lengths of fragments showing CNAs. We set five length ranges of 1 kb-10 kb, 10 kb-100 kb, 100 kb-1 Mb, 1 Mb-10 Mb, and more than 10 Mb and the numbers of CNA fragments showing gain, loss, or both detected in each category for each tumor were independently counted. All the combinations showing statistical significance ($P < .05$) are summarized in Table 3. Higher NG, ER negativity, and a higher Ki-67 labeling index correlated significantly with higher CNA prevalence in fragments of more than 10 Mb. In addition, tumors with higher NG showed a significantly higher prevalence of CNA in fragments of 10 kb-100 kb and 100 kb-1 Mb, whereas menopausal status and lymphovascular invasion showed significant correlation with higher CNA prevalence in fragments of 100 kb-1 Mb and 1 kb-10 Mb, respectively.

However, when we focused on lengths of less than 10 kb, tumors with ER positivity were identified to correlate with higher prevalence of gain and loss of CNA fragments, which is a unique finding only obtained by this CNV-oriented CGH array. This is in sharp contrast with the results that ER negativity was associated with a significantly high prevalence of loss or gain + loss of fragments with lengths of more than 10 Mb. The median number of fragments with less than 10 kb detected in 20 tumors was 124. Thus, to further investigate the significance of shorter CNA fragments of less than 10 kb in breast tumorigenesis, we examined the correlation between the clinicopathological parameters and the number of CNA fragments of less than 10 kb in 20 tumors. As shown in Table 4, the tumors with high numbers of CNA fragments of less than 10 kb showed significant correlations with lower NG (NG1 or 2, $P = .030$) and ER positivity ($P = .0043$). Quantitative PCR analysis of five fragments exhibiting short CNAs of less than 10 kb in the CNV array analysis revealed that the tumors with lower NG or ER positivity preferentially gave higher incidence of CNAs in 72 additional breast cancers (Table S8), again providing supporting evidence to the findings by CNV array analysis of 20 breast cancers.

TABLE 2 Correlations between the rate of aberrant copy number alteration (CNA) probes and clinical features of breast cancer patients

| Clinical feature | | Rate of aberrant CNA probes | | P value |
		<35%	≥35%	
Age (y)	<50	6	7	NS
	≥50	1	6	
Menopause	Pre	1	6	NS
	Post	6	7	
pT stage	T1	2	2	NS
	T2	5	11	
Lymph node metastasis	0	4	7	NS
	≥1	3	6	
Nuclear grade	1	2	0	<.05
	2	5	10	
	3	0	3	
Histological grade	I	1	2	NS
	II	5	6	
	III	1	5	
Lymphovascular invasion	–	5	8	NS
	+	2	5	
Estrogen receptor	+	7	7	<.05
	–	0	6	
Progesterone receptor	+	6	5	NS
	–	1	8	
HER2	–	7	10	NS
	+	0	3	
Ki-67 labeling index	<20%	7	5	<.05
	≥20%	0	8	
TP53 mutation status	WT	1	5	NS
	Mutation	6	8	

P values calculated with χ^2 test.
HER2, human epidermal growth factor receptor 2; NS, not significant.

4 | DISCUSSION

The aims of this study are: (i) providing an overview of somatic CNAs at the sites of CNVs in breast cancer; (ii) understanding the clinicopathological features of CNAs on the basis of relevant genes; and (iii) identifying possibly novel cascades of breast carcinogenesis triggered by CNAs. For this purpose, we examined CNAs of the DNA in 20 invasive breast cancers in comparison with DNA from PBL from the same patients using CNV-oriented CGH array in more than 412 000 sites, because CNVs are highly polymorphic between individuals. Copy numbers were also adjusted by tumor content in each case. In addition, as CNV regions are known to be vulnerable to systematic signal bias or noise due to their complex cross-hybridization and strong signal intensity, we validated the results of CNV array by two independent approaches: (i) comparison of the number of sequencing reads by the next generation sequencing analysis of specific genes; and (ii) Q-PCR of the relevant DNA fragments. Coincident results between the CNV array and next generation sequencing is shown in Table S2. Quantitative PCR assay was then carried out as another

Clinical feature		CNA length	CN status	Median no. of CNA regions	P value
ER	+	1-10 k	Gain + loss	76.5	.034
	−			21.5	
NG	1	10-100 k	Loss	4.5	.026
	2 or 3			30	
NG	1	10-100 k	Gain + loss	14	.021
	2 or 3			76.5	
NG	1	100 k-1 M	Gain	15.5	.042
	2 or 3			84.5	
NG	1	100 k-1 M	Loss	8	.021
	2 or 3			76.5	
Menopausal status	Pre	100 k-1 M	Loss	43	.037
	Post			179	
NG	1	100 k-1 M	Gain + loss	23.5	.042
	2 or 3			169	
Lymphovascular invasion	−	1-10 M	Loss	21	.045
	+			118	
NG	1 or 2	>10 M	Gain	7	.034
	3			19	
ER	+	>10 M	Loss	5.5	.016
	−			19.5	
Ki-67 labeling index	<20	>10 M	Loss	5.5	.024
	≥20			19.5	
ER	+	>10 M	Gain + loss	14.5	.011
	−			36.5	
NG	1 or 2	>10 M	Gain + loss	15	.032
	3			39	
Ki-67 labeling index	<20	>10 M	Gain + loss	14.5	.041
	≥20			34	

P values calculated with Wilcoxon rank-sum test.

ER, estrogen receptor; NG, nuclear grade.

approach to validate three DNA fragments on chromosomal loci 17q12, 1p36.3, and 3q26.3, detected as short CNAs of less than 10 kb by this CNV array analysis. As shown in Figure S2, marked copy gain of a DNA fragment on chromosome 17q12 in tumor no. 12 detected by CNV array was confirmed by Q-PCR analysis as a significantly increased DNA signal. Furthermore, six samples of tumor DNA exhibiting a copy gain of 1p36.3 by CNV array gave increased signals and three samples showing copy loss of 3q26.3 by CNV array gave decreased signals in Q-PCR analysis. Additional supportive evidence was obtained that two of two tumors showing copy gain of the *HER2* gene by CNV array gave 3+ signals in the HER2 assay, as shown in Table S9A. Thus, we concluded that the CNV array used in this study provided highly quantitative results that were comparable to other quantitative approaches. In this connection, Q-PCR analysis of the *HER2* gene also produced quantitative results that were well correlated with protein expression of HER2 by IHC (Table S9A, Figure S3).

Initial screening of CNVs in 20 breast cancers revealed that somatic CNAs were present in 39.9% of the CNV sites examined. As shown in Figure S1, most of the CNAs were detected on the continuous DNA fragments showing gains or losses of chromosomal regions, suggesting that most findings obtained from the analysis of these CNAs are based on the gross chromosomal abnormality and are essentially the same as those obtained by conventional CGH array analysis. In fact, we identified frequent gains of 1q21-22, 8q21-24, 1q44, and 3q11 and losses of 16q22-24 and 17p13, all of which had already been reported previously by conventional CGH analyses.[14,15] We confirmed that these chromosomal loci also provided hotspots of CNAs by Q-PCR analyses of 72 additional breast cancers (Table S3). However, these somatic alterations of copy number could be independent of germline susceptibility to breast cancer, because polymorphic copy gain of *APOBEC3B*, *GSTM1*, and *RAD51C* has been shown to be associated with an increased risk of breast cancer, whereas not gain but loss of these

TABLE 4 Clinicopathological features of breast cancer and number of copy number alteration (CNA) fragments shorter than 10 kb

| Clinical feature | | No. of CNA fragments <10 kb | | |
		<124	≥124	P value
Age (y)	<50	7	6	NS
	≥50	3	4	
Menopause	Pre	3	4	NS
	Post	7	6	
pT stage	T1	1	3	NS
	T2	9	7	
Lymph node metastasis	0	4	7	NS
	≥1	6	3	
Nuclear grade	1 or 2	7	10	<0.05
	3	3	0	
Histological grade	I	1	2	NS
	II	6	5	
	III	3	3	
Lymphovascular invasion	−	5	8	NS
	+	5	2	
Estrogen receptor	+	5	9	<0.05
	−	5	1	
Progesterone receptor	+	6	7	NS
	−	4	3	
HER2	−	8	9	NS
	+	2	1	
Ki67 labeling index	<20%	5	7	NS
	≥20%	5	3	

P values calculated with χ^2 test.
HER2, human epidermal growth factor receptor 2; NS, not significant.

loci was observed at a high incidence in the 20 examined breast cancers.

We then picked up aberrant CNA fragments and corresponding genes within the fragments showing statistical significance with several clinicopathological features of breast cancer. It is noteworthy that the largest number of genes (161 genes) correlated with NG1 or NG2 vs NG3 ($P < .01$ or $P < .001$), whereas no gene showed significant correlation between NG1 vs NG2 or NG3, suggesting that NG3 is a distinctive phenotype.

Next, GO analysis of these genes was carried out to elucidate the biological significance of aberrant CNAs in each phenotype. As shown in Tables S5 and S6, copy number gain of six genes, *GP6C*, *LYN*, *BAX*, *PRKDC*, *TPD52*, and *KIR3DL1*, showed a highly significant correlation with patients younger than 50 years. Accelerated glycolysis with enhanced activity of glucose-6-phosphatase and activation of LYN tyrosine kinase, TPD52 oncoprotein, and an immune inhibitory natural killer cell receptor, KIR3DL1, were reported in breast cancer, although their specific roles in early onset breast cancer are not reported.[20-23] Imbalanced expression of BAX, a member of the BCL2 proapoptotic protein family, is also shown to be involved in breast cancer.[24] *PRKDC* encodes the catalytic subunit of the DNA-dependent protein kinase and is involved in DNA DSB repair and recombination in cooperation with Ku70/Ku80 proteins.[25] As *BRCA1/2*, causal genes of hereditary breast and ovarian cancer, are also involved in the repair of DNA DSB, imbalanced expression of *PRKDC* might play a role in early onset of breast carcinogenesis.

The HER2-positive and postmenopausal breast cancer showed a significant correlation with gain of 14-15 genes and loss of 48-64 genes, respectively. As shown in Table S5, most of them were transcription factors and annotated to "transcription," "regulation of transcription," "DNA-dependent regulation of transcription," and "regulation of RNA metabolic process." As numerous genes, especially a large family of zinc-finger proteins, were identified probably due to their clustered location on the chromosomal region 19q13, identification of the pathobiological significance of each gene would be difficult. However, it is interesting that gain of the *SOX5* gene on chromosomal regions 12p12.1 was also picked up in HER2-positive tumor because SOX5 is known to be involved in cell proliferation, invasion, and epithelial-mesenchymal transition of breast cancer cells, although possible involvement in HER2-positive breast cancer has not yet been reported.[26] Similarly, it would be interesting that loss of tumor suppressor-like genes *DEDD2* and *RBL1* were identified in postmenopausal breast cancer.[27,28] Although these findings were obtained by CNV array analysis of only three of 20 breast cancers showing HER2-positive phenotype, supporting evidence was obtained by Q-PCR analysis of 69 additional breast cancers. As shown in Table S10, CNAs of three genes listed in Table S6 were preferentially observed in 11 tumors with HER2 positivity in comparison with 58 tumors with HER2 negativity ($P = .008$). These results suggest that molecular pathways related to these four GO terms, including transcriptional regulations and RNA metabolism, are involved in carcinogenesis of HER2-positive breast cancer.

The prevalence of CNAs in these tumors by CNV analysis varied from 6% to 60% and tumors and were categorized into two groups; 13 tumors showed a high prevalence (35% or higher) of CNAs, whereas seven tumors had a low prevalence (less than 35%) of CNAs (Table 2). Comparison of clinicopathological features between these two groups of tumors indicated that higher NG, ER negativity, and an elevated Ki-67 labeling index correlated significantly with a high prevalence of CNAs, again coincident with the results of previous studies using conventional CGH array (Table 2).[14,15] Miyaguchi et al reported correlations between CNA length and aberrant gene expressions found in the CNAs in hepatocellular carcinoma and colorectal cancer.[29] Our results would also support that long CNA is associated with more malignant features of breast cancer. Furthermore, tumors with high prevalence of CNAs with gross chromosomal abnormality appeared to be caused by CIN, which is known to accelerate the proliferation of aneuploid tumor cells.[30] Several studies have reported a high incidence of CIN in solid tumors,[2,31,32] including breast cancers.[15,33-39] Endesfelder et al reported the correlation

between CIN and ER negativity in young breast cancer patients and hypothesized that CIN might be characteristic of younger-onset ER-negative breast cancer.[40] Our data also supports the importance of CIN in breast carcinogenesis. In this connection, it is well known that the *TP53* mutation causes chromosomal instability of tumors and we found significant correlation between them in a Q-PCR analysis of 69 tumors, although analysis of 20 tumors did not show a significant difference (Tables 2 and S7).[41]

In the present study, using the CGH array that exclusively targets established CNV sites, we were able to detect short spike-like CNAs of less than 10 kb, which could not be detected by conventional CGH array analysis. Our preliminary analysis suggests that these short CNAs of less than 10 kb are detected not only in breast cancer but also several other cancers, including cholangiocarcinoma, oral squamous cell carcinoma, and bladder carcinoma, when examined by the same CNV array platform. These results suggest that the short CNVs would play a role in breast and other carcinogenesis. Interestingly, we found that these short CNAs with gain or loss of less than 10-kb fragments were observed at a significantly higher rate in breast cancer with ER positivity and with nuclear grade 1 or 2 in the CNV array analysis of 20 breast cancers as well as in Q-PCR analysis of 72 breast cancers (Tables 4 and S8).

While large continuous CNAs are suggested to link with CIN mechanistically, the molecular basis as well as the function of the short CNAs has not yet been elucidated. The spike-like short CNAs would be generated by specific structural instabilities of DNA sequences functioning as *cis*-acting factors. Alternatively, some enzymes that serve to maintain the accurate replication or repair of specific CNVs might be dysregulated in cancer cells, functioning instead as *trans*-acting factors to generate short CNAs. It is well known that microsatellite sequences consisting of one to four nucleotide repeats trigger replication errors of DNA when the mismatch repair function is abrogated in cancer cells.[1] *Alu* repetitive sequences are also shown to be often present at the boundary of deleted or inserted DNA fragments in several genetic disorders and cancer cells.[42-44] In contrast, neither the boundary sequences of the short CNAs of less than 10 kb, nor *trans*-acting factors or some enzymatic dysfunction leading to CNAs in cancer cells have been identified yet. The present study would be the first report to suggest that spike-like CNAs would trigger ER-positive breast carcinogenesis through activating or inactivating a limited number of oncogenes or tumor suppressor genes on the DNA fragments, respectively, through allelic imbalance and resultant expression imbalance of genes that promote ER-related pathways. Further investigation would be required to elucidate the novel cascade as well as the underlying mechanisms of tumorigenesis associated with short CNAs in breast cancer.

ORCID

Daisuke Matsubara https://orcid.org/0000-0002-6233-6840
Tatsuhiro Shibata https://orcid.org/0000-0002-0477-210X
Yoshinori Murakami https://orcid.org/0000-0002-2826-4396

REFERENCES

1. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998;396:643-649.
2. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet*. 2020;21:44-62.
3. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85-97.
4. Wellcome Trust Case Control C, Craddock N, Hurles ME, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*. 2010;464:713-720.
5. Zhao X, Li C, Paez JG, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Can Res*. 2004;64:3060-3071.
6. Yoshihara K, Tajima A, Adachi S, et al. Germline copy number variations in BRCA1-associated ovarian cancer patients. *Genes Chromosomes Cancer*. 2011;50:167-177.
7. Singh RR, Patel KP, Routbort MJ, et al. Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. *Br J Cancer*. 2014;111:2014-2023.
8. Tsai P-C, Huang S-W, Tsai H-L, et al. The association between DNA copy number aberrations at chromosome 5q22 and gastric cancer. *PLoS One*. 2014;9:e106624.
9. Pilleron S, Sarfati D, Janssen-Heijnen M, et al. Global cancer incidence in older adults, 2012 and 2035: A population-based study. *Int J Cancer*. 2019;144:49-58.
10. Brufsky AM. Predictive and prognostic value of the 21-gene recurrence score in hormone receptor-positive, node-positive breast cancer. *Am J Clin Oncol*. 2014;37:404-410.
11. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375:717-729.
12. Kumaran M, Cass CE, Graham K, et al. Germline copy number variations are associated with breast cancer risk and prognosis. *Sci Rep*. 2017;7:14621.
13. Pelttari LM, Shimelis H, Toiminen H, et al. Gene-panel testi ng of breast and ovarian cancer patients identifies a recurrent RAD51C duplication. *Clin Genet*. 2018;93:595-602.
14. Brewster AM, Thompson P, Sahin AA, et al. Copy number imbalances between screen- and symptom-detected breast cancers and impact on disease-free survival. *Cancer Prev Res (Phila)*. 2011;4:1609-1616.
15. Endesfelder D, Burrell R, Kanu N, et al. Chromosomal instability selects gene copy number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res*. 2014;74:4853-4863.
16. The Japanese Breast Cancer Society. General rules for clinical and pathological recording of breast cancer. 2012.
17. Tsuda H, Akiyama F, Kurosumi M, Sakamoto G, Watanabe T. Establishment of histological criteria for High0-risk node-negative

18. Andersson R, Bruder CEG, Piotrowski A, et al. A segmental maximum a posteriori approach to genome-wide copy number profiling. *Bioinformatics*. 2008;24:751-758.

19. Jusakul A, Cutcutache I, Yong CH, et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov*. 2017;7:1116-1135.

20. Martinez-Outschoorn UE, Peiris-Pagés M, Pestell RG, Sotgia F, Lisanti MP. Cancer metabolism: a therapeutic perspective. *Nat Rev Clin Oncol*. 2017;14:11-31.

21. Tornillo G, Knowlson C, Kendrick H, et al. Dual mechanisms of LYN kinase dysregulation drive aggressive behavior in breast cancer cells. *Cell Rep*. 2018;25:3674-3692.

22. Zhang Z, Wang J, Gao R, et al. Downregulation of MicroRNA-449 promotes migration and invasion of breast cancer cells by targeting tumor protein D52 (TPD52). *Oncol Res*. 2017;25:753-761.

23. Varker KA, Terrell CE, Welt M, et al. Impaired natural killer cell lysis in breast cancer patients with high levels of psychological stress is associated with altered expression of killer immunoglobin-like receptors. *J Surg Res*. 2007;139:36-44.

24. Fister S, Gunthert AR, Aicher B, , Paulini KW, Emons G, Grundker C. GnRH-II antagonists induce apoptosis in human endometrial, ovarian, and breast cancer cells via activation of stress-induced MAPKs p38 and JNK and proapoptotic protein Bax. *Cancer Res*. 2009;69:6473-6481.

25. Sun G, Yang LE, Dong C, Ma B, Shan M, Ma B. PRKDC regulates chemosensitivity and is a potential prognostic and predictive marker of response to adjuvant chemotherapy in breast cancer patients. *Oncol Rep*. 2017;37:3536-3542.

26. Pei XH, Lv XQ, Li HX. Sox5 induces epithelial to mesenchymal transition by transactivation of Twist1. *Biochem Biophys Res Commun*. 2014;446:322-327.

27. Muthu M, Somagoni J, Cheriyan VT, et al. Identification and testing of novel CARP-1 functional mimetic compounds as inhibitors of non-small cell lung and triple negative breast cancers. *J Biomed Nanotechnol*. 2015;11:1608-1627.

28. Simpson DS, Mason-Richie NA, Gettler CA, Wikenheiser-Brokamp KA. Retinoblastoma family proteins have distinct functions in pulmonary epithelial cells in vivo critical for suppressing cell growth and tumorigenesis. *Cancer Res*. 2009;69:8733-8741.

29. Miyaguch K, Fukuoka Y, Mizushima H, et al. Genome-wide integrative analysis revealed a correlation between lengths of copy number segments and corresponding gene expression profile. *Bioinformation*. 2011;7:280-284.

30. He Q, Au B, Kulkarni M, et al. Chromosomal instability-induced senescence potentiates cell non-autonomous tumourigenic effects. *Oncogenesis*. 2018;7:62.

31. Cucco F, Servadio A, Gatti V, et al. Mutant cohesin drives chromosomal instability in early colorectal adenomas. *Hum Mol Genet*. 2014;23:6773-6778.

32. Quimbaya M, Raspé E, Denecker G, et al. Deregulation of the replisome factor MCMBP prompts oncogenesis in colorectal carcinomas through chromosomal instability. *Neoplasia*. 2014;16:694-709.

33. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer*. 2006;6:96.

34. Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR. Genomic instability in breast cancer: pathogenesis and clinical implications. *Mol Oncol*. 2010;4:255-266.

35. Rummel S, Valente AL, Kane JL, Shriver CD, Ellsworth RE. Genomic (in)stability of the breast tumor microenvironment. *Mol Cancer Res*. 2012;10:1526-1531.

36. Heselmeyer-Haddad K, Berroa Garcia LY, Bradley A, et al. Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression. *Am J Pathol*. 2012;181:1807-1822.

37. de Vargas Wolfgramm E, Alves LNR, Stur E, et al. Analysis of genome instability in breast cancer. *Mol Biol Rep*. 2013;40:2139-2144.

38. Vollebergh MA, Klijn C, Schouten PC, et al. Lack of genomic heterogeneity at high-resolution aCGH between primary breast cancers and their paired lymph node metastases. *PLoS One*. 2014;9:e103177.

39. Przybytkowski E, Lenkiewicz E, Barrett MT, et al. Chromosome-breakage genomic instability and chromothripsis in breast cancer. *BMC Genom*. 2014;15:579.

40. Endesfelder D, McGranahan N, Birkbak NJ, et al. A breast cancer meta-analysis of two expression measures of chromosomal instability reveals a relationship with younger age at diagnosis and high risk histopathological variables. *Oncotarget*. 2011;23:6773-6778.

41. Gronroos E, López-García C. Tolerance of chromosomal instability in cancer: mechanisms and therapeutic opportunities. *Cancer Res*. 2018;78:6529-6535.

42. Perez-Cabornero L, Borras Flores E, Sanz MI, et al. Characterization of new founder Alu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype. *Cancer Prev Res (Phila)*. 2011;4:1546-1555.

43. Su P, Wang YE, Cooper DN, et al. Disclosing the hidden structure and underlying mutational mechanism of a novel type of duplication CNV responsible for hereditary multiple osteochondromas. *Hum Mutat*. 2015;36:758-763.

44. Nikiforov YE, Koshoffer A, Nikiforova M, Stringer J, Fagin JA. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene*. 1999;18(46):6330-6334.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.