Ductal carcinoma in situ of the breast : cancer precursor or not?

Visser, L.L.

Lindy L. Visser
Marlous Hoogstraat
Tycho Bismeijer
Frank Nieboer
Michiel de Maaker
Petra Kristel
Lennart Mulder
Marielle Kreté
Wim Brugman
Michael Schaapveld
Marjanka K. Schmidt
Elinor Sawyer
Alastair Thompson
Andrew Futreal
Helen Davies
Serena Nik-Zainal
Lodewyk F. A. Wessels
Esther H. Lips*
Jelle Wesseling*
On behalf of the PRECISION team

Manuscript in preparation

*: Shared last author

CHAPTER 6

Approximately 30% of invasive ipsilateral breast cancers after treatment of ductal carcinoma in situ may be new independent primary cancers

Abstract

Ductal carcinoma *in situ* (DCIS) is regarded as a precursor of invasive breast cancer (IBC). Yet, not all DCIS progresses into IBC and proof of direct progression of DCIS to IBC is lacking. Therefore, we assessed the clonal relatedness between the initial DCIS and subsequent ipsilateral invasive breast cancer (IBC) by combined analysis of somatic copy number changes and mutations. This revealed that a substantial percentage (29.5%; i.e. 23 out of 78) of subsequent ipsilateral IBC are likely not clonally related to the preceding DCIS and therefore could be new primary tumors. This challenges the dogma that almost all subsequent ipsilateral IBC represent outgrowth of the initial DCIS lesion. If true, DCIS might not just be an IBC precursor, but could be an "independent" risk factor for developing IBC as well. This will substantially impact current clinical DCIS risk estimates, if validated by ongoing confirmatory analyses.

Main

Ductal carcinoma *in situ* (DCIS) is a non-invasive breast lesion with the potential to progress to invasive breast cancer (IBC) within a time period varying from a few years to decades.[1] Only a subset of DCIS will undergo the transition from *in situ* to invasive cancer, and not all invasive cancers arise from DCIS.[2,3] Factors mediating the progression of DCIS to IBC are still largely unknown. Before the introduction of population breast cancer screening, DCIS was rarely diagnosed, but nowadays it represents about 20-25% of all newly diagnosed breast cancers.[4–6]

Although DCIS is considered a precursor of IBC, definite proof of its progression to IBC is still lacking. The fact that most DCIS is treated, hinders studies on the natural course of DCIS. Treatment of DCIS consists of breast conserving surgery (BCS), often followed by radiotherapy and/or endocrine therapy, or even mastectomy. After breast conserving treatment, recurrent disease may occur. The subsequent ipsilateral IBC could be progression of residual DCIS that was left behind after BCS, or may be a new primary tumor that is unrelated to the preceding DCIS.

Traditionally, assessment whether a subsequent cancer is related to the preceding one, is primarily based on comparison of clinical and histopathologic information, i.e time interval, location, morphology, tumor grade, and immunohistochemical marker expression.[7–11] In recent years, genetic markers have started to play an increasingly important role in assessing the clonal relatedness of two lesions.[12–16]

In the present study, we evaluated the clonal relatedness of 78 patient-matched DCIS and subsequent ipsilateral IBC. This will help us to provide evidence of direct progression from DCIS to IBC. The most ideal patient group for this study would have been a group of women with untreated DCIS that subsequently developed ipsilateral IBC. However, this is unethical and almost all women with DCIS are treated. Therefore, we investigated women with DCIS and subsequent ipsilateral IBC, who were treated for DCIS with BCS alone. This patient group was part of a nation-wide population-based cohort including all women diagnosed with and treated for DCIS with BCS alone in the Netherlands between 1989 and 2005, with a median follow-up time of 12.0 years.[17] Formalin-fixed paraffin-embedded (FFPE) tissue specimens of patient-matched DCIS and subsequent ipsilateral IBC were retrieved from across the Netherlands, and pathology review and molecular profiling was performed. Patterns of shared and unique somatic copy number alterations and mutations were evaluated between the patient-matched DCIS and subsequent IBC. Pairs classified as having a shared clonal origin, showing one or more shared somatic events and only few to none sample-specific somatic events, were considered true recurrences. Pairs classified as having independent origin, showing only few to none shared somatic events and one or more sample-specific events, were considered new primary tumors.

Table 1. Overview of clinical and histopathologic characteristics of the primary DCIS and matched subsequent invasive breast cancer

	Primary DCIS							Time to event (years)	Same quadrant?	Invasive breast cancer						
Patient ID	Age (years)	Lesion size (mm)	Margin status	Histologic grade	ER	PR	HER2			Histologic subtype	Stage	Grade	ER	PR	HER2	Discordance
9698	61.9	20	free	poor	+	+	+	0.5	unknown	NST	1	3	+	+	+	
6325	60.2	NA	unknown	intermediate	+	+	-	0.5	yes	mixed	2B	2	+	+	-	
3630	48.4	NA	free	intermediate	+	+	-	0.7	no	NST	3A	1	+	+	-	location, ER
8417	37.7	NA	involved	intermediate	+	+	+	1.0	no	NST	2A	2	+	-	-	location, PR, HER2
1569	68.9	NA	free	poor	NA	NA	NA	1.0	no	NST	2A	3	-	-	-	location
4863	58.4	NA	free	poor	-	-	+	1.1	no	NST	1	3	-	-	+	location
10048	52	14	free	intermediate	-	-	-	1.1	unknown	NST	4	2	-	-	-	
4590	72.4	20	unknown	intermediate	+	+	-	1.4	yes	NST	1	2	+	+	-	
6280	55.9	NA	free	poor	+	+	-	1.9	no	NST	1	2	+	+	-	location, grade
9663	51.5	NA	free	poor	+	-	-	1.9	unknown	NST	3A	3	+	+	-	PR
1800	66.2	NA	free	poor	-	-	+	1.9	unknown	NST	1	2	-	-	-	grade, HER2
7114	73.7	NA	free	intermediate	+	+	-	1.9	yes	NST	1	3	+	+	-	grade
971	56	11	free	intermediate	+	-	-	2.1	yes	NST	1	1	+	+	-	grade, PR
9371	50.4	NA	free	intermediate	-	-	+	2.1	unknown	NST	1	2	-	-	+	
8088	60.1	NA	unknown	intermediate	-	-	+	2.3	unknown	NST	1	2	-	-	-	HER2
6929	43.8	7	free	intermediate	+	+	-	2.3	unknown	NST	2B	3	+	+	-	grade
783	66.4	NA	free	intermediate	+	+	-	2.4	yes	NST	2A	2	+	+	-	
384	39.3	8	unknown	intermediate	+	+	-	2.8	yes	NST	2B	2	+	-	-	PR
6921	49.6	15	free	poor	+	-	+	2.9	yes	NST	1	3	+	-	+	
2010	55.5	NA	free	well	+	+	-	3.0	no	NST	2A	3	+	-	-	location, grade, PR
6890	53.8	NA	unknown	intermediate	+	+	-	3.0	unknown	NST	2A	2	+	+	-	
2002	66.9	NA	free	poor	-	-	+	3.0	yes	NST	1	3	-	-	+	
293	58.8	20	free	poor	+	+	-	3.1	yes	NST	2A	1	+	+	-	grade
88	74.7	NA	free	intermediate	+	+	-	3.5	yes	NST	2A	2	+	+	-	
3507	62.6	30	involved	intermediate	-	-	+	3.7	yes	NST	1	1	-	-	+	grade
5736	68.3	25	free	intermediate	+	+	-	4.2	yes	NST	1	2	+	+	-	
7323	56.6	24	free	well	+	+	-	4.2	yes	NST	2A	2	+	+	-	grade
9355	67.5	NA	unknown	well	+	-	-	4.2	yes	NST	1	2	+	+	-	grade, PR
7199	57.9	NA	free	intermediate	+	+	-	4.3	unknown	NST	3B	1	+	+	-	grade

Table 1. Overview of clinical and histopathologic characteristics of the primary DCIS and matched subsequent invasive breast cancer

	Primary DCIS									Invasive breast cancer						
Patient ID	Age (years)	Lesion size (mm)	Margin status	Histologic grade	ER	PR	HER2	Time to event (years)	Same quadrant?	Histologic subtype	Stage	Grade	ER	PR	HER2	Discordance
6919	46.4	NA	unknown	poor	+	+	+	4.6	yes	NST	2A	3	+	-	+	PR
8126	58.5	NA	free	poor	-	-	+	4.6	yes	NST	1	3	+	-	-	ER, HER2
7097	47.8	NA	involved	intermediate	+	+	-	5.0	unknown	mixed	2B	2	+	+	-	
3956	47.1	5	free	well	+	-	-	5.0	yes	mixed	2A	2	+	+	-	grade, PR
5230	46.4	NA	free	intermediate	+	+	-	5.1	unknown	NST	NA	2	+	+	-	
9413	51.7	NA	free	intermediate	+	-	+	5.1	yes	NST	1	3	+	-	+	grade
9253	56	NA	unknown	well	+	+	-	5.3	yes	mixed	2A	3	+	+	-	grade
3275	53	15	free	well	+	+	-	5.3	yes	NST	1	1	+	+	-	
455	86.7	NA	free	intermediate	+	+	-	5.4	yes	NST	2B	2	+	+	-	
4879	59.7	NA	free	intermediate	+	+	+	5.5	yes	NST	1	3	-	-	-	grade, ER, PR, HER2
7831	66.6	NA	free	poor	-	-	+	5.5	no	NST	4	3	-	-	+	location
8759	33.9	NA	free	intermediate	-	-	+	6.1	unknown	NST	1	3	+	+	+	grade, ER, PR
4965	58	NA	involved	poor	+	+	+	6.3	yes	lobular	4	3	+	-	+	PR
3800	63.9	NA	unknown	intermediate	+	+	-	6.3	yes	NST	2B	2	+	-	-	PR
3386	86.5	NA	unknown	intermediate	-	-	-	6.4	yes	NST	2A	3	+	-	-	grade, ER
1678	69	8	involved	intermediate	+	+	-	6.6	yes	NST	1	2	+	+	-	
9076	65.5	NA	involved	intermediate	+	+	-	6.7	yes	mixed	1	1	+	+	-	grade
2364	53	NA	free	poor	+	-	-	6.9	yes	NST	1	3	+	+	+	PR, HER2
5613	34.5	22	free	well	+	+	+	7.0	unknown	NST	3A	2	+	-	+	grade, PR
567	53.8	NA	involved	well	+	+	-	7.1	yes	NST	4	2	-	-	+	grade, ER, PR, HER2
8300	49.7	10	free	intermediate	+	+	-	7.1	yes	NST	3A	2	+	+	-	
7993	58.1	NA	free	intermediate	+	-	+	7.3	unknown	NST	1	3	+	+	+	grade, PR
809	51.6	10	free	intermediate	+	-	+	7.3	yes	NST	3A	3	+	-	+	grade
471	59.6	17	free	poor	-	-	+	7.4	yes	NST	3A	2	+	+	-	grade, ER, PR, HER2
8206	58.5	15	free	intermediate	+	+	-	7.5	yes	NST	1	3	+	+	-	grade
4901	56.6	10	free	intermediate	+	NA	-	8.0	yes	mixed	1	3	+	+	-	grade
216	51.2	NA	involved	intermediate	+	-	-	8.0	unknown	NST	1	2	+	-	-	
7829	65.3	NA	free	intermediate	+	-	+	8.1	unknown	NST	1	2	+	-	+	

Table 1. Overview of clinical and histopathologic characteristics of the primary DCIS and matched subsequent invasive breast cancer

	Primary DCIS									Invasive breast cancer						
Patient ID	Age (years)	Lesion size (mm)	Margin status	Histologic grade	ER	PR	HER2	Time to event (years)	Same quadrant?	Histologic subtype	Stage	Grade	ER	PR	HER2	Discordance
7784	80.7	NA	involved	intermediate	+	-	-	8.4	no	NST	1	3	+	-	-	location, grade
9196	57.1	NA	free	intermediate	+	-	-	8.4	yes	NST	2B	2	+	-	-	
820	70.3	15	free	well	+	+	-	8.7	yes	NST	1	1	+	+	-	
9337	50.7	NA	free	intermediate	+	+	-	8.7	unknown	NST	1	3	-	-	-	grade, ER, PR
264	58	15	free	intermediate	+	-	-	8.8	no	NST	1	3	-	-	+	location, grade, ER, HER2
5753	69.5	NA	free	intermediate	+	+	-	9.1	unknown	NST	1	3	+	+	-	grade
6737	43.8	12	involved	well	+	+	-	9.7	unknown	NST	2B	1	NA	NA	NA	
7815	73.3	8	free	intermediate	+	+	-	9.9	no	NST	2A	3	+	+	-	location, grade
9255	63.2	NA	unknown	intermediate	+	-	-	10.4	unknown	NST	2B	2	+	+	-	PR
9530	54.1	NA	free	intermediate	+	-	-	10.4	yes	NST	2A	2	+	-	-	
8168	55	NA	unknown	intermediate	+	+	-	10.8	unknown	NST	1	3	+	+	-	grade
277	50.1	NA	free	intermediate	+	+	-	10.9	unknown	NST	1	2	+	+	-	
3543	76.9	4	free	well	+	+	-	11.7	yes	NST	2A	2	+	+	-	grade
8210	45.7	2	free	intermediate	+	+	-	13.7	no	NST	2A	3	+	+	-	location, grade
6287	54.5	12	involved	intermediate	+	+	-	13.9	unknown	NST	1	2	+	+	-	
9117	38.2	NA	unknown	poor	NA	NA	NA	13.9	yes	NST	2A	3	-	-	-	
9202	71.1	NA	free	intermediate	-	-	+	14.4	no	NST	1	2	+	-	+	location, ER
2553	42.9	NA	unknown	well	+	+	-	15.3	unknown	mixed	4	2	+	+	-	grade
2802	37.5	NA	unknown	well	+	+	-	15.5	yes	NST	3A	2	+	-	-	grade, PR
9127	60.6	NA	unknown	poor	NA	NA	NA	15.7	unknown	NST	2A	3	-	-	+	
10018	44.7	NA	unknown	intermediate	+	+	-	17.0	unknown	NST	2A	3	+	-	-	grade, PR

NA: not assessable; NST: invasive carcinoma of no special type.

Results

Patient characteristics

We made use of a nation-wide population-based cohort including all women diagnosed and treated for DCIS in the Netherlands between 1989 and 2005. Details on this cohort are described before.[17] In brief, clinicopathological information was obtained from the Netherlands Cancer Registry (NCR) and the nation-wide histopathology and cytopathology data network (PALGA). The cohort comprises 10,090 women of which 2,658 were treated with BCS alone. Patients did not receive tamoxifen or other endocrine adjuvant treatment.

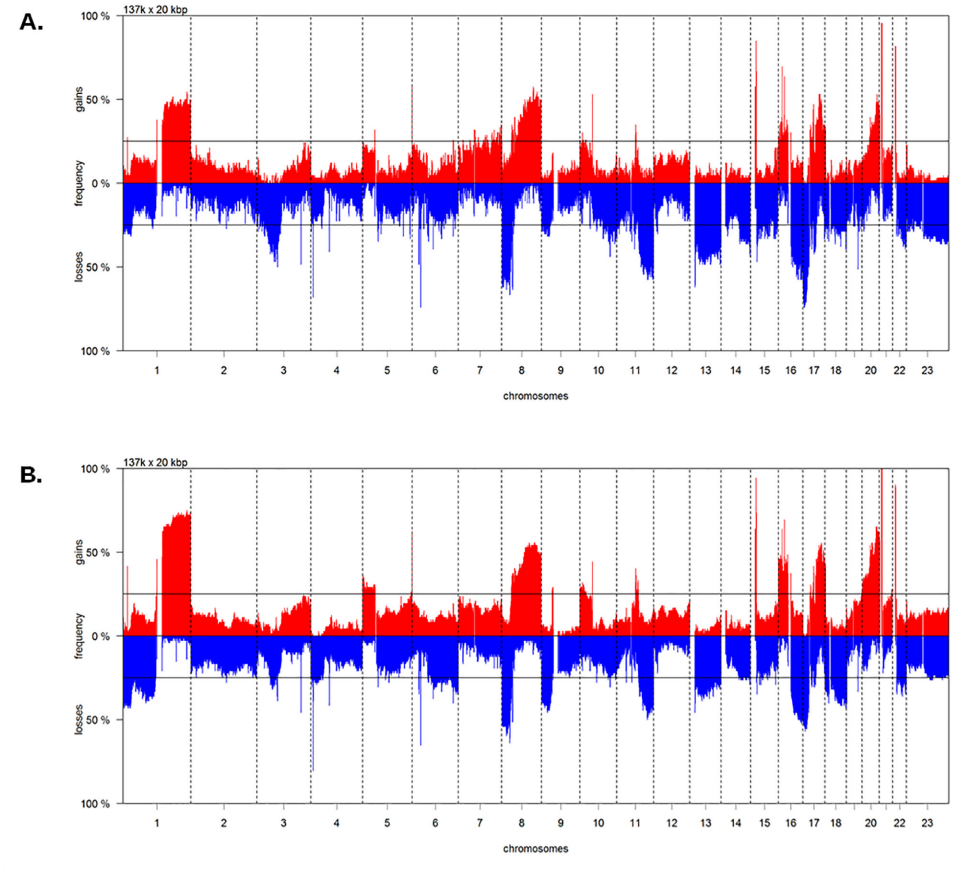

Figure 1. Genome-wide frequency plots of DNA copy number gains (red) and losses (blue) of DCIS and IBC samples separately. Both homozygous and heterozygous copy number events are present ranging from very focal to whole arm events. (**A**) Frequency plot of 66 DCIS samples and (**B**) frequency plot of 72 IBC samples. Noisy copy number profiles were excluded.

Fifteen years after DCIS diagnosis, 374 ipsilateral IBC events were observed after DCIS treatment with BCS alone (cumulative incidence of ipsilateral IBC of 15.4%). FFPE tissue specimens of patient-matched DCIS and subsequent IBC were successfully retrieved and pathologically re-examined of 155 women with an ipsilateral IBC event.[18] DNA and RNA of sufficient yield and quality was successfully extracted from both DCIS and IBC of 78 women.

Clinical characteristics of the 78 patients are shown in Table 1. Mean time to IBC event was 6.3 years (range 0.5-17.0). Mean age at DCIS diagnosis was 57.5 years (range 33.9-86-7). Histological grade of DCIS was well differentiated in 16.7% (13/78), intermediately differentiated in 61.5% (48/78), and poorly differentiated in 21.8% (17/78) of the cases. 79.5% of DCIS was ER positive and 29.2% was HER2 positive.

Discordances in histopathological characteristics

Histopathological features were compared between patient-matched DCIS and IBC, including tumor location, histological grade, and the expression of ER, PR, and HER2 (Table 1). 65.4% of the patients (51/78) showed discordances between the DCIS and IBC lesion, with 25 of the 78 patients (32.1%) showing two or more discordant changes. Most changes were found when comparing grade (47.4%; 37/78 patients), while tumor location differed in 23.1% (12/52 patients; for 27 patients tumor location of DCIS and/or IBC was unknown), ER status in 12.2% (9/74 patients), PR status in 28.8% (21/73 patients), and HER2 status in 12.2% (9/74 patients).

Copy number alterations and mutations

DNA copy number analysis using low coverage whole genome sequencing (CNVseq) was performed for all DCIS and IBC pairs to identify regions of DNA copy number gain (red) and loss (blue) that occurred in at least 25% of the DCIS and IBC lesions in our study group. In both DCIS and IBC lesions, frequent DNA gains were identified on chromosomes 1q, 8q, 16p, 17q, and 20q, while DNA losses were detected on 1p, 3p, 8p, 11q, and 16q (Figure 1). These results are in line with DNA gains and losses frequently identified in breast cancer.[19] Overall, although the frequencies of allelic changes are slightly lower in the *in situ* lesions, the general pattern of broad genetic instability is similar.

Ultra-deep sequencing (average >2000x) was performed to assess mutations in breast cancer driver genes, using a custom 53-gene panel. PIK3CA, TP53, NF1, BRCA2, and GATA3 were most frequently mutated in DCIS in 33.3%, 28.7%, 9.3%, 9.3% and 8.3% of the lesion, respectively. In the IBCs, we found TP53, PIK3CA, BRCA2, GATA3, and KMT2C to be the most frequently mutated genes, as they were affected in 31.7%, 27.0%, 11.1%, 9.5%, and 9.5% of the lesions, respectively.
and events unique to the DCIS or IBC lesion.

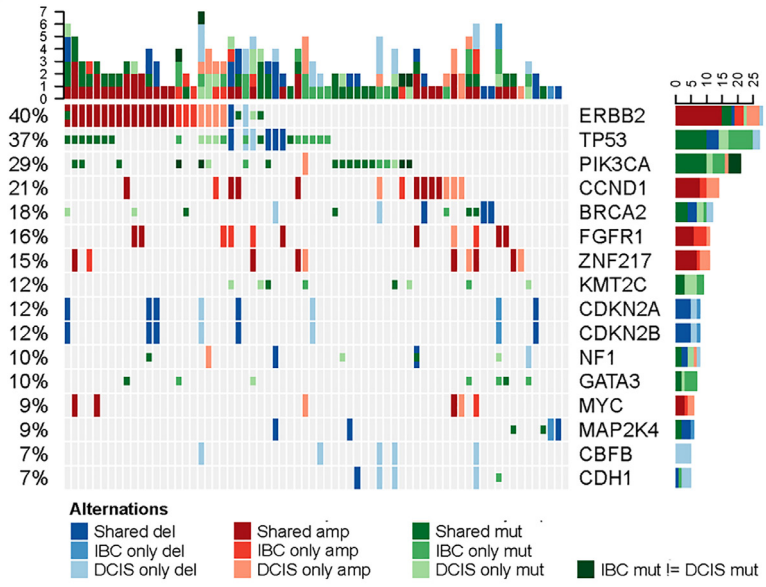

Figure 2. OncoPrint of 68 patient-matched DCIS and subsequent iIBC. Ten patients did not harbor any somatic events in the 53 assessed genes. IBC mut != DCIS mut, gene mutation present in both lesions but different mutation.

In Figure 2, an overview is given of the most frequent somatic events found in patient-matched DCIS and IBC (for full list of genes see Supplementary Figure S1 and Supplementary File 1). A distinction was made between somatic events shared by the DCIS and IBC lesion

Assessment of clonal relatedness

We determined if the patient-matched DCIS and IBC pairs had a shared clonal origin or represented two independent lesions, by evaluating patterns of shared and unique somatic copy number alterations and mutations (representative example shown in Figure 3). For this, we made use of two independent methods. First, we used the Clonality R package which is based on genome-wide copy number data. A cutoff for clonal origin and independence was set on $p<0.1$ as suggested by the developers of the method. Second, we developed a panel-based clonality score making use of CNA and mutation data, restricted to our panel of 53 driver genes. The score consists of the sum of shared aberrations divided by the sum of and unique aberrations for every patient, corrected for the probability of observing these aberrations in breast cancer. Both true pairs (test set) and artificial pairs (independent pairs; reference set) were analyzed using the two methods. The artificial pairs are obtained by constructing pairs of lesion from different patients. Since these pairs are by definition

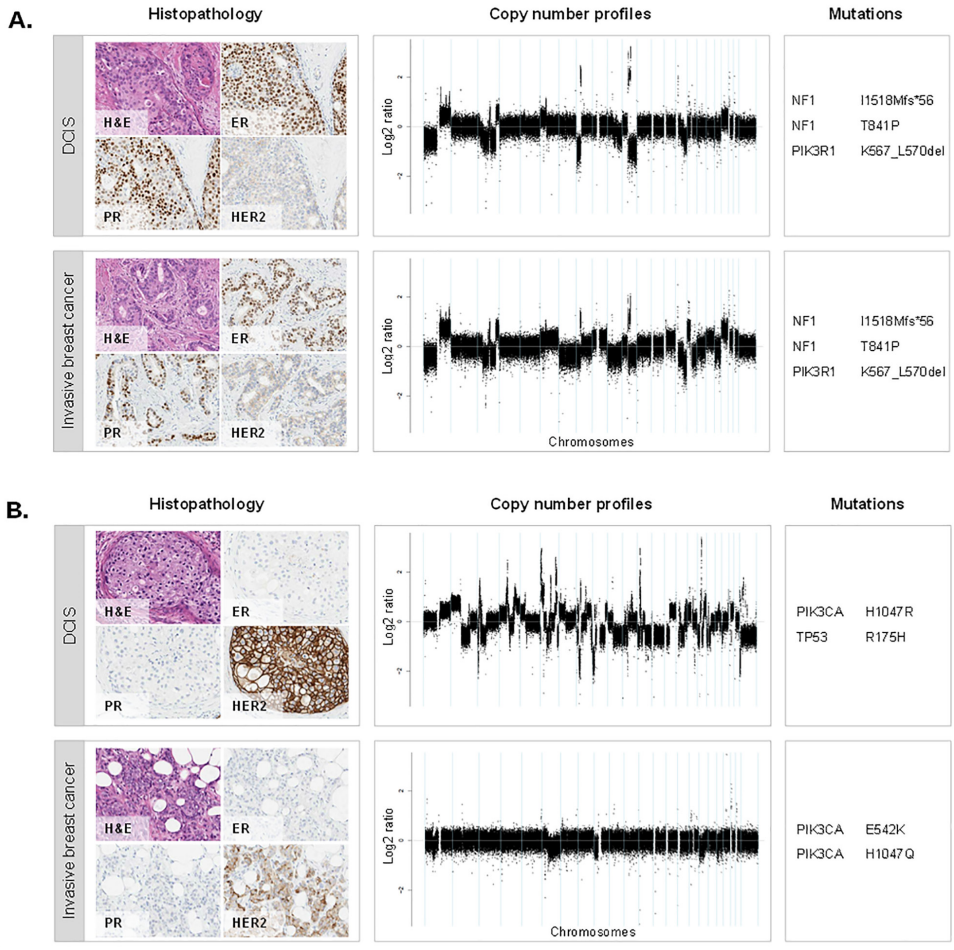

Figure 3. Example of a clonal (**A**) and independent (**B**) DCIS IBC pair

independent they provide the reference distribution, which is used to permit calculation of the *p* values of the real paring. A cutoff for clonal origin and independence was set on the 99th percentile of the artificial pairs' scores. Pairs classified as "independent" based on the Clonality package and with a low panel-based clonality score were considered new primary tumors. Pairs classified as "clonal" and with a high panel-based clonality score were considered true recurrences.

Based solely on copy number data, 45 of 78 pairs were classified as clonally related and 33 of 78 suspected of having an independent origin (Clonality package; Figure 4; Supplementary Figure S2; Supplementary File 2). Using our panel-based clonality score, 43 pairs were classified as clonal and 25 were likely to be independent (n=10 no somatic events found;

Supplementary Figure S3). The agreement between the two methods was 79% (54 of 68, with 10 unknowns; Supplementary Figure S4). 14 of 68 pairs showed discordant classification based on clonality package and panel-based clonality score. With this, 41 IBCs (52.6%) were classified as "true recurrences" and 23 IBCs (29.5%) were likely to be "new primary tumors". For 14 pairs (17.9%) our method was inconclusive and were classified as "equivocal" (Figure 4).

Next, we assessed the clinicopathological characteristics of true recurrences and new primary tumors. Discordant grade was more often observed among new primary tumors (low (1-2) and high (3) grade; p=0.049; Figure 4; Supplementary Table S1). Discordant ER status (p=0.021) was higher among new primary tumors as compared to true recurrences. Furthermore, new primary IBCs trended to originate more often from HER2-positive DCIS index lesions (p=0.054) and were more often of HER2-negative type (p=0.009). Interestingly, involved margins (p=0.70), size of the DCIS index lesion (p=0.32), and discordant tumor location (p=0.72) are not associated with a higher rate of new primary tumors. No difference was found when comparing IBC stage between true recurrences and new primary IBCs (p=0.52).

Calculating the proportion of new primary tumors after DCIS based on epidemiologic data

Lastly, we estimated the chance for a woman with DCIS to develop a new primary tumor, using epidemiologic data. From a previous study by our group, making use of the same patient cohort, we know that: 1) fifteen years after DCIS diagnosis, cumulative incidence of ipsilateral IBC was 15.4% after BCS alone; and 2) that the cumulative incidence of contralateral invasive breast cancer (cIBC) at 15 years was 6.4% and did not differ by treatment, period of diagnosis, or age group.[17] We assumed that the incidence of new primary tumors is similar to the incidence of cIBC. This led us to the following proportions of true recurrences and new primary tumor: 1) after DCIS treatment with BCS alone, 41.6% of the subsequent ipsilateral IBCs (6.4% of 15.4%) are new primary tumors and the remaining 58.4% are true recurrences, clonally related to the initial DCIS index lesion. These numbers support our estimate that 29.5% of subsequent ipsilateral IBCs might very well be new primary tumors based on the molecular characterization.

Discussion

This study included a unique series of 78 patient-matched DCIS and subsequent ipsilateral IBC, that occurred during a median follow-up time of 12 years. Clonality analysis based on genomic aberrations showed that 52.6% of the pairs (41/78) are clonally related. This provides evidence for direct progression from *in situ* to invasive breast disease. Strikingly, our results also showed that 29.5% of the subsequent IBCs (23/78) may be new primary tumors. This high rate of new primary tumors after DCIS treated by BCS alone indicates that: 1) *true* recurrence rates after DCIS are likely to

be overestimated; and, if true, 2) DCIS might not just be a precursor, but could also be risk lesion of IBC, as has been estimated

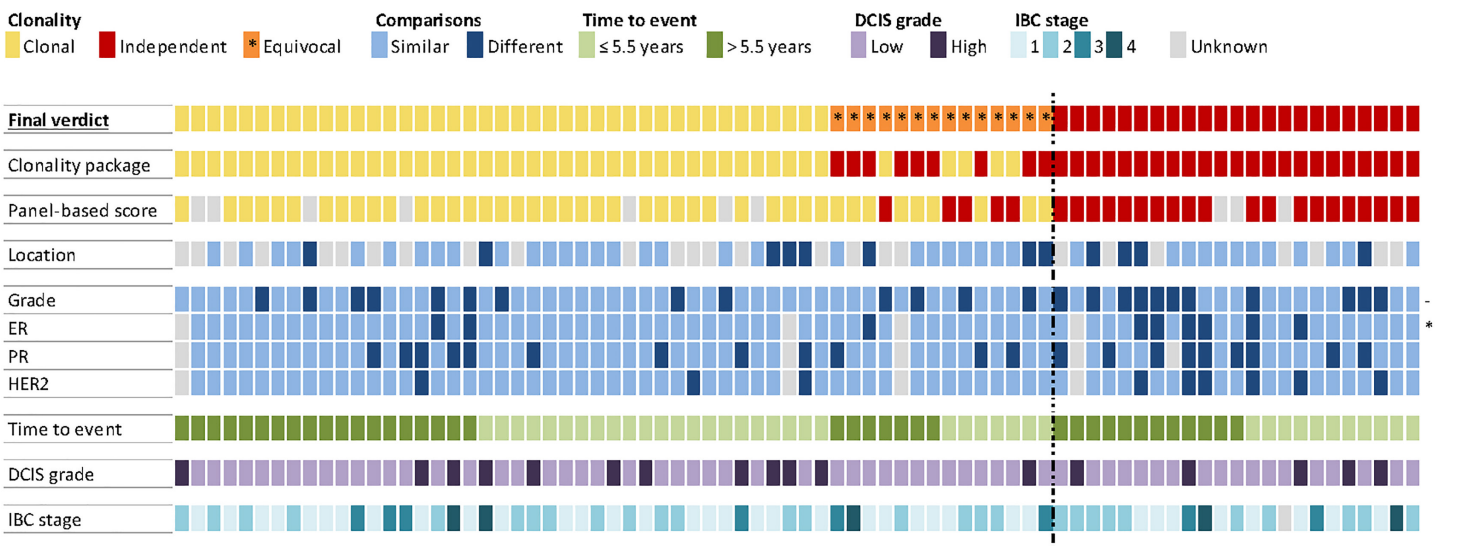

Figure 4. Overview figure of all datatypes available in this study. Discordant ER status (*) was higher among independent pairs compared to clonal pairs (p=0.021). Discordant grade (-) trended to be more often observed among independent pairs (low (1-2) vs. high (3) grade; p=0.054).

for lobular carcinoma in situ (LCIS).

These findings may impact personalized risk stratification for women with DCIS substantially, as two different risk issues have to be addressed. First, the risk of the DCIS lesion progressing to IBC. Second, the risk to develop a non-clonally related, most likely independent new IBC. As a consequence, prognostic factors studies so far might have an intrinsic limitation, as they focus on DCIS features predicting the risk of developing subsequent IBC. However, it is far less likely that the risk of a subsequent independent, new primary IBC is dictated by the initial DCIS lesion.

Currently, an ipsilateral tumor after BCS is diagnosed as a local recurrence, as it is considered outgrowth of residual cancer cells outside the surgical margins. However, studies on tumor adjacent histologically normal tissue raise the possibility that non-malignant, pre-cancerous cells within these tissues can also contribute to subsequent ipsilateral IBC.[20–24] This hypothesis is further acknowledged by two concepts which have been recognized to underlie the development of breast cancer: the sick lobe theory and the field cancerization concept.[25–29] The sick lobe theory postulates that breast carcinoma, both *in situ* and invasive, are lobular diseases, meaning that synchronous and metachronous lesions appear in a single lobe of the breast. Both concepts imply that the epithelial lining within a lobe can consists of cells which have undergone early genetic events (first "hit"), either during mammary development (sick lobe theory) or at unspecified time points (field cancerization concept), but have not acquired all the changes (second "hit") necessary for tumorigenesis. This breast lobe constitutes the field cancerized tissue or "sick lobe", and is hypothesized to be the mediator of disease progression.

An alternative explanation for finding such a high rate of new primary tumors is tumor heterogeneity. It could be that not all cells in a DCIS lesion harbor the same genetic aberrations. When analyzing bulk samples, this intra-lesional heterogeneity could be missed. The issue of intra-DCIS heterogeneity complicates clonality analysis and makes that we cannot rule out that new primary tumors originate from the index lesion.

Previous studies have attempted to distinguish NP from TR after IBC treated with BCS, with or without RT, and have reported different proportions of patients with NP, ranging from 38-52%.[7–9,11,30,31] One other study reported a NP rate of 15%.[10] All the studies reported that TRs occur sooner than NPs. Furthermore, some studies showed that TRs have worse survival rates,[7,9,10,30,31] and were associated with involved margins.[10,30] Patients with NPs more often developed cIBC.[7,9,10] Our results on DCIS do not confirm these associations.

There is no consensus about which type of data (e.g. copy number, mutation, histology) and especially which statistical method is most suitable to distinguish clonal recurrences from independent primary tumors.[16] Many studies have been published on tumor clonality using different data types and statistical methods,[12,13,16,32–34] whereas other studies defined their own methods and cohort-specific cut-offs.[14,15] Yet, studies comparing clonality assessment based on clinical and molecular data have shown that the clinical assessment

of clonality is inaccurate for most ipsilateral breast tumor recurrences.[16,35,36] Furthermore, Bierman and colleagues have shown that the type of molecular data analysis had a stronger impact on clonality determination than the analytical methods used.[16] Data suggests that mutations evolve gradually over long periods of time, generating extensive clonal diversity.[37] In contrast, CNAs are acquired at early stages of tumorigenesis, which makes them the most stable type of biological data you can used for clonality assessment, but depending on the methodology, it is difficult to identify CNAs present in a (small) subset of cells in a sample.[37,38]

For the DCIS and IBC pairs in our cohort, both copy number and mutation data were available to evaluate clonal relatedness of the lesions. The results from these two methods were mostly concordant. However, in several cases, our panel-based clonality score demonstrated evidence of clonality, while the clonality package did not. This may be due to higher noise levels in the data of these pairs. In case of discordance, pair were classified as equivocal.

Our study group comprised of patients with DCIS diagnosed between 1989 and 2004, which were treated with BCS alone. Regarding this time period, we have to keep in mind that treatment strategies and screening techniques for DCIS have evolved over the years. This may have impacted treatment or other care for these patients and with this possibly also the rates of new primary tumors. Also, adjuvant radiotherapy is nowadays standard of care after BCS for DCIS. Radiotherapy significantly reduce the risk of recurrence (both *in situ* and invasive) after DCIS.[39,40] Fifteen years after DCIS diagnosis, cumulative incidence of ipsilateral IBC was 15.4% after BCS alone and 8.8% after BCS followed by RT.[17] Based on our calculations using epidemiological data, we estimated that also the proportion of true recurrences is lower when radiotherapy is added to the treatment of women with DCIS as compared to women treated with BCS alone: 72.7% of the subsequent ipsilateral IBCs (6.4% of 8.8%) are new primary tumors and the remaining 27.3% are true recurrences. Lastly, it is important to note that we do not know what the recurrence rate is if the women in our cohort were left untreated.

Our study has several strengths. First, we made use of a unique, well-annotated set of 78 patient-matched DCIS and subsequent IBC pairs derived from a nation-wide population-based cohort of DCIS patients treated by BCS alone. Differences between DCIS and IBC could not be caused by treatment. Second, using microdissection we achieved a high neoplastic cell purity for our molecular studies. This improved the intensity of detected CNAs and small cell populations. With this we ruled out that a lack of clonal relatedness could be due to a lack of tumor cells. Third, we reached an average coverage >2000x for variant calling in the DCIS sample using PanelSeq. With this high coverage, mutations in potential DCIS subclones should have been picked up. In addition, the positions of variants solely called in the IBC lesions were checked in the matching DCIS lesion and vice versa, to reassure the absence of this variant in the paired sample. Fourth, we took along loci-specific probability

of CNA and mutation to correct for the frequency in which an aberration is found in breast cancer. Fifth, our epidemiological calculation based on same source population supports our experimental results.

Our study also has some limitations. First, with our panel-based clonality score we compared CNA and mutations based on a gene panel. This method covered only 53 genes and thus represent a part of the full picture. Whole exome sequencing would have given us more information, although with the cost of lower sequencing depth, but is not feasible for this cohort of old-FFPE derived DNA. Yet, the analysis using the clonality package did gave us a whole genome-based assessment. Second, due to the lack of normal DNA, we used germline databases (GNOMAD and GoNL) to remove common germline variants from our PanelSeq data. Because of this, we cannot fully exclude the possibility that germline mutations that are not recognized as such are called as somatic mutations. This is probably also why we detected such a high rate of BRCA2 mutations. Third, intra-lesional heterogeneity could have influence the outcome of our clonality analysis due to biological differences in different parts of a tumor and subclone evolution.

In conclusion, we provided proof of the direct progression of DCIS to IBC. Additionally, we found that many ipsilateral IBCs after BCS alone treated-DCIS are new primary tumors, challenging the current dogma that almost all subsequent ipsilateral IBC are due to progression of the initial DCIS. Furthermore, our findings will have major impact on assessing the risk of subsequent ipsilateral IBC after diagnosis and treatment of DCIS.

More research is needed to assess the role of DCIS heterogeneity in the outcome of our clonality analysis. Single cell or multi-regional sequencing could be used to explore the existence of minority DCIS subclones. Furthermore, better understanding of mammary field cancerization ("sick lobe theory") could improve informed decision-making in patient risk stratification and treatment. Ultimately, this may reveal markers that could be applied to identify and create a "map" of the affected lobe(s) within the breast and could be used complemented to margin assessment.[41] Yet, one has to bear in mind that choosing optimal resection margins required balancing the competing risks of morbidities associated with more radical surgery with the increased rapid recurrence from a cancerized field left *in situ*.[28] Lastly, we recommend validation of our results using an independent cohort of women treated for DCIS, or a group of women that did not receive any treatment for DCIS.

Article information

Author's contributions: LLV, MH, MS, MKS, LFAW, EHL, and JW designed the study. LLV, FN, MdM, PK, and LM set-up and performed DNA and RNA extractions. MK and WB set-up and performed library preparations and sequencing. LLV, MH, and TB analysed that data. LLV, MH, TB, MS, MKS, ES, AT, AF, HD, SNZ, LFAW, EHL, and JW interpreted the data. LLV wrote the report. All authors provided comments and approved the final version.

Funding: This work was jointly funded by Cancer Research UK and the Dutch Cancer Society (grant number C38317/A24043, to J. Wesseling).

Acknowledgements: We thank all patients of which stored residual material was used in this study, all collaborating hospitals and in particular the pathology departments, and all persons who helped in the process of data collection and analysis. We would like to specifically acknowledge Otto Visser, Annemarie Eeltink, and the registration teams of the Netherlands Comprehensive Cancer Organization for collection of the data for the Netherlands Cancer Registry. We thank Lucy Overbeek and PALGA, the nation-wide histopathology and cytopathology data network and archive, for providing pathology data and for their help in the collection of the residual patient material. We acknowledge the staff of the NKI-AVL Core Facility Molecular Pathology & Biobanking for their technical support and the staff of the NKI-AVL Genomics Core Facility and Ronald van Marion of the Erasmus University Medical Center for their sequencing support.

Supplementary data: Supplementary File 1-5 will be available after publication of the article.

Methods

Study population

We made use of a nation-wide population-based cohort including all women diagnosed and treated for DCIS in the Netherlands between 1989 and 2005, with a median follow-up time of 12.0 years. Details on this cohort are published elsewhere.[17] In brief, clinicopathological information was obtained from the Netherlands Cancer Registry (NCR) and the nation-wide histopathology and cytopathology data network (PALGA). The cohort included 10,090 women of which 2,658 were treated with BCS alone. Patients did not receive tamoxifen or other anti-hormonal adjuvant treatment. Fifteen years after DCIS diagnosis, 374 iIBC events were observed after DCIS treatment with BCS alone (cumulative incidence of iIBC of 15.4%). Formalin-fixed paraffin-embedded (FFPE) tissue specimens of patient-matched DCIS and subsequent iIBC were successfully retrieved and pathologically re-examined of 155 women.[18] DNA and RNA, of sufficient

yield and quality, was successfully extracted from both DCIS and IBC of 78 women. Data on tumor location (ICD-10) was available for n=52 pairs. DCIS lesion size (n=21) and margin status (n=61) was extracted from pathology reports. Immunohistochemical data on ER, PR, and HER2 was available for n=74, n=73, and n=74 pairs, respectively.[42]

The study was approved by the review board on the NKR (request K12.281; January 3, 2013) and PALGA (LZV990; April 16, 2013). Secondary use of tissue and data for this study was done under an opt-out regime which is conform Dutch regulations and the Code of Conduct of Federa-COREON.

Tissue processing and microdissection

For laser microdissection, 8-µm-thick FFPE tissue section were mounted on PolyEthylene Naphthalate (PEN) slides. Prior to sectioning, the slides were UV-treated for 30 minutes at 254 nm, to overcome the membrane's hydrophobic nature and to allow better section adherence. Sections were incubated overnight at 56°C. After deparaffinization and rehydration, FFPE sections were stained with 1% toluidin blue for 5 seconds to allow identification of epithelial areas of DCIS, followed by a washing step in distilled water and dehydration in serial ethanol dilutions (70, 90, 100%).

Microdissection of in situ lesions was performed using a laser microdissection (LMD) microscope (Leica Microsystems). Invasive lesions were microdissected from 10-µm-thick sections either using the LMD system or manually using a scalpel. The number of sections per case used for microdissection varied by lesions size and cellularity, with an average of 8 sections per in situ lesion (range 2-15) and an average of 9 sections per invasive lesion (range 5-10). The microdissected tissue fragments were collected in the cap of a 0.5ml tube containing 30µl PKD buffer, the first reagent of the extraction method. Tissue fragments were stored at 4°C for a maximum of two days until DNA and RNA extraction or were processed immediately.

DNA and RNA extraction

DNA and RNA were simultaneously extracted from (laser) microdissected tissue fragments using the Qiagen AllPrep DNA/RNA FFPE kit (QIAGEN), following manufacturer's instructions. DNA samples were RNase treated. DNA concentrations were measured using the Qubit dsDNA High-Sensitivity assay and Qubit 2.0 Fluorometer (Life Technologies). Quality and quantity of total RNA from FFPE tissue was assessed using a Bioanalyzer 2100 (Agilent) and the percentage RNA fragments >200bp (DV200) were determined.

RNA samples were submitted for RNA sequencing and DNA samples were submitted for low coverage whole genome sequencing and panel sequencing (PanelSeq). In cases with sufficient DNA, we performed both assays. In cases with insufficient DNA for both assays, only low coverage whole genome sequencing was performed. An overview of the available datasets per sample is presented in Supplementary File 3.

Whole transcriptome RNA sequencing (RNAseq)

Total RNA samples were converted to strand-specific libraries using TruSeq RNA Access Library Prep kit (Illumina, RS-301-2001/2), according manufacturer's instructions, protocol part #15049525 Rev. B. The libraries were sequenced with 65-bp single-reads using a HiSeq 2500 instrument (Illumina). With

this, an average sequence depth was reached of 20 million reads per sample. Reads were aligned to the reference genome GRCh38 (hg38), using BWA-MEM.[43]

Low coverage whole genome sequencing (CNVseq)

DNA was sheared to 160-200bp using the Covaris S220 Focused-ultrasonicator, cleaned with SPRI beads, analyzed for proper length on Perkin Elmer GX and library prepped with the KAPA hyper prep kit (KAPA Biosystems), protocol KR0961-v5.16. Agilent S5XT-2 (1-96) adapters with Illumina P5 and P7 sequences were used, containing 8bp Agilent indices. All samples received 9 PCR cycles, a SPRI beads cleanup and Perkin Elmer GX quality control.

Each sample was sequenced single-end 65bp on a HiSeq 2500 instrument (Illumina), one pool per lane, to obtain low coverage (0.2x) whole genome sequencing data. Reads were aligned to the reference genome GRCh38 (hg38) using BWA-MEM.[43] For each sample, per bin of 20kb, reads of a mapping quality of >37 were counted and read counts were rated against a similar mapping of all attainable 65bp sequences on the reference genome. Sample counts were corrected per bin for local GC effects using a non-linear loess fit of mappability over 0.8 on autosomes. Reference values were scaled according to the slope of a linear fit, forced to intercept at the origin, of reference mappability after GC correction. Bins with a mappability below 0.2 or overlapping ENCODE blacklisted regions were removed from further analysis.[44]

Copy number segmentation was performed using DNAcopy;[45] genome-wide copy number frequency plots were created using CGHcall.[46] Segments were identified as lost or gained at log2 ratios of <-1 and >1.5, respectively. To clean up the data, we removed losses in noisy samples (>1000 segments). Lost and gained segments were subsequently annotated with overlapping genes using BEDTools intersect.[47]

Panel sequencing

Samples were screened for mutations with the Ion Torrent Personal Genome Machine (PGM; Life Technologies), using a custom-made amplicon panel encompassing 2778 amplicons covering 53 genes including hotspot mutation regions and single nucleotide polymorphisms (SNPs). Genes included in this panel were selected based on a recent publication describing frequently found breast cancer drivers and by consulting the METABRIC and ICGC/TCGA breast cancer datasets.[19,48,49] Amplicons for the multiplex PCR assay were designed using the IonTorrent AmpliSeq Designer tool, aiming for 150-bp amplicons and allowing efficient amplification of fragmented DNA isolated from FFPE tissue specimens.

Samples analyzed using the IonTorrent AmpliSeq custom 53-gene panel were processed according to the Ion AmpliSeq Library Kit Plus protocol (ThermoFisher Scientific). The multiplexed PCR was split into two reactions of 10 ng of DNA, allowing the amplification and sequencing of overlapping amplicons, required to obtain full sequence coverage of large exons. Each sample was barcoded using IonTorrent Xpress bar-coded adapters, allowing multiplexed sequencing. A total of 19 PCR cycles were performed on the FFPE samples. Ten samples were multiplexed on an Ion 540 Chip and sequenced on the Ion S5XL Semiconductor sequencer. Reads were aligned to the reference genome GRCh37 (hg19) using the Torrent Mapping Alignment Program, and variant calling was performed using Torrent Variant Caller (TVC) version 5.6.

Variant data in VCF format was first translated to GRCh38 and annotated using bedtools, picard (https://broadinstitute.github.io/picard/command-line-overview.html), samtools, bcftools and VEP, and further analyzed in R, employing vcfR, and tidy verse.[47,50–52] True somatic variants, identified via filtering during which low quality variants (variant allele frequency [VAF] <10%, coverage <100x, and a quality [QUAL] of <1000), artifacts (found in >90% of samples), and germline variants (>5 cases in GNOMAD and GoNL; except for BRCA1, BRCA2, and TP53) were removed, and validated in RNAseq data using samtools mpileup.[50,53] Details regarding the amplicon panel design, performance, and filtering strategy are provided in Supplementary File 4-5.

Based on germline mutation data, we confirmed that patient-matched DCIS and IBC pairs indeed originated from the one patient.

Statistical analysis

We determined if the patient-matched DCIS and subsequent iIBC pairs had a shared clonal origin or represented two independent lesions, by evaluating patterns of shared and unique somatic copy number alterations and mutations. For this, two independent methods were used. Clonality package was used to define the likelihood ratio with individual comparisons (LR2) and LR2 *p*-value.[54] Using this package, pairs were classified as either clonal or independent, based on copy number data. Both true pairs (test set) and artificial pairs (independent pairs; reference set) were analyzed using the clonality package. The artificial pairs are obtained by constructing pairs of lesions from different patients. Since these pairs are definitely independent they provide the reference distribution, which is used to permit calculation of the *p* values of the real paring. A cutoff for clonal origin and independence was set on $p<0.1$. In addition, we developed a panel-based clonality score making used of CNA and mutation data. This score is restricted to our panel of 53 driver genes and was computed as follows: 1) loci-specific probabilities of observing a somatic mutation/CNA were obtained from the METABRIC breast cancer dataset;[19,49] 2) a score was computed based on the sum of shared and unique aberrations for every patient, corrected for the probability and squared to give less weight to highly common, and more weight to highly rare aberrations (e.g. PIK3CA mut 17% = (1-0.17)^2); 3) the following formula was applied: score = log(("shared-score" +1) / ("unique-score" +1)). The panel-based clonality score was computed for both true pairs (test set) and artificial pairs (reference set). The reference set included 80 extra DCIS samples without a matching IBC lesion. A cutoff for clonal origin and independence was set on the 99th percentile of the artificial pairs' scores.

Pairs classified as "independent" based on the Clonality package and with a low panel-based clonality score were considered new primary tumors (NP). Pairs classified as "clonal" and with a high panel-based clonality score were considered true recurrences (TR).

We used t-test and Fisher's exact test to ascertain differences in clinicopathological characteristics between TR and NP. *P* values ≤0.05 were considered statistically significant.

References

1. Barnes NLP, Ooi JL, Yarnold JR, Bundred NJ. Ductal carcinoma in situ of the breast. *BMJ*. 2012;344:e797.
2. To T, Wall C, Baines CJ, Miller AB. Is Carcinoma in Situ a Precursor Lesion of Invasive Breast Cancer? *Int J Cancer*. March 2014:1-7.
3. Levi F, Randimbison L, Te V-C, La Vecchia C. Invasive breast cancer following ductal and lobular carcinoma in situ of the breast. *Int J cancer*. 2005;116(5):820-823.
4. Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst*. 2002;94(20):1546-1554.
5. Netherlands Comperhencive Cancer Organization (IKNL). DCIS incidence trends over time, the Netherlands.
6. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975-2014*. National Cancer Institute. Bethesda, MD
7. Huang E, Buchholz TA, Meric F, et al. Classifying local disease recurrences after breast conservation therapy based on location and histology: new primary tumors have more favorable outcomes than true local disease recurrences. *Cancer*. 2002;95(10):2059-2067.
8. Freedman GM, Anderson PR, Hanlon AL, Eisenberg DF, Nicolaou N. Pattern of local recurrence after conservative surgery and whole-breast irradiation. *Int J Radiat Oncol Biol Phys*. 2005;61(5):1328-1336.
9. Nishimura S, Takahashi K, Akiyama F, et al. Classification of ipsilateral breast tumor recurrence after breast-conserving therapy: new primary cancer allows a good prognosis. *Breast Cancer*. 2005;12(2):112-117.
10. Komoike Y, Akiyama F, Iino Y, et al. Analysis of ipsilateral breast tumor recurrences after breast-conserving treatment based on the classification of true recurrences and new primary tumors. *Breast Cancer*. 2005;12(2):104-111.
11. Panet-Raymond V, Truong PT, McDonald RE, et al. True recurrence versus new primary: an analysis of ipsilateral breast tumor recurrences after breast-conserving therapy. *Int J Radiat Oncol Biol Phys*. 2011;81(2):409-417.
12. Banelli B, Casciano I, di Vinci A, et al. Pathological and molecular characteristics distinguishing contralateral metastatic from new primary breast cancer. *Ann Oncol*. 2009;21(6):1237-1242.
13. Begg CB, Ostrovnaya I, Carniello JVS, et al. Clonal relationships between lobular carcinoma in situ and other breast malignancies. *Breast Cancer Res*. 2016;18(1):1-11.
14. Bollet MA, Servant N, Neuvial P, et al. High-Resolution Mapping of DNA Breakpoints to Define True Recurrences Among Ipsilateral Breast Cancers. *JNCI J Natl Cancer Inst*. 2008;100(1):48-58.
15. Hwang ES, Nyante SJ, Chen YY, et al. Clonality of lobular carcinoma in situ and synchronous invasive lobular carcinoma. *Cancer*. 2004;100(12):2562-2572.
16. Biermann J, Parris TZ, Nemes S, et al. Clonal relatedness in tumour pairs of breast cancer patients. *Breast Cancer Res*. 2018;20(1):96.
17. Elshof LE, Schaapveld M, Schmidt MK, Rutgers EJ, van Leeuwen FE, Wesseling J. Subsequent risk of ipsilateral and contralateral invasive breast cancer after treatment for ductal carcinoma in situ: incidence and the effect of radiotherapy in a population-based cohort of 10,090 women. *Breast Cancer Res Treat*. 2016;159(3):553-563.
18. Visser LL, Elshof LE, van de Vijver K, et al. Discordant marker expression between invasive breast carcinoma and corresponding synchronous and preceding DCIS. *Am J Surg Pathol*. 2019.
19. Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403).
20. Li Z, Moore DH, Meng ZH, Ljung B-M, Gray JW, Dairkee SH. Increased risk of local recurrence is associated with allelic loss in normal lobules of breast cancer patients. *Cancer Res*. 2002;62(4):1000-1003.
21. Larson PS, De las Morenas A, Bennett SR, Cupples LA, Rosenberg CL. Loss of heterozygosity or allele imbalance in histologically normal breast epithelium is distinct from loss of heterozygosity or allele imbalance in co-existing carcinomas. *Am J Pathol*. 2002;161(1):283-290.

22. Heaphy CM, Bisoffi M, Fordyce CA, et al. Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors. *Int J cancer*. 2006;119(1):108-116.
23. Forsberg LA, Rasi C, Pekar G, et al. Signatures of post-zygotic structural genetic aberrations in the cells of histologically normal breast tissue that can predispose to sporadic breast cancer. *Genome Res*. 2015;25(10):1521-1535.
24. Takeshima H, Ushijima T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *npj Precis Oncol*. 2019;3(1):7.
25. Tot T. DCIS, cytokeratins, and the theory of the sick lobe. *Virchows Arch*. 2005;447(1):1-8.
26. Heaphy CM, Griffith JK, Bisoffi M. Mammary field cancerization: molecular evidence and clinical importance. *Breast Cancer Res Treat*. 2009;118(2):229-239.
27. Lochhead P, Chan AT, Nishihara R, et al. Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. *Mod Pathol*. 2015;28(1):14-29.
28. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer*. 2018;18(1):19-32.
29. Tan MP, Tot T. The sick lobe hypothesis, field cancerisation and the new era of precision breast surgery. *Gland Surg*. 2018;7(6):611-618.
30. Haffty BG, Carter D, Flynn SD, et al. Local recurrence versus new primary: clinical analysis of 82 breast relapses and potential applications for genetic fingerprinting. *Int J Radiat Oncol Biol Phys*. 1993;27(3):575-583.
31. Smith TE, Lee D, Turner BC, Carter D, Haffty BG. True recurrence vs. new primary ipsilateral breast tumor relapse: an analysis of clinical and pathologic differences and their implications in natural history, prognoses, and therapeutic management. *Int J Radiat Oncol Biol Phys*. 2000;48(5):1281-1289.
32. Nemes S, Danielsson A, Parris TZ, et al. A diagnostic algorithm to identify paired tumors with clonal origin. *Genes Chromosomes Cancer*. 2013;52(11):1007-1016.
33. Andrade VP, Ostrovnaya I, Seshan VE, et al. Clonal relatedness between lobular carcinoma in situ and synchronous malignant lesions. *Breast Cancer Res*. 2012;14(4):R103.
34. Mauguen A, Seshan VE, Ostrovnaya I, Begg CB. Estimating the probability of clonal relatedness of pairs of tumors in cancer patients. *Biometrics*. 2018;74(1):321-330.
35. Goldstein NS, Vicini FA, Hunter S, et al. Molecular clonality determination of ipsilateral recurrence of invasive breast carcinomas after breast-conserving therapy: comparison with clinical and biologic factors. *Am J Clin Pathol*. 2005;123(5):679-689.
36. McGrath S, Antonucci J, Goldstein N, et al. Long-term patterns of in-breast failure in patients with early stage breast cancer treated with breast-conserving therapy: a molecular based clonality evaluation. *Am J Clin Oncol*. 2010;33(1):17-22.
37. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;512(7513):155-160.
38. Gao R, Davis A, McDonald TO, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*. 2016;(August):1-15.
39. Wapnir IL, Dignam JJ, Fisher B, et al. 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*. 2011;103(6):478-488.
40. Donker M, Liti??re S, Werutsky G, et al. Breast-conserving treatment with or without radiotherapy in ductal carcinoma In Situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial. *J Clin Oncol*. 2013;31(32):4054-4059.
41. Lebya K, Garcia-Smith R, Swaminathan R, et al. Towards a personalized surgical margin for breast conserving surgery-Implications of field cancerization in local recurrence. *J Surg Oncol*. 2017;115(2):109-115.
42. Visser LL, Elshof LE, Schaapveld M, et al. Clinicopathological Risk Factors for an Invasive Breast Cancer Recurrence after Ductal Carcinoma In Situ-A Nested Case-Control Study. *Clin Cancer Res*. April 2018:clincanres.0201.2018.
43. Li H. Aligning sequence reads, clone

sequences and assembly contigs with BWA-MEM. March 2013.

44. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
45. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657-663.
46. Van de Wiel M, Vosse S. CGHcall: Calling aberrations for array CGH tumor profiles. 2018.
47. Quinlan AR. BEDTools : the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinforma*. 2014.
48. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47-54.
49. Pereira B, Chin S-F, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*. 2016;7(May):11479.
50. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
51. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
52. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour*. 2017;17(1):44-53.
53. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
54. Ostrovnaya I, Seshan VE, Olshen AB, Begg CB. Clonality: an R package for testing clonal relatedness of two tumors from the same patient based on their genomic profiles. *Bioinformatics*. 2011;27(12):1698-1699.

Supplementary tables and figures

Table S1. Comparison of cliniopathological charactistics between clonal pairs and independent pairs

	Clonal pairs n = 41		Independent pairs n = 23		*P*	
Age at DCIS diagnosis in years, mean (range)	58.5	(33.9-86.5)	58.0	(38.2-86.7)	0.87	a
DCIS lesion size in mm, mean (range)	15.9	(4-30)	11.75	(2-20)	0.21	a
DCIS lesion size						
1-10 mm	5	(12.2%)	4	(17.4%)		
11-20 mm	5	(12.2%)	4	(17.4%)		
>20 mm	4	(9.8%)	0	(0.0%)	0.32	b
Unknown	27	(65.9%)	15	(65.2%)		
Margin status						
Free	27	(65.9%)	15	(65.2%)		
Involved	7	(17.1%)	2	(8.7%)	0.70	b
Unknown	7	(17.1%)	6	(26.1%)		
Time to iIBC event in years, mean (range)	5.6	(0.5-15.7)	6.8	(0.5-17.0)	0.24	a
Same breast quadrant						
Yes	20	(44.4%)	12	(52.2%)		
No	5	(11.1%)	4	(17.4%)	0.72	b
Unknown	16	(35.6%)	7	(30.4%)		
Grade *						
Concordant	32	(78.0%)	12	(52.2%)		
Discordant	9	(22.0%)	11	(47.8%)	0.049	b
ER						
Concordant	37	(90.2%)	16	(69.6%)		
Discordant	2	(4.9%)	6	(26.1%)	0.021	b
NA	2	(4.9%)	1	(4.3%)		
PR						
Concordant	30	(73.2%)	12	(52.2%)		
Discordant	9	(22%)	9	(39.1%)	0.14	b
NA	2	(4.9%)	2	(8.7%)		
HER2						
Concordant	36	(87.8%)	16	(69.6%)		
Discordant	3	(7.3%)	6	(26.1%)	0.06	b
NA	2	(4.9%)	1	(4.3%)		

Table S1. Comparison of cliniopathological charactistics between clonal pairs and independent pairs

	Clonal pairs n = 41		Independent pairs n = 23		*P*	
DCIS						
Histologic grade *						
Low	30	(73.2%)	18	(78.3%)		
High	11	(26.8%)	5	(21.7%)	0.77	b
ER status						
Positive	31	(75.6%)	18	(78.3%)		
Negative	8	(19.5%)	4	(17.4%)	1.00	b
NA	2	(4.9%)	1	(4.3%)		
PR status						
Positive	20	(48.8%)	14	(60.9%)		
Negative	19	(46.3%)	7	(30.4%)	0.29	b
NA	2	(4.9%)	2	(8.7%)		
HER2 status						
Positive	17	(41.5%)	4	(17.4%)		
Negative	22	(53.7%)	18	(78.3%)	0.054	b
NA	2	(4.9%)	1	(4.3%)		
IHC subtype						
HR+ HER2-	21	(51.2%)	17	(73.9%)		
HR+ HER2+	10	(24.4%)	1	(4.3%)		
HR- HER2+	7	(17.1%)	3	(13%)		
HR- HER2-	1	(2.4%)	1	(4.3%)	0.12	b
NA	2	(4.9%)	1	(4.3%)		
IBC						
Histologic type						
NST	39	(95.1%)	21	(91.3%)		
Lobular	1	(2.4%)	0	(0.0%)		
Mixed	1	(2.4%)	2	(8.7%)	0.71	b
Status						
1	21	(51.2%)	8	(34.8%)		
2A	10	(24.4%)	6	(26.1%)		
2B	4	(9.8%)	4	(17.4%)		
3A	4	(9.8%)	1	(4.3%)		
3B	0	(0.0%)	1	(4.3%)		
4	2	(4.9%)	2	(8.7%)	0.52	b
NA	0	(0.0%)	1	(4.3%)		
Status (3)						
1	21	(51.2%)	8	(34.8%)		
Other	20	(48.8%)	14	(60.9%)	0.30	b
NA	1	(2.4%)	1	(4.3%)		
Grade *						
Low	21	(51.2%)	13	(56.5%)		
High	20	(48.8%)	10	(43.5%)	0.80	b

Table S1. Comparison of cliniopathological charactistics between clonal pairs and independent pairs

	Clonal pairs n = 41		Independent pairs n = 23		*P*	
ER status						
Positive	33	(80.5%)	16	(69.6%)		
Negative	8	(19.5%)	7	(30.4%)	0.366	b
PR status						
Positive	19	(46.3%)	12	(52.2%)		
Negative	22	(53.7%)	11	(47.8%)	0.795	b
HER2 status						
Positive	17	(41.5%)	2	(8.7%)		
Negative	24	(58.5%)	21	(91.3%)	0.009	b
IHC subtype						
HR+ HER2-	22	(53.7%)	16	(69.6%)		
HR+ HER2+	11	(26.8%)	0	(0.0%)		
HR- HER2+	6	(14.6%)	2	(8.7%)		
HR- HER2-	2	(4.9%)	5	(21.7%)	0.006	b

Abbreviations: mm, millimeter; iIBC, ipsilateral invasive breast cancer; ER, estrogen receptor; PR, progesterone receptor; NA, not assessable (NAs were not included in the analysis); HR, hormone receptor (ER and /or PR); NST, no special type.
*: Low grade group consists of grade 1 and 2, and high grade group consists of grade 3.
Comparisons between clonal and independent pairs were made by t-test ([a]) or Fisher's exact test ([b]).

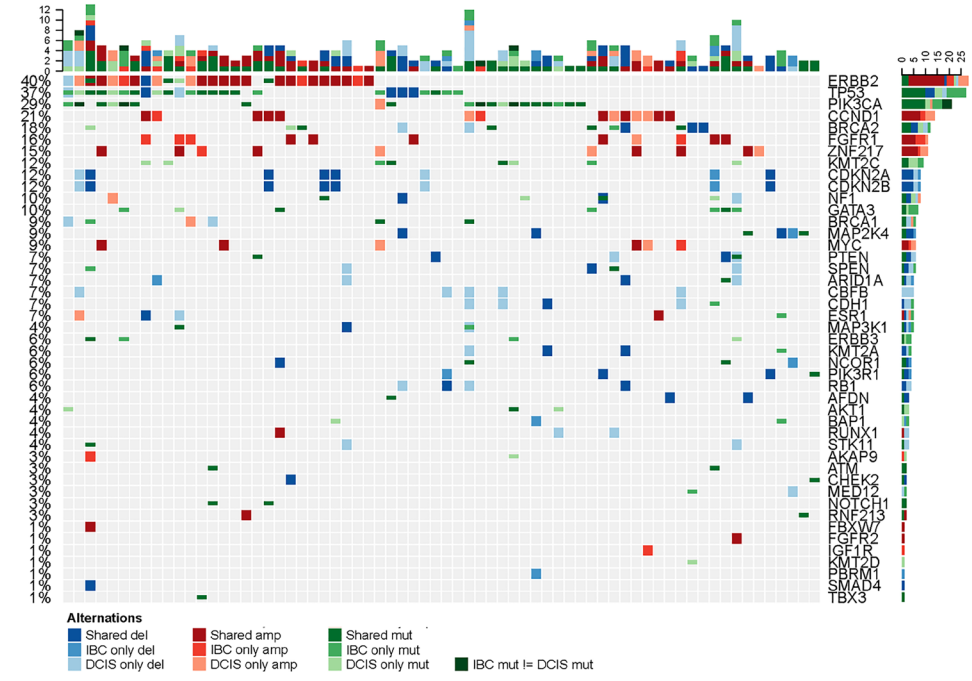

Figure S1. OncoPrint of 68 patient-matched DCIS and subsequent iIBC. Ten patients did not harbor any somatic events in the 53 assessed genes. IBC mut != DCIS mut, gene mutation in both lesions but different mutation.

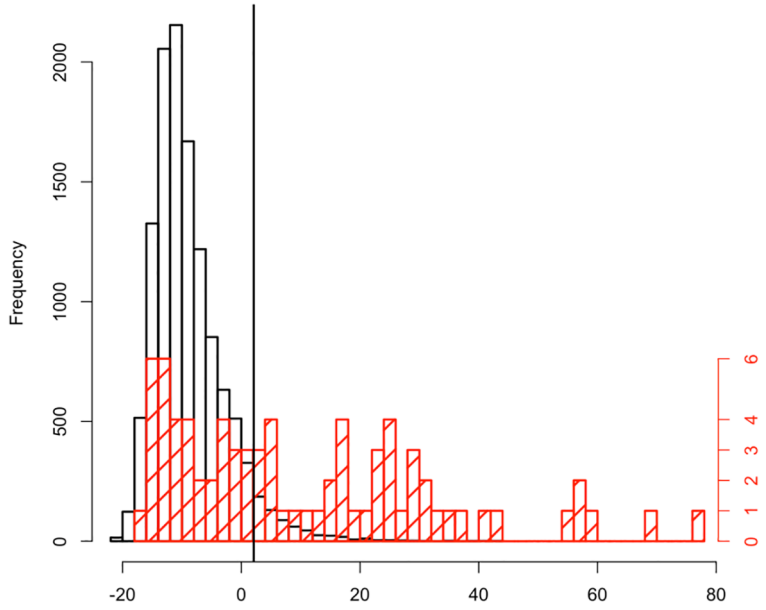

Figure S2. Distribution of the measure of clonality outputted by the Clonality package. Clonality measure of true pairs (test set) is depicted in red and artificial pairs (reference set) in black. The scale of the horizontal axis is the log-likelihood (LogLR), the measure of evidence in favor of clonality. The vertical axes represent the frequencies of pairs: left axis, black (reference set) histogram; right axis, red (test set) histogram. A cutoff for clonal origin or independence was set on p<0.1 (black vertical line).

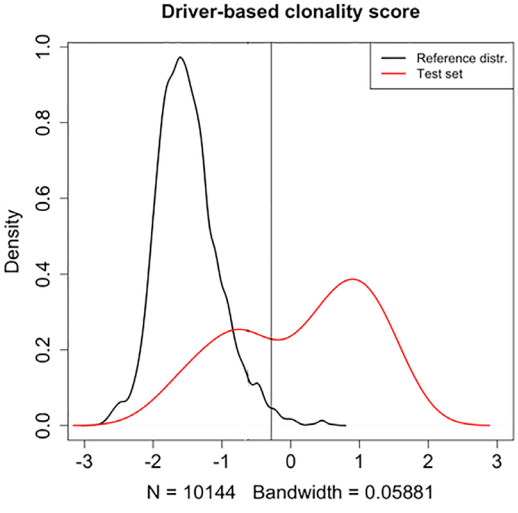

Figure S3. Distribution of the measure of clonality outputted by the panel-based clonality score. Clonality measure of true pairs (test set) is depicted in red and artificial pairs (reference set) in black. A cutoff for clonal origin and independence was set on the 99th percentile of the artificial pairs' scores (black vertical line). N=10144 is number of pairs in the reference distribution.

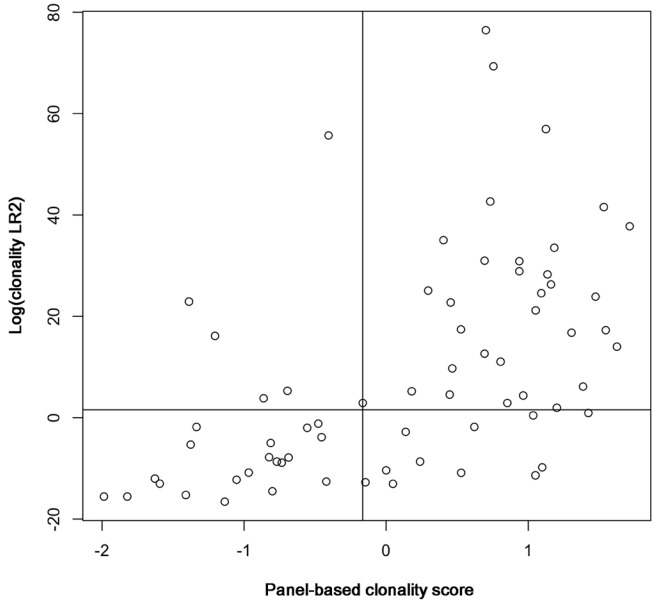

Figure S4. Correlation plot of clonality measure outputted by the clonality package (LR2 value; y-axis), based on whole genome copy number data, and the panel-based clonality score score (x-axis) which us based on copy number and mutation data, restricted to 53 breast cancer driver genes. Lines represent the used cutoffs. Pairs with a value above the cutoff are clonal and pairs with a value below the cutoff are independent. Agreement between the two methods is 79%.