



# Genomic landscape of ductal carcinoma in situ and association with progression

Chieh-Yu Lin<sup>1,7</sup> · Sujay Vennam<sup>1</sup> · Natasha Purington<sup>2</sup> · Eric Lin<sup>1</sup> · Sushama Varma<sup>1</sup> · Summer Han<sup>2</sup> · Manisha Desai<sup>3</sup> · Tina Seto<sup>4</sup> · Nicholas J. Wang<sup>5</sup> · Henning Stehr<sup>1</sup> · Megan L. Troxell<sup>1</sup> · Allison W. Kurian<sup>6</sup> · Robert B. West<sup>1</sup> 

Received: 3 May 2019 / Accepted: 7 August 2019 / Published online: 17 August 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

**Purpose** The detection rate of breast ductal carcinoma in situ (DCIS) has increased significantly, raising the concern that DCIS is overdiagnosed and overtreated. Therefore, there is an unmet clinical need to better predict the risk of progression among DCIS patients. Our hypothesis is that by combining molecular signatures with clinicopathologic features, we can elucidate the biology of breast cancer progression, and risk-stratify patients with DCIS.

**Methods** Targeted exon sequencing with a custom panel of 223 genes/regions was performed for 125 DCIS cases. Among them, 60 were from cases having concurrent or subsequent invasive breast cancer (IBC) (DCIS + IBC group), and 65 from cases with no IBC development over a median follow-up of 13 years (DCIS-only group). Copy number alterations in chromosome 1q32, 8q24, and 11q13 were analyzed using fluorescence in situ hybridization (FISH). Multivariable logistic regression models were fit to the outcome of DCIS progression to IBC as functions of demographic and clinical features.

**Results** We observed recurrent variants of known IBC-related mutations, and the most commonly mutated genes in DCIS were *PIK3CA* (34.4%) and *TP53* (18.4%). There was an inverse association between *PIK3CA* kinase domain mutations and progression (Odds Ratio [OR] 10.2,  $p < 0.05$ ). Copy number variations in 1q32 and 8q24 were associated with progression (OR 9.3 and 46, respectively; both  $p < 0.05$ ).

**Conclusions** *PIK3CA* kinase domain mutations and the absence of copy number gains in DCIS are protective against progression to IBC. These results may guide efforts to distinguish low-risk from high-risk DCIS.

**Keywords** Ductal carcinoma in situ · *PIK3CA* · Copy number variant · Breast cancer

## Introduction

Ductal carcinoma in situ (DCIS) of the breast is a risk factor as well as a precursor lesion for invasive breast cancer (IBC). The current standard of treatment for DCIS involves surgery

in combination with radiation therapy and/or endocrine therapy [1–4]. However, this treatment paradigm was developed based on the natural history of DCIS and IBC in the premammography screening era. In recent decades, DCIS detection rates have increased significantly due to contemporary, advanced screening imaging modalities [5]. Concurrently, multiple studies have demonstrated that only 13–52% of patients with DCIS eventually develop subsequent IBC [3,

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10549-019-05401-x>) contains supplementary material, which is available to authorized users.

✉ Robert B. West  
rbwest@stanford.edu

<sup>1</sup> Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

<sup>2</sup> Department of Medicine, Quantitative Sciences Unit, Stanford University, Stanford, CA, USA

<sup>3</sup> Department of Medicine and of Biomedical Data Science, Quantitative Sciences Unit, Stanford University, Stanford, CA, USA

<sup>4</sup> Research Information Technology, Stanford University School of Medicine, Stanford, CA, USA

<sup>5</sup> Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, USA

<sup>6</sup> Departments of Medicine and of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA

<sup>7</sup> Department of Pathology and Immunology, School of Medicine, Washington University in St. Louis, St. Louis, MO, USA

6], suggesting that many patients with DCIS may not require extensive treatment, and raising the concern that some DCIS patients are being overtreated.

Therefore, a precision medicine approach to stratify risk of developing IBC among DCIS patients is critically needed. The discovery of genomic features that correlate with high-risk and low-risk DCIS would be important for tailored IBC screening and prevention. A number of studies, including ours, have examined the changes in the genome during the progression of breast neoplasia to IBC. The findings suggest that single nucleotide variations (SNVs) and copy number alterations (CNAs) both are acquired over a series of genomic events, which occur over the course of development of the IBC [7–11].

In our prior studies, we have examined genomic changes in hyperplasia, DCIS, and IBC by targeted sequencing [8], whole genome sequencing [7], and fluorescence in situ hybridization (FISH) [12] to identify genomic changes. These studies have found recurrent genomic changes in pre-invasive neoplasia, both CNAs and SNVs, which have also been identified in IBC. Some events, like *PIK3CA* mutations, can occur quite early in the neoplastic timeline [13] while others, like *ERBB2* amplification, occur later [14]. However, our hypothesis is that there is no single genomic feature that correlates with the transition from DCIS to IBC; instead, it is likely that a constellation of features or higher-order features, such as genome complexity or gene pathway alterations, are driving progression.

In this study, we investigated the mutational profiles of DCIS cases that do not develop IBC over a long follow-up interval (median 13 years), and DCIS cases that are initially associated with IBC or later develop IBC. We hypothesize that by combining molecular signatures with clinicopathologic features, we can elucidate the biology of breast cancer progression, and risk-stratify patients with DCIS.

## Materials and methods

### Patient population

Available cases were identified in the Department of Pathology at the Stanford University Hospital (SUH) from 2000 to June 2011. DCIS cases with adequate tissue for research sampling and confirmed follow-up were categorized as follows: (1) DCIS-only group: DCIS and no development of IBC over a median follow-up of 13 years (average 11 years) or (2) DCIS + IBC group: DCIS with concurrent or subsequent IBC present. There is no restraint imposed for the size of associated invasive cancer or the detection methods (screen-detect v.s. palpable mass, etc.). Surgical samples with sufficient tissue were collected with Health Insurance Portability and Accountability

Act (HIPAA)-compliant Stanford University Institutional Review Board (IRB) approval (Protocol number 32496). Clinical data were obtained from the Oncoshare breast cancer research database, which has been described previously [15, 16].

### Generation of targeted-capture libraries

Hematoxylin- and eosin-stained slides were reviewed to confirm the diagnosis of DCIS. The areas with abundant amounts of DCIS were chosen for examination. For cases with DCIS and invasive carcinoma present in the same specimen, we carefully selected the area with abundant DCIS away from the invasive component. The DCIS samples were acquired by taking 3–10 numbers of 2-mm cores from the corresponding areas of paraffin blocks (from cases on tissue microarrays TA 239, 419, 420, and 445). The thickness of the tissue in each core is approximately 2–3 mm, so we do not anticipate a major contamination from invasive carcinoma in deeper content of the core. Only tumor samples were analyzed, and no paired normal samples were included. The DNA was extracted using RecoverAll Total Nucleic Acid Isolation kit (Ambion). Targeted libraries of genomic DNAs were generated using the Agilent SureSelect XT kit and Agilent Automation Systems NGS system. Additional information is provided in the supplementary materials and methods.

### Sequencing data analysis

The sequencing data were analyzed using a custom pipeline. In brief, reads were aligned to the hg19 human reference genome assembly using BWA [17]. Duplications were marked using Picard Tools V1.118 (<http://broadinstitute.github.io/picard>). Cases with less than 92 M aligned reads were discarded. Insertion–deletion realignment and base recalibration were achieved using GATK v.3.3-0 [18]. The somatic variant calls were carried out using an ensemble approach with four variant callers: MuTect [19], VarScan2 [19], VarDict [20], and Freebayes. [21] Calls present in at least two out four callers were accepted. The variant annotation was done using ANNOVAR [22] and custom scripts. The full list of nonsynonymous variants is provided in Supplementary Table S6. Bedtools coverage was used to create a histogram of coverage for each feature in the BED file and a summary histogram of all of features in the BED file. These histograms were then plotted using R to show the percentage of capture regions covered at any given depth for every individual sample. The sequencing data were uploaded to the NCBI Sequence Read Archive.

## Fluorescence in situ hybridization (FISH)

FISH was performed as previously described [12], and additional details are provided in the Supplementary Materials and methods. Total test-probe green counts (1q32.1, 8q24.21, 11q13.11) were compared with red (2q37.3) control-probe counts, which are frequently unaltered in breast cancer [23]. The signals were evaluated according to two parameters: signals per cell and ratio of test probe to control probes. Cases were scored as gain at the locus if the target to control probe ratio was greater than 1.5 or the number of test signals was greater than three per cell.

## Statistical analysis

Descriptive statistics were presented of baseline characteristics among the two DCIS groups (DCIS-only group and DCIS + IBC group). The student *t* test was used to assess whether differences in mutation burden existed between the DCIS alone and DCIS with IBC cohorts.

Two main multivariable logistic regression models were fit to characterize the association between DCIS progression to IBC and clinical and demographic features. The first model fits DCIS progression to IBC as a function of categorical copy number status, age at diagnosis, race/ethnicity, DCIS nuclear grade, tumor size, margins, surgery type, and an indicator for lack of *PIK3CA* mutation in the kinase region (*PIK3CA*-KD). The second model additionally adjusted for ER+ status and HER2+ status. Odds ratios and 95% confidence intervals were reported. Further details on additional models fit are available in the supplementary materials and methods. All statistical analyses were performed in R (Version 3.2.2, Vienna, Austria) [24].

## Results

### Patient population and demographic data

Cases of DCIS and documented follow-up in two cohorts were identified for genomic studies: (1) DCIS and no development of IBC (DCIS-only group) or (2) DCIS with concurrent/subsequent IBC events (DCIS + IBC group). A total of 125 DCIS cases were successfully analyzed, including 65 cases (52%) in the DCIS-only group and 60 cases (48%) in the DCIS + IBC group. In the DCIS + IBC group, 5 DCIS cases had IBC detected at a later time point. All patients were female, with median age at DCIS diagnosis of 51 years (range 29–89 years, Table 1). DCIS was characterized histologically and immunophenotypically; more than half (77 cases, 62%) demonstrated high-grade nuclei. The median tumor size was 2.4 cm (range 0.4–13.0 cm). The majority (72%) of cases had tumor margins of 0.2 cm and above.

Sixty-one percent of surgeries were mastectomies. Eighty-one cases (65%) were positive for estrogen receptor (ER). Seventy-two cases (58%) were negative for HER2 (0, 1+, or 2+ by immunohistochemistry), and 27 cases (21.6%) were positive for HER2 (3+ by immunohistochemistry). The univariable analysis of these parameters for association with IBC is presented in supplementary Table S2, and the detailed clinicopathologic data are provided in supplementary Table S7.

### The genomic profile of DCIS is similar to IBC

Targeted sequencing for common mutations in IBC was performed on 125 cases of DCIS. The targeted regions include SNPs and coding exons of known breast-cancer/pan-cancer-related genes, including *APC*, *AR*, *ATM*, *BAP1*, *BRAF*, *CCND1*, *CHD1*, *CDKN1A*, *CTNNA1*, *DICER1*, *DNMT3A*, *EGFR*, *ERBB2*, *FOXA1*, *GATA3*, *IDH1/2*, *KRAS*, *MED12*, *MYB*, *NF1*, *NOTCH1*, *PIK3CA*, *PTEN*, *RBI*, *VHL*, and *WT1* (full list in Supplementary Materials and methods). The average read count obtained per case was 5,604,491. Additional quality control data are presented in Supplementary Table S5 and Supplementary Fig. 1. We excluded calls with coverage less than 30 reads. To focus on mutations having tumor suppressive or oncogenic effects, we limited the analysis to recurrent position mutations identified in the TCGA dataset that are frequently mutated in IBC. The most commonly mutated genes were *PIK3CA* (34.4%) and *TP53* (18.4%) (Fig. 1a). We observed no significant difference of mutational burdens in known breast cancer-associated genes between the two groups of DCIS cases ( $p > 0.05$  for all the genes) (Fig. 1b). We further analyzed variants based on their locations within the functional domains in each gene [25–27]. We identified “hotspots” of somatic mutations in the helical domain and kinase domain of *PIK3CA* for both DCIS-only and DCIS + IBC groups (Fig. 2). There was a significant enrichment of *PIK3CA* kinase domain mutations (*PIK3CA*-KD mutations) in the DCIS-alone group ( $p = 0.029$ ). Analysis of other domains in *PIK3CA* genes found no statistically significant differences in recurrent mutations outside the kinase domain. Similar domain analysis was also performed for other frequently mutated genes, *TP53* and *GATA3*, and no further predictive mutational profiles were identified.

In addition to SNV and small insertion–deletion mutations identified by targeted exon sequencing, we also interrogated larger-scale copy number variations in selected “hotspot” genomic areas in the two DCIS groups (Tables 1 and S1). Three chromosomal loci were measured by FISH, 1q32, 8q24, and 11q13, based on prior genomic data on invasive breast cancer [28] and DCIS [12]. The current study cohort consists of 73 cases (58.4%) from the previously published cohort [12], and 52 new cases that have not been analyzed

**Table 1** Baseline features of the DCIS cases

Characteristic	DCIS-only ( <i>n</i> = 65)	DCIS + IBC ( <i>n</i> = 60)	Total ( <i>n</i> = 125)
<b>Age at diagnosis (years)</b>			
Less than 40	10 (15%)	11 (18%)	21 (17%)
40–49	19 (29%)	15 (25%)	34 (27%)
50–64	23 (35%)	23 (38%)	46 (37%)
65 and older	13 (20%)	11 (18%)	24 (19%)
<b>Race/ethnicity</b>			
NH Asian/Pacific Islander	11 (17%)	10 (17%)	21 (17%)
NH white	47 (72%)	43 (72%)	90 (72%)
Other	6 (9%)	4 (7%)	10 (8%)
<b>Surgery type</b>			
Lumpectomy	27 (42%)	21 (35%)	48 (38%)
Mastectomy	37 (57%)	39 (65%)	76 (61%)
No surgery	1 (2%)	0	1 (1%)
<b>ER status<sup>a</sup></b>			
Positive	46 (71%)	35 (58%)	81 (65%)
Negative	10 (15%)	19 (32%)	29 (23%)
Missing	9 (14%)	6 (10%)	15 (12%)
<b>HER2 status<sup>a</sup></b>			
Positive	15 (23%)	12 (20%)	27 (22%)
Negative	37 (57%)	35 (58%)	72 (58%)
Missing	13 (20%)	13 (22%)	26 (21%)
<b>Tumor size (cm)</b>			
Median (range)	2.4 (0.4–10.9)	2.4 (0.7–13.0)	2.4 (0.4–13.0)
<b>Tumor margins</b>			
<0.2 cm	19 (29%)	14 (23%)	33 (26%)
0.2 cm +	46 (71%)	44 (73%)	90 (72%)
<b>DCIS nuclear grade</b>			
Low	5 (8%)	2 (3%)	7 (6%)
Intermediate	22 (34%)	16 (27%)	38 (30%)
High	38 (58%)	39 (65%)	77 (62%)
<b>Copy number variations</b>			
None	20 (31%)	5 (8%)	25 (20%)
Gene 1q only	10 (15%)	13 (22%)	23 (18%)
Gene 8q24 only	1 (2%)	6 (10%)	7 (6%)
Gene 11q13 only	2 (3%)	1 (2%)	3 (2%)
Two of three gains	10 (15%)	20 (33%)	30 (24%)
All three gains	10 (15%)	12 (20%)	22 (18%)

DCIS ductal carcinoma in situ, IBC invasive breast cancer, NH non-Hispanic

<sup>a</sup>Equivocal is included in the negative category

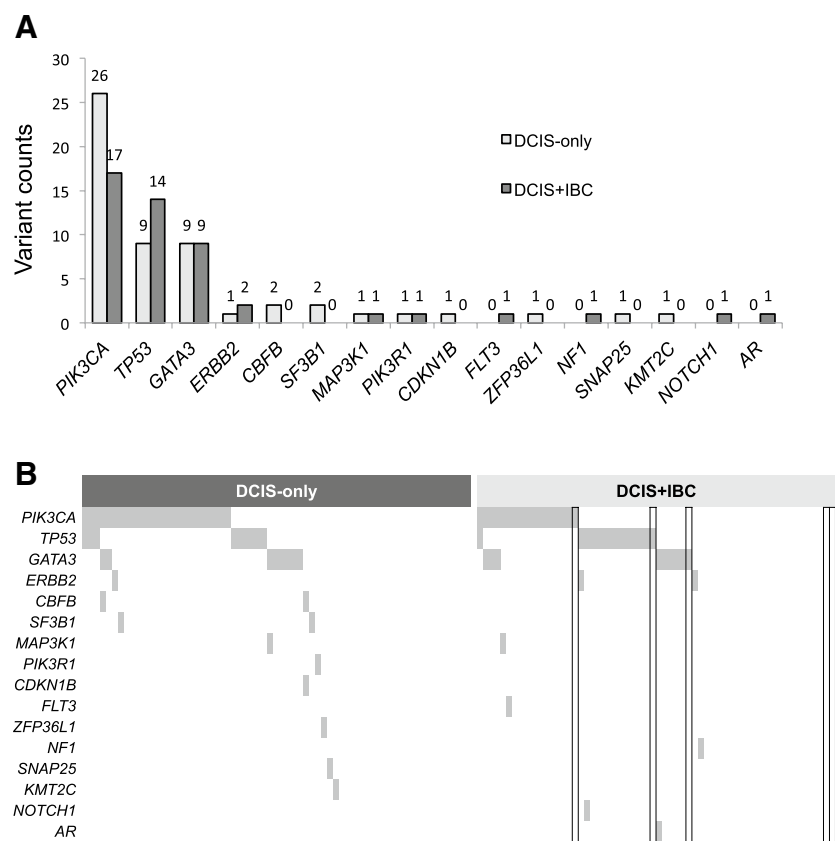
before. Consistent with our prior data, frequency of 1q32 gain is the highest in the cohort (59.0%), followed by 8q24 (48.2%) and 11q13 (32.5%).

### Multivariable analysis demonstrates strong correlation between *PIK3CA-KD* mutation and risk of progression

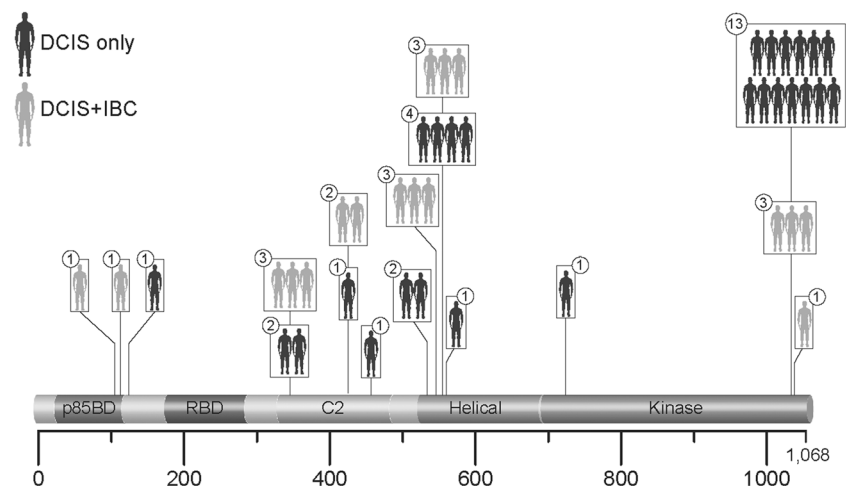
We performed multivariable analysis, examining the association of IBC dependent on variables including *PIK3CA-KD*

mutational status, copy number gains, age, race, nuclear grade, tumor size, margins, and surgery type (Table 2). After removing cases with missing data, we had 97 complete DCIS cases. We identified a statistically significant association between lack of *PIK3CA-KD* mutation and increased risk of IBC ( $p < 0.05$ ). Patients without *PIK3CA-KD* mutations were 4.52 times (confidence interval: 1.05–25.27) as likely to have IBC compared to subjects with the mutation. This association was also statistically significant in the univariable analysis (Supplementary Table S2). When subdividing

**Fig. 1** Genomic landscape of DCIS. **a** Distributions of known recurrent IBC-associated variants are (a) displayed in a bar graph with the number of cases denoted above the bars; **b** displayed in a heat map. In the group of DCIS + IBC, cases with subsequent IBC events are highlighted in the boxes



**Fig. 2** Distribution of *PIK3CA* mutations in two groups of DCIS cases. The number in the circles depicts the number of DCIS cases harboring the mutation in that particular position. Of note, one of the DCIS-only cases exhibited two *PIK3CA*-KD variants



the DCIS + IBC group into the cases with synchronous or recurrent IBC (Supplementary Table S8), similar trend was noticed. The number of cases in the group of DCIS + subsequent IBC is not powerful enough to draw a conclusion.

When additionally controlling for ER and HER2 status (complete case number = 81, Table 3), the association between lack of *PIK3CA*-KD mutation and IBC risk remained statistically significant: DCIS patients without the *PIK3CA*-KD mutations were 10.22 times as likely to be progress to IBC as compared to DCIS with *PIK3CA*-KD

mutations ( $p < 0.05$ , Table 3). Similar results were observed when DCIS nuclear grades were grouped into a two-tiered system (low-grade vs non-low-grade DCIS; high-grade vs non-high-grade DCIS) (data not shown). In addition, the inverse association between any *PIK3CA* mutation and IBC risk became statistically significant (OR 4.66,  $p < 0.05$ , data not shown).

The presence of genomic copy number gain (1q32 only, 8q24 only, or two or three of three gains) was also associated with increased risk of progression to IBC (Table 3). In

**Table 2** Multivariable associations with DCIS progression to IBC ( $n=97$ )

Characteristic	Odds ratio (95% CI)
<b>Copy number variations</b>	
None	Reference
Gene 1q only	5.44** (1.16, 30.46)
Gene 8q24 only	21.92** (2.00, 603.68)
Gene 11q13 only	1.71 (0.06, 29.98)
Two of three gains	7.61*** (1.90, 36.86)
All three gains	2.73 (0.58, 14.82)
<b>Age at diagnosis</b>	
Less than 40	0.45 (0.09, 2.17)
40–49	0.86 (0.25, 2.88)
50–64	Reference
65 and older	0.88 (0.21, 3.76)
<b>Race/ethnicity</b>	
NH Asian/Pacific Islander	1.67 (0.19, 16.46)
NH white	1.55 (0.26, 10.58)
Other	Reference
<b>DCIS nuclear grade</b>	
Low	Reference
Intermediate	0.54 (0.05, 6.33)
High	0.78 (0.07, 9.41)
<b>Tumor size</b>	1.08 (0.85, 1.40)
<b>Tumor margins &lt; 0.2 cm</b>	1.61 (0.53, 5.10)
<b>Lumpectomy</b>	0.77 (0.26, 2.21)
<b>No <i>PIK3CA-KD</i> mutation</b>	4.52** (1.05, 25.27)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

addition, there were a trend that, but not statistically significant, overexpression of HER2 is inversely associated with the risk of IBC (OR 0.28,  $p < 0.1$ ). This trend became statistically significant (OR 0.24,  $p < 0.05$ ) when modeling any *PIK3CA* mutations instead of *PIK3CA-KD* mutations in the multivariable analysis (data not shown).

ER status and DCIS nuclear grade are important pathological features that are routinely examined in the clinical setting. Therefore, we investigated the interaction between ER/nuclear grade and the *PIK3CA* mutation status. The association between lack of *PIK3CA-KD* mutations and progression to IBC was not modified by ER status (Supplementary Table S3). The association between lack of any *PIK3CA* mutations (including *PIK3CA-KD* mutations and other *PIK3CA* mutations) and progression to IBC was modified by ER status ( $p < 0.05$ , Supplementary Table S4). There was a statistically significant interaction between DCIS nuclear grade and the presence of *PIK3CA-KD* mutation for the association with IBC (data not shown). The effect of *PIK3CA-KD* mutation appeared to be dependent on the nuclear grade: the association with no progression to IBC was stronger in the group of high-grade DCIS patients.

**Table 3** Multivariable associations with DCIS progression to IBC include ER and HER2 statuses ( $n=81$ )

Characteristic	Odds ratio (95% CI)
<b>Copy number variations</b>	
None	Reference
Gene 1q only	9.31** (1.57, 76.66)
Gene 8q24 only	45.96** (2.58, 1,917.18)
Gene 11q13 only	2.89 (0.07, 83.36)
Two of three gains	15.68*** (2.75, 129.49)
All three gains	10.85** (1.36, 120.55)
<b>Age at diagnosis</b>	
Less than 40	0.32 (0.05, 2.06)
40–49	0.78 (0.17, 3.47)
50–64	Reference
65 and older	1.77 (0.32, 10.77)
<b>Race/ethnicity</b>	
NH Asian/Pacific Islander	0.91 (0.07, 12.41)
NH white	2.74 (0.34, 28.03)
Other	Reference
<b>DCIS nuclear grade</b>	
Low	Reference
Intermediate	0.40 (0.02, 6.56)
High	0.68 (0.03, 13.12)
<b>No <i>PIK3CA-KD</i> mutation</b>	10.22** (1.61, 101.71)
<b>ER positive</b>	0.57 (0.11, 2.75)
<b>HER2 positive</b>	0.30 (0.07, 1.23)
<b>Tumor size</b>	0.84 (0.59, 1.16)
<b>Tumor margins &lt; 0.2 cm</b>	2.04 (0.54, 8.93)
<b>Lumpectomy</b>	0.83 (0.20, 3.24)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

## Discussion

While the genomic landscape of IBC has been extensively studied [29–33], the molecular profiling of DCIS is still under investigation. Limited data have been published for DCIS genomic profiling, using either next-generation sequencing or array CGH [9, 34–39]. To the best of our knowledge, this is the largest cohort of genomic profiling with longitudinal clinical follow-up of DCIS cases that did not progress. *PIK3CA*, *TP53*, and *GATA3* are among the most commonly mutated genes in DCIS, and chromosome 1q and 8q copy number gains are frequently identified in DCIS, as seen in prior studies [34–36, 39]. Of note, in these previous reports, there was no recurrent mutation in a single gene that could stratify the risk of progression to IBC. However, further analysis based on protein domains was not performed in these previous publications.

We demonstrated a novel finding that the somatic mutations in the *PIK3CA* kinase domain provide predictive value of DCIS progression to IBC. The presence of mutations in

this particular domain is associated with lower risk of concurrent or subsequent IBC. Previous studies have investigated the role of *PIK3CA* mutations in in situ breast cancers. In a small cohort of lobular carcinoma in situ (LCIS) without invasive lobular carcinoma (ILC) versus LCIS with associated ILC, the presence of *PIK3CA* mutations was not correlated with progression [40]. One of the studies showed that in ER-positive/HER2-negative DCIS, *PIK3CA* “hotspot mutations” were more prevalent in DCIS associated with IBC, compared with DCIS alone [41]. However, these “hotspot mutations” queried in this study include mutations in C2, helical, and kinase domains. This difference could account for the different conclusion drawn in our study, as we demonstrated that specifically *PIK3CA* kinase domain mutations are associated with lack of progress.

The finding that activating mutations in *PIK3CA* and overexpression of HER2 oncogenes are correlated with a tendency not to progress to IBC is unexpected. Variants of these two oncogenes are quite prevalent in IBC. In cultures and animal model systems, they demonstrate biologic influences that promote neoplastic growth, invasion, or metastasis [42–44]. Given that conventional models of cancer-associated progression with oncogenic mutations, it is surprising that two prominent oncogenes would be inversely correlated with the progression of DCIS to IBC.

One possible explanation for our findings is that alterations in specific pathways (such as HER2, *PIK3CA*) allow the cells to overcome immediate biological constraints in the process of tumorigenicity, rather than specifically promoting the DCIS to IBC transition. After the initial biologic challenges have been overcome, their influence on progression may be diminished. That they remain at high incidence in the invasive carcinoma may be in part due to the genomic difficulty or biological ambivalence in removing these somatic changes. In fact, *PIK3CA* mutations are extremely common in hyperplastic lesions of the breast [45]. HER2 is often amplified in DCIS and other noninvasive breast lesions, with a higher rate of amplifications in preinvasive lesions than in IBC [46–49]. A large cohort study with long-term follow-up data from Sweden showed that HER2-positive DCIS has lower risk of progression to IBC compared to HER2-negative DCIS [50]. Similar results have been reported independently [51]. For *PIK3CA*, previous researchers have found that in a subset of paired DCIS alone and DCIS with IBC samples, the *PIK3CA* mutations were present in DCIS alone but not DCIS with IBC, or with lower alternative allele frequency in the IBC component [37, 41, 52, 53]. *PIK3CA* KD mutations (exon 20 mutations) have also been observed in preneoplastic lesions (usual ductal hyperplasia, columnar cell change, or atypical ductal hyperplasia), while paired IBC lesions lack such mutations [45]. These results and our current findings suggest that selection for

HER2 amplification or *PIK3CA* mutations may address neoplastic challenges that occur well before the transition from in situ to invasive cancer.

Notably, the genomic features we found to correlate with risk of progression consisted of aneuploidies or large amplicons. While there are a number of known and suspected oncogenes present in the chromosomal regions with recurrent copy number alterations, it is not clear whether a single driving event is responsible for these somatic changes. We speculate that higher-order function attributable to the gross change in genomic composition (e.g., copy number gains that likely have widespread effects on cellular function and genomic instability) may influence progression rather than a more precise influence on a specific gene function or related pathway. This observation has also been made by other studies in the literature [9, 54]. In our prior whole genome sequencing study, clonally related progression was marked by recurrent events of aneuploidy at the earliest stages and successive DNA copy number events throughout progression to invasion [7].

There are some limitations of our study. First, while some associations in the multivariable analyses were significant, those with large confidence intervals should be carefully interpreted. An independent cohort is required to validate these findings. Also, more than half (57%) of the DCIS-only patient received mastectomy specimen. The choice of mastectomy over lumpectomy potentially could drastically decrease the rate subsequent IBC event. Indeed, in our cohort, only 5 patients subsequently developed invasive disease. In addition, without paired normal controls, we could not reliably distinguish between germline and somatic events. However, we focused on ‘hotspot’ mutations with known impacts on protein functions, such as *PIK3CA* kinase domain mutations, which are easily recognized. Moreover, there are several other genes that are mutated at frequencies that are likely to be significant for a clinical classifier, such as *TP53*, *GATA3*, and *MAPK3*. However, beyond these, most genes are mutated at less than 1% and are thus unlikely to be useful as clinical biomarkers of progression.

The strength of our cohort is the long-term comprehensive follow-up data, to ensure DCIS-only cases were indeed without invasive or metastatic events. These cases are considered as “low-risk” clinically. It is important to identify this type of DCIS patients, who carry low-risk of progression and could consider forgoing extensive treatment. In this study, we hypothesized that the molecular signatures of the “low-risk” DCIS are distinct from the “high-risk” DCIS. DCIS patients with either subsequent or synchronous IBC are considered “high-risk.” We acknowledge that the risk of progress in the “high-risk” DCIS group may be heterogeneous, and that DCISs with synchronous IBCs have even higher risk and may carry a different pathological mechanism and molecular profiles compared the DCISs with

subsequent IBCs. This is an interesting, important, yet separate hypothesis that is beyond the scope of this current study.

Future studies are necessary to provide deeper understanding of this novel finding of the *PIK3CA-KD* predictive value in DCIS progression. Previous literature has suggested that different *PIK3CA* variants could cause different downstream signaling pathway alterations and biological functions in *in vitro* systems [55, 56]. One key goal is to map the occurrence and evolution of these genomic alterations in early neoplastic and precursor lesions, as well as in paired invasive and metastatic samples. This would help us to understand the roles of *PIK3CA-KD* mutations and copy number gains in breast cancer progression. It is also imperative to combine transcriptional and proteomics analyses, in order to interrogate the downstream effects related to the *PIK3CA-KD* mutations.

Our novel findings that *PIK3CA-KD* mutations are associated with relative lack of DCIS progression to IBC, coupled with other traditional and novel risk factors, contribute to knowledge enhancement of the sequence and mechanisms of breast cancer progression. These data also begin to demonstrate the possibility of an integrated clinicopathologic-molecular risk classifier of DCIS. For example, women with low nuclear grade, ER+, *PIK3CA-KD* mutant DCIS may have particularly low risk of progression. Larger studies with more complete histologic, immunohistochemical, proteomic, molecular, treatment and long-term follow-up data are necessary to build risk -assessment models of precision to counsel patients, tailor therapy, and reduce the overtreatment of the more indolent forms of DCIS.

**Acknowledgements** Short-read sequencing assays were performed by the OHSU Massively Parallel Sequencing Shared Resource. We thank Norman Cyr for his artistic contribution to Fig. 2.

**Author contributions** RBW conceived and designed the study. CYL conceived and carried out experiments, as well as analyzed the data. S Vennam and HS analyzed the sequencing data (S Vennam: single-nucleotide alteration; HS: copy number variations). EL and S Varma carried out the FISH experiments. NJW carried out the targeted exon sequencing. NP, SH, MD, and TS performed biostatistics analysis. MLT conceived the study. AK developed and led the Oncoshare data resource. All authors were involved in writing the paper and gave final approval to the submitted and published versions.

**Funding** This work is supported by NIH R01CA193694, the Susan and Richard Levy Gift Fund; the Suzanne Pride Bryan Fund for Breast Cancer Research; the Breast Cancer Research Foundation; the Jan Weimer Junior Faculty Chair in Breast Oncology; the BRCA Foundation; and the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California. The project was supported by an NIH CTSA award number UL1 RR025744. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the state-wide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract

HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors, and endorsements by the University or State of California, the California Department of Health Services, the National Cancer Institute, or the Centers for Disease Control and Prevention or their contractors and subcontractors are neither intentional nor should be inferred.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interests.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. As stated in the Materials and Methods section, the study was performed with Health Insurance Portability and Accountability Act (HIPAA)-compliant Stanford University Institutional Review Board (IRB) approval (Protocol number 32496). Because archival tissue was used, a waiver of consent was obtained.

## References

1. Meijnen P, Oldenburg HS, Peterse JL et al (2008) Clinical outcome after selective treatment of patients diagnosed with ductal carcinoma in situ of the breast. *Ann Surg Oncol* 15:235–243
2. Cutuli B, Cohen-Solal-le Nir C, de Lafontan B et al (2002) Breast-conserving therapy for ductal carcinoma in situ of the breast: the French Cancer Centers' experience. *Int J Radiat Oncol Biol Phys* 53:868–879
3. Wapnir IL, Dignam JJ, Fisher B et al (2011) Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst* 103:478–488
4. Cuzick J, Sestak I, Pinder SE et al (2011) Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: long-term results from the UK/ANZ DCIS trial. *Lancet Oncol* 12:21–29
5. Ernster VL, Barclay J, Kerlikowske K et al (1996) Incidence of and treatment for ductal carcinoma in situ of the breast. *JAMA* 275:913–918
6. Fisher ER, Dignam J, Tan-Chiu E et al (1999) Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) eight-year update of Protocol B-17: intraductal carcinoma. *Cancer* 86:429–438
7. Newburger DE, Kashef-Haghighi D, Weng Z et al (2013) Genome evolution during progression to breast cancer. *Genome Res* 23:1097–1108
8. Weng Z, Spies N, Zhu SX et al (2015) Cell-lineage heterogeneity and driver mutation recurrence in pre-invasive breast neoplasia. *Genome Med* 7:28
9. Johnson CE, Gorringer KL, Thompson ER et al (2012) Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. *Breast Cancer Res Treat* 133:889–898



10. Rane SU, Mirza H, Grigoriadis A et al (2015) Selection and evolution in the genomic landscape of copy number alterations in ductal carcinoma in situ (DCIS) and its progression to invasive carcinoma of ductal/no special type: a meta-analysis. *Breast Cancer Res Treat* 153:101–121
11. Kroigard AB, Larsen MJ, Laenkholtm AV et al (2015) Clonal expansion and linear genome evolution through breast cancer progression from pre-invasive stages to asynchronous metastasis. *Oncotarget* 6:5634–5649
12. Afghahi A, Forgo E, Mitani AA et al (2015) Chromosomal copy number alterations for associations of ductal carcinoma in situ with invasive breast cancer. *Breast Cancer Res* 17:108
13. Troxell ML, Brunner AL, Neff T et al (2012) Phosphatidylinositol-3-kinase pathway mutations are common in breast columnar cell lesions. *Mod Pathol* 25:930–937
14. Xu R, Perle MA, Inghirami G et al (2002) Amplification of Her-2/neu gene in Her-2/neu-overexpressing and -nonexpressing breast carcinomas and their synchronous benign, premalignant, and metastatic lesions detected by FISH in archival material. *Mod Pathol* 15:116–124
15. Weber SC, Seto T, Olson C et al (2012) Oncoshare: lessons learned from building an integrated multi-institutional database for comparative effectiveness research. *AMIA Annu Symp Proc* 2012:970–978
16. Kurian AW, Mitani A, Desai M et al (2014) Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. *Cancer* 120:103–111
17. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
18. McKenna A, Hanna M, Banks E et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
19. Cibulskis K, Lawrence MS, Carter SL et al (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31:213–219
20. Lai Z, Markovets A, Ahdsmaki M et al (2016) VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* 44:e108
21. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. [arXiv:12073907](https://arxiv.org/abs/12073907)
22. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164
23. Gao Y, Niu Y, Wang X et al (2009) Genetic changes at specific stages of breast cancer progression detected by comparative genomic hybridization. *J Mol Med (Berl)* 87:145–152
24. R development Core Team (2015) R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna
25. Porta-Pardo E, Godzik A (2014) e-Driver: a novel method to identify protein regions driving cancer. *Bioinformatics* 30:3109–3114
26. Porta-Pardo E, Garcia-Alonso L, Hrabe T et al (2015) A pan-cancer catalogue of cancer driver protein interaction interfaces. *PLoS Comput Biol* 11:e1004518
27. Yang F, Petsalaki E, Rolland T et al (2015) Protein domain-level landscape of cancer-type-specific somatic mutations. *PLoS Comput Biol* 11:e1004147
28. Curtis C, Shah SP, Chin SF et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486:346–352
29. Cancer Genome Atlas N (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490:61–70
30. Horlings HM, Weigelt B, Anderson EM et al (2013) Genomic profiling of histological special types of breast cancer. *Breast Cancer Res Treat* 142:257–269
31. Ciriello G, Gatza ML, Beck AH et al (2015) Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 163:506–519
32. Desmedt C, Zoppoli G, Gundem G et al (2016) Genomic characterization of primary invasive lobular breast cancer. *J Clin Oncol* 34:1872–1881
33. Yates LR, Gerstung M, Knappskog S et al (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 21:751–759
34. Abba MC, Gong T, Lu Y et al (2015) A molecular portrait of high-grade ductal carcinoma in situ. *Cancer Res* 75:3980–3990
35. Kim SY, Jung SH, Kim MS et al (2015) Genomic differences between pure ductal carcinoma in situ and synchronous ductal carcinoma in situ with invasive breast cancer. *Oncotarget* 6:7597–7607
36. Pang JB, Savas P, Fellowes AP et al (2017) Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. *Mod Pathol* 30:952–963
37. Hernandez L, Wilkerson PM, Lambros MB et al (2012) Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intratumour genetic heterogeneity and clonal selection. *J Pathol* 227:42–52
38. Liao S, Desouki MM, Gaile DP et al (2012) Differential copy number aberrations in novel candidate genes associated with progression from in situ to invasive ductal carcinoma of the breast. *Genes Chromosomes Cancer* 51:1067–1078
39. Hwang ES, Lal A, Chen YY et al (2011) Genomic alterations and phenotype of large compared to small high-grade ductal carcinoma in situ. *Hum Pathol* 42:1467–1475
40. Shah V, Nowinski S, Levi D et al (2017) PIK3CA mutations are common in lobular carcinoma in situ, but are not a biomarker of progression. *Breast Cancer Res* 19:7
41. Sakr RA, Weigelt B, Chandraratnam S et al (2014) PI3 K pathway activation in high-grade ductal carcinoma in situ—implications for progression to invasive breast carcinoma. *Clin Cancer Res* 20:2326–2337
42. Zhao JJ, Liu Z, Wang L et al (2005) The oncogenic properties of mutant p110alpha and p110beta phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc Natl Acad Sci USA* 102:18443–18448
43. Zardavas D, Phillips WA, Loi S (2014) PIK3CA mutations in breast cancer: reconciling findings from preclinical and clinical data. *Breast Cancer Res* 16:201
44. Fry EA, Taneja P, Inoue K (2017) Oncogenic and tumor-suppressive mouse models for breast cancer engaging HER2/neu. *Int J Cancer* 140:495–503
45. Ang DC, Warrick AL, Shilling A et al (2014) Frequent phosphatidylinositol-3-kinase mutations in proliferative breast lesions. *Mod Pathol* 27:740–750
46. Allred DC, Clark GM, Molina R et al (1992) Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. *Hum Pathol* 23:974–979
47. Latta EK, Tjan S, Parkes RK et al (2002) The role of HER2/neu overexpression/amplification in the progression of ductal carcinoma in situ to invasive carcinoma of the breast. *Mod Pathol* 15:1318–1325
48. Jang M, Kim E, Choi Y et al (2012) FGFR1 is amplified during the progression of in situ to invasive breast carcinoma. *Breast Cancer Res* 14:R115
49. Park K, Han S, Kim HJ et al (2006) HER2 status in pure ductal carcinoma in situ and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence in situ hybridization and immunohistochemistry. *Histopathology* 48:702–707

50. Borgquist S, Zhou W, Jirstrom K et al (2015) The prognostic role of HER2 expression in ductal breast carcinoma in situ (DCIS); a population-based cohort study. *BMC Cancer* 15:468
51. Hoque A, Sneige N, Sahin AA et al (2002) Her-2/neu gene amplification in ductal carcinoma in situ of the breast. *Cancer Epidemiol Biomark Prev* 11:587–590
52. Miron A, Varadi M, Carrasco D et al (2010) PIK3CA mutations in in situ and invasive breast carcinomas. *Cancer Res* 70:5674–5678
53. Dunlap J, Le C, Shukla A et al (2010) Phosphatidylinositol-3-kinase and AKT1 mutations occur early in breast carcinoma. *Breast Cancer Res Treat* 120:409–418
54. Gorringer KL, Hunter SM, Pang JM et al (2015) Copy number analysis of ductal carcinoma in situ with and without recurrence. *Mod Pathol* 28:1174–1184
55. Bhat-Nakshatri P, Goswami CP, Badve S et al (2016) Molecular insights of pathways resulting from two common PIK3CA mutations in breast cancer. *Cancer Res* 76:3989–4001
56. Dogruluk T, Tsang YH, Espitia M et al (2015) Identification of variant-specific functions of PIK3CA by rapid phenotyping of rare mutations. *Cancer Res* 75:5341–5354

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