



TITLE:

# Genetic and clinical landscape of breast cancers with germline BRCA1/2 variants

AUTHOR(S):

Inagaki-Kawata, Yukiko; Yoshida, Kenichi; Kawaguchi-Sakita, Nobuko; Kawashima, Masahiro; Nishimura, Tomomi; Senda, Noriko; Shiozawa, Yusuke; ... Miyano, Satoru; Ogawa, Seishi; Toi, Masakazu

---

CITATION:

Inagaki-Kawata, Yukiko ...[et al]. Genetic and clinical landscape of breast cancers with germline BRCA1/2 variants. *Communications Biology* 2020, 3: 578.

ISSUE DATE:

2020-10-16

URL:

<http://hdl.handle.net/2433/255639>

RIGHT:

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.








# COMMUNICATIONS BIOLOGY

## ARTICLE


<https://doi.org/10.1038/s42003-020-01301-9>

OPEN

## Genetic and clinical landscape of breast cancers with germline *BRCA1/2* variants

Yukiko Inagaki-Kawata<sup>1,2</sup>, Kenichi Yoshida<sup>1</sup> , Nobuko Kawaguchi-Sakita<sup>3</sup>, Masahiro Kawashima<sup>2</sup>, Tomomi Nishimura<sup>1,2</sup>, Noriko Senda<sup>2</sup>, Yusuke Shiozawa<sup>1</sup>, Yasuhide Takeuchi<sup>1,4,5</sup>, Yoshikage Inoue<sup>1</sup>, Aiko Sato-Otsubo<sup>1</sup>, Yoichi Fujii<sup>1</sup>, Yasuhito Nannya<sup>1</sup>, Eiji Suzuki<sup>2</sup>, Masahiro Takada<sup>2</sup>, Hiroko Tanaka<sup>6</sup> , Yuichi Shiraishi<sup>7</sup>, Kenichi Chiba<sup>7</sup>, Yuki Kataoka<sup>8</sup>, Masae Torii<sup>9</sup>, Hiroshi Yoshibayashi<sup>9</sup>, Kazuhiko Yamagami<sup>10</sup>, Ryuji Okamura<sup>11</sup>, Yoshio Moriguchi<sup>12</sup> , Hironori Kato<sup>13</sup>, Shigeru Tsuyuki<sup>14</sup>, Akira Yamauchi<sup>15</sup>, Hirofumi Suwa<sup>16</sup>, Takashi Inamoto<sup>17</sup>, Satoru Miyano<sup>6,7</sup>, Seishi Ogawa<sup>1,4,18</sup>  & Masakazu Toi<sup>2</sup> 

The genetic and clinical characteristics of breast tumors with germline variants, including their association with biallelic inactivation through loss-of-heterozygosity (LOH) and second somatic mutations, remain elusive. We analyzed germline variants of 11 breast cancer susceptibility genes for 1,995 Japanese breast cancer patients, and identified 101 (5.1%) pathogenic variants, including 62 *BRCA2* and 15 *BRCA1* mutations. Genetic analysis of 64 *BRCA1/2*-mutated tumors including TCGA dataset tumors, revealed an association of biallelic inactivation with more extensive deletions, copy neutral LOH, gain with LOH and younger onset. Strikingly, *TP53* and *RB1* mutations were frequently observed in *BRCA1*- (94%) and *BRCA2*- (9.7%) mutated tumors with biallelic inactivation. Inactivation of *TP53* and *RB1* together with *BRCA1* and *BRCA2*, respectively, involved LOH of chromosomes 17 and 13. Notably, *BRCA1/2* tumors without biallelic inactivation were indistinguishable from those without germline variants. Our study highlights the heterogeneity and unique clonal selection pattern in breast cancers with germline variants.

<sup>1</sup> Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan. <sup>2</sup> Department of Breast Surgery, Kyoto University, Kyoto, Japan. <sup>3</sup> Department of Clinical Oncology, Kyoto University Hospital, Kyoto, Japan. <sup>4</sup> Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan. <sup>5</sup> Department of Diagnostic Pathology, Kyoto University, Kyoto, Japan. <sup>6</sup> Laboratory of Sequence Analysis, Human Genome Centre, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. <sup>7</sup> Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. <sup>8</sup> Hospital Care Research Unit/Department of Respiratory Medicine, Hyogo Prefectural Amagasaki General Medical Center, Amagasaki, Japan. <sup>9</sup> Department of Breast Surgery, Japanese Red Cross Wakayama Medical Center, Wakayama, Japan. <sup>10</sup> Department of Breast Surgery, Shinko Hospital, Kobe, Japan. <sup>11</sup> Department of Breast Surgery, Yamatotakada Municipal Hospital, Yamatotakada, Japan. <sup>12</sup> Department of Breast Surgery, Kyoto City Hospital, Kyoto, Japan. <sup>13</sup> Department of Breast Surgery, Kobe City Medical Center General Hospital, Kobe, Japan. <sup>14</sup> Department of Breast Surgery, Osaka Red Cross Hospital, Osaka, Japan. <sup>15</sup> Department of Breast Surgery, Kitano Hospital, Osaka, Japan. <sup>16</sup> Department of Breast Surgery, Hyogo Prefectural Amagasaki General Medical Center, Amagasaki, Japan. <sup>17</sup> Tenri Health Care University, Tenri, Japan. <sup>18</sup> Department of Medicine, Centre for Haematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden. ✉email: [sogawa-tyk@umin.ac.jp](mailto:sogawa-tyk@umin.ac.jp); [toi@kuhp.kyoto-u.ac.jp](mailto:toi@kuhp.kyoto-u.ac.jp)

Germline predisposition plays a substantial role in breast cancer, the most prevalent cancer in women. Management and prevention of breast cancer would, therefore, benefit from better knowledge and understanding of the genetic cause behind such familial predisposition<sup>1</sup>. Previous studies reported that pathogenic germline mutations account for 10.7% of breast cancer cases in a Western cohort<sup>2</sup> and 9.2% of those in a Chinese cohort<sup>3</sup>. Breast cancer is also prevalent in Japan, affecting 116.3 per 100,000 women, where germline predisposition has been confirmed or suspected in as many as 5.7% of cases<sup>4</sup>. However, genetic studies of germline mutations that result in a predisposition to breast cancer are limited in the Japanese population<sup>4-6</sup>. In particular, the effects of pathogenic germline variants on somatic mutations and clinical/pathological phenotypes of accompanying breast cancers are poorly understood. Exception is the well-established mutational signatures associated with germline mutations in *BRCA1/2* and *PALB2*<sup>7,8</sup>, which are key genes in DNA repair by homologous recombination (HR) of DNA double strand breaks<sup>9</sup>. Although a previous study analyzed tumors with germline *BRCA1/2* mutations in terms of presence or absence of biallelic inactivation<sup>10</sup>, its genetic and clinical impact on breast cancers have not been fully elucidated.

In this study, therefore, we investigated pathogenic germline variants in 11 genes implicated in hereditary breast cancer, which were *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *CDH1*, *STK11*, *NF1*, *PALB2*, *ATM*, *CHEK2*, and *NBN*<sup>1,11-18</sup>, for 1995 unselected Japanese women with breast cancer, using targeted-capture sequencing of pooled DNA (Supplementary Fig. 1a). For those patients for whom tumor samples were available, the somatic alterations in the tumor were also interrogated in order to link the genetic features of the germline risk alleles to the associated tumor clinical presentations. In particular, we investigated the effects of biallelic inactivation of *BRCA1/2* genes on the somatic mutations, and copy number (CN) abnormalities (CNAs) and clinical features of the resulting breast cancers.

## Results

**Prevalence of germline mutations.** The characteristics of the study participants are shown in Table 1. Of the 1995 patients analyzed, 101 (5.1%) carried pathogenic germline variants, comprising 54 (53.5%) frameshift insertion or deletions (indels), 35 (34.7%) nonsense mutations, 7 (6.9%) missense mutations, 4 (4.0%) splice site mutations, and 1 (1.0%) synonymous mutation (Supplementary Fig. 1b). With respect to the 11 genes of interest, 77 of the total (3.9%) patients carried a variant in *BRCA2* ( $n = 62$ , 3.1%) or *BRCA1* gene ( $n = 15$ , 0.8%), followed by *PALB2* ( $n = 9$ , 0.5%), *TP53* ( $n = 4$ , 0.2%), *PTEN* ( $n = 4$ , 0.2%), *CHEK2* ( $n = 3$ , 0.2%), *ATM* ( $n = 3$ , 0.2%), and *NF1* ( $n = 1$ , 0.05%) (Fig. 1a, Supplementary Table 1). No pathogenic variants were identified in *CDH1*, *STK11*, and *NBN*. None of the patients carried two or more pathogenic variants. The genetic mutations were widely distributed along the entire coding region of the genes, except for missense mutations on the DNA binding domain of *TP53* (Supplementary Fig. 2). Some variants were detected in  $\geq 2$  patients, such as L63X in *BRCA1*<sup>19</sup> and c.5576\_5579del in *BRCA2*<sup>3,4</sup>, which have been previously suggested as founder mutations in the Asian population<sup>19,20</sup>.

**Characteristics of patients carrying germline variants.** The profiles of patients carrying germline variants in each gene are shown in Supplementary Table 2. Pathogenic variants were more frequently identified in patients with a family history of breast cancer ( $n = 41$ , 11.0%), compared with those without ( $n = 50$ , 3.4%) ( $P < 0.00001$ ). Of the analyzed genes, *BRCA2* was the most frequently mutated in both patients with and without a family

**Table 1 Characteristics of 1995 patients enrolled in this study.**

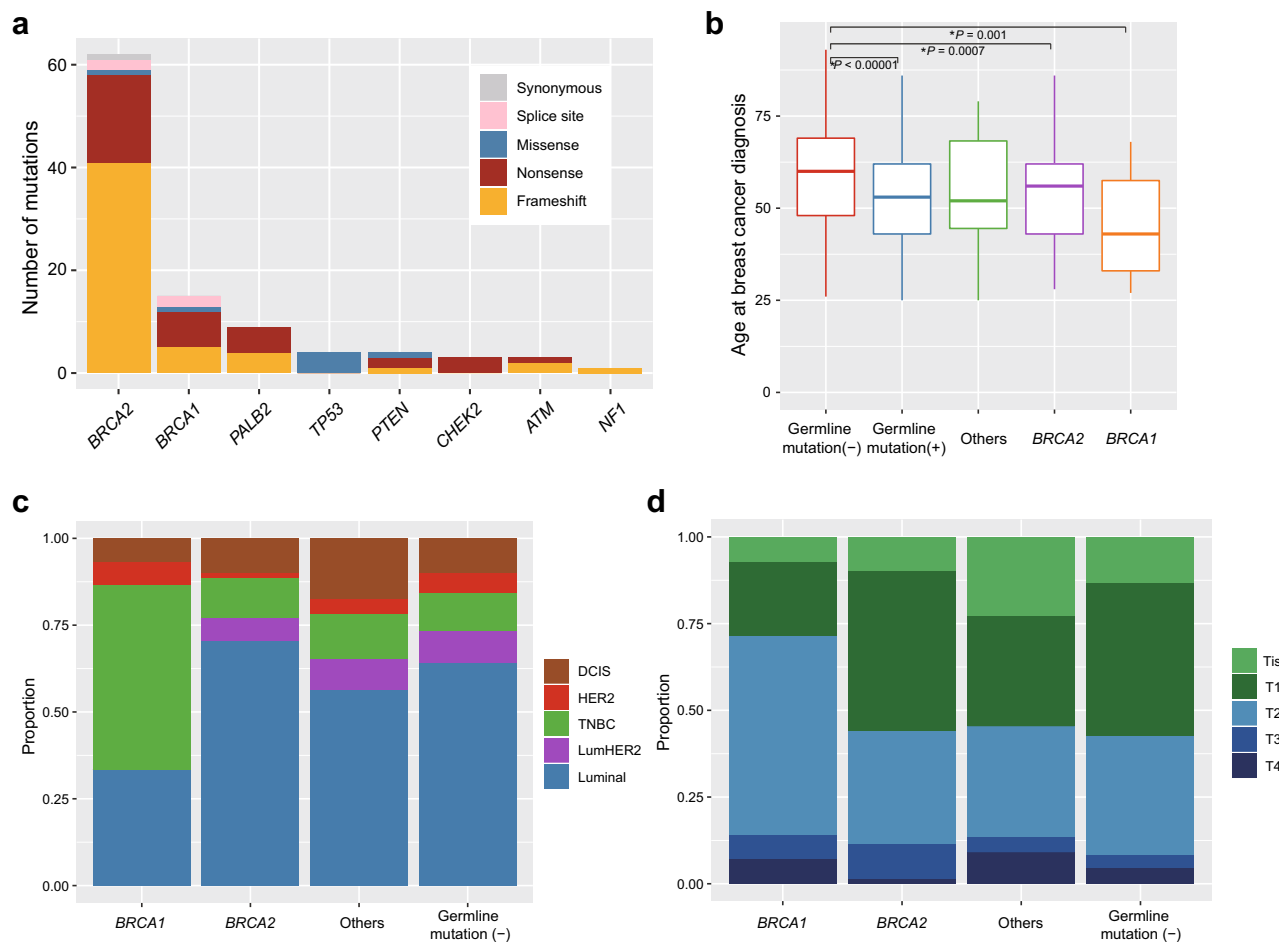
	Number of patients
<b>Age</b>	
-35	51
36-45	317
46-55	461
56-65	522
66	644
<b>Histology</b>	
IDC	1484
ILC	74
DCIS	212
Others	106
<b>Phenotype</b>	
ER+	1540
ER-	377
HER2+	310
HER2-	1382
Ki67 high (>14)	967
Ki67 low	721
<b>Histological grade</b>	
3	146
2	419
1	217
<b>Clinical stage</b>	
0	254
I	774
II	733
III	142
IV	53
Unknown	39
<b>Past history</b>	
Ovarian cancer	14
<b>Family history<sup>a</sup></b>	
Breast cancer	348
Ovarian cancer	33

Abbreviations: IDC Invasive ductal carcinoma, ILC invasive lobular carcinoma, DCIS ductal carcinoma in situ.  
<sup>a</sup>Family history refers to reported breast or ovarian cancer in first or second-degree relative.

history (Supplementary Table 3). A quarter of the patients with germline variants did not fulfill the NCCN criteria<sup>21</sup> for assessment as high-risk for genetic or familial cancers.

The median age at diagnosis of patients with pathogenic germline variants was 53 years, which was younger than that of patients with no pathogenic variants (60 years) ( $P < 0.00001$ ) (Fig. 1b, Supplementary Fig. 3a). *BRCA1* germline mutations were associated with younger age at diagnosis (median age, 43 years), compared with *BRCA2* (median age, 56 years;  $P = 0.08$ ), and other 6 genes (median age, 52 years;  $P = 0.08$ ). Among early onset breast cancer patients who were diagnosed before 35 years of age, prevalence of germline variants in *BRCA1*, *BRCA2*, and other genes was 9.8%, 17.1%, and 4.9%, respectively (Supplementary Fig. 3b).

As previously reported<sup>3</sup>, *BRCA1* variant carriers were more likely to have triple-negative (estrogen receptor (ER) negative, progesterone receptor (PR) negative, and HER2 negative) diseases, compared with those with other germline variants ( $P = 0.0007$ ) and those without germline mutations ( $P = 0.0001$ ) (Fig. 1c). In contrast, there was no obvious tumor subtype associated with *BRCA2* variant carriers. Advanced (T2-T4) breast cancers were more common in *BRCA1*-mutated cases, compared with those with pathogenic germline mutations in other genes ( $P = 0.08$ ) or those with no pathogenic variants detected ( $P = 0.05$ ) (Fig. 1d), especially among younger patients



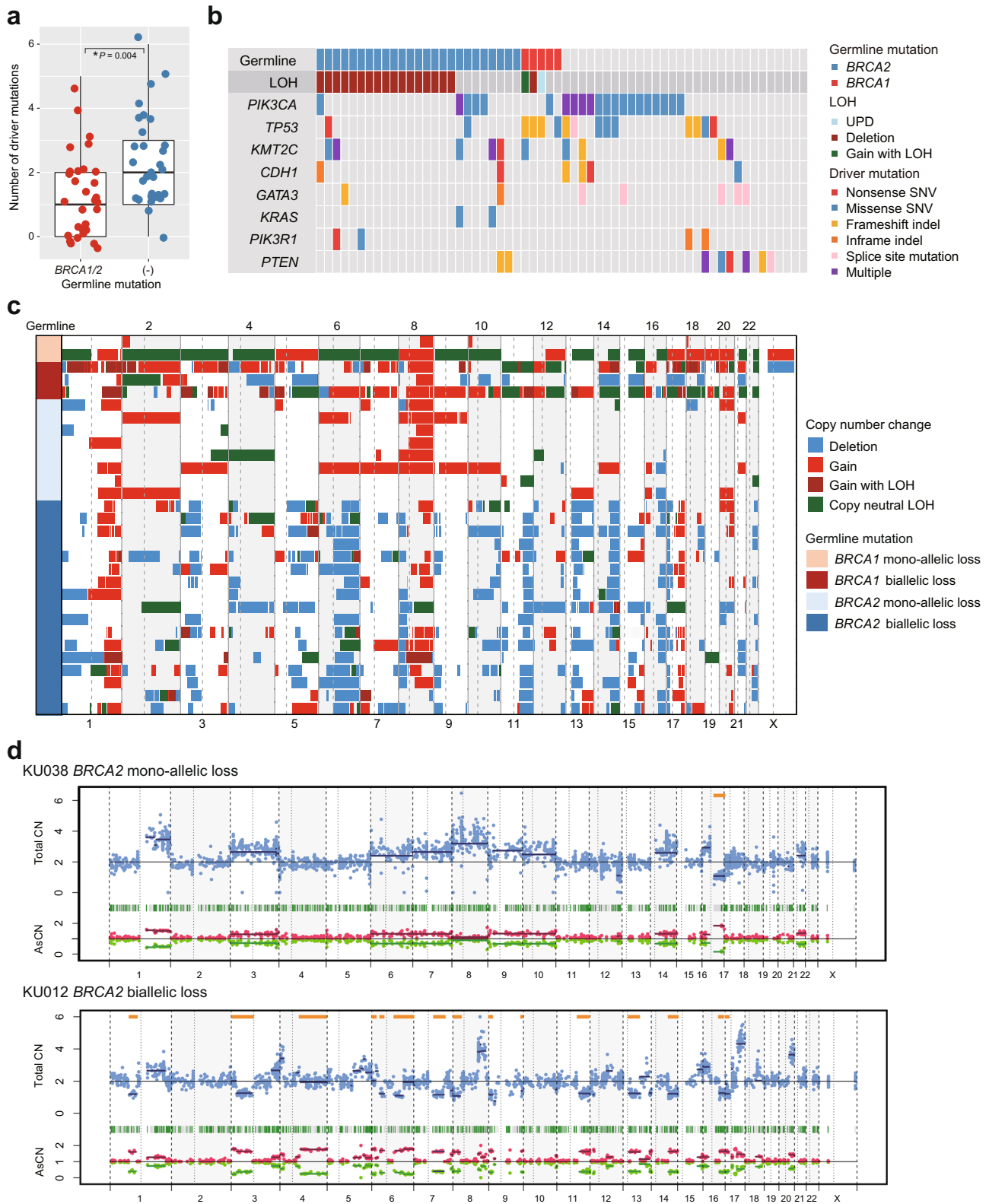
**Fig. 1** Clinical characteristics of breast tumors harboring pathogenic germline variants. **a** Numbers and types of germline variants identified in each gene. **b** Box-and-whisker plots of age at breast cancer diagnosis for tumors with germline variants ( $n = 101$ ), including *BRCA1* ( $n = 15$ ), *BRCA2* ( $n = 62$ ), and others ( $n = 24$ ), and those without ( $n = 1892$ ). The boxes indicating median and interquartile range, and the whiskers denoting the range. Asterisks indicate significance difference (Mann–Whitney  $U$  test:  $P < 0.05$ ). **c** Distribution of phenotype according to the status of germline mutations. DCIS ductal carcinoma in situ, TNBC triple-negative breast cancer, LumHER2 luminalHer2. **d** Distribution of  $t$  factor for tumors according to the status of germline mutations. Tis carcinoma in situ.

(Supplementary Fig. 3c). Lymph node involvement did not significantly correlate with the status of pathogenic germline variants (Supplementary Fig. 3d). Survival data were available for 1387 (69.5%) of 1995 patients with a median follow-up of 3 years (range: 0–35 years). No prognostic impact of germline mutations was demonstrated for overall and disease-free survival in both univariate and multivariate analyses (Supplementary Fig. 4, Supplementary Table 4).

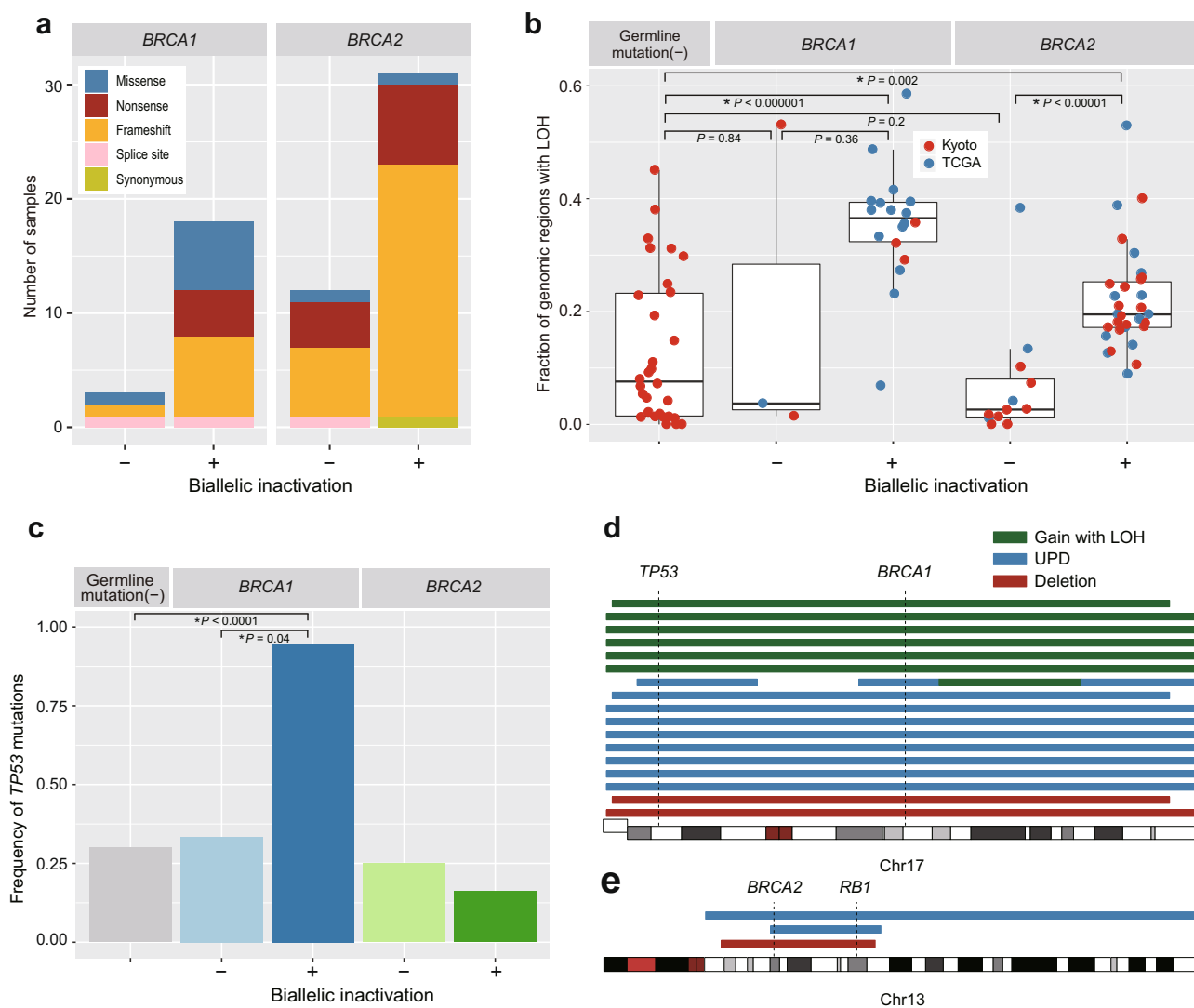
**Somatic alterations in tumors with germline *BRCA1* and *BRCA2* variants.** Tumor samples were obtained from 30 patients with pathogenic germline variants in *BRCA2* ( $n = 25$ ) and *BRCA1* ( $n = 5$ ), as well as from an additional 30 patients without pathogenic germline mutations. Somatic mutations in common breast cancer drivers and CNAs were analyzed for these samples using targeted panel sequencing (Supplementary Table 5). In total, 19 of 30 samples with germline variants in *BRCA1/2* had one or more somatic mutations in 18 driver genes with a median of 1 mutation/sample, which was significantly smaller than those without germline variants (median 2 mutations/sample,  $P = 0.004$ ) (Fig. 2a, Supplementary Table 6). For tumors with germline *BRCA1/2* mutations, somatic mutations were most frequently detected in *PIK3CA* ( $n = 6$ ), *TP53* ( $n = 6$ ) and *KMT2C* ( $n = 6$ ) (Fig. 2b, Supplementary Table 7). All samples had CNAs,

regardless of the presence or absence of a pathogenic germline mutation. Even though the two-hit hypothesis of tumorigenesis predicts that majority of cases will have biallelic inactivation of the relevant cancer predisposing loci, biallelic inactivation of the predisposing alleles was found in only 20 cases (67%), while the remaining 10 retained an intact allele (mono-allelic inactivation). For each of the 20 cases, biallelic inactivation was caused by loss-of-heterozygosity (LOH) affecting the relevant germline variant loci. Indeed, nearly all cases involved deletions in *BRCA2* (17/17) and *BRCA1* (1/3), followed by copy-neutral LOH (uniparental disomy) ( $n = 1$ ) and gain with LOH ( $n = 1$ ). These results suggest that LOH is the predominant mechanism of biallelic inactivation of *BRCA1/2* genes. Interestingly, one tumor with a germline *BRCA2* variant also harbored a low allele frequency of somatic *BRCA2* mutation in addition to LOH, suggesting clonal heterogeneity in the tumor over time, with independent events leading to biallelic inactivation.

**Mono-allelic vs. biallelic inactivation of *BRCA1/2*.** *BRCA1/2* genes normally function in DNA repair and their deleterious mutation has been linked to HR deficiency. Hence, in our evaluation of tumors with mono-allelic and biallelic inactivation of *BRCA1/2*, we first analyzed CNAs (Fig. 2c). As seen for representative cases in Fig. 2d, samples with mono-allelic *BRCA1/2*



**Fig. 2 Genetic landscape of breast cancers with pathogenic germline variants in BRCA1/2.** **a** Numbers of driver mutations in tumors with germline variants in BRCA1/2 ( $n = 30$ ) and those without ( $n = 30$ ). Asterisks indicate significance difference (Mann-Whitney  $U$  test:  $P < 0.05$ ). **b** The landscape of somatic variants in tumors with germline variants in BRCA1/2 ( $n = 30$ ) and those without ( $n = 30$ ). Distribution and types of recurrently mutated genes in tumors with germline BRCA1/2 mutations are shown. **c** Genome-wide CN changes in tumors with germline variants in BRCA1/2. Red: gain, dark red: gain with LOH, green: copy neutral LOH, blue: deletion. **d** Representative CN plots of tumors with and without biallelic inactivation. Regions with LOH are shown in orange lines. CN copy number, AsCN allele specific CN.



**Fig. 3 Genetic characteristics of breast tumors harboring germline *BRCA1/2* mutations with and without biallelic inactivation. a** Types of germline variants in *BRCA1/2* according to the status of biallelic inactivation ( $n = 64$ ). **b** Fraction of genomic regions with LOH in tumors without germline variants ( $n = 30$ ) and in those harboring germline *BRCA1* mutations with ( $n = 18$ ) and without ( $n = 3$ ) biallelic inactivation and those with germline *BRCA2* mutations with ( $n = 31$ ) and without ( $n = 12$ ) biallelic inactivation. Asterisks indicate significance difference (Mann-Whitney  $U$  test:  $P < 0.05$ ). **c** Frequency of *TP53* mutations in tumors without germline variants ( $n = 30$ ) and those harboring germline *BRCA1/2* mutations with and without biallelic inactivation ( $n = 64$ ). Asterisks indicate a significance difference (Fisher test:  $P < 0.05$ ). **d** LOHs on chromosome 17 in patients with germline *BRCA1* mutations ( $n = 17$ ). LOHs identified in a single patient are shown on a line. **e** LOHs on chromosome 13 in patients with germline *BRCA2* mutations ( $n = 3$ ).

lesions mainly showed CN gains with rare deletions. In contrast, biallelic *BRCA1* inactivation showed extensive gains with or without LOH, and those with biallelic *BRCA2* inactivation were characterized by extensive deletions or copy neutral LOHs. Although no specific patterns of CNAs have been reported in breast cancers with biallelic *BRCA1/2* inactivation, these observations led us to hypothesize that specific genomic rearrangements caused by HR deficiency could cause characteristic CN changes. Combining the Japanese cases with 34 cases from TCGA breast cancer cases, we tested this hypothesis using a total of 64 cases with germline *BRCA1/BRCA2* variants. In total, biallelic inactivation of *BRCA1/2* loci was observed in 49 of 64 (77%) patients: 18/21 (86%) cases with *BRCA1* and 31/43 (72%) cases with *BRCA2* germline mutations (Fig. 3a). Tumors with mono-allelic *BRCA1* and *BRCA2* mutations are less common in the TCGA cases, compared with the Japanese cases, particularly those with mono-allelic *BRCA1* mutations (1/16 vs. 2/5 for *BRCA1* and 4/18 vs. 8/25 for *BRCA2*). The predominant role of LOH was also

confirmed in TCGA cases, in which LOH explained 93% of biallelic inactivation in both *BRCA1*- and *BRCA2*-mutated cases, whereas biallelic inactivation via compound germline and somatic mutations were found only in two cases.

We next investigated characteristic patterns of mutations and structural variants (SVs) associated with biallelic *BRCA1* or *BRCA2* inactivation, focusing on mutational signatures and SVs using whole-exome sequencing (WES) data in the TCGA cohort. Four predominant mutational signatures were identified using pmsignature<sup>22</sup> (Supplementary Fig. 5a). Of these, the mutational signature caused by deficient HR (Sig\_3) was more frequent in tumors with biallelic *BRCA2* inactivation than those without germline mutations, which concurs with previous reports<sup>7,8</sup> (Supplementary Fig. 5b, c). In analysis of SVs for tumors with biallelic inactivation of *BRCA1* and *BRCA2*, compared to tumors without germline variants, increased occurrence of tandem duplications and deletions (for *BRCA1* inactivation) and deletions (for *BRCA2* inactivation) were observed (Supplementary Fig. 5d,



e). By contrast, tumors with mono-allelic inactivation of either *BRCA* gene did not show an increase in Sig<sub>3</sub> mutations and deletions/tandem duplications. In addition, tumors with biallelic *BRCA2* inactivation exhibited significantly more extensive LOH lesions compared with those with mono-allelic inactivation and those without germline variants (Fig. 3b). Tumors with biallelic *BRCA1* inactivation also tended to have more extensive LOH than those without germline variants, but a comparison with mono-allelic and biallelic *BRCA1* inactivation was inconclusive due to the small number of patients with tumors of this mono-allelic category. These results suggest that biallelic *BRCA1/2* inactivation causes extensive CNAs, in addition to small SVs.

Strikingly, except for one case, which displayed compound germline and somatic *BRCA1* mutations, all but one tumors with biallelic *BRCA1* inactivation (17/18) harbored *TP53* mutations (Fig. 3c, Supplementary Fig. 5f). The *TP53* mutations were accompanied by high variant allele frequency and loss of an intact chromosome 17, leading to biallelic *TP53* inactivation (Fig. 3d, Supplementary Fig. 6a). Of added interest in this regard is the observation that some tumors with biallelic *BRCA2* inactivation, commonly accompanied by LOH of chromosome 13, also exhibited concomitant biallelic inactivation of *RB1*, which was mutated in 3 cases (Fig. 3e, Supplementary Fig. 6b). *RB1* mutations were more frequent in tumors with biallelic loss of *BRCA2* (3/31, 9.7%) than those without (22/858 cases in our cohort and TCGA dataset, 2.6%) ( $P = 0.05$ ). These results suggest that loss of chromosomes 17 and 13 play an important role in the development of breast cancer in patients with mutated *BRCA1* and *BRCA2*, though inactivating *TP53* and *RB1*, respectively. However, the number of tumors identified in our study that exhibited *RB1* mutation in addition to biallelic loss of *BRCA2* is small ( $n = 3$ ), and further studies are warranted to confirm the association of these genetic lesions.

Finally, we evaluated the clinical characteristics of patients with mono-allelic and biallelic *BRCA1/2* inactivation. Patients with biallelic inactivation showed a significantly younger onset than those without (median age at diagnosis: 47 vs. 59.6 years) ( $P = 0.01$ ) (Fig. 4a), with no significant difference between patients with biallelic *BRCA1* and *BRCA2* inactivation. Although not significant, tumors having biallelic *BRCA1* inactivation tended to have more advanced (T2–T4) ( $P = 0.09$ ) and triple-negative breast cancer ( $P = 0.55$ ) (Fig. 4b, c, Supplementary Table 8). We also analyzed tumors with germline *BRCA1* and *BRCA2* variants for classification into PAM50 gene expression subtypes using TCGA samples. In accordance with previous reports, tumors with biallelic *BRCA1* inactivation were more frequently classified as basal-type<sup>23,24</sup>, compared to those without (Fig. 4d). In strong contrast to biallelic *BRCA1/2* inactivation, samples with mono-allelic *BRCA1/2* inactivation were not associated with younger age at onset, or an increase in triple-negative or basal-type tumors (Supplementary Table 8). The mutation status of *BRCA1/2* or the presence or absence of biallelic involvement of these genes did not affect overall or disease-free survival both in univariate and multivariate regression analyses (Supplementary Fig. 7, Supplementary Table 9).

## Discussion

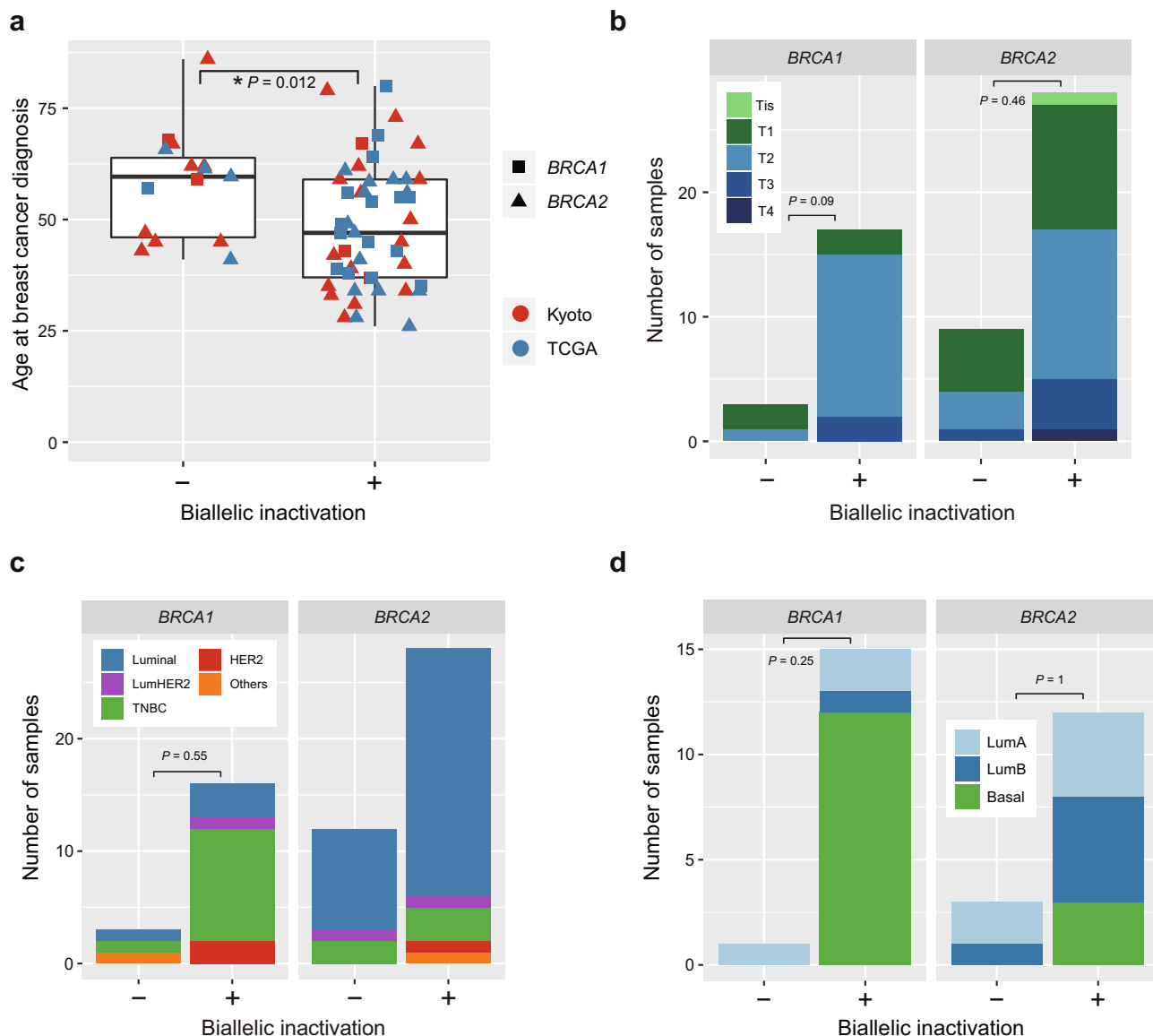
In the current study, pathogenic germline mutations were detected in 5.1% of 1995 unselected Japanese breast cancer patients, which was equivalent to the frequency in the previous report of Japanese cohort<sup>4</sup>. The incidence rate of *BRCA1/2* variants was also relatively similar with those reported in Japanese patients<sup>4</sup> and Chinese population<sup>3</sup>. Given that the current study was restricted to detection of SVs and other variants that are not registered in ClinVar, the actual prevalence of pathogenic

germline variants might be underestimated. Importantly, a half of the cases (50/101) was negative for a family history of breast cancer, or did not fulfill the NCCN criteria for high-risk of familial cancer, indicating the importance of investigating germline DNA, even among sporadic breast cancer patients.

In line with the previous reports<sup>8,10,25</sup>, tumors with mono-allelic *BRCA1* and *BRCA2* mutations were frequently observed in our cohort. Tumors with biallelic *BRCA1/2* inactivation had unique genetic features, in terms of CNA and *BRCA*-associated mutational signature and SVs, which were not seen in those with mono-allelic inactivation. Although the only significant clinical difference between tumors with mono-allelic and biallelic *BRCA1/2* inactivation was age at onset, the frequency of advanced stage, triple-negative or basal tumors tended to be higher in tumors with biallelic inactivation. The correspondence of biallelic *BRCA1/2* inactivation with earlier age of onset conflicts with results of a previous study<sup>10</sup>. Although both studies included TCGA samples, here, we carefully removed low quality samples and whole genome-amplified samples from the analysis. Furthermore, we considered compound germline and somatic mutations of *BRCA1/2* as contributors to biallelic inactivation, which likely increased the sensitivity of the dataset to detection of a correlation between the status of *BRCA1* and *BRCA2* and the clinical variables (Supplementary Fig. 8).

Tumors with mono-allelic *BRCA2* mutations and those without *BRCA1/BRCA2* mutations did not differ in their clinical presentation or analyses of additional genetic effects. In particular, mono-allelic tumors did not show an enhanced *BRCA*-related mutational signature or increase in SVs, which was seen in tumors with biallelic *BRCA1/2* inactivation. Nevertheless, mono-allelic germline *BRCA1/2* mutations show a significant enrichment in breast cancer patients, compared with the general population. Despite having only sequenced a portion of the tumors with *BRCA1/2* variants in our cohort (30/77), a mono-allelic loss-of-function mutation in *BRCA1* and *BRCA2* was more frequent than within the control cohort<sup>4</sup>:  $\geq 2/1995$  (0.1%) vs.  $5/11,241$  (0.04%) for *BRCA1* and  $\geq 8/1995$  (0.4%) vs.  $15/11,241$  (0.13%) for *BRCA2*. Thus, mono-allelic mutation alone does seem to play a role in the development of breast cancer, which is supported by several biological studies showing the effects of haploinsufficiency of *BRCA1/2* in carcinogenesis<sup>26,27</sup>. This raises the question as to whether or not platinum<sup>28</sup> or poly (ADP-ribose) polymerase (PARP) inhibitor<sup>29</sup> are also effective against tumors with mono-allelic mutations. A lower sensitivity of PARP inhibitors to cells with heterozygous *BRCA* mutations have indeed been reported using in vitro<sup>30</sup> and mouse models<sup>31</sup>. In contrast to this, however, Jonsson et al.<sup>32</sup> reported that the allelic status did not affect the response of in tumors with germline *BRCA1/2* mutations to PARP inhibitors<sup>32</sup>. Further evaluation of these drugs in the context of biallelic inactivation, is required in the future, incorporating clinical follow-up of patients, or clinical trials.

Finally, our study has revealed an intriguing linked mechanism of biallelic inactivation of *TP53* with *BRCA1* and *RB1* with *BRCA2*, respectively. Tumors showing biallelic inactivation of both *BRCA1* and *TP53* genes were almost invariably associated with loss of normal chromosome 17. Of interest, a previous study using fluorescence in situ hybridization (FISH) and immunohistochemistry of *TP53* on tumors with biallelic inactivation of both *BRCA1* and *TP53* have demonstrated that *TP53* mutation occurs before LOH of the intact chromosome 17 as there remained cells with two chromosome 17 alleles and mutated *TP53*. The subsequent biallelic *BRCA1* inactivation by LOH thus leads to biallelic inactivation of both *TP53* and *BRCA1* simultaneously<sup>25</sup>. A similar scenario is suggested here for *RB1* and *BRCA2* on chromosome 13, leading to simultaneous biallelic inactivation of the two genes via a deletion of part of chromosome 13.



**Fig. 4 Phenotypes of breast tumors harboring germline *BRCA1/2* mutations with and without biallelic inactivation.** **a** Age at diagnosis of breast cancers harboring germline *BRCA1/2* mutations with ( $n = 49$ ) and without ( $n = 15$ ) biallelic inactivation. The asterisk indicates significance difference (Mann-Whitney  $U$  test:  $P < 0.05$ ). **b-d** Distribution of T factor: **(b)**, in subtypes based on immunohistochemistry **(c)** in PAM50 mRNA-expression subtypes **(d)** of breast cancers harboring germline *BRCA1/2* variants with and without biallelic inactivation. The difference in frequency of advanced (T2–T4) tumors, triple-negative breast cancer (TNBC), and basal-type tumors between tumors with and without biallelic inactivation were tested by the two-sided Fisher’s exact test. Tis carcinoma in situ, LumHER2 luminalHer2, LumA luminalA, LumB luminalB.

In summary, we revealed that breast tumors with pathogenic germline *BRCA1/2* variants show different genetic and clinical characteristics depending on the presence or absence of biallelic inactivation of these genes. Along with the recent data of different impact of mono-allelic and biallelic somatic *TP53* mutations in myelodysplastic syndromes<sup>33</sup>, our data highlights the importance of allelic status of cancer driver genes.

## Methods

**Patients and samples.** A total of 2136 breast cancer patients were enrolled in this study, treated between September 2011 and October 2016 at Kyoto Breast Cancer Research Network institutions, consisting of Kyoto University Hospital and 17 affiliated institutions. Among these, 1995 cases fulfilled the following inclusion criteria (Supplementary Fig. 1a): female; sufficient amount of high quality genomic DNA; pathological diagnosis of breast cancer; clinical data of at least one from age of onset, histology, phenotype, grade, clinical stage, past history, and family history.

These cases were collected consecutively with no selection bias. Family history was defined as the presence of one or more first- or second-degree relatives with

breast and/or ovarian cancer. ER, PR, and HER2 status was determined by immunohistochemistry and/or FISH using breast tumor tissue obtained from a core needle biopsy or taken during surgery. For HER2 status, an immunohistochemistry score of 0 and 1+ was considered negative, whereas 3+ was considered positive. Tumors with a score of 2+ were evaluated further by FISH.

Written informed consent was obtained from all participants. The study was reviewed and approved by the Ethics Committees of Kyoto University Graduate School and Faculty of Medicine and Kyoto University Hospital and was performed in accordance with the Declaration of Helsinki (2013 revision).

**Targeted sequencing of pooled DNA samples.** Genomic DNA samples were extracted from peripheral blood samples of patients using a Gentra Puregene Kit (Qiagen). After adjusting the concentration of each genomic DNA sample to 50 ng/ $\mu$ L, each sample from 10 to 20 consecutive patient extractions were combined into one DNA pool<sup>34</sup>, generating a total of 106 DNA pools. Pooled DNA samples were analyzed using targeted-capture sequencing of 11 genes implicated in hereditary breast cancer using the SureSelect Custom kit (Agilent Technologies). RNA probes were designed to cover all coding regions and intron–exon boundaries of the 11 breast cancer susceptibility genes. Captured libraries were sequenced on a



HiSeq 2000 (Illumina). To confirm the sensitivity of pooled sequencing, we first generated 2 pools from 40 patients. These 40 patients were individually genotyped and 60 private germline variants in 11 genes were detected. Next, 2 pools were sequenced at 1436–1476 $\times$ , and 58/60 (97%) of variants were of variant allele frequency (VAF) > 0.005. Subsequently, 106 DNA pools were sequenced. The mean coverage of pooled DNA samples was 1644 $\times$  (1267 $\times$ –2316 $\times$ ). Reads were aligned to the reference human genome GRCh37. Germline variations were called using EBCall (Empirical Bayesian mutation Calling, <https://github.com/friend1ws/EBCall>)<sup>35,36</sup> with the following parameters: (i) a variant allele frequency (VAF)  $\geq 0.005$ ; (ii) a  $P$  value < 0.01 (by EBCall); (iii) removal of SNPs in ESP, the 1000 genomes project, ExAC and HGVD (<http://www.genome.med.kyoto-u.ac.jp/SnpDB/>) with a minor allele frequency of  $\geq 0.001$ ; (iv) support from  $\geq 2$  reads. Variants were annotated using Annovar (<http://annovar.openbioinformatics.org/en/latest/>). Three *BRCA1/2* mutations already detected in clinical purposes were successfully identified. For each pool with positive variant calls, the variant-positive samples were interrogated by amplicon-sequencing or Sanger sequencing of all samples in the corresponding pool with germline variants. Finally, 101 out of 117 germline mutations were validated. High allele frequencies of variants were supported by both of sequencing chromatograms and the VAFs of amplicon sequencing. Except for the long (20 bp) deletion variants, the VAFs of deep sequencing were  $\geq 0.33$ , supporting the germline nature of these mutations.

**Variant classification of germline variants.** Truncating mutations (nonsense mutations or frameshift indels) were considered as pathogenic, except for low-risk truncating mutations, such as the K3326X mutations of *BRCA2*. Missense, synonymous and splice site mutations registered as “pathogenic” or “likely pathogenic” in ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>)<sup>37,38</sup> were also considered as pathogenic variants in this study (Supplementary Fig. 1b).

**Targeted sequencing of tumor samples.** To identify and characterize somatic mutations in tumors from patients with germline variants in *BRCA1/2*, 36 tumor samples with germline variants in *BRCA1/2* and those without pathogenic germline mutations ( $n = 35$ ) were analyzed by target sequencing using a SureSelect system (Agilent). All the tumor samples were collected prior to treatment. For formalin-fixed paraffin-embedded (FFPE) samples, a KAPA Hyper Prep Kit (KAPA Biosystems, Wilmington, MA) was also used before target enrichment. RNA probes were designed to capture 115 genes associated with breast cancer (Supplementary Table 5) and 1275 SNP sites for the measurement of genomic CNs. Based on the allele frequency of mutations and CN changes, we excluded 11 samples with a lower tumor cell fraction from further analysis. Finally, 30 samples with germline variants in *BRCA1/2*, including frozen tissues ( $n = 5$ ) or FFPE samples ( $n = 25$ ), and 30 tumors without germline mutations were analyzed. The mean coverage of fresh frozen and FFPE samples were 599 $\times$  (347 $\times$ –1253 $\times$ ) and 293 $\times$  (112 $\times$ –557 $\times$ ), respectively. Somatic mutations were analyzed using EBCall, with the following parameters: (i) removal of SNPs in ESP, the 1000 genomes project, ExAC and HGVD with a minor allele frequency of  $\geq 0.001$ ; (ii) support from  $\geq 5$  reads in the tumor; (iii) a VAF  $\geq 0.02$ ; (iv) a  $P$  value < 0.001 (by EBCall); (v) support from reads mapped to both strands. Variants with a VAF  $\geq 0.4$  were removed as germline SNPs, except for loss-of-function mutations in tumor suppressor genes and gain-of-function mutations reported in the COSMIC database. Synonymous variants were also excluded as germline variants. Mapping errors were removed by visual inspection on the Integrative Genomics Viewer (<http://www.broadinstitute.org/igv/>) browser. Finally, mutations in 28 driver genes reported in a previous study<sup>6</sup> (Supplementary Table 5) and hot spot mutations reported in the COSMIC database with  $\geq 10$  mutated tumors, including *KRAS* and *CDKN2A* mutations, were considered as driver mutations. To confirm the accuracy of mutation calling, we also called single nucleotide variants using MuTect<sup>39</sup> with unmatched control samples with the following parameters: (i) removal of SNPs in ESP, the 1000 genomes project, ExAC and HGVD with minor allele frequency of  $\geq 0.001$ ; (ii) support from  $\geq 5$  reads in a tumor; (iii) a VAF  $\geq 0.02$ ; (iv) a tumor\_alt\_fpr\_max > 0. Variants with VAF  $\geq 0.4$  and synonymous variants were also removed as germline SNPs, except for loss-of-function mutations in tumor suppressor genes and gain-of-function mutations reported in the COSMIC database. We confirmed a high concordance rate between mutations call using both methods, except for a small number of variants with low VAF (Supplementary Fig. 9a).

CN changes were analyzed using CNACS ([https://github.com/papaemmelab/toil\\_cnacs](https://github.com/papaemmelab/toil_cnacs))<sup>40</sup>, in which the total number of sequencing reads covering each bait region or SNP probe, and the allele frequency of the heterozygous SNP were used as input data. For the identification of regions with LOH, in addition to deletions and copy-neutral LOHs called by CNACS, we identified gains with LOH based on the estimated tumor purities by total and allele specific (As) CN (Supplementary Fig. 9b). For regions with gain (CN = 3), tumor purity was estimated as follows: (by total CN) Purity = Total CN - 2; (by As CN) Purity =  $2 \times (1 - \text{As CN}) / \text{As CN}$  (gain without LOH); and Purity =  $2 \times (1 - \text{As CN}) / (2 + \text{As CN})$  (gain with LOH).

We used Control-FREEC<sup>41</sup> with the contaminationAdjustment option, which corrects for contamination by normal cells, to confirm the LOH status of *BRCA1/2* loci determined by CNACS. The median of the total and As CN of probes within *BRCA1/2*, estimated by CNACS and Control-FREEC, were well correlated (Supplementary Fig. 9c). To further confirm the accuracy of CN calling, we called CN changes using SNP array karyotyping for fresh frozen samples ( $n = 5$ ). SNP

array-based CN analysis was performed using CNAG software<sup>42,43</sup>. SNP array analysis also provided an almost identical CN profile, including LOH of the *BRCA1* and *BRCA2* loci (Supplementary Fig. 10), and CNACS detected 57/61 CN alterations identified by SNP array.

**Analysis of TCGA dataset.** Samples subjected to whole-genome amplification were excluded from analysis to accurately define CN changes and SVs. Sequencing data of 829 WES of breast cancer tumors in the TCGA dataset were downloaded. Variants with a VAF  $\geq 0.2$  in the germline control sample and with <0.001 minor allele frequency in ESP, the 1000 genomes project, ExAC and HGVD were considered as germline mutations. Tumors with pathogenic germline variants of *BRCA1/2* ( $n = 38$ ) were defined in the same way as our cohort, and those with somatic *BRCA1/2* mutations were not included. Four samples were excluded for the following reasons: low quality of CN data ( $n = 3$ ) and low tumor purity ( $n = 1$ ). Somatic variants were detected using EBCall with the following parameters: (i) VAF in tumor samples  $\geq 0.05$ ; (ii) a  $P$  value <  $10^{-1.3}$  (by Fisher’s test); (iii)  $P$  value < 0.0001 (by EBCall). Mutational signatures were analyzed using pmsignature<sup>22</sup>, and three samples with a high number of artifacts, including TCGA-A2-A0T5, TCGA-A2-A0T6, and TCGA-A7-A0DB were excluded from the analysis. CN changes were also detected from the WES data using CNACS. SVs were analyzed using GenomSV (<https://github.com/Genomon-Project/GenomonSV>) as previously reported<sup>44</sup> with additional filters, (i) a frequency in the germline sample < 0.02; (ii) a  $P$  value <  $10^{-1.5}$  (by Fisher’s test); (iii) a length of overhang  $\geq 100$ . SVs identified in other samples were also removed as germline variants or errors. Clinical information and PAM50 mRNA subtypes of these samples were extracted from the TCGA database.

**Statistics and reproducibility.** A comparison of categorical variables between mutation carriers and noncarriers was made using the Fisher’s exact test or chi-square test where appropriate. For continuous variables, the Mann–Whitney  $U$  test was used for group comparisons.  $P$  values less than 0.05 were considered statistically significant. The overall survival time for all patients was determined from the date of diagnosis of breast cancer to the time of last follow-up, or death, by examining medical records. Survival was estimated using the Kaplan–Meier product-limit method and differences were tested for statistical significance using the log-rank test. We performed multivariate regression analysis using the Cox proportional hazards model. All analyses were performed using JMP Pro 14.0.0 software.

No statistical methods were used to determine sample size since this is an exploratory study. We enrolled as many patients as possible who provided consent for our study during the enrollment period between September 2011 and October 2016. A total of 2136 breast cancer patients were enrolled in this study.

**Data access.** Targeted sequencing data of 106 pooled DNAs and 60 breast cancer samples have been deposited at the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega/>) under the accession Nos. EGAS00001004630 and EGAS00001004182, respectively.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Received: 23 March 2020; Accepted: 15 September 2020;  
Published online: 16 October 2020

## References

- Easton, D. F. et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **372**, 2243–2257 (2015).
- Tung, N. et al. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J. Clin. Oncol.* **34**, 1460–1468 (2016).
- Sun, J. et al. Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. *Clin. Cancer Res.* **23**, 6113–6119 (2017).
- Momozawa, Y. et al. Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. *Nat. Commun.* **9**, 4083 (2018).
- Sugano, K. et al. Cross-sectional analysis of germline *BRCA1* and *BRCA2* mutations in Japanese patients suspected to have hereditary breast/ovarian cancer. *Cancer Sci.* **99**, 1967–1976 (2008).
- Nakamura, S. et al. Prevalence and differentiation of hereditary breast and ovarian cancers in Japan. *Breast Cancer* **22**, 462–468 (2015).
- Polak, P. et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genet.* **49**, 1476–1486 (2017).
- Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
- Nielsen, F. C., van Overeem Hansen, T. & Sørensen, C. S. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat. Rev. Cancer* **16**, 599–612 (2016).

10. Maxwell, K. N. et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.* **8**, 319 (2017).
11. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
12. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
13. Banerji, S. et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409 (2012).
14. Shah, S. P. et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
15. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
16. Sanchez-Garcia, F. et al. Integration of genomic data enables selective discovery of breast cancer drivers. *Cell* **159**, 1461–1475 (2014).
17. Antoniou, A. C. et al. Breast-cancer risk in families with mutations in PALB2. *N. Engl. J. Med.* **371**, 497–506 (2014).
18. Couch, F. J. et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J. Clin. Oncol.* **33**, 304–311 (2015).
19. Sekine, M. et al. Mutational analysis of BRCA1 and BRCA2 and clinicopathologic analysis of ovarian cancer in 82 ovarian cancer families: two common founder mutations of BRCA1 in Japanese population. *Clin. Cancer Res.* **7**, 3144–3150 (2001).
20. Kim, Y. C. et al. Prevalence and spectrum of BRCA germline variants in mainland Chinese familial breast and ovarian cancer patients. *Oncotarget* **7**, 9600–9612 (2016).
21. Daly, M., Pilarski, R. & Berry, M. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®Genetic/Familial High-Risk Assessment: Breast and Ovarian Version 1, 2018).
22. Shiraishi, Y., Tremmel, G., Miyano, S. & Stephens, M. A simple model-based approach to inferring and visualizing cancer mutation signatures. *PLoS Genet.* **11**, e1005657 (2015).
23. Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100**, 8418–8423 (2003).
24. Foulkes, W. D. et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J. Natl Cancer Inst.* **95**, 1482–1485 (2003).
25. Martins, F. C. et al. Evolutionary pathways in BRCA1-associated breast tumors. *Cancer Discov.* **2**, 503–511 (2012).
26. Skoulidis, F. et al. Germline Brca2 heterozygosity promotes Kras(G12D)-driven carcinogenesis in a murine model of familial pancreatic cancer. *Cancer Cell* **18**, 499–509 (2010).
27. Konishi, H. et al. Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc. Natl Acad. Sci. USA* **108**, 17773–17778 (2011).
28. Tutt, A. et al. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. *Nat. Med.* **24**, 628–637 (2018).
29. Robson, M. et al. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. *N. Engl. J. Med.* **377**, 523–533 (2017).
30. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
31. Drew, Y. et al. Therapeutic potential of poly(ADP-ribose) polymerase inhibitor AG014699 in human cancers with mutated or methylated BRCA1 or BRCA2. *J. Natl Cancer Inst.* **103**, 334–346 (2011).
32. Jonsson, P. et al. Tumour lineage shapes BRCA-mediated phenotypes. *Nature* **571**, 576–579 (2019).
33. Bernard, E. et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat. Med.* **26**, 1549–1556 (2020).
34. Calvo, S. E. et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat. Genet.* **42**, 851–858 (2010).
35. Shiraishi, Y. et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41**, e89 (2013).
36. Yoshida, K. et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* **45**, 1293–1299 (2013).
37. Pritchard, C. C. et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N. Engl. J. Med.* **375**, 443–453 (2016).
38. Zhang, J. et al. Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
39. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
40. Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
41. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

42. Nannya, Y. et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
43. Yamamoto, G. et al. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
44. Kataoka, K. et al. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature* **534**, 402–406 (2016).

## Acknowledgements

We are grateful to the TCGA Research Network (<http://cancergenome.nih.gov/>) for providing the data analyzed in this manuscript. We would like to thank all of the Kyoto Breast Cancer Research Network (KBRCN) participants, including Mitsuru Tanaka, Susumu Mashima, Tsuyoshi Tachibana, Shingo Sakata, Fumiaki Yotsumoto, Norimichi Kan, and Nobuhiko Shinkura for their help with providing DNA samples and clinical information of the patients in their hospital. We also thank Maki Nakamura for her technical assistance. This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Health, Labor and Welfare of Japan (15H05909) [S.O.]. This research was also partially supported by Astra Zeneca Co., Ltd. We would like to thank Enago ([www.enago.jp](http://www.enago.jp)) for the English language review.

## Author contributions

Y.I.-K., K. Yoshida, N.K.-S., S.O., and M. Toi conceived the project. Y. Shiozawa, H.T., Y. Shiraishi, A.S.-O., Y.N., K.C., and S.M. developed the bioinformatics pipelines of the sequencing data. Y.I.-K. prepared samples and performed the sequencing experiments. Y.I.-K., K. Yoshida, T.N., Y. Shiozawa, Y.T., and Y.I. performed the sequencing data analyses. Y.F. performed SNP array karyotyping. M. Takada and Y. Kataoka helped statistical analyses. M.K., N.S., E.S., M. Takada, M. Torii, H.Y., K.Y., R.O., Y.M., H.K., S.T., A.Y., H.S., and T.I. collected the specimens and clinical information. Y.I.-K., K. Yoshida, S.O., and M. Toi wrote the paper. Y.I.-K. and K. Yoshida generated the figures and tables. S.O. and M. Toi led the entire project. All authors participated in discussions and interpretation of the data and results.

## Competing interests

The authors declare the following competing interests: Masakazu Toi: Research grant: Astra Zeneca, Pfizer, Eisai, Chugai Pharma, Astellas Pharma, Daiichi Sankyo, Taiho Pharmaceutical, C & C Research Laboratories, Ono Pharmaceutical, Shimadzu, Nippon-Kayaku, AFI technology, Japan Breast Cancer Research Group. Consulting or Advisory Role: Daiichi Sankyo, Kyowa Hakko Kirin, Athenex Oncology and Bertis. Honoraria as lecture fee: AstraZeneca, Bayer, C & C Research Laboratories, Chugai Pharma, Daiichi Sankyo, Eisai, Eli Lilly, Genomic Health, Konica Minolta, Kyowa Hakko Kirin, MSD, Novartis, Pfizer, Sanofi, Shimadzu, Taiho Pharmaceutical, Takeda, Yakult. Travel and accommodation expenses: Genomic Health Inst., Eli Lilly and companies. Others: A member of board of directors (no salary) Organization for Oncology and Translational Research, Japan Breast Cancer Research Group and Kyoto Breast Cancer Research Network.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-020-01301-9>.

**Correspondence** and requests for materials should be addressed to S.O. or M.T.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020